Charles University

Faculty of Science

Doctoral Study Program

Molecular and Cellular Biology, Genetics and Virology

Ravindra Naraine, MSc.

Bioinformatics analysis of sequences required for localization of RNA during development

Bioinformatická analýza sekvencí pro lokalizaci RNA ve vývoji

Doctoral Thesis

Supervisor: **Radek Šindelka, Ph.D.**

Institute of Biotechnology

of the Czech Academy of Sciences

Prague, 2023

Declaration

I hereby declare that I am the author of this dissertation, all the sources and literature are properly cited and the content of this thesis or its major part was not previously used for obtaining of the same or other academic degree.

Prohlášení

Prohlašuji, že jsem doktorskou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce nebyla předložena k získání jiného nebo stejného akademického titulu.

Prague, June 15th, 2023 Ravindra Naraine

Table of Contents

Acknowledgment

I wish to express my sincerest appreciation to my supervisor, Dr. Radek Sindelka for allowing me to be part of this amazing project, and for mentoring me as a member of his prestigious research team. I thank the Laboratory of Gene Expression and in extension the Institute of Biotechnology, and all of its past and present members for welcoming me to their team and creating an environment for both continued growth and development. Thank you for the opportunity to learn, the many moments filled with laughter, and the team-building experiences that welcomed me to your team.

I would like to also thank the funding agencies for their financial support throughout the project and the model organisms that are the unsung heroes of biology.

I am grateful to my friends here and abroad who have seen me through this journey. They have been a source of solace and motivation throughout.

Most importantly I would like to thank my family for their consistent and loving support and guidance throughout the many years.

Abstract

The development of a complex organism from the fusion of two cells (oocyte and sperm) has been a fascinating aspect of developmental biology. It is now known that certain spatially and temporally regulated molecules tightly regulate embryogenesis. The asymmetrical gradient of these molecules within a given cell or within groups of cells helps to guide the differentiation of certain parts of the developing embryo. In fishes and frogs, the establishment of the maternal animal-vegetal transcript gradient within the egg produces the first developmental axis and subsequent formation of the ectoderm, mesoderm, and endoderm regions. Despite this important process, most of our knowledge on this initial animal-vegetal distribution in vertebrates has been limited primarily to the *Xenopus laevis* model, involved the analysis of only few transcripts and also analyzed only polar regions of the oocyte. This thesis aims to address this deficit in knowledge by leveraging high throughput analysis (RNA sequencing) to characterize and compare the maternal transcriptome and its sub-compartmentalization within the egg of four distantly related models. Additionally, we analyzed different stages of oocyte maturation to determine where the observed localization occurs. The models used were the African clawed frog (*Xenopus laevis*), axolotl (*Ambystoma mexicanum*), sturgeon (*Acipenser ruthenus*) and zebrafish (*Danio rerio*). We found that the animal-vegetal gradient contained subprofiles comprising of extremely animal, animal, central, vegetal, and extremely vegetal transcripts. There was very low conservation of vegetal transcripts across all models while much more were seen within the animal region. Regardless of the low vegetal transcript conservation, the enriched putative localization motifs within the vegetal transcripts were very similar and therefore point to a potentially common localization mechanism. We were able to also characterize when the transcript localization occurs during oocyte maturation. We observed both the early and late vegetal pathways, and also for the first time early and late animal pathways. Degradation and *de novo* synthesis of transcripts during the oogenesis were found. Our results point to a maternal transcript landscape that is sublocalized within the egg, dynamically altered during oocyte maturation, but show low conservation of localization across species. In general, we observed similar germ plasm related vegetally localized transcripts, cell cycle related animally localized transcripts, shared universal localization motifs and pathways. This may point to some conservation with a common ancestor or functional convergence.

Abstrakt

Vývoj komplexního organismu po spojení dvou buněk (vajíčka a spermie) je jedním z fascinujících aspektů vývojové biologie. Je známo, že embryogeneze je řízena pomocí prostorové a časové produkce molekul, které tvoří asymetrické gradienty v buňce nebo ve skupině buněk a pomáhají tak regulovat diferenciaci jednotlivých částí vyvíjejícího se embrya. Animálně-vegetativní gradienty transkriptů ve vajíčcích ryb a žab dávají vzniknout první vývojové ose, která je následně využita v tvorbě ektodermu, mezodermu a endodermu. Přes důležitost tohoto procesu je většina vědomostí o distribuci molekul podél animálně-vegetativní osy u obratlovců získána studiem modelu drápatky, založena na několika málo transkriptech a analyzovány jsou zejména póly vajíček. Tato práce doplňuje dosavadní poznatky využitím RNA sekvenování k charakterizaci a srovnání maternálního transkriptomu a jeho rozložení v rámci vajíček čtyř vzdáleně příbuzných modelových organismů - drápatky vodní (Xenopus laevis), axolotla mexického (Ambystoma mexicanum), jesetera malého (Acipenser ruthenus) a dánia pruhovaného (Danio rerio). Nalezli jsme několik animálně-vegetativních gradientů a můžeme je rozdělit do podskupin extrémně animální, animální, centrální, vegetativní a extrémně vegetativní. Existuje velmi nízká shoda u vegetativních transkriptů mezi studovanými modely, zatímco opačná situace je u animálních. Přes nízkou podobnost u vegetativní podskupiny jsou nabohacené lokalizační motivy velmi podobné a naznačují potenciálně shodný mechanismus lokalizace. Také jsme byli schopni charakterizovat lokalizaci RNA během zrání oocytu. Zjistili jsme časnou i pozdní dráhu vegetativní lokalizace a navíc jsme objevili časnou a pozdní dráhu pro animální lokalizaci. Byly zjištěny také degradace a de novo syntézy transkriptů v rámci oogeneze. Naše výsledky ukazují na distribuci maternálních transkriptů ve vajíčku, dynamické změny během zrání oocytu, avšak nízkou podobnost lokalizace u studovaných modelů. Obecně jsme pozorovali podobnosti ve vegetativní lokalizaci transkriptů důležitých pro zárodečnou plasmu, animální lokalizaci u transkriptů regulujících buněčný cyklus, sdílené lokalizační motivy a dráhy. To vše naznačuje určitou hladinu konzervace pocházející od společného předka, nebo nezávislou funkčně získanou podobnost.

List of abbreviations

ACE	after common era
Cap-Seq	chromophore-assisted proximity labeling and sequencing
clampFISH	click-amplifying fluorescence *in situ* hybridization
CLIP	cross-linking and immunoprecipitation
DREME	discriminative regular expression motif elicitation
ERPIN	easy RNA profile identification
FISH	fluorescence *in situ* hybridization
GO	gene ontology
Halo-Seq	halo-mediated localization Sequencing
ISH	*in situ* hybridization
lncRNA	long non-coding RNA
MBT	midblastula transition
MCES	mining and combining emerging substrings
MCLS	mitochondrial cloud localization element
MEME	multiple expectation maximizations for motif elicitation
MERFISH	multiplexed error-robust fluorescence *in situ* hybridization
METRO	messenger transport organizer
miRNA	microRNA
mRNA	messenger ribonucleic acid
NormQ	normalization by RT-qPCR
nt	nucleotide
PCR	polymerase chain reaction
PGCs	primordial germ cells
polyA	polyadenylic acid
polyT	polythymine
PWM	position weight matrix
RBP	RNA binding protein
RNA	ribonucleic acid

RNASeq	RNA sequencing
RNP	ribonucleoprotein
RT-qPCR	reverse transcription quantitative real-time PCR
SeqFISH	sequential Fluorescence *in situ* hybridization
smFISH	single molecule fluorescence *in situ* hybridization
smiFISH	single molecule inexpensive fluorescence *in situ* hybridization
UTR	untranslated region

1 Introduction

The asymmetrical localization of biomolecules serves as an essential mechanism that permits the occurrence of certain events within the cell. These events can range from aspects of the developmental process during embryogenesis, to mounting an impulse along nerve cells (reviewed by Palacios and Johnston, 2001). The destabilization of these asymmetrical gradients can therefore result in restrictions in these processes and in most cases be detrimental to the organism. One of the most interesting and essential examples of the use of asymmetrical biomolecules is during the process of embryogenesis. The development of organisms from an egg (single cell) to a multicellular organism is a complex process driven by both the spatial and temporal requirements of asymmetrically localized transcripts. Some organisms like the amphibians and fishes require the establishment of maternal transcript and protein gradients to help direct the development of the matured egg during embryogenesis (Elinson and del Pino, 2012; Howley and Ho, 2000). The first gradients are established during oogenesis and result in what is known as an initial animal-vegetal gradient along the egg. The establishment of such gradients can be driven either through active transportation to the site of interest or the selective spatial degradation of the transcripts (Buxbaum et al., 2014). Despite the importance of this process, not much is known about the full scale of the establishment or number of maternal transcripts that show asymmetry during the animal-vegetal gradient formation. Historically, most of the information on the developmental process as it relates to asymmetrical maternal transcripts has been studied using invertebrates like the fruit fly (*Drosophila melanogaster*) or the vertebrate frog (*Xenopus laevis*) (Gavis, 1997). As a result, very little is known about whether the observed asymmetry is conserved across the taxonomic families or genera, and whether it forms the basis for some conserved mechanism during development. Additionally, most of these studies were done before the establishment of high throughput sequencing. As a result, only a few maternal transcripts have been fully characterized. It is only recently that high throughput analysis has started to be leveraged to fully explore the contents of the egg. However, they have still been limited to only the typical models like the zebrafish (*Danio rerio*) or *Xenopus* (Cuykendall and Houston, 2010; Holler et al., 2021).

To address this deficit in knowledge, our research aimed to fully characterize the asymmetry within the matured egg of several typical and atypical model organisms: African clawed frog (*X. laevis*), axolotl (*Ambystoma mexicanum*), sturgeon (*Acipenser ruthenus*) and zebrafish (*D.*

rerio). In addition to describing these transcripts, we aimed to identify whether there are key similarities that have been conserved evolutionarily between the amphibians and fishes. Given that these transcripts must have some mechanism that establishes its gradient, we also aimed to identify any putatively enriched motifs that might serve as localization elements. Perhaps more novel, we aimed to fully characterize the spatial and temporal changes in the gradients of the maternal transcripts during different stages of the oogenesis process. Our work addresses the fundamental gradients that propel the developmental process and does so at a global scale while also addressing whether it has been evolutionary conserved. My key role during these projects was leveraging bioinformatic analysis to optimize and analyze useful information from these big datasets.

This thesis is structured to introduce the reader to what is already known about the animal-vegetal transcript gradient within the egg, its constituents, and the localization mechanisms. The majority of this information will stem from the *X. laevis* model as it was one of the most characterized of the vertebrate models historically. The reader will then be introduced to the major aims of the thesis and a summary of the publications that addressed each of these aims. Finally, the results from these publications will be discussed in relation to previous knowledge in the field.

2 Literature overview

2.1 History

The development of a new organism has been an area of intrigue as far back as during the times of Hippocrates (reviewed by Lopata, 2009). During this period, the prevailing theory on this subject came from Aristotle and Galen (~210 ACE), where they postulated that organisms were created in-uterine through the mixture of the male semen and the woman's blood/semen (reviewed by Lopata, 2009). This theory and variants of it prevailed for over 2000 years, until the 19th century when several key research propelled the discovery of the ovum, sperm, fertilization and meiosis (reviewed by Lopata, 2009). During this period two main theories arose, that of the preformist who believed that the sperm was the major factor for human development, whereby a small homunculus was already pre-formed within the sperm versus epigenesis whereby the organism developed from an initial formless state through the reshaping of the mother's vital blood and the sperm (reviewed by Speybroeck et al., 2006). It took several centuries for this theory to be challenged. The 17th century's inception of the microscope allowed for the discovery of semen, and later research in amphibians showed that fertilizing an egg with a single sperm created a fetus (reviewed by Lopata, 2009). Later Karl Ernst von Baer microscopically discovered the mammalian ovum using ovarian follicles from dogs. However, aspects of fertilization and inheritance were still unknown at this time and came with the later discoveries that included discovery of the cell, nucleus, meiotic division, and the role of the chromosome in heredity (reviewed by Lopata, 2009).

With the discovery that the egg and sperm are important for the development of a new organism, there was the question as to how these singular cells were driving such a complex process and also the organization of the body structure (reviewed by Horder, 2010; reviewed by Lopata, 2009). During the 19th century, 'recapitulation' was popularized by Ernst Haeckel whereby it was believed that the embryo goes though the ancestral adult stages during the process of embryogenesis (reviewed by Horder, 2010). The "germinal selection" theory later proposed by August Weismann, postulated that separate inherited determinants control different parts of the egg and future embryo. This was reinforced by Wilhelm Roux in 1881, where he found in frogs that only half an embryo developed after eliminating one of the cells from the two cell stage. However, this contrast from what Driesch in 1891 discovered later, whereby he found that a

complete normal embryo could be formed from the individual cells of the early zygote in sea urchins (reviewed by Horder, 2010).

Later work from Herbst and Hans Spemann helped to resolve this conflicting issue on how supposedly "unordered" cells may direct the control of embryogenesis (reviewed by Horder, 2010). Their work on embryonic induction proposed that one part of the developing embryo can induce the next part's development. On of the most pivotal moments in developmental biology came in 1924 with the discovery of the 'organiser' by Spemann (reviewed by Horder, 2010). It was a small region of the early amphibian embryo that controlled the entire emergence of the patterned differentiation throughout the embryo. However, before the establishment of the 'organiser', the dorsal mesoderm needs to be first established (reviewed by De Robertis et al., 2000). This dorsal mesoderm, in amphibians and fishes are predefined by the animal-vegetal gradient and subsequent chain reaction events following fertilization (reviewed by De Robertis et al., 2000). As a result, understanding the initial asymmetrical organization of the egg helps to better understand the subsequent chain reactions that are triggered due to this established gradient.

2.2 Benefits of a polarized cell

Transcript localization has already been described in many organisms, including the budding yeast (eg: *ASH1* mRNA) (Bertrand et al., 1998), in bacteria (eg: *lacY*) (Nevo-Dinur et al., 2011), in protist (Naegleria - eg: α-tubulin) (Han et al., 1997), neurons (eg: *Arc* mRNA) (Steward et al., 1998) and enterocytes (eg: *LPH* mRNA) (Barth et al., 1998). It is believed that polarization of transcripts may in fact be present in most, if not all cells. This is because it is more energy efficient to localize transcripts rather than their larger molecular mass protein products (reviewed by Jansen, 2001; Martin and Ephrussi, 2009). Additionally, the mRNA translation can be activated directly at the location where it is utilized (reviewed by Jansen, 2001; Martin and Ephrussi, 2009). It is also easier to associate the required transcript products to similar spatial areas for the production of supermolecular structures (reviewed by Jansen, 2001). The localization of mRNAs that code for toxic protein products can also protect the cell from death by localizing these transcripts to areas where its products may not be detrimental (reviewed by Martin and Ephrussi, 2009).

Short nucleotides present within localized transcripts may serve as binding sites to RNA binding protein (RBP) that direct it to the cytoskeletal framework, or to other proteins that either stabilize

it or induce its degradation (reviewed by Buxbaum et al., 2014). The nucleotides that direct the transcripts to the cytoskeletal network are known as localization motifs or zipcodes and have been historically identified as usually on average greater than 50 nucleotides long (Figure 1) (reviewed by Taliaferro, 2022). This differs from the usual 4 to 8 nucleotides usually found for other regulatory elements (reviewed by Taliaferro, 2022). These zipcodes are usually found in the 3'UTR (untranslated region) sequences like in the *gdf1* (previously *vg1*/*dvr1*) mRNA (Figure 2) (reviewed by Mowry and Melton, 1992). However, there are few examples of localization motifs present in the 5'UTR (eg: *neff1* (previously *nif*)) (Claußen et al., 2004) and also exonic sequences (eg: *ASH1*) (Chartrand et al., 1999) (Figure 2). Additionally, the secondary structure of the UTRs may also serve as a localization motif (eg: *bcd*) (Figure 2) (Macdonald and Kerr, 1997). Cells may utilize these motifs or similar, for either active transport to the area of interest, passive diffusion of the transcript which then becomes selectively bounded to already spatially fixed proteins, or selective degradation and stabilization in a particular area (Figure 3) (reviewed by Buxbaum et al., 2014).

The misregulation of transcript localization has serious consequences to the proper functioning of the cell (Ephrussi and Lehmann, 1992). In *Drosophila*, the misregulation of specific localized transcripts that define the anterior/posterior axis results in developmental patterning defects (Ephrussi and Lehmann, 1992). In mammals, particularly neurons, localization of transcripts is required for proper cellular functioning. The misregulation (mutations in RBPs) have been associated with neurological diseases such as fragile X syndrome (Ephrussi and Lehmann, 1992), spinal muscular atrophy (Ephrussi and Lehmann, 1992), and amyotrophic lateral sclerosis (Ephrussi and Lehmann, 1992).

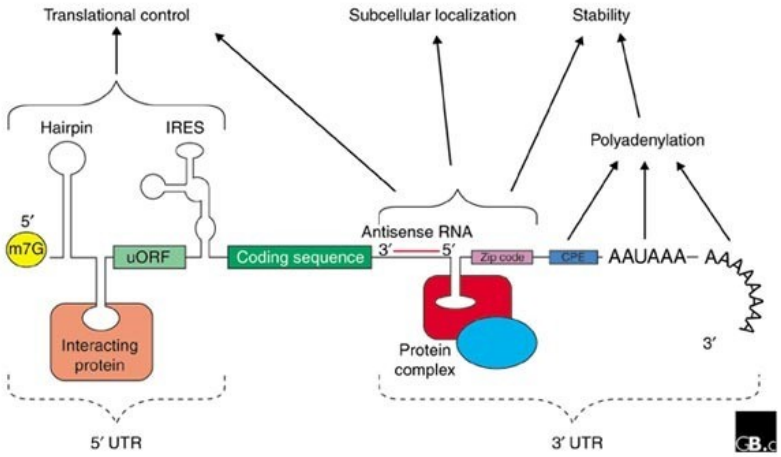


Figure 1: Common sites for the regulatory elements found in the UTR sequences of transcripts, including the motifs for localization. (From: Mignone et al., 2002)

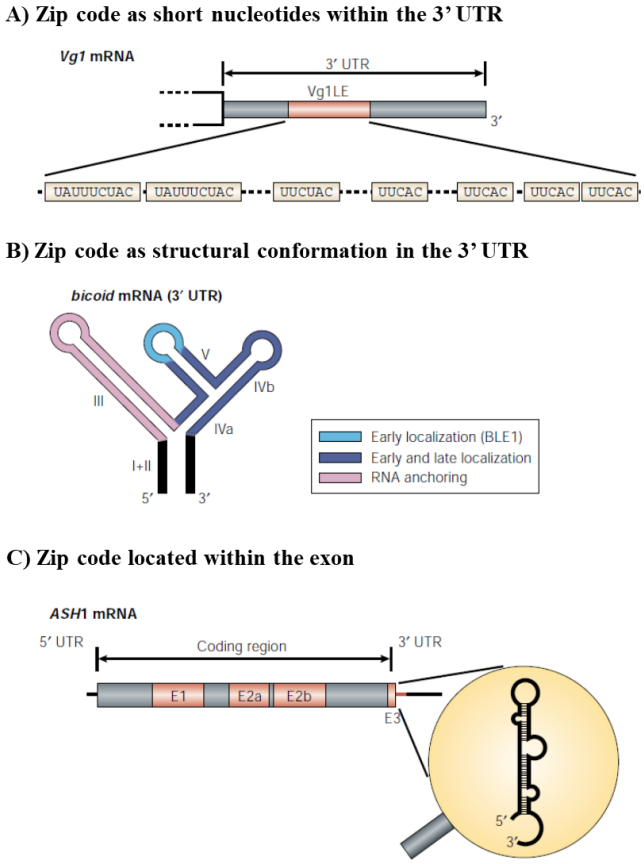


Figure 2: Examples of some of the different types of localization elements. (From: Jansen, 2001)

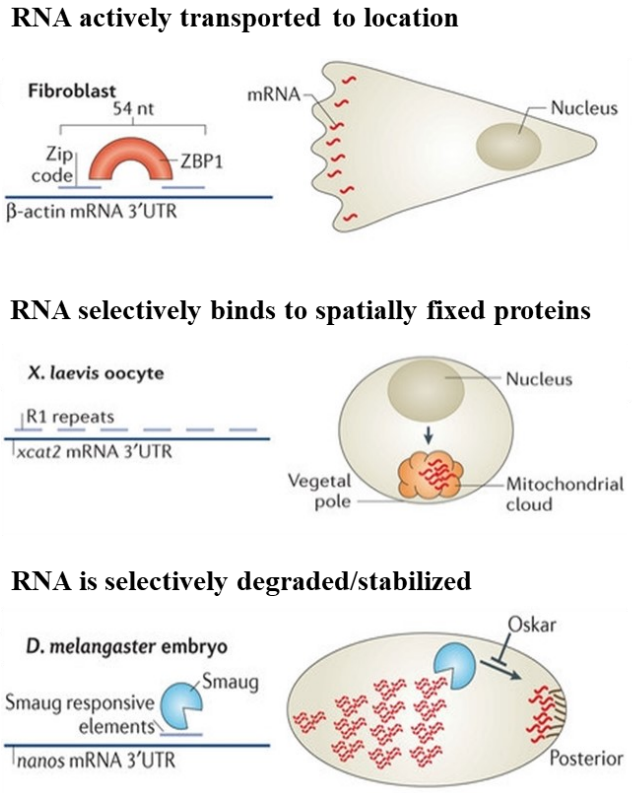


Figure 3: The different techniques used by the cell to create transcript gradients. **A)** The transcript is actively transported using localization elements found primarily in the 3'UTR sequence that binds to the RNA binding protein complex. Transportation is facilitated by the cytoskeletal framework. **B)** Spatially fixed proteins bind to homogeneously localized transcripts causing them to accumulate within a particular region. **C)** Transcripts are preferentially degraded in one region while stabilized and therefore accumulated at another. (From: Buxbaum et al., 2014)

2.3 Models to study maternal transcript localization

The two models (*X. laevis* and *D. rerio*) have become increasingly used in research (Figure 4). *Xenopus laevis* and *D. rerio* have a longer history as model organisms (reviewed by Houston, 2017). Their fast reproduction, large size and large eggs make them ideal models (reviewed by Blum and Ott, 2019; Link and Megason, 2008). In particular many of the fundamental knowledge on development has been derived from the *X. laevis* and *D. melanogaster* model (reviewed by Gavis, 1997). Our laboratory has introduced two additional models, the *A. mexicanum* and *A. ruthenus* for transcript localization analysis. The *A. mexicanum* and *A. ruthenus* are more recent models due to their tissue regenerative capacity and commercial

(caviar) potential, respectively (Romanenko et al., 2015; Voss et al., 2009). Therefore, almost nothing is known about their maternal transcriptome at a high throughput level.

These four models are phenotypically and evolutionary diverse species. The earliest ancestor shared between the fishes with fins and tetrapodomorpha was approximately 380 million years ago (Clack, 2009). This is a significant evolutionary distance separating the amphibians from the fishes and is clearly apparent in the great variations observed in the body shapes, sizes, age of maturity and genome sizes (comparison in discussion). Alteration in the genomes of these models' ancestors have also contributed to their diversity. There was a genome duplication of the ancestors of the *D. rerio* and also *A. ruthenus* (Du et al., 2020; Van De Peer et al., 2009). In the *X. laevis* there was an alloploid event whereby two progenitor species with 2n=18 genomes merged to produce the ancestor of the now allotetraploid *X. laevis* (Session et al., 2016). One of the sub-genomes underwent drastic alterations, such as changes in its histone and DNA methylation, intrachromosomal rearrangement and gene loss (Session et al., 2016). More interestingly, some of the homoeologous genes that are maternally transcribed showed possibility of neofunctionalization, and in some cases loss of expression (Session et al., 2016). The presence of these homoeologs offers a unique opportunity to study what cis-acting elements have been altered in previously related transcripts, that might relate to a localization or functional change.

The *A. mexicanum* genome represents one of the largest ever sequenced (Keinath et al., 2015). However, its size is primarily a result of repetitive elements and intronic expansion, and its protein coding genes are actually very similar to those of the other models (Keinath et al., 2015; Nowoshilow et al., 2018). Although all of these species have a clear visible animal-vegetal axis (except *D. rerio*), their cleavage pattern during development is different. *Acipenser ruthenus* and amphibian embryos have complete division of the embryos and involves the yolk region (holoblastic cleavage), while in teleosts (eg: *D. rerio*) the cellular divisions do not involve most of the yolk regions (meroblastic cleavage) (Ballard, 1981; reviewed by Bordzilovskaya and Dettlaff, 1991; reviewed by Elinson, 2009; Kimmel et al., 1995).

Such a diverse evolutionary difference may be reflected in the animal-vegetal gradient of their eggs. Previous results already observed that some localized transcripts were different between *D. rerio* and *Xenopus*, and even between *X. laevis* and its related species *X. tropicalis* (Claußen

et al., 2015; Holler et al., 2021). However, given that this process is not only essential but most likely rooted early within the ancestral lineage, it is also possible that there lies a universal convergent evolution for some aspects of the developmental process (True and Haag, 2001).

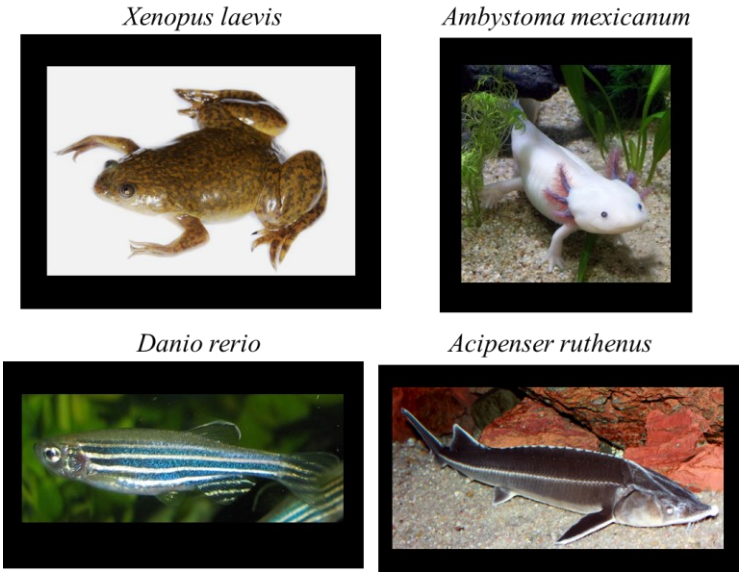


Figure 4: The models used during this thesis (not to scale). The amphibian models *Xenopus laevis*, *Ambystoma mexicanum* and the fish models *Acipenser ruthenus* and *Danio rerio*. (Images are from https://commons.wikimedia.org.)

2.3.1 Early development and transcript localization

It is now known that copious amounts of maternal transcripts and proteins are deposited within the egg that direct the cell to differentiate into multiple ordered cells (reviewed by Houston, 2017). Much of what is known about the vertebrate transcript localization along the animal-vegetal axis has been derived from the *Xenopus* model.

The initial gradient that forms the animal-vegetal axis underlies the framework from which the other major body components are derived (Figure 5**)**. These components are represented as the three primary germ layers (reviewed by Flachsova et al., 2013). The endoderm layer develops from the vegetal part of the egg and leads to the development of the gastrointestinal, respiratory, and urinary systems. The mesoderm originates from the meridial part through the interactions of vegetal and the mid-animal regions and leads to the development of the notochord, axial skeleton, cartilage, connective tissue, trunk muscles, kidneys, and blood. The ectoderm is derived from the animal part of the egg and leads to the development of the nervous system,

epidermis and various neural crest-derived tissues (reviewed by Kiecker et al., 2015; reviewed by Schnapp et al., 1997).

The animal layer which comprises of the oocyte nuclear region is easily identifiable due to the dark pigmentation of protein granules. The vegetal layer contains the yolk region and maternal transcripts that define the endoderm and interact with the transcripts along the mid-animal regions to define the mesoderm (reviewed by Elinson, 2006; Houston, 2017; Kloc, 2014; Schnapp et al., 1997). Upon fertilization, the sperm usually enters at the animal end and initiates a streaming of microtubules which causes the rotation of the vegetal cortical cytoplasm (Figure 5**)**. This results in the shift of the extreme vegetal transcripts along the animal-vegetal regions to be deposited to the dorsal mesoderm (Figure 5**)** (reviewed by Elinson, 2006; Houston, 2017; Kloc, 2014; Schnapp et al., 1997). The interactions of these transcripts with other genes help to initiate the formation of the dorsal ventral region (Figure 5**)** (reviewed by Houston, 2017; Kloc, 2014; Schnapp et al., 1997). The first twelve cell divisions post fertilization occur synchronously and does not involve *de novo* transcription (Anderson et al., 2017; Flachsova et al., 2013; reviewed by Kloc, 2014). The following cellular division signals the start of the *de novo* transcription and also the occurrence of asynchronous cellular division and represents what is regarded as the mid-blastula transition (MBT) (reviewed by Elinson, 2006; Houston, 2017; Kloc, 2014; Schnapp et al., 1997). At the 32-cell stage, early gastrulation begins, where the regions form the more complex parts of the embryo (Figure 5**)**. The initial animal-vegetal established gradient is therefore key to the establishment of the chain reaction for the later gastrulation events.

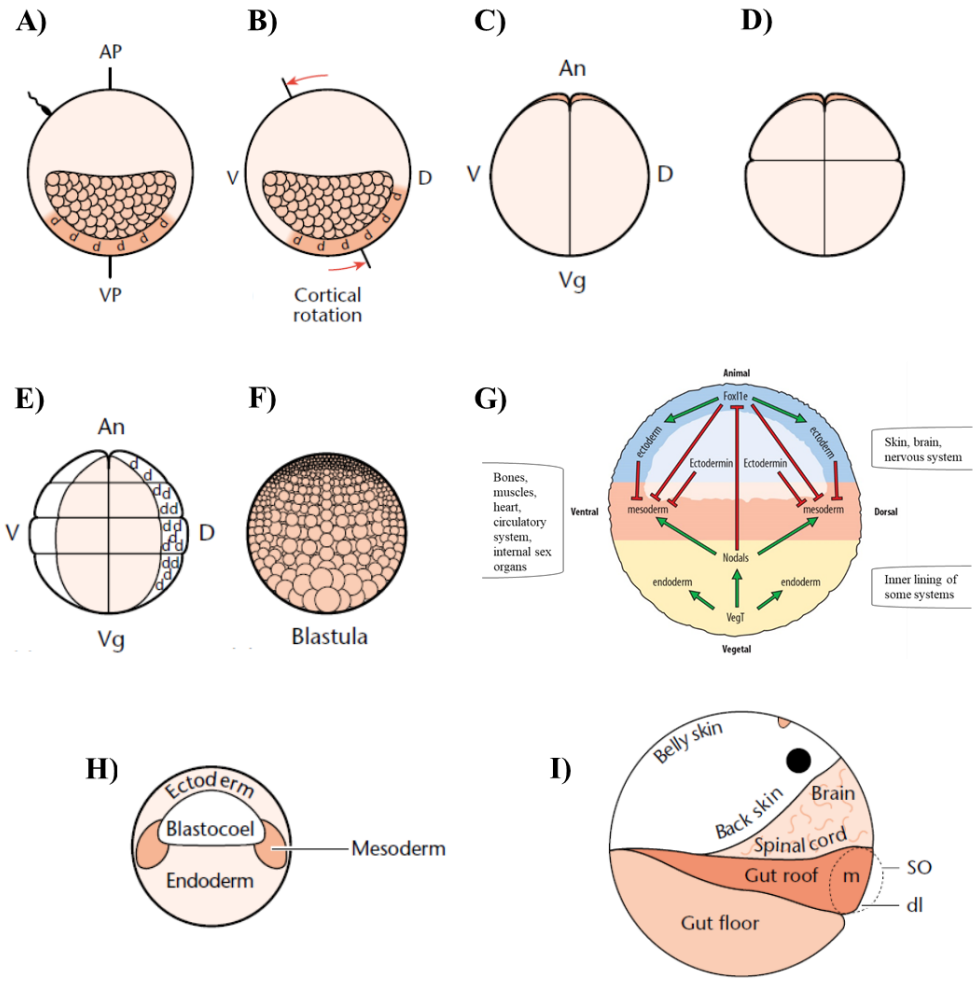


Figure 5: A) The animal-vegetal polarity of the unfertilized *Xenopus laevis* cell. **B)** cortical rotation after the fertilization. The vegetal transcripts in the cortex rotate to the dorsal axis. **C)** 4-cell and D) 8-cell development. **E)** 32-cell embryo. The eight cells along the dorsal side contain dorsal information. **F)** Blastula stage. **G)** The cell fates for the different regions of the cell. **H)** The differentiation of the three germ layers within the blastula. I) Fate maps during gastrulation. The Spemann organizer (so) is found on the animal region of the dorsal lip (dl). AP – animal pole; VP – vegetal pole; An - Animal; Vg - vegetal; D - dorsal; V - ventral. (Adapted from Elinson, 2006).

2.4 Animal and vegetal genes

Prior to the onset of high throughput sequencing, only about 36 transcripts were identified as being localized within the *X. laevis* egg. Twenty were characterized as being vegetal while 16 to the animal hemisphere (reviewed by King et al., 2005). The first 4 localized transcripts in *X. laevis* were identified only in 1985 by Rebagliati, M. R. and colleagues (Figure 6) (Rebagliati

et al., 1985). These comprised of the 3 animal genes (*zfand4* (previously *An1*), *pax6* (previously *An2*), *ddx3x* (previously *An3*)) and the vegetal gene *vg1* (Rebagliati et al., 1985). Twelve other transcripts were later discovered between 1992-1996 (Elinson et al., 1993; Hinkley et al., 1992; Houston et al., 1998; Hudson and Woodland, 1998; Hudson et al., 1996; Kloc et al., 1993, 1991; Ku and Melton, 1993; Mosquera et al., 1993; Reddy et al., 1992; Zhang and King, 1996).

The vegetal transcripts that localize early to the cortex (eg: *nanos1*-previously *xcat2*) were found to be take part mainly in germ plasm function, germ cell development and migration (reviewed by Kloc, 2014; Kloc and Etkin, 2005). Those that migrate later were found to be involved in the development of the germ layers (reviewed by Houston, 2017; Kloc, 2014; Kloc and Etkin, 2005; Schnapp et al., 1997). Several of the vegetal genes are in fact transcription factors and proteins that activate the expression of other genes (Owens et al., 2017). The animal localized genes are believed to persist in the animal blastomeres where they are involved in the embryonic germ layers and D-V axis formation, cell polarity and cell signalling (reviewed by Kloc, 2014). Tables 1, 2, 3 and 4 show the summary of some of these asymmetrical maternal transcripts in the *X. laevis* egg and their proposed functions.

Table 1: Examples of early vegetal genes and their proposed functions. (Extracted from: Kloc, 2014)

RNA	Function
nanos1 (previously *xcat2, nos1*)	translational repressor of endoderm fate mRNAs in PGCs
dazl	translational regulator that affects migration and differentiation of PGCs, relives miRNA- repression of mRNAs for PGC
ddx25 (previously deadsouth)	proper distribution of germplasm in dividing PGCs
pgat (previously xpat)	germ plasm formation; recruits mitochondria
germes	organizes and regulates function of germ plasm
otx1 (previously *xotx1*)	head development in the embryo

rbpms2 (previously *hermes*, *rbpms*)	RBP that functions during oocyte maturation. negatively regulates *spdya* (previously *ringo*)*, nanos1, mos*
wnt11 (previously *wnt11a*, *wnt11r*)	formation of the embryonic dorsal axis
trim36 (previously *haprin*)	formation of the embryonic dorsal axis
ctdspl (previously *xnif*, *scp3*)	neuron-specific *Xenopus* intermediate filament protein
cdx2 (previously *xcad2*)	organizer-specific homeobox transcription factor
grip2	involved in PGC development and migration
exd2	Germ plasm localized
fgfr2, *igfr1b*	Growth factor receptors
acvr1 (previously *alk2*), *tob2*, *smad7*	bone morphogenetic protein (BMP)
pes1 (previously *gptl1*)	Noncoding RNA
xlsirts	Noncoding RNA, cytokeratin and actin integrity within the vegetal cortex

Table 2: Example of late vegetal genes and their proposed functions (Extracted from Birsoy et al., 2006; Kloc, 2014; Krishnakumar et al., 2018)

RNA	**Function**
*gdf1 (*previously *vg1, dvr1)*	Tgfβ family member; induce anterior mesendodermal genes (Bmp and Wnt antagonists)
*Vegt (*previously *brat)*	cytokeratin and actin integrity within the vegetal cortex
velo1 (previously *Xvelo1*)	potential ortholog of *D. rerio* buc (scaffold for RBP)
spire1 (previously *pEg6*)	actin nucleation

Table 3: Example of intermediate vegetal genes and their proposed functions (Extracted from: Kloc, 2014)

RNA	Function
plin2 (previously *fatvg*, *adfp*)	Vesicular trafficking protein; proper localization of germ plasm determinants; cytokeratin and actin integrity within the vegetal cortex
dnd1 (previously *dead end*)	RBP for dorsal migration of PGCs
ddx59 (*centroid*)	Possible regulation of germ plasm-stored mRNPs

Table 4: Example of animal genes and their proposed functions (Extracted from: Kloc, 2014)

RNA	Function
CK2α	early localization, signalling during dorsal axis specification
bub3 (previously *xbub3*)	later localization; similar to mitotic checkpoint gene
adrm1 (previously *xoom*)	later localization; membrane protein
pabpc1 (previously *pabp*)	later localization; polyA binding protein
ets1 and *ets2*	later localization; transcription factors
XGβ1	later localization;
ddx3x (previously *An3*)	ubiquitous in animal hemisphere

2.5 High throughput analysis of animal-vegetal axis

The limited number of discovered localized transcripts was primarily due to the techniques used during those times, namely *in situ* hybridization, immunostaining, or RT-qPCR analysis on whole or bilaterally sections oocytes. Not only did these techniques limit the number of transcripts that could be assessed, but also the spatial resolution. To improve the spatial resolution, Šindelka and colleagues in 2008 implemented a new technique called RT-qPCR tomography (Figure 6) (Sindelka et al., 2008). In this technique, they cryo-sectioned the *X. laevis* egg into 35 sections ranging from the extreme animal pole to the extreme vegetal pole and queried via RT-qPCR the profiles of 18 known localized transcripts (Sindelka et al., 2008). They showed that there were clear sub-localization patterns where many vegetal and animal transcripts were in the animal and vegetal hemispheres respectively (Sindelka et al., 2008). This

method has served as the basis for the subsequent high throughput sequencing based methods outlined for this thesis and also for many other more recent publications.

Before the onset of sequencing based high throughput methods, microarrays allowed for the assessment of more transcripts within the animal and vegetal regions (Figure 6). However, they were limited by the number of probes that could be assayed for the transcriptome. Cuykendall and Houston in 2010 utilizes this method to assess the vegetal cortices from stage VI oocytes (Cuykendall and Houston, 2010). They identified 275 RNAs with increased (greater than 2.9-fold) abundance in the vegetal cortex versus the whole oocyte. A large portion of these genes (70%) had no known functions. The remaining represented members with RNA binding potential, zinc fingers, transcription factors, maternal signalling events, bone morphogenetic protein signalling pathway, metabolic proteins, and endoplasmic reticulum (ER) function. Therefore, the vegetal cortex was even more polarized than previously thought and it represented a region with a potentially high metabolic activity (Cuykendall and Houston, 2010).

Only six major publications (three of which originates from this thesis) using high throughput RNASeq (RNA sequencing) techniques for assessing of the maternal transcript in the *X. laevis* model can be found since the first publication in the year 2015 (Figure 6) (Bezawork-Geleta et al., 2018; Claußen et al., 2015; Holler et al., 2021; Iegorova et al., 2022; Naraine et al., 2022; Owens et al., 2017). During the start of this thesis there was only one available publication, with the others published closer to its completion (Figure 6).

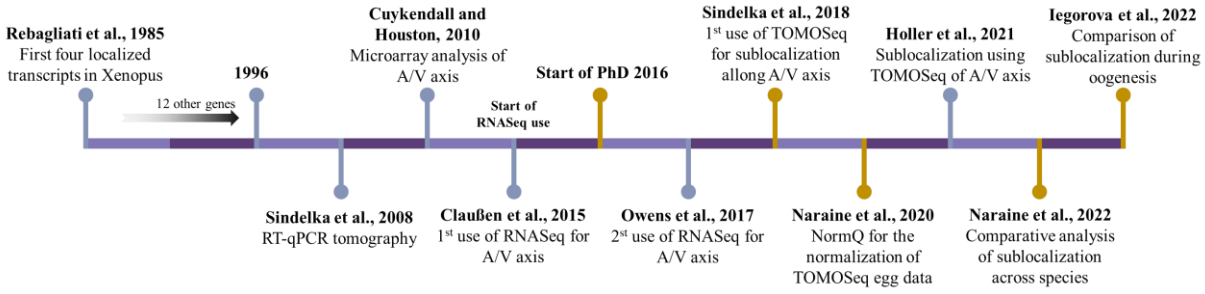


Figure 6: Basic timeline of the assessment of transcripts within the animal-vegetal axis of the *Xenopus laevis* with particular focus on the use of high throughput sequencing.

The first major high throughput RNASeq analysis on the *X. laevis* maternal transcript asymmetry within the egg was carried out by Claußen and colleagues in 2015 (Claußen et al., 2015). They analysed the animal and vegetal halves of the *X. laevis* and *X. tropicalis* eggs to

determine transcript enrichments within each section, and their research represents one of the first usage of next generation sequencing for analysis of the animal-vegetal axis. For the *X. laevis*, they were able to identify an additional 324 vegetally enriched genes with 2-fold enrichment relative to the animal sphere (Claußen et al., 2015). More uniquely they were able to identify a large 338 number of transcripts with 2-fold enrichment in the animal hemisphere. Many of the previously described animal genes were however found at a lower fold change. Some of the animal localized transcripts they confirmed using qPCR and also *in situ* hybridization in staged *X. laevis* oocytes using cloned cDNA fragments of randomly selected animal transcripts. Only half of the vegetal transcripts of the *X. tropicalis* was shared with those from the *X. laevis*, with also similar low concordance for the animal genes. However, the genes that were vegetally conserved had enrichment of GO terms associated with germ cell development, regulation of gene expression, and vesicular and lipid transport processes (Claußen et al., 2015). The animally conserved transcripts showed only one main enriched biological process, "response to drug". They also observed that there was a low overlap in the number of enriched genes between the *X. laevis* and *X. tropicalis* oocytes. They postulated that this weak level of conservation may be in line with the hourglass model for embryogenesis, whereby more significant alterations in gene expression is permitted before and after the critical period of development during mid-embryogenesis (Claußen et al., 2015; Kalinka and Tomancak, 2012). They also postulated that functional redundancy of localized transcripts may allow for the developmental divergence between species. Another suggestion is that the required gene may be restricted to its intended location at some point during embryogenesis even if not at the initial animal-vegetal axis (Claußen et al., 2015).

Owens and colleagues in 2017 also performed next generation RNA sequencing analysis of the animal and vegetal polar regions from *X. laevis* egg (Figure 6) (Owens et al., 2017). They found 5717 differentially expressed transcripts (q-value < 0.05) of which at a 4-fold difference, 411 were vegetal and at 10-fold difference 27 were animal localized. The vegetal genes were functionally enriched for developmental processes, signalling regulation, localization, phosphate metabolic processes, cellular protein metabolic processes, cell cycle, and gamete generation. Gene regulatory network analysis of the vegetal genes identified 47 genes that formed a direct interaction network with four major transcription factor hubs, *e2f1*, *irf8* (previously *icsbp1*), *esrra* (previously *nr3b1*), *ep300* (previously *p300*). The target genes are

involved in protein-modifying enzymes, receptors, ligands, and transcription factors and co-factors with known previous associations in regulating the cell cycle (*e2f1*), endoderm specification (*sox7*), metabolic pathways (*esrra*) and lineage commitment (*irf8*). They also analysed for miRNAs (microRNAs) and found eight, but none showed any asymmetry. However, a potential of 13 vegetally enriched targets were found for 7/8 of these miRNAs. They postulated that any regulation of embryonic patterning by miRNAs are potentially due to their targeting of these vegetal localized transcripts. They also queried published 8-cell data and found that 27 of their observed vegetal transcripts remained vegetally enriched even after fertilization and cortical rotation. The identified 15 previously unidentified animal transcripts were involved in signalling (*dand5* (previously *cerl2*), *ifrd2*, *slc18a2* (previously *vmat2*), *spata13*, *acaca*, *tmem192*, *ssr1* (previously *trapa*), *prr11*), gene expression (*pou2f1* (previously *oct1*)), cell division (*rmdn3* (previously *fam82c*)) and metabolism (*adpgk*, *prrg4*, *prkag1*). However, unlike the vegetal transcripts, no gene regulator networks were identifiable (Owens et al., 2017).

The previous high throughput research showed that there were even more maternal localized transcripts than previously discovered. However, they were still limited to two sectional analysis and in most cases the polar regions only. Our research was the first to assess the potential for even more subcellular localization. Recent research by Holler et al., 2021, also looked at subcellular localization within the *X. laevis* and *D. rerio* egg.

2.6 Germ plasms

The gametes originate from the development of primordial germ cells (PGCs) (Kloc et al., 1998). The region for the development of these specificized cells are usually already predefined at the oocyte level as vegetally deposited transcripts, usually using the messenger transport organizer (METRO) pathway (Kloc et al., 1998). This conservation of the germ plasm is also apparent in its morphological similarity between species (Beams and Kessel, 1974; Cuykendall and Houston, 2010). However, in mammalian and urodele (eg: *A. mexicanum*) embryos, PGCs are created through inductive signalling rather than predefined asymmetrical gradients (Cuykendall and Houston, 2010). Additionally other anurans, like the *R. pipiens* and *E. coqui* oocytes appear to lack a mitochondrial cloud which is usually associated with the germ plasm (Elinson et al., 2011; Nath et al., 2005). However, the proteins involved in PGC formation, like Vasa, appear to be conserved and important in the development of these regions (Cuykendall

and Houston, 2010). Therefore, regardless of the mode of accumulation, the similar determinants are required at the given region for correct development (Cuykendall and Houston, 2010).

In vertebrates, the germ plasm has been extensively studied in *X. laevis*. The germ plasm determinants include those coding RBPs: *nanos1*, *dazl*, *ddx25*, *dnd1* and *rbpms2*; and those encoding novel proteins and lip droplet associated proteins such as *pgat*, *germes* and *plin2* (Table 1) (Cuykendall and Houston, 2010). Owens et al., 2017 found a total of 17 germline components of which *pgat*, *sybu*, *otx1*, *tob2* and *efnb1* were identified previously by Cuykendall and Houston, 2010 and De Domenico et al., 2015. These PGCs migrate to the dorsal side during cortical rotation following the fertilization of the egg in *X. laevis*. They become transcriptionally active at the gastrula stage, with some still persisting all the way to the tailbud stage (De Domenico et al., 2015; Owens et al., 2017).

2.7 Pathways and mechanisms of RNA localization

Vertebrate localization of transcripts that make up the animal-vegetal gradient has been done primarily using the *Xenopus* model (reviewed by Houston, 2017). Using this model, the study of a few localized transcripts based on *in situ* hybridization, has resulted in the identification of two major pathways (the early and late pathways) (Figure 7, 8) (Forristall et al., 1995; Kloc and Etkin, 1995). Early pathways represent mRNAs that are localized to the mitochondrial cloud/Balbiani body early during oogenesis and follow the general localization of the germ plasm (Figure 7). This pathway is known as the METRO or messenger transport organizer pathway (Kloc and Etkin, 1995). It usually results, through the passive diffusion and entrapment of transcripts to the mitochondrial cloud (Figure 7) (Chang et al., 2004). The late pathway RNAs do not localize to the mitochondrial cloud but instead to the broader vegetal cortex later during oogenesis (Figure 8). These transcripts are usually actively transported using some localization machinery and utilize dynein and kinesin motor proteins for localization (Betley et al., 2004; Claußen et al., 2015; Messitt et al., 2008). Most of the animally localized transcripts have not been investigated thoroughly and are thought to primarily become animally localized due to passive accumulation (Drummond et al., 1985; Kloc, 2014; Rand and Yisraeli, 2001).

Many components, including cis-acting localization elements, trans-acting proteins, and different cellular organelles and cytoskeletal systems involved in localizing early and late pathway RNAs have been identified for some of the localized transcripts (reviewed by Houston,

2013). However, although clusters of short nucleotide repeats (cis-acting elements) have been identified that interact with RBPs and have some effects on localization, no unique sequences have been identified that by itself determines the localization pattern (reviewed by Houston, 2013).

2.7.1 Early pathway RNAs

The first signs of asymmetry within the *X. laevis* can already be observed at stage I of oogenesis (Figure 7). Here a spherical structure, known as the Balbiani body or mitochondrial cloud, which contains mitochondria, Golgi complex, endoplasmic reticulum, lipids and pigment granules, localize near to the oocyte's nucleus (Figure 7) (Heasman et al., 1984; reviewed by Kloc et al., 2004). Early pathway mRNAs localize in the mitochondrial cloud during the early stages and then utilize the METRO pathway to migrate to the vegetal pole by stages II-IV (Figure 7). These transcripts include, *nanos1*, *dazl*, *pgat*, *Xlsirts* and *ddx25* (Table 1) (Forristall et al., 1995; reviewed by King et al., 1999; Kloc et al., 1993; Kloc and Etkin, 1995).

The localization of these early pathway genes is a bit complex as there is no clear consensus as to the components that facilitate the process (reviewed by Houston, 2013). Furthermore, it was found that the early pathway mRNAs can also colocalize by the late pathway if injected in stage III oocytes (reviewed by Houston, 2013). In some genes, like *nanos1* a 240 nucleotide 3'UTR element was discovered that is required for the localization to the mitochondrial cloud (Chang et al., 2004; reviewed by Houston, 2013; Zhou and King, 1996). However this UGCAC repeated elements is not sufficient alone for localization (Betley et al., 2002; reviewed by Houston, 2013). Complicating matters, is that the *nanos1* localization element and the late vegetal RNA *gdf1* localization element bind to the same set of proteins, including Igf2bp3 (previously Vg1RBP, Vera) and Ptbp1 (previously VgRBP60, hnRNPI) (Chang et al., 2004; reviewed by Houston, 2013). Additionally, other germ plasm transcripts (eg: *Xlsirts* and *dazl*), lack similar motifs in the Mitochondrial cloud localization element (MCLS) but yet localize similarly as *nanos1* (Betley et al., 2002; reviewed by Houston, 2013). *Xlsirt* however does utilizes repeat elements (80 nt long) for localization to the Mitochondrial cloud (reviewed by Houston, 2013).

Some of these early transcripts (*nanos1*, *pgat*, *ddx25*) migrate along with RNA and protein rich structures known as germinal granules (reviewed by Houston, 2013; Kloc et al., 2002). Their association and migration are usually facilitated by a germinal granule localization element that is adjacent to the MCLS. Specific motifs for granule localization has not been detected as yet

(reviewed by Houston, 2013). The migration of the early transcripts is not restricted by the depolymerization of the microtubules or microfilaments, indicating that it is anchored and translocated using some other mechanism (reviewed by Houston, 2013). However, these cytoskeletal structures are important in ensuring the localization fidelity of the transcripts once they have reached the vegetal cortex region (reviewed by Houston, 2013; Kloc and Etkin, 1995).

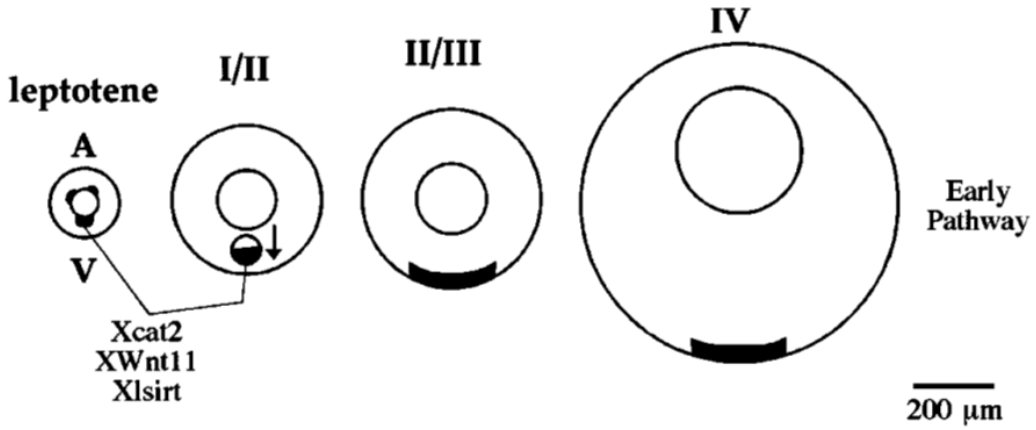


Figure 7: Schematic of the early vegetal pathway. Early localized transcripts become entrapped in the mitochondrial cloud or germinal granules. The mitochondrial cloud then migrates to the vegetal region where the transcripts then become secured to the vegetal cortex via microtubules or microfilaments. The example genes include *nanos1* (previously *Xcat2*), *wnt11b* (previously *XWnt11*) and *Xlsirt*. (From: Schnapp et al., 1997)

2.7.2 Late pathway RNAs

The late pathway is characterized by transcripts that localize to the vegetal region later during oogenesis, particularly in stage III-IV (Figure 8) (reviewed by Rand and Yisraeli, 2001). Historically about nine transcripts have been characterized as using the late pathway (Table 2). These include *gdf1*, *vegt*, *ldlrap1*, *zfp36l2*, *bicc1*, and *velo1* which encodes a protein similar to *D. rerio* Buc (Table 2) (reviewed by Houston, 2013).

The mRNA *gdf1* was the first transcript to be characterized as using the late pathway and serves as an example of the mechanism utilized. At stage I, it shows no localization pattern and is instead uniformly distributed throughout the cytoplasm. By the late stage II/III, it can be found within a wedge-shaped domain, stretching from the germ vesicle to the vegetal cortex (Figure 8) (Deshler et al., 1997; Forristall et al., 1995; reviewed by Houston, 2013; Kloc and Etkin, 1995; Melton, 1988). Here the gdf1 is colocalized with the endoplasmic reticulum (ER) where

it then later translocated and becomes anchored to the vegetal cortex during stages III-IV (Alarcón and Elinson, 2001; reviewed by Houston, 2013; Yisraeli et al., 1990). The anchoring of the gdf1 is believed to be facilitated by cortical cytokeratin filaments and cortical membranous elements (Alarcón and Elinson, 2001; reviewed by Houston, 2013; Pondel and King, 1988).

The initial accumulation of *gdf1* to the ER-wedge is not dependent on the cytoskeleton. However, microtubule and microtubule motors are required during the later transportations during stages III/IV (reviewed by Houston, 2013; Yisraeli et al., 1990). The association of the mRNAs with the motor protein and with the ribonucleic proteins might be facilitated by localization elements. In *gdf1*, two short but repeated motifs up to 340 nt UUUCU (VM1) and UUCAC (E2) have been associated with its localization to the vegetal cortex (Deshler et al., 1997; Gautreau et al., 1997; reviewed by Houston, 2013; Lewis et al., 2004).

The trans-acting proteins that interact with these localization elements have also been determined. The Insulin-like growth factor 2 mRNA-binding protein 3 (Igf2bp3) is associated with the endoplasmic reticulum and binds to UUCAC motifs in the *gdf1* localization element (Deshler et al., 1998, 1997; reviewed by Houston, 2013). The presence of UUCAC motifs have also been observed in many species of chordates and may indicate a commonly shared mechanism for localization (Betley et al., 2002). Other proteins associated with the localization include Polypyrimidine tract binding protein 1 (Ptbp1) which bind to the UUUCU motifs, and also the proteins Hnrnpab (previously 40LoVe), Dazap1/Prrp, and Staufen homolog 1 (Stau1) (reviewed by Houston, 2013). Interestingly both Hnrnpab and Dazap1 also bind to animally enriched transcripts (e.g., *zfand4*) (reviewed by Houston, 2013; Kroll et al., 2009; Zhao et al., 2001).

The *vegt* is another late pathway RNA (reviewed by Houston, 2013). Its 3'UTR sequence contains a cluster of UUCAC rich motifs (300 nt long) that is important for localization and interact with Igf2bp3 (reviewed by Houston, 2013). Other associated localized RNA-binding proteins include Dazap1, Khsrp (previously Fbp2), and Hnrnpab. The other late pathway RNAs *ldlrap1* (previously *xarh-alpha*), and *zfp36l2* also contain clusters of CAC-containing motifs (Betley et al., 2002; Zhou et al., 2003), while *bicc1* has both UUUCU and UUCAC motifs (reviewed by Houston, 2013).

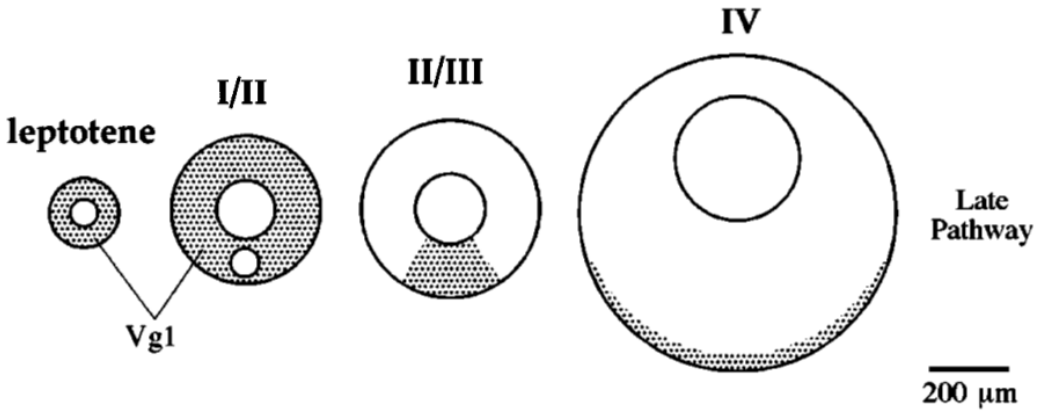


Figure 8: Schematic of the late vegetal pathway. Late localized transcripts are initially homogeneous within the egg. However, they become bounded to RNA binding protein complexes that direct them to the cytoskeletal pathway for active transport towards the vegetal pole. They form an initial wedge shape, followed by a later dispersion along vegetal hemisphere. Example RNA *gdf1* (previously *vg1*). (From: Rand and Yisraeli, 2001)

2.7.3 Intermediate/dual-pathway RNAs

The intermediate pathway represents mRNAs that utilize a combination of the METRO and late pathway (reviewed by Houston, 2013). They usually enter the Balbiani body in stage II oocytes and localize more broadly in the vegetal cortex during stages IV-V (reviewed by Houston, 2013). These transcripts include *plin2*, *dnd1*, *grip2*, *trim36*, *rbpms*/*rbpms2* (Table 3) (reviewed by Houston, 2013). Only a few of these genes have been investigated in detail (*plin2*, *dnd1*, *grip2*), with a single 25 nt localization element found for plin2 that is absent in other mRNAs (Chan et al., 1999; reviewed by Houston, 2013).

2.7.4 Animal localization

No localization mechanisms have been determined for the animally localized transcripts (reviewed by Houston, 2013). However, four distinct localization patterns have been observed for some of the animal transcripts to the animal hemisphere region (reviewed by Rand and Yisraeli, 2001). Some transcripts like the *zfand4*, *pax6*, *ddx3x* and *XGβ1* can be found throughout the animal region (Figure 9) (Devic et al., 1996; Rebagliati et al., 1985). The RNAs *pabpc1* and *igf2bp3* can be found in the animal pole cortex (Schroeder and Yost, 1996; Zhang et al., 1999), while *fn1* (previously *fibronectin*) is localized perinuclearly (Oberman and Yisraeli, 1995; reviewed by Rand and Yisraeli, 2001). Lastly, *sptan1* (previously *alpha-spectrin*) is localized in the germ vesicle and along the animal cortex (Carotenuto et al., 2000;

reviewed by Rand and Yisraeli, 2001). Although the mechanism for animal localization is not known, based on the sudden drop of *igf2bp3* RNAs at the vegetal pole during localization at the animal pole, there is an indication that animal localized mRNA might be preferentially stabilized at the animal pole while degraded at the vegetal pole (reviewed by Rand and Yisraeli, 2001). Additionally, localization signals have been found in the coding and non-coding regions of the animal localized transcript *CK2α* mRNA, pointing to a possible active localization (reviewed by Kloc, 2014).

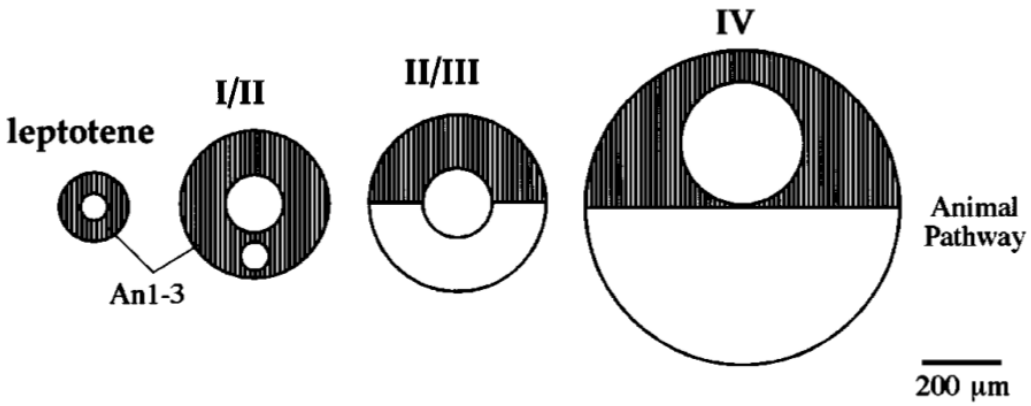


Figure 9: Schematic of the migration of the animal transcripts, *zfand4* (previously *An1*), *pax6* (previously *An2*) and *ddx3x* (previously *An3*). These transcripts are originally homogeneous within the egg. However, during the later stages, they become accumulated within the whole animal hemisphere. (From: Schnapp et al., 1997)

2.7.5 Localization in other models

Similarities in maternal transcript asymmetry have been observed and described previously. The genes *dazl* and *ddx25* are vegetally localized to the germ plasm in *Eleutherodactylus coqui* and *X. laevis* (Elinson et al., 2011). Additionally, in both *Xenopus* and *Rana*, the early (*vegt*) and late (*dazl*) pathway mRNAs are localized to the vegetal cortex, indicating that these two distinct phylogenetic groups may have had a common ancestor with this observed transcript localization pattern (200 MYA ago) (reviewed by Houston, 2013). However, contrastingly the late vegetal pathway genes *vegt* and *gdf1* are animally localized in *E. coqui* (Beckham et al., 2003; reviewed by Houston, 2013). This contrast is particularly interesting as the *Eleutherodactylus* diverged from *Rana* within the *Neobatrachia* clade, about 150 MYA and therefore indicates the altered localization was a more recent event (Elinson et al., 2011; reviewed by Houston, 2013).

Not much is known about the maternal transcripts in urodeles, of which *A. mexicanum* is a member. These organisms have been found to lack the germ plasm and instead define their PGCs through somatic induction, similar to mammals (reviewed by Elinson and del Pino, 2012; reveiewed by Houston, 2013). However, both *dazl* and *vegt* have been observed in *A. mexicanum*, but neither are localized (Bachvarova et al., 2004; reviewed by Elinson and del Pino, 2012; Nath and Elinson, 2007). As we go further along the evolutionary tree, to the fishes, there are also differences observed. The *gdf1* mRNA in *D. rerio* becomes more animally localized later in the oogenesis while it is vegetal in *X. laevis* (Bally-Cuif et al., 1998; reviewed by Pelegri, 2003). However, *dazl* is localized vegetally, similar to *X. laevis* (Maegawa et al., 1999; reviewed by Pelegri, 2003). However, the requirement of similar localized products to the dorsal region for the establishment of the Spemann organizer in *X. laevis* and *D. rerio* indicate that despite the initial differences in key localized transcripts, that the required products are translocated to their intended conserved sites at some point (De Robertis et al., 2000).

In *D. rerio* most of the mRNAs are ubiquitously distributed, but become animally enriched following egg activation and cytoplasmic streaming (reviewed by Houston, 2013; Howley and Ho, 2000). There exist however three main classes of localized RNAs during oogenesis (reviewed by Houston, 2013). One class is uniform during early oocytes but become animally localized during stage II-IV oocytes and is believed to be analogous to the late pathway in *X. laevis* (reviewed by Houston, 2013). The second group, like the early pathway of the *X. laevis*, localize initially using the mitochondrial cloud during stage I. However, it later relocalizes to the animal cortical band of the matured egg. The third category of RNAs also utilizes the mitochondrial cloud and localizes and anchors to the vegetal pole. The localization mechanisms are not as well studied as those in *Xenopus*. Cis-acting elements in the 3'UTR region of *dazl* (third category RNA), were however found that mediated localization to the mitochondrial cloud, vegetal cortex and cleavage furrow of the germplasm (reviewed by Houston, 2013; Kosaka et al., 2007).

In mammals, there does not appear to be a clear asymmetrical inheritance, but instead regulative interactions that control embryonic development (reviewed by Houston, 2013). In mouse the three main tissue types are instead established through position and cell polarity dependent mechanisms. Although there also exist a mitochondrial cloud in the oocyte of the mouse, it is dispersed symmetrically within the oocyte and does not appear to have a role in cell polarity.

RNA localization however still plays a role during embryogenesis in mammalian development. The earliest role is the specification of the trophectoderm in the blastocyte. The accumulation of caudal type homeobox 2 (*Cdx2*) mRNAs to the future apical side of 8 to 16-cell blastomeres is believed to help establish the trophectoderm fate (reviewed by Houston, 2013). Not much is known about the maternal transcript asymmetry within human oocytes. Similar high throughput analysis at the subcellular level as was done on *X. laevis* is difficult for mammalian models due to their very small sizes and also lack of animal pigmentation.

2.8 Methods used to assess subcellular localization

The accurate assessment of the transcriptome at the single cell level remains a methodological and computationally daunting task. As such it is not surprising that surveying the transcriptome at an even greater resolution, that is at the subcellular level, is a difficult and usually very low throughput endeavor. Historically, the assessment of transcripts within the egg was limited to just a few transcripts and only assessed two sections, the animal and vegetal cortex regions of large eggs. However, recent technological developments have allowed for the increased resolution at the subcellular level, allowing for several sections and also the assessment of the whole transcriptome. Such tools have revealed the large magnitude of transcripts that are localized within the cell and how essential it appears for the proper functioning of the cell. The current approaches to assess subcellular localization can be divided into either imagining or sequence based (Fazal and Chang, 2019; Taliaferro, 2022, 2019). However, many of these techniques have yet to be utilized or adapted for the assessment of localization within the egg.

2.8.1 Image based approaches

The traditional approach to identify localized genes was based primarily on hybridization-based techniques and differential display PCR (Hudson et al., 1996; reviewed by Taliaferro, 2019). The early ISH (*in situ* hybridization) approaches utilized long probes that were detected through autoradiography or immunohistochemistry (reviewed by Taliaferro, 2019). However, their signals were very weak and dispersed and were more suitable for the assessment of expression within tissues. The later development of smFISH (single molecule FISH) and its derivatives (smiFISH - single molecule inexpensive FISH, clampFISH - Click-amplifying FISH), which utilized hybridization probes combined with fluorescence microscopy, greatly improved the resolution of detection (Raj et al., 2008; reviewed by Taliaferro, 2019). However, the necessity of multiple transcript specific probes makes this method expensive when assessing multiple

transcripts (reviewed by Taliaferro, 2019). More recently approaches that utilized the MS2 RNA-hairpin system have become a common method for the assessment of mRNA spatial distribution (reviewed by Fazal and Chang, 2019; Taliaferro, 2019; Wu and Jaffrey, 2020). The mRNAs are tagged with multiple MS2 hairpins, which then recruit a fusion protein that contains a fluorescent protein (reviewed by Wu and Jaffrey, 2020). The aggregated signal from the fluorescent protein allows for the detection of the mRNA at the subcellular level. This method and similar approaches have been used to assess localized transcripts and its translation (reviewed by Fazal and Chang, 2019). Another approach utilizes RNA aptamers whose specific secondary conformation binds to a fluorescent molecule (Paige et al., 2011; reviewed by Taliaferro, 2019). However, these conformations are not as stable as the M2 hairpin and may also alter the processing of the RNA (reviewed by Taliaferro, 2019). Most recently CRISPR-related approaches that utilize Cas9 and Cas13a bounded to fluorescent protein and utilize guide RNA specific to the transcript of interest have been used to assess RNAs. However, these methods are limited to one single fluorescent molecule per Cas protein and therefore their approach in single molecule imaging is still lacking (reviewed by Taliaferro, 2019).

These previous image-based methods are limited to the assessment of just a few transcripts. However, more high throughput image methods have since been developed that utilize sequential FISH along with super microscopy (MERFISH - Multiplexed Error Robust Fluorescence in situ Hybridization, SeqFISH - sequential Fluorescence In Situ Hybridization) (Fazal and Chang, 2019; Lubeck et al., 2014; Taliaferro, 2022, 2019; Xia et al., 2019). These methods utilize multiple rounds of hybridization of DNA oligonucleotides to the complementary RNAs of interest (reviewed by Fazal and Chang, 2019; Taliaferro, 2022, 2019). Each round of hybridization is marked by a different fluorescent probe and the assessment of a transcript can be read out based on the fluorescent sequential code used to mark it. The number of transcripts that can be assessed is limited by the number of combinations of fluorophores that can be achieved (Fazal and Chang, 2019; Taliaferro, 2022, 2019).

2.8.2 Sequence-based approaches

The development of high throughput sequencing has allowed for the assessment of the full breath of expressed transcripts within cells and tissues. Its utilization on tissues and single cells has been well established (reviewed by Jovic et al., 2022). However, to allow for subcellular resolution, there requires the need for some fractionation of the cell to allow spatial

identification. Given how small the cell is and the limited resolution and contents capable of being extracted, this method has been limited primarily to eggs and cells of large sizes, eg: *X. laevis* eggs or neuronal cells. Many of these newly developed high throughput methods have unfortunately not been adapted to the egg as yet. However, given their unprecedented resolution, their future adaptation will most likely provide an unprecedented spatial awareness of maternal transcript asymmetry.

The initial subcellular demarcation can be achieved in several ways including either mechanical or biochemical fractioning of the cell (reviewed by Fazal and Chang, 2019; Taliaferro, 2022, 2019). These include, cellular fractionation using tomography, *in situ* sequencing using patterned arrays, cellular fractionation using microporous membranes, microfluidic systems, laser capture microdissection and biochemical methods (reviewed by Taliaferro, 2022). Most of these methods have however been limited to cell types with large and long morphologies and only a few have been used for the assessment of the egg.

Cell fractionation using tomography has recently been used quite frequently for the assessment of the transcriptome spatial location within the egg and embryos of several model organisms (Junker et al., 2014; Sindelka et al., 2010). It involves the cryosectioning of the cell or tissue into defined regions. These regions can then be sequenced separately and the transcriptome for each section explored (Junker et al., 2014; Sindelka et al., 2010). This method allows for precision over the spatiotemporal parameters of fractionation (reviewed by Taliaferro, 2022). In *in situ* sequencing using patterned arrays (Stereo-seq), the tissue can be overlayed onto a chip containing DNA nanoballs (Chen et al., 2022; reviewed by Taliaferro, 2022). These barcoded nanoballs allow for the detection of sequences (usually 400 unique) at resolution of hundreds of nanometers (Reviwed by Taliaferro, 2022). This method has not been applied to the oocyte but to the expression patterns within the developing *Drosophila* embryo (reviewed by Wang et al., 2022). Cellular fractionation using microporous membranes has been applied primarily to neuronal cells (reviewed by Taliaferro, 2022). It deals with the mechanical separation of the cell body from their projections (reviewed by Taliaferro, 2022, 2019). The cells can be grown on a membrane whereby the projections grow through the pore and can then be easily isolated from their cellular body. Transcripts from each portion are then sequenced and queried. The microfluidic culture system extends on this method and allows for a more fine-tuned control and compartmentalization of the cultured neuron. Laser capture and microdissection allows a high

level of precision when extracting a region of the cell/tissue (reviewed by Taliaferro, 2022, 2019). It has however been used primarily on tissues, but has also been adapted for neuronal subcompartments, and also the apical and basolateral compartments of intestinal epithelial cells (Moor et al., 2017; reviewed by Taliaferro, 2022; Zivraj et al., 2010). However, it is limiting in that it results in a low yield in the recovered RNA (reviewed by Taliaferro, 2022). Biochemical cellular fractionation is a unique method compared to the previous and relies on the fractionation of cells using detergent containing buffers and differential centrifugation (reviewed by Taliaferro, 2022, 2019). RNA is extracted from each of these fractions and sequenced using high throughout sequencing. This technique has been used to identify transcripts localized to cytoskeletal substructures (Blower et al., 2007; Sharp et al., 2011) and organelles (Kaltimbacher et al., 2006; reviewed by Taliaferro, 2022). However, this technique does not necessarily give spatial resolution as the co-purification of RNAs into the same biochemical fraction may have originated from different regions (reviewed by Taliaferro, 2022). To improve on these limitations, RNA proximity labelling eg: APEX-Seq, Cap-Seq (chromophore-assisted proximity labeling and sequencing) and Halo-Seq (Halo-mediated localization Sequencing), utilize a protein marker that specifically localizes to a subcellular region of interest (reviewed by Fazal and Chang, 2019; Taliaferro, 2022, 2019). The RNAs in this region becomes labeled by the created reactive chemical species.

2.8.3 Mechanism of the zipcode identification

The historical investigation into the mechanisms of transcript localization usually involved the systematic testing of portions of the localized transcript (usually the 3'UTR) attached to a reporter transcript (reviewed by Taliaferro, 2022). Once the RNA fragment has induced localization, the region is sequentially shortened until the most likely zipcodes are identified. Such methods are extremely labor intensive and therefore limited to assessment of only a few transcripts (reviewed by Taliaferro, 2022). Additionally, in the case of localization in oocytes, the localization machinery might be temporally activated and therefore the correct zipcode associated with the reporter transcript may not necessarily induce a localization if they are not present at the correct time. Furthermore, localization may also require the presence and interactions of multiple zipcodes or the formation of intricate secondary structures. Given these limitations, this type of experiment is limited to very low throughput and the results are not necessarily extrapolatable at a global transcriptional level.

Another approach to identifying localization motifs represents the *de novo* identification of shared zipcodes within commonly localized transcripts (reviewed by Taliaferro, 2022). Once transcripts have been identified to be localized to a particular region, their sequences can be scanned for shared motifs that are lacking in the unlocalized transcripts. Although these sequences will still need to be validated experimentally as described for the classical experimentations, they serve as an efficient and high throughput method for the exploration and identification of potential mechanisms. The *in silico* identification of motifs is however a computationally intensive process. It requires assessing similarities shared within the sequences of a group of RNAs that would otherwise not occur by just random chance. There are two common approaches for such identification, the enumeration approach, and the probabilistic approach (reviewed by Ardekani, 2010). The enumeration approach searches every combination of an oligonucleotide of a given length across each sequence. The iteration across all combinations ensures that it will discover an optimum motif. However, the more sequences, the more time intensive the process becomes as all combinations are queried. Popular algorithms based on this approach are DREME (Discriminative Regular Expression Motif Elicitation) (Bailey, 2011), CisFinder (Sharov and Ko, 2009), Weeder (Pavesi et al., 2004), FMotif (Jia et al., 2014), and MCES (Mining and Combining Emerging Substrings) (Ardekani, 2010; reviewed by Jia et al., 2014). This method can be further subdivided into several classifications such as either simple word enumeration, cluster-based method, tree-based method, graph theoretic-based method, hashing-based method, fixed candidates and modified candidates-based method (reviewed by Ardekani, 2010).

The probabilistic approach method creates a position weight matrix (PWM) that is built on the probability of encountering the motif in question from randomly sampled k-mers within the transcripts of interest (reviewed by Ardekani, 2010). This method includes the deterministic approach used by MEME (Multiple Expectation maximizations for Motif Elicitation) (Bailey and Elkan, 1994) and the stochastic approach used by BioProspector (reviewed by Ardekani, 2010; Liu, 2001). Other tools including, RNAalifold (Yu et al., 2015), RNAMotif (Yu et al., 2015), RNALfold (Hofacker et al., 2004) and ERPIN (Easy RNA Profile IdentificatioN) (Lambert et al., 2004) have also been developed to identify similarities in secondary structures that may be shared across localized transcripts (reviewed by Hamilton and Davis, 2011). However, these methods are computationally intensive when analyzing many transcripts and

long UTR sequences (Huang et al., 2019). Additionally, depending on the modelled parameters, the secondary structures may differ (Vicens and Kieft, 2022).

A more targeted approach for the detection of motifs relies on the sequencing of regions that are bounded to RBPs (reviewed by Hafner et al., 2021). CLIP (cross-linking and immunoprecipitation) technique and its derivatives utilize cross-linking of the RNAs and their interacting proteins using UV radiation. The cross linked RNA regions are then purified using immunoprecipitation or some similar methods. The ribonucleoprotein (RNP) complexes are further purified, and the RNA fragments released by digestion of the RBP. The fragments are then sequenced and mapped to the genome. The regions may represent areas that interact with RBP and be involved in the localization process (reviewed by Hafner et al., 2021). However, some of the limitations of this method include its requirement for high-quality antibodies against the protein of interest and also challenges in obtaining the high-quality RNP complexes (Reviewed by Hafner et al., 2021; Ramanathan et al., 2019).

2.8.4 RNA Sequencing technology

The development of high throughput sequencing has ushered in an age of renewed discovery from previously experimentally limiting methods. The Illumina sequencing platform has become one of the most popular sequencing platforms to assess the full transcriptome from samples (reviewed by Levy and Myers, 2016). The library preparations require extraction of the total transcripts, followed by fragmentation, fragment size selection, adapter ligation and then enrichment (Bronner and Quail, 2019). Given that 3'UTRs of eukaryotic mRNAs are usually rich in polyA (polyadenylic acid) regions, a polyT (polythymine) capture technique and enrichment of the 3'UTR regions are usually employed (reviewed by Chen et al., 2020; Passmore and Coller, 2022). However, this usually excludes some of the lncRNAs (long non-coding RNAs) which do not have a polyA rich 3'UTR (Yang et al., 2011). Additionally, in oocytes, many maternal transcripts have been found to be either unadenylated or partially adenylated to alter their stability and translational potential (Winata and Korzh, 2018). Therefore, the polyT capture method may not be the most appropriate for assessing maternal transcripts. Some techniques have solved this by utilizing random priming which can target any aligned region of the fragmented transcripts (Boone et al., 2018). Therefore, previous experiments related to RNASeq of the maternal transcripts that utilized polyA capture may have missed the potential contributions of some RNA transcripts (Claußen et al., 2015).

There are other factors that also need to be considered after sequencing. These include the correct mapping of the fragments to the genome or transcriptome, and the correct normalization of the produced counts table. The normalization of the count table represents an essential part as it seeks to minimize the effects introduced from library size differences (Dillies et al., 2013; Jiang et al., 2011). However, there are advantages and disadvantages with each normalization technique and therefore those that best fit the given experimental type should be selected (Evans et al., 2018).

One of the primary assumptions of these normalization methods is that a large portion of the transcripts is not differentially expressed. DESeq2, one of the popularly used tool for differential analysis, also follows this assumption and utilizes the median-of-ratios method to normalize the library size by scaling the data using a calculated size factor (Love et al., 2014). However, such assumptions on the distribution of the data can become erroneous when there is a global shift in the expression of transcripts between the experimental conditions. Such issues have already been described for cancer studies or the comparison between very different tissue types (Chen et al., 2016; Lovén et al., 2012; Xu and Zhang, 2016). One such sample that may be affected by a global shift is the egg. Given that some eggs show active polarization of transcripts at the subcellular level and also given the bulged shape of the egg, there is already an inherent bias in where the transcripts are present. The total RNA between sections along the animal-vegetal axis are not the same (Sindelka et al., 2010). Therefore, it is possible that given these limitations that previous RNASeq analysis involving cryosectioned eggs might have missed the complete asymmetrical transcriptome.

3 Aims of the Thesis

The aim of this thesis is to map out the localization of maternal transcripts along the animal-vegetal gradient across several species. It targets the utilization of bioinformatic approaches to identify these biological factors. This thesis consists of the work from four main publications. The motivation to pursue this project lies in the deficit in our knowledge on the fundamental localization drivers coordinating development. Historically the study of embryogenesis and developmental biology have always been truncated and lagging behind the other fields. Additionally, most of our knowledge on maternal transcript asymmetry is limited to a few model organisms and genes. The onset of high throughput sequencing has opened up the opportunity to assess the full breadth of the transcriptome landscape within tissues and cells. Despite this methodological availability, very few publications have addressed the embryogenesis process with even fewer looking at the subcellular levels. RNA localization is a cellular property that extends beyond just the egg. Its requirement is essential across all cells and its perturbation can lead to serious defects. As a result, understanding what transcripts are involved and the potential mechanisms driving its localization, serves as the first pillar to truly understand the complex properties driving development and other cellular processes.

The main objectives outlined in this thesis:

1) To create a novel approach to normalize data from bulk-RNASeq data that addressed the innate issues that affect normalization of data from the egg. **(Publication I, II, III)**
2) Assess and compare the maternal transcript asymmetry within the eggs of several divergent species. **(Publication I, II)**
3) Identify putative localization pathways when transcript asymmetry occurs during oocyte maturation between divergent species. **(Publication IV)**

4 Experimental Part

During my PhD study, I was part of the authorship of eight publications, three of which I am shared first author. Although these publications utilized skills and techniques learnt during the process of my PhD studies, I selected four publications directly related to my PhD topic as the base of my thesis (Publication I-IV).

4.1 List of publications

Publication I: Asymmetric distribution of biomolecules of maternal origin in the *Xenopus laevis* egg and their impact on the developmental plan.

Sindelka, R., Abaffy, P., Qu, Y., Tomankova, S., Sidova, M., **Naraine, R.**, Kolar, M., Peuchen, E., Sun, L., Dovichi, N. and Kubista, M. (2018). Sci. Reports 2018 81, 8, 1–16.

doi:10.1038/s41598-018-26592-1

Journal: Scientific Reports, IF_{2018} 4.29

Publication II: Evolutionary conservation of maternal RNA localization in fishes and amphibians revealed by TOMO-Seq.

Naraine, R.*, Iegorova, V.*, Abaffy, P., Franek, R., Soukup, V., Psenicka, M. and Sindelka, R. (2022). Dev. Biol., 489, 146–160.

doi:10.1016/j.ydbio.2022.06.013

*Shared first author

Journal: Developmental Biology, IF_{2022} 3.15

Publication III: NormQ: RNASeq normalization based on RT-qPCR derived size factors

Naraine, R.*, Abaffy, P.*, Sidova, M., Tomankova, S., Pocherniaieva, K., Smolik, O., Kubista, M., Psenicka, M. and Sindelka, R. (2020). Comput. Struct. Biotechnol. J., 18, 1173–1181.

doi:10.1016/j.csbj.2020.05.010

*Shared first author

Journal: Computational and Structural Biotechnology, IF_{2022} 7.27

Publication IV: Comparison of RNA localization during oogenesis within *Acipenser ruthenus* and *Xenopus laevis*

Iegorova, V.*, **Naraine, R.***, Psenicka, M., Zelazowska, M. and Sindelka, R. (2022). Front. Cell Dev. Biol., 10, 1–16.

doi:10.1002/jez.b.22802

*Shared first author

Journal: Frontiers in Cell and Developmental Biology, IF_{2022} 6.08

Other unrelated publications:

1) Abaffy, P., Tomankova, S., **Naraine, R.**, Kubista, M. and Sindelka, R. (2019). The role of nitric oxide during embryonic wound healing. BMC Genomics 20. doi:10.1186/s12864-019-6147-6
2) Bezawork-Geleta, A., Wen, H., Dong, L., Yan, B., Vider, J., Boukalova, S., Krobova, L., Vanova, K., Zobalova, R., Sobol, M., Hozak, P., Novais, S.M., Caisova, V., Abaffy, P., **Naraine, R.**, Pang, Y., Zaw, T., Zhang, P., Sindelka, R., Kubista, M., Zuryn, S., Molloy, M.P., Berridge, M. V., Pacak, K., Rohlena, J., Park, S. and Neuzil, J. (2018). Alternative assembly of respiratory complex II connects energy stress to metabolic checkpoints. Nat. Commun. 2018 91 9, 1–17. doi:10.1038/s41467-018-04603-z
3) Gazo, I., **Naraine, R.**, Lebeda, I., Tomčala, A., Dietrich, M., Franěk, R., Pšenička, M. and Šindelka, R. (2022). Transcriptome and Proteome Analyses Reveal Stage-Specific DNA Damage Response in Embryos of Sturgeon (*Acipenser ruthenus*). Int. J. Mol. Sci. 23. doi:10.3390/ijms23126392
4) Magalhaes-Novais, S., Blecha, J., **Naraine, R.**, Mikesova, J., Abaffy, P., Pecinova, A., Milosevic, M., Bohuslavova, R., Prochazka, J., Khan, S., Novotna, E., Sindelka, R., Machan, R., Dewerchin, M., Vlcak, E., Kalucka, J., Stemberkova Hubackova, S., Benda, A., Goveia, J., Mracek, T., Barinka, C., Carmeliet, P., Neuzil, J., Rohlenova, K. and Rohlena, J. (2022). Mitochondrial respiration supports autophagy to provide stress resistance during quiescence. Autophagy 18, 2409–2426. doi:10.1080/15548627.2022.2038898

Manuscripts in preparation:

5) Sindelka, R., Abaffy, P., Zucha, D., **Naraine, R.**, Kraus, D., Netusil, J., Smetana Jr., K., Lukas, L., Beduya Endaya, B., Neuzil, J., Psenicka, M., Kubista, M. (unpublished). Characterization of regeneration initiating cells during Xenopus laevis tail regeneration

6) Šimková, K., **Naraine, R.**, Vintr, J., Soukup, V., Sindelka, R. (unpublished). Maternal RNA localization during early development of Ambystoma mexicanum

4.2 Publication summary

Although several research have studied the maternal transcript localization, they have been limited in the models used, extent of maternal transcripts assessed and the resolution of transcript detection. The results from this thesis address these deficits in our knowledge by analyzing multiple divergent models using the most up-to-date high-throughput sequencing techniques. This allowed us to fully characterize the maternal transcript asymmetry, identify their sub-localized profiles, the level of conservation between models and also the putative localization motifs present in their UTR sequences. Additionally, we have also fully characterized when this asymmetry is mounted during oogenesis and discovered transcript depletion and synthesis during the oocyte maturation. In the process of this research, we have also created a new method for normalization of transcript count data from the egg, as it represents an atypical single cell sample. The method fully allows all the transcripts to be characterized and reduces the biases produced from other normalization methods. Below is presented a list of the publications with a brief summary of their findings.

Publication 1:

- Asymmetric distribution of biomolecules of maternal origin in the *Xenopus laevi*s egg and their impact on the developmental plan.
    - Sindelka, R., Abaffy, P., Qu, Y., Tomankova, S., Sidova, M., **Naraine, R.**, Kolar, M., Peuchen, E., Sun, L., Dovichi, N. and Kubista, M. (2018) Sci. Reports 2018 81, 8, 1–16.

Although the *X. laevis* has been used as a model for developmental biology, the assessment of its maternal transcripts has been primarily limited to just two sections along the animal-vegetal regions or limited historically to just a few transcripts. In this publication we assessed the asymmetry of the full maternal transcripts (mRNA, lncRNA) and proteins within subsections (5 sections) of the egg ranging from the most animal to the most vegetal regions. We observed copious amounts of asymmetrical maternal transcripts that showed 4 distinctive localization profiles. We identified 14828 asymmetrical maternal transcripts, (>2x more than previously described). 424 were found in our newly described extreme animal section, the majority (14176)

in the animal section, 194 in the vegetal and 34 in the extreme vegetal. There was also asymmetry in the proteins with 322 in the animal, 761 in vegetal and 1468 evenly distributed. Also, the lncRNAs were also assessed and the following asymmetry observed, 12 extremely animal, 136 animal, 10 vegetal, 2 extremely vegetal. There was no clear correlation between the protein sequences and the maternal mRNAs. The mRNA gradient for a few animal transcripts (*actb*, *clic5* and *dicer1*) were observed to become more uniform in the blastula stage. Several enriched animal and vegetal putative localization motifs were identified. We also observed different localizations of homoeologous genes which may indicate some neofunctionalization of these homologous transcripts.

Contributions: Bioinformatic analysis of the data. Performed and analyzed *de novo* motif discovery and analysis. Carried out analysis for known regulatory motifs and miRNA binding site analysis of localized transcripts. Proofread and contributed to the methodology, discussion, and main text.

Publication 2:

- Evolutionary conservation of maternal RNA localization in fishes and amphibians revealed by TOMO-Seq.
    - **Naraine, R.***, Iegorova, V.*, Abaffy, P., Franek, R., Soukup, V., Psenicka, M. and Sindelka, R. (2022) Dev. Biol., 489, 146–160.

The maternal transcripts and its asymmetry in the matured egg has only been assessed recently using high throughput analysis of all transcripts for a few models, in particular *D. rerio* and *X. laevis*. These two models represent two evolutionarily divergent models with very different cleavage patterns, meroblastic and holoblastic respectively. However, there has not been a comprehensive comparison of the maternal transcript asymmetry across several evolutionarily distant vertebrate models. In this publication we assessed the asymmetry of maternal transcripts in two amphibians (*X. laevis*, *A. mexicanum*) and two fishes (*D. rerio*, *A. ruthenus*). We assessed the sub-localization profiles by analyzing the transcripts from within 5 sections ranging from the extreme animal to the extreme vegetal region. We compared the similarity between the transcript asymmetry, localization patterns and putative localization motifs. In this publication we observed extensive asymmetrical transcripts for all models except *D. rerio*. The transcripts could be placed into the previously found profiles (extremely animal, animal, vegetal, extremely

vegetal). We were able to identify a new central profile primarily in the *A. mexicanum* and the *A. ruthenus*. There was very little conservation between the members within each localization profile, with the most similarity observed within the animal localized transcripts. Despite the low similarity, many of the putative localization motifs found within the transcripts were similar across all species, indicating similar mechanisms for transportation or cis-acting control. Dendrograms analyzing the dissimilarity between the transcript and their localization pattern showed that that the amphibians grouped together, and the fishes grouped together, with the exception of the *D. rerio* animal transcripts which grouped more closely with the amphibian vegetal transcripts. Our localized transcripts correlated well with previous recent research published for *X. laevis* and *D. rerio*. Our findings showed that a few key maternal transcripts are conserved in their localization across the divergent models, and they are overall probably controlled by different initial steps.

Contributions: Carried out complete downstream bioinformatic analysis, included data normalization, profile detections, motif discovery, *in silico* functional analysis and cross species pattern discovery. Contributed to the writing of the publication and analysis of the data for discussion.

Publication 3:

- NormQ: RNASeq normalization based on RT-qPCR derived size factors.
    - **Naraine, R.***, Abaffy, P.*, Sidova, M., Tomankova, S., Pocherniaieva, K., Smolik, O., Kubista, M., Psenicka, M. and Sindelka, R. (2020). Comput. Struct. Biotechnol. J., 18, 1173–1181.

Normalization of bulk-RNASeq transcript counts is essential to mitigate the effects from biases during the sequencing process (eg: library size). However, most normalization procedures, eg: median-of-ratios, assume that the majority of the transcripts between the conditions are not differentially expressed and are of similar transcript heterogeneity. This is however not true for all tissue types, for example in tumor cells or comparing very contrasting cell types like from the testes with the epidermis. This limitation can also be observed for the egg, where there is the presence of a large proportion of maternal transcripts and also a disproportionation in their localization across the sections. To address this issue, we developed a new tool, based on normalization of the raw counts using RT-qPCR data from stably differential expressed or

uniformly expressed transcripts. The new method, NormQ (normalization by RT-qPCR), yielded significantly better results on modelled data whereby 90% of the expected differentially localized transcripts (DLTs) were detected with the correct profiles versus just 37% using the median-of-ratios method. On real data from the *X. laevis* and *A. ruthenus* egg, more DLTs were identified and RT-qPCR validation of selected transcripts, showed that 87.5% (*X. laevis*) and 95.5% (*A. ruthenus*) DLTs matched their expected profiles versus when the median-of-ratios method was used 50% and 72.7% respectively and even outperformed the spike-in control which was at 86.4%. The NormQ method therefore serves as an ideal and easily implementable method for normalization of transcript count data from bulk-RNASeq. In relation to the eggs, it allowed for the characterization of the animal group as they represented the majority of the transcript and would have otherwise not been observed when using the median-of-ratios method.

Contributions: Created the final codes and algorithms for the novel method NormQ. Performed the bioinformatics analysis described in the paper. I prepared and wrote the manuscript.

Publication 4:

- Comparison of RNA localization during oogenesis within *Acipenser ruthenus* and *Xenopus laevis*.
    - Iegorova, V.*, **Naraine, R.***, Psenicka, M., Zelazowska, M. and Sindelka, R. (2022) Front. Cell Dev. Biol., 10, 1–16.

Our previous publication showed that there exist copious amounts of asymmetrically localized maternal transcripts in both the matured eggs of the *X. laevis* and *A. ruthenus*. However, information on when these gradients are established during oogenesis is very limited in *X. laevis* to just a few genes and practically non-existent for *A. ruthenus*. In this publication we analyzed the temporal establishment of the animal-vegetal gradient at different stages of the oocyte up to maturity. We discovered that many transcripts have a temporal preference for the formation of their asymmetry. We were able to identify the previously early and late vegetal pathways in both models. Specifically, we found that there were vegetal transcript gradients that were established only during the early stages, only during the late stages or starts in the early phase and becomes more pronounced during the later stages. More interestingly we observed early and late animal gradient pathways for both models, something previously not discovered. Some of these temporal transcripts were conserved between both models. Additionally, we also observed *de*

novo transcription and transcript degradation within the oocyte, pointing to an active transcriptional landscape even during the egg maturation.

Contributions: Performed all downstream bioinformatic analysis of the datasets. Contributed to the analysis and interpretation of the results and also to the writing of the text and discussion.

5 Methods

The following methodological approaches were utilized during the course of this dissertation:

Performed by coauthors:
Cryosectioning of oocytes (Papers I, II, III, IV)

RNA extraction (Papers I, II, III, IV)

RT-qPCR analysis (Papers I, II, III, IV)

RNASeq library preparation (Papers I, II, III, IV)

RNASeq quality control, sequence alignment and gene quantification (Papers I, II, III, IV)

Performed by me:
RNASeq normalization, quality assessment, profile clustering and characterization (Papers I, II, III, IV)

NormQ development (Paper I, II, III)

Motif discovery and enrichment analysis (Papers I, II, IV)

Enrichment of gene ontology (GO) (Papers I, II, III, IV)

6 Discussion

This dissertation represents the first study to comprehensively compare the sublocalization of maternal transcripts across multiple distance models. In this research, we found that many maternal transcripts show differential localized profiles within the egg of our models. Our research analyzed the *X. laevis*, *A. mexicanum*, *A. ruthenus* and *D. rerio* for subsectional compartmentalization in the egg. This led to the discovery of discrete localization categories of extreme animal, animal, central, vegetal, and extreme vegetal transcripts. Publication I (Sindelka et al., 2018) represents the first high throughput RNASeq characterization of subcellular localization in *X. laevis* while publication II (Naraine et al., 2022) for *A. mexicanum* and *A. ruthenus*. Publication II represent the second RNASeq dataset on the subcellular localization in *D. rerio* egg with the first being from the recent publication from Holler et al., 2021. Even before analyzing for asymmetry, all models showed differences in their transcript quantities per egg section (Publication II: Naraine et al., 2022). All of the matured eggs contained copious amounts of maternal transcripts as is seen typically for organisms that produce large eggs (Harry and Zakas, 2023).

Our results amplify the notion of the diversity in the early developmental controls between different organisms. Although the developmental process is an extremely essential and coordinated process, it appears that there has not been a clear overall conservation in the maternal transcripts that constitute the animal-vegetal gradients. The egg is however exceptionally dynamic and contains multiple sublocalized transcript profiles. Despite the low conservation, we were able to identify several transcripts related to germ cell development and transcription factors that were conserved in all four models. Additionally, many animal transcripts involved in different aspects of cell cycle regulation, RNA transport, and cytoskeleton organization were conserved amongst the four model. This points to a possible conservation of these key processes. Additionally, we observed evidence to a possible conservation of localization pathways. We characterized for the first time an early and late animal localization pathway and also more transcript representatives for the previously known early and late vegetal localization pathways. Several transcripts showed the same pathway usage between the *X. laevis* and *A. ruthenus*. Additionally, there were shared *de novo* and degraded transcripts during oocyte maturation. We observed several previously undescribed universally

conserved putative localization motifs in the UTR sequences of the shared vegetal and animal transcripts.

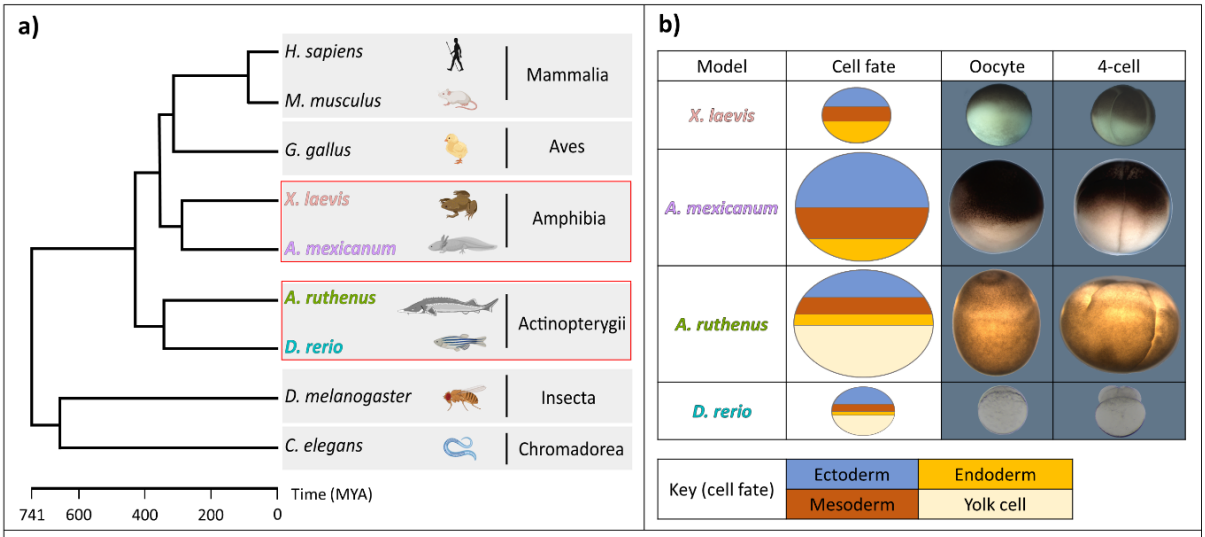
6.1 New normalization method

The development of a new way to normalize the data allowed for the detection and correct profiling of more localized transcripts. As such our NormQ method serves as an essential part for our study (Publication III: Naraine et al., 2020). Prior to our studies very few genes were discovered to be differentially localized to the animal pole (Publication III: Naraine et al., 2020). This was due to the effect of the normalization method used by these publications, particularly the median-of-ratios method. The median-of ratios method assumes that between conditions, the total transcripts are similar, heterogeneity is low, and that the majority of the transcripts show very low differences. However, these conditions are not always representative between a given sample, especially for samples from very contrasting tissue types, tumor tissue or in our case the egg (reviewed in the introduction). Each section of the egg, just by its size difference contains a different amount of transcript (Publication II: Naraine et al., 2022). Additionally, the active localization of transcripts creates a disparity in the level of transcript heterogeneity between the sections. As such normalizing the data forces the copious amounts of animal transcripts to be seen as not being localized and everything becomes normalized relative to their location. To solve this issue and obtain an absolute localization pattern, we utilized the more sensitive RT-qPCR methodology to assay stably expressed genes and obtain a relative fold difference separating their expression in each section. To reduce the bias introduced from one gene, we found the geometric mean of this fold change across several stable transcripts. We tested this method both *in vivo* and *in silico* with a 0.99 AUC match to the expected DLTs relative to the 37% in the traditional median-of-ratios method. Our method allowed for not only the discovery of more novel transcripts but also the correct profiling of these transcripts (94% vs 37% for the NormQ and median-of-ratios respectively). As such this method was a vital implementation to improve the characterization of the subcellular localization of transcripts within the egg.

6.2 New and diverse models

There are many similarities and differences between these models as it relates to the characteristics of their eggs and development, and they are summarized in the introduction and also Figure 10 (Publication II: Naraine et al., 2022). Evolutionarily these organisms are quite

diverse, and our research observed that even the animal-vegetal axis is reflective of this level of divergence. As many of these models were new, they offered a unique challenge to analyse. Some of these include the lack of a transcriptome for *A. ruthenus*, which was solved by creating an annotated *de novo* transcriptome. Other aspects involved determining the orthologs between models which was achieved by comparing the protein sequence similarity between the different models.



c)

Model	Maturation	Size	Ploidy	Haploid	Genome	Proteins	Egg size	Fertilization	Early cell division	Cleavage pattern
X. laevis	1-2 years	10 cm	Allotetraploid	18	3.1 x 10^9 bp	45099	1.2-1.5 mm	monospermic	1:30 h	holoblastic
A. mexicanum	12 months	30 cm	Diploid	14	32 x 10^9 bp	23251	1.85-3 mm	polyspermic	1 h	holoblastic
A. ruthenus	3-9 years	125 cm	Diploid	~59±1	3.74 x 10^9 bp	47424	1.9-2.5 mm	polyspermic	3 h	holoblastic
D. rerio	2.5 months	3.5 cm	Diploid	25	1.5 x 10^9 bp	25672	0.7-0.9 mm	monospermic	40 min	meroblastic

Figure 10: "The taxonomic and developmental relationship between the different models studied. **A)** Timetree (v. 4.0) taxonomic tree of select animal models. The model organisms used for this research are highlighted. **B)** Images of the egg, 4-cell stages, and cell fate maps. **C)** Genomic and egg developmental characteristics." (From: Publication II: Naraine et al., 2022)

6.3 Comparison of our datasets with previously published dataset

Our profiled genes correlated well with many of the detected genes from previously published datasets (Figure 11). These comprised of three RNASeq (Claußen et al., 2015; Holler et al., 2021; Owens et al., 2017) and one microarray (Cuykendall and Houston, 2010) datasets for *X. laevis* while one RNASeq dataset (Holler et al., 2021) for *D. rerio*.

Despite the many similarities with our *X. laevis* vegetal dataset (60-100% - transcripts with gene symbols and present in at least two datasets), there were discrepancies. Only 31 vegetal transcripts were similar in all datasets, while an additional 74 for just the RNASeq datasets (Figure 11). This could be due to different thresholds used to define a differentially localized transcript, the different methods for normalization of the data (median-of-ratios vs NormQ), the differences in experimental design (two sections vs multiple, different annotations), and also the innate biological differences between eggs of the same species (Harry and Zakas, 2023). Such comparative differences are in fact useful as it converges the data towards the list of genes with the strongest conserved differences regardless of experimental differences, experimenter, model conditions and sporadic maternal gene expressions. Almost all (99%) of the vegetally localized transcripts as determined by Holler et al., 2021 was identifiable in our *D. rerio* dataset (Figure 11) (Publication II: Naraine et al., 2022).

A large portion (>95%) of the previously identified animal transcripts were detected in our dataset for both the *X. laevis* and *D. rerio* datasets (Figure 11) (Publication II: Naraine et al., 2022). Our detection of almost all of the previously known animal transcripts is not surprising as our dataset contained the most comprehensive characterization for the animal transcripts due to the implementation of our new normalization method (Publication III: Naraine et al., 2020). The NormQ method allowed for determination of the absolute localization pattern of the transcripts rather than the relative (Publication III: Naraine et al., 2020). This prevented the normalization of the data onto the majority of the transcripts (which would be animal transcripts) and allowed for the detection and quantification of them fully.

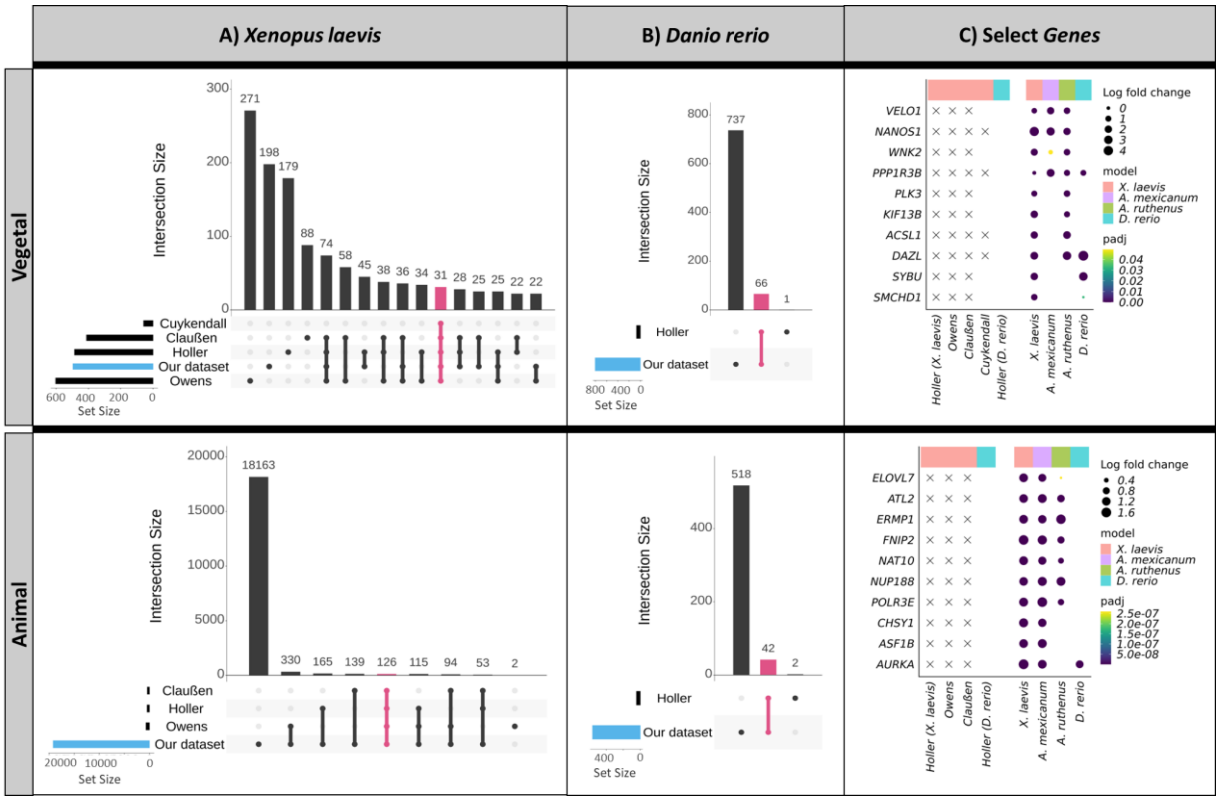


Figure 11: "Comparative analysis between the maternal asymmetry observed within each studied model and with data from previous publications. Overlap of asymmetrical maternal transcripts shared between this dataset and **A)** Three published datasets on the *Xenopus laevis* egg, **B)** One published dataset on the *Danio rerio* egg. Highlighted in red are the subset of transcripts that are shared across all datasets. **C)** The differentially localized presence within all four analyzed models for the top most significant transcripts derived from the overlap between all datasets/all RNA-Seq datasets for the *X. laevis* model from panel A." (From: Publication II: Naraine et al., 2022)

6.4 Evolutionary conservation of maternal transcripts

Overall, there was a very poor conservation of the localization pattern of transcripts across the models. We observed an overlap of 1015 differentially localized transcripts that were conserved across all four models (Figure 12) (Publication II: Naraine et al., 2022). The amphibians shared the highest numbers of DLTs (7531) while the fishes only 1240. A large overlap (7275) in orthologous maternal transcripts within the eggs have already been observed between divergent species like *Mus musculus*, *Bos taurus*, and *X. laevis* using microarray analysis (Vallée et al., 2008).

The animal genes showed the greatest amount of conservation across the models with 339 genes being shared, while only 7 genes for the vegetal (Table 5). Holler et al., 2021 also found a low conservation of the vegetal genes (9) across *D. rerio*, *X. laevis* and *X. tropicalis*. We were able to observe 8 of these 9 genes as vegetal between *D. rerio* and *X. laevis* in our dataset. However, both *A. mexicanum* and *A. ruthenus* showed animal localization for many of these vegetal genes (Publication II: Naraine et al., 2022). The shared animal transcripts were involved in cell cycle regulation, RNA transport, cytoskeleton organization, the heart, cerebellum and skeletal system development, while the vegetal transcripts were involved in aspects of metabolic pathways, transcription factors and protein complexes. Such regulatory disparity makes sense as many genes related to the cell division process were enriched in the animal region. Similar enriched GO terms were observed by other researchers for the animal and vegetal enriched genes in *X. laevis* (Claußen et al., 2015; Cuykendall and Houston, 2010; Owens et al., 2017) (reviewed in introduction). Some key PGC markers were enriched in the vegetal region but were different or absent in *A. mexicanum* and *D. rerio* and might represent transcripts that are relocalized in the later zygotic stages or after fertilization (Publication II: Naraine et al., 2022).

One of the unique discoveries was the observation of a novel group of genes that were centrally enriched within the oocyte. These transcripts were absent or limiting in the *X. laevis* and *D. rerio* but showed a distinct profile in the *A. ruthenus* and *A. mexicanum* models. However, gene ontology analysis of these genes did not give any useful information on their function.

The distance matrix tree of the differences in localization patterns, placed the animal transcripts of the *X. laevis* and *A. mexicanum* closest followed by the *A. ruthenus*, matching their evolutionary divergence (Publication II: Naraine et al., 2022). This was also observed for their vegetal genes. However, both the animal and vegetal transcripts of the *D. rerio* both branched

within the vegetal clad, showing a greater relationship of their transcripts with the vegetal genes of the other species. Interestingly, our research has demonstrated clearly that the yolk for all the models is rich in maternal transcripts. This is even true for the *D. rerio* vegetal pole which does not take part in the cellular division during cleavage but instead serves solely as a nutrient deposit. The unique cleavage pattern of *D. rerio* might be a reason for the observed shift in the placement of the vegetal transcript, whereby the transcripts are shifted to the mid to upper region of the egg.

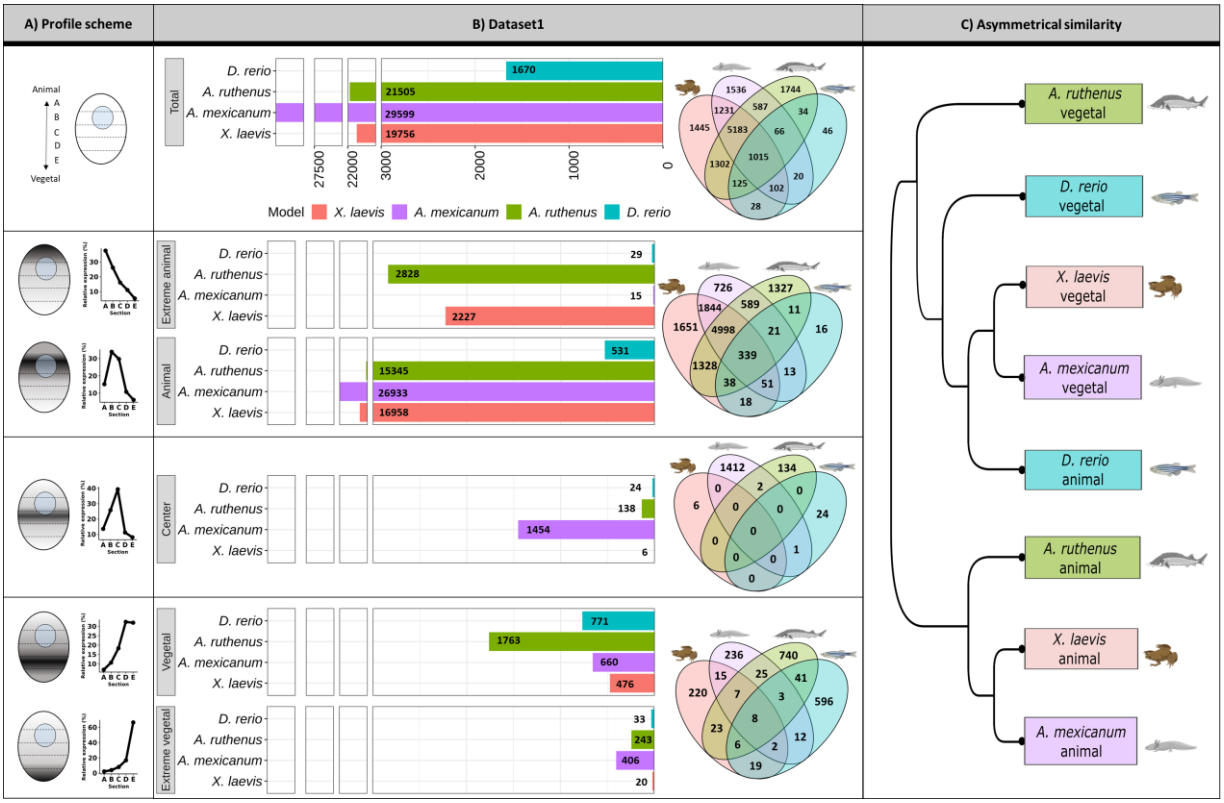


Figure 12: "Localization of maternal differentially localized transcripts (DLTs). **A)** Schematic of the five main localization profiles detected for the maternal transcripts within the egg. The darker colored regions within the egg represent higher transcript count saturation. The line graphs represent the median expression of the transcripts within all models that showed the strongest distributions for each profile. **B)** Distribution of DLTs with a localization profile for each model (differentially localized transcript: padj <0.1, overall oocyte transcript count >20). Venn diagram shows the overlap in the number of unique gene symbols or orthologous units. **C)** Dendogram showing the similarity between the shared asymmetry amongst the analyzed models. The list of transcripts used comprised of only those with an assigned orthologous unit

number that did not have a paralog in a contrasting localization profile." (From: Publication II: Naraine et al., 2022)

Table 5: Vegetally localized transcripts found in the egg of all four models (*Xenopus laevis*, *Ambystoma mexicanum*, *Acipenser ruthenus*, *Danio rerio*). Their functions/putative functions are derived from the works of Holler et al., 2021 and review by Kloc and Kubiak, 2014. The localization pattern is for *X. laevis* and *A. ruthenus* based on our Publication IV: Iegorova et al., 2022.

Genes	Function	Localization pattern (***X. laevis***; **A. *ruthenus*)**
dnd1	germ cell development	late vegetal - predefined
grip2	germ cell development	early vegetal
hwa	dorsal organizer and body axis formation	early vegetal
ppp1r3b	potential germ plasm determinant	late vegetal-predefined; early vegetal
rbpms2	germ cell development	late vegetal-predefined
rnf38	germ cell development	early vegetal
rtn3	germ plasm RNA	late vegetal

6.5 Conservation of putative localization elements

Our research has discovered the presence of several shared, class specific and species specific putative motifs that might serve as the localization mechanism for the movement of the vegetal and animal transcript. Many of these have not been previously described (Publication II: Naraine et al., 2022; Publication I: Sindelka et al., 2018).

The localization of transcripts is driven by localization mechanisms involving the RNP complexes and the actin filament cytoskeleton. Localization element or zipcodes which are usually found in the UTR sequence of transcripts, serve as a proxy for the interaction with the RNP (reviewed in the introduction). We analyzed a curated list of the localized transcripts for each model to identify any unique motifs over-represented in a given set of profiled genes (Figure 13). We observed 7 3'UTR motifs and 15 5'UTR motifs significantly enriched for the vegetal transcripts shared across all models (Publication II: Naraine et al., 2022).

Only one of these have been previously described, a CAC rich motifs which we found present in the 3'UTR sequences of the vegetal transcripts shared across the models, or other variant of

them enriched between amphibians, between fishes or just specific to the model (Figure 13). This widespread enrichment of CAC motifs has already been described in the chordates and in some genes (eg: *grip2*) in *X. laevis* (Betley et al., 2002). Holler et al., 2021, also found CAC rich motifs enriched in *D. rerio*, *X. laevis* and *X. tropicalis*. It is interesting that despite the low conservation among the vegetal transcripts they all shared a similar zipcode which may indicate a similar mechanism for localization.

Although a few animal transcripts have been previously described in *X. laevis*, no mechanisms have been indicated for their localization. We found for the first time, one (AAGTATCT) 3'UTR motif present in the shared animal transcripts amongst all models (Figure 13). Similar to the vegetal transcripts, conserved motifs were also observed that were specific to the fishes, amphibians and individual models, and these may form the basis for their localization (Publication II: Naraine et al., 2022). All *de novo* identified motifs showed no significant (e-value < 0.01) match to any known motifs in the Ray2013_rb- p_All_Species dataset (Publication II: Naraine et al., 2022). Therefore, such novel motifs warrant further functional investigations to determine if they are indeed required for localization or some other biological role.

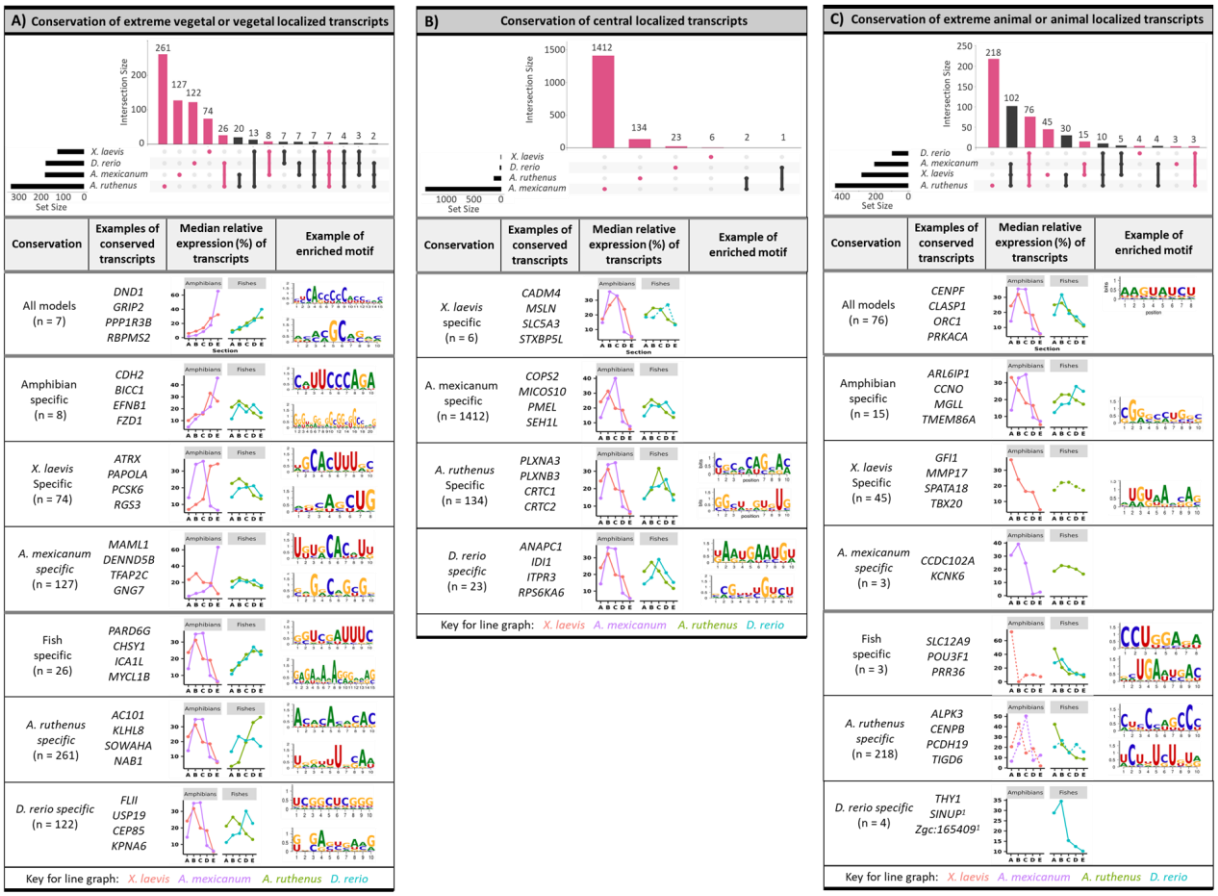


Figure 13: "Summary of the conservation of the differentially localized transcripts amongst the different models. **A)** Conservation of extreme vegetal or vegetal localized transcripts. **B)** Conservation of central localized transcripts. **C)** Conservation of extreme animal or animal localized transcripts. The Upset plots show the number of transcripts shared between the different models. The tables show the number of conserved localized transcripts for the specific models, examples of these conserved transcripts, the median expression of the transcripts within this group for all models and an example of a significantly enriched motif within the 3'UTR (top) and 5'UTR (bottom) sequences. Gene nomenclature follows those for human transcripts. [1]*Danio rerio* transcript nomenclature." (From: Publication II: Naraine et al., 2022)

6.6 Maternal transcript asymmetry during oogenesis

As stated previously the establishment of these localization patterns are driven by some localization mechanisms. In *X. laevis*, three pathways, the early, intermediate, and late have been described for a few vegetal transcripts (reviewed in introduction). Given that we observed copious amounts of sub-localized transcripts, we set out to look for when during the oogenesis process is this localization mounted (Publication IV: Iegorova et al., 2022). We analyzed 4 stages (III, IV, early V, late V) of the oocyte in *X. laevis* and two stages (early vitellogenic, late vitellogeneic) in the *A. ruthenus*. Our results showed that there is a clear temporal dependence for transcript localization and that some transcripts are even utilized, or *de novo* synthesized during the oogenesis process. All oocyte stages had approximately similar amounts of expressed maternal transcripts (*X. laevis*: 15800-16100, *D. rerio*: 21800-22600). However, the number of DLTs increased starkly from the immature to the mature stage (*X. laevis*: 450 to 1331, *D. rerio*: 3184 to 13901) (Publication IV: Iegorova et al., 2022). Therefore transcripts became more localized as the oocyte matured.

We were able to identify both the early and late vegetal pathways in both the *X. laevis* and *A. ruthenus* (Figure 14). Additionally, we were able to sub-divide the late pathway into primarily two groups, one that is established only in the late stage and the other which is slowly established as the oocyte matures (Figure 14). The most interesting results came from the animal transcripts. It was previously postulated that the animal gradient might be established due to diffusion. However, we were able to identify both an early and late animal pathway (Figure 14). The shift of the nucleus is already observed at the given analyzed stages and therefore should not be a contributing factor. Full details on the profiles and their constituents can be found in Publication IV: Iegorova et al., 2022.

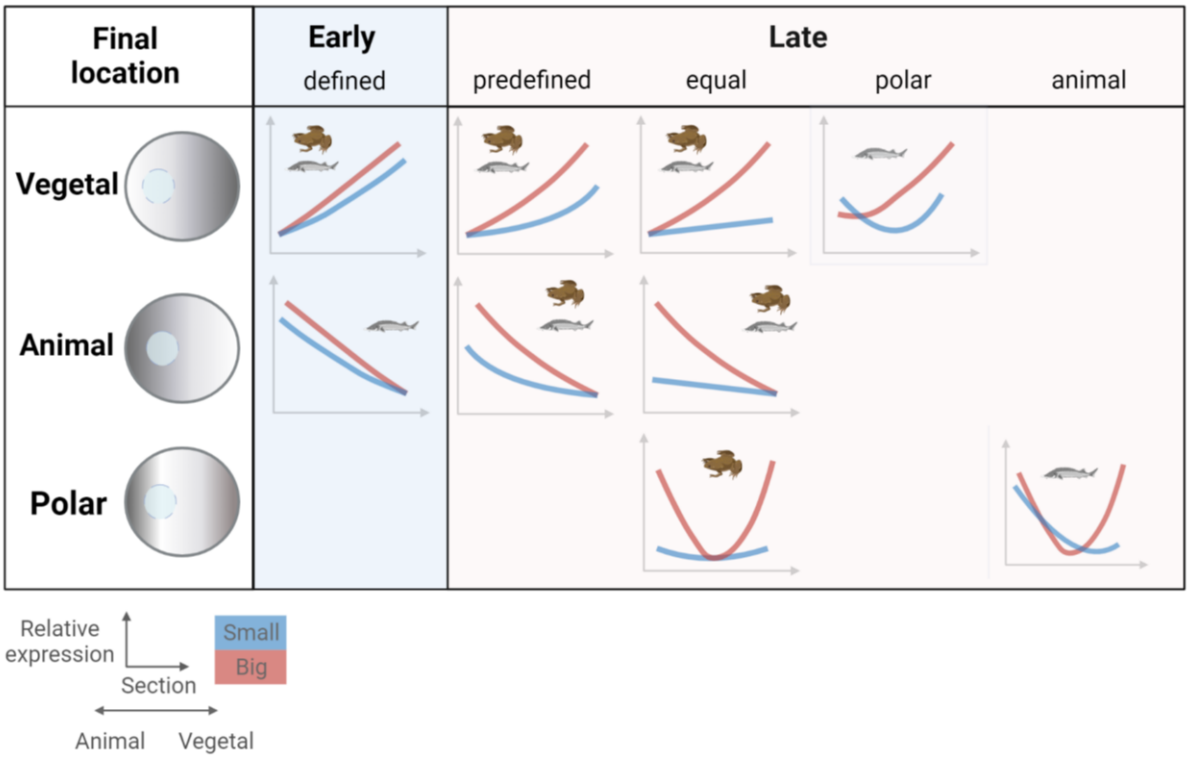


Figure 14: Simplified schematic of the localization patterns observed during the late stages of oogenesis in *Xenopus laevis* and *Acipenser ruthenus*. Three final localization profiles of vegetal, animal, and polar (high abundance in animal and vegetal region but low in middle region) were observed. The early pathway was already defined in the small oocyte. The late pathway contained four different subprofile whereby there was a different/reduced profile in the immature oocyte and the establishment of the final profile in the later stages: 1) early establishment of the profile in immature oocyte with more defined profile in later stage; 2) homogeneous distribution during immature oocyte with established profile in later stage; 3) polar distribution in immature oocyte with establishment of final profile in later stage; 4) animal profile in immature oocyte with establishment of final profile in later stage. (Image created using Biorender)

Between the two models very few (12) genes showed a similar shared early vegetal pathway, while 80 for the late vegetal pathway and 36 for the late animal pathway. The gene ontology analysis of these genes did not yield any significant results, however the ontology terms associated with these genes correlated with their perceived roles in those particular regions. Terms associated with the cytoskeletal organization, microtubule-based movement and germ cell development were associated with the genes that localized to the vegetal side early during oocyte maturation. Other genes associated with establishment of localization, cell morphogenesis and embryo development were associated with the late vegetal pathway. The shared late animal genes were related to cell cycle regulation and DNA repair (Publication IV: Iegorova et al., 2022).

Another interesting discovery was the fact that there were copious amounts of new transcripts being produced during the oogenesis process and also being degraded (Publication IV: Iegorova et al., 2022). Although this *de novo* synthesis behavior has already been described in *X. laevis*, it has been a commonly held assumption that *de novo* transcription is primarily activated during mid-blastula stage (Session et al., 2016). Even in humans it is now becoming apparent that *de novo* transcription is already active for some genes shortly after fertilization versus the previously thought 8-cell stage (Asami et al., 2022). These *de novo* and degraded transcripts in our models also showed sub-localized profiles, hinting to an active remodeling of the transcript profiles (Publication IV: Iegorova et al., 2022). Fifty shared transcripts were degraded while 27 were *de novo* synthesized in both the *X. laevis* and *A. ruthenus*. There were no enriched GO terms for these shared genes and further functional analysis will be required to better understand their role during the oocyte maturation. Given the shared similar developmental pattern (holoblastic) between *X. laevis* and *A. ruthenus*, it is not surprising to observe that there is a shared asymmetry of the maternal transcripts in the matured eggs and that there appears to be a shared conservation in the localization mechanisms that drive them to these locations.

7 Limitations

All methods have their limitations that introduce some level of bias to the data, and as such the conclusions. At each stage of the analysis of the data for this thesis, the effects of these biases have been attempted to be addressed or reduced as much as possible.

One of the biases can be from variations in the biological samples. However, the utilization of biological replicates from different females and not the same females allowed for the removal of variant transcripts that are not important for the biological process of embryogenesis. Given that this process is highly regulated, most alterations may result in the death of the organism. The maternal age of the organism may also be a source of variation, as it is believed that the cytoskeletal framework and transcript constituents may change with increased age (Bebbere et al., 2022; Ntostis et al., 2022). Although this was not taken into consideration, the varied age group and requirement for reproducible clustering of replicates helps focus on the elements that are still conserved regardless of stochastic variations and degenerative variations.

DESeq2 method for differential expression analysis on data that follows a log-normal distribution, represents one of the best methods in regard to its FDR control, power and stability when using different sample sizes (Li et al., 2022). Although it performs well on low number of samples (3), it performs better on larger sample sizes (6) (Li et al., 2022). However, most studies have used the lower end of the spectrum between 3-4 samples per condition as it maximizes the number of conditions that can be analyzed and also fits more informative data within the cost of the sequencing for a whole experiment. Our experiments have also ranged between three and five samples per condition but in some limited cases as low as two samples per condition after removal of outliers or recategorization.

One of the major hurdles to overcome was that of the limited annotations for many transcripts for each of the models. Also given that many of the models used are not typical models (*A. mexicanum*, *A. ruthenus*), the orthology between species has not been fully established. To reduce the effect from this we did extensive ortholog analysis to compare the protein sequences of the transcripts across the four model and also against the human proteome (Publication II: Naraine et al., 2022). This allowed us to map many more transcripts to their most likely gene symbol and also to their closest ortholog across the models. Another hurdle was the lack of a transcriptome for the *A. ruthenus* during the time of our publication II: Naraine et al., 2022.

However, this was solved by creating a *de novo* transcriptome which allowed us to align our RNASeq fragments for transcript quantification. One major issue for the *A. mexicanum* was the presence of many paralogs of the same gene. We prefiltered the transcript counts to remove low counts present in all samples (Publication II: Naraine et al., 2022). This would remove many of these paralogs and focus on those that are more significantly expressed within the egg.

It is important to note that despite the detection of many enriched motifs, they were not exclusively observed for a particular profiled set of transcripts. Therefore, similar to other previous reports, there may be multiple factors that determine the localization of the transcript rather than a singular motif. One of these could be the secondary structure or combination of motifs. However, the prediction of the secondary structure can be difficult and prone to variability (as reviewed in the introduction). One of the major limitations is the computational time required to *de novo* predict and align the structure across multiple transcripts. However, newer software are slowly starting to be able to do this a bit more efficiently (Guarracino et al., 2021).

8 Conclusion

The work of this thesis represents the first high throughput transcriptome sequencing analysis at the subcellular level for *X. laevis*, *A. mexicanum* and *A. ruthenus*. Prior to our work, only three RNASeq analysis were done on *X. laevis* and one major one on *D. rerio* (see review in introduction). However, most of these only analyzed the animal and vegetal polar regions. Our findings have shown that the egg is an exceptionally transcript rich environment with many sub-localized profiles. However, many of the localization patterns are not conserved between the models and become even more apparent as you diverge further taxonomically. Regardless of these differences, some conservation is observable. At least seven genes that are known to be associated with the germ plasm are conserved vegetally in all models. Additionally, a larger 339 genes are conserved in their animal localization with functions related to cell cycle regulation. These processes and transcription factors may represent highly conserved mechanisms that are still utilized from an ancestrally common organism. Regardless of the dissimilarity in localization profiles for many transcripts, the observation of similar zipcodes within similarly localized transcripts points to a common conserved mechanism for localization. Additionally, the observation of similar localization patterns through oocyte maturation also points to this potentially evolutionary conserved mechanism. Our work also showed that the transcriptome is still quiet dynamic within the maturing egg and that *de novo* transcriptome and degradation is still actively reshaping the maternal transcript landscape.

The lack of key PGC markers in the *A. mexicanum* points to an establishment of these determinant to their correct location at another stage during embryogenesis. Future studies from our lab have already started to look at the asymmetrical transcript gradients at the 4-cells, 64-cells, and 1000-cells stage for *A. mexicanum*, *A. ruthenus* and several other fish models. Such work will help to track how the observed animal-vegetal gradients become altered after fertilization.

In summary our work enforces the need to study more models for the developmental process as there is a clearly observable difference between related and distant models. Also, the assumption of the lack of asymmetry in models that utilize cellular induction to define their developmental axes may warrant further investigation with the onset of these newer and more advanced sub cellular tools.

9 References

Alarcón, V.B., Elinson, R.P., 2001. RNA anchoring in the vegetal cortex of the Xenopus oocyte. J. Cell Sci. 114, 1731–1741. https://doi.org/10.1242/jcs.114.9.1731

Anderson, G.A., Gelens, L., Baker, J.C., Ferrell, J.E., 2017. Desynchronizing Embryonic Cell Division Waves Reveals the Robustness of Xenopus laevis Development. Cell Rep. 21, 37–46. https://doi.org/10.1016/j.celrep.2017.09.017

Ardekani, A.M., 2010. Review of Different Sequence Motif Finding Algorithms. Avicenna J. Med. Biotechnol. 2.

Asami, M., Lam, B.Y.H., Ma, M.K., Rainbow, K., Braun, S., VerMilyea, M.D., Yeo, G.S.H., Perry, A.C.F., 2022. Human embryonic genome activation initiates at the one-cell stage. Cell Stem Cell 29, 209-216.e4. https://doi.org/10.1016/j.stem.2021.11.012

Bachvarova, R.F., Masi, T., Drum, M., Parker, N., Mason, K., Patient, R., Johnson, A.D., 2004. Gene expression in the Axolotl germ line: Axdazl, Axvh, Axoct-4, and Axkit. Dev. Dyn. 231, 871–880. https://doi.org/10.1002/dvdy.20195

Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27, 1653–1659. https://doi.org/10.1093/BIOINFORMATICS/BTR261

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers.

Ballard, W.W., 1981. Morphogenetic movements and fate maps of vertebrates. Integr. Comp. Biol. 21, 391–399. https://doi.org/10.1093/icb/21.2.391

Bally-Cuif, L., Schatz, W.J., Ho, R.K., 1998. Characterization of the zebrafish Orb/CPEB-related RNA-binding protein and localization of maternal components in the zebrafish oocyte. Mech. Dev. 77, 31–47. https://doi.org/10.1016/S0925-4773(98)00109-9

Barth, J.A., Li, W., Krasinski, S.D., Montgomery, R.K., Verhave, M., Grand, R.J., 1998. Asymmetrical localization of mRNAs in enterocytes of human jejunum. J. Histochem. Cytochem. 46, 335–343. https://doi.org/10.1177/002215549804600307

Beams, H.W., Kessel, R.G., 1974. The Problem of Germ Cell Determinants. Int. Rev. Cytol. 39, 413–479. https://doi.org/10.1016/S0074-7696(08)60944-4

Bebbere, D., Coticchio, G., Borini, A., Ledda, S., 2022. Oocyte aging: looking beyond chromosome segregation errors. J. Assist. Reprod. Genet. 39, 793–800. https://doi.org/10.1007/s10815-022-02441-z

Beckham, Y.M., Nath, K., Elinson, R.P., 2003. Localization of RNAs in oocytes of Eleutherodactylus coqui, a direct developing frog, differs from Xenopus laevis. Evol. Dev. 5, 562–571. https://doi.org/10.1046/j.1525-142X.2003.03061.x

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., Long, R.M., 1998. Localization of ASH1 mRNA particles in living yeast. Mol. Cell 2, 437–445. https://doi.org/10.1016/S1097-2765(00)80143-4

Betley, J.N., Frith, M.C., Graber, J.H., Choo, S., Deshler, J.O., 2002. A ubiquitous and

conserved signal for RNA localization in chordates. Curr. Biol. 12, 1756–1761. https://doi.org/10.1016/S0960-9822(02)01220-4

Betley, J.N., Heinrich, B., Vernos, I., Sardet, C., Prodon, F., Deshler, J.O., 2004. Kinesin II Mediates Vg1 mRNA Transport in Xenopus Oocytes. Curr. Biol. 14, 219–224. https://doi.org/10.1016/j.cub.2004.01.028

Bezawork-Geleta, A., Wen, H., Dong, L., Yan, B., Vider, J., Boukalova, S., Krobova, L., Vanova, K., Zobalova, R., Sobol, M., Hozak, P., Novais, S.M., Caisova, V., Abaffy, P., Naraine, R., Pang, Y., Zaw, T., Zhang, P., Sindelka, R., Kubista, M., Zuryn, S., Molloy, M.P., Berridge, M.V., Pacak, K., Rohlena, J., Park, S., Neuzil, J., 2018. Alternative assembly of respiratory complex II connects energy stress to metabolic checkpoints. Nat. Commun. 9. https://doi.org/10.1038/s41467-018-04603-z

Birsoy, B., Kofron, M., Schaible, K., Wylie, C., Heasman, J., 2006. Vg1 is an essential signaling molecule in Xenopus development. Development 133, 15–20. https://doi.org/10.1242/dev.02144

Blower, M.D., Feric, E., Weis, K., Heald, R., 2007. Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. J. Cell Biol. 179, 1365–1373. https://doi.org/10.1083/jcb.200705163

Blum, M., Ott, T., 2019. Xenopus: An undervalued model organism to study and model human genetic disease. Cells Tissues Organs 205, 303–312. https://doi.org/10.1159/000490898

Boone, M., De Koker, A., Callewaert, N., 2018. Survey and summary capturing the "ome": The expanding molecular toolbox for RNA and DNA library construction. Nucleic Acids Res. 46, 2701–2721. https://doi.org/10.1093/nar/gky167

Bordzilovskaya, N.P., Dettlaff, T.A., 1991. The Axolotl Ambystoma mexicanum. Anim. Species Dev. Stud. 203–230. https://doi.org/10.1007/978-1-4615-3654-3_8

Bronner, I.F., Quail, M.A., 2019. Best Practices for Illumina Library Preparation. Curr. Protoc. Hum. Genet. 102, 1–48. https://doi.org/10.1002/cphg.86

Buxbaum, A.R., Haimovich, G., Singer, R.H., 2014. In the right place at the right time: visualizing and understanding mRNA localization. Nat. Rev. Mol. Cell Biol. 16, 95–109. https://doi.org/10.1038/nrm3918

Carotenuto, R., Vaccaro, M.C., Capriglione, T., Petrucci, T.C., Campanella, C., 2000. alpha-Spectrin has a stage-specific asymmetrical localization during Xenopus oogenesis. Mol. Reprod. Dev. 55, 229–239. https://doi.org/10.1002/(SICI)1098-2795(200002)55:2<229::AID-MRD13>3.0.CO;2-6

Chan, A.P., Kloc, M., Etkin, L.D., 1999. fatvg encodes a new localized RNA that uses a 25-nucleotide element (FVLE1) to localize to the vegetal cortex of Xenopus oocytes. Development 126, 4943–4953. https://doi.org/10.1242/DEV.126.22.4943

Chang, P., Torres, J., Lewis, R.A., Mowry, K.L., Houliston, E., King, M. Lou, 2004. Localization of RNAs to the Mitochondrial Cloud in Xenopus Oocytes through Entrapment and Association with Endoplasmic Reticulum. Mol. Biol. Cell 15, 4669–

4681. https://doi.org/10.1091/mbc.e04-03-0265

Chartrand, P., Meng, X.H., Singer, R.H., Long, R.M., 1999. Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. Curr. Biol. 9, 333–338. https://doi.org/10.1016/S0960-9822(99)80144-4

Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., Qiu, X., Yang, J., Xu, J., Hao, S., Wang, X., Lu, Huifang, Chen, X., Liu, X., Huang, X., Li, Z., Hong, Y., Jiang, Y., Peng, J., Liu, Shuai, Shen, M., Liu, C., Li, Q., Yuan, Y., Wei, Xiaoyu, Zheng, H., Feng, W., Wang, Zhifeng, Liu, Y., Wang, Zhaohui, Yang, Y., Xiang, H., Han, L., Qin, B., Guo, P., Lai, G., Muñoz-Cánoves, P., Maxwell, P.H., Thiery, J.P., Wu, Q.F., Zhao, F., Chen, B., Li, M., Dai, X., Wang, S., Kuang, H., Hui, J., Wang, L., Fei, J.F., Wang, O., Wei, Xiaofeng, Lu, Haorong, Wang, B., Liu, Shiping, Gu, Y., Ni, M., Zhang, W., Mu, F., Yin, Y., Yang, H., Lisby, M., Cornall, R.J., Mulder, J., Uhlén, M., Esteban, M.A., Li, Y., Liu, L., Xu, X., Wang, J., 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell 185, 1777-1792.e21. https://doi.org/10.1016/j.cell.2022.04.003

Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., Tyler, J.K., 2016. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Mol. Cell. Biol. 36, 662–667. https://doi.org/10.1128/MCB.00970-14

Chen, Li, Yang, R., Kwan, T., Tang, C., Watt, S., Zhang, Y., Bourque, G., Ge, B., Downes, K., Frontini, M., Ouwehand, W.H., Lin, J. wen, Soranzo, N., Pastinen, T., Chen, Lu, 2020. Paired rRNA-depleted and polyA-selected RNA sequencing data and supporting multi-omics data from human T cells. Sci. Data 7, 1–7. https://doi.org/10.1038/s41597-020-00719-4

Clack, J.A., 2009. The Fish-Tetrapod Transition: New Fossils and Interpretations. Evol. Educ. Outreach 2, 213–223. https://doi.org/10.1007/s12052-009-0119-2

Claußen, M., Horvay, K., Pieler, T., 2004. Evidence for overlapping, but not identical, protein machineries operating in vegetal RNA localization along early and late pathways in Xenopus oocytes. Development 131, 4263–4273. https://doi.org/10.1242/dev.01283

Claußen, M., Lingner, T., Pommerenke, C., Opitz, L., Salinas, G., Pieler, T., 2015. Global analysis of asymmetric RNA enrichment in oocytes reveals low conservation between closely related Xenopus species. Mol. Biol. Cell 26, 3777–3787. https://doi.org/10.1091/mbc.E15-02-0115

Cuykendall, T.N., Houston, D.W., 2010. Identification of germ plasm-associated transcripts by microarray analysis of Xenopus vegetal cortex RNA. Dev. Dyn. 239, 1838–1848. https://doi.org/10.1002/dvdy.22304

De Domenico, E., Owens, N.D.L., Grant, I.M., Gomes-Faria, R., Gilchrist, M.J., 2015. Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing. Dev. Biol. 408, 252–268. https://doi.org/10.1016/j.ydbio.2015.06.010

De Robertis, E.M., Larraín, J., Oelgeschläger, M., Wessely, O., 2000. The establishment of spemann's organizer and patterning of the vertebrate embryo. Nat. Rev. Genet. 1, 171–

181. https://doi.org/10.1038/35042039

Deshler, J.O., Highett, M.I., Abramson, T., Schnapp, B.J., 1998. A highly conserved RNA-binding protein for cytoplasmic mRNA localization in vertebrates. Curr. Biol. 8, 489–496. https://doi.org/10.1016/S0960-9822(98)70200-3

Deshler, J.O., Highett, M.I., Schnapp, B.J., 1997. Localization of Xenopus Vg1 mRNA by Vera Protein and the Endoplasmic Reticulum. Science (80-.). 276, 1128–1131. https://doi.org/10.1126/science.276.5315.1128

Devic, E., Paquereau, L., Rizzoti, K., Monier, A., Knibiehler, B., Audigier, Y., 1996. The mRNA encoding a β subunit of heterotrimeric GTP-binding proteins is localized to the animal pole of Xenopus laevis oocyte and embryos. Mech. Dev. 59, 141–151. https://doi.org/10.1016/0925-4773(96)00588-6

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N.S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 14, 671–683. https://doi.org/10.1093/bib/bbs046

Drummond, D.R., McCrae, M.A., Colman, A., 1985. Stability and movement of mRNAs and their encoded proteins in xenopus oocytes. J. Cell Biol. 100, 1148–1156. https://doi.org/10.1083/jcb.100.4.1148

Du, K., Stöck, M., Kneitz, S., Klopp, C., Woltering, J.M., Adolfi, M.C., Feron, R., Prokopov, D., Makunin, A., Kichigin, I., Schmidt, C., Fischer, P., Kuhl, H., Wuertz, S., Gessner, J., Kloas, W., Cabau, C., Iampietro, C., Parrinello, H., Tomlinson, C., Journot, L., Postlethwait, J.H., Braasch, I., Trifonov, V., Warren, W.C., Meyer, A., Guiguen, Y., Schartl, M., 2020. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. Nat. Ecol. Evol. 4, 841–852. https://doi.org/10.1038/s41559-020-1166-x

Elinson, R.P., 2009. Nutritional endoderm: A way to breach the holoblastic-meroblastic barrier in tetrapods. J. Exp. Zool. Part B Mol. Dev. Evol. 312, 526–532. https://doi.org/10.1002/jez.b.21218

Elinson, R.P., 2006. Cleavage and Gastrulation in Xenopus laevis Embryos . eLS. https://doi.org/10.1038/npg.els.0004222

Elinson, R.P., del Pino, E.M., 2012. Developmental diversity of amphibians. Wiley Interdiscip. Rev. Dev. Biol. https://doi.org/10.1002/wdev.23

Elinson, R.P., King, M. Lou, Forristall, C., 1993. Isolated vegetal cortex from Xenopus oocytes selectively retains localized mRNAs. Dev. Biol. 160, 554–562. https://doi.org/10.1006/dbio.1993.1329

Elinson, R.P., Sabo, M.C., Fisher, C., Yamaguchi, T., Orii, H., Nath, K., 2011. Germ plasm in Eleutherodactylus coqui, a direct developing frog with large eggs. Evodevo 2, 20. https://doi.org/10.1186/2041-9139-2-20

Ephrussi, A., Lehmann, R., 1992. Induction of germ cell formation by oskar. Nature 358, 387–392. https://doi.org/10.1038/358387a0

Evans, C., Hardin, J., Stoebel, D.M., 2018. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief. Bioinform. 19, 776–792. https://doi.org/10.1093/bib/bbx008

Fazal, F.M., Chang, H.Y., 2019. Subcellular Spatial Transcriptomes: Emerging Frontier for Understanding Gene Regulation. Cold Spring Harb. Symp. Quant. Biol. 84, 31–45. https://doi.org/10.1101/sqb.2019.84.040352

Flachsova, M., Sindelka, R., Kubista, M., 2013. Single blastomere expression profiling of Xenopus laevis embryos of 8 to 32-cells reveals developmental asymmetry. Sci. Rep. 3, 1–6. https://doi.org/10.1038/srep02278

Forristall, C., Pondel, M., Chen, L., King, M.L., 1995. Patterns of localization and cytoskeletal association of two vegetally localized RNAs, Vg1 and Xcat-2. Development 121, 201–208. https://doi.org/10.1242/dev.121.1.201

Gautreau, D., Cote, C.A., Mowry, K.L., 1997. Two copies of a subelement from the Vg1 RNA localization sequence are sufficient to direct vegetal localization in Xenopus oocytes. Development 124, 5013–5020. https://doi.org/10.1242/dev.124.24.5013

Gavis, E.R., 1997. Expeditions to the pole: RNA localization in Xenopus and Drosophila. Trends Cell Biol. 7, 485–492. https://doi.org/10.1016/S0962-8924(97)01162-8

Guarracino, A., Pepe, G., Ballesio, F., Adinolfi, M., Pietrosanto, M., Sangiovanni, E., Vitale, I., Ausiello, G., Helmer-Citterich, M., 2021. BRIO: A web server for RNA sequence and structure motif scan. Nucleic Acids Res. 49, W67–W71. https://doi.org/10.1093/nar/gkab400

Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., Zavolan, M., 2021. CLIP and complementary methods. Nat. Rev. Methods Prim. 1. https://doi.org/10.1038/s43586-021-00018-1

Hamilton, R.S., Davis, I., 2011. Identifying and Searching for Conserved RNA Localisation Signals, in: Methods. pp. 447–466. https://doi.org/10.1007/978-1-61779-005-8_27

Han, J.W., Park, J.H., Kim, M., Lee, J.H., 1997. mRNAs for microtubule proteins are specifically colocalized during the sequential formation of basal body, flagella, and cytoskeletal microtubules in the differentiation of Naegleria gruberi. J. Cell Biol. 137, 871–879. https://doi.org/10.1083/jcb.137.4.871

Harry, N.D., Zakas, C., 2023. Maternal patterns of inheritance alter transcript expression in eggs. BMC Genomics 24, 191. https://doi.org/10.1186/s12864-023-09291-8

Heasman, J., Quarmby, J., Wylie, C.C., 1984. The mitochondrial cloud of Xenopus oocytes: The source of germinal granule material. Dev. Biol. 105, 458–469. https://doi.org/10.1016/0012-1606(84)90303-8

Hinkley, C.S., Martin, J.F., Leibham, D., Perry, M., 1992. Sequential expression of multiple POU proteins during amphibian early development. Mol. Cell. Biol. 12, 638–649. https://doi.org/10.1128/mcb.12.2.638-649.1992

Hofacker, I.L., Priwitzer, B., Stadler, P.F., 2004. Prediction of locally stable RNA secondary structures for genome-wide surveys. Bioinformatics 20, 186–190. https://doi.org/10.1093/bioinformatics/btg388

Holler, K., Neuschulz, A., Drewe-Boß, P., Mintcheva, J., Spanjaard, B., Arsiè, R., Ohler, U., Landthaler, M., Junker, J.P., 2021. Spatio-temporal mRNA tracking in the early zebrafish embryo. Nat. Commun. 12, 1–13. https://doi.org/10.1038/s41467-021-23834-1

Horder, T., 2010. History of Developmental Biology. eLS 1–12. https://doi.org/10.1002/9780470015902.a0003080.pub2

Houston, D.W., 2017. Vertebrate axial patterning: From egg to asymmetry, Advances in Experimental Medicine and Biology. https://doi.org/10.1007/978-3-319-46095-6_6

Houston, D.W., 2013. Regulation of cell polarity and RNA localization in vertebrate oocytes, in: International Review of Cell and Molecular Biology. Elsevier Inc., pp. 127–185. https://doi.org/10.1016/B978-0-12-407694-5.00004-3

Houston, D.W., Zhang, J., Maines, J.Z., Wasserman, S.A., King, M. Lou, 1998. A Xenopus DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of Drosophila boule. Development 125, 171–180. https://doi.org/10.1242/dev.125.2.171

Howley, C., Ho, R.K., 2000. mRNA localization patterns in zebrafish oocytes. Mech. Dev. 92, 305–309. https://doi.org/10.1016/S0925-4773(00)00247-1

Huang, L., Zhang, H., Deng, D., Zhao, K., Liu, K., Hendrix, D.A., Mathews, D.H., 2019. LinearFold: Linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search. Bioinformatics 35, i295–i304. https://doi.org/10.1093/bioinformatics/btz375

Hudson, C., Woodland, H.R., 1998. Xpat, a gene expressed specifically in germ plasm and primordial germ cells of Xenopus laevis. Mech. Dev. 73, 159–168. https://doi.org/10.1016/S0925-4773(98)00047-1

Hudson, J.W., Alarcón, V.B., Elinson, R.P., 1996. Identification of new localized RNAs in the Xenopus oocyte by differential display PCR. Dev. Genet. 19, 190–198. https://doi.org/10.1002/(SICI)1520-6408(1996)19:3<190::AID-DVG2>3.0.CO;2-4

Iegorova, V., Naraine, R., Psenicka, M., Zelazowska, M., Sindelka, R., 2022. Comparison of RNA localization during oogenesis within Acipenser ruthenus and Xenopus laevis. Front. Cell Dev. Biol. 10, 1–16. https://doi.org/10.3389/fcell.2022.982732

Jansen, R.P., 2001. mRNA localization: message on the move. Nat. Rev. Mol. Cell Biol. 2, 247–256. https://doi.org/10.1038/35067016

Jia, C., Carson, M.B., Wang, Y., Lin, Y., Lu, H., 2014. A new exhaustive method and strategy for finding motifs in ChIP-enriched regions. PLoS One 9. https://doi.org/10.1371/journal.pone.0086044

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., Oliver, B., 2011. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 21, 1543–1551. https://doi.org/10.1101/gr.121095.111

Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., Luo, Y., 2022. Single-cell RNA sequencing technologies and applications: A brief overview. Clin. Transl. Med. 12. https://doi.org/10.1002/ctm2.694

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., Van Oudenaarden, A., 2014. Genome-wide RNA Tomography in the Zebrafish Embryo. Cell 159, 662–675. https://doi.org/10.1016/j.cell.2014.09.038

Kalinka, A.T., Tomancak, P., 2012. The evolution of early animal embryos: Conservation or divergence? Trends Ecol. Evol. 27, 385–393. https://doi.org/10.1016/j.tree.2012.03.007

Kaltimbacher, V., Bonnet, C., Lecoeuvre, G., Forster, V., Sahel, J.A., Corral-Debrinski, M., 2006. mRNA localization to the mitochondrial surface allows the efficient translocation inside the organelle of a nuclear recoded ATP6 protein. Rna 12, 1408–1417. https://doi.org/10.1261/rna.18206

Keinath, M.C., Timoshevskiy, V.A., Timoshevskaya, N.Y., Tsonis, P.A., Voss, S.R., Smith, J.J., 2015. Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. Sci. Reports 2015 51 5, 1–13. https://doi.org/10.1038/srep16413

Kiecker, C., Bates, T., Bell, E., 2015. Molecular specification of germ layers in vertebrate embryos. Cell. Mol. Life Sci. 2015 735 73, 923–947. https://doi.org/10.1007/S00018-015-2092-Y

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., Schilling, T.F., 1995. Stages of embryonic development of the zebrafish. Dev. Dyn. 203, 253–310. https://doi.org/10.1002/aja.1002030302

King, M. Lou, Messitt, T.J., Mowry, K.L., 2005. Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. Biol. Cell 97, 19–33. https://doi.org/10.1042/BC20040067

King, M. Lou, Zhou, Y., Bubunenko, M., 1999. Polarizing genetic information in the egg: RNA localization in the frog oocyte. BioEssays 21, 546–557. https://doi.org/10.1002/(SICI)1521-1878(199907)21:7<546::AID-BIES3>3.0.CO;2-Z

Kloc, M., 2014. Polarity of Xenopus Oocytes and Early Embryos, in: Xenopus Development. John Wiley & Sons, Inc, Oxford, pp. 60–74. https://doi.org/10.1002/9781118492833.ch4

Kloc, M., Bilinski, S., Etkin, L.D., 2004. The Balbiani Body and Germ Cell Determinants: 150 Years Later. Curr. Top. Dev. Biol. 59, 1–36. https://doi.org/10.1016/S0070-2153(04)59001-4

Kloc, M., Dougherty, M.T., Bilinski, S., Chan, A.P., Brey, E., King, M. Lou, Patrick, C.W., Etkin, L.D., 2002. Three-dimensional ultrastructural analysis of RNA distribution within germinal granules of Xenopus. Dev. Biol. 241, 79–93. https://doi.org/10.1006/dbio.2001.0488

Kloc, M., Etkin, L.D., 2005. RNA localization mechanisms in oocytes. J. Cell Sci. https://doi.org/10.1242/jcs.01637

Kloc, M., Etkin, L.D., 1995. Two distinct pathways for the localization of RNAs at the vegetal cortex in Xenopus oocytes. Development 121, 287–297. https://doi.org/10.1242/DEV.121.2.287

Kloc, M., Kubiak, J.Z., 2014. Xenopus Development, Xenopus Development. https://doi.org/10.1002/9781118492833

Kloc, M., Larabell, C., Chan, A.P.Y., Etkin, L.D., 1998. Contribution of METRO pathway localized molecules to the organization of the germ cell lineage. Mech. Dev. 75, 81–93. https://doi.org/10.1016/S0925-4773(98)00086-0

Kloc, M., Reddy, B.A., Miller, M., Eastman, E., Etkin, L.D., 1991. xl21: A localized maternal transcript in Xenopus laevis. Mol. Reprod. Dev. 28, 341–345. https://doi.org/10.1002/mrd.1080280405

Kloc, M., Spohr, G., Etkin, L.D., 1993. Translocation of repetitive RNA sequences with the germ plasm in Xenopus oocytes. Science (80-.). 262, 1712–1714. https://doi.org/10.1126/science.7505061

Kosaka, K., Kawakami, K., Sakamoto, H., Inoue, K., 2007. Spatiotemporal localization of germ plasm RNAs during zebrafish oogenesis. Mech. Dev. 124, 279–289. https://doi.org/10.1016/J.MOD.2007.01.003

Krishnakumar, P., Riemer, S., Perera, R., Lingner, T., Goloborodko, A., Khalifa, H., Bontems, F., Kaufholz, F., El-Brolosy, M.A., Dosch, R., 2018. Functional equivalence of germ plasm organizers. PLoS Genet. 14, 1–29. https://doi.org/10.1371/journal.pgen.1007696

Kroll, T.T., Swenson, L.B., Hartland, E.I., Snedden, D.D., Goodson, H. V, Huber, P.W., 2009. Interactions of 40LoVe within the ribonucleoprotein complex that forms on the localization element of Xenopus Vg1 mRNA. Mech. Dev. 126, 523–538. https://doi.org/10.1016/j.mod.2009.03.007

Ku, M., Melton, D.A., 1993. Xwnt-11: a maternally expressed Xenopus wnt gene. Development 119, 1161–1173. https://doi.org/10.1242/dev.119.4.1161

Lambert, A., Fontaine, J.F., Legendre, M., Leclerc, F., Permal, E., Major, F., Putzer, H., Delfour, O., Michot, B., Gautheret, D., 2004. The ERPIN server: An interface to profile-based RNA motif identification. Nucleic Acids Res. 32, 160–165. https://doi.org/10.1093/nar/gkh418

Levy, S.E., Myers, R.M., 2016. Advancements in Next-Generation Sequencing. Annu. Rev. Genomics Hum. Genet. 17, 95–115. https://doi.org/10.1146/annurev-genom-083115-022413

Lewis, R.A., Kress, T.L., Cote, C.A., Gautreau, D., Rokop, M.E., Mowry, K.L., 2004. Conserved and clustered RNA recognition sequences are a critical feature of signals directing RNA localization in Xenopus oocytes. Mech. Dev. 121, 101–109. https://doi.org/10.1016/j.mod.2003.09.009

Li, D., Zand, M.S., Dye, T.D., Goniewicz, M.L., Rahman, I., Xie, Z., 2022. An evaluation of RNA-seq differential analysis methods. PLoS One 17, 1–19.

https://doi.org/10.1371/journal.pone.0264246

Link, B.A., Megason, S.G., 2008. Zebrafish as a model for development. Source B. Model. Biomed. Res. 103–112. https://doi.org/10.1007/978-1-59745-285-4_13

Liu, X., 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Symp. A Q. J. Mod. Foreign Lit. 138, 127–138. https://doi.org/10.1142/9789814447362_0014

Lopata, A., 2009. History of the egg in embryology. J. Mamm. Ova Res. 26, 2–9. https://doi.org/10.1274/jmor.26.2

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., Young, R.A., 2012. Revisiting Global Gene Expression Analysis. Cell 151, 476–482. https://doi.org/10.1016/j.cell.2012.10.012

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., Cai, L., 2014. Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361. https://doi.org/10.1038/nmeth.2892

Macdonald, P.M., Kerr, K., 1997. Redundant RNA recognition events in bicoid mRNA localization. Rna 3, 1413–1420.

Maegawa, S., Yasuda, K., Inoue, K., 1999. Maternal mRNA localization of zebrafish DAZ-like gene. Mech. Dev. 81, 223–226. https://doi.org/10.1016/S0925-4773(98)00242-1

Martin, K.C., Ephrussi, A., 2009. mRNA Localization: Gene Expression in the Spatial Dimension. Cell 136, 719–730. https://doi.org/10.1016/j.cell.2009.01.044

Melton, D.A., 1988. Translocation of a localized maternal mRNA to the vegetal pole of xenopus oocytes. Nature 328, 80–82. https://doi.org/10.1038/328080a0

Messitt, T.J., Gagnon, J.A., Kreiling, J.A., Pratt, C.A., Yoon, Y.J., Mowry, K.L., 2008. Multiple Kinesin Motors Coordinate Cytoplasmic RNA Transport on a Subpopulation of Microtubules in Xenopus Oocytes. Dev. Cell 15, 426–436. https://doi.org/10.1016/j.devcel.2008.06.014

Mignone, F., Gissi, C., Liuni, S., Pesole, G., others, 2002. Untranslated regions of mRNAs. Genome Biol 3, 0004–1. https://doi.org/10.1186/gb-2002-3-3-reviews0004

Moor, A.E., Golan, M., Massasa, E.E., Lemze, D., Weizman, T., Shenhav, R., Baydatch, S., Mizrahi, O., Winkler, R., Golani, O., Stern-Ginossar, N., Itzkovitz, S., 2017. Global mRNA polarization regulates translation efficiency in the intestinal epithelium. Science (80-.). 357, 1299–1303. https://doi.org/10.1126/science.aan2399

Mosquera, L., Forristall, C., Zhou, Y., King, M.L., 1993. A mRNA localized to the vegetal cortex of Xenopus oocytes encodes a protein with a nanos-like zinc finger domain. Development 117, 377–386. https://doi.org/10.1242/dev.117.1.377

Mowry, K.L., Melton, D.A., 1992. Vegetal messenger RNA localization directed by a 340-nt

RNA sequence element in Xenopus oocytes. Science (80-.). 255, 991–994. https://doi.org/10.1126/science.1546297

Naraine, R., Abaffy, P., Sidova, M., Tomankova, S., Pocherniaieva, K., Smolik, O., Kubista, M., Psenicka, M., Sindelka, R., 2020. NormQ: RNASeq normalization based on RT-qPCR derived size factors. Comput. Struct. Biotechnol. J. 18, 1173–1181. https://doi.org/10.1016/j.csbj.2020.05.010

Naraine, R., Iegorova, V., Abaffy, P., Franek, R., Soukup, V., Psenicka, M., Sindelka, R., 2022. Evolutionary conservation of maternal RNA localization in fishes and amphibians revealed by TOMO-Seq. Dev. Biol. 489, 146–160. https://doi.org/10.1016/j.ydbio.2022.06.013

Nath, K., Boorech, J.L., Beckham, Y.M., Burns, M.M., Elinson, R.P., 2005. Status of RNAs, localized in Xenopus laevis oocytes, in the frogs Rana pipiens and Eleutherodactylus coqui. J. Exp. Zool. Part B Mol. Dev. Evol. 304, 28–39. https://doi.org/10.1002/jez.b.21020

Nath, K., Elinson, R.P., 2007. RNA of AmVegT, the axolotl orthologue of the Xenopus meso-endodermal determinant, is not localized in the oocyte. Gene Expr. Patterns 7, 197–201. https://doi.org/10.1016/j.modgep.2006.07.003

Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., Amster-Choder, O., 2011. Translation-independent localization of mRNA in E. coli. Science (80-.). 331, 1081–1084. https://doi.org/10.1126/science.1195691

Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., Falcon, F., Knapp, D., Powell, S., Cruz, A., Cao, H., Habermann, B., Hiller, M., Tanaka, E.M., Myers, E.W., 2018. The axolotl genome and the evolution of key tissue formation regulators. Nature 554, 50–55. https://doi.org/10.1038/nature25458

Ntostis, P., Iles, D., Kokkali, G., Vaxevanoglou, T., Kanavakis, E., Pantou, A., Huntriss, J., Pantos, K., Picton, H.M., 2022. The impact of maternal age on gene expression during the GV to MII transition in euploid human oocytes. Hum. Reprod. 37, 80–92. https://doi.org/10.1093/humrep/deab226

Oberman, F., Yisraeli, J.K., 1995. Two protocols for nonradioactive in situ hybridization to Xenopus oocytes. Trends Genet. https://doi.org/10.1016/S0168-9525(00)89006-X

Owens, D.A., Butler, A.M., Aguero, T.H., Newman, K.M., Van Booven, D., King, M. Lou, 2017. High-throughput analysis reveals novel maternal germline RNAs crucial for primordial germ cell preservation and proper migration. Development 144, 292–304. https://doi.org/10.1242/dev.139220

Paige, J.S., Wu, K.Y., Jaffrey, S.R., 2011. RNA Mimics of Green Fluorescent Protein. Science (80-.). 333, 642–646. https://doi.org/10.1126/science.1207339

Palacios, I.M., Johnston, D.S.T., 2001. Getting the message across: The intracellular localization of mRNAs in higher eukaryotes. Annu. Rev. Cell Dev. Biol. 17, 569–614. https://doi.org/10.1146/annurev.cellbio.17.1.569

Passmore, L.A., Coller, J., 2022. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nat. Rev. Mol. Cell Biol. 23, 93–106. https://doi.org/10.1038/s41580-021-00417-y

Pavesi, G., Mereghetti, P., Mauri, G., Pesole, G., 2004. Weeder web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res. 32, 199–203. https://doi.org/10.1093/nar/gkh465

Pelegri, F., 2003. Maternal Factors in Zebrafish Development. Dev. Dyn. 228, 535–554. https://doi.org/10.1002/dvdy.10390

Pondel, M.D., King, M.L., 1988. Localized maternal mRNA related to transforming growth factor beta mRNA is concentrated in a cytokeratin-enriched fraction from Xenopus oocytes. Proc. Natl. Acad. Sci. 85, 7612–7616. https://doi.org/10.1073/pnas.85.20.7612

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., Tyagi, S., 2008. Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5, 877–879. https://doi.org/10.1038/nmeth.1253

Ramanathan, M., Porter, D.F., Khavari, P.A., Alto, P., 2019. Methods to study RNA-protein interactions 16, 225–234. https://doi.org/10.1038/s41592-019-0330-1.Methods

Rand, K., Yisraeli, J., 2001. RNA Localization in Xenopus Oocytes, in: Current Opinion in Microbiology. pp. 157–173. https://doi.org/10.1007/978-3-540-40025-7_10

Rebagliati, M.R., Weeks, D.L., Harvey, R.P., Melton, D.A., 1985. Identification and cloning of localized maternal RNAs from xenopus eggs. Cell 42, 769–777. https://doi.org/10.1016/0092-8674(85)90273-9

Reddy, B.A., Kloc, M., Etkin, L.D., 1992. The cloning and characterization of a localized maternal transcript in Xenopus laevis whose zygotic counterpart is detected in the CNS. Mech. Dev. 39, 143–150. https://doi.org/10.1016/0925-4773(92)90041-H

Romanenko, S.A., Biltueva, L.S., Serdyukova, N.A., Kulemzina, A.I., Beklemisheva, V.R., Gladkikh, O.L., Lemskaya, N.A., Interesova, E.A., Korentovich, M.A., Vorobieva, N. V., Graphodatsky, A.S., Trifonov, V.A., 2015. Segmental paleotetraploidy revealed in sterlet (Acipenser ruthenus) genome by chromosome painting. Mol. Cytogenet. 8, 1–13. https://doi.org/10.1186/s13039-015-0194-8

Schnapp, B.J., Arn, E.A., Deshler, J.O., Highett, M.I., 1997. RNA localization in Xenopus oocytes. Semin. Cell Dev. Biol. 8, 529–540. https://doi.org/10.1006/scdb.1997.0178

Schroeder, K.E., Yost, H.J., 1996. Xenopus poly (A) binding protein maternal RNA is localized during oogenesis and associated with large complexes in blastula. Dev. Genet. 19, 268–276. https://doi.org/10.1002/(SICI)1520-6408(1996)19:3<268::AID-DVG10>3.0.CO;2-W

Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S.J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J.B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S.S., van Kruijsbergen, I., Shu, S., Carlson, J., Kinoshita,

T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S. V., Suzuki, Y., Haramoto, Y., Yamamoto, T.S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D.S., Flajnik, M.F., Houston, D.W., Shendure, J., DuPasquier, L., Vize, P.D., Zorn, A.M., Ito, M., Marcotte, E.M., Wallingford, J.B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G.J.C., Fujiyama, A., Harland, R.M., Taira, M., Rokhsar, D.S., 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538, 336–343. https://doi.org/10.1038/nature19840

Sharov, A.A., Ko, M.S.H., 2009. Exhaustive search for over-represented DNA sequence motifs with cisfinder. DNA Res. 16, 261–273. https://doi.org/10.1093/dnares/dsp014

Sharp, J.A., Plant, J.J., Ohsumi, T.K., Borowsky, M., Blower, M.D., 2011. Functional analysis of the microtubule-interacting transcriptome. Mol. Biol. Cell 22, 4312–4323. https://doi.org/10.1091/mbc.E11-07-0629

Sindelka, R., Abaffy, P., Qu, Y., Tomankova, S., Sidova, M., Naraine, R., Kolar, M., Peuchen, E., Sun, L., Dovichi, N., Kubista, M., 2018. Asymmetric distribution of biomolecules of maternal origin in the Xenopus laevis egg and their impact on the developmental plan. Sci. Reports 2018 81 8, 1–16. https://doi.org/10.1038/s41598-018-26592-1

Sindelka, R., Jonák, J., Hands, R., Bustin, S.A., Kubista, M., 2008. Intracellular expression profiles measured by real-time PCR tomography in the Xenopus laevis oocyte. Nucleic Acids Res. 36, 387–392. https://doi.org/10.1093/nar/gkm1024

Sindelka, R., Sidova, M., Svec, D., Kubista, M., 2010. Spatial expression profiles in the Xenopus laevis oocytes measured with qPCR tomography. Methods 51, 87–91. https://doi.org/10.1016/j.ymeth.2009.12.011

Speybroeck, L., Waele, D., Vijver, G., 2006. Theories in Early Embryology. Ann. N. Y. Acad. Sci. 981, 7–49. https://doi.org/10.1111/j.1749-6632.2002.tb04910.x

Steward, O., Wallace, C.S., Lyford, G.L., Worley, P.F., 1998. Synaptic activation causes the mRNA for the leg Arc to localize selectively near activated postsynaptic sites on dendrites. Neuron 21, 741–751. https://doi.org/10.1016/S0896-6273(00)80591-7

Taliaferro, J.M., 2022. Transcriptome-scale methods for uncovering subcellular RNA localization mechanisms. Biochim. Biophys. Acta - Mol. Cell Res. 1869, 119202. https://doi.org/10.1016/j.bbamcr.2021.119202

Taliaferro, J.M., 2019. Classical and emerging techniques to identify and quantify localized RNAs. Wiley Interdiscip. Rev. RNA 10. https://doi.org/10.1002/wrna.1542

True, J.R., Haag, E.S., 2001. Developmental system drift and flexibility in evolutionary trajectories. Evol. Dev. 3, 109–119. https://doi.org/10.1046/j.1525-142X.2001.003002109.x

Vallée, M., Aiba, K., Piao, Y., Palin, M.F., Ko, M.S.H., Sirard, M.A., 2008. Comparative analysis of oocyte transcript profiles reveals a high degree of conservation among species. Reproduction 135, 439–448. https://doi.org/10.1530/REP-07-0342

Van De Peer, Y., Maere, S., Meyer, A., 2009. The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10, 725–732. https://doi.org/10.1038/nrg2600

Vicens, Q., Kieft, J.S., 2022. Thoughts on how to think (and talk) about RNA structure. Proc. Natl. Acad. Sci. U. S. A. 119, 1–9. https://doi.org/10.1073/pnas.2112677119

Voss, S.R., Epperlein, H.H., Tanaka, E.M., 2009. Ambystoma mexicanum, the axolotl: A versatile amphibian model for regeneration, development, and evolution studies. Cold Spring Harb. Protoc. 4. https://doi.org/10.1101/pdb.emo128

Wang, M., Hu, Q., Lv, T., Wang, Y., Lan, Q., Xiang, R., Tu, Z., Wei, Y., Han, K., Shi, C., Guo, J., Liu, C., Yang, T., Du, W., An, Y., Cheng, M., Xu, J., Lu, H., Li, W., Zhang, S., Chen, A., Chen, W., Li, Y., Wang, X., Xu, X., Hu, Y., Liu, L., 2022. High-resolution 3D spatiotemporal transcriptomic maps of developing Drosophila embryos and larvae. Dev. Cell 57, 1271-1283.e4. https://doi.org/10.1016/j.devcel.2022.04.006

Winata, C.L., Korzh, V., 2018. The translational regulation of maternal mRNAs in time and space. FEBS Lett. 592, 3007–3023. https://doi.org/10.1002/1873-3468.13183

Wu, J., Jaffrey, S.R., 2020. Imaging mRNA trafficking in living cells using fluorogenic proteins. Curr. Opin. Chem. Biol. 57, 177–183. https://doi.org/10.1016/j.cbpa.2020.07.007

Xia, C., Fan, J., Emanuel, G., Hao, J., Zhuang, X., 2019. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proc. Natl. Acad. Sci. U. S. A. 116, 19490–19499. https://doi.org/10.1073/pnas.1912459116

Xu, Q., Zhang, X., 2016. The influence of the global gene expression shift on downstream analyses. PLoS One 11, 1–13. https://doi.org/10.1371/journal.pone.0153903

Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., Chen, L.L., 2011. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 12, 1–14. https://doi.org/10.1186/gb-2011-12-2-r16

Yisraeli, J.K., Sokol, S., Melton, D.A., 1990. A two-step model for the localization of maternal mRNA in Xenopus oocytes: Involvement of microtubules and microfilaments in the translocation and anchoring of Vg1 mRNA. Development 108, 289–298. https://doi.org/10.1242/dev.108.2.289

Yu, Q., Huo, H., Chen, X., Guo, H., Vitter, J.S., Huan, J., 2015. An Efficient Algorithm for Discovering Motifs in Large DNA Data Sets. IEEE Trans. Nanobioscience 14, 535–544. https://doi.org/10.1109/TNB.2015.2421340

Zhang, J., King, M. Lou, 1996. Xenopus VegTRNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. Development 122, 4119–4129. https://doi.org/10.1242/dev.122.12.4119

Zhang, Q., Yaniv, K., Oberman, F., Wolke, U., Git, A., Fromer, M., Taylor, W.L., Meyer, D., Standart, N., Raz, E., Yisraeli, J.K., 1999. Vg1 RBP intracellular distribution and evolutionarily conserved expression at multiple stages during development. Mech. Dev. 88, 101–106. https://doi.org/10.1016/S0925-4773(99)00162-8

Zhao, W.M., Jiang, C., Kroll, T.T., Huber, P.W., 2001. A proline-rich protein binds to the localization element of Xenopus Vg1 mRNA and to ligands involved in actin polymerization. EMBO J. 20, 2315–2325. https://doi.org/10.1093/emboj/20.9.2315

Zhou, Y., King, M. Lou, 1996. Localization of Xcat-2 RNA, a putative germ plasm component, to the mitochondrial cloud in Xenopus stage I oocytes. Development 122, 2947–2953. https://doi.org/10.1242/dev.122.9.2947

Zivraj, K.H., Tung, Y.C.L., Piper, M., Gumy, L., Fawcett, J.W., Yeo, G.S.H., Holt, C.E., 2010. Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs. J. Neurosci. 30, 15464–15478. https://doi.org/10.1523/JNEUROSCI.1800-10.2010