High Through-put Analysis Reveals Novel Maternal Germline RNAs Critical for PGC Survival, Migration, and Differentiation

Dawn Owens
University of Miami, dagant132@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation
https://scholarlyrepository.miami.edu/oa_dissertations/2004

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
HIGH THROUGH-PUT ANALYSIS REVEALS NOVEL MATERNAL GERMLINE RNAS CRITICAL FOR PGC SURVIVAL, MIGRATION, AND DIFFERENTIATION

By

Dawn Owens

A DISSERTATION

Submitted to the Faculty of the University of Miami in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Coral Gables, Florida

December 2017
UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

HIGH THROUGH-PUT ANALYSIS REVEALS NOVEL MATERNAL GERMLINE
RNAs CRITICAL FOR PGC SURVIVAL, MIGRATION, AND DIFFERENTIATION

Dawn Owens

Approved:

Mary Lou King, Ph.D.
Professor of Cell Biology and Anatomy

Athula Wikramanayake, Ph.D.
Professor and Chair of Biology

David Robbins, Ph.D.
Professor of Surgery

Carlos Moraes, Ph.D.
Professor of Neuroscience

Micheal Sheets, Ph.D.
Professor of Biomolecular Chemistry
University of Wisconsin-Madison

Guillermo Prado, Ph.D.
Dean of Graduate School
Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Mary Lou King.
No. of pages in text. (119)

Xenopus laevis embryonic development requires the selective localization of maternal RNAs to the animal or vegetal pole. These RNAs include somatic and germline determinants, essential in establishing the embryo’s body plan. Such determinants are involved in primary germ layer formation and organization, dorsal/ventral axis specification, anterior/posterior axis specification, and germline development. In Xenopus laevis, germline development requires maternally inherited germplasm, a vegetally localized material that both specifies and protects primordial germ cells (PGCs) from adopting somatic fates throughout embryogenesis. The germline is unique in its ability to maintain heritable genetic material for many metazoans, including humans. It is, therefore, of great interest to understand what keeps the integrity of this cell population from becoming somatic cells. Previous studies utilizing microarray have confidently identified localized RNAs in the Xenopus laevis oocyte, but these studies were limited to known transcripts only. Therefore, we hypothesized that using high through-put RNA sequencing would identify novel vegetally localized transcripts that are required for PGC specification.
In this dissertation, I present substantial computational data derived from our RNA-seq analysis on fully grown *Xenopus laevis* oocytes, revealing 411 (198 annotated) and 27 (15 annotated) mRNAs as being enriched at the vegetal and animal pole, respectively. Of the vegetally and animally enriched mRNAs, ninety were novel and over 4-fold enriched and six were over 10-fold enriched at the vegetal and animal poles, respectively. In addition to mRNAs, identified microRNAs were found to not be localized within our oocyte samples. We also show that within the vegetally enriched transcripts, we are able to construct a gene regulatory network (GRN) linking 47 genes with 5 defining hubs which identify protein-modifying enzymes, receptors, ligands, RNA-binding proteins, transcription factors and co-factors.

To address and validate our vegetally localized transcript set for its inclusion of germ plasm mRNAs, we performed a comparison of our data with that of Butler et al. (2017), who recently identified enriched transcripts in zygotically active PGCs of *Xenopus laevis*. Over half of the 198 vegetally enriched transcripts were zygotically expressed in the PGC sample. Further similarity of both maternal and zygotic germ plasm containing samples was highlighted during GRN analysis. Each sample set generated GRNs containing the transcription factors *sox7* and *e2f1*. Taken together, these data guided expression and functional studies for the investigation of germ plasm enriched RNAs in both sample types.

Outside of computationally generated data, my dissertation work also provides expression and functional data for transcripts identified in maternal and/or zygotic germ plasm containing samples. Transcripts were chosen for Gain-of Function and Loss-of Function experiments by either anticipated biological function in PGC development or
presence in our GRN. Whole mount in situ hybridization analysis confirmed 17 and 11 RNAs were localized to the vegetal pole and the zygotic PGC, respectively. Functional studies provided evidence that efnb1, sox7, e2f1, wwtr1, otx1, parn, p300, and rras2 are all involved in promoting the normal development of PGCs. Additional studies indicate novel functions for efnb1 and sox7 in PGCs, where ephrinB1 (efnb1) plays a role in migration and sox7 activity must be tightly regulated to maintain PGCs.

Here, I present the initial and required steps for identifying germ plasm components involved in Xenopus laevis germline development. However, the full utility of my RNA-seq data has yet to be realized. The proposed pathways operating in both the vegetal pole and zygotic PGC samples highlight where future investigations will be most fruitful.
ACKNOWLEDGEMENTS

Dr. Tristan Agüero was pivotal in my development as a scientist. Without his endless patience and guidance, I would not have attained the skills that I have today, in addition to my scientific approach and critical thinking ability. This individual, single handedly taught me what it is to be a good scientist by way of teaching me how to critically analyze my own work. Additionally, he helped me to understand the importance for finding the best controls for every experiment so that the clearest information can be attained.

Both Drs. Amanda Butler and Karen Newman provided me with the knowledge to rely on and believe in myself as a scientist. Dr. Butler was essential in my development as a writer. With her guidance, my writing ability has evolved into something I am truly proud of. Dr. Newman provided me support in idea development and expertise in experimental design and development.

Credit must be given to Dr. Mary Lou King, who taught me self-determination and self-reliance in addition to the value of independent study. As a mentor, she really taught me that to do great science, one must strive for excellence within the scientific community. A scientific career must not be looked at as a job, but as a duty to better the scientific community by giving the maximal effort within one’s own capacity. Additionally, Dr. Mary Lou King took a chance on me as a graduate student with the knowledge that I
would attempt to become a mother during my graduate career, which I will indefinitely appreciate.

The scientific minds of my committee worked continuously throughout the years to define my studies, ultimately molding me into the scientist I am today. Thank you Drs. Carlos Moraes, Athula Wikramanayake, and David Robbins.

As a colleague and a friend, Natasha Ward of Dr. Thomas Malek’s lab provided immeasurable support and advice during times when I lacked clarity in the lab.

The search committee of the Program in Biomedical Sciences at the Miller School of Medicine, University of Miami took a chance on me. They provided me this career opportunity so that I may be competitive for academic positions and afforded future opportunities that would otherwise be unattainable.

Thank you to my mother, Kimm Gant, who always put me first during the hardest times of our lives. Everything that I am today is because of the moral values which she has instilled into me. If I could ever be as lucky to be even close to the caliber of mother she was to me, my children will truly be blessed.

My husband, Garrett Owens, and two daughters, Madison Raylin and Lilly Ann Owens, have provided me with unsurmountable support during my graduate career. Madison and Lilly both served as my cheerleaders during my hardest days during my graduate studies,
whereas my husband truly made it possible for me to do it all: have a family and pursue a graduate level degree to hopefully provide us a better life in the future.

Additionally, I would like to thank all of the external support that I received during my graduate research: Dr. Mike Gilchrist through the Bioinformatics Workshop (Cold Spring Harbor, NY) gave valuable advice for the initial planning of my RNA-seq work; Dr. Jing Yang was kind to provide X. tropicalis clones (hook2, tob2 and sox7); Dr. Mike Klymkowsky gave us the dominant negative construct for the sox7 experiments; Dr. Lingyu Wang provided excellent technical support for our RNA-seq analysis; Dr. Sawsan Khuri was essential in guiding some of our gene regulatory network analysis; and Dr. Radia Forteza provided me technical support whenever common equipment issues came up.
DEDICATIONS

This dissertation is dedicated to my daughters, Madison and Lilly Owens, in addition to my father, Jackie Lee Gant, for being my cheerleaders when times were tougher than I could bear on my own. My daughters, both with and without words, have always inspired me to push through any obstacle that was in my way during my thesis work. My father believed in me more than I did myself. He also taught me the invaluable lesson that sometimes you can be more productive if you just take the time to relax. These individuals could always give me the extra push I needed with just a simple hug or smile.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF PUBLICATIONS AND COAUTHOR CONTRIBUTIONS</td>
<td>xiii</td>
</tr>
<tr>
<td>CHAPTER 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Specification of the germline</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Germplasm formation in <em>Xenopus laevis</em> during oogenesis</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Germline development during Xenopus embryogenesis</td>
<td>7</td>
</tr>
<tr>
<td>1.4 Germ plasm composition</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 2: High-throughput analysis identifies maternal germline RNAs</td>
<td>24</td>
</tr>
<tr>
<td>2.1 RNA-seq analysis of vegetal and animal poles</td>
<td>27</td>
</tr>
<tr>
<td>2.2 Biological process and network analysis of vegetally enriched transcripts</td>
<td>29</td>
</tr>
<tr>
<td>2.3 miRNA target RNAs are localized at the vegetal pole</td>
<td>33</td>
</tr>
<tr>
<td>2.4 Expression of vegetally localized RNAs during development</td>
<td>34</td>
</tr>
<tr>
<td>2.5 Novel mRNAs enriched at the animal pole</td>
<td>39</td>
</tr>
<tr>
<td>CHAPTER 3: Maternal germline RNAs are crucial for primordial germ cell survival and proper migration</td>
<td>43</td>
</tr>
<tr>
<td>3.1 Overexpression of <em>e2f1, otx1, parn, rras2</em> and <em>wwrt1</em> significantly reduces PGC number</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Embryos depleted of <em>otx1</em> and <em>wwr1</em> show a significant increase in PGC number</td>
<td>46</td>
</tr>
<tr>
<td>3.3 p300 is required for normal PGC development</td>
<td>48</td>
</tr>
<tr>
<td>3.4 <em>efnb1</em> plays an essential role in PGC specification and migration</td>
<td>49</td>
</tr>
</tbody>
</table>
CHAPTER 4: Normal PGC development requires strict regulation of *sox7*  
4.1 Maternal, vegetally-enriched transcripts are re-expressed in PGCs  
4.2 Expression of PGC enriched mRNAs during development  
4.3 Over expression and knockdown of *sox7* in single-cell embryos reduces PGC number  
4.4 PGC-directed *sox7* knockdown or over-expression increases PGC number in early development  
4.5 WISH analysis reveals MO-mediated knockout of *sox7* causes no detected effect on native mRNA expression for select germ plasm enriched transcripts  

Chapter 5: Discussion and Significance  
5.1 RNA-seq analysis reveal novel gene regulatory networks (GRNs)  
5.2 MicroRNAs are equally distributed in vegetal and animal pole samples  
5.3 Efnb1 is required for proper PGC migration and formation  
5.4 Sox7 is required for PGC development  
5.5 The vegetal pole is a signaling center  

Significance  

Chapter 6: Materials and Methods  
6.1 Isolation of animal and vegetal pole samples  
6.2 RNA isolation  
6.3 RNA preparation for Illumina sequencing  
6.4 Data processing and quantification  
6.5 Small RNA analysis and novel non-coding RNA identification  
6.6 Gene name identification and GO analysis
6.7 Whole-mount in situ hybridization (WISH) 97
6.8 PGC-targeted injections for sox7 manipulation 97
6.9 Gain-of-function analysis of vegetally enriched genes of interest 98
6.10 Loss-of-Function analysis of novel germline RNAs 98
References 102
List of Figures

Fig. 1. Schematic of germline formation in *Xenopus laevis.* 12

Fig. 2.1. RNA-seq analysis of vegetal versus animal pole transcripts in stage VI *X. laevis* oocytes. 28

Fig. 2.2. Biological process and network analysis of vegetally enriched transcripts. 32

Fig. 2.3. WISH of vegetal pole transcripts. 36

Figure 2.4. WISH of select vegetal and animal pole transcripts. 37

Fig. 3.1. Overexpression of five out of six selected vegetally enriched mRNAs reduces PGC number. 46

Figure 3.2. Morpholinos (MOs) inhibit expression of their target RNAs 47

Fig. 3.3. MO-mediated knockdown of a subset of vegetally enriched mRNAs increases PGC number. 47

Fig. 3.4. Inhibition of p300 reduces PGC number. 49

Fig. 3.5. *efnb1* is required for normal PGC number and migration. 51

Figure 4.1. Analysis of PGC enriched transcripts. 59

Figure 4.2. PGC-directed Injections. 59

Figure 4.3. WISH analysis of PGC transcripts. 62

Figure 4.4. Morpholino-mediated *sox7* inhibition reduces PGC number in tailbud embryos. 64

Fig. 4.5. Altering *sox7* expression reduces PGC number. 64

Figure 4.6. PGC number increased after PGC-directed *sox7* knockdown or over-expression in early development. 67
Figure 8. Model for EphrinB1/Grip2 scaffold promoting signaling at vegetal pole.
**List of Tables**

Table 1. Summary of Known vegetally localized RNAs found prior to the use of microarray.

Table 2.1. Top 198 vegetally enriched transcripts identified by RNA-seq.

http://www.biologists.com/DEV_Movies/DEV139220/TableS1.xlsx

Table 2.2 All annotated transcripts identified using v6.0 and v7.1 *Xenopus laevis* scaffold sets.

http://www.biologists.com/DEV_Movies/DEV139220/TableS2.xlsx

Table 2.3 All un-annotated transcripts identified using v6.0 and v7.1 *Xenopus laevis* scaffold sets.

http://www.biologists.com/DEV_Movies/DEV139220/TableS3.xlsx

Table 2.4. Gene function of the 40 most highly enriched mRNAs identified at the oocyte vegetal pole.

Table 2.5. GO processes of vegetally enriched transcripts.

Table 2.6. miRNAs and their vegetally localized mRNA targets.

Table 2.7. Gene function of the 15 most highly enriched mRNAs identified at the oocyte animal pole.

Table 4. Maternal versus zygotic PGC-enriched transcripts.
List of Publications and Coauthor Contributions

CHAPTER 2

Dawn Owens performed majority of the experiments, data analysis, figure preparation, planned experiments, conducted whole mount in situ hybridization analysis, in addition to editing, organizing and writing this chapter/manuscript submission.

Amanda Butler performed whole mount in situ hybridization analysis (sox7-gastrula), and helped edit this chapter and write manuscript submission.

Tristan H. Agüero planned OE experiments, figure preparation (2.3), and data analysis.

Karen M. Newman performed sample collection, embryo acquisition, data preparation, and analysis.

Darek V. Booven was responsible for a majority of the Bioinformatics and analysis of RNA-seq raw data.

Mary Lou King conceived and planned experiments, assisted in editing this chapter and writing the manuscript submission.

The majority of this chapter has been published in Development

(http://dev.biologists.org/content/144/2/292)
Dawn Owens performed majority of the experiments, functional analysis (p300, wwt1 and otx1), data analysis, figure preparation, planned experiments, conducted whole mount *in situ* hybridization analysis, in addition to editing, organizing and writing this chapter/ manuscript submission.

Amanda Butler planned experiments, performed functional experiments (efnb1, wwt1, otx1), conducted whole mount *in situ* hybridization analysis, data analysis, and helped edit this chapter and write manuscript submission.

Tristan H. Agüero planned experiments, performed Functional experiments (efnb1, wwt1, otx1, rras2, spire1, e2f1, parn), conducted whole mount *in situ* hybridization analysis, figure preparation, and data analysis.

Karen M. Newman helped with embryo acquisition and experiments (Fig. 3.4).

Mary Lou King conceived and planned experiments, assisted in editing this chapter and writing the manuscript submission.

The majority of this chapter has been published in *Development* (http://dev.biologists.org/content/144/2/292).
CHAPTER 4

Dawn Owens performed functional experiments (sox7-MO), data analysis, figure preparation, planned experiments, majority of whole mount in situ hybridization analysis, in addition to editing, organizing and writing this chapter/manuscript submission.

Amanda Butler was responsible for a majority of the Bioinformatics and GeneGo analysis (Fig.4.1), planned experiments, performed functional experiments, conducted whole mount in situ hybridization analysis, data analysis, and helped edit this chapter and write manuscript submission.

Lingyu Wang analyzed data and provided excellent technical support for RNA-seq analysis.

Karen M. Newman helped with embryo acquisition.

Mary Lou King conceived and planned experiments, assisted in editing this chapter and writing the manuscript submission.

The majority of this chapter has been submitted to Development and is currently in revision (http://dev.biologists.org/content/144/2/292; Butler et al., in Revision, 2017)
Chapter 1

INTRODUCTION

1.1 Specification of the germline

The germline is responsible for maintaining and passing on the genetic information of sexually reproducing organisms to future generations. Located in the mature gonad, germ cells (gametes) possess half of an organism’s chromosomal number (termed as haploid) and combine in a species-specific manner with the haploid cell of the opposite sex during sexual reproduction. Primordial germ cells (PGCs) are the exclusive precursors of germ cells. They arise early in development, and eventually migrate within the dorsal mesentery and reach the presumptive gonad. A prominent characteristic of PGCs is their totipotent potential. What gives PGCs the ability to preserve totipotent potential and provide subsequent generations with genetic information is far from being completely understood.

Specification of the germline has been reported to occur by two separate methods: induction or inheritance of germ plasm. “Keimplasma” or “germ plasm” was a term first coined by the German evolutionary biologist, August Weismann in 1892 and referred to the germ plasm as being the heritable genes of an organism. To date, germ plasm describes a cytoplasmic substance containing electron dense, germinal granule inclusions consisting of proteins and RNAs essential and specific for germline specification (Heasman et al., 1984a). The induction (or epigenesis) model for germline specification requires paracrine, inductive signaling from neighboring tissue to establish a germ cell
fate (Extavour and Akam, 2003). The first investigator to suggest germline specification by way of induction was Kotani, who replaced embryonic mesoderm cells of the presumptive lateral plate from early gastrula stage *Triturus pyrrhogaster* with donor presumptive ectoderm cells (1957). Experimental host embryos formed PGCs, showing that PGCs may have a somatic origin arising from ectodermal tissue. Support of this conclusion was provided by Kocher-Becker and Tiedemann who induced PGCs, including mesodermal and endodermal structures, when they subjected Triturus early gastrula ectoderm to an unknown “vegetalizing factor” from chick embryo extracts (1971). Further experimentation demonstrated that the ventral endoderm of the early embryo gave inductive signals to form a germline, in addition to mesodermal tissues (Nieuwkoop, 1947; Boterenbrood and Nieuwkoop, 1973).

Induction is postulated to be the ancestral form of germline specification and occurs in both human and mouse (Extavour and Akam, 2003). In mouse, extensive study of germline specification has revealed that bone morphogenic (BMP) factors 2, 4 and 8b of the TGFβ superfamily are required for germ cell induction (Lawson et al., 1999; Ying et al., 2000; Ying and Zhao, 2001). Once induced, PGCs form germinal granules around E7.5 that persist until gametogenesis occurs (Chuma et al., 2009; Spiegelman and Bennett, 1973). Thus, PGCs are specified from a group of competent cells, all capable of responding to inductive signals, but only cells receiving appropriate signaling will become PGCs (Agüero et al., 2017; Extavour and Akam, 2003). Therefore, in inductive specification, more plasticity exists for formation of the germline population, although specific cell-cell signaling is required.
The second method of germline specification, inheritance (or preformation), is dependent on maternal determinants localized within the oocyte, not inductive signaling from surrounding tissue. This model of specification has evolved independently in many lineages of vertebrates and is found in diverse organisms including: C. elegans; D. melanogaster; D. rerio; and G. domesticus (Johnson et al., 2003). Anuran studies provided the initial evidence for the requirement of inherited maternal determinants for germline formation in vertebrates (Bounoure, 1939). *Xenopus laevis* (African Clawed Frog) also utilizes the inheritance method for germline specification. *X. laevis* is prized as a model organism for germline investigations for multiple reasons: 1) Fertilization and embryonic development occur externally for relatively easy observations in development; 2) *X. laevis* 1-cell eggs/embryos are large in size (~1.2mm), making manual manipulation possible; 3) Germline specific reporter assays have been developed for *in vivo* study of the PGC population; 4) Biochemical amounts of germ plasm can be extracted for proteomic and nucleic acid studies; 5) Gene manipulation by knockdown and overexpression can be selectively administered to the whole embryo and germline fate-mapped blastomeres; 6) *Xenopus laevis* possesses a genetic advantage when compared to some organisms due to the many genes shared with humans. Many more reasons to utilize *X. laevis* over other model organisms exist, but these are the most prominent for germline investigations.

### 1.2 Germplasm formation in *Xenopus laevis* during oogenesis

In *X. laevis*, germ plasm begins to accumulate at the very beginning of oogenesis, and can be found in close association with a visible aggregation of mitochondria adjacent to the
nucleus towards the future vegetal pole of the stage I oocyte. This aggregation of mitochondria forms in a cellular body termed the Balbiani body (Bb) or mitochondrial cloud (MC) in both zebrafish and Xenopus. The MC is a non-membrane bound structure that is held together by a fibrillar filament material composed of Xvelo, an amyloid-like matrix forming protein (Boke et al., 2016). Actin, intermediate filaments and microtubules, three of the major cytoskeletal proteins, also contribute to the structural integrity of the MC (Gard et al., 1997). The MC also contains golgi, endoplasmic reticulum, lipids, and electron dense germinal granules, which are characteristic of germ plasm. The most prominent function of the MC is to retain germ plasm for germline specification in the future embryo as well as stockpile the maternal supply of mitochondria.

Germ plasm formation is a highly involved process. The earliest stages of germ plasm formation are visualized within the nuclei of oogonia as nuage material forms. Based on careful ultrastructural analysis, nuage is thought to give rise to germinal granules (Reviewed in Agüero et al., 2017; al-Mukhtar and Web, 1971). By stage I, the nuage has moved outside of the nucleus via nuclear pores and into the MC, forming germinal granules there containing germline specific RNAs and proteins (Agüero et al., 2017). Hermes protein has been found within both germinal granules and the surrounding germ plasm matrix, whereas its mRNA is in the matrix only (Song et al., 2007; Agüero et al., 2016). To date, identified RNAs of germinal granules only include nanos, whereas deadsouth, and xpat RNAs are in close apposition to the granules. Mitochondrial RNA and xdazl are found in the germ plasm matrix (Chang et al., 2004; Kloc et al., 2002). The functional significance of the particular distribution of components in the germ plasm
remains unclear. The germinal granule containing portion of the MC is now termed as germ plasm and localizes to the future vegetal pole of the MC (Kloc and Etkin, 1995). Orientation of the MC is important for establishment of the animal/vegetal (A/V) axis in the oocyte. During stage II-III, germ plasm and associated RNAs, now migrate towards and spread over the vegetal cortex via directional expansion of the MC (Wilk et al., 2005).

Establishment of the A/V axis in the stage I oocyte by the MC orientation initiates necessary RNA localization patterning required for the axis (dorsal/ventral, left/right), primary germ layer organization, and germline specification of the future embryo (King, 2014). RNAs found within germ plasm that are required for germline specification have been recorded to utilize 1 of 3 localization pathways: early (METRO), late (Vg1) or both. Originally, it was thought that early pathway RNAs were related to germline determination because of their prevalent association with germ plasm, whereas late RNAs were speculated to be involved in somatic determination (Reviewed in Houston, 2013). Counter to this previous hypothesis, multiple examples of both somatic and/or germline RNAs have now been recorded to use each pathway (Owens et al., 2017; Cuykendall and Houston, 2010). The early and late RNA localization pathways were first identified by Forristall et al. (1995) with their studies on xcat2 (nanos) a germline specific RNA and Vgl (somatic determinant) in X. laevis. These findings were later confirmed and extended by Kloc and Etkin (1995) using the same as well as different RNAs, Xlsirts (germline) and wnt11 (somatic determinant) (See Table 1). The RNAs wnt11 and xcat2 proved to be early localized transcripts that use part of the MC called the message transport organizer (METRO). This region facilitates the translocation of germ plasm and
associated RNAs towards the vegetal pole. The METRO localized transcripts move to the most vegetal position within the MC of a stage I oocyte (Kloc and Etkin, 1995). During oogenesis stages I-III, the MC and vegetally positioned METRO move from close association with the nucleus to become detached and ultimately anchored to the vegetal cortex. At this juncture, germinal granules become anchored by an actin and intermediate filament system (Kloc et al., 2007). A textbook example of early pathway localization can be visualized within the stage I oocyte when whole mount in situ hybridization (WISH) analysis is performed using Xpat, an RNA required for proper germ plasm formation (Hudson and Woodland, 1998; Machado et al., 2005). Outside of the MC, the oocyte also accumulates large quantities of maternally derived proteins (including yolk) and RNAs for use during the initial cleavage stages of embryogenesis, prior to the maternal to zygotic transition.

*Vg1* was the first identified RNA to use an alternative, later acting pathway of RNA localization in the Xenopus oocyte (Weeks and Melton, 1987). *Vg1* was identified by Rabegliati et al. (1985) and codes for a TGF-beta growth factor (Reviewed in Heasman, 2006; See Table 1). RNAs using the late pathway are equally distributed throughout the cytoplasm of the stage I and II oocyte. Other RNAs that have been shown to use the late pathway are *vegt* (meso-endoderm determinant) and *dnd1* (a germline determinant) (Lustig et al., 1996; Stennard et al., 1996; Zhang and King, 1996; Horvay et al., 2006). At stage III, late pathway RNAs begin to move towards the vegetal cortex via the activity of microtubules and dynein/kinesin motors (Kloc et al., 2007; Gagnon et al., 2013). Late pathway RNAs are completely localized to the vegetal cortex by stage IV, and are found
to show a more widespread vegetal distribution in the mature, stage VI oocyte when compared to early pathway RNAs.

Finally, some RNAs utilize both early and late localization pathways for vegetal cortex enrichment (King et al., 1999; Chan et al., 1999). The first RNA characterized to utilize both pathways was $fatvg(\text{plin}2)$, an RNA with functionality in both soma and germline lineages (Chan et al., 1999, 2001, 2007). $\text{hermes}(rbpms2)$ RNA also uses both RNA localization pathways and codes for an RNA binding protein. This gene has a role in myocardial differentiation and initial blastomere cleavages in the early Xenopus embryo, although the later finding has yet to be replicated by other groups (Zearfoss et al., 2004; Gerber et al., 2002; See Table 1). Once at the vegetal cortex of the stage IV oocyte, somatic and germline localized RNAs maintain this position until fertilization.

It is important to note that RNA localization also occurs at the animal pole. Transcripts localized at the animal pole are not as abundant or as diverse as are those localized to the vegetal pole (Cuykendall and Houston, 2010; Owens et al., 2017; Claußen et al., 2015; De Domenico et al., 2015). Localization mechanisms supporting animal pole enrichment of transcripts are not well understood. Identified animal pole transcripts tend to be inherited by blastomeres fated to be neural ectoderm in the developing embryo (Grant et al., 2014; Owens et al., 2017).

### 1.3 Germline development during Xenopus embryogenesis

At the start of embryogenesis, germ plasm remains in subcortical yolk-free ‘isles’ at the vegetal pole (Savage and Danilchik, 1993). Initial cleavages of the early embryo cause partitioning of the germ plasm and vegetally localized somatic determinants into the
resulting, vegetal most blastomeres. Because embryos contain a pre-determined, maternally derived amount of germ plasm, only presumptive PGCs (pPGCs) with sufficient germ plasm, become PGCs. Neighboring somatic cells and pPGCs alike divide synchronously 11-12 times (embryonic stage 8), at which time the mid-blastula transition (MBT) occurs, and further divisions become asynchronous (Newport and Kirschner, 1982a). Daughter cells of pPGCs continue to asymmetrically inherit germ plasm until gastrulation (embryonic stage 10), when germ plasm moves to a perinuclear location and segregation of the germline occurs. At this stage of germline development, both germes and deadsouth (ddx25) RNAs function in the perinuclear orientation of germ plasm, and are required for proper germline formation (Yamaguchi et al., 2013a; Yamaguchi et al., 2013b). The pPGCs are now called PGCs and germ plasm is divided equally in each following cell division, increasing the number of PGCs.

Counter intuitive to what might be expected, Whitington and Dixon (1975) suggest PGCs only divide symmetrically ~3 times prior to migrating towards the genital ridge (embryonic stage 40). The seemingly slower cell cycle and mitotic rate of the PGC population has been suspected to be necessary for delaying their transcriptional activation for proper germline formation to occur (reviewed by Houston and King, 2000). During MBT, zygotic transcription begins for many RNAs in the somatic cells, but the maternal to zygotic transition, which begins at fertilization and ends with full activation of the zygotic genome, occurs later in PGC development (Newport and Kirschner, 1982b; Tadros and Lipshitz, 2009; Langley et al., 2014). At the MBT, PGCs do not initiate zygotic transcription and quiescence is maintained until sometime between mid-gastrula and early neurula (Venkaterama et al., 2010; Lai and King, 2013). The specific factor(s)
which initiate the PGC zygotic program are unknown. Tight regulation of the cell cycle and translational repression of somatic determinants in the PGCs are required for maintaining the germline to prevent somatic fates.

What maintains the tight regulation of many cellular activities of PGCs has been investigated in diverse animal systems. PGC specification and development have been well characterized in both *Drosophila* and *C. elegans*. Using genetic approaches, many genes have been identified that are involved in these processes (Reviewed in Nikolic et al., 2016). Recently, we have learned that translational regulation in Xenopus PGCs may be dictated to a significant degree by the genes Pumilio, Nanos, Xdazl, and Dnd (Reviewed by Agüero, 2017). Nanos translation is inhibited in the oocyte by sequestering and protecting its message within germinal granules. In addition, there is a secondary structure in the *nanos* ORF that prevents initiation events (Lai et al., 2011). After fertilization, Dnd1, an RNA binding protein, likely relieves translational repression of *nanos* RNA at the structural level (Agüero et al., 2017). Nanos functions as a translational repressor when partnered with Pumilio, an RNA-binding protein. The Nanos/Pumilio partnership is broadly conserved in species as diverse as worm and human (Kraemer et al., 1999; Jaruzelska et al., 2003). The Nanos-Pumilio complex aids in germline conservation during early development by translationally repressing targeted mRNAs (like *vegT*) that possess the Pumilio recognition sequence UGUA(N)UA. Embryos deficient in Nanos activity are infertile as PGCs inappropriately express somatic genes and become apoptotic (Lai et al., 2012). The disruption of the Nanos/Pumilio repression complex may allow VegT activity which regulates the expression of many somatic genes (Lai et al., 2012).
MicroRNA mediated translational regulation provides additional complexity to the way in which germline and soma develop separately. In the soma, it has been reported that both germline RNAs *deadsouth* and *dnd1*, when fused to a fluorescent reporter, are destroyed via miRNA mediated degradation (Kataoka et al., 2006; Koebernick et al., 2010). Degradation of germline specific RNAs in the soma must occur for differentiation of somatic tissues in the embryo. Alternatively, in the germline, fluorescent reporter studies show that both proteins Dazl and Dnd can provide protection from miRNA mediated degradation of certain RNAs. Specifically, Dnd protein has been found to inhibit miR-18 mediated degradation of RNAs *dnd, xpat, xdal, nanos*, and *deadsouth* by way of Dnd/ErlB1 binding the 3’UTR region of those RNAs (Koebernick et al., 2010). Another form of possible miRNA protection in Xenopus may be provided by Xdazl. In Zebrafish, Takada and colleagues demonstrated that Dazl provided protective properties for mRNAs against miR-430 degradation (Xenopus homolog is miR-427) within germplasm (2009), and protection of germline RNAs is required in early development.

Translational repression is required for maintaining proper quiescence in the PGC to prevent adoption of a somatic fate (Venkaterama et al., 2010), however, translational activation is also necessary for PGC survival. Xdazl and Dnd1 dictate when the activation of essential germline genes occur, as in Dnd1 activation of *nanos*. The enrichment of Xdazl in the germline is observed until stage ~52, at which time the tadpole PGCs have completed their migration into the presumptive gonads; Xdazl is known to translationally activate mRNAs by way of partnership with poly(A) binding protein-1 (PABP1) and embryonic PABP (ePABP) (Collier et al., 2005). A true “in vivo” germline target mRNA for a Xdazl/PABP complex has yet to be identified and requires
further investigative studies. Notably, Xdazl and Dnd1 have been linked to migration activity of PGCs, a critical process for the proper formation of functional gametes (Horvay et al., 2006; Houston and King, 2000). Studies in which both Xdazl and Dnd1 in Xenopus have been depleted revealed that this depletion causes a failure in migratory activity of PGCs, and results in the loss of PGCs (Houston and King, 2000). These results indicate that RNA-binding proteins mediate vital translationexpression of other migratory factors in PGCs.

Throughout early development of the Xenopus embryo, PGCs go through controlled periods of active and passive migration prior to arriving at the location of the future gonad. Vegetal localization of germ plasm in the oocyte is causative of pPGC formation within vegetal blastomeres of the late blastula. During gastrulation, pPGCs are passively moved with surrounding endoderm cells towards the endodermal core of the embryo. After germ plasm migration to a perinuclear position and the separation of the germ cell lineage during gastrulation, PGCs generally remain within the endoderm until the early tailbud stage (22-24) (Whittington and Dixon, 1975; Houston and King, 2000). Active migration begins after stage 24 with the cells first moving laterally (~33/34), and later becoming dorsally localized within the endoderm (stage 40) (Kamimura et al., 1976; Nishiumi et al., 2005; Whittington and Dixon, 1975). Finally, PGCs migrate out of the endoderm (stage 41) towards the presumptive gonads, along the dorsal mesentery and complete migration by stage 52 (King, 2014). Once in the somatic gonadal environment, PGCs undergo proliferation and differentiation to form functional germ cells (Wylie and Heasman, 1976; Braat et al., 1999). (see Figure 1; Reviewed in Wylie et al., 1985).
Figure 1. Schematic of germline formation in *Xenopus laevis*. Stage I oocyte: germ plasm (green) assembles in MC (red) in close association with the germinal vesicle (GV). Stage II/III oocyte: MC fragments and moves toward the vegetal cortex. Stage VI: germ plasm within the vegetal cortical area. Eight-cell embryo: germ plasm is inherited by vegetal blastomeres shown from the lateral and vegetal pole perspective. Blastula: germ plasm lies near the plasma membrane of four to six cells, the pPGCs. Gastrulation: germ plasm translocates by a microtubule-based mechanism to a perinuclear position. The germline (PGCs) is now segregated from endoderm lineage (yellow, endoderm; red, mesoderm; blue, ectoderm). Tail bud stages 24–34: PGCs begin migration steps clustering, dispersing laterally, directionally migrating dorsally, and, at tadpole 41, reaggregating at the dorsal tip of the endoderm (adapted from Figure 1e–h’ in Terayama K, Kataoka K, Morichika K, Orii H, Watanabe K, Mochii M. Developmental regulation of locomotive activity in Xenopus primordial germ cells. Dev Growth Differ 2013;55(2):217–228.) Tadpole: PGCs migrate along the dorsal mesentery to reach the presumptive gonads.

Proper migration patterns of PGCs are fundamentally important for survival of the germline in the developing embryo. Species including ascidians, worms, flies, zebrafish, frog, mouse, and human all specify PGCs outside of the gonad primordium in early development, making PGC migration pivotal for the later development of functional gametes (Houston and King, 2000; Raz, 2003; Nakamura and Seydoux, 2008; Shirae-Karabayashi et al., 2011; Reviewed in Nikolic et al., 2016). Successful migration of PGCs to the somatic gonadal environment results in necessary signaling, which is not well understood, for initiating controlled proliferative activity of PGCs via meiosis and differentiation within the gonad to form mature germ cells (Reviewed in Nicolik et al., 2016). To date, multiple germ plasm components (xdazl, dnd, germes, grip2) in addition to 3 possible signaling mechanisms (Kif13b/PIP3, CXCR4/SDF1, and Notch/Delta2) conducive to proper PGC migration have been found in Xenopus (Tarbashevich et al., 2011; Tarbashevich et al., 2007; Houston and King, 2000; Morichika et al., 2010; Takeuchi et al., 2010). Tarbashevich and colleagues have found that correct PGC migration requires the scaffolding protein GRIP2 in addition to signaling between Kif13b and PIP3 (2007 and 2011). Mechanistically, the Kif13b and PIP3 interaction was shown to cause polarization of PIP3 signaling for directional migration of PGCs. Notch/Delta signaling provides another mechanism for PGC migration. PGCs have been found to highly express Delta2, whereas Notch1 and Delta1 are expressed equally in both the germline and soma populations of the tailbud embryo (Morichika et al., 2010). The third mechanism utilizing CXCR4 and SDF1 to provide directional cues for proper PGC migration has been found in chicken, mouse, zebrafish, and Xenopus (Knaut et al., 2003; Mizoguchi et al., 2008). In all these model systems, PGCs express the G-coupled-
chemokine receptor CXCR4 with its ligand SDF1 being found in the dorsal mesentery, providing a seemingly conserved chemotaxis response by PGCs to migrate to the dorsal mesentery. All previously mentioned mechanisms and genes have been demonstrated to have an important role in PGC migration, but the entire gene regulatory network that could be responsible for proper germline formation in the embryo is far from complete.

Several studies have investigated how the migration patterns of PGCs are regulated during their journey from the endodermal core to the dorsal mesentery (Nishiumi et al., 2005; Terayama et al., 2013; Dzementsei et al., 2013). Recently, Terayama et al. (2013) reported 4 distinct phases of PGC migration: clustering; dispersing; directional migration; and reaggregation. Similar to other motile cells, PGCs exhibit cellular blebbing controlled in part by F-actin, Myosin II and RhoA/Rho-protein kinase signaling (Terayama et al., 2013). PGC migration is an involved process, requiring temporally appropriate signaling and PGC morphology for future development of the germline. When PGCs improperly migrate and fail to leave the endoderm, apoptosis occurs (Lai et al., 2012; Ikenishi et al., 2007; Köprunner et al., 2001). Incorrect expression of known genes important for PGC migration, like grip2, results in abnormal PGC migration and, most commonly, PGC numbers decrease due to apoptosis (Tarbashevich et al., 2007; Kirilenko et al., 2008). Tarbashevich et al. have shown that GRIP2 morphants experience PGC migration patterns that were outside of the normal distribution boundary between somite 5 and 11 observed at tailbud stage 32 (2007). PGCs outside the normal boundaries for migration may be more likely to fail reaching the dorsal mesentery, remaining in the endoderm, and leading to their demise via apoptosis (Lai et al., 2012; Ikenishi et al., 2007; Köprunner et al., 2001). These results are indicative of improper signaling to the
PGC by the surrounding somatic endoderm, causing the PGCs to either initiate apoptotic events or possibly adopt a somatic cell fate. Unfortunately, this study did not track the mislocalized PGCs after migration into the future gonad, leaving important questions about their fate in the endoderm.

Upon arrival in the presumptive gonads, proliferative events now occur in the PGC population, forming functional germ cells (Reviewed in Seydoux and Braun, 2006). At sexual maturity (taking about 8 months), spermatogenesis or oogenesis cycles are fully functional in the germ cell population. Within the female *X. laevis*, a single cycle begins with the replenishment of the oogonia population resulting from mitotic germline stem cell (GSC) divisions. Resulting daughter cells create one renewed GSC and one oogonia for forming a mature gamete. The oogonia become secondary oogonia, and enter meiosis that results in the appropriate haploid number of alleles (Rasar et al., 2006). Finally, the oogonia become arrested in the cell cycle at G2 prophase, accumulating maternally derived components until a mature, stage VI oocyte develops (Song et al., 2007; Kloc et al, 2004; Reviewed in Seydoux and Braun, 2006). Once again, formation of the maternally derived, all-inclusive germline determinant, germ plasm, is complete and ready for use inside of the stage VI oocyte to maintain the heritable genetic information and germline integrity of the future embryo.

**1.4 Germ plasm composition**

The complete list of RNAs and mechanisms controlling PGC migration, proliferation, and specification in the early embryo still requires extensive investigation. To fully understand germline differentiation, investigation of the source for PGC specification,
germ plasm, is essential. Although it has been found that germ plasm is sufficient for germline specification (Tada et al., 2012), the complete molecular make-up of germ plasm has yet to be determined. Prior to high-throughput analytical tools, investigators had attempted to elucidate the components of germ plasm by available molecular means, including cDNA cloning and microarray analysis.

Unfortunately, using mammalian model systems to study germ plasm is a daunting undertaking because acquiring amounts sufficient for biochemical or gene analysis is difficult. Therefore, scientists have turned to other organisms such as \textit{C. elegans}, \textit{Drosophila}, zebrafish and Xenopus for investigating germ plasm, where it is possible to isolate sufficient amounts for molecular analysis. Initial studies that identified germ plasm-specific components were primarily focused on identifying localized maternal RNAs, and proteins to a lesser extent, required for specifying developmental patterns in the embryo. Investigations using \textit{Drosophila} have identified localized determinants required for anterior/posterior and dorsal/ventral patterning in addition to pole cell specification (St Johnston and Nusslein-Volhard, 1992; Micklem, 1995; St Johnston, 1995; Grunert and St Johnston, 1996; Williamson and Lehmann, 1996). Pole cells are the PGCs of \textit{Drosophila}, which are also determined autonomously. The pole cell-specific determinants identified in \textit{Drosophila} genetically included many molecular components of germ plasm, which have been now cloned (St Johnston and Nusslein-Volhard, 1992; Williamson and Lehmann, 1996). One of the few pole-cell specific determinants, Oskar, appears to be specific to Dipteran insects, having no equivalent in Xenopus germ plasm (Machado et al., 2005). Interestingly, Komiya et al. (1994) utilized the \textit{Drosophila} cDNA clone of Vasa, a DEAD-box helicase required for PGC specification and migration.
(Shimaoka et al., 2017), to identify the first marker for PGCs in Xenopus, XVLG1 (Xenopus vasa-like gene). Specific expression of XVLG1 was localized to the germ cells of embryos at stage 46 and after (Komiya et al., 1994).

Other groups utilized cDNA cloning techniques to identify and investigate localized messages within the *Xenopus laevis* oocyte/egg. As previously stated, germ plasm is known to localize to the oocyte/egg vegetal pole region. Therefore, to identify germ plasm components, the vegetal pole was compared to the animal pole of the oocyte to determine which RNAs are enriched at the vegetal pole, and hence, may be a germ plasm component. By making a cDNA library from animal and vegetal stage VI oocyte tips, Rebagniati et al. discovered the first vegetally localized RNA, *vg1* (1985), that was later identified as a member of the TGF-super family (Weeks and Melton, 1987). Many vegetally localized components have been identified similarly in eggs/oocytes and have been found to be part of the germ plasm (see Table 1). Of these identified RNAs, *xlsirts* was the first early germ plasm marker found in the stage II oocyte of *X. laevis* and has a role in cytoskeletal function for anchoring *vg1* (Kloc et al., 1993; Kloc and Etkin, 1995). Shortly after this discovery, *nanos* (*xcat-2*) was also identified as an early germ plasm marker, and was shown to be expressed in germ plasm as early as stage 1 oocytes up until embryonic stage 34 (in contrast to vasa, which can only be used to identify germ plasm after stage 46) (Mosquera et al., 1993; Komiya et al., 1994; Forristal et al., 1995; Lai et al., 2011). Nanos was later found to be required for proper germ plasm formation, and localization in Xenopus (Mosquera et al., 1993). Although early investigations using single cDNAs of vegetally localized messages gave us some insight into the germ plasm
Table 1: Summary of Known vegetally localized RNAs found prior to the use of microarray. RNAs with previously described vegetal localization are listed with associated gene information. GeneCards was used to indicate gene functions.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>Citation</th>
<th>Unigene</th>
<th>Gene Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>grip2</td>
<td>Glutamate receptor interacting protein 2</td>
<td>Kaneshiro et al. (2007)</td>
<td>Xl.64243</td>
<td>PDZ domain</td>
</tr>
<tr>
<td>nanos1</td>
<td>Nanos homolog 1 (xcat2)</td>
<td>Mosquera et al. (1993)</td>
<td>Xl.1145</td>
<td>RNA binding</td>
</tr>
<tr>
<td>pat</td>
<td>Primordial germ cell-associated transcript protein (xpat)</td>
<td>Machado et al. (2005)</td>
<td>Xl.38</td>
<td>Germ plasm formation, positioning and maintenance</td>
</tr>
<tr>
<td>dazl</td>
<td>Deleted in azoospermia-like</td>
<td>Houston and King (2000)</td>
<td>Xl.311</td>
<td>RNA binding</td>
</tr>
<tr>
<td>germes</td>
<td>RNA binding protein with multiple splicing</td>
<td>Berekelya et al. (2003)</td>
<td>Xl.1742</td>
<td>Nucleic acid binding</td>
</tr>
<tr>
<td>wnt11b</td>
<td>Wingless-type MMTV integration site family, member 11B</td>
<td>Ku and Melton (1993)</td>
<td>Xl.24008</td>
<td>Secreted growth factor/developmental regulator</td>
</tr>
<tr>
<td>acsl1</td>
<td>Acyl-CoA synthetase long-chain family member 1 protein</td>
<td>King et al. (2005)</td>
<td>Xl.15591</td>
<td>Fatty acid CoA ligase</td>
</tr>
<tr>
<td>veget</td>
<td>VEGT protein</td>
<td>Zhang and King (1996)</td>
<td>Xl.1775</td>
<td>T-box transcription factor, mesendodermal determinant</td>
</tr>
<tr>
<td>otx1</td>
<td>Orthodenticle homeobox 1</td>
<td>Pannese et al. (2000)</td>
<td>Xl.781</td>
<td>Homeodomain transcription factor</td>
</tr>
<tr>
<td>ldrap1</td>
<td>Low density lipoprotein receptor adaptor protein 1(arhbeta)</td>
<td>Zhou et al. (2004)</td>
<td>Xl.8355</td>
<td>Adaptor protein CED-6, contains PIB domain</td>
</tr>
<tr>
<td>bicc1</td>
<td>BicC family RNA binding protein 1</td>
<td>Wessely and De Robertis (2000)</td>
<td>Xl.641</td>
<td>Putative RNA-binding protein. Acts as a negative regulator of Wnt signaling. May be involved in regulating gene expression during embryonic development</td>
</tr>
<tr>
<td>efnb1</td>
<td>Ephrin receptor ligand Ephrin-B1 (ephrinB1)</td>
<td>Betley et al. (2002)</td>
<td>Xl.302</td>
<td>Crucial for migration, repulsion and adhesion during neuronal, vascular and epithelial development</td>
</tr>
<tr>
<td>rbpms2</td>
<td>RNA-binding protein with multiple splicing 2 (hermes)</td>
<td>King et al. (2005)</td>
<td>Xl.449</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>plin2</td>
<td>Perilipin family member 2 (fatvg)</td>
<td>Chan et al. (2001)</td>
<td>Xl.29614</td>
<td>Involved in lipid storage and adipose tissue formation. May have role in fat body development in Xenopus.</td>
</tr>
<tr>
<td>gdf1</td>
<td>Transforming growth differentiation factor 1 (vg1)</td>
<td>Rabegliati et al. (1985)</td>
<td>Xl.25780</td>
<td>TGF-beta growth related peptide growth factor</td>
</tr>
<tr>
<td>brtc</td>
<td>Beta-transducin repeat containing (betatrcp)</td>
<td>Hudson et al. (1996)</td>
<td>Xl.968</td>
<td>F-box protein/ubiquitin protein ligase complex component. Functions in phosphorylation dependent ubiquitination</td>
</tr>
<tr>
<td>ddx59</td>
<td>DEAD-box helicase 59 (centroid)</td>
<td>Kloc and Chan (2007)</td>
<td>Xl.3767</td>
<td>ATP-dependent RNA helicase</td>
</tr>
<tr>
<td>xsirt</td>
<td>Family of interdispersed repeat RNAs</td>
<td>Kloc et al. (1993)</td>
<td>NA</td>
<td>IncRNAs containing 79-81nt repeats within Xenopus. Involved in cortical retention of Vg1 RNA.</td>
</tr>
<tr>
<td>velo1</td>
<td>Velo protein 1 (xvelo)</td>
<td>Claussen and Pieler (2004)</td>
<td>Xl.491</td>
<td>Disordered protein with prion like domain at the N-terminal. Crucial for mitochondrial cloud assembly in Xenopus</td>
</tr>
</tbody>
</table>
components, the development of genomic approaches, such as microarray and RNA-seq, provide opportunities to analyze germ plasm rich samples for multiple transcripts at a time.

*Xenopus laevis* is a prime model for investigating germ plasm composition because of its large size (stage VI oocyte diameter is ~1.2mm) and ability for manual manipulation. As discussed previously, germ plasm is tightly attached to the vegetal cortex by actin and intermediate filaments (Pondel and King, 1988; Kloc et al., 2007). These qualities allow for the unique opportunity to obtain samples rich in germ plasm through manual removal of the vegetal cortex from the oocyte. Horvay et al. was the first group to utilize these qualities of *Xenopus* for a genomic approach to identify localized RNAs within the oocyte using microarray technology (2006). However, this group only identified one novel RNA localized to the vegetal pole, *dead end(dnd1)*. Cuykendall and Houston utilized microarray maximally and identified 275 transcripts having a vegetal enrichment of 1.5-fold or higher when compared to intact oocytes (2010). Further investigation was performed on 27 of the identified vegetally localized transcripts for validation, revealing 3 as having germ plasm specific expression (*reticulon 3.1, exd2*, and a putative non-coding RNA).

Cuykendall and Houston (2010) conducted an exceptional study, identifying an unprecedented number of vegetally localized transcripts using microarray analysis, however, this approach has its limitations. First, microarrays are limited to the set of mRNAs used to make the array and therefore is not an unbiased tool for identification. Second, the vegetal samples were compared to the entire oocyte for differential expression. This approach is problematic because the whole oocyte includes the nucleus;
nuclear components are known to drive gene regulation and are rich in transcription factors. If germ plasm contains all the necessary determinants for PGCs, transcription factors for regulating specific transcriptional programs in PGCs should be present. By including the nucleus in the sample for comparison with the vegetal pole, enrichment of such factors would be lost. Future germ plasm-specific studies must take notice of these limitations to more accurately identify germ plasm and vegetally localized determinants in the oocyte.

Although microarrays can test samples for enrichment of many thousands of RNAs at one time, the development of high-throughput analytical tools, such as RNA-seq, have provided an ever-increasing ability for genome-wide studies. With RNA-seq, sequencing data is immortalized and has the sensitivity to detect very low-expressing transcripts within a sample. Acquired data sets can also be used to build novel transcriptomes (termed, de novo) specific to each individual data set. Additionally, data sets can be infinitely compared with previously published data sets of interest. However, limitations for RNA-seq also exist. One of the most prominent limitations is cost: samples for RNA-seq can cost several thousands of dollars to process. Another limitation is that only expressed transcripts will be obtained within data sets, omitting repressed genomic messages. Positive and negative aspects for RNA-seq analysis go beyond what is listed here, but it remains the best available tool for comprehensive transcript expression experiments.

The first group to utilize RNA-seq for investigating localization patterns of RNAs within the *X. laevis* oocyte was Claussen et al. (2015). By investigating animal and vegetal halves of the oocyte, a discontinuity of localized transcripts between *Xenopus*
*laevis* and *Xenopus tropicalis* was revealed. Additionally, 114 new transcripts were found to be vegetally localized in the oocyte with a 2-fold enrichment compared to the animal half \((\log_2 FC \geq 2)\). Although this group identified many vegetally localized RNAs, a specificity for germ plasm association of RNAs was not conducted. However, *in situ* validation studies did reveal that 26/34 tested transcripts localized to the germ plasm of stage I-VI oocytes via either the early (14 transcripts) or late (12 transcripts) pathway.

The work by Claußen et al. undeniably increased the number of identified maternally localized RNAs, but the utility of their study is limited for germ plasm specific information (2015). The first limitation is their analyzed samples include the complete animal or vegetal half of the oocyte. As previously mentioned, nuclear factors will be included, which may limit the finding of vegetally localized transcripts in germ plasm. Additionally, this group used a threshold of \(\log_2 FC \geq 2\) for identifying vegetally-enriched transcripts. The fact that the majority of yolk platelets are deposited in the vegetal hemisphere of the oocyte, limiting cytoplasmic volume, is problematic (Danilchik and Gerhart, 1987; Callen et al., 1980; Rebagliati et al., 1985). A low threshold for vegetally localized RNA identification may incorrectly assign a transcript as localized. This may explain why 8 of the 34 ‘localized’ transcripts that were investigated did not show any localization pattern after *in situ* analysis.

Prior to Claußen et al. (2015), De Domenico et al. (2015) identified localized RNAs within each blastomere of the 8-cell *X. tropicalis* embryo. This group provides significant findings of vegetally localized RNAs that preceded what Claußen et al. reported (2015). In total, 448 vegetally localized transcripts were found using \(\log_2 FC \geq 2\). A major finding of this study, and confirmed by Claußen et al., shows that vegetal localization and
enrichment of many maternal messages remains after fertilization. Once again, maternally inherited messages are essential for germline specification in the embryo. These studies provide information for building gene regulatory networks that may be active in the germ plasm, and necessary for germline development. However, neither of these studies were optimized for investigating germ plasm components. In the following chapters, I will discuss the results from my work, which incorporates the previously mentioned changes necessary to paint a clearer picture of the germ plasm landscape.

Another pressing question in the field of germ cell biology is: what turns on the zygotic transcription program for the germline. We anticipate that a maternal germ plasm component(s) must be necessary to initiate zygotic transcription in the PGCs. Therefore, we can utilize the findings of these studies, to look for possible candidates that may be required maternal, transcription initiators.

Vegetally localized transcription factors are the obvious candidates for investigating possible roles in activating the zygotic PGC transcriptome. A maternally inherited transcription factor(s) is likely responsible for initiating the zygotic transcription program to specify PGCs. Interestingly, among all three studies discussed (Houston et al., 2010; Claußen et al., 2015; and De Domenico et al., 2015), few maternally inherited transcription factors were identified. Of those represented, sox7 was one of the highest expressed by both RNA-seq investigations. This transcription factor belongs to the SoxF TF family and is known to activate mesoendoderm genes in Xenopus (Charney et al., 2017). Sox17, another SoxF type TF, is also known to be required for endoderm specification in humans and frogs (Zhang et al., 2005b). Additionally, SOX17 was shown to regulate the germline genes including DND1, those of the Nanos family, pluripotency
factors including OCT4 (POU5F1), and Nanog (Irie et al., 2015). Taken together, these data suggest that Sox7 may play a similar role in Xenopus PGC development as SOX17 does in human PGCs.

In the following chapters, I report data my colleagues and I derived through our own transcriptome analysis used to elucidate maternal transcripts that may be functional in germ plasm. Using RNA-seq, we identified 411 transcripts, 198 of which were annotated, and 27 transcripts, 15 of which were annotated, to be enriched at the vegetal and animal poles, respectively, of the stage VI oocyte in *X. laevis*. We also validated vegetally-enriched transcripts within our RNA-seq data by WISH, revealing both germ plasm and somatic expression patterns of 17 maternal RNAs, highlighting transcripts of neural origin. Gene regulatory network analysis of our vegetally-enriched transcripts revealed a gene network suspected to be functioning, at least to some extent, within the vegetal pole to establish the future germline. This network guided my thesis work to further investigate *sox7*, and its potential role in PGC development. Additionally, I, with the help of my colleagues, investigated *efnb1* because biological function analysis has shown that it plays a role in neuronal migration. Our investigations show evidence that *efnb1* has a role in the migration activity of PGCs in addition to a novel developmental role. This thesis will discuss these findings in detail in the following chapters.
Chapter 2

HIGH-THROUGHPUT ANALYSIS IDENTIFIES MATERNAL GERMLINE RNAs

OVERVIEW

During oogenesis, hundreds of maternal RNAs are selectively localized to the animal or vegetal pole, including determinants of somatic and germline fates. Although microarray analysis has identified localized determinants, it is not comprehensive and is limited to known transcripts. Here, we utilized high-throughput RNA sequencing analysis to comprehensively interrogate animal and vegetal pole RNAs in the fully grown *Xenopus laevis* oocyte. We identified 411 (198 annotated) and 27 (15 annotated) enriched mRNAs at the vegetal and animal pole, respectively. Ninety were novel mRNAs over 4-fold enriched at the vegetal pole and six were over 10-fold enriched at the animal pole. Unlike mRNAs, microRNAs were not asymmetrically distributed. Whole-mount *in situ* hybridization confirmed that all 17 selected mRNAs were localized. Biological function and network analysis of vegetally enriched transcripts identified protein-modifying enzymes, receptors, ligands, RNA-binding proteins, transcription factors and co-factors with five defining hubs linking 47 genes in a network. We propose potential pathways operating at the vegetal pole that highlight where future investigations might be most fruitful.
BACKGROUND

For many organisms, oogenesis is a protracted affair during which RNAs and proteins are synthesized and stored for later use during embryogenesis. Patterning of the early Xenopus embryo is determined by these components and includes specification of the three axes (animal/vegetal, dorsal/ventral, left/right), the three primary germ layers, and the germ cell lineage (King, 2014). The animal/vegetal (A/V) axis is the first to be established and specifies where the three primary germ layers will arise in the embryo. Visible signs of A/V polarity are obvious in the stage I oocyte as the Balbiani body (BB) (or mitochondrial cloud) forms in close association with the nucleus and faces the future vegetal pole. The BB contains the maternal stockpile of mitochondria as well as the germline determinants embedded within germ plasm. Later in oogenesis, the BB components accumulate at the vegetal pole, becoming tightly associated with the subcortical region.

The identity of the maternal RNAs and proteins that participate in embryonic patterning, and thus normal development, are of great interest. Initial screens selecting mRNAs enriched at either pole identified both somatic determinants, such as vg1, vegt and wnt11, and germ cell determinants including nanos1, deadsouth, xdazl and xpat (Mowry, 1996; King, 2014; Agüero et al., 2017). These mRNAs defined two patterns of RNA localization during oogenesis that appeared to align with their embryonic functions: BB-localized RNAs that function in germline identity (early pathway); and RNAs that are uniformly distributed in stage I but vegetally localized during stages II-IV (late pathway) and function in somatic patterning. However, as the number of known localized RNAs increased, some were found that used both the early and late pathways (hermes, fatvg), or
used the late pathway but, after fertilization, were found only in the germ plasm (dead-end). Loss-of-function studies revealed that these RNAs indeed have multiple functions important to both somatic and germ cell lineages (Houston, 2013).

How the asymmetric distribution of maternal RNA controls embryonic patterning represents a key area of research in developmental biology. Recent microarray data using cortical RNAs as probes have identified several hundred transcripts at the vegetal pole and many fewer localized at the animal pole (Cuykendall and Houston, 2010). Although microarrays have identified transcripts localized to the vegetal cortex and to germ plasm, this type of analysis is limited in sensitivity and to known transcripts. A comprehensive analysis identifying the RNAs, both coding and non-coding, that are significantly enriched at either the animal or vegetal pole is an important first step towards understanding the maternal contribution to embryonic patterning.

In the present study, we utilized high-throughput RNA sequencing (RNA-seq) analysis to interrogate both animal and vegetal pole localized RNAs in the fully-grown oocyte. We identified 411 vegetally localized mRNAs and, of those, 198 were previously identified genes currently in the Xenbase database (Karpinka et al., 2015). Analysis of vegetally enriched transcripts identified receptors, ligands, RNA binding proteins, protein modifying enzymes and transcription factors, as well as defined gene hubs. Functional analysis of key genes confirmed their roles in primordial germ cell (PGC) development. We also identified eight microRNAs (miRNAs), all uniformly distributed, suggesting that early embryonic patterning is not regulated by localized maternal miRNAs but rather their localized mRNA targets. Analysis of noncoding RNAs must await further annotation of the *Xenopus tropicalis* or *laevis* genome. Here, we present a comprehensive
analysis of identified RNAs found enriched at either the animal or vegetal pole. Our findings strongly support the vegetal pole as a major signaling center that patterns the early embryo.

RESULTS

2.1 RNA-seq analysis of vegetal and animal poles

To identify transcripts localized at either the vegetal or animal pole, RNA was isolated from the respective poles (each comprising ~10-20% of total oocyte) of stage VI *X. laevis* oocytes and subjected to RNA-seq analysis. A total of six samples, comprising three vegetal and three oocyte-matched animal poles, were included in the analysis as described in Materials and Methods. The total number of reads for all three samples of vegetal and matched animal poles were virtually identical, revealing sample precision (Fig. 2.1 A). The reads were aligned to the version 7.1 *X. laevis* genome (Xenbase.org) and principal component analysis (PCA) was performed on the normalized results. Two-dimensional PCA showed that transcripts identified in the vegetal pole samples cluster together and away from the animal pole samples, which also cluster together (Fig. 2.1 B). The identified transcripts having an FDR<0.05 and an FPKM≥5 were used to generate a scatter plot (Fig. 2.1 C). The data support and extend previous analyses that show both a greater complexity and fold enrichment of RNAs at the vegetal pole in comparison to the animal pole (Cuykendall and Houston, 2010; De Domenico et al., 2015). As expected, mRNAs known to be localized at the vegetal pole (*nanos1, dazl, ddx25/deadsouth*) and animal pole [*dand5 (coco) and slc18a2 (vtmat2)*] were identified as well as novel mRNAs.
Fig. 2.1. RNA-seq analysis of vegetal versus animal pole transcripts in stage VI X. laevis oocytes. (A) Total read counts for oocyte-paired vegetal and animal pole samples. Bars of the same color represent vegetal (n=3) and animal (n=3) pole samples extracted from the same oocytes. (B) Two-dimensional principal component analysis of vegetal and animal pole transcripts. Vectors V1, V2, V3 and A1, A2, A3 represent vegetal and animal pole samples, respectively. (C) Scatter plot comparing vegetal and animal pole transcripts. (D) Differential expression analysis of vegetal versus animal pole transcripts with FDR≤0.05 and FPKM≥5. GV, germinal vesicle.

Over five thousand transcripts (5717) were found differentially expressed between the animal and vegetal poles based on a minimum q-value of 0.05 and 2-fold change.
Transcripts were considered significantly enriched at the vegetal pole if they had at least a 4-fold increase compared with the animal pole. The stringent criteria set yielded a total of 411 vegetally enriched transcripts, 198 of which were annotated (Fig. 2.1D, Table 2.1 [http://www.biologists.com/DEV_Movies/DEV139220/TableS1.xlsx]). Of the 198 transcripts identified, 38 have been shown to be vegetally localized, with 23 of them being specifically associated with germ plasm (Cuykendall and Houston, 2010; Claussen et al., 2015; De Domenico et al., 2015). Transcripts were considered localized at the animal pole if they were at least 10-fold enriched compared with the vegetal pole. Under these conditions, 27, including 15 annotated, mRNAs were enriched at the animal pole (Fig. 2.1D). All annotated and unannotated transcripts can be found in Table 2.2 (http://www.biologists.com/DEV_Movies/DEV139220/TableS2.xlsx) and Table 2.3 (http://www.biologists.com/DEV_Movies/DEV139220/TableS3.xlsx), respectively.

2.2 Biological process and network analysis of vegetally enriched transcripts

To identify possible gene functions, the 198 annotated vegetal mRNAs were manually data mined using GeneCards (www.genecards.org). Table 2.4 shows the 40 most enriched transcripts in the vegetal pole. Ten categories were established based on function (Fig. 2.2A). The top six categories were: signal transduction (26%), transport (13%), transcription (7%), cytoskeletal related (8%), the ubiquitin pathway (7%) and cell cycle (7%). Enzymes are often key players in regulating gene pathways; therefore, we also identified and categorized the enzymes represented in our vegetally enriched data set. Enzymes represent 60/198 (30%) of localized transcripts. Nine categories of enzymes
Table 2.4. Gene function of the 40 most highly enriched mRNAs identified at the oocyte vegetal pole.
GeneCards was used to indicate gene functions. ‡Nine novel mRNAs enriched at the vegetal pole and associated with germ plasm are identified. *The highest expressing isoform was used when multiple isoforms were identified. V/A, vegetal/animal.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>JGI</th>
<th>Unigene</th>
<th>Fold Change (V/A)</th>
<th>Gene Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>grip2*</td>
<td>Glutamate receptor interacting protein 2</td>
<td>Xelaev16034487m.g</td>
<td>Xl.64243</td>
<td>656.02</td>
<td>PDZ domain</td>
</tr>
<tr>
<td>nanos1</td>
<td>Nanos homolog 1</td>
<td>Xelaev16026544m.g</td>
<td>Xl.1145</td>
<td>391.89</td>
<td>RNA binding</td>
</tr>
<tr>
<td>rras2‡</td>
<td>Related RAS viral (r-ras) oncogene homolog 2</td>
<td>Xelaev16056461m.g</td>
<td>Xl.47312</td>
<td>360.11</td>
<td>Ras-related GTPase</td>
</tr>
<tr>
<td>kif13b*</td>
<td>Kinesin family member 13B</td>
<td>Xelaev16027627m.g</td>
<td>Xl.55605</td>
<td>267.80</td>
<td>Kinesin-like protein</td>
</tr>
<tr>
<td>pat</td>
<td>Primordial germ cell-associated transcript protein</td>
<td>Xelaev16012806m.g</td>
<td>Xl.38</td>
<td>226.17</td>
<td>Germ plasm formation, positioning and maintenance</td>
</tr>
<tr>
<td>dazl</td>
<td>Deleted in azoospermia-like</td>
<td>Xelaev16047101m.g</td>
<td>Xl.311</td>
<td>217.44</td>
<td>RNA binding</td>
</tr>
<tr>
<td>sybu*</td>
<td>Syntabulin (syntaxin-interacting)</td>
<td>Xelaev16066812m.g</td>
<td>Xl.8441</td>
<td>175.30</td>
<td>Nucleolar GTPase/ATPase</td>
</tr>
<tr>
<td>rtn3</td>
<td>Reticulon 3</td>
<td>Xelaev16042166m.g</td>
<td>Xl.57382</td>
<td>162.84</td>
<td>Reticulon</td>
</tr>
<tr>
<td>trank1‡</td>
<td>Tetratricopeptide repeat and ankyrin repeat containing 1</td>
<td>Xelaev16058483m.g</td>
<td>NA</td>
<td>154.58</td>
<td>Translocase of outer mitochondrial membrane complex, subunit</td>
</tr>
<tr>
<td>germes</td>
<td>RNA binding protein with multiple splicing</td>
<td>Xelaev16035521m.g</td>
<td>Xl.17472</td>
<td>141.43</td>
<td>Nucleic acid binding</td>
</tr>
<tr>
<td>wnt11b</td>
<td>Wingless-type MMTV integration site family, member 11B</td>
<td>Xelaev16048139m.g</td>
<td>Xl.24008</td>
<td>139.89</td>
<td>Secreted growth factor/developmental regulator</td>
</tr>
<tr>
<td>trim36*</td>
<td>Tripartite motif containing 36</td>
<td>Xelaev16048562m.g</td>
<td>Xl.6926</td>
<td>127.93</td>
<td>Zinc finger, B-box</td>
</tr>
<tr>
<td>farp2</td>
<td>FERM, RhoGEF and pleckstrin domain protein 2</td>
<td>Xelaev16049032m.g</td>
<td>Xl.53782</td>
<td>111.93</td>
<td>Rho guanine nucleotide exchange factor CDEP</td>
</tr>
<tr>
<td>cnppd1</td>
<td>Cyclin Pas1/PHO80 domain containing 1</td>
<td>Xelaev16027858m.g</td>
<td>Xl.7190</td>
<td>93.24</td>
<td>Cyclin</td>
</tr>
<tr>
<td>slain1‡</td>
<td>SLAIN motif family, member 1</td>
<td>Xelaev16028782m.g</td>
<td>NA</td>
<td>89.96</td>
<td>Microtubule plus-end tracking protein</td>
</tr>
<tr>
<td>srgap1‡</td>
<td>SLIT-ROBO Rho GTPase activating protein 1</td>
<td>Xelaev16070447m.g</td>
<td>NA</td>
<td>78.71</td>
<td>Cdc42-interacting protein CIP4</td>
</tr>
<tr>
<td>ddx25</td>
<td>DEAD (Asp-Glu-Ala-Asp) box helicase 25</td>
<td>Xelaev16075440m.g</td>
<td>Xl.670</td>
<td>75.17</td>
<td>ATP-dependent RNA helicase</td>
</tr>
<tr>
<td>pcox6</td>
<td>Proprotein convertase subtilisin/kexin type 6</td>
<td>Xelaev16063723m.g</td>
<td>Xl.48635</td>
<td>71.99</td>
<td>Protease, proprotein convertase</td>
</tr>
<tr>
<td>gk</td>
<td>Glycerol kinase</td>
<td>Xelaev16073000m.g</td>
<td>Xl.1742</td>
<td>71.20</td>
<td>Ribulose kinase and related carbohydrate kinases</td>
</tr>
<tr>
<td>acsl1</td>
<td>Acyl-CoA synthetase long-chain family member 1 protein</td>
<td>Xelaev16052929m.g</td>
<td>Xl.15591</td>
<td>52.76</td>
<td>Fatty acid CoA ligase</td>
</tr>
<tr>
<td>slc12a9</td>
<td>Solute carrier family 12, member 9</td>
<td>Xelaev16065485m.g</td>
<td>Xl.59798</td>
<td>52.58</td>
<td>Amino acid transporter</td>
</tr>
<tr>
<td>wnk2</td>
<td>WNK lysine deficient protein kinase 2</td>
<td>Xelaev16065595m.g</td>
<td>Xl.50299</td>
<td>51.69</td>
<td>Serine/threonine protein kinase</td>
</tr>
<tr>
<td>vegt*</td>
<td>Vegt protein</td>
<td>Xelaev16018006m.g</td>
<td>Xl.1775</td>
<td>51.34</td>
<td>T-box transcription factor, meso-endodermal determinant</td>
</tr>
<tr>
<td>cdr2l‡</td>
<td>Cerebellar degeneration-related protein 2-like</td>
<td>Xelaev16042542m.g</td>
<td>Xl.9895</td>
<td>50.48</td>
<td>Myosin class II heavy chain</td>
</tr>
<tr>
<td>tesk2‡</td>
<td>Testis-specific kinase 2</td>
<td>Xelaev16001025m.g</td>
<td>Xl.71668</td>
<td>46.74</td>
<td>Endosomal membrane proteins, EMP70</td>
</tr>
<tr>
<td>cd5r2</td>
<td>Cyclin-dependent kinase 5, regulatory subunit 2 (p39)</td>
<td>Xelaev16007720m.g</td>
<td>Xl.31022</td>
<td>46.60</td>
<td>CDK5 kinase activator</td>
</tr>
<tr>
<td>otx1</td>
<td>Orthodenticle homeobox 1</td>
<td>Xelaev16005310m.g</td>
<td>Xl.781</td>
<td>46.51</td>
<td>Homeodomain transcription factor</td>
</tr>
<tr>
<td>Gene ID</td>
<td>Gene Name</td>
<td>JGI</td>
<td>Unigene</td>
<td>Fold Change (V/A)</td>
<td>Gene Function</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>rhobtb2</td>
<td>Rho-related BTB domain containing 2</td>
<td>Xelaev16065046m.g</td>
<td>X1.12379</td>
<td>45.82</td>
<td>Ras-related small GTPase</td>
</tr>
<tr>
<td>tmem65</td>
<td>Transmembrane protein 65</td>
<td>Xelaev16028617m.g</td>
<td>X1.55768</td>
<td>45.36</td>
<td>Uncharacterized conserved protein</td>
</tr>
<tr>
<td>mov10</td>
<td>Mov10 RISC complex RNA helicase</td>
<td>Xelaev16074505m.g</td>
<td>X1.52320</td>
<td>45.22</td>
<td>RNA helicase</td>
</tr>
<tr>
<td>trib1</td>
<td>Tribbles pseudokinase 1</td>
<td>Xelaev16002738m.g</td>
<td>X1.75411</td>
<td>44.66</td>
<td>Serine/threonine protein kinase</td>
</tr>
<tr>
<td>ppp1r2</td>
<td>Protein phosphatase 1, regulatory (inhibitor) subunit 2</td>
<td>Xelaev16041705m.g</td>
<td>X1.52180</td>
<td>44.20</td>
<td>Protein phosphatase 1, regulatory (inhibitor) subunit PPP1R2</td>
</tr>
<tr>
<td>sox7</td>
<td>SRY (sex determining region Y)-box 7</td>
<td>Xelaev16000442m.g</td>
<td>X1.1241</td>
<td>42.25</td>
<td>HMG-box transcription factor</td>
</tr>
<tr>
<td>ldrap1-b</td>
<td>Low density lipoprotein receptor adaptor protein 1</td>
<td>Xelaev16018841m.g</td>
<td>X1.8355</td>
<td>40.88</td>
<td>Adaptor protein CED-6, contains PTB domain</td>
</tr>
<tr>
<td>paqr6†</td>
<td>Progestin and adipoQ receptor family member VI</td>
<td>Xelaev16007491m.g</td>
<td>X1.57243</td>
<td>39.20</td>
<td>Predicted membrane proteins, contain hemolysin III domain</td>
</tr>
<tr>
<td>sf3a1†</td>
<td>Splicing factor 3a, subunit 1, 120kDa</td>
<td>Xelaev16054742m.g</td>
<td>X1.34155</td>
<td>37.67</td>
<td>Required for pre-mRNA splicing</td>
</tr>
<tr>
<td>dnd1</td>
<td>DND microRNA-mediated repression inhibitor 1</td>
<td>Xelaev16016713m.g</td>
<td>X1.29785</td>
<td>37.50</td>
<td>Nuclear ribonucleoprotein R (RRM superfamily)</td>
</tr>
<tr>
<td>fam168b</td>
<td>Family with sequence similarity 168, member B</td>
<td>Xelaev16019587m.g</td>
<td>X1.3240</td>
<td>36.47</td>
<td>RNA polymerase II, large subunit</td>
</tr>
<tr>
<td>hook2</td>
<td>Hook microtubule-tethering protein 2</td>
<td>Xelaev16057970m.g</td>
<td>X1.9790</td>
<td>34.05</td>
<td>Uncharacterized coiled-coil protein</td>
</tr>
<tr>
<td>bicc1</td>
<td>BicC family RNA binding protein 1</td>
<td>Xelaev16022482m.g</td>
<td>X1.641</td>
<td>25.70</td>
<td>Putative RNA-binding protein. Acts as a negative regulator of Wnt signaling. May be involved in regulating gene expression during embryonic development</td>
</tr>
</tbody>
</table>

were identified, with kinases (18%), metabolism related (17%), ubiquitin pathway (13%) and ATPases/GTPases (12%) making up the majority (Fig. 2.2B).

Vegetally localized mRNAs were subject to gene pathway and network analysis by GeneGo. Significantly related gene ontology (GO) processes were grouped into seven categories based on gene expression (Table 2.5). Consistent with a role in embryonic patterning, these categories included: developmental processes, signaling regulation, localization, phosphate metabolic processes, cellular protein metabolic processes, cell cycle, and gamete generation. Interestingly, genes involved in neurogenic processes such
as neuroblast proliferation, including \textit{fgfr2}, \textit{frizzled1} and \textit{ephrinB1} (\textit{efnb1}), were well represented in our data set, composing 12\% of annotated genes (Table 2.2).

We next investigated potential gene networks present within the 198 vegetally enriched transcripts. Using MetaCore analysis (GeneGo), we identified 47 genes that form a direct interaction network (Fig. 2.2C). These genes encode protein-modifying enzymes [\textit{caspase 3 (casp3)}, \textit{cathepsin C (ctsc)}, \textit{pcs6/pase4}, \textit{tesk1/2}, \textit{senp1}], receptors [\textit{fgfr2}, \textit{frizzled1 (fzd1)}, \textit{a2mr/lrp1}], ligands (\textit{wnt11}), and five key transcription factors or co-factors (\textit{e2f1}, \textit{irf8}, \textit{err1/esrra}, \textit{p300/ep300} and \textit{sox7}), the first four of which represent
Table 2.5. GO processes of vegetally enriched transcripts.
All 198 vegetally localized transcripts were subject to GeneGo analysis. GO processes are categorized based on the greatest number of enriched transcripts identified. Seven primary categories are in bold.

<table>
<thead>
<tr>
<th>GeneGo Processes</th>
<th>p-value</th>
<th>FDR</th>
<th># of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental process</td>
<td>3.42E-05</td>
<td>3.28E-03</td>
<td>78</td>
</tr>
<tr>
<td>Cell development</td>
<td>2.58E-04</td>
<td>9.42E-03</td>
<td>32</td>
</tr>
<tr>
<td>Neurotrophin signaling pathway</td>
<td>2.38E-03</td>
<td>3.71E-02</td>
<td>9</td>
</tr>
<tr>
<td>Neuroblast proliferation</td>
<td>3.23E-03</td>
<td>4.56E-02</td>
<td>3</td>
</tr>
<tr>
<td>Regulation of signaling</td>
<td>9.13E-06</td>
<td>1.61E-03</td>
<td>50</td>
</tr>
<tr>
<td>Intracellular signal transduction</td>
<td>3.02E-08</td>
<td>9.56E-05</td>
<td>40</td>
</tr>
<tr>
<td>Transmembrane receptor protein tyrosine kinase signaling pathway</td>
<td>1.09E-04</td>
<td>6.11E-03</td>
<td>17</td>
</tr>
<tr>
<td>Signal transduction by p53 class mediator</td>
<td>1.83E-03</td>
<td>3.14E-02</td>
<td>6</td>
</tr>
<tr>
<td>Regulation of localization</td>
<td>2.80E-06</td>
<td>9.85E-04</td>
<td>42</td>
</tr>
<tr>
<td>Regulation of protein localization</td>
<td>4.15E-06</td>
<td>3.11E-03</td>
<td>19</td>
</tr>
<tr>
<td>Regulation of intracellular transport</td>
<td>7.49E-06</td>
<td>3.48E-03</td>
<td>16</td>
</tr>
<tr>
<td>Regulation of phosphate metabolic process</td>
<td>3.77E-05</td>
<td>2.18E-03</td>
<td>38</td>
</tr>
<tr>
<td>Regulation of phosphorylation</td>
<td>8.16E-05</td>
<td>4.55E-03</td>
<td>28</td>
</tr>
<tr>
<td>Regulation of protein serine/threonine kinase activity</td>
<td>2.07E-04</td>
<td>8.71E-03</td>
<td>14</td>
</tr>
<tr>
<td>Regulation of cellular protein metabolic process</td>
<td>5.22E-07</td>
<td>4.08E-04</td>
<td>38</td>
</tr>
<tr>
<td>Lipid transport</td>
<td>7.32E-05</td>
<td>5.15E-03</td>
<td>10</td>
</tr>
<tr>
<td>Positive regulation of deacetylase activity</td>
<td>8.68E-05</td>
<td>5.39E-03</td>
<td>3</td>
</tr>
<tr>
<td>Regulation of cell cycle</td>
<td>7.08E-06</td>
<td>1.48E-03</td>
<td>24</td>
</tr>
<tr>
<td>Negative regulation of cell cycle</td>
<td>6.11E-06</td>
<td>6.37E-04</td>
<td>14</td>
</tr>
<tr>
<td>Negative regulation of nuclear division</td>
<td>2.88E-05</td>
<td>3.06E-03</td>
<td>6</td>
</tr>
<tr>
<td>Cell cycle checkpoint</td>
<td>2.08E-03</td>
<td>3.37E-02</td>
<td>8</td>
</tr>
<tr>
<td>Gamete generation</td>
<td>6.10E-05</td>
<td>4.55E-03</td>
<td>19</td>
</tr>
<tr>
<td>Sexual reproduction</td>
<td>1.04E-04</td>
<td>6.00E-03</td>
<td>21</td>
</tr>
<tr>
<td>Female gamete generation</td>
<td>2.73E-04</td>
<td>9.75E-03</td>
<td>7</td>
</tr>
<tr>
<td>Spermatogenesis</td>
<td>1.10E-03</td>
<td>2.35E-02</td>
<td>14</td>
</tr>
<tr>
<td>Multicellular organismal reproductive process</td>
<td>1.70E-04</td>
<td>7.97E-03</td>
<td>22</td>
</tr>
</tbody>
</table>

network hubs. These mRNAs were validated as vegetally localized by either RT-qPCR (Fig. 2.2D) or WISH (Fig. 2.3). Published studies on these factors suggest their involvement in regulating the cell cycle (e2f1), endoderm specification (sox7), metabolic pathways (errl) and lineage commitment (irf8) (Costa et al., 2013; Johansen et al., 2016; Minderman et al., 2016; Stovall et al., 2014).

2.3 miRNA target RNAs are localized at the vegetal pole

Germ plasm RNAs must be post-transcriptionally regulated for germline survival (Lai et al., 2012; reviewed by Lai and King, 2013). RNA degradation within germ plasm may be regulated by miRNAs (Bartel, 2004; Yamaguchi et al., 2014). Therefore, we mined our
data to identify vegetally localized miRNAs. Our analysis identified only eight miRNAs that were expressed in both the vegetal and animal poles: 15c, 18a, 19b, 20a, 92a, 363, 427 and 429. Surprisingly, none was significantly enriched at either pole (data not shown).

We next determined if the predicted mRNA targets of the eight identified miRNAs were vegetally localized. Interestingly, predicted targets of 7/8 miRNAs are enriched in the vegetal pole (Table 2.6). In total, 13 vegetally localized target mRNAs (listed in Table 2.6) were identified that contain at least one recognition sequence conserved between X. tropicalis and human for their respective miRNAs. These results suggest that if early embryonic patterning is regulated by miRNA activity, it is not by localizing miRNAs to the vegetal pole but rather by targeting specific vegetally localized RNAs.

2.4 Expression of vegetally localized RNAs during development

We chose 17 transcripts (xpat, efnb1, rras2, mov10, otx1, sox7, spire1, wnk2, e2f1, sybu, atrx, hook2, tob2, rnf38, trank1, wwtr1 and parn) for WISH analysis to determine their expression pattern during embryogenesis (Fig. 2.3, Fig. 2.4A,B). Consistent with our RNA-seq data, 15/17 were expressed exclusively in the vegetal pole of stage IV oocytes (Fig. 2.3, Fig. 2.4A). Not surprisingly, the two low expressing mRNAs, parn and wwtr1 (only 5- to 6.7-fold enriched compared with the animal pole) were not detected in stage IV oocytes but were detected later at blastula stage (Fig. 2.4B). Recently, another group analyzed transcript localization in the 8-cell X. tropicalis embryo by RNA-seq (De Domenico et al., 2015). Comparison of the vegetal/animal blastomeres revealed that 27
of our filtered 198 transcripts remain vegetally enriched after both fertilization and cortical rotation have occurred (De Domenico et al., 2015).

Spatial expression patterns were examined during oogenesis, the pre-midblastula transition (MBT), gastrula, neurula, and tailbud stages to determine whether localized mRNAs contribute to the future germline, soma, or both lineages. During oogenesis, three mRNA localization patterns were detected: the early or METRO pathway (Fig. 2.3A), the late Vg1-like pathway [Fig. 2.3B, Fig. 2.4A (atrx, hook2, tob2), Fig. 2.4B], or both [Fig. 2.3C, Fig. 2.4A (rnf38)] (King et al., 1999). Germ plasm-specific xpat served as a marker for germline expression (Hudson and Woodland, 1998). Regardless of the pathway used, all mRNAs were subsequently found in the germ plasm of embryos as well as in the soma (Fig. 2.3, Fig. 2.4A,B). 12/17 RNAs represented novel germline components, while xpat, sybu, otx1, tob2 and efnb1 were confirmed as previously described (Cuykendall and Houston, 2010; De Domenico et al., 2015).

Table 2.6. miRNAs and their vegetally localized mRNA targets. Vegetally localized mRNA targets are shown with their corresponding non-localized miRNA(s).

<table>
<thead>
<tr>
<th>Gene</th>
<th>FPKM</th>
<th>Fold Change (V/A)</th>
<th>Refseq</th>
<th>Gene Name</th>
<th>miRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SLIT-ROBO Rho GTPase activating protein 1</td>
<td>15c</td>
</tr>
<tr>
<td>srga</td>
<td>28</td>
<td>57.3</td>
<td>NM_020762</td>
<td></td>
<td>19b</td>
</tr>
<tr>
<td>rnf38</td>
<td>351</td>
<td>28.0</td>
<td>NM_022781</td>
<td>Ring finger protein 38</td>
<td>20a</td>
</tr>
<tr>
<td>sulf1</td>
<td>198</td>
<td>26.6</td>
<td>NM_001128204</td>
<td>Sulfatase 1</td>
<td>363</td>
</tr>
<tr>
<td>pgam1</td>
<td>420</td>
<td>20.6</td>
<td>NM_002629</td>
<td>Phosphoglycerate mutase 1 (brain)</td>
<td>427</td>
</tr>
<tr>
<td>fndc3a</td>
<td>102</td>
<td>18.2</td>
<td>NM_001079673</td>
<td>Fibronectin type III domain containing 3A</td>
<td>429</td>
</tr>
<tr>
<td>testk2</td>
<td>52</td>
<td>15.6</td>
<td>NM_007170</td>
<td>Testis-specific kinase 2</td>
<td>92a</td>
</tr>
<tr>
<td>chek1</td>
<td>219</td>
<td>14.7</td>
<td>NM_001114121.2</td>
<td>Checkpoint kinase 1</td>
<td></td>
</tr>
<tr>
<td>tob2</td>
<td>346</td>
<td>14.1</td>
<td>NM_016272</td>
<td>Transducer of ERBB2, 2</td>
<td></td>
</tr>
<tr>
<td>wasl</td>
<td>130</td>
<td>13.2</td>
<td>NM_003941</td>
<td>Wiskott-Aldrich syndrome-like</td>
<td></td>
</tr>
<tr>
<td>papola</td>
<td>250</td>
<td>8.9</td>
<td>NM_032632</td>
<td>Poly(A) polymerase alpha</td>
<td></td>
</tr>
<tr>
<td>cnot4</td>
<td>120</td>
<td>5.4</td>
<td>NM_001190847</td>
<td>CCR4-NOT transcription complex, subunit 4</td>
<td></td>
</tr>
<tr>
<td>ubec2d1</td>
<td>356</td>
<td>5.1</td>
<td>NM_001204880</td>
<td>Ubiquitin-conjugating enzyme E2D1</td>
<td></td>
</tr>
<tr>
<td>rbpms2</td>
<td>218</td>
<td>3.9</td>
<td>NM_194272</td>
<td>RNA binding protein with multiple splicing 2 (Hermes)</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 2.3. WISH of vegetal pole transcripts. The expression of a subset of vegetally enriched mRNAs was analyzed during oogenesis and embryo development by WISH. Expression patterns are grouped according to mRNA pathway: early (A), late (B) or both (C). xpat expression marks germ plasm. Probes, developmental stages and developmental structures are indicated in the illustrations at the top: germ plasm/PGCs (purple), pronephros (pink), ventral blood islands (lime green), eye (black), lens (white), otic vesicle (gray), cranial ganglia (yellow), brachial arches (green), nasal placodes (teal), intersegmental region (brown), notochord (orange), brain and neural tube (blue). Transcripts detected in PGCs at the tailbud stage are shown at higher magnification in insets. WISH analysis was performed on ≥40 total embryos from at least two adult female frogs. Dorsal (D), ventral (V), left (L), right (R), posterior (P), anterior (A), vegetal view (v); MC, mitochondrial cloud; N, nucleus; Oo, oocyte. Scale bars: 100 μm for stage I oocyte; 200 μm in remaining panels.
Figure 2.4. WISH of select vegetal and animal pole transcripts. Expression of a subset of RNAs enriched in the vegetal (A-B) and animal (C) poles were analyzed during oogenesis and development by WISH. Please refer to Fig 2.3 for developmental stage illustration and color key. Probes, developmental stages, and developmental structures are indicated. Transcripts detected in primordial germ cells at the tailbud stage are magnified and embedded in their respective images. Black bars represent 100 μm in st. I oocyte panels, and 200μm in all other panels. Please refer to Fig. 2.3 for use of xpat probe as a positive control for germ plasm localization. Dorsal (D), ventral (V), left (L), right (R), posterior (P), anterior (A), vegetal view (v); MC, mitochondrial cloud; N, nucleus; Oo, oocyte. Scale bars: 100 μm for stage I oocyte; 200 μm in remaining panels.
Except for sox7 and efnb1, germ plasm expression persisted through neurula (Fig. 2.3, Fig. 2.4A,B). During gastrulation, the germline segregates from endoderm and PGCs form a distinct lineage. By neurula, PGCs are transcriptionally active for the first time (Venkatarama et al., 2010). sox7 is expressed through gastrula stages but is lost by neurula suggesting an early role in PGC development. efnb1 is not expressed at gastrula but is re-expressed by neurula, probably as part of a new gene expression program in PGCs. PGC migration towards the dorsal mesentery and organogenesis occur at the early tailbud stage. Only 41% (7/17) of the vegetal transcripts, including xpat, remained expressed in PGCs during tailbud stage (Fig. 2.3). These transcripts included transcription factors (otx1, e2f1), RISC factor (mov10), actin regulator (spire1), Ser/Thr kinase (wnk2), and an adaptor protein that binds kinesin (sybu). They are likely to be zygotic transcripts and might play roles in migration and/or in preserving PGC totipotency.

In addition to PGC expression, 71% (12/17) are also expressed in the eye anlage and the future posterior region in neurula, including the somitogenic mesoderm (Fig. 2.3, Fig. 2.4A,B). In tailbud stages the most notable expression pattern was in neural regions including the eye, cranial ganglia, neural tube, nasal placodes, brain, otic vesicle and the intersegmental region between the somites. Vegetally enriched transcripts that have previously been shown to be involved in neural pathways included efnb1, sybu, wnk2 and otx1 (Colozza and De Robertis, 2014; Bovolenta et al., 2006; Rinehart et al., 2011; Zhang et al., 2015). Taken together, these data suggest that RNAs enriched at the vegetal pole of Xenopus oocytes contribute to both germline and neural specification during development.
2.5 Novel mRNAs enriched at the animal pole

We identified 15 mRNAs, six not previously reported, that were at least ten-fold enriched at the animal pole (Table 2.7). These mRNAs represent the following functional categories: signaling \([dand5, ifrd2, slc18a2, spata13, acaca, tmem192, ssr1, prr11]\), gene expression \((pou2f1)\), cell division \((rmdn3)\) and metabolism \((adpgk, prrg4, prkag1)\). Two mRNAs previously shown to be enriched at the animal pole were chosen for validation by WISH: \(slc18a2\) and \(dand5\) (Fig. 2.4C). \(dand5\) is a TGFβ and Wnt antagonist (Eimon and Harland, 2001; Bates et al., 2013). \(slc18a2\) transports monoamines into secretory vesicles for eventual exocytosis (Nikishin et al., 2012). As expected, both animal pole transcripts were expressed exclusively in the animal pole of pre-MBT embryos (Fig. 2.4C). Animal pole transcripts are expressed primarily in the neural ectoderm at later developmental stages, as previously described (Fig. 2.4C) (Grant et al., 2014). MetaCore (GeneGo) direct interaction pathway analysis did not reveal integrated networks among the animal pole-enriched RNAs.
Table 2.7. Gene function of the 15 most highly enriched mRNAs identified at the oocyte animal pole. GeneCards was used to indicate gene functions. *Six novel mRNAs enriched at the animal pole are indicated.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>v7.1 JGI</th>
<th>Unigene</th>
<th>Ratio A/V</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ifrd2</td>
<td>Interferon-related developmental regulator 2</td>
<td>Xelaev16023388m.g</td>
<td>Xl. 75341</td>
<td>22.84</td>
<td>Interferon-related protein PC4 like.</td>
</tr>
<tr>
<td>sle18a2</td>
<td>Solute carrier family 18 (vesicular monoamine transporter), member 2</td>
<td>Xelaev16012438m.g</td>
<td>Xl.13611</td>
<td>17.73</td>
<td>Transmembrane solute exchange of neurotransmitters.</td>
</tr>
<tr>
<td>spata13*</td>
<td>Spermatogenesis associated 3</td>
<td>Xelaev16016628m.g</td>
<td>Xl.58825</td>
<td>16.42</td>
<td>Invasion-inducing protein TIAM1/CDC24 and related RhoGEF GTPases.</td>
</tr>
<tr>
<td>adpgk</td>
<td>ADP-dependent glucokinase</td>
<td>Xelaev16007044m.g</td>
<td>NA</td>
<td>15.43</td>
<td>Catalyzes glucose to glucose-6-phosphate via ADP.</td>
</tr>
<tr>
<td>c7orf43*</td>
<td>Chromosome 7 open reading frame 43</td>
<td>Xelaev16021226m.g</td>
<td>Xl.56042</td>
<td>15.20</td>
<td>Uncharacterized conserved protein.</td>
</tr>
<tr>
<td>prrg4</td>
<td>Proline rich Gla (G-carboxyglutamic acid) 4 (transmembrane)</td>
<td>Xelaev16039880m.g</td>
<td>Xl.4519</td>
<td>14.75</td>
<td>Trypsin.</td>
</tr>
<tr>
<td>acaca</td>
<td>Acetyl-CoA carboxylase alpha</td>
<td>Xelaev16038229m.g</td>
<td>NA</td>
<td>12.36</td>
<td>Acetyl-CoA carboxylase.</td>
</tr>
<tr>
<td>pou2f1</td>
<td>POU class 2 homeobox 1</td>
<td>Xelaev16026767m.g</td>
<td>Xl.1265</td>
<td>12.28</td>
<td>Oct1 transcription factor.</td>
</tr>
<tr>
<td>prkag1*</td>
<td>Protein kinase, AMP-activated, gamma 2 non-catalytic subunit of AMPK</td>
<td>Xelaev16029267m.g</td>
<td>Xl. 34576</td>
<td>12.00</td>
<td>AMPK helps modulate cellular energy metabolism.</td>
</tr>
<tr>
<td>xkrx</td>
<td>XK, Kell blood group complex subunit-related, X-linked</td>
<td>Xelaev16047837m.g</td>
<td>NA</td>
<td>11.72</td>
<td>Uncharacterized conserved protein.</td>
</tr>
<tr>
<td>tmem192</td>
<td>Transmembrane protein 192</td>
<td>Xelaev16056645m.g</td>
<td>Xl.19528</td>
<td>11.71</td>
<td>Uncharacterized conserved protein.</td>
</tr>
<tr>
<td>ssr1*</td>
<td>Signal sequence receptor, alpha</td>
<td>Xelaev16076712m.g</td>
<td>Xl.60983</td>
<td>11.48</td>
<td>Glycosylated ER receptor that mediates protein translocation.</td>
</tr>
<tr>
<td>prr11*</td>
<td>Proline rich 11</td>
<td>Xelaev16035856m.g</td>
<td>NA</td>
<td>11.10</td>
<td>Rac1 GTPase effector FRL. Cell Cycle.</td>
</tr>
<tr>
<td>dand5</td>
<td>DAN domain family member 5, BMP antagonist</td>
<td>Xelaev16057997m.g</td>
<td>Xl.51884</td>
<td>11.04</td>
<td>Maternal BMP/TGFbeta/Wnt inhibitor.</td>
</tr>
<tr>
<td>rmdn3*</td>
<td>Regulator of microtubule dynamics 3</td>
<td>Xelaev16028196m.g</td>
<td>Xl.9583</td>
<td>10.15</td>
<td>Nuclear localization sequence binding. Cellular calcium regulation.</td>
</tr>
</tbody>
</table>

DISCUSSION

Here we report the first interrogation of RNAs within the vegetal and animal poles by RNA-seq. WISH revealed all 17 selected mRNAs to be localized, providing strong support for the accuracy of the data sets. Our findings underscore the dramatic transcript asymmetry along the A/V axis and the importance of the vegetal pole in initiating
somatic and germline lineages in the early embryo. Importantly, as the annotation of the Xenopus genome improves, our data set can be continually mined to identify spliced variants and currently unknown transcripts.

Five important observations have emerged from our studies. (1) We identified 90 novel mRNAs that were over 4-fold enriched at the vegetal pole and six that were over 10-fold enriched at the animal pole. (2) GeneGo analysis revealed a network encompassing over 20% of the annotated vegetally enriched mRNAs, indicating great connectivity of gene function and localization. Transcription factors/co-factors e2f1, irf8, err1 and p300 defined four regulatory hubs for future analysis (Fig. 2.2C). (3) Unlike mRNAs, localization of maternal miRNAs does not appear to be a strategy employed to regulate gene expression along the A/V axis. (4) Enzymes represented 30% of the 198 enriched annotated mRNAs, underscoring the vegetal pole as a major platform for cell signaling. (5) Well over 10% of the vegetal mRNAs encode components with known functions in neurogenic pathways, including the transmembrane ligand Efnb1 and scaffold protein Grip2. Functions required in PGCs overlap with those required in the nervous system: stem cell maintenance, directed migration, regulated cell-division. It is tempting to speculate that the nervous system co-opted proteins required to build a scaffold to assemble multiprotein signaling complexes from the oocyte and germ plasm.

MetaCore (GeneGo) analysis placed over 20% (47/198) of the known vegetal pole mRNAs within a direct interaction network that identified four major hubs centered around transcription factors e2f1, irf8, err1 and the histone acetyltransferase p300 (Fig. 2.2C). Although the network is built based on data from different systems, it reveals novel regulatory pathways and candidates that can be tested for their functions in
embryogenesis. For example, it remains unknown what restricts microtubule array formation to the vegetal pole during cortical rotation. Our RNA-seq analysis identified JNK (*mapk8*) and slain1 as enriched at the vegetal pole and both have been implicated in microtubule dynamics (GeneCards).

Our RNA-seq analysis has revealed interconnected pathways highlighting the vegetal pole as a major signaling center. Interestingly, *grip2*, which encodes an important scaffolding protein in the nervous system, was the most abundant vegetal mRNA identified in our analysis. Grip2 is likely to represent a key scaffolding protein for the assembly, through its PDZ domains, of multi-protein signaling complexes. The challenge now is to functionally test the pathways revealed by our comprehensive list of localized mRNAs and to define the maternal contributions to both germline and somatic cell fates.
Chapter 3

MATERNAL GERMLINE RNAs ARE CRITICAL FOR PRIMORDIAL GERM CELL SURVIVAL AND PROPER MIGRATION

OVERVIEW

The Xenopus oocyte is a highly-polarized cell along the animal/vegetal axis. During oogenesis, specific maternal RNAs are localized to the vegetal pole that will determine germ layer identity, dorsal/ventral patterning, and the germ cell lineage in the embryo. Using RNA-seq, we have identified RNAs highly enriched at the vegetal pole. Here, we have performed initial functional studies on the vegetally enriched, maternal *ephrinB1* (*efnb1*), *e2f1*, *otx1*, *wwtr1*, *rras2*, *p300* and *parn* mRNAs and show they have novel and important roles in PGC development.

Transcripts for functional investigation were selected based on the following criteria: 1) vegetal enrichment in comparison with the animal (germ plasm free) pole sample, 2) previously known function in literature, and 3) how the transcript fit into the GeneGo proposed gene network of 47 vegetally enriched transcripts (Fig. 2.2C). All chosen transcripts are in our list of 198 vegetally enriched transcripts in the oocyte and are known to be involved in tissue boundary formation, cell cycle, and/or proliferation. To help validate the generated Genego network found within our data set, we chose *e2f1* (transcription factor), *wwtr1* (transcription co-factor) and *p300* (transcription co-factor) for investigation. Finally, we focused on the investigation of *efnb1* and its role in PGC
development of the early embryo. We propose that the chosen vegetally enriched maternal transcripts are required for normal PGC development.

**BACKGROUND**

In *Xenopus laevis*, germ plasm is the required, maternally inherited substance for proper primordial germ cell development. To understand the components of germ plasm and how they interact to protect and ensure the proper development and totipotent potential of the future PGCs of an embryo is of great interest. Initial functional studies that have been guided by microarray data have been the first steps to understanding the tightly regulated protein and mRNA activity in the germ plasm. Houston et al. identified over 200 RNAs to be vegetally localized that include both germline and somatic cell lineages (2010). The increased sensitivity and potential for discovery that is provided by RNA-seq increases our ability to find new localized transcripts in the oocyte. In our previously discussed RNA-seq analysis we identified vegetally localized RNAs that guided the following investigations to determine their possible roles in PGC development.

Genes involved in cell cycle regulation and maintenance of pluripotency have been found to in the germ plasm. Germ plasm is asymmetrically inherited by daughter cells, termed as presumptive PGCs (pPGCs), during the early cleavage stages of *Xenopus laevis* development. At gastrula, germ plasm becomes perinuclear and can be equally distributed in resulting daughter cells, now termed PGCs. Gastrulation is also when the PGCs become zygotically active, where previously they were transcriptionally quiescent (Lai et al., 2012; Venkatarama et al., 2010). Why is the proliferation and cell cycle of pPGCs and PGCs tightly controlled in development? To better understand such stringent
regulation of the cell cycle in germ plasm containing cells, our lab chose the maternal, vegetally enriched transcripts *e2f1*, *rras2*, *parn*, *otx1*, and *wwtr1* for functional study, as previous studies show each to have a role in cell cycle regulation.

The migration activity of PGCs is critical for germline development in many metazoans. *efnb1* mRNA was found in our list of vegetally enriched transcripts and was first discovered to be a germ plasm component by Betley et al. (2002). The Efnb1 ligand is a type I membrane protein and a ligand of Eph-related receptor tyrosine kinases. Ephrin receptors and ephrin ligands have known roles in axonal guidance and migration. These biological functions of *efnb1* helped our lab choose this gene for further functional analysis. Here, we perform Gain-of function and Loss-of function analysis on *efnb1* to determine if this gene affects PGC migration as it does in neuronal cells.

RESULTS

3.1 Overexpression of *e2f1*, *otx1*, *parn*, *rras2* and *wwrt1* significantly reduces PGC number

As a first step towards functional analysis, six vegetally enriched transcripts expressed in PGCs were selected for overexpression studies. One-cell embryos were injected in the vegetal region with in vitro synthesized mRNA of selected transcripts or GFP as control. Tailbud embryos were collected and the number of PGCs per embryo was calculated and compared with GFP-injected controls. Overexpression of the transcription factors *e2f1* and *otx1*, the transcriptional co-activator *wwtr1*, a Ras-like small GTPase *rras2*, and the poly(A)-specific ribonuclease *parn* significantly reduced PGC number, whereas *spire1* had no effect (Fig. 3.1). Importantly, aside from the effects on PGC number, embryos
Fig. 3.1. Overexpression of five out of six selected vegetally enriched mRNAs reduces PGC number. One-cell embryos were injected in the vegetal region with GFP (control) or the indicated transcripts (0.5 ng). Tailbud embryos (stage 32-35) were analyzed for xpat expression by WISH, and representative images are shown (A-G). The number of PGCs per embryo was quantified (H). GFP, n=16; spire1, n=39; e2f1, n=64; otx1, n=27; parn, n=35; rras2, n=48; wwtr1, n=35. *P<0.05, compared with GFP control. Analysis based on at least two independent experiments and shown as a box and whisker plot.

appeared normal. Taken together, these results suggest a specific role in PGC development for otx1, e2f1, wwtr1, rras2 and parn.

3.2 Embryos depleted of otx1 and wwtr1 show a significant increase in PGC number

To further test the function of otx1 and wwtr1, we created loss-of-function morphants by injection of antisense morpholinos (MOs) into one-cell embryos. Both otx1-MO and wwtr1-MO blocked the translation of their respective proteins in a dose-dependent fashion (Fig. 3.2). Injected embryos were collected at tailbud stages and the number of
Figure 3.2. Morpholinos (MOs) inhibit expression of their target RNAs. A) Schematic of the \textit{efnb1}, \textit{otx1}, and \textit{wwtr1} morpholino targeted regions are indicated in red. B) Wheat germ extracts were incubated with either \textit{efnb1}-FL (1ug), \textit{otx1}-FL (250ng), or \textit{wwtr1}-FL (250ng) transcripts in the presence of increasing concentrations of their respective MOs and subject to anti-flag western blot analysis. Quantification of respective flag expression is shown.

Fig. 3.3. MO-mediated knockdown of a subset of vegetally enriched mRNAs increases PGC number. One-cell embryos were injected in the vegetal region with MOs (15 ng) targeting \textit{otx1} or \textit{wwtr1}. Tailbud embryos (stage 32-35) were analyzed for \textit{xpat}. Representative images are shown (A-C). The number of PGCs per embryo was quantified (D). Uninjected control (ctrl), n=32; \textit{otx1}-MO, n=31; \textit{wwtr1}-MO, n=25. *P<0.05 compared with control. Analysis based on at least two independent experiments.
3.3 p300 is required for normal PGC development

The transcriptional co-activator and histone acetyltransferase p300 has been shown to be expressed during oogenesis (Kwok et al., 2006) and to regulate the metabolic state of mammalian germ cells (Boussouar et al., 2014). p300 is enriched at the vegetal pole of stage VI oocytes (Fig. 2.2D) and represents one of the network hubs at the vegetal pole (Fig. 2.2C). We therefore examined whether p300 plays a role in Xenopus PGC development. Fertilized embryos were incubated with DMSO (control) or the p300 small molecule inhibitor C646 until they reached tailbud stages. Treated embryos were then collected, and the number of PGCs per embryo was calculated and compared with controls. Inhibition of p300 significantly reduced PGC number in a dose-dependent manner, suggesting an indirect role in PGC development (Fig. 3.4).
3.4 *efnb1* plays an essential role in PGC specification and migration

Eph receptor tyrosine kinases and their ligands, ephrins, have been shown to be involved in the formation of tissue boundaries, including separation of the germ layers, by regulating migration, adhesion and repulsion during embryonic development (Rohani et al., 2014). However, their specific role(s) in germline development is unknown. Interestingly, *efnb1* expression is upregulated in vegetal versus animal poles of stage VI oocytes. *efnb1* is also expressed in PGCs at neurula (Fig. 2.3A), suggesting a role in this lineage. To assess the role of *efnb1* in PGC development, *efnb1* was overexpressed by injecting flag-tagged *efnb1* (*efnb1-FL*) mRNA into the vegetal region of fertilized...
embryos. No notable changes in morphology or PGC location was observed (Fig. 3.5A). However, overexpression of *efnb1* significantly reduced the number of PGCs compared with the control (Fig. 3.5A). This effect was rescued by co-injection with *efnb1-MO* (Fig. 3.5A). These data suggest that PGC number is not maintained within an environment of excess Efnb1 protein.

We next assessed the effect of *efnb1* inhibition using an *efnb1-MO* as previously described (Moore et al., 2004). No notable changes in morphology or PGC number was observed in embryos injected with *efnb1-MO* compared with scrambled-MO or uninjected controls (Fig. 3.5B). *efnb1* inhibition significantly increased the total number of embryos containing mislocalized PGCs according to the boundaries designated for normal PGC localization, between somites 5 and 11, as described by Tarbashevich et al. (2011). Mislocalization was along the anterior/posterior (A/P) axis, primarily beyond the normal posterior boundary (Fig. 3.5B). PGC mislocalization was rescued by co-expression with an *efnb1* mRNA construct containing conservative mutations in the MO-binding region (*efnb1-FL-rescue*), rendering the MO ineffective (data not shown), confirming the specificity of the effect of *efnb1-MO* (Fig. 3.5B). These data suggest that *efnb1* was essential for the proper migration of PGCs.
Fig. 3.5. *efnb1* is required for normal PGC number and migration. One-cell embryos were injected in the vegetal region with (A) *efnb1*-FL alone (200 pg), or *efnb1*-FL (200 pg) and *efnb1*-MO (16 ng); (B) scrambled-MO (16 ng), *efnb1*-MO alone (16 ng), or *efnb1*-MO (16 ng) and *efnb1*-FL-rescue (200 pg). Tailbud embryos were analyzed for xpat expression by WISH. Representative images are shown on the left. (A) The number of PGCs per embryo was quantified. (B) Percentage of embryos with mislocalized PGCs was calculated. Black lines indicate the boundaries of normal PGC location, between somites 5 and 11. Uninjected control (Ctrl), n=28; *efnb1*-FL, n=28; *efnb1*-FL and *efnb1*-MO, n=28; scrambled-MO, n=25; *efnb1*-MO, n=27; *efnb1*-MO and *efnb1*-FL-rescue, n=28. *P<0.05 compared with uninjected control. #P<0.05 compared with *efnb1*-FL (A) or *efnb1*-MO (B). Analysis based on at least two independent experiments.
DISCUSSION

This study shows the first functional analysis’ guided by a predicted gene network within our top 198 vegetally enriched transcripts in the *Xenopus laevis* oocyte. Utilization of the GeneGo generated gene network of 47 genes (Fig. 2.2C) guided our studies to investigate the transcription factors *otx1* and *e2f1* and the co-transcription factors *wwtr1* and *p300*. Additionally, we investigated *parn*, *rras2*, and *efnb1* based on vegetal enrichment and functional properties reported in somatic tissues. From our observations, we report that *p300*, *otx1*, *e2f1*, *wwtr1*, *rras2*, *parn*, and *efnb1* are germ plasm components that have novel functions and are required for normal PGC development.

*e2f1* constitutes the largest hub by far, connecting 19 other localized mRNAs, including two additional hubs, *err1* and *p300*, and genes involved in cell cycle regulation, DNA replication, pluripotency/ differentiation, and metabolism (Fig. 2.2C). Overexpression of *e2f1* resulted in a loss of PGCs (Fig. 3.1). This effect might be due to misregulation of the cell cycle and/or in initiating somatic differentiation. Consistent with this hypothesis, Zaragoza et al. (2010) showed that the balance of E2f transcription factors, E2f dimerization partners and C/EBPα is critical for proper cell cycle progression. E2f1 has also been shown to mediate proliferation through Wnt signaling by direct interaction with the Fzd1 promoter (Yu et al., 2013). Furthermore, E2f1 activates Pcsk6, which allows for mesoderm induction by activation of VegT (Heasman, 2006). E2f1 also activates Cbx7, a member of the Polycomb repressor PRC1-like complex that plays a pivotal role in the transition from pluripotency to differentiation by regulating Cbx8 and Fzd1 (Klauke et al., 2013; Creppe et al., 2014; Mani et al., 2008; O’Loghlen et al., 2015). Thus, E2f1 in excess may tip the balance towards somatic differentiation in
PGCs, causing their loss. Because of the dominant position that e2f1 holds, we attempted to knockdown its activity by antisense MO injection into fertilized eggs (data not shown). Unfortunately, we could not detect a phenotype, most likely because of a pre-existing maternal supply of E2f1 protein (Peshkin et al., 2015). Future oocyte host transfer studies will investigate the function of E2f1 by depleting the maternal supply.

Interestingly, overexpression of rras2, parn, otx1 and wwtr1 also resulted in PGC loss, while MO-mediated inhibition of otx1 and wwtr1 caused an increase in PGC number (Fig. 3.1, Fig. 3.3). Previous studies involving these transcripts all mention their possible roles in cell cycle regulation, proliferation and/or survival, consistent with our observations. The deadenylase Parn mediates progression through G0/G1 by regulating p53 and p21 expression (Zhang and Yan, 2015). Rras2 has been implicated in cell proliferation by regulating the PI3K (Murphy et al., 2002) and ERK pathways (Larive et al., 2012). Similarly, Otx1 regulates proliferation through the ERK/MAPK pathway, and is necessary for progression through S phase (Li et al., 2016). Wwtr1(Taz) is involved in cell cycle progression, proliferation and survival by regulating cyclin A and Ctgf expression and Casp3 activity (Wang et al., 2014). Taken together, our working hypothesis is that overexpression results in cell cycle checkpoint abnormalities and cell death, whereas loss of function releases the restrained PGC cell cycle clock resulting in more PGCs. Further investigation of these known downstream targets of rras2, parn, otx1 and wwtr1 in PGCs is necessary to deduce the exact mechanisms by which these genes regulate PGC number.
**Inhibition of p300 results in a loss of PGCs**

Here we show for the first time that p300 is necessary for normal PGC development (Fig. 3.4). Pharmaceutical inhibition of p300 caused a significant reduction in PGC number, suggesting a role in proliferation, apoptosis and/or cell cycle regulation. Similar to what has been shown in retinal cells, p300 may promote PGC proliferation and protect PGCs from apoptosis by modulating the activity of Stat1 and Stat3 (Kawase et al., 2016). Alternatively, p300 might be necessary to allow PGCs to pass through G1 of the cell cycle in order to proliferate, consistent with its role in leukemia cells (Gao et al., 2013). Further investigation is necessary to determine the precise molecular mechanisms by which p300 regulates PGC number.

**Efnb1 is required for normal PGC development and migration**

Our functional studies reveal a novel role for *efnb1* in PGC maintenance and migration (Fig. 3.5). *efnb1* is expressed in BBs, suggesting that it is an early component of germ plasm. After MBT, PGC-specific *efnb1* expression is lost but it is re-expressed at neurula (Fig. 2.3A). These observations suggested that *efnb1* might function in both the endoderm and germline lineages. MO-mediated knockdown of *efnb1* did not affect PGC number, but caused PGCs to migrate outside their normal boundaries, primarily into posterior endoderm. By contrast, *efnb1* overexpression decreased the number of PGCs but did not affect migration and development. These results suggest that *efnb1* is also involved in signaling pathways necessary for normal PGC development. Both the mismigration and loss of PGCs could be rescued, indicating specificity of the observed phenotypes (Fig. 3.5).
Ephrin ligands and Eph receptors are known to contribute to the maintenance of vertebrate tissue boundaries (Rohani et al., 2014), to regulate axon migration (Klein and Kania, 2014) and establish A/P gradients required for proper cell migration (Bush and Soriano, 2010). Therefore, \textit{efnb1} knockdown-mediated PGC mismigration may be due to disruption of the A/P axis. Consistent with this interpretation, PGCs were not ectopically found outside of the endoderm nor was PGC number affected. Interestingly, Enfb1 associates directly with Dishevelled and is capable of recruiting it and Grip2 to the plasma membrane (Brückner et al., 1999; Moore et al., 2004; Lee et al., 2006). Grip2 encodes a scaffolding protein known to interact with receptors including Frizzled1 (Korkut et al., 2009; Ataman et al., 2006). Knockdown of Grip2 in the embryo disturbs normal PGC migration (Kirilenko et al., 2008; Tarbashevich et al., 2007), similar to our results with \textit{efnb1} (Fig. 3.5). Taken together, these observations suggest a close physical association between Grip2, Efnb1 and the Wnt signaling components that facilitates correct PGC migration.

Unlike somatic cells, PGCs are known to divide symmetrically only two or three times before exiting the endoderm (Whittington and Dixon, 1975). Disruption of the cell cycle or inappropriate gene expression in PGCs would trigger their cell death (Lai et al., 2012). \textit{efnb1} interacts with at least two signaling pathways: FGF (Moore et al., 2004; Lee et al., 2006) and Wnt (Lien and Fuchs, 2014; Lee et al., 2006). Disruption of these pathways could affect cell proliferation or differentiation, which might explain the effect on PGC number when \textit{efnb1} is overexpressed.
NORMAL PGC DEVELOPMENT REQUIRES STRICT REGULATION OF *sox7*

OVERVIEW

Xenopus primordial germ cells (PGCs) are determined by the presence of maternally derived germ plasm. Germ plasm components both protect PGCs from somatic differentiation and enable a unique, specifying gene expression program to occur in the PGC cell population. Segregation of the germline from the endodermal lineage occurs during gastrulation and PGCs subsequently initiate zygotic transcription. However, the gene-networks that operate to both preserve the potential for totipotency and promote germline differentiation are poorly understood. In the present study, we utilize the RNA-seq analysis from Butler et al. to determine that 83% of maternal, vegetally-enriched transcripts are zygotically re-expressed, and of those over 50% were enriched in PGCs (in Revision, 2017). Initial expression studies using WISH revealed that PGC-enriched transcripts all show zygotic expression in the PGCs of the neurula embryo. It was also found that similar to humans, an F-sox family member (*sox7* in Xenopus and SOX17 in humans) is required for proper PGC development.

BACKGROUND

The germline is the only cell population that has the responsibility of maintaining heritable genetic information in metazoan organisms. In Xenopus, specification of this
highly specialized cell type is established by the inheritance of maternally derived germ plasm. Germ plasm is localized, along with somatic determinants to the vegetal pole during oogenesis (Forristall et al., 1995; Heasman et al., 1984; Kloc and Etkin, 1995; Zhang et al., 1998). After fertilization, germ plasm is partitioned and inherited by the vegetal most blastomeres. Further divisions of these blastomeres results in asymmetric inheritance of germ plasm by one daughter cell, termed as the presumptive PGC. During gastrulation, segregation of the germline occurs when germ plasm becomes perinuclear and subsequent divisions result in symmetrical inheritance of germ plasm by each daughter cell, now termed as PGCs. At this point, the zygotic PGC transcription program is now active, initiated by unknown maternal factors. Because of the inherent importance of further understanding the regulatory mechanisms responsible for germline specification, it is of interest to elucidate what maternal factor(s) are/may be responsible for initiating the zygotic PGC program.

In the previous two chapters, I discuss my RNA-seq analysis of vegetal/animal pole samples to identify maternally inherited transcripts with novel roles in PGC specification. In summary, we defined 198 annotated RNAs highly enriched at the vegetal pole including several known germline and somatic determinants. Further analysis confirmed known germline components such as: xpat, sybu, and otx1, and identified novel components including the transcription factor (TF) sox7. This work guided my lab to pursue an unprecedented RNA-seq investigation of zygotically active PGCs and surrounding endoderm cells.

Butler et al., utilized RNA-seq analysis to determine the zygotic PGC transcriptome in *Xenopus laevis* by comprehensive interrogation of PGC and neighboring
endoderm cell RNAs just after lineage segregation (in Revision, 2017). This investigation identified 1,865 transcripts enriched in PGCs, containing over half of the 198 annotated vegetally enriched transcripts, including sox7. The top 150 most highly expressed PGC-enriched transcripts were submitted to GeneGo to generate a gene regulatory network of 53 genes, all with direct interactions (Figure 4.1; Butler et al., in revision 2017). sox7 was one of the contained transcription factors of this zygotically related gene network, as it was in the maternally derived network. These findings, taken together with my previous results, directed my thesis work to focus on the role of sox7 in PGC specification.

Here, we utilized RNA-seq analysis of zygotically active PGCs to investigate new transcripts in addition to validating/verifying maternal transcript expression profiles of vegetally enriched germ plasm containing samples. Whole mount in situ hybridization was also used to validate zygotically enriched transcripts for expression. We also aimed to identify the function of sox7 in Xenopus PGC development. To elucidate the specific role of sox7 in PGCs, we first directed sox7 knockdown and over expression constructs to the germline by way of injecting single-cell embryos in the vegetal, germ plasm containing pole of the oocyte. These injections targeted the germ plasm, but would also be incorporated into surrounding endodermal tissue. To see a more localized and specific effect to the germline, we also performed PGC-directed injections for sox7 knockdown and overexpression into the germ plasm containing blastomeres of 16- to 32-cell embryos (Fig. 4.2; Butler et al., in Revision 2017). PGC-directed injections restricted effects to only fate mapped regions for PGC and closely neighboring endoderm. Our results indicate that, prior to neurula, sox7 is necessary for correct PGC number formation. These data provide further evidence that sox7 is a critical TF necessary for PGC
Figure 4.1. Analysis of PGC enriched transcripts. GeneGo direct interaction network connecting 53/150 most highly expressed PGC-enriched transcripts. (Butler et al., in Revision 2017)

Figure 4.2. PGC-directed Injections. A) Live 16 – 32 - cell embryo, vegetal pole up. Germ plasm is detected as dark regions in vegetal blastomeres (*). Typical injection sites are shown. B&C) Lineage tracer (red, outlined in black) showing gastrula (st. 11.5, B) and tailbud (st. 33-34, C) progeny injected as in A). xpat WISH (purple, arrows) indicates PGCs. Note: quality of targeting of PGCs. (Butler et al., in Revision 2017)

development, but the initiating factor(s) that turns on the zygotic transcriptome still remains unknown.
RESULTS

4.1 Maternal, vegetally-enriched transcripts are re-expressed in PGCs
To identify zygotic expression of maternal transcripts within the PGC population, we compared the highly expressed, PGC-enriched transcripts identified by Butler et al. with our vegetally enriched gene list (in Revision, 2017). Of the 198 maternal, vegetally-enriched transcripts, ~83% (165/198) were zygotically re-expressed in either endoderm cells, PGCs, or both (GSE102047; Butler et al., in Revision 2017). Only ~2% (4/198) of vegetally-enriched RNAs were re-expressed and enriched in endoderm cells (blue), while ~30% (57/198) were re-expressed in both PGCs and endoderm cells, but not differentially expressed (white) (GSE102047; Butler et al., in Revision 2017). However, over half of the vegetally-enriched transcripts (104/198) were re-expressed and enriched in PGCs (red) (GSE102047; Butler et al., in Revision 2017), which suggests that most inherited, vegetally-enriched RNAs are involved in germline specification. Whether they are essential will require further functional testing. Furthermore, although most PGC-enriched transcripts are zygotic (1,761 / 1,865), half of the 50 most highly expressed PGC-enriched transcripts are maternal (Table 4.1, gray).

4.2 Expression of PGC enriched mRNAs during development
We selected a set of 15 highly expressed, PGC-enriched transcripts to test, by WISH analysis, for expression at gastrula (st. 11.5), neurula (st. 16), and tailbud (st. 33/34) stages. xpat (pgat), was used as a reference for PGC expression at each stage. Consistent with our RNA-seq data, all of the 15 transcripts tested (fer, pphln1, prpsap2, cdc20, fam168b, lmb3, nasp, rtn3, pgam1, tspan1, ppp1r2, xpat, hfoo1, zfyve26, and impad1),
<table>
<thead>
<tr>
<th>ID</th>
<th>Gene symbol</th>
<th>PGC average (CPM)</th>
<th>Endo average (CPM)</th>
<th>FC (PGC vs Endo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xeleav15011470m.g</td>
<td>pgat</td>
<td>89775</td>
<td>564</td>
<td>159</td>
</tr>
<tr>
<td>Xeleav18035291m</td>
<td>ddx25.L</td>
<td>71752</td>
<td>385</td>
<td>186</td>
</tr>
<tr>
<td>Xeleav18023956m</td>
<td>gripL*</td>
<td>38292</td>
<td>130</td>
<td>295</td>
</tr>
<tr>
<td>Xeleav18031029m</td>
<td>dazL</td>
<td>16463</td>
<td>82</td>
<td>202</td>
</tr>
<tr>
<td>Xeleav18034293m</td>
<td>nanos1.L</td>
<td>6457</td>
<td>30</td>
<td>216</td>
</tr>
<tr>
<td>Xeleav18006177m</td>
<td>tex15.L*</td>
<td>5499</td>
<td>53</td>
<td>104</td>
</tr>
<tr>
<td>Xeleav18031232m</td>
<td>trankL</td>
<td>3728</td>
<td>80</td>
<td>46</td>
</tr>
<tr>
<td>Xeleav15043509m.g</td>
<td>trim36</td>
<td>3005</td>
<td>67</td>
<td>45</td>
</tr>
<tr>
<td>Xeleav18017827m</td>
<td>mcm10.L</td>
<td>2998</td>
<td>810</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18047493m</td>
<td>naga.S</td>
<td>2111</td>
<td>642</td>
<td>3</td>
</tr>
<tr>
<td>Xeleav15038666m.g</td>
<td>acsbq2</td>
<td>2088</td>
<td>379</td>
<td>6</td>
</tr>
<tr>
<td>Xeleav18025369m</td>
<td>ptbp2.S</td>
<td>1766</td>
<td>456</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18028249m</td>
<td>kif13b.L*</td>
<td>1648</td>
<td>189</td>
<td>9</td>
</tr>
<tr>
<td>Xeleav18010315m</td>
<td>slc12a3.2.S*</td>
<td>1565</td>
<td>341</td>
<td>5</td>
</tr>
<tr>
<td>Xeleav1504020m.g</td>
<td>acsI</td>
<td>1467</td>
<td>294</td>
<td>5</td>
</tr>
<tr>
<td>Xeleav15016832m</td>
<td>lidrap1-a</td>
<td>1341</td>
<td>107</td>
<td>13</td>
</tr>
<tr>
<td>Xeleav18034461m</td>
<td>bicc1.L*</td>
<td>1169</td>
<td>93</td>
<td>13</td>
</tr>
<tr>
<td>Xeleav18017004m</td>
<td>dnd1.L*</td>
<td>1051</td>
<td>19</td>
<td>54</td>
</tr>
<tr>
<td>Xeleav15069945m.g</td>
<td>xilin2</td>
<td>1015</td>
<td>47</td>
<td>22</td>
</tr>
<tr>
<td>Xeleav18018420m</td>
<td>pksk6.L*</td>
<td>971</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td>Xeleav18046219m</td>
<td>spire1.S</td>
<td>930</td>
<td>11</td>
<td>81</td>
</tr>
<tr>
<td>Xeleav18012344m</td>
<td>mov10.L</td>
<td>890</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td>Xeleav18032097m</td>
<td>suff1.L</td>
<td>861</td>
<td>155</td>
<td>6</td>
</tr>
<tr>
<td>Xeleav18011128m</td>
<td>atp5a1.S</td>
<td>789</td>
<td>122</td>
<td>6</td>
</tr>
<tr>
<td>Xeleav18033736m</td>
<td>impad1.S</td>
<td>750</td>
<td>16</td>
<td>47</td>
</tr>
<tr>
<td>Xeleav18041802m</td>
<td>hspa3.S</td>
<td>690</td>
<td>193</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18033676m</td>
<td>dsc3.S*</td>
<td>649</td>
<td>161</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18031248m</td>
<td>velo1.L</td>
<td>639</td>
<td>36</td>
<td>18</td>
</tr>
<tr>
<td>Xeleav18040158m</td>
<td>zfyve26.L</td>
<td>633</td>
<td>78</td>
<td>8</td>
</tr>
<tr>
<td>Xeleav18039847m</td>
<td>fbx34.L*</td>
<td>626</td>
<td>148</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav15002985m.g</td>
<td>alk4h</td>
<td>580</td>
<td>145</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18023131m</td>
<td>ccb12.2.L*</td>
<td>562</td>
<td>148</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18070621m.g</td>
<td>лимн3</td>
<td>536</td>
<td>99</td>
<td>5</td>
</tr>
<tr>
<td>Xeleav18051996m.g</td>
<td>vegtb*</td>
<td>536</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>Xeleav18006239m</td>
<td>slc30a9.L</td>
<td>532</td>
<td>126</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18045154m.g</td>
<td>parn.L*</td>
<td>513</td>
<td>41</td>
<td>13</td>
</tr>
<tr>
<td>Xeleav15018513m.g</td>
<td>unnamed</td>
<td>495</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Xeleav18034686m</td>
<td>pgam1.L*</td>
<td>476</td>
<td>11</td>
<td>45</td>
</tr>
<tr>
<td>Xeleav15017463m.g</td>
<td>fam168b</td>
<td>462</td>
<td>94</td>
<td>5</td>
</tr>
<tr>
<td>Xeleav18025228m</td>
<td>tsp1a1.S*</td>
<td>438</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Xeleav18023643m</td>
<td>tob2.L</td>
<td>427</td>
<td>121</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav18029909m</td>
<td>pfkb3.S</td>
<td>425</td>
<td>94</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav15030858m.g</td>
<td>slc16a3-b</td>
<td>423</td>
<td>33</td>
<td>13</td>
</tr>
<tr>
<td>Xeleav1806022m</td>
<td>for.L</td>
<td>415</td>
<td>70</td>
<td>6</td>
</tr>
<tr>
<td>Xeleav18020736m</td>
<td>cpebl1.S*</td>
<td>368</td>
<td>7</td>
<td>52</td>
</tr>
<tr>
<td>Xeleav18019487m.g</td>
<td>nasp</td>
<td>353</td>
<td>36</td>
<td>10</td>
</tr>
<tr>
<td>Xeleav18003587m</td>
<td>cdt1.1</td>
<td>351</td>
<td>71</td>
<td>5</td>
</tr>
<tr>
<td>Xeleav18031734m.g</td>
<td>gmnll</td>
<td>343</td>
<td>102</td>
<td>3</td>
</tr>
<tr>
<td>Xeleav15049105m.g</td>
<td>cab39</td>
<td>339</td>
<td>96</td>
<td>4</td>
</tr>
<tr>
<td>Xeleav15061337m.g</td>
<td>cdc20</td>
<td>324</td>
<td>58</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 4. Maternal versus zygotic PGC-enriched transcripts. Maternal, vegetally enriched transcripts re-expressed in PGCs (gray), and new zygotic transcripts (white). FC = fold change; PGC = primordial germ cell sample; Endo = endoderm cell sample; CPM = counts per million. *most highly expressed copy is shown.
Figure 4.3. WISH analysis of PGC transcripts. Expression of selected PGC-enriched mRNAs at post-MBT stages 11.5 (gastrula), 16 (neurula), and 33/34 (tailbud). xpat expression marks germ plasm in PGCs. vegetal view (Vg), dorsal (D), ventral (V), left (L), right (R), anterior (A), posterior (P), germ plasm/PGCs (gp), pronephros (pn), ventral blood islands (vbi), otic vesicle (ov), cranial ganglia (cg), brachial arches (b), nasal placodes (np), intersegmental region (isr), somites (s), notochord (n), brain (b), neural tube (nt). Transcripts detected in PGCs at neurula are shown at higher magnification in insets. WISH analysis was performed on ≥21 total embryos from at least three adult female frogs. Scale bars: 200um

were expressed in PGCs at neurula (12/15 are shown in Fig. 4.3). Interestingly, all but 4 transcripts (pphl1, cdc20, hfoo1, and impad1) were also expressed at stage 11.5, but of those, only 4 transcripts (fam168b, pgam1, ppp1r2, and zfyve26) are vegetally-enriched, maternal transcripts. These data suggest that PGCs initiate zygotic transcription by stage 11.5. Furthermore, only xpat was still expressed in PGCs at tailbud stages. All other PGC-enriched transcripts tested were expressed only in the soma by tailbud stages. The
most represented expression pattern at tailbud was in neural regions, including the eye, cranial ganglia, brain, and neural tube (Fig. 4.3).

4.3 Over expression and knockdown of sox7 in single-cell embryos reduces PGC number

The transcription factor sox7 has been shown to play various roles in embryonic development, including proliferation, differentiation, hematopoiesis, cardiogenesis and vasculogenesis (Stovall et al., 2014). However, its role in PGCs is unknown. sox7 expression is significantly upregulated in the vegetal compared with the animal pole in stage VI oocytes (Fig. 2.1C, Table 2.1) and its expression persisted in PGCs during gastrulation (Fig. 2.3B). To assess the role of sox7 in PGC development, fertilized embryos were injected vegetally with either a sox7-targeted MO (Fig. 4.4) or the dominant-negative transcript sox7dCEnR (Fig. 4.5). Overexpression and rescue experiments were performed using X. tropicalis sox7 (Xtsox7) mRNA (Zhang et al., 2005a). Initial injections of sox7dCEnR or Xtsox7 mRNAs were performed at various concentrations to determine an effective dose that would not cause the phenotypic alterations observed by Zhang et al. (2005a) (data not shown). No notable changes in morphology were observed in embryos injected with sox7-MO, sox7dCEnR (200 pg) and/or Xtsox7 (200 pg) mRNA (Fig. 4.4C, D, Fig. 4.5A-D). The number of PGCs per embryo was calculated and compared with uninjected controls. Both dominant-negative and MO-mediated inhibition and the overexpression of Xtsox7 significantly reduced the
Figure 4.4. Morpholino-mediated sox7 inhibition reduces PGC number in tailbud embryos. A) Schematic of the sox7 morpholino target region (top). Wheat germ extract was incubated with sox7-FL (500ng) in the presence of increasing concentrations of sox7-MO and subject to anti-flag western blot analysis. Quantification of respective flag expression is shown (bottom). B-D) One cell embryos were injected in the vegetal region with sox7-MO (16ng). Tailbud embryos were analyzed for xpat expression by WISH. The number of PGCs per embryo was quantified (B). Representative images are shown (C-D). Uninjected control (ctrl) n=36, sox7-MO n=43. *statistically significant compared to ctrl (p<0.05). Analysis based on at least two independent experiments.

Fig. 4.5. Altering sox7 expression reduces PGC number. One-cell embryos were injected in the vegetal region with sox7dCEnR RNA (200 pg) and Xtsox7 RNA (200 pg), alone and in combination. Tailbud embryos were analyzed for xpat expression by WISH. Representative images are shown (A-D). The number of PGCs per embryo was quantified (E). Uninjected control (ctrl), n=28; sox7dCEnR, n=24; Xtsox7, n=27; sox7dCEnR and Xtsox7, n=28. *P<0.05 compared with uninjected control. #P<0.05 compared with sox7dCEnR. Analysis based on at least two independent experiments.
number of PGCs in tailbud embryos (Fig. 4.4, Fig. 4.5). Expression of both sox7dCEnR and Xtsox7 mRNAs together significantly rescued the effect that sox7dCEnR had on PGC number, presumably because Xtsox7 restored function (Fig. 4.5D,E). These data suggest that the level of sox7 expression must be tightly regulated for proper PGC development.

4.4 PGC-directed sox7 knockdown or over-expression increases PGC number in early development

Historically, mice have been used as the primary model system to study human PGC development (Hayashi et al., 2007; Ohinata et al., 2009; Ohinata et al., 2005; Saitou et al., 2002; Saitou and Yamaji, 2012). Mouse studies have revealed specific gene sets necessary to specify the germ cell lineage, including the critical pluripotency genes: Nanog, Pou5f1 (Oct3/4), and Sox2 (Tang et al., 2016). However, contrary to what has been predicted for human PGC development (based on mouse studies), Irie et al. (2015) identified the F-sox family member, SOX17, as a primary regulator of human primordial germ cell-like fate, not SOX2 (a soxB1 family member). Interestingly, although several sox transcripts were enriched in PGCs (namely sox2, sox3, sox4, sox7, and sox8), sox7, another F-sox family member, is highly enriched (74-fold versus endoderm) and the most highly expressed (269 CPM) sox transcript in Xenopus PGCs (GSE102047; Butler et al, in Revision 2017). Furthermore, sox7 was recently shown to be maternally expressed in germ plasm through gastrulation, and both MO-mediated sox7 inhibition and over-expression of sox7 in the whole embryo results in significantly reduced PGCs at tailbud stages. Therefore, we hypothesized that sox7 is necessary for proper PGC development.
To test our hypothesis, *sox7* MO-mediated inhibition and over-expression of *sox7*-FL mRNA was targeted to PGCs as described in materials and methods (Fig. 4.2). Interestingly, PGC-directed *sox7* inhibition and over-expression resulted in significantly more PGCs during gastrulation (st. 11.5) (Fig. 4.6A&B). These effects were significantly rescued by co-(PGC-directed) injecting a *sox7* mRNA construct that cannot bind the *sox7*-MO (*sox7*-MO rescue) (Fig. 4.6A&B), thereby confirming specificity of the observed phenotype. However, by tailbud stages (st. 33/34), normal PGC numbers were detected in PGC-directed *sox7* knockdown embryos while PGC-directed *sox7* over-expressed embryos had significantly fewer PGCs (Fig. 4.6C&D). The observed increase in PGC number at gastrulation was both remarkable and unexpected as the amount of germ plasm assembled during oogenesis remains constant during development. One possible explanation for more germ plasm-bearing cells after *sox7* mis-expression might be the premature migration of germ plasm to a perinuclear position, resulting in two rather than one cell receiving germ plasm.
Figure 4.6. PGC number increased after PGC-directed sox7 knockdown or over-expression in early development. 16-32-cell embryos were injected as described in materials and methods with sox7-MO (16ng total), sox7-FL RNA (150pg total), or sox7-MO (16ng total) and sox7-FL rescue RNA (150pg total). Gastrula (st. 11.5) and tailbud (st. 33-34) embryos were analyzed for xpat expression by WISH. The number of PGCs per embryo was determined (A&C). Representative images are shown (B&D). ctrl n=23, sox7-MO n=24, sox7-FL n=23, sox7- MO+ sox7-FL rescue (sox7-MO rescue) n=25. * statistically significant compared to uninjected control (*=p<0.05; **=p<0.005). # statistically significant compared to sox7-MO (p<0.05). Analysis shown for one of at least three independent experiments. ctrl = control; MO = morpholino; FL = full length, flag-tagged. Scale bars = 500um (B), or 1mm (D).
4.5 WISH analysis reveals MO-mediated knockout of sox7 causes no detected effect on native mRNA expression for select germ plasm enriched transcripts

One persistent question in germ cell biology is: What is the germ plasm component responsible for initiating the zygotic program in PGCs? As has been previously described, it is known that PGCs remain quiescent until the mid-blastula transition (MBT) (Vanketeraman et al., 2010 and Lai and King, 2013). sox7 is a transcription factor found in the maternally inherited germ plasm of the oocyte and is also expressed in the zygotically active PGC after MBT. With these findings, we proposed that sox7 may be required to initiate the PGC zygotic program at MBT.

From the selected 15 highly expressed, PGC-enriched transcripts which WISH had been previously performed on (Fig. 4.3), we selected 13 transcripts (fer, pphln1, prpsap2, cdc20, fam168b, lmnb3, nasp, pgam1, tspan1, ppp1r2, xpat, hfoo1, and impad1) to test expression patterns post PGC-directed knockdown of sox7 in the 16- to 32-cell embryo. Specifically, we were most interested in the expression pattern of the neurula stage embryo (st.16) to see zygotic expression of the germline and soma. Interestingly, PGC-directed sox7 knockdown embryos demonstrated expression patterns as that of uninjected control stage 16 embryos (data not shown; control stage 16 embryos serve as representative image found in Fig. 4.3). The most obvious conclusion that we can obtain from this analysis is Sox7 is not the germ plasm component responsible for turning on the zygotic PGC program involving the selected transcripts.

DISCUSSION

Here, we utilized RNA-sequencing data from Butler et al. to identify that over 50% of maternal, vegetally-enriched transcripts were enriched in the PGC transcriptome,
including *sox7* (in Revision, 2017). Initial studies concerning *sox7* knockdown and overexpression in single-cell or PGC-directed injections show an important role for *sox7* in PGC development.

About 10% of PGC-enriched transcripts are TFs and thus likely involved in PGC-specification. Surprisingly, the maternal, vegetally-enriched T-box transcription factor known to initiate mesoendodermal differentiation, *vegt* (Xanthos et al., 2001; Zhang et al., 1998; Zhang and King, 1996), is among the PGC-enriched TFs (GSE102047). Heasman, et al. (2001) have shown that *vegt* mRNA is necessary to anchor late pathway mRNAs to the vegetal cortex of the fully-grown Xenopus oocyte. In the prior chapters, we identified *sox7* as a late pathway mRNA, and here show that *sox7* is necessary for proper PGC development (Figs. 4.4-4.6). Thus, *vegt* RNA may be necessary to anchor *sox7* to the vegetal pole during oogenesis. Additionally, Zhang, et al. (2005) revealed that VegT directly regulates *sox7* expression, which induces the nodal-related genes (*xnr1*, *xnr2*, *xnr4*, *xnr5*, and *xnr6*) and the pan-endodermal marker *endodermin*, to specify the endoderm. In addition to its role in endoderm specification, perhaps *vegt* RNA is enriched in PGCs to induce re-expression of *sox7*, which is necessary for proper PGC development. Future studies will test these hypotheses.

Additionally, both overexpression and knockdown of *sox7* in PGC-directed experiments caused a significant increase in PGC number for the gastrula stage embryo. However, knockdown of *sox7* in the tailbud stage embryo resulted in normal PGC numbers, while overexpression caused a significant reduction in PGC numbers. Alternately, in the single-cell injection experiments for both knocking down and overexpressing *sox7*, we see a significant decrease in PGC number in the tailbud embryo.
Further studies will be required to tease apart the mechanisms controlling these differing results. In the following chapter, I will discuss why these effects may have occurred.
Chapter 5

DISCUSSION AND SIGNIFICANCE

In my dissertation, I provide evidence pertaining to some factors that may be responsible for initiating the maternal and zygotic PGC program, freeing this cell population from a translationally controlled quiescent state. What has been uncovered is that germ plasm contains information for specifying PGCs, and the future germline. My thesis work provides the first germ plasm driven studies for investigating maternal and zygotic gene regulatory networks that determine germline specification; that efnb1 and sox7 are required for correct PGC preservation; and that wwtr1, e2f1, otx1, rras2, and parn all have newly found roles in PGC development. Additionally, my work shows that annotated microRNAs are uniformly distributed, but putative mRNA targets are vegetally localized within the oocyte.

5.1 RNA-seq analysis reveal novel gene regulatory networks (GRNs)

Between both studies of maternal and zygotically derived germ plasm samples, we detected GRNs that may function in proper germline specification. First, we submitted the top 198 most expressed RNAs in our vegetally localized, germ plasm samples to GeneGo. After submission, GeneGo generated a predicted GRN consisting of 47 interconnected genes that may be active in PGC specification. Once obtained, we selected genes within the network for investigation that were anticipated to be required for proper germline formation: e2f1, p300, sox7, and wwtr1. In our second RNA-seq analysis, we
similarly generated a GRN with the 150 most highly expressed, PGC-enriched genes, revealing sox7, again, as a possible target. Many more genes from both networks are worthy for investigation, but these are the specific ones my thesis focused on.

GRN generation is continually improving because of 1) new studies revealing novel gene interactions that continue to be published and 2) diligent literature searches that obtain data for building new links of interconnectivity between genes. These improvements are sure to reveal more interactions within both submitted gene lists, identifying more targets for investigations. Additionally, the oocyte study has a plethora of transcript information for identifying future gene networks because of de novo sequencing. De novo sequencing revealed 2265 un-annotated transcripts, with 213 being enriched in the vegetal pole samples. Of the unannotated RNAs, we identified 15 to have unique expression in only one of the pole sample sets; 3 un-annotated transcripts were uniquely found in the animal pole and 12 were found in the vegetal pole sample sets. Within the uniquely vegetal set of RNAs, 4 transcripts have a Fragment-per-Kilobase-of-Transcript-per-Million-mapped-reads (FPKM) values of over 145, with the highest expressing one over 300. This is an exceptional level of expression when one considers that the beautiful in situ data for the germ plasm marker Xpat has FPKM values that are around 315. It is intriguing to think of what the unannotated transcripts might represent. As previously mentioned, we are highly interested in any vegetally localized transcription factors because temporal translational regulation is critically important for PGC specification. What if some of the un-annotated RNAs are these sought-after transcription factors? Functional studies of unannotated transcripts, especially the ones which are vegetally enriched, are sure to provide a better understanding as to the interactive GRN
that is in place to sustain the functionality of the germ plasm. Therefore, it is of interest for future analysis to further analyze these vegetally enriched transcript lists to identify more novel genes that may be required in the regulation of germ plasm.

Another way to increase the applicability of our germ plasm specific GRNs is to align the immortalized RNA-seq data to future *X. laevis* annotations of the genome. One of the largest issues within the annotation sets available to the Xenopus community is that they are not built upon each other, but rather they are individual genome sets (ie. each annotation is its own entity). Therefore, each successive annotation is better in the sense that it contains more annotated RNAs. “Better” does not mean overlapping and inclusive of previous annotated genome sets. This lack of continuity between genome annotations has caused known genes to be missed. It became obvious during our own RNA-seq analysis when our vegetally enriched data set came back without *vg1* expression. Of course, this was of deep concern to us because our data showed it lacked the first found vegetally localized transcript in Xenopus, making us question our data set and the accompanied gene identifications/expression profiles. Thankfully, our bioinformatics resource, Derek Van Booken, took on the challenge of finding *vg1*. Using the *X. tropicalis* reference for *vg1*, Derek was able to find that our vegetal samples contained a 100-fold enrichment of *vg1* when compared to the animal pole. Derek was also able to find the *vg1* transcript in pieces among the v6.0 and v7.1 genome references. For Xenbase annotations to not include this famous, vegetally localized message, only shows that there is much development needed to bring the genome sets to their maximum potential for utility. Specifically, this lack of continuity that has been shown in the genome-builds demonstrates a need for quality control in these data sets.
Within the PGC specific sample set, we generated a GRN of 53 genes with direct interactions from the 150 most highly expressed transcripts. This network revealed 5 transcription factors that have potential for being involved in the activation of the zygotic PGC transcriptome. Of these transcription factors, e2f1 and oct60/25 (Xenopus homologue for Oct3/4), were defined as hubs (genes with 5 or more direct interactions) and a third was sox7, the maternally enriched factor within the vegetal pole samples. This network directed our studies towards further elucidation of the role of sox7 in PGC specification.

Ultimately, the PGC generated GRN compared with the maternal GRN did not lead me to sox7 as the sole initiating factor for the zygotic PGC transcription program. Embryos subjected to morpholino mediated knockdown of sox7 were tested via *in situ* analysis for altered expression of previously selected transcripts (compared to uninjected controls, Fig.4.3). This analysis failed to detect differences in select transcripts (*fer, pphln1, prpsap2, cdc20, fam168b, lmb3, nasp, pgam1, tspn1, ppp1r2, xpat, hfoo1*, and *impad1*), leaving the question open as to the role of sox7 acting alone or together with another factor in initiating zygotic transcription of the selected genes. These results also do not conclusively show that sox7 is not the initiating factor for other zygotically expressed genes. As seen in the PGC specific network, other genes are worth investigating as possible factor(s) responsible for initiating the zygotic PGC transcription program. For example, the Oct3/4 homologues- oct60/oct25 and e2f1. Further investigation is needed to find the ultimate maternal transcription factor(s) that are required for initiating the zygotic activity in the PGC population.
One major consideration that must not be ignored is that both GRN data sets are informative, but not in their complete form. As annotations increasingly become better and interconnectivity is further validated between genes, constructed GRNs will contain more (and possibly better) targets for investigations what is/are the transcription factor(s) responsible for initiating the zygotic program in PGCs.

5.2 MicroRNAs are equally distributed in vegetal and animal pole samples

One aspect of regulation in germ plasm is the accumulation of localized transcripts followed by temporal degradation of non-germline specific transcripts during development. Here, I show evidence that this regulation is not by the localization of miRNAs. Instead, the vegetal localization of mRNAs targeted by miRNA activity may play a role.

In *X. laevis*, most mRNAs are uniformly distributed in the oocyte (Rebagliati et al., 1985). The germ plasm is known to localize to the vegetal pole of the oocyte, where it becomes inherited by the future PGCs of the embryo. Somatic mRNAs can be equally distributed or also vegetally localized within the oocyte. Stringent control mechanisms must be in place within the germ plasm and the cells which inherit it so that PGCs do not acquire a somatic fate but sustain a germline fate. miRNA mediated translation repression may be one of the mechanisms employed to target somatic RNAs within the germ plasm. This possibility fueled our initial investigations to see if miRNAs are vegetally localized within the oocyte.

Small RNA processing of our vegetal/animal pole samples identified 8 miRNAs equally distributed in our vegetal/animal sample sets: 15c, 18a, 19b, 20a, 363, 427, 429,
and 92a. Of these miRNAs, 7 were found to have vegetally localized, predicted mRNA targets. All miRNA analysis utilized *X. tropicalis* homologues of our *X. laevis* transcripts to identify predicted targets. To further specify our targets, we only included ones which share at least one recognition motif between *X. tropicalis* and human. Identification of vegetally enriched targets is one of the first steps in elucidating whether microRNA activity is needed for proper PGC specification.

Suspected targets that must be translationally regulated within germ plasm would most likely be required somatic determination. What better way to translationally regulate these targets than with equally distributed microRNAs that are already in place to act? Somatic differentiation factors were included in the predicted 13 miRNA targets that were vegetally enriched in our data. miRNA targets related to inhibition of proliferation (*sulf1*) and ubiquitination (*rnf38, ube2d1*) were found. Sulf1 decreases proliferative activity in cells by modulating and enhancing the activity of wnt11, another somatic determinant. miRNA mediated degradation of *sulf1* is understandable for inhibition of a somatic fate in the germ cell population and helps to inhibit wnt11 activity in PGCs (Fellgett et al., 2015). Our lab has previously shown that wnt11 translation is repressed in the germline (Lai and King, 2013). It may be possible that *sulf1* is a true miRNA target and is degraded within the germ plasm because it is not needed when wnt11 protein is absent. The ubiquitin related factors *rnf38* and *ube2d1* can regulate p53 ubiquitination, allowing normal cell cycle progression (Scheffner et al., 1994; Sheren and Kassenbrock, 2013). Ubiquitin activity is necessary for clearing of germline determinants in the soma (DeRenzo et al., 2003). A possibility exists that these ubiquitin related factors must be cleared from the germ plasm so that they do not incorrectly ubiquitinate surrounding
germline determinants. Taken together, we suggest that miRNA targeting of *suf1*, *rnf38*, and *ube2d1* could help to maintain the germline and prevent somatic fates.

Vegetally localized factors that may be required for germline development were also identified as predicted miRNA targets. These factors have been linked to cell cycle (*chek1*), germ cell development (*fnc3a* and *tesk2*), neuronal development (*wasl* and *srgap1*) and proliferation (*pgam1*). Counterintuitive of what would be suspected to need miRNA mediated translational suppression, these listed factors could be suspected to be required for proper germline formation. For example, *chek1* is required for the elongation of the cell cycle after MBT occurs in the Xenopus embryo by way of inhibiting the replication factor *drf1* (Collart et al., 2017). Pairing this finding with our knowledge that PGCs divide more slowly than surrounding somatic cells would question the validity of this predicted miRNA target (Whittington and Dixon, 1975). One possibility is that the predicted targets are incorrect because of the analytical processes used to identify these targets. We used the *X. tropicalis* homologs of our identified miRNAs to identify targets. It is possible that the recognition sequence is not contained in the *X. laevis* suspected targets. *X. laevis* and *X. tropicalis* have a large amount of sequence homology within their respective coding domains (~99%), but are not conserved in their un-translated coding regions (UTRs). miRNA binding is usually found in the 3’ UTR (Sidova et al., 2015), causing a discrepancy between our analytical methods for identifying possible miRNA targets in our vegetally localized samples. 3’UTR analysis for identified targets is required to also identify conservations of the binding sequences. The best way to investigate miRNAs in my data set would be to use data bases which are *X. laevis*
specific, ultimately identifying more reliable mRNA targets of found miRNAs within the vegetal/animal sample sets.

5.3 Efnb1 is required for proper PGC migration and formation

Migration events are pivotal for the correct specification of PGCs. Many animals are known to specify the PGC lineage outside of the presumptive gonads. This requires the PGCs to migrate relatively large distances to become correctly incorporated within the embryo to form functional gametes. Therefore, PGC migration is fundamentally important for proper germline development. This understanding has led my thesis work to investigate vegetally enriched maternal transcripts that have known functions in migration activity. Here, I present a novel requirement for efnb1 in normal PGC migration in the tailbud stage embryo.

In Xenopus, PGCs must actively migrate from the endodermal core of the stage 24 tailbud embryo to ultimately reside in the presumptive gonads by stage 52. This migration process has been shown to involve intricate and specific morphology required by PGCs so that they can undergo strict migration patterns within the endoderm to reach a lateral, then dorsal location, arriving in the dorsal mesentery (Terayama et al., 2014). The complete gene regulatory network that is required for PGC migration is not completely understood. To identify possible unknown components of this network, we looked for vegetally enriched transcripts with characterized functions in cellular migration. efnb1 was selected from our list for investigation because of its known activity in neuronal migration.
Efnb1 is a cell surface membrane Ephrin ligand that binds to EphB receptors EphB2, EphB3, and EphB4 to allow proper migration, adhesion, and repulsion in epithelial, vascular and neuronal tissue development (Rohani et al., 2014; Heilbling et al., 2000; Bush and Soriano 2010; and Risley et al., 2009). In addition, *efnb1* has been shown to mediate both FGFR and disheveled signaling responsible for proper cell migration in eye development (Moore et al., 2004; Lee et al., 2006). These known functions of *efnb1* paired with our vegetally enriched data set led us to hypothesize that *efnb1* is required for PGC migration.

To elucidate whether *efnb1* is required for proper migration in the embryo, we performed overexpression and morpholino mediated knock down experiments of this gene in the single-cell embryo. Overexpression of *efnb1* caused a significant reduction in PGC number. Decreases in PGC number may have resulted from improper cell signaling necessary for regulating cell proliferation or differentiation. As previously mentioned, PGCs only divide ~3 times prior to the tailbud stage (Whittington and Dixon, 1975). Bush and Soriano (2010) show that upregulation of *efnb1* can increase cell proliferation by way of the ERK/MAPK signaling cascade. With these results, overexpression of *efnb1* could have increased ERK/MAPK signaling, forcing PGCs to proliferate like the surrounding somatic cells. The increased proliferation would cause a disruption in the normal required quiescent state maintained by PGCs, signaling apoptosis of the cells. Another possible reason as to why *efnb1* overexpression caused a significant decrease in PGC number is related to the canonical Wnt/β-catenin pathway. Canonical signaling by the Wnt/β-catenin pathway requires Disheveled (Dsh) for normal proliferation events to occur. Efnb1 is known to bind Disheveled (Dsh) which could mediate this signaling, causing
decreased proliferation in PGCs (Lee et al., 2006; Lien, et al. 2014). Disruption of the normal cell cycle in PGCs could lead to apoptosis in this lineage.

The activity of *efnb1* in PGCs was next investigated by morpholino mediated knockdown. When *efnb1* was inhibited, PGC number was not affected, although PGC migration patterns were. Embryos that were injected with *efnb1* morpholino showed a significant increase in posterior localization of PGCs within the tailbud embryo. These results may be due to a disruption in normal anterior/posterior gradients of *efnb1* in the embryo. Neuronal guidance has been related to proper gradient distribution of EphrinA ligands and EphA receptor signaling (Sterratt, 2013). It may be that the embryo has a normal anterior/posterior gradient of Ephb receptors, where concentrated EphB is in the posterior somatic tissue. When PGCs containing endogenous EphrinB1 reach the normal posterior boundary of EphB concentration, cell signaling for adhesion (*EphrinB1 → EphB*) could occur, inhibiting extreme posterior migration. When we knocked out *efnb1*, we may lose this adhesion signaling, causing the PGCs to continue to migrate in a posterior direction. Further functional experiments are required to understand the full effect that *efnb1* has on PGC migration and development.

### 5.4 Sox7 is required for PGC development

Here we show a novel role for *sox7* in PGC development. *sox7* was expressed in germ plasm, and its expression persisted in PGCs after segregation from the endoderm lineage (Fig.2. 3B). Surprisingly, both PGC-directed knockdown and overexpression of *sox7* caused a significant increase in PGC number at the gastrula stage embryo (Fig.4.6. A, B). However, increases in PGC number were no longer observed in tailbud stage embryos,
which only possessed normal (knockdown) or significantly decreased (overexpression) numbers of PGCs (Fig. 4.6 C,D). Similarly, when sox7 was overexpressed in single-cell embryos, a significant decrease in PGC number was observed at tailbud stages (Fig. 4.4, Fig. 4.5). Interestingly, sox7 knockdown by injection into the single cell fertilized egg did not mirror the results from sox7 knockdown after PGC-directed injections. In the latter case, a significant decrease in PGC number occurred (Fig. 4.4, Fig. 4.5). It is noteworthy that the single-cell injections were not analyzed for PGC numbers at the gastrula stage. These results strongly suggest that the level and temporal regulation of sox7 activity in PGCs, and perhaps also within the PGC niche, is crucial for their normal development, possibly by activating gene networks specific to germline formation.

In Xenopus, germ plasm is associated with the cell membrane of pPGCs prior to MBT at stage 8. Between stage 9 and 10, germ plasm then localizes to a perinuclear position (Taguchi et al., 2012; Butler et al., in Revision 2017). At this juncture, future PGC divisions are symmetrical, resulting in both daughter cells inheriting germ plasm. One possible explanation for the increased PGC number at gastrula is that sox7 manipulation in expression forced germ plasm into a perinuclear position prematurely, thereby resulting in premature symmetric divisions of pPGCs to yield daughter cells with equal amounts of germ plasm. Butler et al. (in Revision, 2017) recently found that both PGC-directed overexpression and knockdown of sox7 results in premature perinuclear localization of germ plasm. In affected embryos, germ plasm became dissociated with the cell membrane as early as stage 8, allowing premature symmetrical germ plasm inheritance. Premature perinuclear localization of germ plasm may account for the increase in PGC number we observed at gastrula.
Previously, Venkatarama et al. showed that PGCs initiate their zygotic program after localization of germ plasm to a perinuclear position has occurred (2010). These results, including my data, suggest that in addition to perinuclear localization of germ plasm, alteration of sox7 activity in the early embryo may prematurely initiate somatic gene expression in the PGC population. Butler et al. also provided data in support of sox7 functioning in the temporal initiation of the zygotic transcription program in the PGC population (2017). When sox7 was knocked down or overexpressed by PGC-directed injection, a significant increase in transcriptional activity occurred in stage 10 (early gastrula) PGCs when compared to uninjected controls as indicated by an active RNA Polymerase indicator (Butler et al., in Revision 2017). Taken together with our results of increased PGC number occurring in the gastrula embryo, we hypothesized that sox7 may be a required initiator for the zygotic program in PGCs.

I utilized WISH studies to test if knockdown of sox7 could alter expression patterns of RNAs found to be enriched in our zygotically active PGC sample set. Thirteen PGC-enriched transcripts (fer, pphln1, prpsap2, cdc20, fam168b, lmnb3, nasp, pgam1, tspan1, ppp1r2, xpat, hfoo1, and impad1) were selected to test whether PGC-directed sox7 MO-mediated knockdown altered their expression pattern in the neurula stage embryo. Surprisingly, I found that none of the selected transcripts appeared to have altered expression patterns when sox7 activity was knocked down. The most likely explanation for these negative data is that sox7 is not a required initiation factor for the zygotic expression of the selected genes. This explanation does not rule out that sox7 may be required for zygotic expression of other PGC enriched genes. Another possibility is that the sensitivity of WISH did not allow detection of expression patterns of the
expressed transcripts. A known limitation of WISH analysis is it cannot be utilized for the detection of low expressing transcripts or miniscule fold-changes in gene expression. An alternative approach to determine if sox7 is required for initiating the zygotic transcription of PGC-enriched genes would be to perform RNA-seq on zygotically active PGC samples of uninjected control and sox7 knockdown embryos. This approach would resolve the sensitivity issue of the WISH analysis and would determine effects of sox7 depletion on PGC enriched transcripts. 

Even though my WISH analysis did not show any tested transcripts to be affected by sox7 inhibition, other cell signaling mechanisms may have been affected by sox7 manipulation; namely the Wnt signaling pathway and/or genes involved in somatic differentiation. Canonical Wnt signaling has been associated with cell proliferation, and sox7 has been shown to harbor a beta-catenin binding site (Reviewed by Mohammed et al., 2016; Guo et al., 2008). When we knocked down Sox7 protein, we may have in turn, caused an increase in the availability of cytosolic beta-catenin, and thus promoted Wnt-mediated proliferation in the PGC population. As mentioned previously, we also witnessed a similar increase in PGC number when sox7 was over-expressed. Sox7 has been shown to positively regulate somatic genes necessary for endoderm differentiation. These include the nodal-related protein-encoding genes, as well as endodermin and mixer (Zhang et al., 2005a). Thus, sox7 overexpression in PGCs may have resulted in ectopic expression of such somatic genes that are normally translationally silent in PGCs. If so, it is possible that the increased PGC number observed at gastrula could be the result of PGCs becoming more somatic like, and proliferating at a rate similar to surrounding somatic cells of the early embryo. Furthermore, Lai et al. have shown that somatic gene
expression in the PGC population leads to an apoptotic fate (2012). Therefore, the PGCs expressing somatic determinants would eventually die, consistent with our observation that PGC-directed over-expression of \textit{sox7} caused a significant decrease in PGC numbers by tailbud stages.

As mentioned, in both PGC-directed \textit{sox7} overexpression and knockdown embryos, significant increases in PGC number were not maintained in the later tailbud stages when compared to that of uninjected controls. Considering that there is only a finite amount of maternally inherited germ plasm, premature symmetric divisions of pPGCs may result in one or both daughter cells not inheriting enough germ plasm for future specification. Therefore, the increased PGC numbers observed at gastrula in \textit{sox7} miss-expressed embryos may not have been maintained later in development due to insufficient amounts of germ plasm. Lineage-tracing studies would be necessary to determine if this hypothesis is true.

\textit{sox7 MO} injections into fertilized eggs caused a significant reduction in PGC number, unlike the normal numbers recorded for the PGC-directed experiments. Notably, the differences observed in both of these knockdown experiments may be due to the experimental design. The critical difference between these two experimental groups is that the incorporation of \textit{sox7 MO} was restricted to PGCs and closely associated endoderm in the PGC-directed injection group (Fig. 4.6), whereas \textit{sox7 MO} was distributed to the entire embryo in the single-cell injections (Fig. 4.4, Fig. 4.5). Proper mesoderm and endoderm formation requires signaling by VegT, a direct regulator of both maternal and zygotic \textit{sox7} (Charney et al., 2017). \textit{vegT} RNA was found enriched in both maternal and zygotic germ plasm samples. Dysregulation of \textit{sox7} expression in PGCs and
surrounding endoderm could have dysregulated downstream VegT signaling, which is dependent on *sox7*. It has been suggested that signaling by the endoderm is required for proper PGC migration (Morichika et al., 2010). Taken together, if more endoderm was affected by *sox7-MO* (as in the single-cell injections), then appropriate signaling of downstream targets for both VegT and Sox7 may not have occurred. A lack of signal could then result in more undifferentiated endoderm cells, disrupting normal endoderm-PGC signaling. These possibilities may account for the differences observed in the tailbud stages of single-cell versus PGC-directed *sox7* knockdown embryos.

These recent findings, along with the data presented in my dissertation, have led us to hypothesize that Sox7 may be a key transcription factor necessary to specify PGCs in Xenopus. Future studies are necessary to establish the temporal expression of Sox7 protein, and determine which genes Sox7 regulates to specify PGCs.

### 5.5 The vegetal pole is a signaling center

Axis specification of the early vertebrate embryo requires orchestrated inductive signaling to neighboring tissues and cell types within the embryo. Such signaling has been identified to occur in designated signaling “centers” within the early developing embryo. To date, two signaling centers related to axis and body plan organization have been identified: the Spemann-Mangold organizer and the Nieuwkoop center. Here, I propose the identification of an additional signaling center based on my recent RNA-seq analysis of the *Xenopus laevis* oocyte and in silico modeling of gene regulatory networks that may be active in the vegetal pole.
Spemann and Mangold (1924) were the first to identify such a “center”, located within the dorsal lip of the developing salamander embryo. They found that transplantation of the dorsal lip to the ventral side of a competent recipient embryo caused the formation of Siamese twins, or a duplicated axis (Reviewed in De Robertis, 2006). The dorsal lip was termed the “Spemann-Mangold organizer” for the inductive properties that this collection of cells possessed. cDNA cloning of this organizer performed by the De Robertis lab has provided information on many genes that are, in part, responsible for the inductive properties of this organizer. Major identified players to have inductive properties for the differentiation of surrounding tissues are goosecoid (homeobox gene-DNA binding protein(activator)), chordin (anti-BMP), noggin (anti-BMP), and follastatin(anti-BMP). The molecular machinery of the Spemann-Mangold organizer works to antagonize BMP and Wnt signaling in dorsal animal pole blastomeres of the embryo resulting in the default fate neural ectoderm there instead of epidermis. De Robertis and Kuroda also suggested the finding of an oppositional organizer to the Spemann-Mangold organizer, the ventral center (2004). The ventral center secretes proteins of similar biochemical activities as that of the Spemann-Mangold organizer. Interestingly, the ventral center orchestrates high levels of BMP4 and BMP7 signaling to regulate gene activations of similar genes in the Spemann-mangold organizer which require low levels of BMP signaling such as Chordin and Noggin (Kuroda et al., 2004). These oppositely placed centers function to regulate the embryo body plan along the D/V axis.

The formation of the Spemann-Mangold organizer requires signaling from a “pre-organizer”. This signaling is provided by the Nieuwkoop center (NC). Identification of
this center occurred from experiments performed by Pieter Nieuwkoop (and Boterenbrood) in 1973, where he showed that the Spemann-Mangold organizer developed in the animal cells receiving inductive signals by the most dorsal vegetal cells of the early embryo. The Nieuwkoop signaling center (NC) can occur any place within the equatorial zone of the embryo, but its location is ultimately determined by the sperm entry point (Reviewed in Gerhart, 1999). Wherever sperm enters at the animal pole of the egg, the NC forms at the opposite position in the embryo. Signaling of the NC helps to drive the formation of the Spemann-Mangold organizer. In addition to the NC signaling for the Spemann-Mangold organizer, the blastula chordin noggin expressing (BCNE) center forms above the NC in dorsal animal cells (Kuroda and DeRobertis, 2004). Both the NC and the BCNE center provide signaling required for proper Spemann-Mangold organizer formation.

My thesis work suggests the vegetal pole of the oocyte may be a signaling center which sets up the competency to respond to inductive signaling driving the body pattern of the embryo. My findings also build on the already known recognition of the vegetal pole as a site involved in dorsal/ventral axis specification, which normally requires microtubule driven cortical rotation. Taken together, I propose that the vegetal pole is both an organizer and kingpin, consisting of a scaffold framework upon which signal orchestration can initiate dorsal/ventral axis specification.

After fertilization, microtubules form microarrays exclusively at the vegetal pole of the early Xenopus embryo and these are required for translocation of the still unidentified dorsal determinants. How this microtubule network is restricted to the vegetal pole is not fully understood. Mei et al. demonstrated that cortically located Dead-
end protein is required to retain Trim36 RNA, an E3 ubiquitin ligase, within the vegetal cortical region. The retention of trim36 RNA results in Trim36 protein accumulation, which is required for microtubule arrays to form (2013). My RNA-seq data places JNK/mapk8 and slain1 as being vegetally enriched, two possible players in the formation of such arrays. JNK plays a key role in preventing nuclear β-catenin accumulation and thus may regulate the timing of axis specification (Liao et al., 2006). JNK/Mapk8 also increases microtubule stability by phosphorylating Stathmin present in the oocyte (Tararuk et al., 2006; unpub. data). Slain1 binds to microtubule plus-ends and promotes their nucleation and elongation (Van Der Vaart et al., 2012). One testable model is Slain1 is inhibited at the vegetal pole, but upon fertilization, Trim36 degrades the inhibitor(s) thereby allowing microtubule elongation and microarrays to form (Fig. 5).

An important question is what provides the structural integrity at the vegetal pole that accounts for the asymmetric distribution of mRNAs and a platform for signaling. The RNA-seq data reveal possible new components of the dorsal axis network, some of which have been identified as germ plasm components: grip2, syntabulin, kinesin13b, ephrinB1, wnt11b, JNK and the frizzled/LRP receptors. grip2 is the most abundant RNA at the vegetal pole. It seems likely that Grip2 is present as an abundant maternal protein as preventing its translation in the embryo resulted in only mild phenotypes (Kirilenko et al., 2008). Grip2 has been genetically linked to Syntabulin, a kinesin associated cargo adaptor, where it is required for dorsal specification in zebrafish and frogs (Ge et al., 2014; Colozza and De Robertis, 2014). However, syntabulin is also in germ plasm where it may associate with kinesin13b, required for PGC migration (Colozza and De Robertis, 2014; Tarbashevich et al., 2011). Brückner et al. (1999) have suggested that
Figure 8. Model for EphrinB1/Grip2 scaffold promoting signaling at vegetal pole (see text for details). Transmembrane Enfb1/EphrinB1 binds Dishevelled and recruits Grip2 to lipid rafts within germ plasm (tan) ER or the plasma membrane where receptors frizzled and LRP are found. Grip2 interacts with Syntabulin, facilitating kinesin microtubule-transport of vesicles bearing dorsal determinants. Microtubule arrays form after fertilization and require Trim36, Slain and JNK.

the ephrinB1-grip2 interaction is acting as a platform for membrane signaling through action of the many PDZ domains. We propose that Grip2 provides a scaffold at the vegetal pole as well as in the endoplasmic reticulum rich germ plasm that regulates Wnt signaling in both dorsal specification and PGC migration (Fig. 5).

To further stress the signaling capability of the vegetal pole, it also contains required components that ensure endoderm/ mesoderm formation and germline specification. As discussed in detail, germ plasm is located at the vegetal pole and is required for germline development. *vegt* and *vgl* are also enriched at the vegetal pole and are required for proper germ layer development. VegT is an essential signaling molecule that plays a role in forming all three germ layers (Zhang et al., 1998). Outside of signaling, *vegt* RNA is required for anchoring a subset of vegetally enriched RNAs to the
vegetal cortex, like *vg1, wnt11* and *bicaudal-C* (Heasman et al., 2001). Vg1, another essential signaling molecule required for mesoderm formation, also requires actin microfilaments and intermediate filaments in addition to *vegt* RNA, to be properly anchored to the vegetal cortex (Pondel et al., 1988; Yisraeli et al., 1990, Mosquera et al., 1993). These findings on *vg1 /vegt* show that maternally inherited RNAs at the vegetal pole can provide some of the earliest signals (*vg1* and *vegt*) including required structural components (*vegt*), for proper embryo formation. In fact, such early signaling by Vg1 and VegT have been contributed to the induction of the Nieuwkoop center itself (Agius et al., 2000; reviewed in DeRobertis and Kuroda, 2004). Increasing evidence found by others and my lab has led to the idea that the vegetal pole is an earlier signaling center that is active prior to the Nieuwkoop center.

One major objection to considering the vegetal pole as a signaling center is its cortex has not been shown to have inductive activity. As the vegetal cortex is not a tissue that can be transplanted (as the Nieuwkoop center or the Spemann-Mangold organizer) to a host to prove its inductive capability, the question has remained open. Transplantation of vegetally localized germ plasm has been shown to be required and sufficient for PGC determination (Tada et al., 2012), but other inductive properties showing the vegetal pole to be a recognized signaling center have yet to occur. Future investigations are required to further validate the vegetal pole as a signaling center on the same level as the NC and Organizer. The RNA-seq data presented in my dissertation, along with the generated gene regulatory networks, reveals interconnected functional pathways that highlight where future gene knock-down studies would be most instructive.
SIGNIFICANCE

The germline is crucial for the inheritance of genetic material from one filial generation to the next in sexually reproducing organisms. Germ plasm is the maternally inherited stock pile of RNAs and proteins required for germline specification within the future embryo. The primordial germ cell (PGC) population also shares hallmark characteristics of cancer, including proliferation, migration, and totipotent potential. To understand the molecular components of germ plasm is of clinical importance for identifying gene regulatory networks controlling the tight regulation for PGC development. Here, my thesis work provides sox7, efnb1, wwtr1, otx1, p300, parn, e2f1, and rras2 as germ plasm components that are required for normal PGC development. For the very first time a germ plasm driven RNA-seq analysis has yielded a comprehensive data set of transcripts found enriched at the vegetal pole, where germ plasm exists. Furthermore, this data set is immortal in that it is available for future interrogations.

Gene regulatory networks required for germline specification are poorly understood. Additionally, the specific maternal transcription factor(s) required for initiating the zygotic PGC program have yet to be defined. My thesis work provides an extensive list, in addition to initial studies, of enriched transcripts of both maternal and zygotic germ plasm components, giving insight to their function in germline development. My thesis work also yields evidence for previously unseen similarities of gene regulatory networks between neuron and germline lineages. Therefore, my thesis work provides extensive data that can be utilized and applied in disciplines such as neuroscience, germ cell, stem cell, and cancer biology.
6.1 Isolation of animal and vegetal pole samples

*X. laevis* adult frogs were purchased from Xenopus Express. Ovarian tissue was surgically removed from anesthetized females, then teased into small clumps and rinsed in Modified Barth’s Solution. Next, Oocytes were enzymatically released from ovarian tissue and from follicles with 0.2% collagenase (Type 1; Worthington Biochemical Corp.) in calcium free OR2 (Sive et al., 2000). Stage VI oocytes were selected and incubated overnight in OR2 with 1 mM CaCl2 at 18°C. The following day, oocytes were transferred into P10EM solution (Elinson et al., 1993) and oocyte matched vegetal and animal poles (10-20% of total oocyte for each pole) were cut with a razor blade (Cuykendall and Houston, 2010). Prior to collection, the germinal vesicle (GV) was manually removed from animal pole samples to ensure that GV retention of transcripts with different final locations would not contribute to false positives in our RNA-seq analysis. Samples were immediately collected into Eppendorf tubes, frozen on dry ice, then stored at -80°C until RNA was extracted. RNA was extracted and 25 vegetal and animal oocyte-paired poles were collected per frog. Equal concentrations of RNA from the respective poles of oocytes from three different frogs were combined to make one vegetal and one animal pole paired sample. Oocytes from a total of nine frogs were used to form three vegetal and three animal pole samples that were submitted for RNA-seq
analysis. All animal protocols were approved by the Institutional Animal Care and Use Committee of the University of Miami.

6.2 RNA isolation

Total RNA was isolated using Trizol reagent (Invitrogen) as per the manufacturer’s protocol with the following modifications: samples were homogenized before adding Trizol reagent; RNA precipitation was done overnight in isopropanol and the RNA pellets were rinsed three times with ice-cold ethanol then air-dried and reconstituted in nuclease free water. RNA concentration was determined using Nanodrop 2000c (Thermo Scientific). For each sample, total RNA was used to synthesize cDNA via the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). All procedures for RNA isolation were done using filtered pipette tips. X. laevis nanos RNA is exclusively found within the germ plasm and associates with the vegetal cortex (Elinson et al., 1993). To assess both the purity and cortical quality of our samples, we determined the amount of nanos RNA in the vegetal and animal pole samples by semi-quantitative RT-PCR (Lai et al., 2011). Vegetal pole samples with the highest levels of nanos RNA and oocyte-matched animal pole samples with no detectable nanos RNA were selected for RNA-seq analysis.

6.3 RNA preparation for Illumina sequencing

For each sample, 1 μg of total RNA was processed for RNA quality with the Agilent Bioanalyzer 2100. Samples were processed for both RNA-seq and small (sm)RNA-seq. RNA-seq samples were depleted of mitochondrial and ribosomal RNAs with the
ScriptSeq Complete Gold Kit (Illumina) and subjected to ten cycles of PCR prior to RNA sequencing on the Illumina HiSeq 2000 using the reagents provided in the Illumina TruSeq PE Cluster Kit v3 and the TruSeq SBS Kit-HS (200 cycle) kits. Reads aligning to a ribosome-specific reference or mitochondrial sequences represented <5% and 1.28% of the total, respectively. For smRNA-seq, samples underwent 11 cycles of PCR and were then prepped using the Illumina TruSeq Small RNA Sample Preparation Guide (15004197 Rev. D). Cluster generation was performed on the Illumina cBot according to the manufacturer’s recommendations. An average of 10.1 million (animal) and 19.4 million (vegetal) pass-filter paired-end 100 base reads were generated per sample (range: 9.0-11.2 million, animal; 17.8-22.7 million, vegetal).

6.4 Data processing and quantification

Quality control metrics were determined using FastQC software (Babraham Bioinformatics). The total library size was 2.8 Gb, with 95% of the total base pairs above internal FastQC thresholds. Variance between samples was minimal with sample reads ranging from 24.6-26.8 and 21.2-21.5 million read pairs for animal and vegetal pole samples, respectively. Raw reads were aligned using TopHat v2 RNA-seq analysis software. To assess the quality of the reference genome, reads were aligned to both X. laevis (v6.0 and v7.1) and X. tropicalis (v7.1) references. Alignment to the X. laevis genome ranged from 66.42%-71.45% and the X. tropicalis genome ranged from 25.21%-39.27%. Therefore, reads from the X. laevis genome alignment were quantified.

Transcripts were quantified using Cufflinks v2.1. At least 21,555 transcripts were detected in each sample, and of those 13,930 had a fragments per kilobase of transcript
per million mapped reads (FPKM) value ≥5. Differential expression analysis was performed using Cuffdiff to compare vegetal versus animal pole transcripts with an FPKM value ≥5 (Mortazavi et al., 2008). 5717 total transcripts (FDR<0.05, fold change ≥1.95) were differentially expressed. More stringent criteria were then employed to determine transcripts enriched in the vegetal or animal pole. The bulk of large yolk platelets are within the vegetal hemisphere, reducing the yolk-free cytoplasm content there (Danilchik and Gerhart, 1987; Callen et al., 1980; Rebagliati et al., 1985); therefore, we set a criterion of 4-fold enrichment for vegetal versus animal pole transcripts. An inherent bias exists towards the animal pole based on the 4-fold difference in RNA concentrations along this axis; thus, transcripts enriched 10-fold in the animal pole compared with the vegetal pole were considered localized to the animal pole.

6.5 Small RNA analysis and novel non-coding RNA identification

For small RNA analysis, Cutadapt was used to remove adapter sequences from read ends. Reads trimmed to < 20bp were discarded. Next, the reads were aligned to a *X. laevis* specific microRNA (miRNA) reference using TopHat v2. 23% alignment was achieved. Samtools was used to filter then count the aligned reads using the following command:

```
$ samtools view aligned_file.bam miR21 > miR21.sam | wc –l
```

Since *X. laevis* miRNAs are not well annotated, we identified miRNA homologues in *X. tropicalis*. We subsequently queried TargetScan for miRNA targets. Target mRNAs were identified based on highly conserved regions in homologues human genes. Thus *X.*
tropicalis miRNAs that match with human mRNA target homologs of vegetally localized RNAs from our RNA-seq analysis were reported.

To identify novel long non-coding (Inc) RNAs, Trinity software (Trinity version 2012-06-08) was used for de novo assembly of the transcripts in each sample (Grabherr et al., 2011). These assemblies were analyzed by Coding-Potential Assessment Tools (CPAT) and Coding Potential Calculator (CPC), which determined that there were no viable transcripts. Further analysis of the potential function of the IncRNAs was done using advanced modeling, but nothing significant could be detected based on the current platforms. Further analyses of IncRNAs will have to await better annotation of the Xenopus genome.

6.6 Gene name identification and GO analysis

Joint Genome Institute (JGI) annotation was used for gene name identification (Xenbase.org). To verify gene identity, separate BLAST alignments (NCBI) for the scaffolds were also performed. Genes homologous to those of humans were submitted to GeneCards for summary information (genecards.org).

The total transcript set of all vegetal pole up-regulated genes (≥4 fold enrichment; total = 198) was submitted to GeneGO v6.18 (MetaCore Bioinformatics and Pathway Analysis software from Thomson Reuters, https://portal.genego.com/) for enrichment and pathway analysis as previously described (Egan et al., 2014). Pathway analysis was determined using the direct interaction and analyze network building algorithms. Un-annotated sequences represented approximately half of the total RNAs enriched at either
the animal (12/27) or vegetal poles (213/411) (Fig. 1D) and these are listed at http://www.biologists.com/DEV_Movies/DEV139220/TableS2.xlsx.

6.7 Whole-mount in situ hybridization (WISH)

*X. laevis* embryos were obtained as described in Sive et al., 2000. WISH was performed as described in Agüero et al., 2012 with the exception of day 1, which was performed as described in Colozza and De Robertis, 2014. Plasmids containing full-length clones were purchased from GE Dharmacon (*spire1, atrx, wnk2, rras2, mov10, trank1, sybu, otx1, e2f1, rtn3, pgam1, ppp1r2, zfyve26 and dand5*) and Transomic Technologies (*efnb1, slc18a2 fer, pphln1, cdc20, lmn3b, nasp, tspan1, hfoo1, and impad1*), and provided by Dr. Jing Yang (*hook2, tob2, and sox7*). Inserts were verified by restriction digestion and sequencing. Inserts were PCR amplified using the following primer sets: T7/Sp6 (*efnb1, rras2, mov10, tob2, sybu, sox7, rnf38, spire1, hook2, e2f1, trank1, otx1, pphln1, cdc20, fer, h1foo, impad1, lmn3b, rtn3, fam168b, pgam1, zfyve26, prpsap2and slc18a2*) or T3/Sp6 (*ppp1r2, tspan1, nasp, wnk2, atrx, and dand5*). Xpat clone was synthesized as described in Lai et al., 2012. Antisense probes containing Digoxigenin-11-UTP were synthesized using the T7, SP6 or T3 RNA polymerase (NEB). Primer sequences: T7: 5’-taatacgactcactaatag-3’; T3: 5’-attaacacctcactaaag-3’; Sp6: 5’-attagtgacactatag-3’.

6.8 PGC-targeted injections for sox7 manipulation

For targeting the germ plasm more specifically, sox7-MO (16ng total), sox7-FL (150pg), sox7-MO (16ng total) and sox7-FL-rescue (150pg) together was injected in the germ plasm bearing blastomeres of 16-32-cell embryos. (Fig. 4.2)
6.9 Gain-of-function analysis of vegetally enriched genes of interest

cDNA clones encoding full-length genes were obtained from GE-Dharmacon (\textit{parn}, \# MXL1736-202774978; \textit{rras2}, \# MXL1736-202773625; \textit{wwtr1}, \# MXL1736-202784975; \textit{e2f1}, \# MXL1736-202809814; \textit{otx1}, \# MXT1765-202789330; \textit{spire1}, \# MXL1736-202787536), transOMIC (\textit{efnb1}, ID: TCL1007) or as a gift (pCS107-Xt\textit{sox7}, from Dr. Jing Yang, University of Illinois at Urbana). Synthetic capped RNAs for microinjection were obtained by \textit{in vitro} transcription using the mMESSAGE mMACHINE SP6, T7 or T3 Kit (Ambion). DNA templates were linearized or amplified by PCR, transcribed and 0.5 ng of each mRNA was injected in the vegetal pole of 1-cell embryos as follows: pCS107-Xt\textit{sox7} (SP6/NotI); pCMV-SPORT6-Xlp\textit{arn} (SP6/NotI), pCMVSPORT6-Xl\textit{rras2} (SP6/NotI); pExpress1-Xle\textit{e2f1} (SP6/PCR: T7/SP6); pExpress1-Xl\textit{otx1}(SP6/PCR: T7/SP6); pCR4-TOPO-Xl\textit{efnb1} (T3/PCR: T3/T7); pCMV-SPORT6-Xl\textit{spire1} (SP6/NotI). All clone identities were confirmed by sequencing. All results shown are representative of at least two independent experiments. All embryos were fixed for WISH at stage 32-35 (tailbud). Embryos were staged according to Nieuwkoop and Faber (1956). WISH was performed using the germ plasm marker \textit{xpat} to see PGC differences of injected RNAs. Both sides of each embryo were used for counting PGC number.

6.10 Loss-of-Function analysis of novel germline RNAs

Morpholinos targeting the following RNAs were purchased from gene tools: \textit{efnb1}-MO (5’-CCACCTGTGCCGGATGGAAGGGCTCC), \textit{sox7}-MO (5’-GTCATTATTCCAACCTGACTTGCTGA), \textit{otx1}-MO (5’-TAGGACACATCATGCTCAAGGGCTGGAT), and \textit{wwtr1}-MO (5’-
TGTTACAACAGCTACTTCCCAAGGC). To test MO efficiency full length *X. laevis* efnb1 was PCR amplified from stage VI oocyte cDNA with the following forward and reverse primers: 5’- GGATAATACAAAGCTGGTTTCTGTG and 5’- CTTTGCTCTGTGATTGGATTG, respectively, and subsequently a flag tag (FL) was added using the following primers: forward: 5’- CAGCATGAATTTCGAGTTTAGCAGCTGAGGGCAAG; reverse: 5’- ATATTCTCGAGCTACTTGCGTCATCGTCTTTGTAGTCTTTGTAGTAAATGTGTTT GCAGG. The flag-tagged efnb1 insert was then cloned into PCS2+ using EcoRI/XhoI to make efnb1-FL. Additionally, *otx1* and *wwtr1* (GE-Dharmacon) were cloned into flag(FL)-pCS2+ using BamHI/Clai and Clai/Clai to generate *otx1*-FL and *wwtr1*-FL, respectively. Flag-tagged full length *sox7* (*sox7*-FL) was synthesized and purchased from Genewiz. efnb1-FL-rescue and sox7-FL-rescue were generated from efnb1-FL and *sox7*-FL, respectively, by introducing conservative mutations in the region that binds the morpholinos (efnb1-FL-rescue: 5’ATGGAgGGtCTtCGGCGTCTTCTC; sox7-FL-rescue: 5’-TaAGtAAaTCcGTgGGcATcATGAC), rendering the respective morpholinos ineffective. Mutations were introduced using the Q5 Site-Directed Mutagenesis Kit (New England Biosciences, E0554) according to the manufacturer’s protocol.

All morpholinos were tested for knockdown efficiency using the Wheat Germ Extract kit (Promega, REF: L4380) according to the manufacturer’s protocol. Translation of efnb1-FL, *otx1*-FL, *wwtr1*-FL, and *sox7*-FL were detected by Western blot analysis for flag (Figs. S2 and S3). Primary antibody used: Monoclonal mouse ANTI-FLAG (Sigma, REF: F1804); secondary antibody used: Anti-Mouse IgG HRP Conjugate (Promega, REF: W402B).
Single-cell embryos were injected vegetally with *otx1*-MO (15ng), *wwtr1*-MO (15ng), *efnb1*-MO (16ng), *efnb1*-FL (200pg), *efnb1*-FL (200pg) and *efnb1*-MO (16ng) together, *efnb1*-FL-rescue (200pg) and *efnb1*-MO (16ng) together, or scramble-MO (16ng) at the one-cell stage. Embryos were also injected with *sox7*-MO (16ng), *sox7*-dCEnR (*X. laevis*) (200pg), Xt*sox7* (*X. tropicalis*) (200pg), or *sox7*-dCEnR (200pg) and Xt*sox7* (200pg) mRNAs at the one-cell stage. PGC-targeted injections were performed on 16-32 cell embryos with the following: *sox7*-MO (16ng total), *sox7*-FL (150pg), *sox7*-MO (16ng total) and *sox7*-FL-rescue (150pg) together. *X. laevis* *sox7* dominant negative construct (*sox7*-dCEnR) and *X. tropicalis* *sox7* WT (Xt*sox7*) were generous gifts from Dr. Mike Klymkowsky (University of Colorado, Boulder) and Dr. Jing Yang (University of Illinois at Urbana), respectively. In *sox7*-dCEnR the transacting domain was deleted and the engrailed transcriptional repression domain was inserted downstream of the HMG box (Zhang et al., 2005). Synthetic capped mRNAs for microinjection were obtained by *in vitro* transcription using the mMMESSAGE mMACHINE SP6 or T3 Kit (Ambion). DNA templates were linearized then transcribed as follows: *efnb1*-FL (ApaI/SP6); *efnb1*-FL-rescue (ApaI/SP6); *sox7*-dCEnR (XhoI/T3); Xt*sox7* (NotI/SP6).

To inhibit p300 activity, embryos were incubated with the small molecule inhibitor of p300, C646, or DMSO as a control. To prevent reported light induced retinopathy by C646 (Kawase et al., 2016), embryos were protected from light.

Injected embryos were collected and fixed at stages 11.5 (PGC-directed injections) and 33/34 (PGC-directed or single-cell), and staged according to Nieuwkoop and Faber (1956). Treated embryos were collected at 33/34, similarly. PGCs were identified using
WISH against xpat as previously described (Lai, et al., 2012). At stage 11.5, PGC number per embryo was calculated by manually counting PGCs in each embryo. The yolk plugs along with the endodermal cores, which contain the PGCs, were removed from stage 11.5 embryos and dissected into individual cells to ensure all PGCs were counted. Note: PGCs on the surface of the yolk plug stained purple, and internal PGCs stained bluish-purple.

All results shown are representative of at least three independent experiments. The p(t)-values were determined using a two-tailed unpaired Student’s t-test. P-values <0.05 were considered significant. In the tailbud stage embryo (stage 33/34), PGC number per tailbud embryo was calculated by counting PGCs on both sides of each embryo. An embryo was considered to have mis-localized PGCs if PGCs were located anterior to somite 5 or posterior to somite 11 as described in Tarbashevich et al. (2011). All results shown are representative of at least two independent experiments. The p(t)-values were determined using a two-tailed unpaired Student’s t-test. P-values <0.05 were considered significant.
REFERENCES


Chan A.P., Kloc M., Bilinski S., Etkin L.D., (2001). The vegetally localized mRNA fatvg is associated with the germ plasm in the early embryo and is later expressed in the fat body. Mech Dev. 100,137–140.


Kaneshiro, K., Miyauchi, M., Tanigawa, Y., Ikenishi, K., Komiyai, T., (2007). The mRNA coding for Xenopus glutamate receptor interacting protein 2 (XGRIP2) is maternally transcribed, transported through the late pathway and localized to the germ plasm. Biochem Biophys Res Commun 355, 902-906.


