

Spatiotemporal RNA expression in the *Caenorhabditis elegans* germline

vorgelegt von

M. Sc.

Asija Diag, geb. Sirucic

von der Fakultät III - Prozesswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften

- Dr. Ing.-

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Roland Lauster

Gutachter: Prof. Dr. Juri Rappsilber

Gutachter: Prof. Dr. Nikolaus Rajewsky

Gutachter: Dr. Baris Tursun

Gutachter: Dr. Jan Philipp Junker

Tag der wissenschaftlichen Aussprache: 03. September 2019

Berlin 2020

Zusammenfassung

Raumzeitlich beschränkte RNA-Expression ist ein zentraler, konservierter Mechanismus, welcher zum Erhalt von Gewebefunktionen, -morphologie und -integrität dient. Während der Entwicklung ist die Lokalisierung von RNA-Molekülen in subzellulären Kompartimenten ein weitverbreiteter Mechanismus zur Regulation des Gleichgewichts von Zell-Proliferation und -Differenzierung. Eine Störung in der Regulation dieses Gleichgewichts ist häufig mit Entwicklungsdefekten und Krankheiten wie zum Beispiel Krebs assoziiert. Allerdings ist unser Wissen darüber, wie raumzeitliche Verteilung von RNAs die Zell-Proliferation und -Differenzierung beeinflusst, limitiert.

In dieser Studie verwendeten wir die Keimbahn des Fadenwurms *Caenorhabditis elegans*, welches ein *in-vivo*-System zur Untersuchung von Proliferation und Differenzierung darstellt. Wir haben eine auf Kryoschnitten basierende Sequenzierungstechnik etabliert und optimiert, um RNA-Expression in nahezu Einzelzell-Auflösung und in Abhängigkeit von der Position entlang der Keimbahn darzustellen. Mit dieser Technik haben wir über 10,000 mRNAs und über 300 miRNAs, aber auch potentiell neue miRNAs und neue 3' UTR Isoformen detektiert und quantifiziert. Die meisten dieser RNAs wiesen hoch organisierte Lokalisierungsmuster auf. Interessanterweise waren nicht nur mRNAs und miRNAs lokalisiert, sondern auch endogene siRNAs und piRNAs. Zusätzlich haben wir gezeigt, dass die raumzeitliche Genexpression in der tumorösen und nur proliferierenden *gld-2 gld-1* Doppelmutante stark gestört war, was die Wichtigkeit der mRNA-Lokalisierung für den Erhalt des Proliferations- und Differenzierungsgleichgewichts unterstreicht. Darüber hinaus haben wir durch den Vergleich der Genexpression von Wildtyp und Mutante das PIE-1 Protein, welches ein Repressor der Transkription ist, als mögliches Schlüsselmolekül für den Erhalt und die Regulation der raumzeitlichen Genexpression identifiziert. Zusätzlich haben wir gezeigt, dass differenzielle 3' UTR-Nutzung, welche einen Schlüsselmechanismus zur Regulation der Genexpression in Raum und Zeit darstellt, ein häufiges Phänomen entlang der Wildtyp-Keimbahn ist. Erstaunlicherweise war die raumzeitliche Regulation der 3' UTR-Wahl in der *gld-2 gld-1*-Mutante signifikant gestört. Unsere Daten suggerieren, dass *cpsf-4* und *fipp-1* an der Regulation der differenziellen 3' UTR-Nutzung von fast 1,000 Genen in der Keimbahn beteiligt sind.

Abschließend haben wir ein physikalisches 3-D-Keimbahn-Modell erstellt, welches unsere ganzen Daten integriert und somit ein benutzerfreundliches Interface für die Erforschung von raumzeitlicher RNA-Expression während der Keimzellen-Proliferation und -Differenzierung darstellt (<https://shiny.mdc-berlin.de/spacegerm/>).

Abstract

Spatiotemporally-restricted RNA expression is a crucial, conserved mechanism for maintaining tissue viability, morphology, and integrity. In particular, localization of RNA molecules to subcellular compartments is a prevalent mechanism for regulating the balance between cell proliferation and differentiation during development. Perturbations in the regulation of this balance are often associated with developmental defects and diseases such as cancer. However, our understanding of how spatiotemporal distribution of RNAs shapes cell proliferation and differentiation is still limited.

In this study, we used the *Caenorhabditis elegans* hermaphrodite germline as an *in vivo* model for proliferation and differentiation. We adapted and optimized a cryo-cut-sequencing technique to determine, at near single cell resolution, RNA expression as a function of position along the germline. With this technique, we detected and quantified over 10,000 mRNAs and over 300 miRNAs but also numerous putative novel miRNAs and novel 3' UTR isoforms. Most of these RNAs displayed highly organized localization patterns. Interestingly, not only mRNAs and miRNAs but also endogenous siRNAs and piRNAs exhibited distinct localization patterns. In addition, we showed that the spatiotemporal gene expression was strongly perturbed in the tumorous and solely proliferating *gld-2 gld-1* double mutant germline indicating the importance of mRNA localization in maintaining the balance between proliferation and differentiation. Moreover, by comparing the gene expression of wild type and mutant, we identified PIE-1, a transcriptional repressor, as a potential key player in maintaining and regulating the spatiotemporal gene expression. Furthermore, we showed that differential 3' UTR usage, a key mechanism in regulating gene expression in space and time, is a common phenomenon along the germline. Surprisingly, the spatiotemporal regulation of the 3' UTR choice was significantly disrupted in the *gld-2 gld-1* mutant. Our data suggests that *cpsf-4* and *fipp-1* are involved in the regulation of differential 3' UTR usage for almost 1,000 genes in the germline. Finally, we constructed a physical 3D germline model that integrates our data providing a user-friendly interface for exploring spatiotemporally-restricted RNA expression during germ cell proliferation and differentiation (<https://shiny.mdc-berlin.de/spacegerm/>).

Contents

Acknowledgments	1
Abbreviations	3
Statement of contributions	5
1 Introduction	6
1.1 Spatiotemporal restriction of gene expression	6
1.2 Posttranscriptional regulation of spatiotemporally-restricted gene expression	7
1.3 The <i>Caenorhabditis elegans</i> germline as a powerful <i>in vivo</i> model	9
1.4 Cryo-cut method allows spatiotemporal reconstruction of gene expression	12
1.5 Aims of the thesis	13
2 Materials and Methods	15
2.1 Chemicals.....	15
2.2 Enzymes.....	16
2.3 Commercially available assays.....	17
2.4 Equipment	17
2.4.1 Consumables	17
2.4.2 Devices	17
2.5 Synthetic oligonucleotides	18
2.6 Buffers, solutions and growth media	20
2.7 Strains	21
2.8 Experimental details	21
2.8.1 Embedding and cryo-sectioning.....	21
2.8.2 mRNA library preparation	22
2.8.3 Small RNA library preparation	23
2.8.4 Poly(A)+-selected library preparation	23
2.8.5 Ribosomal RNA depleted total RNA library preparation	23
2.8.6 Probe preparation for mRNA in situ hybridization (ISH)	24
2.8.7 mRNA ISH	24
2.8.8 small RNA ISH	25
2.8.9 TaqMan® assays	25
2.8.10 Nested PCR.....	25
2.8.11 Physical gonad model.....	26
2.9 Computational pipeline	29
2.9.1 Data pre-processing	29
2.9.2 Mapping of reads to the <i>C. elegans</i> genome	30
2.9.3 3' extension of transcript annotation.....	30
2.9.4 Transcriptome pre-processing.....	30
2.9.5 Isoform-specific transcript abundance estimation	31
2.9.6 Data processing.....	31
2.9.7 Aligning cryo-cuts of different samples to a single coordinate system	31
2.9.8 Integration of replicate data.....	32
2.9.9 Data analysis of the small RNA transcriptome.....	32
3 Results	34

3.1	Cryo-cut based sequencing method is reproducible and reliable	34
3.2	Cryo-cut-seq creates a spatiotemporal RNA map of the <i>C. elegans</i> gonad.....	38
3.3	Germline-specific miRNA family co-localizes with putative targets.....	41
3.4	Germline-specific small RNA sequencing predicts novel miRNAs.....	42
3.5	Spatiotemporal gene expression is highly dynamic in the germline.....	48
3.6	Spatiotemporal gene expression is lost in the germline tumour mutant.....	50
3.7	Hundreds of 3' UTR isoforms extended in the germline	53
3.8	3' UTR choice is spatially regulated in the germline.....	55
3.9	Differential 3' UTR isoform usage is perturbed in <i>gld-2 gld-1</i> double mutants.....	56
3.10	Other small RNAs are spatially organized in the germline	57
3.11	A physical gonad model represents RNA localization in the germline	60
4	Discussion.....	63
4.1	mRNAs and small RNAs are spatiotemporally organized in the germline	63
4.2	Identification of novel miRNAs with spatiotemporal expression in the germline	65
4.3	Localized RNA expression is important for cell proliferation and differentiation.....	66
4.4	Differential 3' UTR isoform usage is strongly regulated in the germline	68
4.5	SPACEGERM: An interactive online data visualization tool	70
5	References	73
	Appendix.....	84
	Curriculum Vitae	88
	Selbstständigkeitserklärung	91

Acknowledgments

I would like to thank:

Nikolaus Rajewsky, for the great supervision, fruitful scientific discussions, constant motivation and support during my whole PhD.

Marcel Schilling, who joined the project to analyse the spatiotemporal expression of mRNAs and 3' UTRs and to create a physical 3D gonad model.

Filippos Klironomos, who teamed up with us to analyse the spatiotemporal expression of small RNAs.

Salah Ayoub, for helping with the establishment of the small RNA protocol but also for any technical help needed.

Margareta Herzog, for her help with the dissection of *C. elegans* gonads and any help concerning *C. elegans* maintenance and cultivation.

Andrei Filipchik, for supporting the project with his analysis about chimeric interactions of the novel miRNAs.

Kathrin Theil, *David Koppstein* and *Agnieszka Rybak-Wolf*, for critically reading of the manuscript and very helpful comments.

My committee meeting members *Jan Philipp Junker* and *Baris Tursun*, for helpful discussions and comments and my special thank goes to *Jan Philipp*, for introducing me to the tomo-seq technique.

The *whole Rajewsky lab* for helpful discussions and support. It was a great pleasure to work in this wonderful research atmosphere.

Alexandra Tschernycheff and *Nadine Ewald* for support with administrative issues.

The *Computational Systems Biology (CSB)* graduate school and the Helmholtz association, for funding and an exciting PhD program.

And most importantly, a special thank goes to my *amazing family*, for their constant support and motivation to reach my scientific goals. Without you, all this would not be possible.

Abbreviations

APA	Alternative polyadenylation
bp	base pair
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
cDNA	complementary desoxyribonucleic acid
CGC	Caenorhabditis Genetics Center
CPM	Counts per million
CPSF-4	Cleavage and polyadenylation specificity factor 4
CstF	Cleavage stimulation factor
cytoPAP	cytoplasmic poly(A)-polymerase
DAPI	4',6-diamidin-2-phenylindol
DNA	Desoxyribonucleic acid
DIG	Digoxigenin
dNTP	deoxynucleotide triphosphate
DTC	Distal tip cell
DTT	Dithiothreitol
e.g.	<i>exempli gratia</i>
EDTA	Ethylenediaminetetraacetic acid
ERCC	External RNA Controls Consortium
ESC	Embryonic stem cell
Fip1	Factor interacting with poly(A) polymerase
gc	germ cell
gcd	germ cell diameter
GFP	Green fluorescent protein
<i>gld-1/-2</i>	defective in germline development 1/2
<i>glp-1 (gf)</i>	abnormal germline proliferation (gain of function) 1
GTF	Gene Transfer Format
HF	High Fidelity
i.e.	<i>id est</i>
iPAR-CLIP	<i>in vivo</i> PAR-CLIP
iPSC	induced pluripotent stem cell
v/ISH	virtual/ <i>in situ</i> hybridization
IVT	<i>In vitro</i> transcription
kb	kilobase
LB	Lysogeny broth
LITE-Seq	Low-Input 3'-Terminal sequencing

LNA	Locked nucleic acid
LOESS	Locally estimated scatterplot smoothing
LoF	Loss-of-function
LTR	Long terminal repeats
miRNA	microRNA
mRNA	messenger RNA
NGM	Nematode growth medium
nov	novel
nt	nucleotide
PAR-CLIP	Photo-activatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PAS	Polyadenylation signal
PBS	Phosphate buffered saline
PBS-T	PBS with Tween
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
piRNA	piwi-interacting RNA
PNK	Polynucleotide kinase
RBP	RNA-binding protein
RdRP	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RNAi	RNA interference
RNP	Ribonucleoprotein
RPM	Reads per million
rRNA	ribosomal RNA
siRNA	small interfering RNA
smRNA	small RNA
SPACEGERM	Spatial <i>C. elegans</i> Germline Expression of mRNA and miRNA
SSC	Saline sodium citrate
TEA	Triethanolamine
TPM	Transcripts per million
tRNA	transfer RNA
TSV	Tab-separated values
UCSC	University of California, Santa Cruz
UMI	Unique molecular identifier
UTR	Untranslated region

Statement of contributions

Most of this study was published as:

Diag, A., Schilling, M., Klironomos, F., Ayoub, S., and Rajewsky, N. (2018). Spatiotemporal m(i)RNA Architecture and 3' UTR Regulation in the *C. elegans* Germline. *Dev. Cell*.

The methods section, figures 2, 4, 6 - 12, 14 -15, 17, 19, 20 - 22, 26 - 29, 32 - 33 and table 1 were published in a similar or slightly modified way in Diag et al., 2018. Figure legends and text paragraphs that were adopted for this thesis were marked with the reference "Diag et al., 2018".

Individual contributions for this project are as follows:

I designed and performed all laboratory experiments. I established and led the development of the project and wrote the paper manuscript together with Nikolaus Rajewsky, and with input from the other authors. Marcel Schilling performed the computational analysis for the mRNA data, designed the interactive data visualization tool, and constructed the 3D germline model together with my input. Filippos Klironomos helped with the computational analysis for the small RNA data. Salah Ayoub helped with the establishment of the small RNA library preparation and sequencing protocol. I led the interpretation of the data together with Nikolaus Rajewsky. Nikolaus Rajewsky supervised the project.

1 Introduction

1.1 Spatiotemporal restriction of gene expression

Targeting mRNA molecules to subcellular compartments is an essential mechanism for regulating gene expression in both space and time (reviewed in Besse and Ephrussi, 2008; Medioni et al., 2012). The localization of mRNAs is a conserved mechanism which is crucial for establishing functionally distinct compartments by different cell types. The first localized maternal mRNAs were discovered in *Drosophila* and *Xenopus* oocytes, which established that spatial restriction of RNA can cause early embryonic patterning (Berleth et al., 1988; Frigerio et al., 1986; Rebagliati et al., 1985). Although early studies speculated that mRNA localization to particular compartments is specifically restricted to a small subset of cell types such as oocytes, recent genome-wide analyses demonstrated that subcellular localization is a widespread mechanism in many other cell types as well (reviewed in Besse and Ephrussi, 2008). For example, in *Drosophila* embryos, 71 % of the 3,000 investigated transcripts were expressed in spatially distinct patterns, and in neurons, over 100 transcripts were restricted to neuronal processes (reviewed in Besse and Ephrussi, 2008; Lécuyer et al., 2007; Martin and Zukin, 2006). Localization of mRNAs has many advantages. It is more energy efficient than localizing proteins and one mRNA can serve as a template for multiple rounds of translation. Thus, synthesized proteins do not have to individually be transported to the site of action. Additionally, it enables local regulation of protein synthesis preventing proteins from acting ectopically at other sites (reviewed in Jansen, 2001; Martin and Ephrussi, 2009). The mechanisms by which mRNAs are localized to distinct compartments are not fully understood. However, recent live-cell imaging studies have suggested three possible mechanisms: mRNAs can be locally protected from degradation, transcripts can diffuse followed by local entrapment, and/or they can be actively transported along the cytoskeleton (reviewed in Medioni et al., 2012). Regardless of the mechanism used, targeting mRNAs to distinct compartments requires cis-acting elements encoded in the RNA. These 'localization elements' are primarily found in the 3' untranslated region (UTR), although they can also infrequently occur in the 5' UTR or coding region. The cis-acting elements are recognized and bound by RNA-binding proteins (RBPs) forming the ribonucleoprotein (RNP) complex (Besse and Ephrussi, 2008). Correspondingly, recent studies showed that alternative polyadenylation and, hence, the choice of the 3' UTR isoform leads to spatially-restricted gene expression and cell fate transition (Berkovits and Mayr, 2015; Brumbaugh et al., 2018; Ciolli Mattioli et al., 2018; reviewed in Mayr, 2017, 2018). Differential 3' UTR isoform usage is especially important for the balance between cell proliferation and differentiation. It has been shown that proliferating cells predominantly express shorter 3' UTR isoforms while differentiating cells use mainly the longer 3' UTR isoform (Lackford et al., 2014; Mayr and Bartel, 2009; Sandberg et al., 2008; Shepard et al., 2011). Proliferation and differentiation are distinct cellular states that are

fundamental for maintaining the health state of an organism. Differentiated cells have a limited capacity to proliferate while proliferative cells are usually not differentiated (Gingold et al., 2014). Maintaining the balance between proliferation and differentiation is of crucial importance to all tissues as unlimited proliferation can impede differentiation which can cause severe diseases such as developmental defects and cancer. Previous studies demonstrated that this balance is dependent on transcription, *i.e.*, dysregulation of transcriptional regulators impedes differentiation and leads to unlimited proliferation of cells (Gonda and Ramsay, 2015). However, our understanding of how mRNA localization and especially differential 3' UTR isoform usage regulate the balance *in vivo* is limited. Therefore, investigating the mRNA localization is critical for understanding the identity of cells, what makes them unique, and how cells keep the homeostasis and integrity of every tissue. Especially, germ cells and early developmental stages offer a powerful system to study mRNA localization as gene regulation in these systems relies on maternally-deposited mRNAs before transcription is switched on again during embryogenesis.

1.2 Posttranscriptional regulation of spatiotemporally-restricted gene expression

The state of proliferation and differentiation is defined by cell type-specific gene expression, which is itself regulated via diverse transcriptional and posttranscriptional mechanisms. One important posttranscriptional mechanism for modulating gene expression is alternative polyadenylation (APA), which is highly regulated during development and deregulated in disease. In humans, more than half of the genes use APA to generate isoforms with different 3' UTRs (reviewed in Mayr, 2017, 2018). Although some regions in 3' UTRs are often conserved, the 3' UTR length varies widely between different organisms and correlates with genome size. For example, the median 3' UTR length is ~ 60 nt in unicellular organisms such as bacteria and archaea, and in yeast it is only slightly higher at 150 nt (Mayr, 2017). In multicellular organisms, however, the 3' UTR length increases with the organismal complexity, *e.g.*, worms have a median 3' UTR length of 140 nt while human 3' UTRs are on average 1,200 nt long (Mayr, 2017).

As already mentioned, proliferative and differentiating cells express different 3' UTR isoforms which impact stability, translation and/or intracellular localization of mRNAs. While stem and cancer cells use short 3' UTRs, differentiated cells preferentially express long 3' UTRs (Brumbaugh et al., 2018; Lackford et al., 2014; Mayr and Bartel, 2009; Sandberg et al., 2008; Shepard et al., 2011; Sood et al., 2006). So far, different mechanisms have been proposed to regulate APA. A well-established one is the regulation by binding of CPSF-4 (cleavage and polyadenylation specificity factor 4) and Fip1 (factor interacting with poly(A) polymerase 1) to the polyadenylation signal (PAS) (Lackford et al., 2014). CPSF-4 and Fip1 are responsible for the recognition of the PAS and interact with the poly(A) polymerase and other factors

(Kaufmann et al., 2004). CPSF-4- and Fip1-mediated APA regulation depends on Fip1-RNA interaction and the distance between the proximal and distal PAS (Fig. 1) (Lackford et al., 2014). It is thought that the distal PAS generally resembles a canonical PAS (AAUAAA), and is therefore preferentially used (Lackford et al., 2014; West et al., 2018). For genes whose distal and proximal PAS are far from each other, as it can occur in embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs), lower levels of CPSF-4/Fip1 promote the recognition of the stronger, distal PAS while higher levels result in the binding of the weaker, proximal PAS (Fig. 1A) (Lackford et al., 2014). For genes whose distal and proximal PAS are close to each other, a direct competition for binding of CPSF-4/Fip1 and other mRNA 3' processing factors between those PASs exists: At higher levels of CPSF-4/Fip1, binding to the region between the two PASs can result in the inhibition of CstF (cleavage stimulation factor, involved in PAS recognition) binding and, thus, may block the recognition of the proximal PAS (Fig. 1B). In contrast, lower levels of CPSF-4/Fip1 allows binding by CstF and, thus, a recognition of the proximal PAS (Fig. 1B) (Lackford et al., 2014). This model of APA regulation provides one example of how the usage of distal and proximal PAS can be regulated and how it can influence the ESC and iPSC state. However, it remains to be investigated if this model holds true for other systems, especially in systems where the distance between the distal and proximal PAS is always marginal such as in worms.

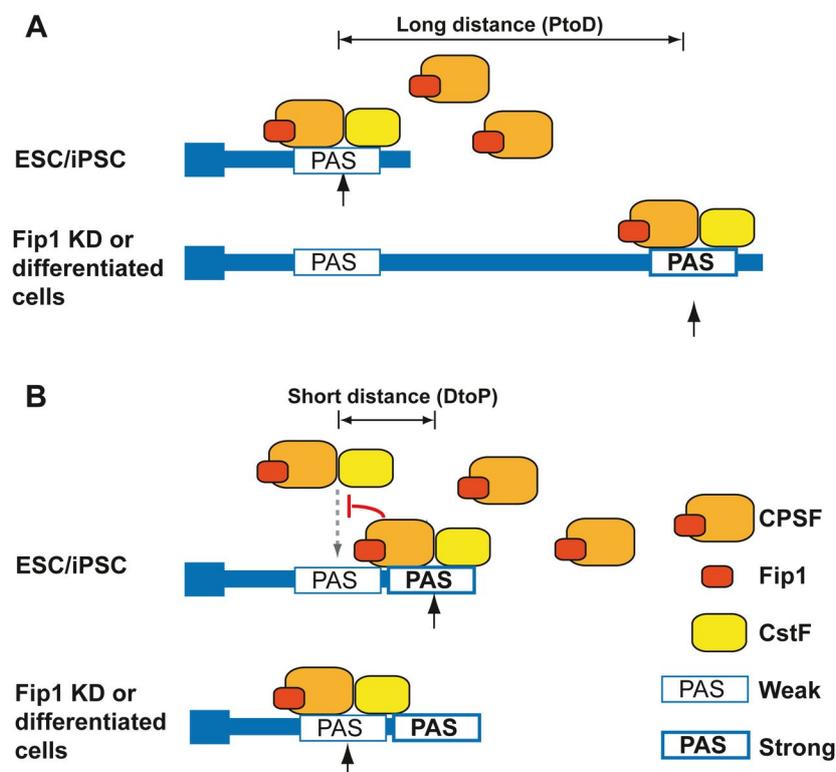


Figure 1. A model for regulation of differential 3' UTR usage. (A) At long distances between distal and proximal PAS, lower levels of CPSF-4/Fip1 result in the recognition of the stronger, distal PAS while higher levels result in the usage of the weaker, proximal PAS. **(B)** At short distances, higher CPSF-4/Fip1 levels and binding to the region between distal and proximal PAS can block the binding by CstF and, thus, inhibit the weaker, proximal PAS while

at lower levels CstF binding can occur allowing recognition of the proximal PAS by CPSF-4/Fip1. ESC, embryonic stem cell; iPSC, induced pluripotent stem cell; PAS, polyadenylation signal; CstF, cleavage factor I. (Lackford et al., 2014)

3' UTR functions are predominantly effected by RBPs that bind specific RNA sequence motifs within them. However, most of these binding elements for RBPs within 3' UTRs are not known due to limitations of experimental approaches that enable identification of these motifs. In addition to RBPs, a class of small RNAs, termed microRNAs (miRNAs) can bind to 3' UTRs thereby posttranscriptionally regulating the stability and translation of the corresponding mRNAs (reviewed in Bartel, 2018; Jens and Rajewsky, 2015). MiRNAs are ~ 22 nt non-coding RNAs which are processed in multiple steps from stem-loops of longer RNA transcripts via Drosha in the nucleus and later on via Dicer in the cytoplasm. After processing, the mature miRNA is loaded into an Argonaute (AGO) protein to form the silencing complex. Binding of the silencing complex to a 3' UTR results mainly in translational inhibition and/or degradation of the corresponding mRNA target. However, a recent study by Dallaire and colleagues demonstrated that miRNAs can also have a stabilizing effect on their targets such as the miR-35 family in the *C. elegans* germline (Dallaire et al., 2018). The number of miRNA genes in a certain organism varies with the complexity of the corresponding organism. So far, 147 canonical miRNAs have been identified in *C. elegans*, 164 in *Drosophila* while 475 have been found in mouse and even 519 in humans (Bartel, 2018). Nevertheless, and despite the large number of identified miRNAs, the cellular and physiological function remains still poorly understood for most of them. A previous study by Stoeckius and colleagues discovered that small RNA expression is highly dynamic during *C. elegans* embryogenesis (Stoeckius et al., 2009). Furthermore, an important function that was suggested is the regulation of the proliferation and differentiation balance in the *C. elegans* germline (Bukhari et al., 2012; Ding et al., 2008). In the study of Bukhari and colleagues, it was shown that loss-of-function of *alg-1* and *alg-2*, two miRNA-specific Argonaute proteins in *C. elegans*, causes a reduced mitotic region and less proliferative cells (Bukhari et al., 2012). In contrast, in humans the expression of miRNAs is reduced in tumours compared to healthy tissue, suggesting that a loss of miRNAs impedes differentiation and promotes proliferation (Lu et al., 2005). It remains to be investigated how certain miRNAs regulate the proliferation and differentiation balance in dependence of their spatiotemporal expression in a tissue or a cell.

1.3 The *Caenorhabditis elegans* germline as a powerful *in vivo* model

Caenorhabditis elegans is a tiny, free-living nematode found worldwide. Adult worms are ~ 1 mm long, whereas newly-hatched larvae are 0.25 mm long (reviewed in Corsi et al., 2015). *C. elegans* is a transparent organism, *i.e.*, subcellular components are easily visualized using microscopy. The life cycle takes about 3 days at 25 °C from egg to egg-laying adult via four larval stages (L1-L4), although *C. elegans* can grow at a broad temperature range (from 16 °C

to 25 °C) (reviewed in Corsi et al., 2015). The nematode exists predominantly as a self-fertilizing hermaphrodite and males appear at a frequency of 0,1 % - 0.2 %, only (reviewed in Corsi et al., 2015). These features and the advantage of easy cultivation, as this multicellular organism needs only a bacterial food source, makes *C. elegans* a powerful model for eukaryotic *in vivo* genetic studies. With an invariant number of somatic cells in the hermaphrodite, *i.e.*, 959 cells, one has tracked each cell fate between fertilization and adulthood to generate an entire cell lineage (Sulston et al., 1983). In addition, a complete genome sequence is available and many discoveries in *C. elegans* have relevance for the study of human health and disease as 60 % - 80 % of human genes have orthologs in the *C. elegans* genome (reviewed in Corsi et al., 2015).

One exceptionally interesting and insufficiently-studied tissue of *C. elegans* is its gonad (Fig. 2). Located in the centre of the body alongside the intestine, the gonad of hermaphrodites consists of an anterior and a posterior gonad arm, two mirror-image U-shaped tubes (Corsi et al., 2015). These two gonad arms harbour the germline where the oocytes and sperm develop (Corsi et al., 2015). The germline in each gonad arm is divided into different zones. Close to the germline niche, the distal tip cell (DTC), proliferative germ cells form a syncytial tissue (Hirsh et al., 1976). At a defined distance from the DTC, germ cells start differentiation by leaving the mitotic cell cycle and entering meiosis (Hirsh et al., 1976). This switch from proliferation to differentiation is called the mitosis-to-meiosis transition. Thus, the *C. elegans* germline offers an excellent *in vivo* model for studying the cell proliferation and differentiation balance in a spatiotemporal manner. Each gonad arm comprises at least 1,000 germ cells, of which the first 40 develop into 160 sperm cells during L4 stage and the other germ cells develop into oocytes during adulthood (Hubbard and Greenstein, 2000).

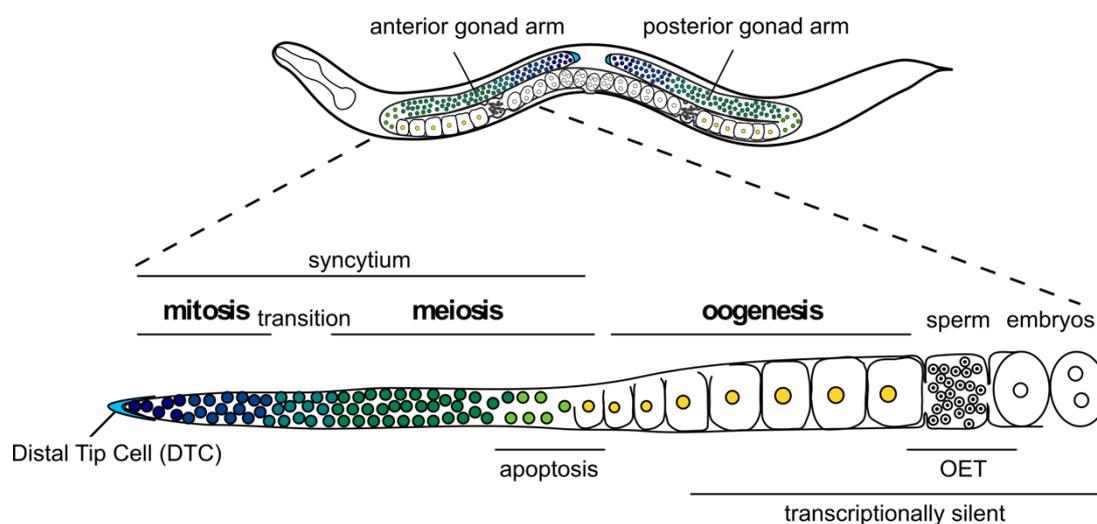


Figure 2. Schematic overview of the *Caenorhabditis elegans* germline. Germ cells form a syncytial tissue and are arranged in a distal-to-proximal organization that reflects successive developmental germ cell stages (mitosis, meiosis, oogenesis and fertilization). Close to the germline niche, the distal tip cell (DTC), proliferative cells are located while more proximally, at a defined distance from the DTC, cells differentiate to oocytes and sperm (Corsi

et al., 2015). Most of the germ cells undergo apoptosis around the loop region. Bulk of RNA synthesis takes place during pachytene stage of meiosis whereas no transcription occurs from late stage oocytes up to 4-cell-stage embryos (Evsikov et al., 2006; Stoeckius et al., 2014). OET: Oocyte-to-embryo transition. (Diag et al., 2018)

As part of the intrinsic oogenesis program, many early germ cells undergo programmed cell death, *i.e.*, apoptosis (Gartner et al., 2008). Only few germ cells differentiate to become oocytes.

On the transcriptome level, bulk synthesis occurs in the pachytene stage, the late stage of meiosis I. From the late stage oocyte (diakinesis) over the oocyte-to-embryo transition (OET) to the 4-cell stage embryo, cells are transcriptionally silent, thus, depending solely on maternal mRNAs that are transcribed earlier in the germline (Evsikov et al., 2006; Stoeckius et al., 2014). Essential developmental changes determine the OET, *i.e.*, fertilization of oocytes by sperm gives rise to an embryonic genome and later fully-developed organism. This process, called “genetic reprogramming”, remains poorly understood (Evsikov et al., 2006). However, post-transcriptional regulation, which is known to control gene expression in early embryogenesis, has been suggested to also play a major role during OET (Stoeckius et al., 2014).

Recent studies focused mainly on the identification of proteins that regulate the switch between cell proliferation and differentiation in the germline. Amongst these, GLP-1, GLD-1 and GLD-2 have been identified as key players in regulating the mitosis-to-meiosis transition (Hansen et al., 2004a, 2004b; Hubbard, 2007; Schmid et al., 2009). GLP-1 is a member of the LIN12/Notch family of receptors that promotes cell proliferation and inhibits meiotic entry by binding to its ligand LAG-2 (expressed in the DTC) and inhibiting the accumulation of GLD-1 in the very distal part of the gonad (Hansen et al., 2004a; Hubbard, 2007). Binding to its ligand results in the cleavage of the intracellular part of GLP-1, following a translocation to the nucleus where it binds LAG-1 (Hansen et al., 2004a; Kadyk and Kimble, 1998). The newly-formed complex may cause the transcription of proliferation-promoting and/or meiosis-inhibiting genes. At a certain distance from the DTC, the proliferation signal and level of GLP-1 decreases allowing the accumulation of GLD-1 and GLD-2. GLD-1 is an RBP which promotes meiotic entry and inhibits mitosis/proliferation by binding and inhibiting the translation of multiple mitosis-promoting mRNAs (Brenner and Schedl, 2016; Jungkamp et al., 2011; Lee and Schedl, 2010). GLD-2 is a cytoplasmic poly(A) polymerase (cytoPAP) which promotes the translation of meiotic entry genes by polyadenylation of these transcripts (Nousch et al., 2017; reviewed in Schmid et al., 2009). Intensive studies on factors that regulate the mitosis-to-meiosis transition have added more key regulators such as GLD-3 and NOS-3. However, most of the studies focused only on proteins and due to limited techniques, less is known about the spatiotemporal RNA expression and how it regulates the proliferation-to-differentiation switch. Previous studies already suggested that small RNAs such as miRNAs can control proliferation and differentiation in *C. elegans* and that 3'UTRs and not promoters are the main determinants of transcript levels in the *C. elegans* germline (Bukhari et al., 2012; Ding et al., 2008; Merritt et

al., 2008). Recent studies also indicated that miR-58 and miR-35, a highly-expressed miRNA specific to the germline, are involved in regulation of apoptosis during *C. elegans* development (Miska et al., 2007; Sherrard et al., 2017). Thus, miRNAs may also participate in regulation of the programmed cell death during oogenesis. More small RNA classes, such as siRNAs (small interfering RNAs) and piRNAs (PIWI-interacting RNAs) are important for the germline viability as these two classes are responsible for transposon silencing and the defence against foreign DNA (Ruby et al., 2006; Weick and Miska, 2014).

It is of great importance to determine the spatiotemporal localization of RNAs in the germline in order to gain insights into their local function and how they influence the germline viability and integrity. Resolving the spatiotemporal RNA architecture of the germline will help to understand what controls the proliferation and differentiation balance and it might be a crucial step to understand what happens during dysregulation in a disease state such as cancer. However, the low RNA content of the *C. elegans* germline, the fact that germ cells are connected through a syncytium (not suitable for single-cell techniques), and the lack of sequencing protocols for low input materials have, thus far, impeded, a system-wide, spatiotemporally-resolved identification of mRNA and small RNA expression during germ cell proliferation and differentiation.

1.4 Cryo-cut method allows spatiotemporal reconstruction of gene expression

Different techniques have been developed to identify the spatiotemporal patterning of transcripts in an organ, cell, or even subcellular compartment. Most of these are microscopy-based techniques such as *in situ* hybridization (ISH) or immunohistochemistry. However, these techniques are limited to analysis of a handful of genes at the same time. Thus, a global spatiotemporal transcriptome is difficult to obtain with microscopy-based techniques, only. In contrast, single-cell RNA sequencing allows the identification of the global transcriptome but lacks spatiotemporal resolution. It is obvious that a combination of imaging and single-cell RNA sequencing would be necessary to determine genome-wide spatiotemporal expression. Previous studies combined ISH with RNA sequencing opening the possibility to investigate the spatial transcriptome on a genome-wide scale (Ke et al., 2013; Lee et al., 2014). However, this technique has still a low detection efficiency and investigation of intact tissue is not yet possible. A further attempt to visualize and quantify RNA at the same time was made by Ståhl and colleagues, who combined histology with reverse transcription primers containing unique positional barcodes (Ståhl et al., 2016). These molecular barcodes maintain the two-dimensional positional information of the transcript from tissue sections. Although this technique captures the global transcriptome and is highly reproducible and sensitive, the design of barcoded arrays remains very expensive and laborious. To circumvent the listed limitations above, Junker and colleagues developed a cryo-cut based sequencing technique,

called tomo-seq, which allows the detection of genome-wide spatial gene expression even in 3D (Junker et al., 2014). As a model system they used the zebrafish embryo and sectioned it into 50-100 thin slices (Fig. 3).

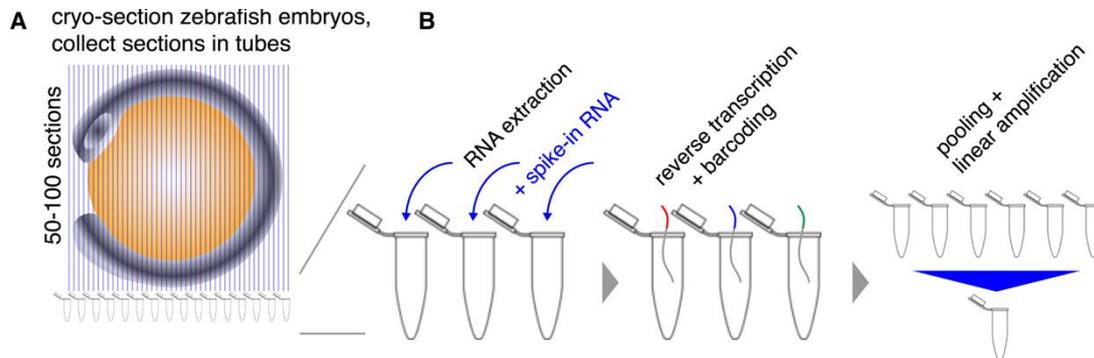


Figure 3. Schematic overview of the tomo-seq approach. (A) A zebrafish embryo (grey) containing the yolk sac (orange) was cryo-sectioned into 50-100 slices collecting each slice separately in a tube. (B) Each slice was subjected to the CEL-seq protocol (Hashimshony et al., 2012). RNA was extracted from each slice, followed by reverse transcription, barcoding and pooling for linear amplification via *in vitro* transcription. (Junker et al., 2014)

Each slice was subjected to the CEL-seq protocol which allowed for barcoding of each section and, thus, finally pooling of all slices in a single tube (Hashimshony et al., 2012). Using the CEL-seq protocol, Junker and colleagues also minimized the amplification bias associated with PCR-based methods by taking advantage of an *in vitro* transcription for linear amplification. Furthermore, the method is suitable for very low input material. Tomo-seq demonstrated high reproducibility and sensitivity showing high accuracy in detecting the spatial transcriptome of the zebrafish embryo on a genome-wide scale (Junker et al., 2014). It is a comparatively cheap method which can be used for any intact tissue. By cutting the tissue in different orientations and by using mathematical image reconstruction, Junker and co-workers reconstructed the spatial gene expression in 3D. However, this method is so far limited to the detection of polyadenylated transcripts as it relies on oligo(dT) primers for transcript capture. Thus, further optimization will be needed in order to investigate the global RNA profile of a tissue in a spatiotemporal manner.

1.5 Aims of the thesis

This study aimed at investigating the spatiotemporal RNA architecture of the *C. elegans* hermaphrodite germline during germ cell proliferation and differentiation. For this purpose, I adapted the tomo-seq technique for the *C. elegans* gonad (Fig. 4). With this, I wanted to (1) quantify the number of genes that display spatiotemporal patterning across the gonad, and (2) determine in which compartment of the germline these genes are expressed. Following this approach, and by comparing to mutant germlines, we hoped to determine which genes are essential for the balance between proliferation and differentiation and, thus, to characterize

their molecular function. As previous studies showed that APA and 3' UTRs are essential for proliferation and differentiation, we resolved our sequencing data to a transcript-isoform level in order to dissect the isoform distribution during progression of cells through proliferation and differentiation in the germline. Going one step further, I expanded the tomo-seq technique for the detection of other RNA classes such as small RNAs including miRNAs, siRNAs and piRNAs as these RNAs play a crucial role in posttranscriptional regulation of gene expression in the germline. Finally, we created a 3D physical gonad model that integrates our data set providing a user-friendly interface and a “universal” coordinate system of our data.

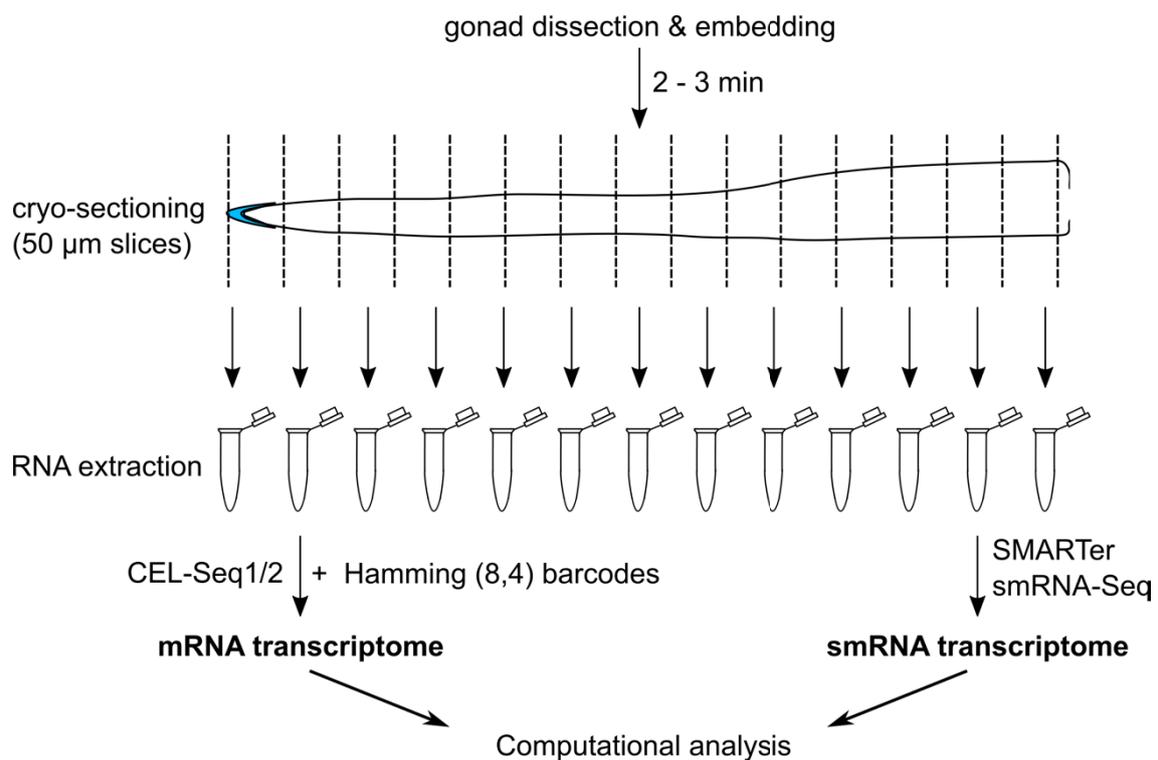


Figure 4. Schematic overview of the tomo-seq technique for the *C. elegans* gonad. Gonads of young adult *C. elegans* hermaphrodites were dissected and embedded in a tissue freezing medium. Following rapid freezing at -80 °C, gonads were cryo-sectioned at 50 µm resolution. After RNA extraction, each slice was subjected either to the CEL-seq1/2 protocol for detection of the mRNA transcriptome or to the SMARTer smRNA-Seq protocol for the investigation of the small RNA transcriptome. Finally, the spatiotemporal RNA architecture of the germline was analysed computationally. (Diag et al., 2018)

2 Materials and Methods

2.1 Chemicals

Acetic anhydride, $\geq 99\%$	Sigma-Aldrich, Germany
Affi-Gel Blue Gel	Bio-Rad, USA
Agar-Agar Kobe I, pulv. 500 g	Carl Roth, Germany
Agarose NEEO Ultra-Quality	Carl Roth, Germany
Ammonium thiocyanate	Carl Roth, Germany
Bactotryptone	Thermo Fisher Scientific, USA
Blocking Solution	Roche, Swiss
CaCl ₂ , $\geq 96\%$	Carl Roth, Germany
Chloroform, 99 %	Sigma-Aldrich, Germany
Cholesterol	Sigma-Aldrich, Germany
Cryomatrix	Thermo Fisher Scientific, USA
Denhardt's solution	Thermo Fisher Scientific, USA
Dextran sulphate	Sigma-Aldrich, Germany
Digoxigenin (DIG) RNA labelling mix	Roche, Swiss
Dithiothreitol (DTT)	Invitrogen, USA
dNTP Mix	Fermentas, USA
EDTA	VWR, USA
Ethanol, $> 99,5\%$	Carl Roth, Germany
Ethidium bromide	Carl Roth, Germany
External RNA Controls Consortium (ERCC) Spike-In mix I	Thermo Fisher Scientific, USA
FastAP Thermosensitive Alkaline Phosphatase, 10 x	Thermo Fisher Scientific, USA
Formamide	Merck, Germany
GeneRuler 100 bp Plus DNA Ladder	Thermo Fisher Scientific, USA
GeneRuler 100 kb Plus DNA Ladder	Thermo Fisher Scientific, USA
Glycerine, $> 99,5\%$	Carl Roth, Germany
Guanidinium thiocyanate	Carl Roth, Germany
GycoBlue Coprecipitant	Thermo Fisher Scientific, USA
HCl, 25 %	Carl Roth, Germany
HEPES, $\geq 99,5\%$	Carl Roth, Germany
High fidelity (HF) Phusion buffer, 5 x	New England Biolabs, USA
Isopropanol, $\geq 99,8\%$	Carl Roth, Germany
KCl, $> 99,5\%$	Carl Roth, Germany
KH ₂ PO ₄ , $\geq 98\%$	Carl Roth, Germany
Levamisol Hydrochloride Vetranal	Sigma-Aldrich, Germany

Maleic acid, > 99 %	Sigma-Aldrich, Germany
MgCl ₂ Hexahydrate, ≥ 98 %	Carl Roth, Germany
MgSO ₄ , ≥ 99 %	Carl Roth, Germany
NaCl, > 99,8 %	Carl Roth, Germany
Normal goat serum	Abcam, England
Paraformaldehyde (PFA)	Carl Roth, Germany
Phosphate buffered saline (PBS)	Genaxxon Bioscience, Germany
PIPES, Free Acid, Ultrol Grade, 99 %	Calbiochem, USA and Canada
Polynucleotide kinase (PNK) buffer, 10 x	Thermo Fisher Scientific, USA
Prolong gold	Invitrogen, USA
RiboLock RNase Inhibitor	Thermo Fisher Scientific, USA
Roti-Aqua-Phenol	Carl Roth, Germany
Saline Sodium Citrate (SSC), 20 x	Sigma-Aldrich, Germany
Sodium acetate, water free	Carl Roth, Germany
T7 transcription buffer, 10 x	Roche, Swiss
Triethanolamine, ≥ 99 %	Sigma-Aldrich, Germany
Tris, ≥ 99,9 %	Carl Roth, Germany
Triton X 100	Carl Roth, Germany
tRNA from yeast	Genaxxon Bioscience, Germany
Tryptone/Peptone from casein 1 kg	Carl Roth, Germany
Tween 20	Carl Roth, Germany
Yeast extract	Sigma-Aldrich, Germany

2.2 Enzymes

Anti-Digoxigenin (DIG)-Alkaline Phosphatase (AP)	Roche, Swiss
DNase I	Promega, USA
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific, USA
Maxima H minus Reverse Transcriptase	Thermo Fisher Scientific, USA
NBT-BCIP solution	Roche, Swiss
Phusion High-Fidelity DNA Polymerase	New England Biolabs, USA
Polynucleotide kinase (PNK)	Thermo Fisher Scientific, USA
Proteinase K	Roche, Swiss
RNase H	Biozym Scientific, Germany
Superscript III Reverse Transcriptase	Invitrogen, USA
T4 DNA ligase, truncated	New England Biolabs, USA
T7 RNA polymerase	Roche, Swiss
TURBO DNase	Thermo Fisher Scientific, USA

2.3 Commercially available assays

Agencourt AMPure XP	Beckman Coulter, USA
Agencourt RNAClean XP	Beckman Coulter, USA
Agilent RNA 6000 Nano Kit	Agilent Technologies, USA
Agilent RNA 6000 Pico Kit	Agilent Technologies, USA
Custom TaqMan MicroRNA assay	Thermo Fisher Scientific, USA
High sensitivity DNA Kit	Agilent Technologies, USA
MessageAmp II aRNA Amplification Kit	Thermo Fisher Scientific, USA
NucleoSpin Gel and PCR Clean-Up Kit	Macherey-Nagel, Germany
SMARTer smRNA-Seq Kit	Clontech Laboratories, USA
TaqMan Universal Master Mix II, no UNG-1	Thermo Fisher Scientific, USA
TruSeq small RNA Kit	Illumina, USA
TruSeq-stranded mRNA LT Kit	Illumina, USA

2.4 Equipment

2.4.1 Consumables

Cover Slips 24 x 24 mm	Carl Roth, Germany
DNA/RNA LoBind Tubes 1.5 ml	Eppendorf, Germany
Falcon 25 ml High Clarity PP Centrifuge Tube	Corning Life Sciences, USA
Falcon 50 ml High Clarity PP Centrifuge Tube	Corning Life Sciences, USA
Multiply- μ Strip 0.2 ml	Sarstedt, Germany
Petri dishes	Sarstedt, Germany
Polysine slides	Thermo Fisher Scientific, USA
Tissue-Tek Cryomold	Sakura, Japan

2.4.2 Devices

Agilent 2100 Bioanalyzer	Agilent Technologies, USA
BluePippin	Sage Science, USA
Centrifuge 5430 R/5415 R	Eppendorf, Germany
E-Box Gel Documentation Imaging	Vilber, Germany
Fluorescence microscope BZ-X	Keyence, Japan
Gel chamber	Peqlab, Germany
GeneTouch Thermal Cycler	Biozym Scientific, Germany
HiSeq 2500	Illumina, USA
Light microscope	Leica, Germany
Mini centrifuge	Benchmark Scientific, USA
NanoDrop ND-100 Spectrophotometer	Peqlab, Germany

NextSeq 500	Illumina, USA
PowerPac Basic Power Supply	Bio-Rad, USA
StepOnePlus Real-Time PCR System	Thermo Fisher Scientific, USA
Thermomixer compact	Eppendorf, Germany
VortexGenie 2	Scientific Industries, USA
Zeiss AX10 with AxioCam	Zeiss, Germany

2.5 Synthetic oligonucleotides

Following oligonucleotides (Eurofins, Luxembourg) were used to generate DIG-labelled RNA probes for *in situ* hybridization (ISH):

gene	forward (5' to 3')	reverse (5' to 3')
<i>cpg-1</i>	TGCATTCCTTGTTGCATCGG	GATCCTAATACGACTCACTATAGGGAGAA CATCCTCCAATCGCGT
<i>iff-1</i>	GCTCGCTCATGGATCCAGAA	GATCCTAATACGACTCACTATAGGGCAGG CAGCAACGACTTGAAC
<i>rpl-17</i>	CGGAAAACAGCACCAAGTCG	GATCCTAATACGACTCACTATAGGGGATG TGGCATGGGGAAGACA
DY3.8	GCATGGTTACGTCCGGATCT	GATCCTAATACGACTCACTATAGGGACCT TGAGCAATGATGGCCA
<i>nos-2</i>	GCCGTCGCCAACATTTGATT	GATCCTAATACGACTCACTATAGGGGATT CCTCGCGAGCTCTTCA
<i>perm-4</i>	CCCATCCGTCCCAGCTAATC	GATCCTAATACGACTCACTATAGGGTGAG CAGCTGGAGAGGATCT
<i>perm-2</i>	CCTGAGAAGCAATCCGACGT	GATCCTAATACGACTCACTATAGGGTCGT CGTCCTTCTTGTGGTG
<i>pos-1</i>	ACAACGATTTCTGAGCGGT	GATCCTAATACGACTCACTATAGGGACAG AGCCGTCTTGAATGCA
<i>pie-1</i>	ACAACCACGGAAGCTCGAAT	GATCCTAATACGACTCACTATAGGGGTTG GTGGTGGCTGAGAGAA
<i>cey-2</i>	CAGCCATTACCACCGGACTT	GATCCTAATACGACTCACTATAGGGTCTT GCGTTGCTTTCTTGGC
<i>gld-2</i>	ACTTCTGGCGGTCTCAATGG	GATCCTAATACGACTCACTATAGGGTGTT GCTGCTGATTGTTGCC

DIG-labelled locked nucleic acid (LNA) probes for small RNA ISH were purchased from Exiqon (now Qiagen, Netherlands). The sequences are listed below:

miRNA	DIG-labelled LNA probe (5' to 3')
250-3p	AGTTGACTGTGATT
35-3p	TGCTAGTTTCCACCCGGT
40-3p	AGCTGATGTACACCCGGT
61-3p	ATGAGTAACGGTTCTAGTCA
nov-1-3p	GGATATTCAATTTTTTTTTC

TaqMan miRNA probes were purchased from Thermo Fisher Scientific (USA).

miRNA	TaqMan target sequence
nov-63-3p	GUCACCUUGAGGUCAGGCAGGC
nov-72-3p	CCUGGGAAUCCUGGAUGU
nov-82-5p	AGCGGAAAUUGCUACAAUAUUC
let-7-5p	commercial (supplied by manufacturer)
miR-1-5p	commercial (supplied by manufacturer)
miR-3p	commercial (supplied by manufacturer)

Oligonucleotides used for nested PCR were purchased from Eurofins (Luxembourg) or Biotex (Germany). Sequences are listed below. The primer-sequence used for cDNA synthesis was GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTTVN (5' to 3').

gene	first nested PCR (5' to 3')	second nested PCR (5' to 3')
C49F5.3	CTACCCGAGCAAGTGGATCT	GACCACGCATCCATCCTTTC
<i>uba-2</i>	TTGCGACGATTTCCAACAGG	AAATGGAAATGAGGCGGTGG
<i>ztf-20</i>	ATTTGTGCGGAATGTTCTGGG	CGTCAAGTCCGGTGTTTTGA
F22D6.2	TAGTCCGAAGCGCAGTGTTT	GATTTCCGTTCTGCGGATC
K09H9.2	ATCGGTGATGTGAAGCTGGA	CCTGCCTGAGATCAACCTGA
AP	GACTCGAGTCGACATCG	-
MAP	-	CGACATCGATTTTTTTTTTTT

Eight of the 3' UTR extensions were validated with conventional PCR instead of nested PCR. Oligonucleotide sequences used for conventional PCR were purchased either from Eurofins (Luxembourg) or Biotex (Germany) and are listed below.

gene	forward (5' to 3')	reverse (5' to 3')
<i>exos-4.2</i>	CAATACAGCCTTCAGACCGC	CGGAGAAGGCAAAGGGAAG
<i>nhl-2</i>	GAGAACCATCACAGACCCCA	ACATTTGAAAGCGACGGGAC

<i>rec-8</i>	TCAATGAATTCACCTGCGGATCA	CACCCGGGTCATATTATTCACG
<i>par-4</i>	TGGAGCGAATGAGAACAGGA	AAATCGAGAACTGCCGGAA
<i>spin-4</i>	ACAGCAACACAGACAACGAC	CGAATATTTCTCTACCAGCCGC
Y92H12BR.7	GGAATCCCAAATGCCGAACC	TTTTGGCACTTATTCGGGTAAAA
W05F2.7	TCTACGAACAGCACGCAAAC	GCAAGAACGAAAGAGCAACC
<i>lmd-2</i>	CGTCGTCATCAGCAACAGAG	GACCCAGTGCTTCCAAAGG

2.6 Buffers, solutions and growth media

Nematode growth medium (NGM)

17.5 g/l agar, 3 g/l NaCl, 2.5 g/l peptone and 0.005 g/l cholesterol; autoclaved and cooled down to 55 °C while stirring. Following, 1 mM MgSO₄, 1 mM CaCl₂, 25 mM and K₂PO₄ (pH 6) were added under the sterile hood. NGM solution was dispensed into petri dishes. The petri dishes were stored at 4 °C for several weeks.

LB growth medium for bacteria

10 g bacto tryptone, 5 g yeast extract, 10 g NaCl, adjusted to pH 7.5 and autoclaved.

OP50-seeded NGM plates

NGM plates were seeded with OP50 bacteria grown in LB medium.

Trizol

38 % phenol, 800 mM guanidinium thiocyanate, 400 mM ammonium thiocyanate, 100 mM sodium acetate (pH 5.0) and 5 % glycerine.

Sperm salt buffer

100 mM PIPES pH 7.0; 90 mM NaCl; 50 mM KCl; 40 mM CaCl₂; 20 mM KH₂PO₄

PBS-T

PBS containing 0.1 % Tween 20

TEA buffer

1.3 % triethanolamine; 0.25 % acetic anhydride

Prehybridization buffer

10 mM HEPES pH 7.5; 600 mM NaCl; 50 mM DTT; 1 mM EDTA; 1 x Denhardt's solution; 100 µg/ml tRNA; 50 % formamide

Hybridization buffer

Prehybridization buffer containing 10 % dextran sulphate

Posthybridization buffer

5 x SSC containing 50 % formamide

Maleic acid buffer

11.6 g/l maleic acid; 9,76 g/l NaCl; 0.1 % Triton X, pH 7.5

TMN buffer

0.1 M Tris-HCl pH 9.5; 0.1 M NaCl; 50 mM MgCl₂; 1 % Tween 20

TEA buffer for small RNA *in situ* hybridization

1.3 % triethanolamine; 0.6 N HCl; 0.27 % acetic anhydride

Blocking solution for small RNA *in situ* hybridization

PBS-T containing 5 % normal goat serum

2.7 Strains

(Diag et al., 2018)

All *Caenorhabditis elegans* strains (Caenorhabditis Genetics Center, USA) were cultured by standard techniques (Brenner, 1974). Worms were maintained at 16 °C on *Escherichia coli* OP50-seeded NGM plates. The following strains were used in this study: N2 Bristol wild type, *gld-2(q497) gld-1(q485)/hT2 [bli-4(e937) let-?(q782) qls48] (I;III)*, *gld-2(q497)/dpy-5(e61 unc-13(e-51) I* and *glp-1(ar202) III*. The strains were provided by the Caenorhabditis Genetics Center (CGC). For all experiments, young adult or L4 staged hermaphrodites were used.

2.8 Experimental details

(Diag et al., 2018)

2.8.1 Embedding and cryo-sectioning

Gonads of wild type and mutants were dissected according to Francis and Nayak (Schedl lab) by cutting off the head or the tail with a syringe needle to extrude the anterior or posterior gonad arm, respectively. The gonad, still attached to the worm body, was transferred to a specimen mold (Tissue-Tek[®] cryomold[®]) filled with tissue freezing medium. This medium is very viscos, facilitating the stretching of the gonad and the separation from the worm body. Once the gonad was stretched, distal tip end and proximal end (at the end of oogenesis) were marked with AffiGel[®] blue beads (Bio-Rad). Next, the specimen mold was rapidly frozen at

- 80 °C for 1 min and subsequently fixed in the cryotome to cut the gonad into slices of desired resolution (here 50 µm). Each slice of the gonad was collected in an individual LoBind Eppendorf® tube and immediately transferred to dry ice. RNA extraction of each slice was performed according Junker et al. (2014) with minor modifications. Briefly, 500 µl of self-made trizol were added to each slice including 0.5 µl GlycoBlue and 2 µl ERCC Spike-In mix (1:10,000 - 1:50,000). Samples were mixed thoroughly and incubated for 5 min at room temperature. Subsequently, 100 µl of chloroform were added to each sample, mixed well, incubated for 5 min at room temperature and centrifuged at 12,000 g for 15 min at 4 °C. Following, the aqueous phase (~ 250 - 300 µl) was carefully transferred to a new LoBind Eppendorf® tube and 250 - 300 µl of isopropanol were added. Samples were mixed thoroughly and incubated over night at -20 °C. Afterwards, samples were centrifuged for 10 min at 12,000 g and at 4 °C, supernatant was removed and RNA pellet (should appear blue) was washed with 75 % ethanol. Following the last centrifugation step at 7,500 g for 5 min and at 4 °C, supernatant was removed and pellet was either resuspended in 1.2 µl barcoded oligo(dT) primer in order to proceed with mRNA library preparation or in 7 µl nuclease free water for small RNA library preparation. All experiments were performed in biological and technical triplicates for wild type and replicates for mutants for each gonad arm, *i.e.*, anterior and posterior gonad arm.

2.8.2 mRNA library preparation

Reverse transcription and *in vitro* transcription (IVT) were performed with the Ambion™ MessageAmp™ II kit according to the CEL-seq method (Hashimshony et al., 2012) and the tomo-seq method (Junker et al., 2014), except that all purification steps were performed using Agencourt® AMPure® or RNAClean® XP beads according to CEL-seq2 minimizing loss of material (Hashimshony et al., 2016). In brief, after resuspending each slice in 1.2 µl barcoded oligo(dT) primer, first strand synthesis was performed for 2 h at 42 °C. Following, slices were pooled and second strand synthesis was performed for 2 h at 16 °C. The resulting cDNA was purified with Agencourt® AMPure® XP beads according manufacturer's instructions. Subsequently, IVT was performed for 13 h at 37 °C to amplify the RNA which afterwards was fragmented and purified with RNAClean® XP beads according manufacturer's instructions. Library preparation was performed using the Illumina TruSeq® small RNA kit following the tomo-seq protocol (Junker et al., 2014). This included phosphatase and polynucleotide kinase (PNK) treatment, purification with RNAClean® XP beads, 3' adapter ligation, reverse transcription, PCR amplification and final purification with Agencourt® AMPure® XP beads. Unlike CEL-seq1/2 and tomo-seq, unanchored oligo(dT) barcodes used in this study were designed according to a Hamming [8,4] code allowing for barcode correction after sequencing (Bystrykh, 2012). Anchored oligo(dT) barcodes according to CEL-seq and tomo-seq

(Hashimshony et al., 2012; Junker et al., 2014) were used for uncut samples (N2_mRNA_uncut_A/P) and the first replicates of cut anterior and posterior gonad arm samples (N2_mRNA_A1/P1). Libraries (with 20 % - 30 % of PhiX spike-in DNA) were sequenced on the NextSeq 500 in a paired end mode. Except for samples N2_mRNA_A1/P1 and N2_mRNA_uncut_A/P, the number of cycles for read 1 was decreased to sequence barcode and UMI only.

2.8.3 Small RNA library preparation

Small RNA libraries were only performed for the wild type N2 strain. Library preparation was performed for each slice separately using the SMARTer smRNA-Seq kit for Illumina from Clontech® according to manufacturer's instruction. Briefly, total RNA was extracted from each slice as described above (see 2.8.1). Next, total RNA was polyadenylated using a synthetic poly(A) polymerase. Subsequently, cDNA was synthesized and amplified according manufacturer's instructions. The PCR product was purified using the NucleoSpin Gel and PCR Clean-Up kit according manufacturer's instructions. Afterwards, libraries were validated with the Agilent 2100 Bioanalyzer using the High Sensitivity DNA kit according manufacturer's instructions. Finally, size selection was performed with the BluePippin system (range: 148 bp - 185 bp) in order to enrich for small RNAs. The small RNA libraries were pooled and sequenced on HiSeq 2500, 1 x 50 cycles.

2.8.4 Poly(A)+-selected library preparation

For the poly(A)+-selected library, several gonads were dissected and pooled. Total RNA extraction of pooled gonads was performed as described above (see 2.8.1). Library preparation was performed with the Illumina TruSeq® stranded mRNA LT kit according to manufacturer's instruction. Sequencing was performed on the NextSeq 500.

2.8.5 Ribosomal RNA depleted total RNA library preparation

For the ribosomal RNA-depleted (ribodepleted) total RNA library several gonads were dissected and pooled. Ribosomal RNA (rRNA) was depleted using an RNase H-based protocol adapted from Adiconis *et al.* (2013). The extracted total RNA from dissected gonads was mixed with the same amount of a DNA oligonucleotide pool consisting of a 50 nt long oligonucleotide mix covering the reverse complement of the entire length of each *C. elegans* rRNA (5S rRNA, 5.8S rRNA, 18S rRNA, 26S rRNA). Following, samples were incubated with 1 U of RNase H for 30 min at 45 °C and purified with RNAClean® XP beads according manufacturer's instructions. The purified RNA was treated with DNase according to the TURBO DNase rigorous treatment procedure (Thermo Fisher Scientific). After the final purification step with RNAClean® XP beads, rRNA-depleted RNA samples were fragmented and library preparation

was performed with Illumina TruSeq[®] stranded mRNA LT kit. Paired end sequencing was performed on the NextSeq 500.

2.8.6 Probe preparation for mRNA in situ hybridization (ISH)

Digoxigenin (DIG)-labelled anti-sense RNA probes were prepared by IVT using a PCR generated DNA template. PCR primers were designed using Primer3 (Koressaar and Remm, 2007; Untergasser et al., 2012) to amplify a 300-500 nt fragment from the cDNA prepared from whole worm samples. The T7 promoter sequence was added to the reverse primer to produce later an antisense probe by IVT. Primer sequences are provided in. PCR fragments were cleaned-up using Agencourt[®] AMPure[®] XP beads and *in vitro* transcription was performed with 0.5 - 1 µg DNA template using the T7 RNA polymerase and a DIG-RNA labelling mix (Roche). The remaining DNA template was digested with DNase I and the RNA probe was precipitated with sodium acetate and ethanol for at least 30 min at -80 °C. After centrifugation the RNA pellet was washed with 75 % ethanol and probe integrity was checked on an agarose gel. The concentration of each RNA probe was adjusted to 50 ng/µl using 10 mM Tris-HCl/formamide solution (1:1).

2.8.7 mRNA ISH

Worms were washed several times in sperm salt buffer only and in the final step in sperm salt buffer containing levamisole. Up to 15 worms were transferred to a poly-L-lysine coated slide containing 8 µl of sperm salt and gonads were dissected according to Francis and Nayak (Schedl lab) with minor modifications. After dissection 8 µl of 4 % paraformaldehyde (PFA) were added to the dissected gonads, a cover slip was put on top and the slide was incubated for 2 min. Following, the slide was incubated on dry ice for at least 20 min and the coverslip was flipped away using a razor blade under the coverslip (freeze and crack method). Slides were immediately immersed in ice-cold 100 % ethanol for 2 min, rehydrated in an ethanol series (90 %, 70 %, 50 %, 20%), following washing with PBS containing 0.2 % Tween 20 for 30 min. Permeabilization of gonads was achieved with proteinase K treatment (1 µg/ml) for 5 min. Slides were washed in PBS-T, fixed for 20 min in 4 % PFA, washed again with PBS-T, incubated in TEA buffer (always prepared fresh), following final washing steps in PBS-T. Slides were prehybridized in prehybridization buffer for 1 h at 50 °C. Slides were hybridized over night at 50 °C in hybridization buffer containing 0.5 - 1 µg/ml denaturated DIG-labelled antisense RNA probe (denaturation at 95 °C for 10 min). Slides were washed at 50 °C for 10 min with following solutions: posthybridization buffer (posthyb); 75 % posthyb buffer + 25 % 2 x SSC, 0.1 % Triton X; 50 % posthyb buffer + 50 % 2 x SSC, 0.1 % Triton X; 25 % posthyb buffer + 75 % 2 x SSC, 0.1 % Triton X; 2 x SSC, 0.1 % Triton X; 0.22 x SSC, 0.1 % Triton X. Following, slides were washed in maleic acid buffer and afterwards incubated in 1 % blocking solution

(Roche) diluted in maleic acid buffer for 1 h. Slides were incubated in Anti-DIG-AP (Roche, 1:2500) over night at 4 °C. After several washes with maleic acid buffer and TMN buffer (always prepared fresh), the signal was developed using NBT/BCIP (diluted in TMN buffer) solution. Time of development depended on the expression of the corresponding RNA and took up to 24 h for very lowly expressed RNAs. The background was removed with dehydration and rehydration in an ethanol series (samples were fixed before in 4 % PFA for 20 min again). For mounting, some μ l of prolong gold (Invitrogen) were dropped on a coverslip and then inverted onto the slide. The edges were sealed with nail polish.

2.8.8 small RNA ISH

Gonad preparation and the prehybridization procedure was the same as for mRNA ISH. The TEA buffer contained additionally 0.06 N HCl and 0.27 % acetic anhydride. For small RNA ISH, DIG-labelled LNA (Locked Nucleic Acid) probes (former: Exiqon, now: Qiagen) were used and the prehybridization and hybridization temperature was set according to manufacturer's instruction (20 - 25 °C below the melting temperature of the LNA probe). LNA probes were denaturated at 95 °C for 1 - 5 min prior hybridization. Prehybridization (without probe) was done for 1 h and hybridization (with 10 - 25 nM of LNA probe) over night. Slides were washed several times with 2 x SSC buffer and with 0.2 x SSC buffer. Following, the slides were washed with PBS-T and incubated for 1 h in blocking solution. Slides were incubated in Anti-DIG-AP (Roche, 1:2,000) over night at 4 °C. After several washes with PBS-T and TMN buffer, signal developing, background removal and mounting was performed according to mRNA ISH.

2.8.9 TaqMan[®] assays

The TaqMan[®] assay was used to validate the novel miRNA predictions. TaqMan[®] probes were designed with the Custom TaqMan[®] Small RNA Assay Design Tool (ThermoFisher). TaqMan[®] assays were performed according manufacturer's instruction for gonad and whole worm samples.

2.8.10 Nested PCR

Nested PCR was performed according to the Cold Spring Harbor Protocols (Sambrook and Russell, 2006). 0.5 - 1 μ g of whole worm and gonad RNA were used as input RNA for the cDNA synthesis using TAP-VN as a primer. For the first nested PCR, 4 μ l of 1:5 diluted cDNA was used. The first PCR was performed with a gene-specific forward primer and AP as a reverse primer. PCR products were purified using Agencourt[®] AMPure[®] XP beads and 10 - 20 ng of purified PCR were used for the second nested PCR. The second PCR was performed with a second gene-specific primer and MAP as a reverse primer. Annealing temperature was calculated using the NEB Tm Calculator (BioLabs). The PCR products from the second PCR

were separated by agarose gel, purified and Sanger-sequenced to confirm the identity of the bands. Nested PCR was used for 3' UTR extension validation. Alternatively, conventional PCR by designing the forward primer in the second last exon (to distinguish from genomic DNA) and the reverse primer in the 3' UTR extension was used for validation (using whole worm RNA only).

2.8.11 Physical gonad model

To be able to assign the relative distal-to-proximal coordinates used for the spatially resolved gene expression profiles, a physical model of the *C. elegans* germline was built using a custom R script. The following assumptions were made for that model: i) Cells are approximately spherical. ii) Germ cells form a single layer tube within the distal part of the gonad arm. iii) The diameter of the gonad is minimal under the constraint of encompassing all germ cells. This enables a direct conversion between the number of cells in a germ cell layer and the diameter of that cell layer (given the size of a single germ cell) using basic geometry:

$$\phi_G(d_L(l)) := \frac{\phi_g}{2 \sin\left(\frac{\pi}{N_{g_L}(l)}\right)} \cdot 2 + \phi_g = \left(1 + \left[\sin\left(\frac{\pi}{N_{g_L}(l)}\right)\right]^{-1}\right) \phi_g$$

were ϕ_g is the diameter of a germ cell, $l \in \mathbb{N}^+$ is the germ cell layer (one-based), $N_{g_L}(l)$ is the number of germ cells in layer l , $d_L(l)$ is the distance of the center of layer l to the distal tip cell (DTC) ($d_L(l) := \left(l - \frac{1}{2}\right) \phi_g$) and $\phi_G(d)$ is the diameter of the gonad arm at distance d from the DTC.

The (modelled constant) diameter of a single germ cell was set to 4.6 μm (Maciejowski et al., 2006). Based on our own measurements and results by Hirsh and colleagues (Hirsh et al., 1976) the total length of a stretched-out gonad arm was defined as 650 μm . At this distance to the distal tip cell (DTC) (*i.e.*, at the proximal end), the gonad must fit a fully mature oocyte, while at the distal-most end only a single germ cell needs to be fit in the gonad arm. To get a rough estimate of the size of a fully matured oocyte, the number of cells per embryo (558) (Wolke et al., 2007) was multiplied with the volume of a single germ cell. Given the equality in diameter of embryonic cells and germ cells and the equality in volume of the mature oocyte and the embryo, this gives a direct estimate for the size of the oocyte. To be able to model the gonad diameter in-between those extreme boundaries, we measured four gonad arms based on microscopic images (Table S1). Using these measurements at discrete points, a spline fit was used to model the radius of the gonad arm as a function of the distance to the DTC. Using this fit, the outline of the stretched-out gonad arm was modeled as a solid of revolution around the distal-to-proximal axis:

$$v_G(d_s, d_e) = \pi \int_{d_s}^{d_e} r_G(d)^2 dd$$

where d_s and d_e denote the distance to the DTC of the start and the end of the interval of interest, respectively, and $v_G(d_s, d_e)$ is the volume of the corresponding part of the gonad arm. Based on the assumptions introduced above, the distal arm was filled with 1,002 germ cells in layers maximizing the number of cells per layer under the constraint given by the corresponding gonad diameter:

$$n_{g_L}(l) := \begin{cases} 0 & \text{if } \phi_L(l) < \phi_g \\ 1 & \text{if } \phi_L(l) < 2\phi_g \\ \left\lfloor \frac{\pi}{\arcsin\left(\frac{\phi_g}{\phi_L(l) - \phi_g}\right)} \right\rfloor & \text{otherwise} \end{cases}$$

where $l \in \left\{1, \left\lfloor \frac{d_B}{w_L} \right\rfloor\right\} \subset \mathbb{N}^+$ ($d_B = 389.3 \mu\text{m}$ representing the distance to the DTC of the bend and $w_L := \phi_g$ the width of a germ cell layer) is the germ cell layer of interest (one-based), $n_{g_L}(l)$ is the number of germ cells in that layer, and $\phi_L(l)$ is the minimal diameter of the gonad in the interval containing germ cell layer l :

$$\phi_L(l) := \phi_{G_L}(l) = \min_{d \in [(l-1)w_L, lw_L] \subset \mathbb{R}} 2r_G(d)$$

where $r_G(d)$ is the radius of the gonad at distance d from the DTC according to the spline model.

The total number of distal germ cells was derived from the total number of distal germ cell layers which was determined by comparing the cumulative number of cells up to each potential layer to the expected number of germ cells $\tilde{N}_g := 1,000$:

$$N_g := n_{g_L}^\Sigma(N_L)$$

with

$$n_{g_L}^\Sigma(l) := \sum_{\lambda=1}^l n_{g_L}(\lambda)$$

and

$$N_L := \arg \min_{l=1}^{\lfloor \frac{d_B}{w_L} \rfloor} |\tilde{N}_g - n_{g_L}^\Sigma(l)|$$

The mean distance in-between cells within the same layer resulting from this model was used as distance in-between germ cell layers:

$$w_L := \phi_g + \overline{d_g}$$

with

$$\overline{d_g} := \sum_{\lambda=1}^{\lfloor \frac{d_B}{w_L} \rfloor} \frac{d_{g_L}(\lambda)}{\lfloor \frac{d_B}{w_L} \rfloor}$$

where

$$d_{g_L}(l) := \frac{\phi_L(l)}{1 + \frac{1}{\sin\left(\frac{\pi}{N_g}\right)}} - \phi_g$$

Germ layers were annotated functionally based on literature (Brenner and Schedl, 2016; Fox et al., 2011). The proximal gonad arm was filled with 8 oocytes, maximizing the diameter of each oocyte under the constraint of the corresponding gonad diameter. The proximal end of the distal germ cell layers and the distal end of the distal-most oocyte defined the boundaries of the loop region. Assuming steady-state with an apoptotic rate of 90 % (Brenner and Schedl, 2016), the loop region was filled with 100 germ cells in layers (uniformly spread across the loop region).

In order to determine the germ cell migration time, migration speed data was used from Wolke and colleagues (Wolke et al., 2007).

Those migration speeds were assigned to germ cell layers and oocytes as follows:

- i) The switch to pachytene happens between the transition zone and the meiotic zone.
- ii) The switch to 'proximal of the loop' happens at the bend (398.3 μm from the DTC).

- iii) In between, the pachytene was split into thirds to assign 'early', 'mid', and 'late' pachytene.
- iv) The speed of particles entering oocytes was assigned to the very last germ cell layer in the loop region.
- v) Oocytes were assigned the speed of particles right after entering oocytes.

For each germ cell layer and oocyte, 10 speed values were sampled from a zero-truncated normal distribution with the corresponding mean and standard deviation reported by Wolke and colleagues (Wolke et al., 2007).

Those speed samples were smoothed by calculating the running mean of 100 adjacent values (10 adjacent germ layers / oocytes).

The smoothed speed samples were used to fit a natural spline function mapping the distance to the DTC d [μm] to the migration speed $v(d)$ [$\frac{\mu\text{m}}{\text{min}}$].

To translate migration speeds to migration times, for each distance (in 1 μm steps) along the gonad arm, the migration speed in $\frac{\mu\text{m}}{\text{min}}$ was calculated, inverted and summed up cumulatively to get the total migration time from the DTC to the given distance.

These discrete measurements were used to fit the final spline model mapping the distance to the DTC d [μm] to the migration time from the DTC $t(d)$ [h].

2.9 Computational pipeline

(Diag et al., 2018)

2.9.1 Data pre-processing

Raw sequencing basecalls were demultiplexed and converted to FASTQ format using `bcl2fastq v2.18.0.12` pooling reads across lanes (`--no-lane-splitting`). No adapter trimming was performed at this stage by not specifying adapter sequences in the sample sheet CSV file. To avoid masking of the short read 1 (barcode and UMI), the `--mask-short-adapter-reads=10` option was used. 3' reads (read 2) were annotated with their corresponding (corrected) barcode and UMI sequences (read 1) using custom scripts. Reads with identical barcode, UMI and sequence were collapsed and the unique reads were assigned to per-slice FASTQ files by barcode. Small RNA reads were subject to two rounds of 3' end trimming by `flexbar v. 2.5` (Dodt et al., 2012): The first round to remove 3' adapters, the second to remove the poly(A)-tail added during the library preparation (using 10 A's as 'adapter sequence'). 3' nucleotides with low basecall quality scores were trimmed using `flexbars --pre-trim-phred=30` option and the 3 nucleotides 5' overhang introduced by the template-switching polymerase were trimmed using a custom awk script also discarding reads with a remaining length < 18 nts.

2.9.2 Mapping of reads to the *C. elegans* genome

RNA-seq reads were mapped to the ce11/WBcel235 genome assembly using STAR_2.5.1b (Dobin et al., 2013) and an index with splice junction information from the Ensembl 82 transcriptome annotation. Alignments were sorted using sambamba v0.4.7 (Tarasov et al., 2015). Coverage tracks were generated using bedtools v2.23.0 (Quinlan and Hall, 2010) via the `genomecov` command specifying the `-split` and `-bg` options for splice-aware BedGraph output and splitting by strand using the `-strand` parameter. The total number of mapped reads per sample/splice was determined using the `flagstat` command of samtools 0.1.19-96b5f2294a (Li et al., 2009) and converted to the corresponding `-scale` parameter for `bedtools genomecov` for reads per million-mapped (RPM) normalization. Coverage BedGraph files were converted to BigWig format using `bedGraphToBigWig` v 4.

2.9.3 3' extension of transcript annotation

The identification of downstream coverage peaks for 3' extension of the WS260 transcriptome was performed using a custom R script: For each protein coding gene, the intergenic distance to the next downstream protein coding, ncRNA, lincRNA, pseudogene, rRNA or snoRNA gene (on the same strand) was calculated. Intergenic regions longer than 10 kb were truncated and the RPM-scaled genome coverage per sample of those regions was extracted from the BigWig files generated before. The per-sample coverage vectors were averaged per genomic position and binarized into uncovered regions (< 5 RPM mean coverage) and covered regions (≥ 5 RPM mean coverage). Covered regions with a length of ≥ 50 nucleotides were considered as coverage peaks. Per downstream intergenic region, the downstream-most coverage peak was selected for the 3' extension of the corresponding upstream gene. Only downstream extensions with a length up to 3 kb were considered for downstream analyses. For each gene with a downstream extension, all annotated transcript isoforms extending to the 3' most genomic position of the corresponding gene were kept and got their 3' UTRs extended by until the 3' position of the respective downstream peak. Those 3' extended transcripts were exported to a GTF file and merged with the WS260 transcriptome annotation using custom awk scripts.

2.9.4 Transcriptome pre-processing

To enable the assignment of 3' end RNA-seq reads to transcript isoforms, the 3' extended WS260 transcriptome annotation was pre-processed using a series of custom R scripts: 3' A's were trimmed from all annotated transcripts as they would be indistinguishable from poly(A)-tails. The resulting transcripts were truncated to the 3' most 500 nts. Transcript isoforms with the same genomic coordinates and internal structure were collapsed and enumerated by decreasing corresponding (max.) 3' UTR length.

2.9.5 Isoform-specific transcript abundance estimation

RNA-seq reads were assigned to transcripts using kallisto 0.43.1 (Bray et al., 2016): For 3' reads, an index of the collapsed transcriptome annotation described above was used. For full-length coverage reads (poly(A)+ and ribodepleted total RNA-seq libraries), an index of the full 3' extended transcriptome annotation was used. For all libraries, the `--bias` was passed to `kallisto quant`. For single end reads, additionally the `--single`, `--fragment-length=1` and `--sd=1` options were used. All libraries were sequenced with a first-strand-reverse stranded protocol. Thus, poly(A)+ and ribodepleted total RNA-seq samples were analyzed in `--rf-stranded` mode. The 3' reads, while presented to kallisto as single-end reads, originally were sequenced as read 2, therefore resembling first-strand-forward single-end data. Thus, for these libraries the `--fr-stranded` mode of `kallisto quant` was used. Per-isoform read counts were exported to TSV files using the `--plaintext` option.

2.9.6 Data processing

The raw read counts per transcript isoform and slice/sample were further processed using a custom R script: Though the whole annotated transcriptome was quantified to check for specificity of the experimental and computational approach, downstream analyses were limited to protein coding transcripts only. For gene-level analyses, isoform-level read counts were summed across all isoforms of a given gene. To compensate for differences in sequencing depth, raw read counts were normalized to counts-per-million (CPM). For full-length coverage protocols (poly(A)+ and ribodepleted total RNA-seq) an additional correction for the transcript length was performed, resulting in transcripts per million (TPM) estimates. Slice-data were arranged from distal-to-proximal by the known order of their barcodes and assigned to a relative position scale representing each slice by its center and accounting for differences in the number of slices per sample.

2.9.7 Aligning cryo-cuts of different samples to a single coordinate system

As the start- and endpoint of gonad slicing was not precisely the same for all replicates, all slices of different replicates were aligned to a common coordinate system. This was achieved by comparing per-sample LOESS (locally estimated scatterplot smoothing) fits of abundance estimates across slices with ISH images of certain genes in the germline. Therefore, the gene profile of one replicate was fixed according to the corresponding *in situ* image and other replicates were aligned to the fixed replicate. This was done for ~ 30 gene profiles and the median of the shifting for those 30 profiles was calculated and used for all gene profiles.

2.9.8 Integration of replicate data

The aligned discrete per-gene/isoform spatial expression profiles of individual replicates were used to fit a continuous consensus profile using LOESS with a span of 0.4 through a custom R script. Slices with less than 10,000 reads assigned to the transcriptome ('dropout-slices') were excluded from the fitting procedure. For visualization, 50 equidistant points along the distal-to-proximal axis were inferred from those fits. For downstream analyses, only 20 points were used to reflect the actual resolution of the data more conservatively. All data (incl. dropout-slices) are available through the interactive data exploration interface published alongside this study.

2.9.9 Data analysis of the small RNA transcriptome

The trimmed libraries were first mapped with bowtie2 (version 2.3.3.1) (Langmead and Salzberg, 2012) using the parameters `--very-fast-local --phred33 --local` to the *E. coli* genome (NC_000913.3, K-12, MG1655) in order to remove *E. coli* RNA contamination. The cleaned-up libraries were then mapped with STAR (version 2.5.3a) (Dobin et al., 2013) to the WBcel235/ce11 genome assembly using the Ensembl 87 annotation and the parameters `-alignIntronMax 140000 -alignSJDBoverhangMin 17 --alignSplicedMateMapLmin 30 --outFilterMultimapNmax 5 --outFilterMismatchNmax 2 --outFilterMatchNmin 17 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0`.

Sense and antisense read counting on features was done using HTSeq (version 0.9.1) (Anders et al., 2015) with the parameters

```
-a 0 -m intersection-nonempty --nonunique=all
--secondary-alignments=score combined with -s yes for sense and with
-s reverse for antisense counts.
```

Known and novel miRNAs were identified separately using the cleaned-up libraries and the miRDeep2 algorithm (version 2.0.0.7) (Friedländer et al., 2012) with the miRBase21 reference. First, miRDeep2 was ran on the pooled libraries. Then, the novel miRNA predictions found were added to the miRBase21 reference. Consequently, the combined reference of known and novel miRNAs was used for a second run of miRDeep2 on each library separately and on the pooled library as well. This way we unified the expression estimates of known and novel miRNAs under a common measure of counts per million of mapped reads (CPM).

The miRNA-target correlation analysis used robust linear regression based on the MM-estimator in order to reduce the effect of outliers (Koller and Stahel, 2011). All miRNAs were divided into families based on their 2 - 7 nt 6mer seeds (reverse-complemented). Putative target genes were identified by counting miRNA 7mer seeds on all of their unique and longest 3' UTR isoforms. The 3' UTR isoform with the maximum number of 7mer seeds was taken as

representative for that miRNA-target gene interaction. The miRNA 7mer seeds were chosen to be either the reverse-complement of the miRNA 2 - 8 nts or the reverse-complement of the miRNA 2 - 7 nts immediately followed by an A (Bartel, 2009). The control list of targets was generated by mutating the 3rd and 4th nucleotides of these 7mer seeds. Robust linear regression was done by summing the LOESS smoothed CPMs among the miRNA family members on each LOESS point and using this summarized family-wise expression with the corresponding target smoothed expression. In order for a correlation to be considered both the family-wise miRNA expression and the target expression were demanded to be commonly non-zero in at least 25 % of the LOESS points.

For the analysis of siRNAs and piRNAs, the alignments overlapping sense and antisense annotated exons were counted (ce11 Ensembl annotation WBcel235) using HTseq (version 0.6.0) (Anders et al., 2015) with the following parameters

```
-f bam -r pos -a 0 -i gene_id -m intersection-nonempty -nonunique=all
-secondary-alignments=score -supplementary-alignments=score combined
with -s yes for sense and -s reverse for antisense overlap counts. For each gene the
relative antisense coverage  $f = \text{antisense} / (\text{sense} + \text{antisense})$  was computed and
weighted by antisense TPM expression to produce an antisense expression score
 $\text{score} = (f - 0.5) 10 \log_{10}(1 + \text{antisense TPM})$ . Using the gene annotations, we
identified the piRNA annotated genes and their corresponding sense and antisense coverage.
```

For the analysis of 21U, 22G and 26G RNAs, all reads of lengths 21, 22 and 26 nts long with their corresponding 5' nucleotides being U, G and G, respectively, were isolated from the mapped small RNA-seq libraries using a custom python script. Again, the exon coverage was computed based on these read alignments of all annotated exons using HTseq (Anders et al., 2015) as it was done before for the siRNA analysis. Furthermore, this procedure was repeated using the RepeatMasker (developed by Smit, A.F.A., Hubley, R., and Green P. (2013-2015); see <http://www.repeatmasker.org/>) UCSC (Kent et al., 2002) annotation for ce11 in order to identify sense and antisense coverage on repeat elements like Simple repeats or long terminal repeats (LTRs).

3 Results

3.1 Cryo-cut based sequencing method is reproducible and reliable

To investigate the spatial and temporal distribution of RNA molecules in the *Caenorhabditis elegans* germline, I adapted and optimized a previously published cryo-cut-based sequencing method, called tomo-seq, for the *C. elegans* gonad (Junker et al., 2014). Therefore, I dissected the gonad, embedded, and cryo-cut it from the distal tip cell up to the spermatheca at 50 μm resolution. Following RNA extraction, each individual section was subjected to the CEL-seq1/2 protocol that relies on polyadenylated transcripts (Hashimshony et al., 2012, 2016). However, CEL-seq1/2 was not suitable for small RNAs, as small RNAs are not polyadenylated. Hence, I established the SMARTer smRNA sequencing protocol for detection of small RNAs in the sectioned gonad.

To ensure that the embedding, freezing, and cryo-sectioning itself do not harm the germ cells, I embedded the gonad of a *his-72::gfp* reporter strain and cut it into slices of 14 μm . The *his-72* gene encodes for an H3 histone, one of the core proteins of chromatin in eukaryotic cells. Thus, the *his-72::gfp* reporter strain enables the visualization of chromatin of intact germ cell nuclei under a fluorescence microscope. As shown in Figure 5, germ cell nuclei and gonad were not destroyed by embedding, freezing, and/or cryo-cutting.

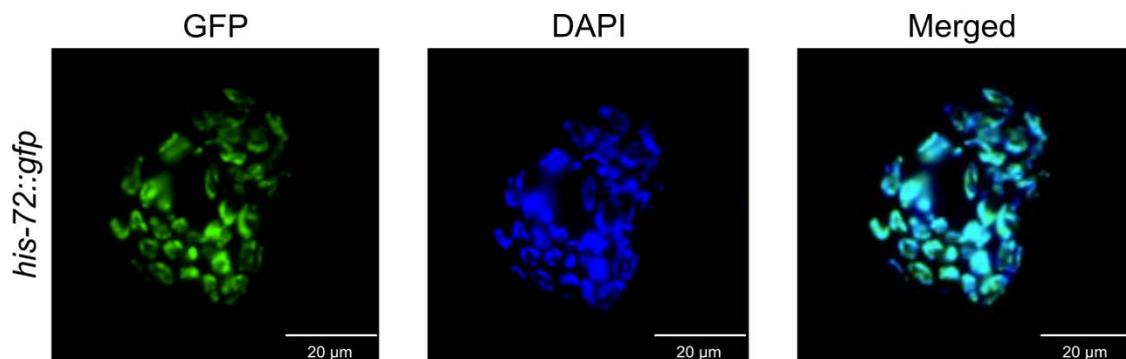


Figure 5. Embedding, freezing, and cryo-sectioning of the gonad do not harm the germ cells. Fluorescence images of a slice from the *C. elegans* gonad arm (150 μm from distal tip cell) using a *his-72::gfp* reporter strain. The *his-72* gene encodes an H3 histone that is part of the chromatin structure. Hence, the chromatin of each intact germ cell nuclei is stained in green. DAPI stains all germ cell nuclei in blue. Additionally, an overlay of GFP and DAPI signal is shown (Merged). Slices were cut at 14 μm resolution. Scale bar: 20 μm .

After confirming that the experimental procedure of gonad dissection and cryo-cutting did not harm the germ cells, I proceeded with the library preparation and sequencing. Using the CEL-seq1 protocol only, did not result in any mRNA library, as most of the material was lost during the experimental procedure. This was likely due to the purification steps with columns. Therefore, I combined the CEL-seq1 protocol with the purification steps of CEL-seq2 where Agencourt AMPure XP beads were used instead of columns. Consequently, I reduced the material loss during purification steps and succeeded in producing appropriate libraries of an

average size of 400 bp. All mRNA libraries were sequenced on the Illumina NextSeq 500 machine and sequences were mapped to the *C. elegans* transcriptome. As shown in Figure 6 for all mRNA samples most reads can be assigned to known *C. elegans* transcripts (Fig. 6A). The mappability varied between 67 % and 85 % (average: 75.8 %) dependent on the loss of material during the experimental procedure. Furthermore, the majority of quantified transcripts were in line with a poly(A)-selection profile (Fig. 6B) (Diag et al., 2018). This was expected, as the barcoded primers used for the reverse transcription reaction consisted of a 24 bp poly(T)-stretch, hence, capturing only RNAs with a poly(A) tail, *i.e.*, primarily mRNAs. The Pearson correlation across all transcripts, averaged over all sections for two independent biological and technical replicates, demonstrated that the cryo-cut-based approach is highly reproducible ($r = 0.96$) (Fig. 6C). In order to investigate if the cutting procedure would alter the transcript detection, I performed the same protocol with an uncut gonad. In fact, the uncut gonad arm transcript expression correlated well with the expression averaged across all slices ($r = 0.94$) (Fig. 6D). Moreover, we reconstructed the whole-gonad expression profile *in silico*.

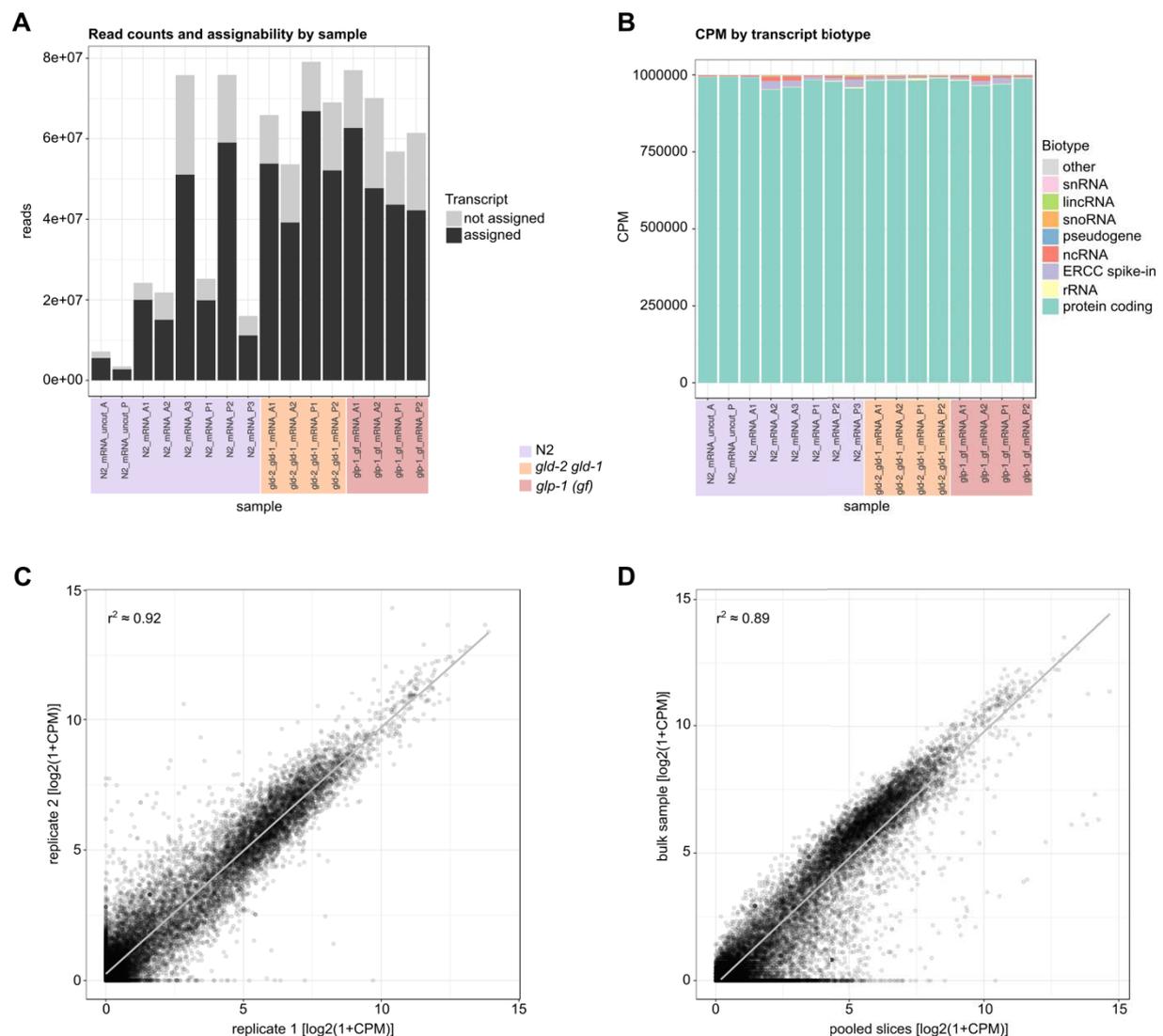


Figure 6. Cryo-cut based sequencing method for detection of spatiotemporal mRNA expression is reproducible and reliable. (A) Read counts and assignability of reads for each biological and technical replicate of N2, *gld-2 gld-1* double mutant and *glp-1 (gf)* mutant. **(B)** Transcript biotype distribution over the fraction of mapped reads for each biological and technical replicate of N2, *gld-2 gld-1* double mutant and *glp-1 (gf)* mutant. **(C)** Linear correlation (Pearson's r) across all transcripts, summed and averaged over all sections for two biological replicates. **(D)** Linear correlation (Pearson's r) across all transcripts of uncut (bulk) sample and sliced samples (summed and averaged over all sections for all biological replicates). (Diag et al., 2018)

To reliably estimate the abundance of RNAs and for later normalization, I added a certain amount of an ERCC RNA spike-in mix to each slice of the mRNA samples. As shown in Figure 7 the estimated spike-in abundance reflected the known concentration (Diag et al., 2018). However, some lower abundant ERCC Spike-Ins were not or at a very low level detected (Fig. 7A). This might be explained by the low amount of ERCC Spike-In used in this study. Nevertheless, comparing our cryo-cut based sequencing approach to other sequencing methods such as poly(A)+ RNA-sequencing or ribosomal RNA depleted total-RNA-sequencing showed that our method correlates well with other sequencing methods ($r = 0.95$ and $r = 0.94$, respectively) (Fig. 7B) (Diag et al., 2018).

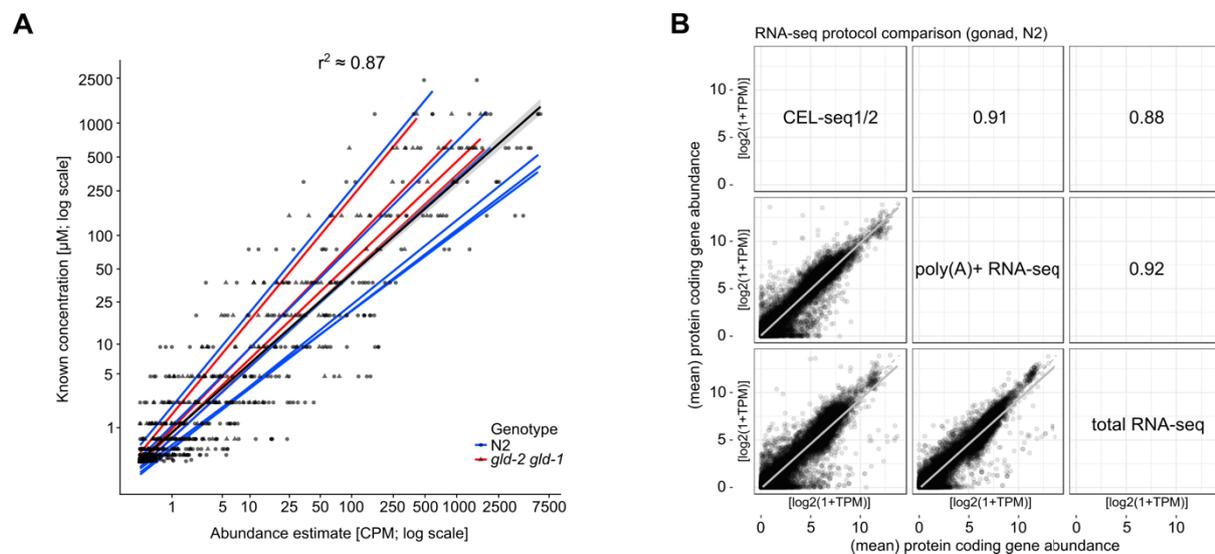


Figure 7. Cryo-cut based sequencing method for detection of spatiotemporal mRNA expression is reliable and compares well with other sequencing methods. (A) Linear correlation (Pearson's r) of known ERCC Spike-In concentrations and estimated spike-in abundance for N2 (red line) and *gld-2 gld-1* double mutant (blue line). **(B)** Linear correlation (Pearson's r) across all genes for different sequencing approaches, *i.e.*, CEL-seq1/2, poly(A)+ RNA-seq and total RNA-seq. (Diag et al., 2018)

Small RNA libraries were sequenced on the Illumina HiSeq 2500 machine and mappability was analysed separately for each slice of each replicate, as samples were not barcoded and pooled. In summary, the mappability of small RNA libraries was lower compared to the mRNA libraries, on average 59 % (Fig. 8A). However, the fraction of reads obtained for miRNA sequences was decent enough to detect most of the expressed miRNAs in *C. elegans*.

Additionally, the reproducibility between two independent biological and technical replicates was high ($r = 0.88$) (Fig. 8B) and whole-gonad small RNA expression correlated well with the small RNA expression averaged across all slices ($r = 0.91$) (Fig. 8C).

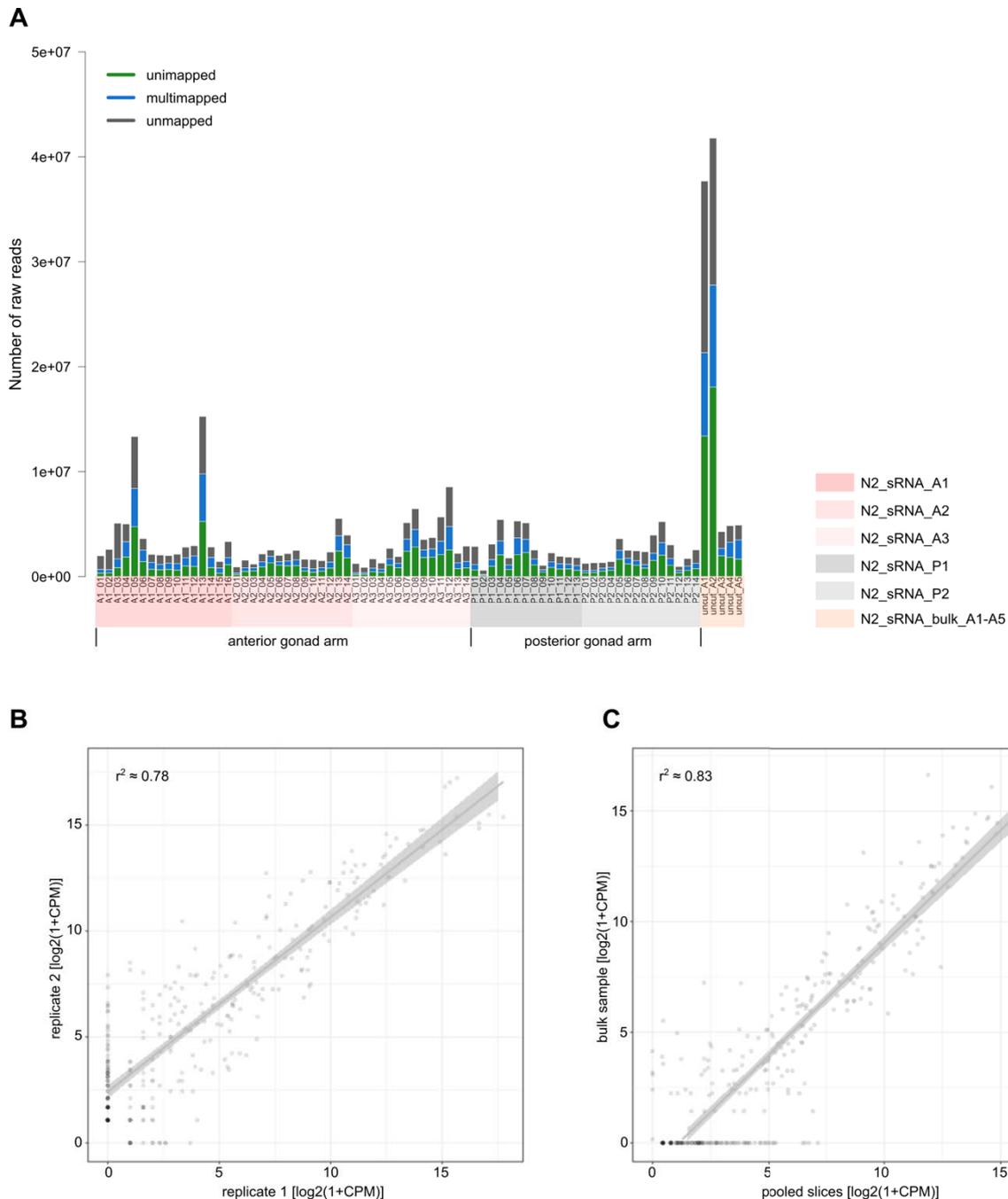


Figure 8. Cryo-cut based sequencing method for detection of spatiotemporal small RNA expression is reproducible and reliable. (A) Read counts and assignability of reads for each biological and technical replicate. **(B)** Linear correlation (Pearson's r) across all miRNAs of *in silico* pooled slices for two biological replicates. **(C)** Linear correlation (Pearson's r) across all miRNAs of uncut (bulk) sample and *in silico* pooled slices. (Diag et al., 2018)

Overall, our data demonstrated high reproducibility and reliability for the detection of mRNAs as well as for small RNAs.

3.2 Cryo-cut-seq creates a spatiotemporal RNA map of the *C. elegans* gonad

The *C. elegans* hermaphrodite possesses two gonad arms, an anterior and a posterior gonad arm, two mirror-image U-shaped tubes. The morphology and structure are exactly the same for both arms. However, we asked the question if anterior and posterior arm differ in gene expression. Therefore, I dissected, embedded, and cut both arms. In concordance with the literature, we did not observe major differences above background between both arms (Fig. 9) (Diag et al., 2018). It is likely that some of the detected differences reflect noise, as the differences decreased with increasing expression levels. Additionally, we did not observe any differences in small RNA expression. Thus, we decided to treat anterior and posterior samples as biological replicates and pooled both in order to increase the statistical power for further downstream analysis.

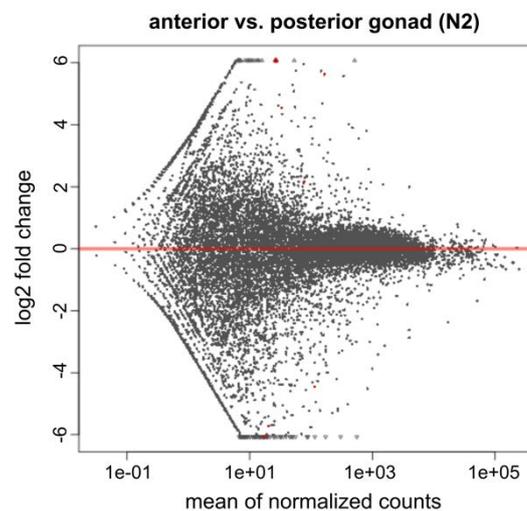


Figure 9. Anterior and posterior gonad arm do not differ in gene expression. Comparison of all N2 samples on the gene level by gonad arm (anterior or posterior) using DESeq2 (Love et al., 2014). Observable gene expression differences between both arms reflects probably noise, as differences decrease with increasing expression levels. (Diag et al., 2018)

The biological and technical replicates were marginally shifted and/or compressed to each other due to different cutting start- and endpoints of these samples (Diag et al., 2018). Therefore, all slices of different replicates needed to be aligned to a common coordinate system before pooling them (Diag et al., 2018). This was achieved by comparing the gene expression profile of certain mRNAs with *in situ* hybridization (ISH) images. Following this approach, one sample was fixed according to the ISH images and the other replicates were aligned to this fixed sample. This was done for ~ 30 genes, the median of shifting was calculated, and used for further downstream analysis. The small RNA samples were not shifted to each other. Thus, a *de novo* alignment to a common coordinate system was not necessary.

Following, we investigated the expression profiles of single mRNAs and miRNAs in detail. We observed that mRNAs as well as miRNAs localized in distinct expression patterns (Fig. 10) (Diag et al., 2018). These expression patterns were consistent with ISH images I performed in this study (Fig. 10). In addition, Figure 10A and 10C show how well the raw data of all replicates align to each other demonstrating high reproducibility.

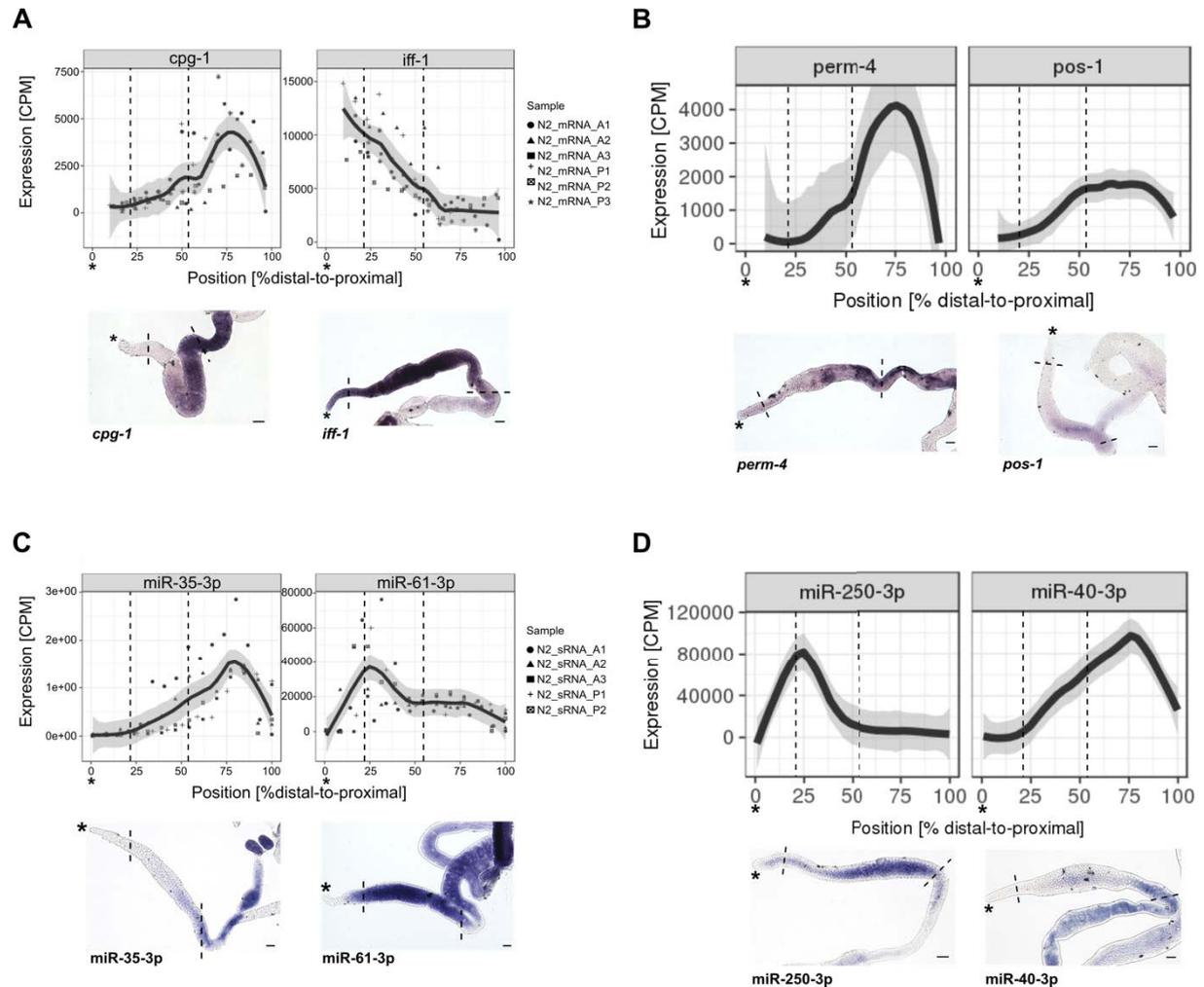


Figure 10. mRNAs and miRNAs are localized in the germline. (A) Spatial expression of *cpg-1* and *iff-1* from distal-to-proximal. $n = 6$ independent experiments (N2_mRNA_A1-A3 and N2_mRNA_P1-P3) for wild type N2, LOESS \pm standard error (SE). Corresponding *in situ* hybridization (ISH) images of *cpg-1* and *iff-1* (lower panel). (B) Spatial expression of *perm-4* and *pos-1* from distal-to-proximal. $n = 6$ independent experiments, LOESS \pm SE. Corresponding ISH images of *perm-4* and *pos-1* (lower panel). (C) Spatial expression of miR-35-3p and miR-61-3p from distal-to-proximal. $n = 5$ independent experiments (N2_sRNA_A1-A3 and N2_sRNA_P1-P2) for wild type N2, LOESS \pm SE. Corresponding ISH images of miR-35-3p and miR-61-3p (lower panel). (D) Spatial expression of miR-250-3p and miR-40-3p from distal-to-proximal. $n = 5$ independent experiments, LOESS \pm SE. Corresponding ISH images of miR-250-3p and miR-40-3p (lower panel). Asterisk: Distal tip cell. Scale bar: 20 μ m. Dashed lines represent the different zones in the germline. (Diag et al., 2018)

Interestingly, in most cases both RNA classes were either localized to the mitotic or to the oogenesis region. The meiotic region seemed like a transition zone where genes required for

mitosis decreased and genes required for oogenesis or embryogenesis increased in expression.

To investigate if the localized expression patterns hold true on a global level, we clustered the expression of germline specific genes (Wang et al., 2009) and of all expressed and detected miRNAs, respectively (Fig. 11) (Diag et al., 2018). Indeed, mRNAs and miRNAs showed localized expression patterns on a global scale. Moreover, we observed again predominantly two distinct clusters, *i.e.*, genes or miRNAs localizing to the mitotic zone and genes or miRNAs localizing to the late meiotic/oogenesis compartment.

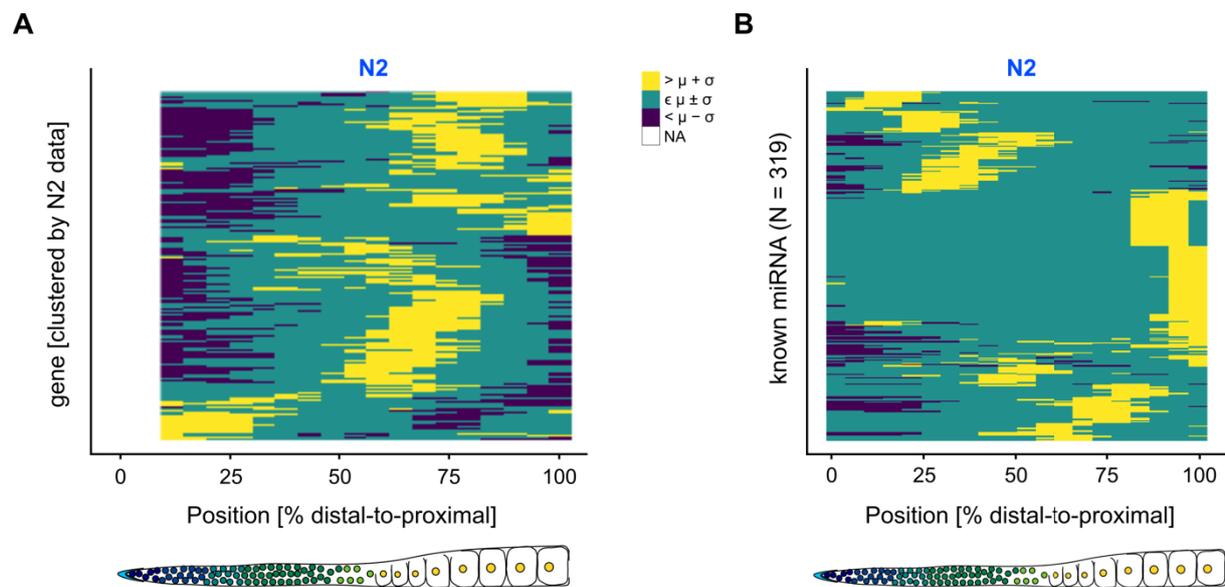


Figure 11. Cryo-cut-seq generates a spatiotemporal RNA map of the *C. elegans* gonad. (A) Hierarchical clustering of germline specific genes by linear correlation ($1 - \text{Pearson's } r$) for N2. NA: No data. **(B)** Hierarchical clustering of known miRNAs by linear correlation ($1 - \text{Pearson's } r$) for N2. μ : Mean; σ : Standard deviation. (Diag et al., 2018)

For the very distal part of the gonad arm we did not have any mRNA expression data. This can be explained by two scenarios: The RNA content is very low in this region. Additionally, it is possible that I cut empty medium upstream of the embedded gonad. It is unlikely due to loss of material, as we observed ERCC Spike-In reads for these slices and the ERCC Spike-In mix was added at the beginning of the experimental procedure to each sample before RNA extraction (Diag et al., 2018). To determine the starting point of RNA expression, we used *in situ* data as a guide as described earlier (Diag et al., 2018). Interestingly, for the small RNA data we observed expression in every slice. An explanation for this might be that the small RNA data was generated later while having more practice in embedding and cutting samples without collecting empty slices.

Nevertheless, with the cryo-cut-seq method we generated a spatiotemporal mRNA and miRNA map of the *C. elegans* gonad and the expression patterns correlated well with ISH images.

3.3 Germline-specific miRNA family co-localizes with putative targets

As we observed similar expression patterns for mRNAs and miRNAs along the *C. elegans* gonad, we next asked the question if miRNAs co-localize with their putative targets. We focussed for this analysis on the miR-35 family as this family is specific for the germline (Miska et al., 2007). Indeed, our sequencing data and corresponding ISH images showed that miR-35 family members (miR-35-3p and mir-40-3p) are enriched in the germline and located around the loop region (late pachytene stage/early oogenesis) (Fig. 10C and 10D). All members of the miR-35 family share the same 6mer seed sequence that is located in the 2 - 7 nts region of each miRNA. The seed region is important for mRNA recognition, binding, and destabilization. Hence, we decided to sum up the expression of all family members and to correlate the summed-up expression with the corresponding expression of putative targets. As miRNAs bind their targets primarily in the 3' untranslated region (UTR), putative targets were identified by their 3' UTR carrying at least one 7mer seed (2 - 8 nts of the miRNA) or one 6mer seed (2 - 7 nts of the miRNA) with an A opposite the first miRNA nucleotide (Bartel, 2009; Diag et al., 2018). Interestingly, we observed that the miR-35 family members correlated positively with their putative targets (Fig. 12). This was expected as a co-localization (positive correlation) is a necessity for *in vivo* miRNA:mRNA interaction (Diag et al., 2018; Shkumatava et al., 2009). Additionally, a recent study by Dallaire and colleagues demonstrated that miR-35-mediated translational repression has a stabilizing effect on its mRNA targets in the germline requiring a co-localization of both RNA classes (Dallaire et al., 2018).

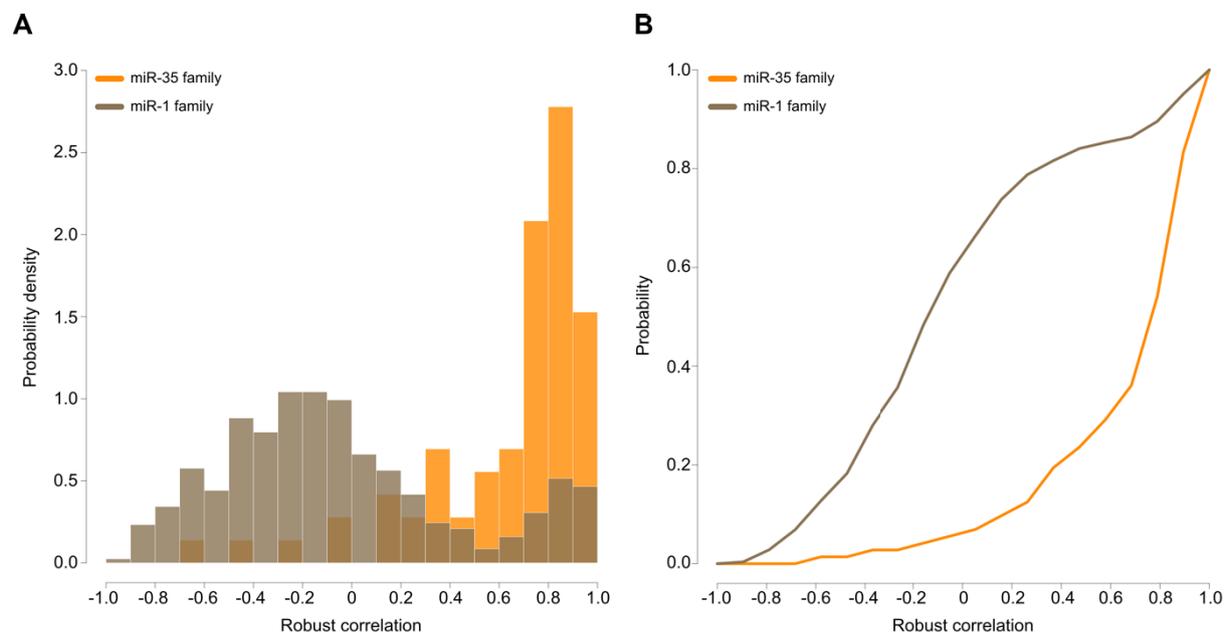


Figure 12. A germline-specific miRNA family co-localizes with its predicted targets. (A) Histogram of robust correlation of miR-35 and miR-1 family members with their putative mRNA targets, respectively. **(B)** Density of robust correlation coefficients of miR-35 and miR-1 family members with their putative mRNA targets, respectively. (Diag et al., 2018)

We performed the same analysis for a non-germline specific miRNA family, the miR-1 family. In contrast to the miR-35 family members, the miR-1 family did not show any prominent co-localization with its putative targets (Fig. 12). This suggests an miRNA:mRNA interaction outside the germline (Diag et al., 2018).

3.4 Germline-specific small RNA sequencing predicts novel miRNAs

With our small RNA sequencing approach, we were able to detect miRNAs that were very lowly expressed and/or are limited to a specific region in the gonad. These miRNAs would have been missed by using already established small RNA sequencing methods (*e.g.* Illumina TruSeq small RNA sequencing), as these require high input material (*e.g.* whole worm samples) and the signal of a lowly expressed and to a specific region localized miRNA (in the germline) would have been diluted out (Diag et al., 2018). As these miRNAs would be essential for regulating spatial expression, we aimed to screen our germline-specific small RNA sequencing data for putative novel miRNAs (Diag et al., 2018). For this purpose, we ran miRDeep2 (Friedländer et al., 2012) on our data. Indeed, we predicted 83 putative novel precursor miRNAs (142 mature and star miRNAs) (Tab. 1) (Diag et al., 2018). For quantification of these and known miRNAs, we included the mature and precursor sequence of the novel predictions to the miRBase21 reference and re-ran miRDeep2 on this data set (Diag et al., 2018). This analysis revealed that known and putative novel miRNAs were quantified spatially (Fig. 13). This was true for all replicates ($n = 5$) demonstrating the reproducibility of our sequencing approach (appendix, Fig. 30 and Fig. 31). Interestingly, for the known miRNAs we observed a fraction ($\sim 6\%$ of all detected miRNAs) that was highly expressed in the germline while all the other miRNAs were expressed moderately (Fig. 13A). Amongst the highly expressed known miRNAs we found the miR-35 family members, in accordance with the fact that this family is highly and specifically expressed in the germline. The putative novel miRNAs displayed the same expression pattern with $\sim 8\%$ being very highly expressed (Fig. 13B). However, and in contrast to the known miRNAs, these highly expressed novel miRNAs were restricted to the very distal part of the gonad, suggesting an important functionality of these in the mitotic region. Additionally, a high fraction of the novel predictions ($> 40\%$) was very lowly expressed in the germline (≤ 100 CPM on average) explaining why they were missed in previous studies.

Overall, with our germline specific small RNA sequencing approach we were able to quantify miRNAs reproducibly in space and time along the gonad. Additionally, we predicted many putative novel miRNAs that were missed by previous attempts to identify novel miRNAs.

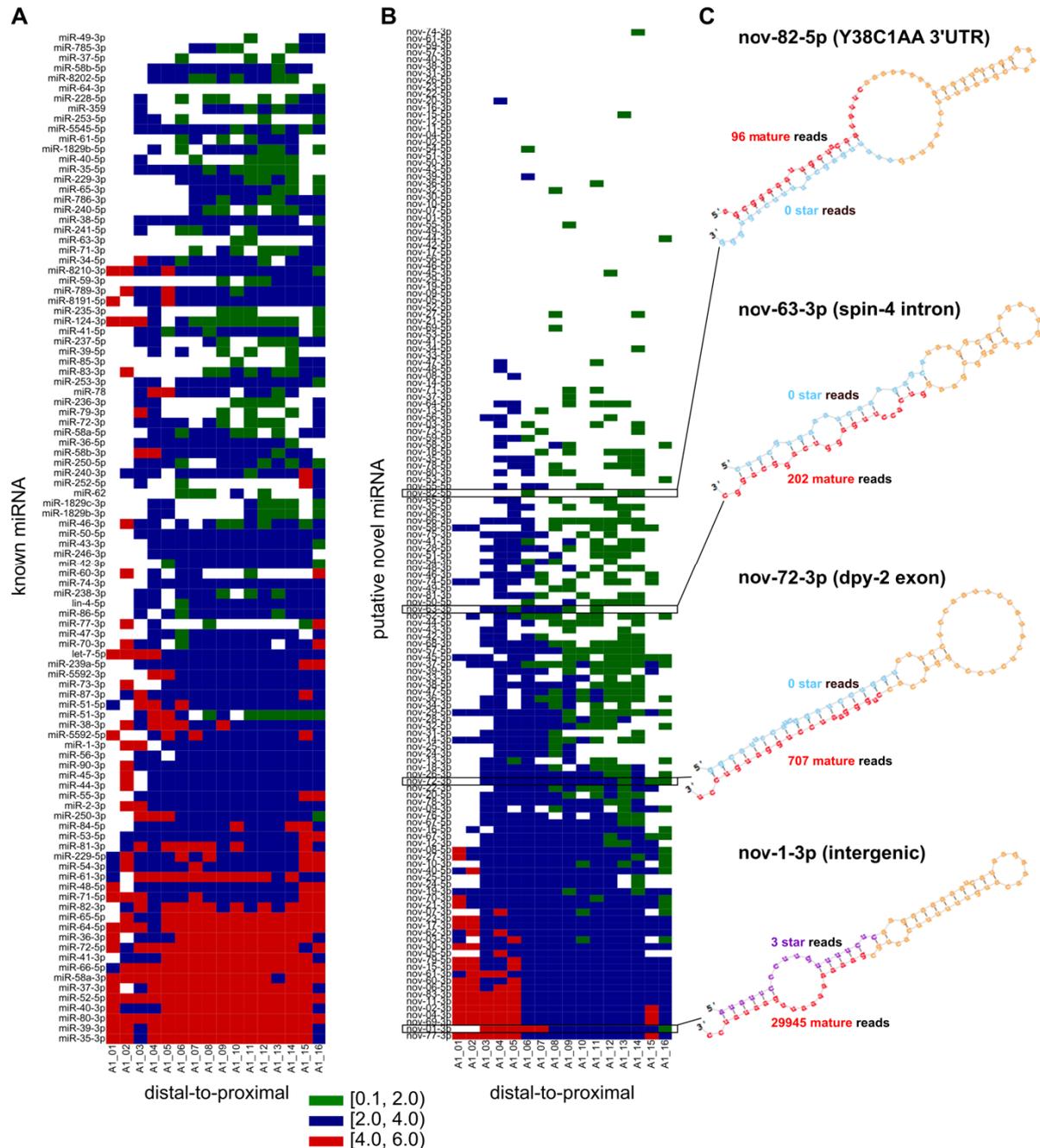


Figure 13. Known and putative novel miRNAs are quantified in space and time along the gonad. (A) Heatmap of the top 100 expressed known mature miRNAs ($\log_{10}[1+\text{CPM}]$) detected in this study. Each row represents one miRNA and each column one slice of one replicate (A1_01-16), ordered from distal-to-proximal. **(B)** Heatmap of all putative novel miRNAs ($\log_{10}[1+\text{CPM}]$) detected in this study. Each row represents one miRNA and each column one slice of one replicate (A1_01-16), ordered from distal-to-proximal. **(C)** Four examples of identified novel miRNAs of different *C. elegans* genomic origin. Reduced miRDeep2 (Friedländer et al., 2012) plots show the precursor hairpin structure and the coverage of mature (red), star (blue, violet), and loop (yellow) sequences (Diag et al., 2018).

We next focused more on the novel miRNA predictions. We investigated structure, read coverage, and putative targets in more detail. For this purpose, we randomly picked four novel candidates out of the 83 predicted ones with different expression levels (Fig. 13C). All four

candidates showed a miRNA-like hairpin structure when folding their pre-miRNA sequence *in silico* (Fig. 13C) (Diag et al., 2018). Moreover, star and mature sequence largely complemented each other. To exclude the possibility that the chosen novel candidates could be technical artefacts, I aimed to validate them by using the TaqMan assay or small RNA ISH (see Materials and methods for more details). With these approaches, I validated three of the four candidates (Fig. 14A and 14B) (Diag et al., 2018). I could not validate nov-82-5p probably because this candidate was very lowly expressed (< 100 CPMs) and the low detection signal was out of the TaqMan assay range (Fig. 13B and 14A). Consistently, the expression levels (in CPMs) correlated well with corresponding C_T values measured by the TaqMan assay (expression of mature miRNAs) (Fig. 14C) (Diag et al., 2018).

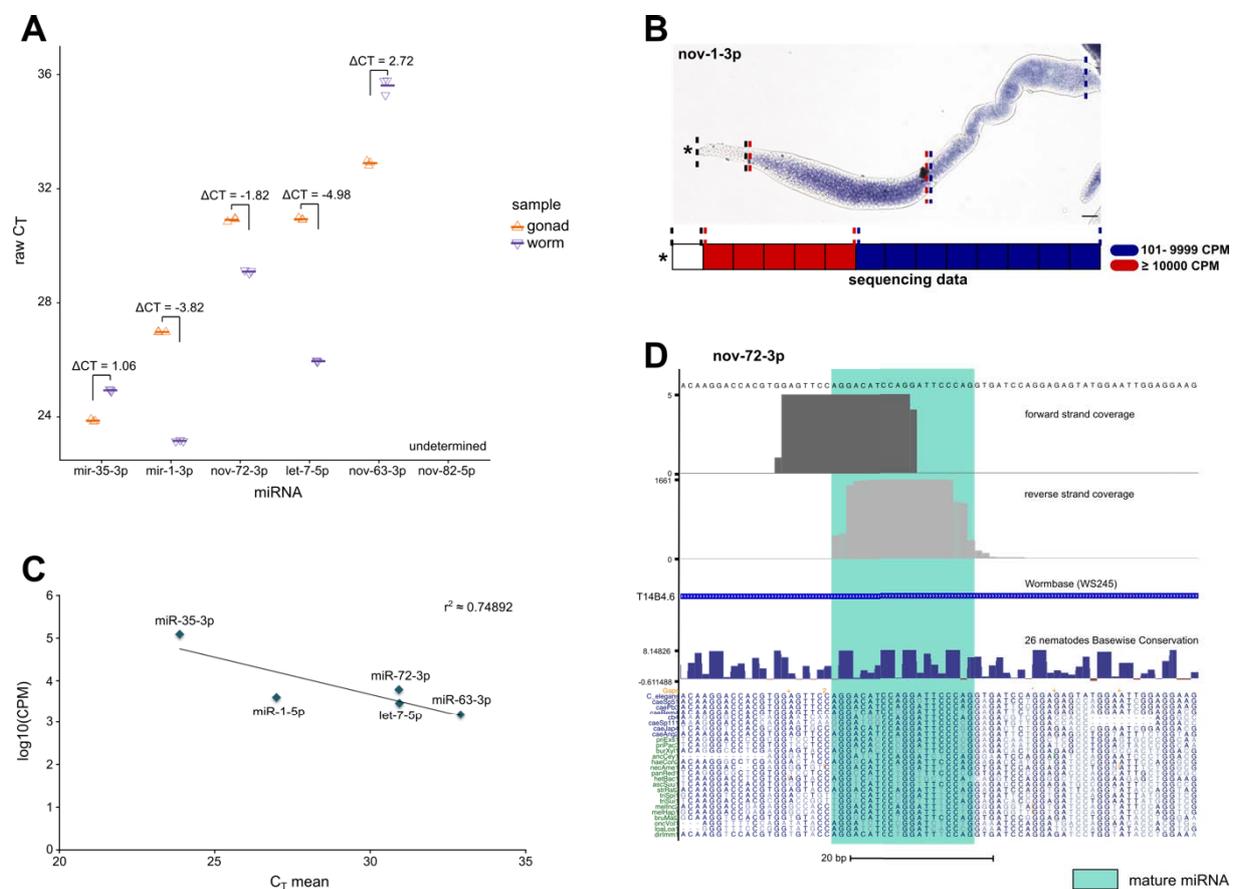


Figure 14. Novel miRNA predictions exhibit miRNA-like features. (A) TaqMan assay validation of known miRNAs expressed in *C. elegans* (mir-35-3p, mir-1-3p and let-7-5p) and novel miRNA predictions (nov-63-3p, nov-72-3p and nov-82-5p). $n = 3$ independent experiments of gonad and whole worm sample, respectively. Triangles: raw C_T values; Bars: mean values. **(B)** *In situ* hybridization image of nov-1-3p with corresponding spatial sequencing data. Asterisk: Distal tip cell. Scale bar: 20 μ m. **(C)** Correlation of expression (CPM) of known miRNAs (mir-35-3p, mir-1-3p and let-7-5p) and novel miRNA predictions (nov-63-3p and nov-72-3p) with corresponding C_T values measured by TaqMan assay (expression of mature miRNAs). **(D)** Genome browser track showing read coverage of predicted novel miRNA candidate, nov-72-3p with reads stacking up mostly on the mature sequence (light blue) at aligned 5' positions. Forward strand coverage is indicated in dark grey and reverse strand coverage is indicated in light grey. Conservation across different species is displayed at nucleotide resolution. (Diag et al., 2018)

For nov-63-3p we observed gonad specific expression, as this candidate was enriched in the gonad compared to whole worm (Fig. 14A) (Diag et al., 2018). In contrast, nov-72-3p seemed to be enriched in the whole worm compared to the gonad (Fig. 14A). Additionally, we observed high conservation of mature nov-72-3p amongst nematodes and other species. Furthermore, the read coverage was miRNA-like with reads piling up on the mature sequence at the aligned 5' position (Fig. 14D). This was true for all four candidates (appendix, Fig. 32) (Diag et al., 2018).

To examine putative targets of the novel predictions, we used miRNA:mRNA chimeric data published previously by our lab and generated via ALG-1 iPAR-CLIP (Grosswendt et al., 2014). These data were generated using L3 staged worms. Thus, we were not able to detect any putative targets for most of the predicted novel miRNAs, as we used young adults in this study (Diag et al., 2018). Nevertheless, for nov-72-3p we discovered an interesting interaction pattern as it binds other miRNAs generating a miRNA-miRNA duplex (appendix, Fig. 32) (Diag et al., 2018). While the strongest interaction was observed between nov-72-3p and miR-52-5p, we could not detect any significant positive correlation between nov-72-3p and its miRNA-targets (Diag et al., 2018). The miRNA-miRNA duplex probably plays an important role in other tissues of the worm rather than the germline, as we already showed that nov-72-3p is enriched in the whole worm compared to the gonad (Fig. 14A) (Diag et al., 2018). However, we found nov-72-3p, nov-63-3p and nov-82-5p covered by reads from DCR-1 PAR-CLIP data (Rybak-Wolf et al., 2014) and ALG-1 iPAR-CLIP data (Grosswendt et al., 2014) supporting existence and functionality of those (Tab. 1).

Table 1. Putative novel miRNAs. ALG-1 and DCR-1 represent the coverage of novel miRNAs by reads from DCR-1 PAR-CLIP (Rybak-Wolf et al., 2014) and ALG-1 iPAR-CLIP data (Grosswendt et al., 2014). CPMs are listed for mature and star novel miRNA. (Diag et al., 2018)

mature	star	ALG-1	DCR-1	mature CPM	star CPM
nov-1-3p	nov-1-5p	0	0	5712,98	0,68
nov-2-3p	nov-2-3p	0	0	4096,73	0,34
nov-3-5p	nov-3-3p	8	35	831,45	15,21
nov-4-3p	nov-4-5p	1	3	5005,24	0,34
nov-5-5p	nov-5-3p	0	0	896,34	1,35
nov-6-5p	nov-6-3p	1	0	2374,01	37,18
nov-7-3p	nov-7-5p	0	2	681,38	0,68
nov-8-5p	nov-8-3p	0	1	401,19	3,72
nov-9-3p	nov-9-5p	0	2	280,19	1,35
nov-10-3p	nov-10-5p	0	2	437,02	0,68
nov-11-3p	nov-11-5p	2	0	3213,23	0,34
nov-12-3p	nov-12-5p	1	0	376,52	0,34

mature	star	ALG-1	DCR-1	mature CPM	star CPM
nov-13-3p	nov-13-5p	0	11	143,64	7,44
nov-14-3p	nov-14-5p	0	0	132,49	5,07
nov-15-3p	nov-15-5p	0	0	1170,78	0,34
nov-16-5p	nov-16-3p	0	4	353,87	0,34
nov-17-3p	nov-17-5p	0	0	702	1,01
nov-18-3p	nov-18-5p	0	0	171,02	23,32
nov-19-3p	nov-19-5p	0	0	470,81	1,35
nov-20-5p	nov-20-3p	1	5	239,63	0,34
nov-21-3p	nov-21-5p	3	1	634,06	1,69
nov-22-3p	nov-22-5p	1	7	238,96	0,34
nov-23-3p	nov-23-5p	1	0	701,32	0,34
nov-24-3p	nov-24-5p	0	2	66,92	230,17
nov-25-3p	nov-25-5p	1	2	66,92	230,17
nov-26-3p	nov-26-5p	0	2	231,52	0,34
nov-27-3p	nov-27-5p	0	0	432,96	1,69
nov-28-3p	nov-28-5p	0	0	113,23	47,66
nov-29-5p	nov-29-3p	1	0	112,55	1,35
nov-30-3p	nov-30-5p	0	0	855,78	0,68
nov-31-5p	nov-31-3p	0	0	126,07	0,34
nov-32-5p	nov-32-3p	0	1	119,99	0,68
nov-33-3p	nov-33-5p	0	2	95,99	2,03
nov-34-3p	nov-34-5p	2	1	106,13	2,03
nov-35-3p	nov-35-5p	0	0	24,67	37,18
nov-36-3p	nov-36-5p	0	1	103,76	0,68
nov-37-5p	nov-37-3p	1	8	88,55	6,08
nov-38-5p	nov-38-3p	0	3	97,34	0,34
nov-39-5p	nov-39-3p	0	0	90,24	0,68
nov-40-5p	nov-40-3p	0	0	438,71	0,34
nov-41-3p	nov-41-5p	1	3	43,94	2,03
nov-42-3p	nov-42-5p	0	1	65,91	1,01
nov-43-3p	nov-43-5p	0	2	65,23	0,68
nov-44-5p	nov-44-3p	0	0	61,51	1,01
nov-45-5p	nov-45-3p	3	0	83,48	1,35
nov-46-3p	nov-46-5p	1	0	55,09	1,35
nov-47-5p	nov-47-3p	0	0	100,38	2,37
nov-48-3p	nov-48-5p	0	0	54,75	2,7

mature	star	ALG-1	DCR-1	mature CPM	star CPM
nov-49-5p	nov-49-3p	0	0	58,81	1,01
nov-50-5p	nov-50-3p	0	0	60,16	0,68
nov-51-5p	nov-51-3p	0	3	48,33	0,68
nov-52-3p	nov-52-5p	0	1	61,51	1,69
nov-53-3p	nov-53-5p	0	3	28,39	2,03
nov-53-3p	nov-53-5p	0	3	28,39	2,03
nov-54-3p	nov-54-5p	0	0	52,73	0,68
nov-55-5p	nov-55-3p	0	0	28,73	1,01
nov-56-3p	nov-56-5p	0	0	10,82	1,35
nov-57-5p	nov-57-3p	2	3	83,14	0,34
nov-58-3p	nov-58-5p	0	1	20,62	39,21
nov-59-5p	nov-59-3p	0	0	17,91	0,34
nov-60-3p	nov-60-5p	0	1	2044,14	0
nov-61-3p	nov-61-5p	23	29	1341,8	0,34
nov-62-3p	nov-62-5p	6	31	611,42	0
nov-63-3p	nov-63-5p	1	10	60,84	0
nov-64-5p	nov-64-3p	5	7	7,1	0
nov-65-3p	nov-65-5p	0	1	32,11	0
nov-66-3p	nov-66-5p	0	12	37,85	0
nov-67-3p	nov-67-5p	0	0	375,84	148,38
nov-68-3p	nov-68-5p	1	14	76,05	0
nov-69-3p	nov-69-5p	1	3	5041,06	2,03
nov-70-3p	nov-70-5p	1	3	542,81	0
nov-71-3p	nov-71-5p	0	42	6,08	0
nov-72-3p	nov-72-5p	10	20	238,62	0
nov-73-3p	nov-73-5p	0	5	15,55	0
nov-74-5p	nov-74-3p	1	6	57,46	0,34
nov-75-3p	nov-75-5p	1	5	43,94	0
nov-76-3p	nov-76-5p	0	2	313,99	0
nov-77-3p	nov-77-5p	4	3	5941,79	0
nov-78-3p	nov-78-5p	0	0	280,19	25,69
nov-79-5p	nov-79-3p	1	1	1150,5	0
nov-80-3p	nov-80-5p	0	2	28,05	0
nov-81-3p	nov-81-5p	2	0	59,15	0
nov-82-5p	nov-82-3p	37	1	29,07	0
nov-83-3p	nov-83-5p	0	3	2487,24	0

Overall, with our germline-specific small RNA sequencing we predicted a high fraction of novel miRNAs being localized throughout the gonad. Further in-depth studies are necessary to investigate the functionality of these predictions.

3.5 Spatiotemporal gene expression is highly dynamic in the germline

Based on the fact that most of the mRNAs showed a localization pattern along the gonad, we were specifically interested in the mRNAs that localize either to the distal mitotic or to the late meiotic/oogenesis compartment. Investigating these two clusters in detail, we observed that the expression of genes encoding for a ribosomal protein (*rpl* and *rps* genes) was restricted to the distal gonad arm (mitotic region/transition zone) (Diag et al., 2018). Figure 15 shows *rpl-17* as an example. This expression decreased towards the proximal gonad arm (early oogenesis) and was completely absent in the very proximal gonad arm (late oogenesis). I confirmed the expression profile of *rpl-17* with ISH (Fig. 15). This was to some extent expected, as these genes have been previously reported to be involved in cell proliferation (Chen and Ioannou, 1999; Zhang et al., 2014). Additionally, we found other genes that are required for germ cell proliferation, such as *iff-1*, to be localized in the distal region of the gonad (Fig. 10A). In contrast to the *rpl* and *rps* genes, we found genes that predominantly localized to the proximal gonad arm (Diag et al., 2018). Among these genes, we recovered expression patterns of known gene groups, e.g., for GLD-1 targets as well as unknown ones like *pie-1*, *nos-2* and *cey-2* (Fig. 15). The expression of those unknown ones was low in the mitotic region and increased towards the oogenesis region indicating an important role of these in the differentiation process. The expression profiles were consistent with ISH images (Fig. 15). Of note, the ISH image of *pie-1* showed expression throughout the gonad. However, the ISH procedure is not quantitative, obscuring a clear statement about the expression level of *pie-1* in the gonad. Furthermore, the expression level of *pie-1* between the distal and proximal gonad arm was very low (~ 200 CPM).

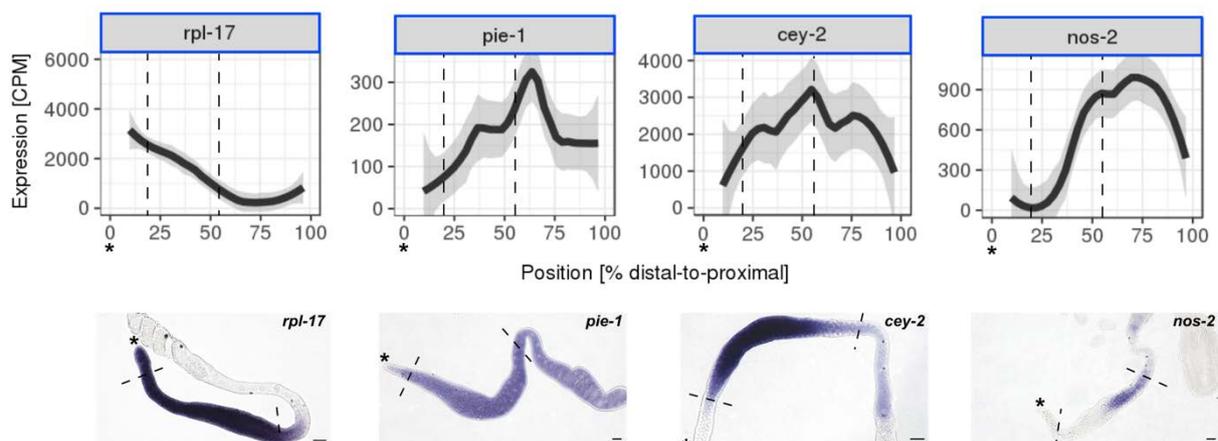


Figure 15. Spatiotemporal gene expression is highly dynamic in the germline. Spatial expression of *rpl-17*, *pie-1*, *cey-2* and *nos-2* in wild type N2 from distal-to-proximal (upper panel). n = 6 independent experiments for N2, LOESS \pm standard error. *In situ* hybridization images of *rpl-17*, *pie-1*, *cey-2* and *nos-2* in wild type N2 (lower panel). Scale bar: 20 μ m. Dashed lines represent different zones in the germline. Asterisk: Distal tip cell. (Diag et al., 2018)

Previous studies showed evidence that marks of actively expressed chromatin are absent from the X chromosomes in proliferating and early meiotic germ cells, but are present during oogenesis (Kelly et al., 2002; Strome et al., 2014). This finding led to the hypothesis that X chromosomes are transcriptionally inactive during all phases of germ cell development except for the oogenesis. To test this hypothesis, we investigated the spatiotemporal expression of X-linked transcripts in the germline. WormBase has over 6,000 annotated X-linked genes and for most of these we observed expression in the proximal gonad arm, *i.e.*, the oogenesis (Fig. 16A). However, we examined a subgroup of X-linked genes that were expressed in the distal gonad arm, *i.e.*, mitotic and early meiotic region (Fig. 16B). Recent studies already speculated that transcription from the X chromosome may occur in germ cells at all stages as primordial germ cells express certain X-linked genes and germlines lacking oogenic germ cells transcribe $\sim 15\%$ of the genes on the X chromosome (Spencer et al., 2011; Tabuchi et al., 2011). Nonetheless, in these studies, a spatiotemporal resolution of the expression from the X-chromosome in the germline was missing.

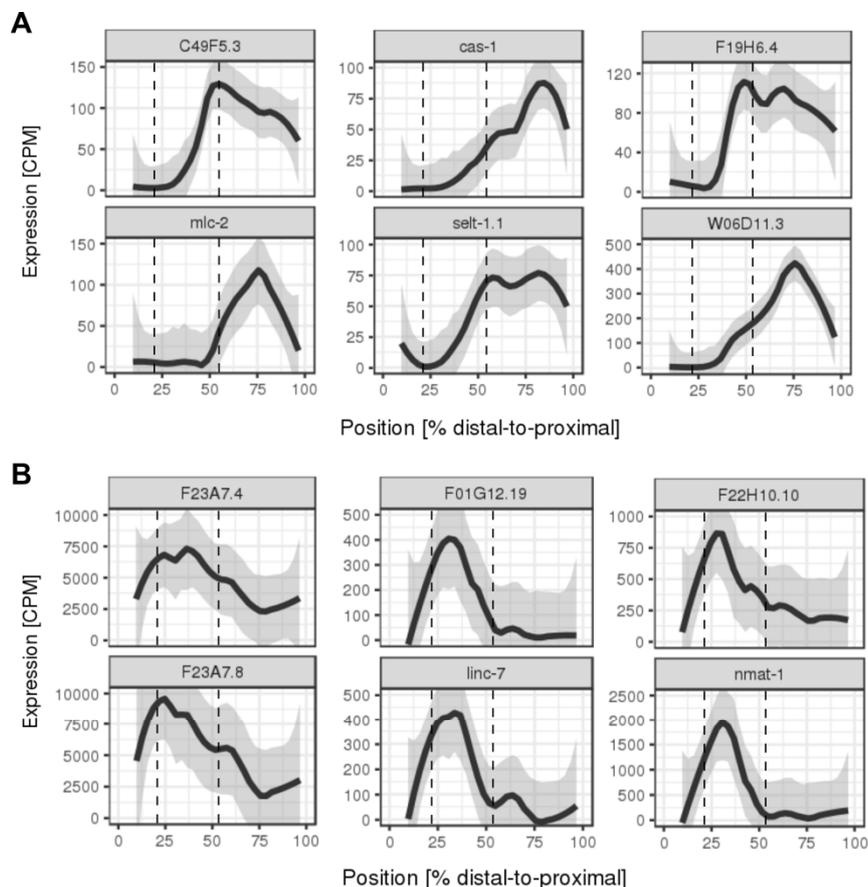


Figure 16. A subgroup of X-linked genes is expressed in the proliferating and early meiotic region. (A) Spatial expression of six X-linked genes (C49F5.3, *cas-1*, F19H6.4, *mlc-2*, *selt-1.1* and W06D11.3) in wild type N2 from distal-to-proximal. **(B)** Spatial expression of six X-linked genes (F23A7.4, F01G12.19, F22H10.10, F23A7.8, *linc-7* and *nmat-1*) in wild type N2 from distal-to-proximal. n = 6 independent experiments for N2, LOESS \pm standard error. Dashed lines represent different zones in the germline.

In summary, our data demonstrated that the spatiotemporal gene expression is highly dynamic in the germline with genes being either expressed in the distal mitotic region or late meiotic/oogenesis region.

3.6 Spatiotemporal gene expression is lost in the germline tumour mutant

To examine whether the mRNA localization is crucial for the proliferation and differentiation balance in the germline, we performed the cryo-cut-sequencing approach for a perturbed system with an impaired proliferation and differentiation balance. We investigated the spatiotemporal gene expression of the double mutant *gld-2 gld-1* which reveals only one third of the meiotic entry resulting in an exclusively proliferating and tumorous germline (Brenner and Schedl, 2016; Fox et al., 2011; Hansen et al., 2004b). Furthermore, GLD-1 and GLD-2 are important regulators of the mitosis-to-meiosis transition (Brenner and Schedl, 2016; Lee and Schedl, 2010; Millonigg et al., 2014; Schmid et al., 2009).

Clustering the expression of germline specific genes (Wang et al., 2009), clustered by wild type N2 data, uncovered that the spatiotemporal gene expression is lost in the *gld-2 gld-1* double mutant (Fig. 17A) (Diag et al., 2018). However, clustering the same genes in the *gld-2 gld-1* mutant revealed that some genes were localized in the germline (appendix, Fig. 33). As expected genes required for proliferation, such as *rpl* and *rps* genes, were constantly expressed throughout the tumorous germline, while genes required for differentiation like *pie-1*, *nos-2* or *cey-2* were downregulated or not expressed at all (Fig. 17B) (Diag et al., 2018). I confirmed this result with ISH images (Fig. 17B) This phenomenon was not observed for the *gld-2* single mutant in accordance with previous studies that *gld-1* and *gld-2* single mutants enter meiosis normally and that meiotic entry was only disturbed in double mutants (Brenner and Schedl, 2016; Hansen et al., 2004b). For the *gld-2* single mutant we observed still high expression of *cey-2* and slight expression for *nos-2*, while *cey-2* expression was downregulated and *nos-2* expression was completely absent in the *gld-2 gld-1* double mutant compared to wild type (Fig. 17B and 17C). Moreover, as expected most of the GLD-1 targets were downregulated in the *gld-2 gld-1* double mutant.

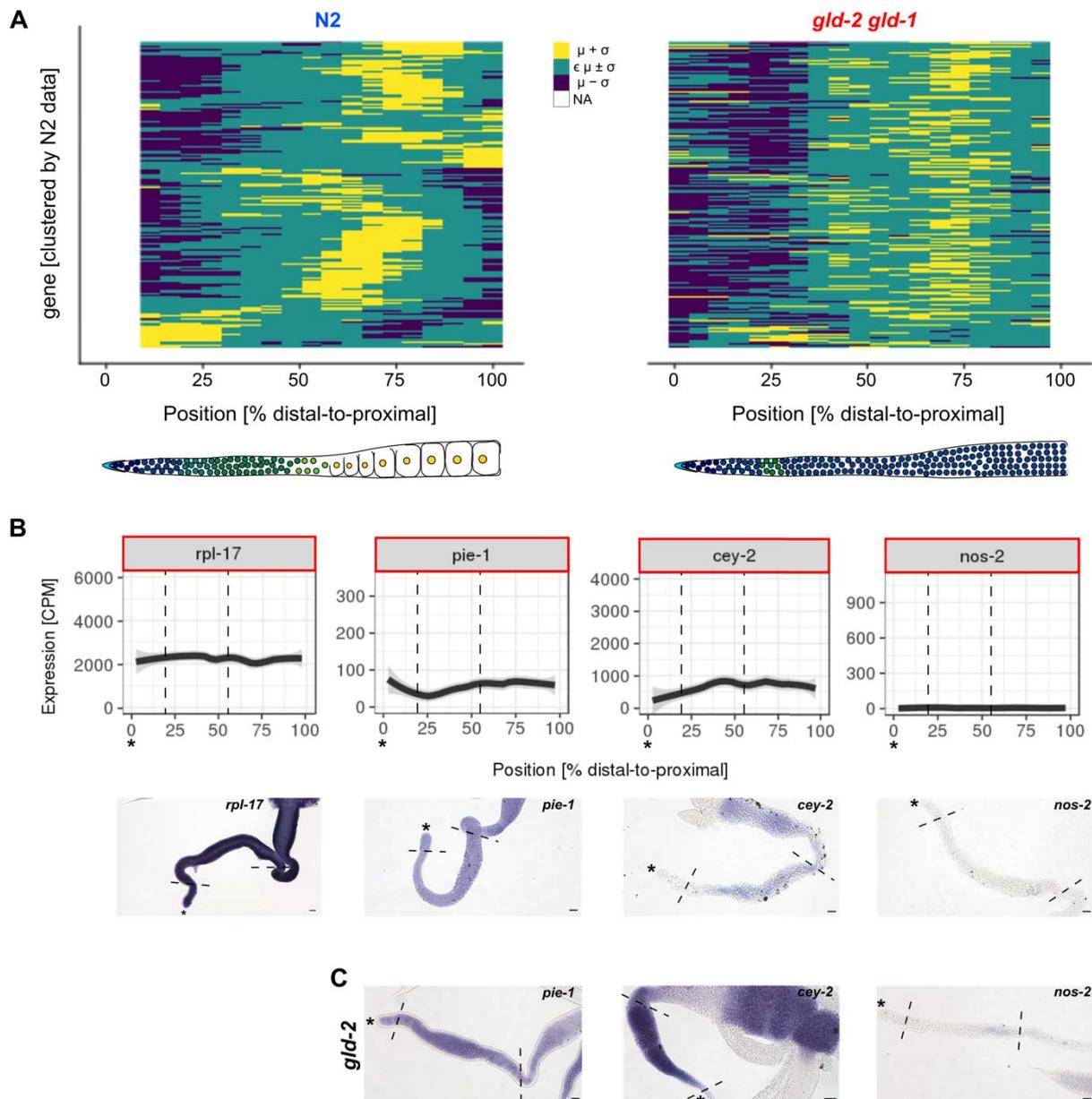


Figure 17. Spatiotemporal gene expression is lost in the germline tumour mutant. (A) Hierarchical clustering of germline specific genes by linear correlation ($1 - \text{Pearson's } r$) for N2 and *gld-2 gld-1* double mutant. μ : Mean; σ : Standard deviation. NA: No data. **(B)** Spatial expression *rpl-17*, *pie-1*, *cey-2* and *nos-2* in *gld-2 gld-1* double mutants from distal-to-proximal (upper panel). $n = 4$ independent experiments for *gld-2 gld-1*, LOESS \pm standard error. *In situ* hybridization (ISH) images of *rpl-17*, *pie-1*, *cey-2* and *nos-2* in *gld-2 gld-1* (lower panel). **(C)** ISH images of *pie-1*, *cey-2* and *nos-2* in *gld-2* single mutant. Scale bar: 20 μm . Dashed lines represent different zones in the germline. Asterisk: Distal tip cell. (Diag et al., 2018)

Above we showed that X-linked genes are dynamically expressed at all stages in the germline. However, sperm-enriched and germ cell-intrinsic genes are under-represented on the X-chromosome leading to the hypothesis that most of the genes required for viability and function of germ cells are not located on the X chromosome but more on the autosomes (Reinke et al., 2000; Strome et al., 2014). To test this hypothesis, we investigated the spatiotemporal expression of the X-linked genes in the *gld-2 gld-1* mutant that lacks a viable and functional germline. Interestingly, the majority of X-linked genes were downregulated in the mutant or

completely absent indicating that these genes are required for proper germline development (Fig. 18). This holds true for the X-linked genes that were expressed in the distal and proximal gonad arm of the wild type (Fig. 16 and Fig 18). However, the X-linked genes that were expressed in the wild type in the distal gonad arm were still expressed at a moderate level in the mutant while genes that were expressed in the proximal gonad arm in the wild type were almost absent in the mutant (Fig.16 and Fig. 18). Kelly and colleagues previously showed that oocyte-enriched genes are found on the X-chromosome at a frequency that is comparable to autosomes. As the *gld-2 gld-1* mutant lacks oocytes, we speculate that the downregulation or absence of X-linked genes could be one reason for this phenotype or a consequence of undifferentiated germ cells.

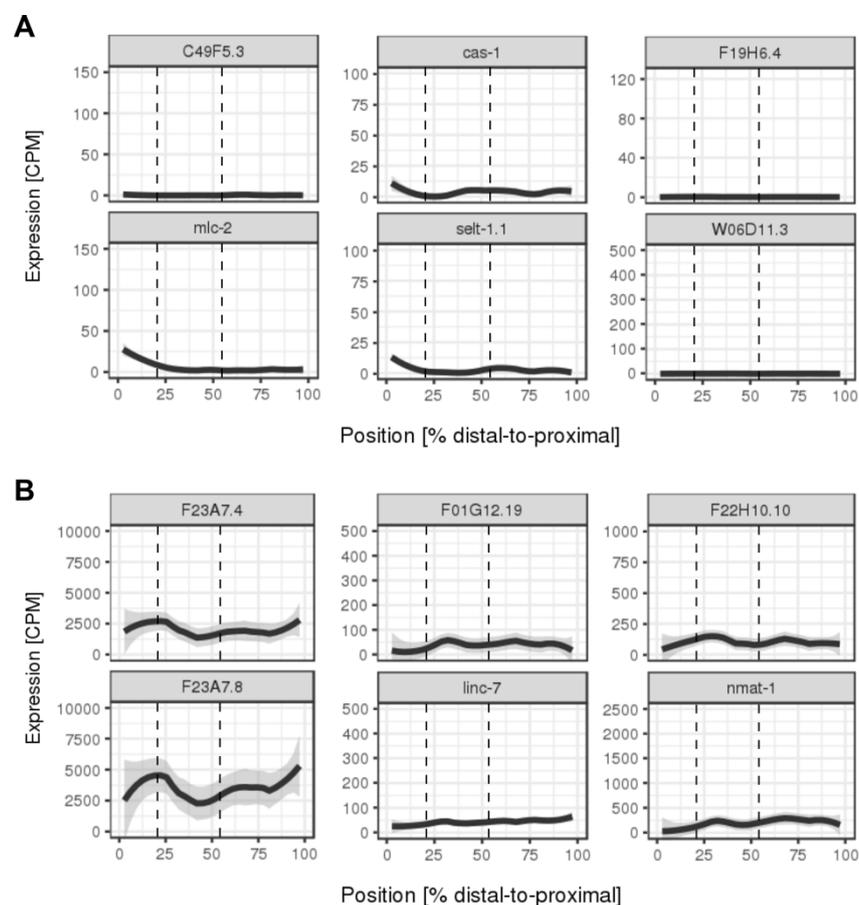


Figure 18. X-linked genes are downregulated in the *gld-2 gld-1* double mutant. (A) Spatial expression of six X-linked genes (C49F5.3, *cas-1*, F19H6.4, *mlc-2*, *selt-1.1* and W06D11.3) in *gld-2 gld-1* mutant from distal-to-proximal. **(B)** Spatial expression of six X-linked genes (F23A7.4, F01G12.19, F22H10.10, F23A7.8, *linc-7* and *nmat-1*) in *gld-2 gld-1* mutant from distal-to-proximal. n = 4 independent experiments for *gld-2 gld-1*, LOESS \pm standard error. Dashed lines represent different zones in the germline.

Additionally, we dissected the gonad of the *glp-1 gain of function (gf)* mutant which possesses a prolonged mitotic region resulting in a larger proliferative zone (reviewed in Hansen and Schedl, 2013). The mutant is temperature sensitive resulting in an inducible tumorous

phenotype (Diag et al., 2018). However, the tumour was located around the loop region impeding the dissection and embedding of these gonads. Hence, we induced the phenotype for a short time in order to avoid tumour formation (Diag et al., 2018). This led to the result that the spatiotemporal gene expression resembled the wild type animal (appendix, Fig. 33B) (Diag et al., 2018). Hence, we did not follow up on this mutant.

Overall, we showed that the spatiotemporal gene expression is important for the proliferation and differentiation balance, as the spatiotemporal gene expression is lost in a solely proliferating and tumorous germline mutant.

3.7 Hundreds of 3' UTR isoforms extended in the germline

Previous studies showed that mature germ cells are transcriptionally silent from the late stage oocytes (diakinesis) to the 4-cell stage embryos, where also the oocyte-to-embryo (OET) transition takes place (Evsikov et al., 2006; Stoeckius et al., 2014). Thus, posttranscriptional regulation, which is known to control gene expression in early embryogenesis, has been suggested to play a major role particularly in this region. We already indicated in this study that miRNAs are important posttranscriptional regulators of gene expression in the germline, particularly the miR-35 family that co-localizes with its targets around the late pachytene/diakinesis stage suggesting an *in vivo* interaction. Besides miRNAs being important posttranscriptional regulators, Merritt and colleagues previously showed that 3' untranslated regions (UTRs) are the main determinants of gene expression in the *C. elegans* germline (Merritt et al., 2008). Thus, we focussed in this study on different 3' UTR isoforms and their ability to control mRNA localization in the germline.

For this purpose, we used the advantage of our approach that enables sequencing of transcripts from the 3' UTR end. As expected most of the reads mapped to genomic loci annotated as 3' ends of protein coding genes (Diag et al., 2018). However, we also observed coverage downstream of the annotated 3' UTR ends (Fig. 19) (Diag et al., 2018).

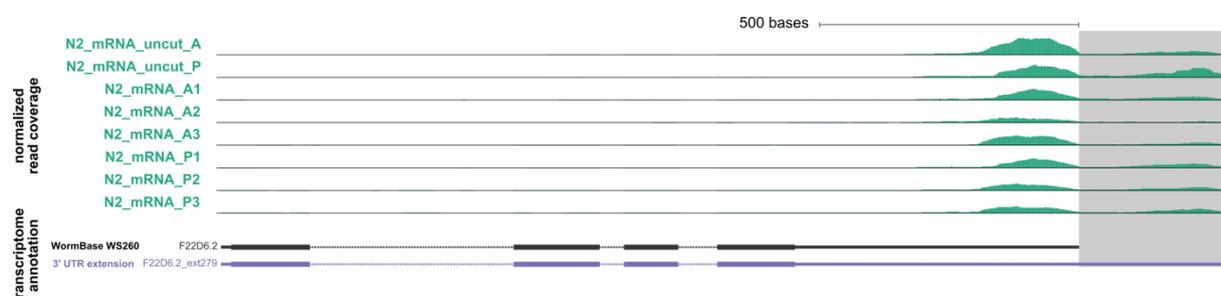


Figure 19. Some genes show coverage downstream of the annotated 3' UTR end. Genome browser track example gene (F22D6.2) with downstream extension of the annotated 3' UTR. (Diag et al., 2018)

This led us to hypothesize that specifically these genes have a longer 3' UTR. Thus, we extended the 3' UTR annotation for 499 genes where we detected an intergenic peak

(coverage) less than 10 kb downstream of an upstream gene (Diag et al., 2018). The median length of a 3' UTR is around 130 nt to 140 nt and the majority of 3' UTRs are not longer than 600 nt in *C. elegans* (Jan et al., 2011; Mangone et al., 2010). Thus, we discarded all 3' UTR extensions above 3 kb for downstream analysis, leaving 419 candidates considered as novel 3' UTRs (Fig. 20A). Out of these, I randomly chose 13 candidates and aimed to validate them by nested or conventional PCR (see Methods) (Fig. 20B). I was able to validate nine out of the 13 candidates (Fig. 20C, 20D, 20E, 20F, 20G and 20H). Additionally, I confirmed their existence by Sanger-sequencing. In two cases I could not validate the isoform of the original 3' UTR annotation (WS260) suggesting that the previous annotation is incorrect and only our extended 3' UTR isoform exists (Fig. 20D and 20E).

Overall, we detected hundreds of novel 3' UTR isoforms, validated nine of them, and extended the 3' UTRome of the *C. elegans* germline.

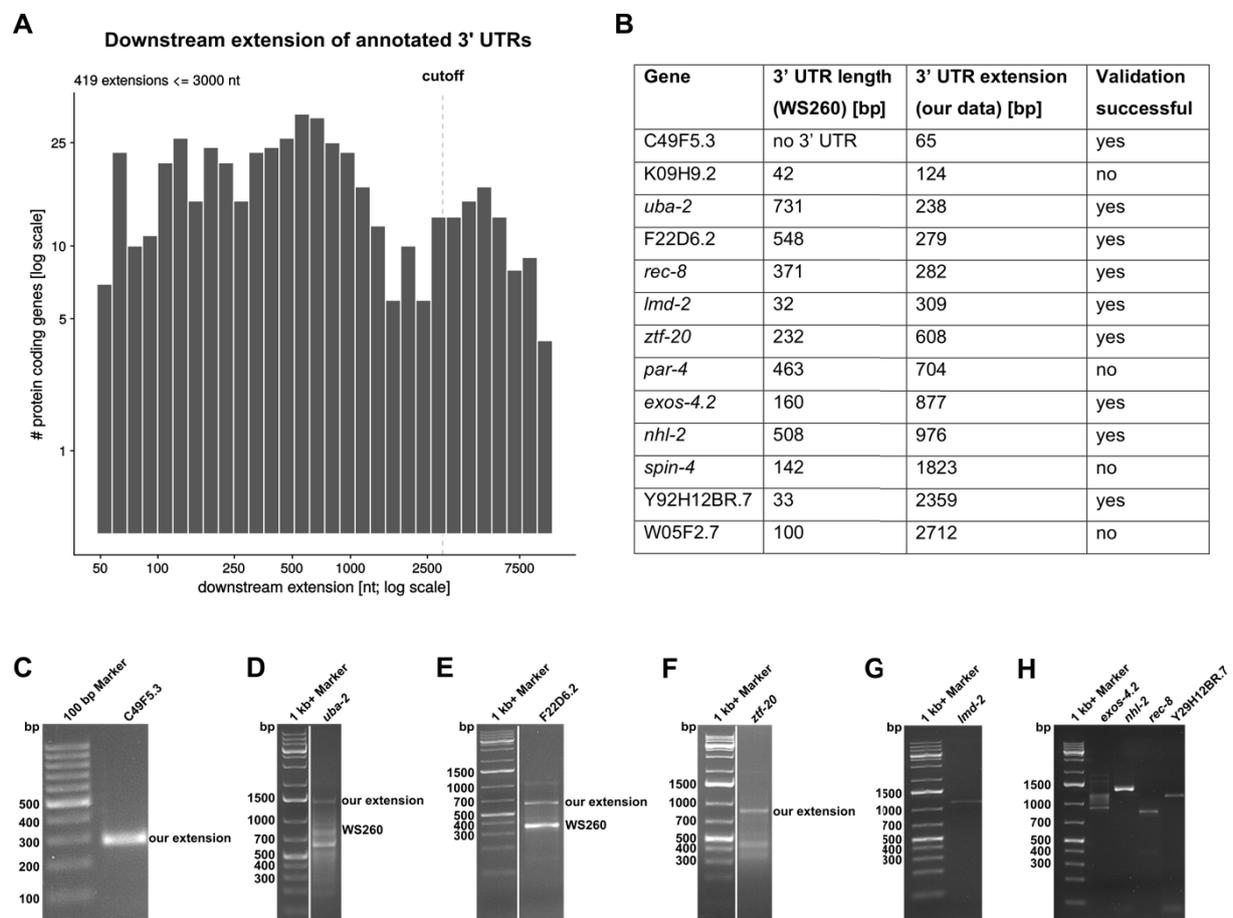


Figure 20. Hundreds of 3' UTR isoforms are extended in the germline. (A) Summary of all downstream extensions of annotated 3' UTRs. Only candidates with an extension smaller or equal to 3000 nt were considered for further analysis and validation. (B) Table summarizing candidates that were chosen for downstream extension validation with annotated 3' UTR length (WS260), downstream extension and result of validation. (C) Validation of downstream extension of C49F5.3, (D) *uba-2*, (E) F22D6.2 and (F) *ztf-20* by nested PCR. Marker: 100 bp or 1 kb+ gene ruler. Our extension and WS260 (if existent) annotation are indicated. White line represents eliminated irrelevant lanes from the gel of candidates that could not be validated (empty lanes). (G) Validation of downstream

extension of *lmd-2*, (H) *exos-4.2*, *nhl-2*, *rec-8* and Y29H12BR.7 annotated 3' UTR by conventional PCR. Marker: 1 kb+ gene ruler. (Diag et al., 2018)

3.8 3' UTR choice is spatially regulated in the germline

We next investigated if different 3' UTR isoforms of the same gene are differentially used in the germline, as 3' UTRs are important regulators of subcellular mRNA localization. Therefore, we quantified the change of the (relative) 3' UTR usage along the germline for 910 genes with at least 2 isoforms (Diag et al., 2018). Intriguingly, we observed a set of genes that mainly used the distal polyadenylation signal (PAS) in the distal gonad arm resulting in the usage of the longer 3' UTR (Diag et al., 2018). The proximal PAS was used preferentially in the proximal arm resulting in the usage of the shorter 3' UTR (Fig. 21). The switch between the usage of distal and proximal PAS happened around the loop region (late pachytene). This phenomenon was not observed when we investigated the gene level summed up over all isoforms.

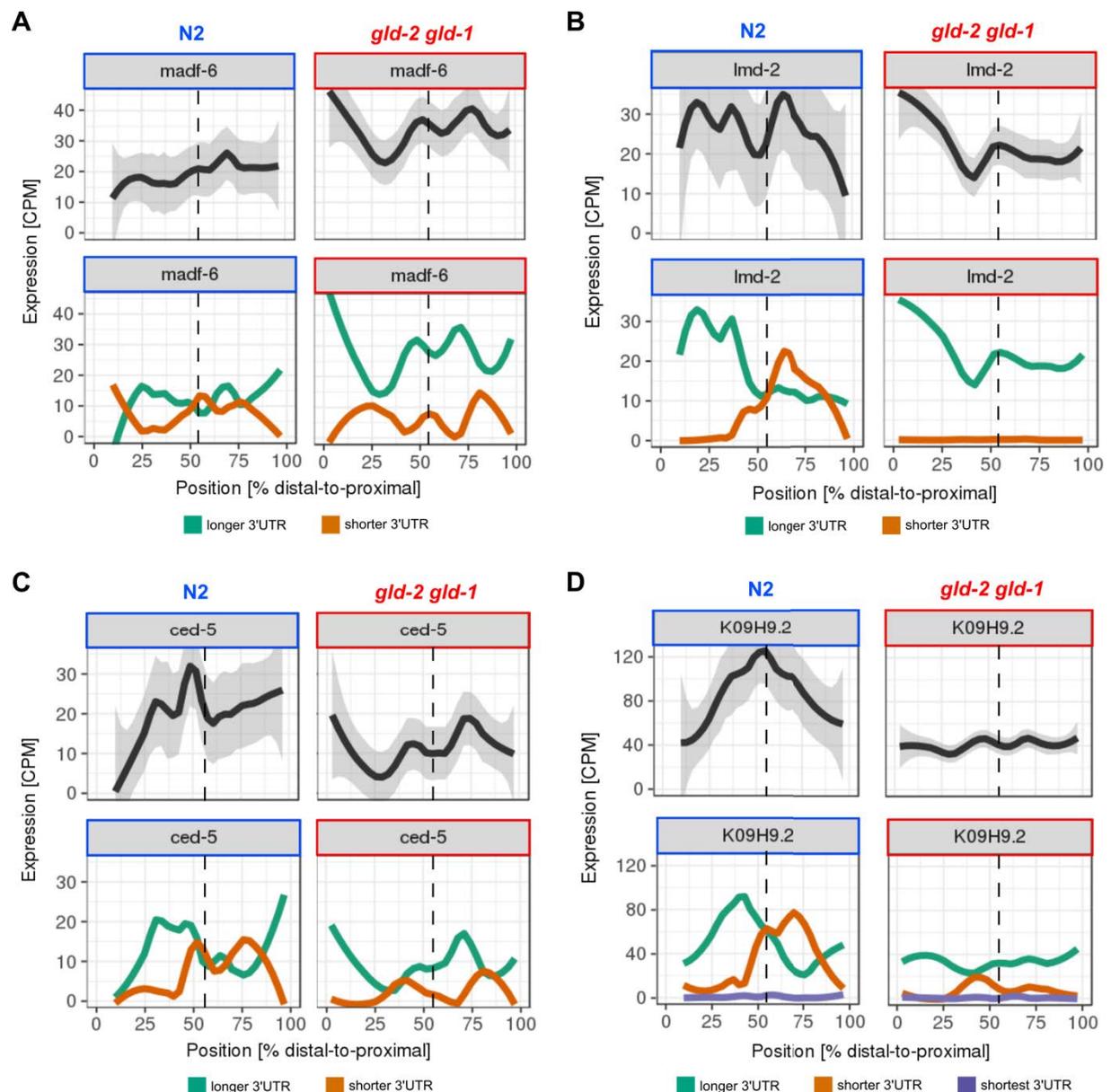


Figure 21. 3' UTR choice is spatially regulated in the germline. (A) Spatial expression of *madf-6*, (B) *lmd-2*, (C) *ced-5*, (D) and K09H9.2 in wild type N2 and *gld-2 gld-1* double mutant from distal-to-proximal at gene and isoform level. $n = 6$ independent experiments for N2 and $n = 4$ for *gld-2 gld-1*, LOESS \pm SE for gene level and LOESS only for isoform level. Longest 3' UTR is marked in turquoise, shorter 3' UTR in orange and the shortest 3' UTR in purple. Dashed line, bend/loop region of the germline. (Diag et al., 2018)

3.9 Differential 3' UTR isoform usage is perturbed in *gld-2 gld-1* double mutants

To further explore the regulatory potential of 3' UTRs, we quantified the (relative) 3' UTR usage along the germline for the same 910 genes with at least 2 isoforms for the tumorous *gld-2 gld-1* double mutant. As observed for the spatiotemporal gene expression the differential 3' UTR usage was significantly impaired in the *gld-2 gld-1* mutants (Fig. 21). In particular, the 3' UTR isoform usage switch from distal to proximal PAS was completely absent in the mutant. Instead, the distal PAS was used throughout the mutant germline, resulting in the expression of solely the long isoform (Fig. 21) (Diag et al., 2018). Moreover, the mutant showed less 3' UTR variability, e.g., most of the 910 genes did not switch the isoform usage along the gonad (Fig. 22A).

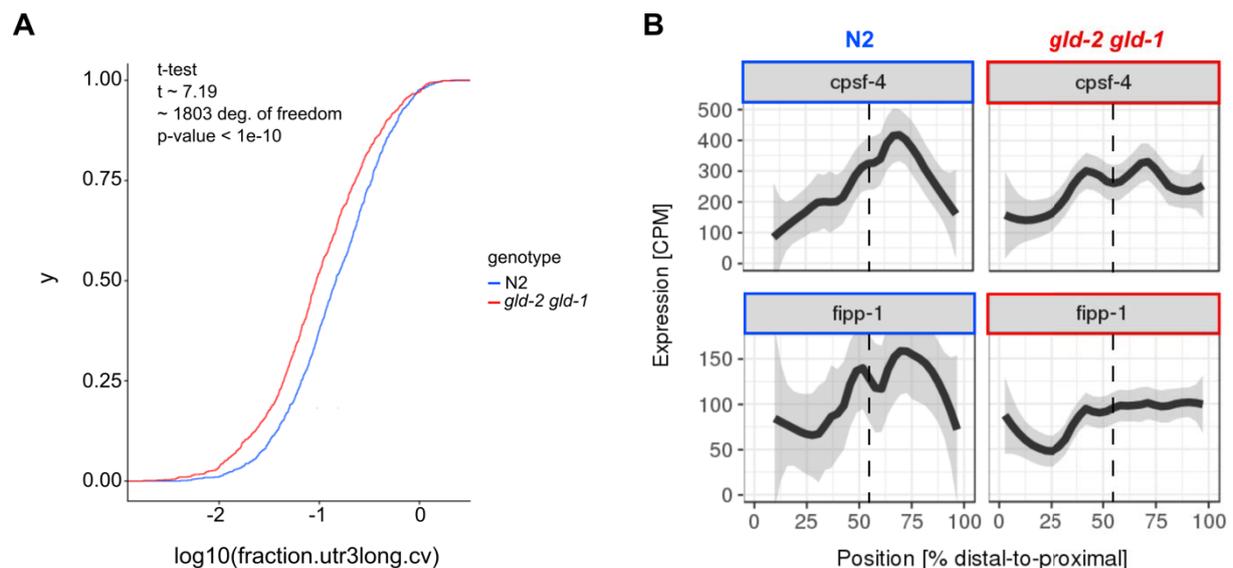


Figure 22. Differential 3' UTR isoform usage is perturbed in the *gld-2 gld-1* double mutants. (A) Comparison of the cumulative densities of 3' UTR variability distribution between N2 (blue line) and *gld-2 gld-1* double mutant (red line). 3' UTR variability was measured by the coefficients of variation (CV) of the contribution of longer 3' UTR to the total expression of the top two expressed (on average) isoforms per gene. 919 genes with several isoforms expressed at 5 CPM or higher on average in either condition were considered for the analysis. Nine genes with CV's below the 0.1st percentile of the log normal fit in either condition were excluded from the analysis. (B) Spatial expression of *cpsf-4* and *fipp-1* in wild type N2 and *gld-2 gld-1* double mutant from distal-to-proximal. $n = 6$ independent experiments for N2 and $n = 4$ for *gld-2 gld-1*, LOESS \pm SE. Dashed line marks the bend/loop region of the germline. (Diag et al., 2018)

The results above indicate that alternative polyadenylation (APA) plays a crucial role in maintaining the tissue integrity of the germline. Hence, we examined factors that are important for APA in more detail. We encountered two factors that were previously reported to be

involved in APA regulation, Fip1 and CPSF-4 (Lackford et al., 2014). Fip1 (*C. elegans* FIPP-1) and CPSF-4 are components of the cleavage and polyadenylation complex that recognize the PAS and interact with the poly(A) polymerase and other factors (Diag et al., 2018; Kaufmann et al., 2004). We found both factors to be expressed in the wild type germline with increasing expression level towards the oogenesis region (Fig. 22B). The maximal expression level was detected around the loop region. Interestingly, the expression level of both factors did not increase towards the oogenesis region in the *gld-2 gld-1* double mutant germline and stayed constant around the loop region (Fig. 22B). Considering that the 3' UTR isoform usage switch occurred around the loop region in the wild type germline but was absent in the *gld-2 gld-1* mutant germline, we suggested that the 3' UTR switch is dependent on the level of *fipp-1* and *cpsf-4*. With increasing expression level of APA factors, the weaker, proximal PAS was used resulting in the usage of the short isoform. As the expression level of these factors did not increase in the *gld-2 gld-1* mutant, only the stronger, distal PAS was used resulting in the usage of the long isoform.

Altogether, our data indicated that 3' UTR choice and length are regulated in the germline depending on the level of APA regulating factors.

3.10 Other small RNAs are spatially organized in the germline

The SMARTer smRNA-seq used in this study abolishes adapter ligation and, therefore, does not rely on a 5' monophosphate or a 3' hydroxyl group for small RNA capture. Hence, it enables capturing various classes of RNAs which do not possess these chemical properties. This was of particular interest as two additional small RNA classes apart from miRNAs play a crucial role in posttranscriptional regulation in the germline: the endogenous small interfering RNAs (endo-siRNAs) and the PIWI-interacting RNAs (piRNAs).

Endo-siRNAs are small RNAs of a length of 22 nt or 26 nt showing a 5' guanosine bias and are, thus, named as 22G and 26G RNAs, respectively (Ruby et al., 2006). These RNAs are mainly antisense to annotated protein-coding genes and are processed in a Dicer-independent manner. Instead, they are a product of transcription of spliced mRNA templates by an RNA-dependent RNA polymerase (RdRP) (Ruby et al., 2006). Endogenous siRNAs are primarily required for transposon silencing. However, some of the siRNAs, especially 26G siRNAs, can silence sperm-enriched genes, initiating the transition to oogenesis (Ruby et al., 2006). In order to test if we truly capture siRNAs with our experimental approach, we investigated the relative antisense coverage for all genes and for the protein-coding genes as well as the size distribution of the detected siRNAs. As expected, the majority of protein-coding genes seemed to be covered by siRNAs more antisense than sense suggesting that siRNA production is widespread (Fig. 23A). Additionally, we observed that 75 % of the antisense alignments were equal or less than 22 nts (median is 21 nts). This percentage went up to 84 % for reads less

or equal to 26 nts supporting our hypothesis that we truly capture 22G and 26G siRNAs. Apart from protein-coding genes being covered by siRNAs we also detected miRNAs, lincRNAs (long non-coding RNAs) and ncRNAs (non-coding RNAs) covered antisense by siRNAs. Interestingly, we observed protein-coding genes with a dominant sense exon coverage, thereby evading siRNA regulation (Fig. 23B). We functionally annotated these genes using DAVID (Huang et al., 2009b, 2009a) and found pathways, such as ribosomal small subunit assembly, translation, and nematode larval development to be significantly enriched (Fig. 23B). Moreover, processes involved in embryo development ending in birth or egg hatching were significantly enriched suggesting that genes required for embryogenesis escape inhibition by siRNAs.

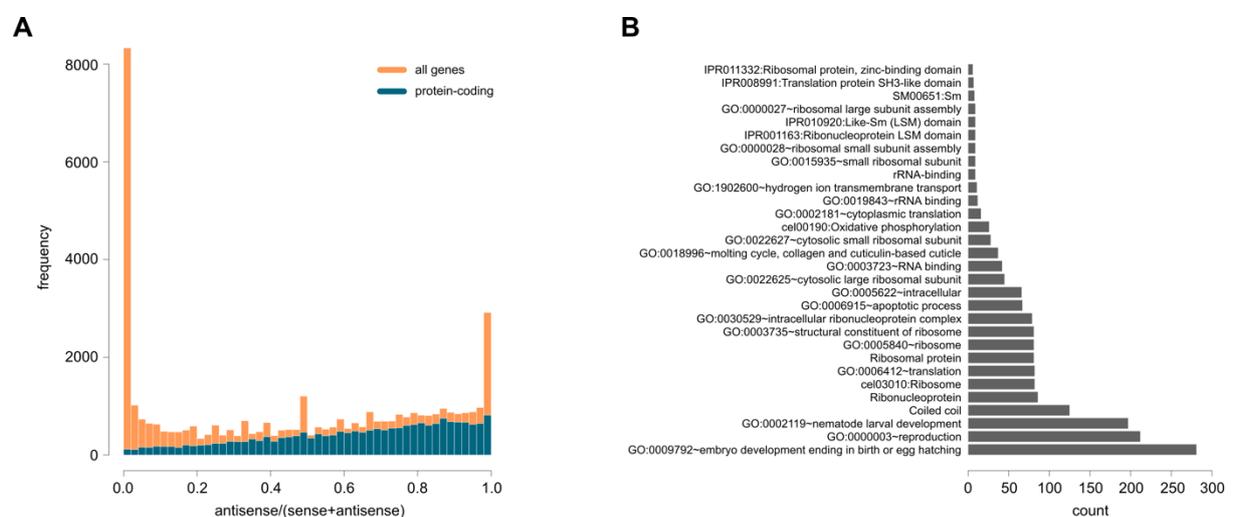


Figure 23. siRNAs are mainly antisense to annotated protein-coding genes but some protein-coding genes evade siRNA regulation. (A) Distribution of the relative antisense coverage for all genes and for protein-coding genes, only. **(B)** Functional annotation of 1213 protein-coding genes using DAVID (Huang et al., 2009b, 2009a). These genes possess a dominant sense exon coverage below the 20th percentile of the negative scores and, thus, evade siRNA regulation.

Another class of small RNAs exclusively expressed in the germline are piRNAs, which silence transposons and exogenous transgenes. In the *C. elegans* germline, piRNAs are exactly 21 nt long single-stranded RNAs with a 5' U bias and are, thus, termed 21U RNAs (Ruby et al., 2006). In contrast to siRNAs, the majority of piRNAs are transcribed from two distinct clusters on chromosome IV and are sense-oriented (Ruby et al., 2006). After 5' and 3' end processing they are bound by the single functional *C. elegans* Piwi homolog PRG-1. Subsequently, PRG-1:piRNA interaction with its targets leads to the generation of secondary siRNAs by RdRPs, also known as 22G RNAs, inducing further gene silencing (Weick and Miska, 2014).

Therefore, we investigated whether we capture *bona fide* the germline specific piRNAs in our data. This was not trivial, as for some piRNAs an enzymatic treatment is necessary to remove the phosphate in order to allow efficient polyadenylation and subsequent capture of piRNAs.

We isolated reads from the small RNA alignments of lengths 21 nt, 22 nt, and 26 nt containing 5' U, G, and G nucleotides, respectively, and computed feature coverage with HTseq (Anders et al., 2015). We observed that sense 21U-RNAs overwhelmingly covered annotated piRNAs (Fig. 24A), whereas their antisense counterparts covered protein-coding genes. This verified that we captured the majority of piRNAs with our experimental approach. Furthermore, we found a large portion of sense 26G RNAs covering tRNAs (transfer RNAs) and a large portion of sense 22G RNAs covering pseudogenes. Additionally, we did not observe any differences in the sense and antisense coverage of RepeatMasker (developed by Smit, A.F.A., Hubley, R., and Green P. (2013-2015); see <http://www.repeatmasker.org/>) transposable elements across the three RNA species (Fig. 24B).

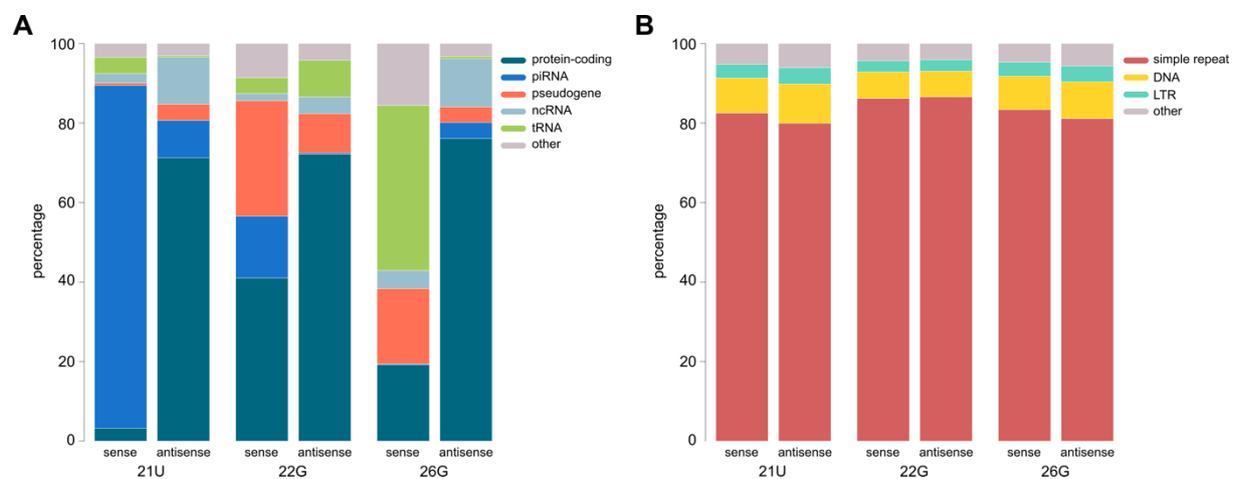


Figure 24. Sense 21U-RNAs cover annotated piRNAs whereas their antisense counterparts cover protein coding genes. (A) Isolated small RNA alignments of lengths 21 nt, 22 nt and 26 nt with their 5' nucleotide being U, G and G, respectively. Using these alignments, the feature coverage was computed with HTseq (Anders et al., 2015) showing the percentage of hits across main feature types for the three RNA species in the sense and antisense direction. **(B)** Sense and antisense coverage of RepeatMasker (developed by Smit, A.F.A., Hubley, R., and Green P. (2013-2015); see <http://www.repeatmasker.org/>) transposable elements across the three RNA species.

After confirming that we accurately capture siRNAs and piRNAs, we investigated the localization patterns of these two RNA species throughout the gonad. Interestingly, both RNA classes displayed distinct localization patterns (Fig. 25). While siRNAs were mainly localized to the early and late meiotic stage, piRNAs seemed to localize more distally in the mitotic region including the transition zone (Fig. 25).

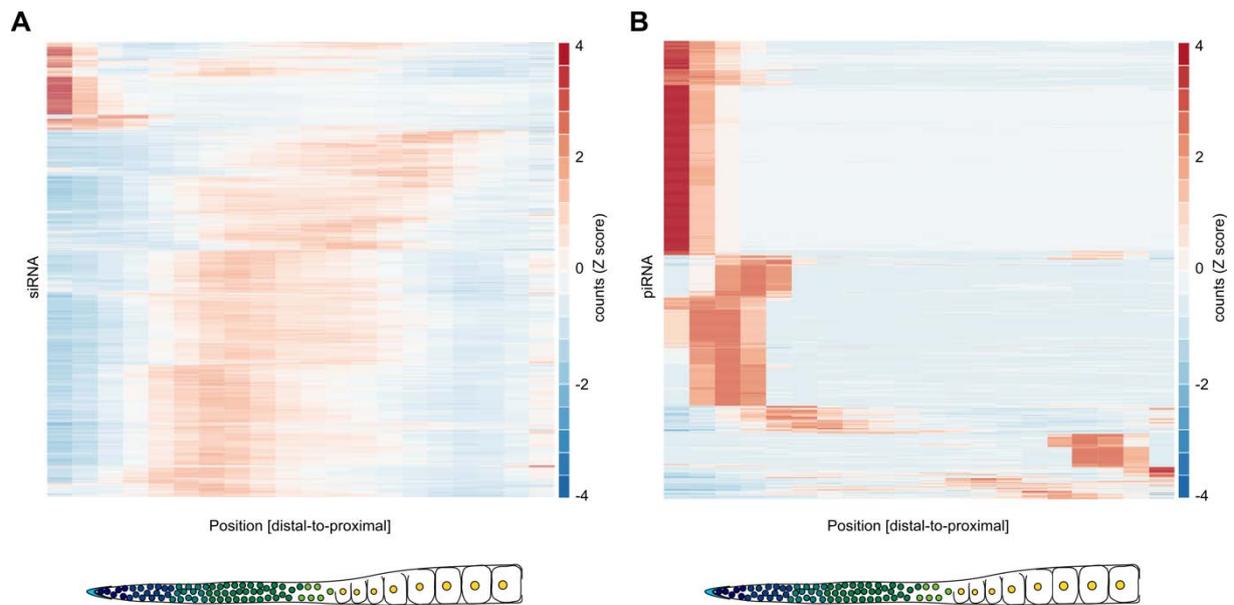


Figure 25. piRNAs and 22G siRNAs are spatially organized throughout the gonad. (A) Hierarchical clustering of detected siRNAs based on Euclidean distances of antisense siRNA z-scores across the gonad. (B) Hierarchical clustering of all detected piRNAs based on Euclidean distances of antisense siRNA z-scores across the gonad. The small RNA expression across slices and replicates were smoothed using LOESS smoothing. The smoothed values were centred and scaled converting them to z-scores

Overall, we demonstrated that in addition to mRNAs and miRNAs, we also captured other small RNA classes which revealed regular localization patterns throughout the gonad.

3.11 A physical gonad model represents RNA localization in the germline

In this study, we generated a data set that contains spatiotemporal mRNA and small RNA expression in the germline at near single-cell resolution. Hence, we thought it would be helpful to build a physical model of the germline that can serve as a universal framework on which these data can be visually displayed and compared. For this purpose, we systematically collected all data available from the literature containing information about the size, composition, and germ cell migration speed of each compartment in the germline (Brenner and Schedl, 2016; Diag et al., 2018; Fox et al., 2011; Hansen and Schedl, 2013; Hirsh et al., 1976; Hubbard, 2007; Maciejowski et al., 2006; Wolke et al., 2007). Furthermore, we used our own ISH images to determine the size and diameter of the gonad. Combining the data from the literature with our own measurements we were able to construct an *in silico* 3D physical gonad model integrating *in vivo* mRNA and miRNA expression throughout germ cell proliferation and differentiation (Methods and Fig. 26A, 26B and 26C) (Diag et al., 2018).

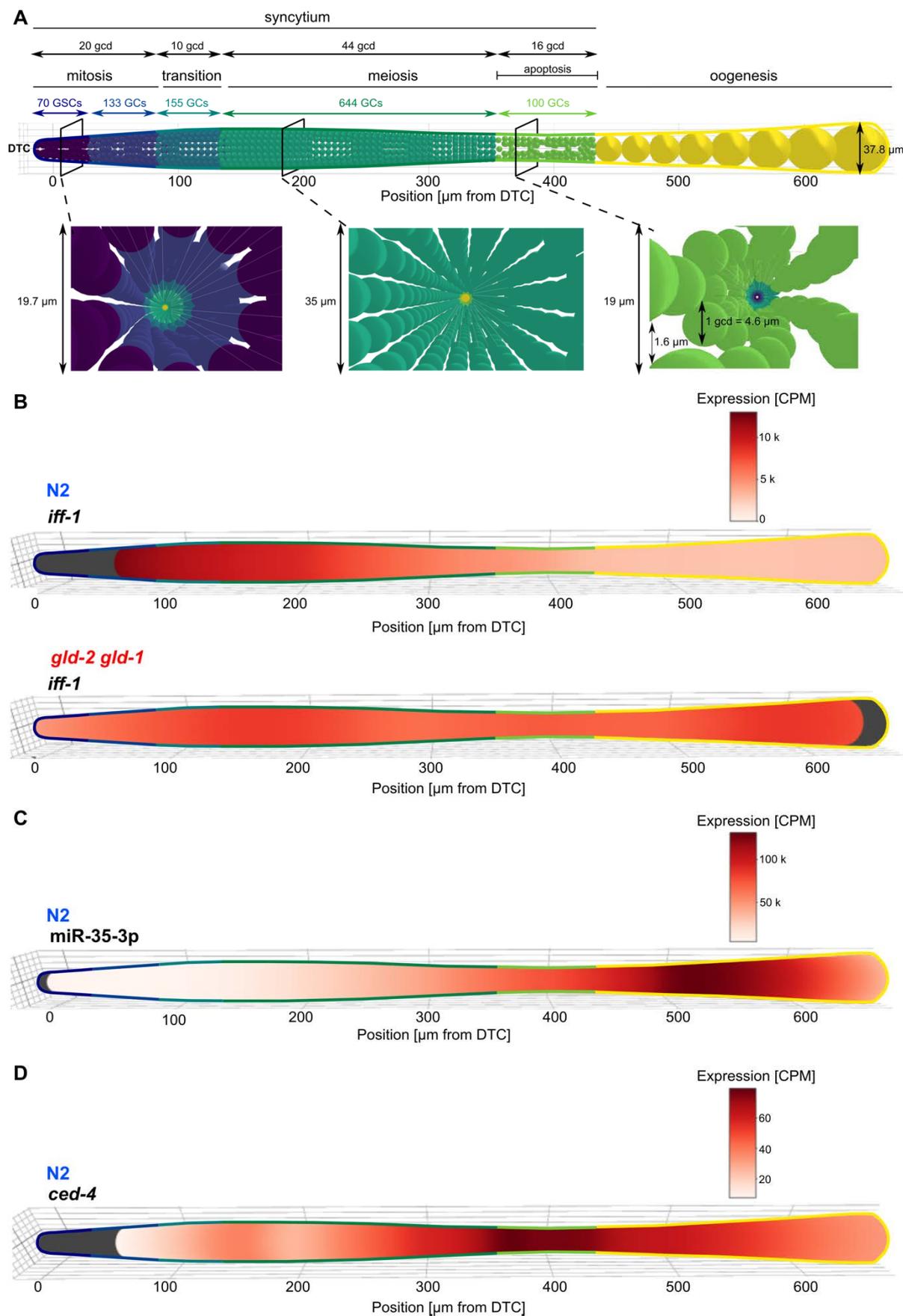


Figure 26. A physical gonad model represents RNA localization in the germline. (A) 3D germline model with assigned sizes of each zone in germ cell diameter (gcd) and corresponding germ cell (GC) numbers. Three cross

sections are shown at 70 μm , 200 μm and 380 μm from the distal tip cell (DTC). **(B)** 3D germline model representing *in vivo* expression of *iff-1* in N2 and *gld-2 gld-1* double mutant. **(C)** 3D germline model representing *in vivo* expression of miR-35-3p in N2. **(D)** 3D germline model representing *in vivo* expression of *ced-4* in N2. Grey: No data. (Diag et al., 2018)

The model contains the size of each zone in the germline with corresponding germ cell number. Thus, we used this model to assign the different zones to our expression profiles and ISH images (Fig. 10, 14-18 and 21-22) (Diag et al., 2018). Additionally, the model represents *in vivo* mRNA expression in the tumorous *gld-2 gld-1* double mutant (Fig. 26B) (Diag et al., 2018). In order to validate our model, we looked up apoptotic gene markers whose expression are known in the germline. Most of the germ cells undergo apoptosis around the bend region implying a high expression of apoptotic genes (Gartner et al., 2008). Indeed, we found *ced-4*, one of the main genes of the core apoptotic machinery, to have its highest expression around the bend region (Fig. 26D) (Diag et al., 2018). The bend region starts approximately at a distance of 350 μm from the distal tip cell which is in accordance with our assignment of the bend region in the physical gonad model. The physical gonad model, virtual *in situ* hybridizations (vISH), and the spatiotemporal resolved data for mRNAs and miRNAs generated in this study are provided as an interactive data visualization tool called SPACEGERM (Spatial C. elegans germline expression of mRNA and miRNA) at <https://shiny.mdc-berlin.de/spacegerm/> (Diag et al., 2018).

In summary, we believe that our 3D physical gonad model and especially SPACEGERM will serve as a useful resource for further studies.

4 Discussion

In this study, we created a spatiotemporal RNA expression map of the *Caenorhabditis elegans* germline by dissecting, shock-freezing, and cryo-cutting the gonad at 50 μm resolution, and sequencing each slice separately (Diag et al., 2018). This revealed new potential mechanisms of RNA localization and function during cell proliferation and differentiation.

4.1 mRNAs and small RNAs are spatiotemporally organized in the germline

The development of the *C. elegans* germline and its classification in proliferating and differentiating cells is largely characterized by spatiotemporal expression of certain proteins such as GLP-1, GLD-1 or GLD-2 (Crittenden et al., 2006; Nusch and Eckmann, 2013; Schmid et al., 2009). Although previous studies provided evidence that individual 3' UTRs and small RNAs, such as miRNAs, play crucial roles in maintaining the germline morphology and integrity, a global spatiotemporal mRNA and small RNA transcriptome of the germline has heretofore not been generated (Bukhari et al., 2012; Merritt et al., 2008). This was mainly due to the lack of techniques that enable sequencing of extremely low input material. Single-cell techniques are not suitable for the germline, as most of the germ cells are connected via a syncytium that impedes the dissociation of germ cells into single cells. Therefore, I adapted and optimized a cryo-cut-based sequencing technique, called tomo-seq, and combined it with a single-cell protocol for low amounts of RNA (CEL-Seq1/2) that allows the determination of the spatiotemporal RNA expression in the germline (Diag et al., 2018; Hashimshony et al., 2012, 2016; Junker et al., 2014). I dissected single gonads, and embedded them in a tissue-freezing medium, followed by rapid freezing and cryo-cutting at 50 μm resolution. Subsequently, each slice was sequenced separately and the spatiotemporal distribution of RNAs was analyzed computationally. We showed that our method is reliable and reproducible and that anterior and posterior gonad arms do not differ in RNA expression (Fig. 6, 7, 8 and 9).

With our cryo-cut-sequencing method we showed that mRNAs and small RNAs, such as miRNAs, siRNAs and piRNAs are organized in distinct localization patterns (Fig. 10, 11, 15 and 27). We recovered known and novel expression profiles of gene groups. *Rpl* and *rps* genes that encode ribosomal proteins localized to the distal proliferative part of the gonad arm (Fig. 15). This was expected as *rpl* and *rps* genes are known to be transcribed and required for proliferation (Chen and Ioannou, 1999). In contrast, X-linked genes lack histone posttranslational modifications associated with active transcription in the proliferative and early meiotic region of the germline and are therefore expressed proximally in the oogenic region (Kelly et al., 2002; Strome et al., 2014; Xu et al., 2016) (Fig. 16). However, we found a fraction of X-linked genes to be expressed in the proliferative region indicating that these genes probably escape chromosome-wide silencing (Fig. 16). This was previously suggested by

Spencer and colleagues and Tabuchi and co-workers without knowing the exact spatial distribution of the X-linked transcripts in the germline (Spencer et al., 2011; Tabuchi et al., 2011). Spencer and colleagues observed that primordial germ cells express a fraction of X-linked genes while Tabuchi and co-workers showed that ~ 15 % of X-linked genes are transcribed in germlines lacking oogenic germ cells (Spencer et al., 2011; Tabuchi et al., 2011). Supporting their observations, we found a subgroup of X-linked genes to be expressed in the tumorous *gld-2 gld-1* double mutant that lacks oocytes (Fig. 18). Nevertheless, one cannot exclude active transport of the encoded transcripts to the distal gonad arm, since we only measured mRNA levels at steady-state in each zone of the germline. Furthermore, we observed that not only mRNAs, but also miRNAs, siRNAs, and piRNAs that are important for posttranscriptional regulation in the germline were spatiotemporally organized (Fig. 11 and 25). Moreover, the germline-specific miRNA family miR-35 was co-localized with its predicted corresponding mRNA targets, strongly suggesting *in vivo* interaction of these two classes in the germline (Fig. 12). In contrast, a non-germline-specific miRNA family, miR-1, did not show any prominent co-localization pattern with its predicted targets (Fig.12). We hypothesize that the miR-35 family may have a threshold function in which it keeps the expression level of its mRNA targets at a certain level below which protein production is restrained (Mukherji et al., 2011; Sood et al., 2006). However, another hypothesis might be that the miR-35 family does not lead to target degradation in the germline as suggested by Dallaire and colleagues (Dallaire et al., 2018).

Recently, two studies have published spatiotemporal maps of RNA expression in the *C. elegans* germline. In the first study, West and colleagues cut the gonad in three distinct regions, *i.e.*, distal mitotic proliferation zone, meiotic region, and developing oocytes, and combined it with a new Low-Input 3'-Terminal sequencing method (LITE-Seq) (West et al., 2018). The second one by Tzur and co-workers used laser capture microscopy to cut the gonad into ten slices and combined it with CEL-Seq for low RNA input (Tzur et al., 2018). Both studies had a lower spatiotemporal resolution compared to our near single-cell cryo-cutting approach (~ 15 slices per gonad arm) and only mRNA expression was assessed. Moreover, West and colleagues pooled gonads to obtain enough material for sequencing. The lower resolution of both studies impeded the detection of lowly or transiently-expressed mRNAs. In our study, we detected over 10,000 mRNAs expressed spatiotemporally in the gonad while Tzur et al. found only over 6,000 and West et al. roughly 10,000 in isolated gonads (Tzur et al., 2018; West et al., 2018). Interestingly, in agreement with our study West and co-workers reported the preferential expression of *rpl* and *rps* genes in the distal mitotic proliferation zone, and Tzur and colleagues observed a distinct expression pattern for some of the X chromosome-linked genes in the distal mitotic region of the gonad. Our sequencing approach uses rapid dissection and shock-freezing of the gonad, thereby avoiding buffer contamination and RNA leaking. Tzur

et al. had an issue with mRNA contamination from the gut that was transferred with the buffer, and West et al. faced the problem of RNA loss of lowly-expressed transcripts due to cutting without freezing. Moreover, West et al. cut the gonad 'by eye', decreasing the precision as the location of the different zones of the germline vary slightly between gonads and are difficult to see with the naked eye.

Overall, our cryo-cut-seq approach demonstrated high reliability and reproducibility enabling the identification of the spatiotemporal RNA transcriptome in the germline. While we showed that this approach can be used to assess spatiotemporal maps of the germline, another study, similarly based on cryo-cut-sequencing, determined spatial mRNA expression in whole worms (Ebbing et al., 2018). We believe that our method can be applied to assess spatial RNA expression of many other tissues and organs.

4.2 Identification of novel miRNAs with spatiotemporal expression in the germline

So far, the detection of known and novel miRNAs from *C. elegans* was, depending on the question, either done using whole worms or whole gonads for sequencing, as existing small RNA sequencing protocols required high RNA input material, *i.e.*, at least 1 μ g. Thus, identification of miRNAs expressed at very low levels and/or in a specific region in the germline was not possible with these standard approaches, as these would contribute only a very tiny fraction of the total amount of isolated small RNAs. We hypothesize that miRNAs expressed in specific regions in the germline may be prime candidates for key regulators of spatiotemporal gene expression in the germline. In this study, we combined our cryo-cut approach with a new small RNA sequencing protocol, the SMARTer smRNA-Seq, for low input material (≤ 1 ng). We screened our germline-specific small RNA data for putative novel miRNAs and, hence, potential regulators of spatiotemporal expression. We identified 83 putative novel miRNAs, the majority of which were very lowly expressed and displaying highly organized spatiotemporal expression in the germline (Fig. 13, 14 and Tab. 1). Further, we validated three of these novel miRNAs using orthogonal techniques. Interestingly, a subgroup of the novel miRNAs such as nov-63-3p had an A-stretch in the seed region, revealing a potential novel class of miRNAs. Moreover, by using chimeric data published previously by our lab we discovered a new binding pattern for nov-72-3p, as it binds other miRNAs creating a miRNA-miRNA duplex (Grosswendt et al., 2014). The miRNA-miRNA duplex was previously predicted for several miRNAs computationally by Lai and colleagues but lacked experimental evidence until now (Lai et al., 2004). They postulated two possible reasons for the miRNA-miRNA duplex. On the one hand, a miRNA-miRNA duplex may stabilize both miRNAs by preventing their degradation. On the other hand, the duplex may tether the miRNA away from its targets. However, we could not prove any of these two hypotheses. The mature miRNA of

nov-72-3p mapped antisense to one exon of the *dpy-2* locus but the first 17 nt including the seed region mapped to ribosomal RNA transcripts as well. Thus, the genomic locus of nov-72-3p is still not unambiguously defined impeding further experiments and analyses. Additionally, it is important to investigate and uncover the role and function of the other putative novel miRNAs in the development of the germline by doing loss-of-function and further cryo-cut-seq experiments.

4.3 Localized RNA expression is important for cell proliferation and differentiation

In this study, we demonstrated that mRNAs are spatiotemporally organized in the wild type germline. Interestingly, investigating the spatiotemporal mRNA expression of the *gld-2 gld-1* double mutant revealed that the highly-organized expression is strongly perturbed in this germline. Meiotic entry is impaired in *gld-2 gld-1* double mutants and germ cells are constantly proliferating throughout the germline (Brenner and Schedl, 2016; Fox et al., 2011; Hansen et al., 2004b). This leads to a sterile and tumorous phenotype due to lack of germ cells that differentiate to oocytes. In accordance, *rpl* and *rps* genes that are known to be required for proliferation are expressed throughout the germline (Fig. 17). Moreover, we found genes that are required for differentiation to be downregulated or completely absent in the *gld-2 gld-1* mutant (Fig. 17). In comparison, these genes were highly expressed in the proximal gonad arm in the wild type where cells differentiate into oocytes (Fig. 15). Among these genes, we detected *pie-1*, *cey-2* and *nos-2* (Fig. 15 and 17). *Pie-1* encodes a CCCH-type zinc finger protein which acts as a repressor of RNA polymerase II-dependent gene expression in the early germline development (Seydoux and Dunn, 1997; Seydoux et al., 1996; Tenenhaus et al., 2001). Additionally, it was shown that PIE-1 is associated with P granules in the germline maintaining the expression of class II mRNAs, mRNAs that are located to P granules such as *nos-2* and *cey-2* (Tenenhaus et al., 2001). *Nos-2* is required for the switch from spermatogenesis to oogenesis while the protein of *cey-2* is involved in translation, RNA processing and transcript stabilization (Arnold et al., 2014). We speculate that *pie-1* may also act in the germline as transcriptional inhibitor where germ cells become transcriptionally silent during late oogenesis until the 4-cell stage embryo (Evsikov et al., 2006; Stoeckius et al., 2014). Supporting this hypothesis, it was shown that PIE-1 in HeLa cells can inhibit transcription directly suggesting a conserved mechanism (Batchelder et al., 1999). However, the localization of *pie-1* in the proximal gonad arm is important as it is maternally provided to the embryos where it is essential for germline cell fate determination. The co-localization of *pie-1* with *nos-2* and *cey-2* supports the fact that PIE-1 is responsible for maintaining the expression of these two genes. Moreover, the same genes were downregulated (*pie-1* and *cey-2*) or completely absent (*nos-2*) in the *gld-2 gld-1* mutant germline (Fig. 17). This was expected, as PIE-1 is required for maintaining the expression of *nos-2* and *cey-2* and

downregulation of *pie-1* probably causes the downregulation of its protein resulting in the downregulation or absence of *cey-2* and *nos-2*. A previous study by Kim and colleagues showed evidence that *pie-1* is a target of GLD-2 (Kim et al., 2010). Thus, we hypothesized that GLD-2 is a prime candidate for regulating *pie-1* and, therefore, *nos-2* and *cey-2* expression and transcription. However, investigating the single *gld-2* mutant did not show the same phenotypic nor spatiotemporal behavior as *gld-2 gld-1* double mutant suggesting that there might be additional regulators or that regulation is carried out on a translational level. Furthermore, *gld-1* and *gld-2* single mutants enter meiosis normally and do not display the phenotype of *gld-2 gld-1* double mutants (Brenner and Schedl, 2016). As our double mutant had a *gld-1* loss-of-function, most of the GLD-1 targets were downregulated as well. GLD-1 is known to be important for the meiotic progression and oogenesis (Lee and Schedl, 2010). Thus, we cannot exclude that GLD-1 and its targets are involved in the phenotypic behavior and perturbation of spatiotemporal gene expression. Further experiments are necessary to investigate if *pie-1* can phenocopy the *gld-2 gld-1* double mutant phenotype, *i.e.*, RNAi of *pie-1*. Nevertheless, our results demonstrated that loss-of-function of *gld-2* and *gld-1* leads to a disruption of the spatiotemporal gene expression that may cause the proliferation and differentiation imbalance (Fig. 27). Moreover, we observed that X-linked genes which are important for oogenesis and, thus, differentiation were downregulated in the mutant germline as well. Yet, we cannot say if the proliferation and differentiation imbalance is caused by the spatiotemporal disturbance or *vice versa*. However, it is clear from comparing our wild type and mutant data that localized RNA expression is important for the germline integrity and morphology and, therefore, for the balance between germ cell proliferation and differentiation.

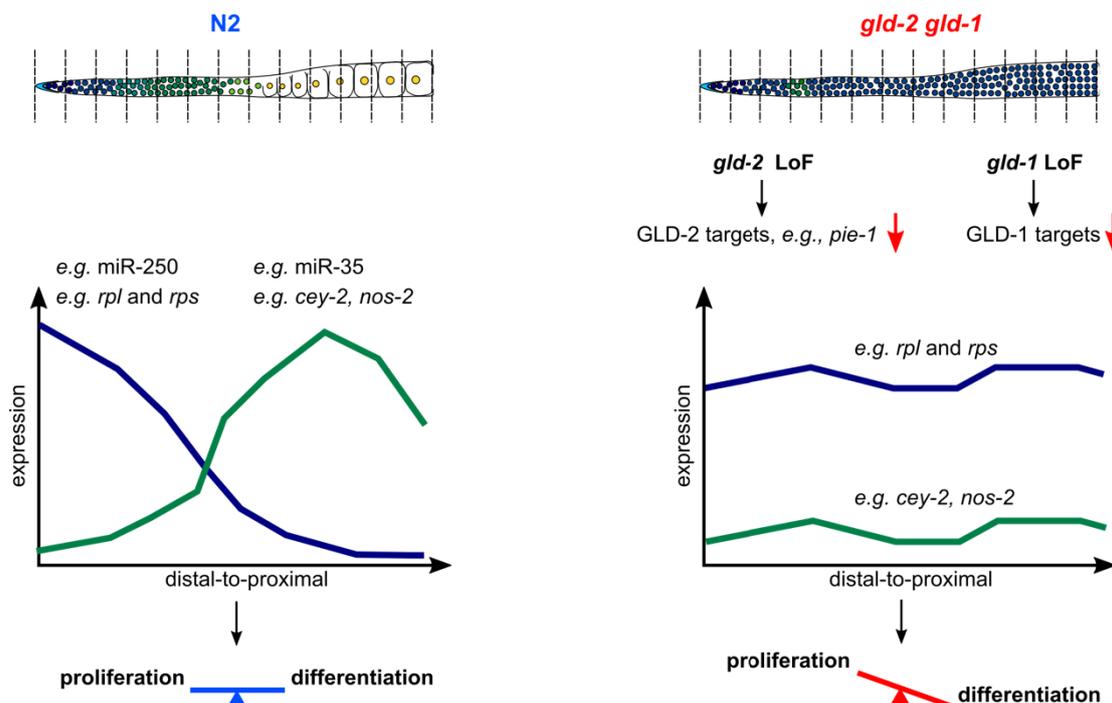


Figure 27. Model for spatially restricted gene expression. Schematic overview of mRNA and miRNA localization in wild type N2 and mRNA localization in *gld-2 gld-1* double mutant germline, indicating potential regulators of spatially restricted gene expression. In the wild type germline, mRNAs and miRNAs localize either to the distal gonad arm or to the proximal gonad arm maintaining the proliferation and differentiation balance. In contrast, in the *gld-2 gld-1* double mutant mRNAs are either expressed throughout the germline or downregulated which may be a cause of the proliferation and differentiation imbalance. LoF, loss-of-function. (Diag et al., 2018)

4.4 Differential 3' UTR isoform usage is strongly regulated in the germline

Our cryo-cut-seq approach generates libraries primed from poly(A) tails of transcripts. Thus, the read coverage is limited to the 3' end of the transcripts. However, we observed some coverage peaks downstream of annotated 3' ends, indicating longer 3' untranslated regions (UTRs) for these transcripts (Fig. 19). Indeed, we extended the 3' UTR annotation for 419 genes and I validated nine of them by PCR (Fig. 20). Previous studies annotating alternative isoforms in *C. elegans* used different 3' UTR sequencing methods in order to investigate the function and role of 3' UTRs in *C. elegans* development (Jan et al., 2011; Mangone et al., 2010). In both studies, the 3' UTR annotation was extended compared to the WormBase annotation and we found an overlap with our own study. However, they used and sequenced whole worms in their studies and some of the germline-specific genes with an extended 3' UTR that we detected were missed.

Following the 3' UTR extension, we investigated the spatiotemporal behavior of genes with different 3' UTR isoforms throughout the gonad. Recent studies showed evidence that proliferating cells use predominantly the proximal polyadenylation signal (PAS) resulting in the usage of the shorter 3' UTR isoform while differentiated cells tend to use the distal PAS leading to the usage of the longer 3' UTR (Mayr and Bartel, 2009; Sandberg et al., 2008; Sood et al., 2006; West et al., 2018). The germline is divided into proliferating and differentiated cells offering a very useful model system to study differential 3' UTR isoform usage. Of note, in most cases our sequencing approach did not provide the resolution to distinguish between different isoforms as we sequenced ~ 500 nt fragments while the mean 3' UTR length of *C. elegans* transcripts is 211 nt (Jan et al., 2011; Mangone et al., 2010). Nevertheless, we quantified the relative changes of the 3' UTR usage through the gonad for nearly 1,000 genes. At this point, we believe that due to our technical limitations, this number might be only a subset of all 3' UTR length switches. Already West and colleagues investigated the differential 3' UTR usage in the *C. elegans* germline (West et al., 2018). However, their study is not comparable to ours, as they cut the gonad only in three parts, while we had ~ 15 slices resulting in a much higher resolution.

In contrast to previous studies mentioned above, we observed genes that predominantly used the distal PAS in the distal gonad arm where proliferating cells are located, whereas the proximal PAS was used in the proximal gonad arm where differentiated cells occur (Fig. 21). The switch from long 3' UTR to short 3' UTR isoform usage occurred around the late pachytene stage (loop region). Correspondingly, we observed that the expression of *cpsf-4* (human

CPSF4L) and *fipp-1* (human FIP1L1), two components of the cleavage and polyadenylation complex that are involved in alternative polyadenylation (APA) regulation, increased around the loop region (Fig. 22). Both factors are involved in the recognition of the PAS and interaction with the poly(A) polymerase and other factors during APA (Kaufmann et al., 2004). Furthermore, it is thought that the proximal PAS is weaker than the distal one, in most cases the canonical one (AAUAAA) (Lackford et al., 2014; West et al., 2018). Intriguingly, the switch from distal to proximal PAS usage was completely absent in the *gld-2 gld-1* double mutant. Instead, only the distal PAS was used throughout the mutant germline (Fig. 21). Concomitantly, the expression of *cpsf-4* and *fipp-1* did not increase around the late pachytene stage but stayed constant throughout the gonad (Fig. 22). Thus, we speculate that the expression level of these two factors is crucial for the differential 3' UTR isoform usage in the germline (Fig. 28). A similar phenomenon was previously observed by Lackford and colleagues in embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) for CPSF-4 and FIP1 (Lackford et al., 2014).

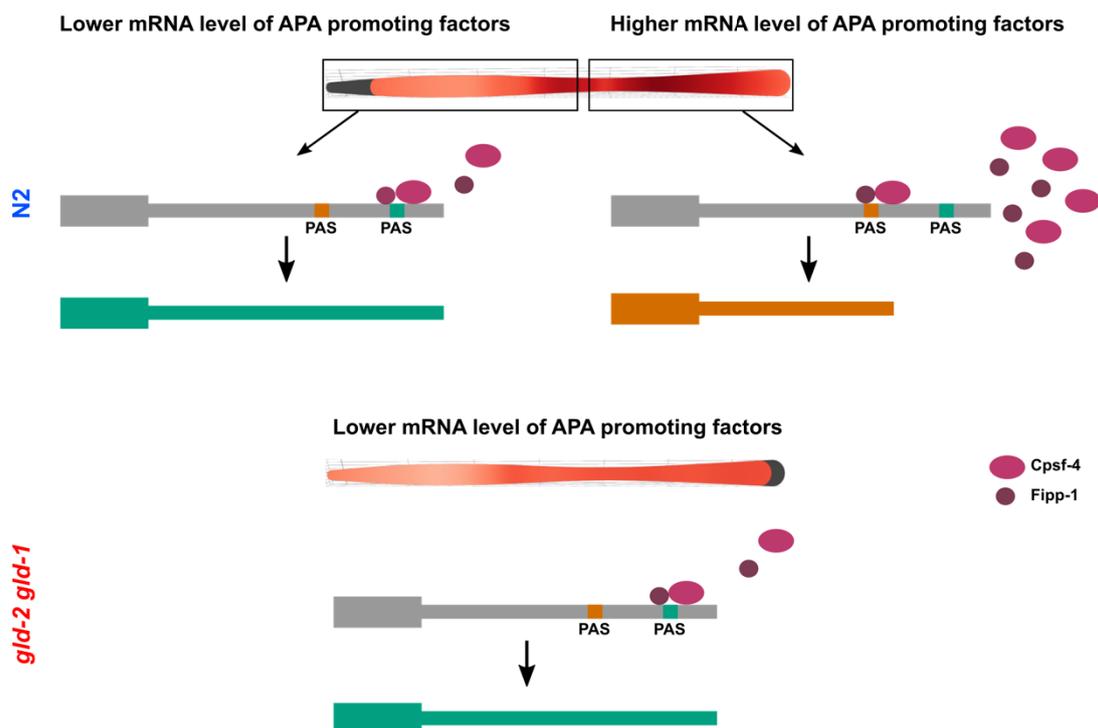


Figure 28. Model for differential 3' UTR isoform usage in the germline. Schematic model for differential 3' UTR isoform usage across the germline. Depending on the concentration of *cpsf-4* (vISH is shown) and *fipp-1*, two factors involved in alternative polyadenylation (APA), a fraction of genes use the longer 3' UTR isoform in the distal gonad arm while the shorter one is used in the proximal gonad arm. In *gld-2 gld-1* double mutants the concentration of *cpsf-4* (vISH is shown) and *fipp-1* stays constant and only the longer 3' UTR isoform is used. (Diag et al., 2018)

We hypothesize the following: If the expression level of *cpsf-4* and *fipp-1* is low, the probability that the weak proximal PAS will be bound decreases resulting in the predominant usage of the strong distal PAS and, thus, in the abundance of the long 3' UTR isoform. In contrast, if the

expression level of these two factors increase, the probability for binding the weak proximal PAS correspondingly increases resulting in the usage of the proximal PAS and the shorter 3' UTR isoform. Lackford and colleagues additionally observed that the usage of distal or proximal PAS was also dependent on the distance between these two PASs. At long distances (~ 1 kb), high levels of CPSF-4 and FIP1 would lead to the usage of the proximal PAS and low levels would result in the usage of the distal PAS as observed in our study. At a shorter distance (> 300 bp), high levels of CPSF-4 and FIP1 may block the binding of other mRNA 3' processing factors, thereby impeding the recognition of the proximal PAS. In contrast, we did not observe any dependence of PAS usage on the distance between distal and proximal PAS, as the distance for the investigated genes was in general very short (\leq 100 nt). Further supporting our hypothesis, the expression level of *cpsf-4* and *fipp-1* was constant throughout the *gld-2 gld-1* double mutant germline and only the long 3' UTR isoform was used (Fig. 21, 22 and 28). Moreover, we generally observed a lower 3' UTR variability in the mutant germline compared to the wild type, *i.e.*, most of the 1,000 genes that switch 3' UTR isoform usage in the wild type germline did not switch in the mutant germline (Fig. 22). We hypothesize that perturbation in the differential 3' UTR isoform usage might be involved in the proliferation and differentiation imbalance in the mutant germline. Further experiments such as RNAi of *cpsf-4* and *fipp-1* are needed in order to prove our hypothesis. However, RNAi experiments of both factors result in an embryonic lethal phenotype impeding any further investigations due to lack of progeny. Nevertheless, the embryonic lethal phenotype demonstrates the importance of both factors for germline viability and integrity.

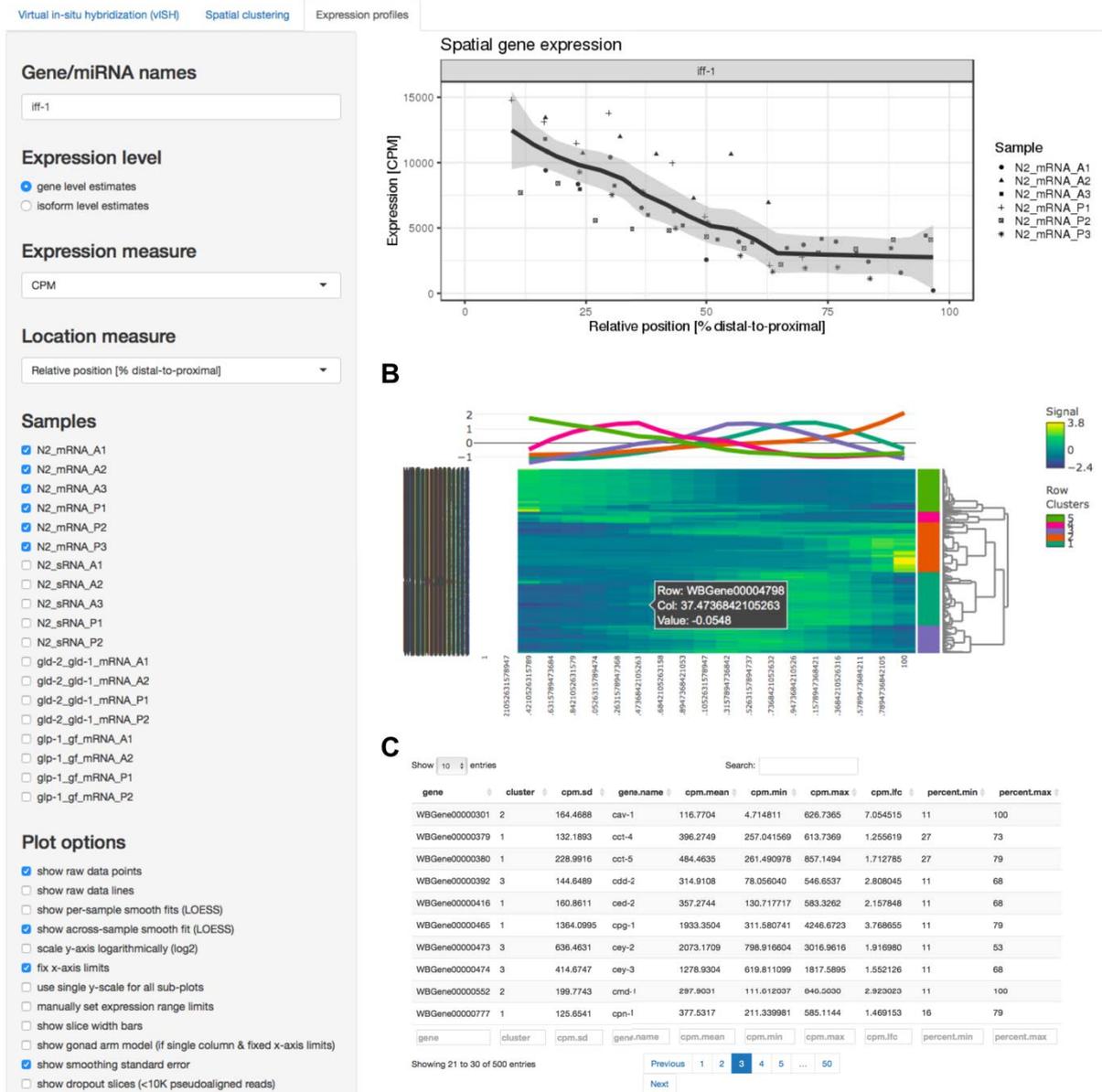
In order to increase the resolution to distinguish between different isoforms of a gene, it might be helpful to change the sequencing method. In this study, we sequenced barcode and UMI only from the 5' end (12 bp) and roughly 60 bp from the 3' end of the transcript (see methods). However, for isoform detection it would be useful to get information from both ends of the transcript, *i.e.*, from 5' and 3' end of the transcript. Thus, it would be necessary to sequence ~ 100 bp from 5' and 3' end of the fragment, respectively, in order to increase the detection resolution and to determine which isoform of the gene is used. This would help to increase the number of detected isoforms and to understand the role and function of different isoforms in the germline.

4.5 SPACEGERM: An interactive online data visualization tool

In this study, we created a data set of spatiotemporal mRNA and small RNA expression in the *C. elegans* germline. In order to make this data set easily accessible to the scientific community, we developed an interactive data visualization tool, named SPACEGERM (Spatial C. e*legans* germline expression of mRNA and miRNA) (Diag et al., 2018). SPACEGERM facilitates exploration of spatiotemporal RNA expression in the germline in well-defined

coordinates, both as raw data and analyzed and integrated (“virtual *in situ* hybridization”) into our constructed 3D physical gonad model (Fig. 26 and 29). Moreover, wild type and mutant data can be compared and evaluated.

A



D

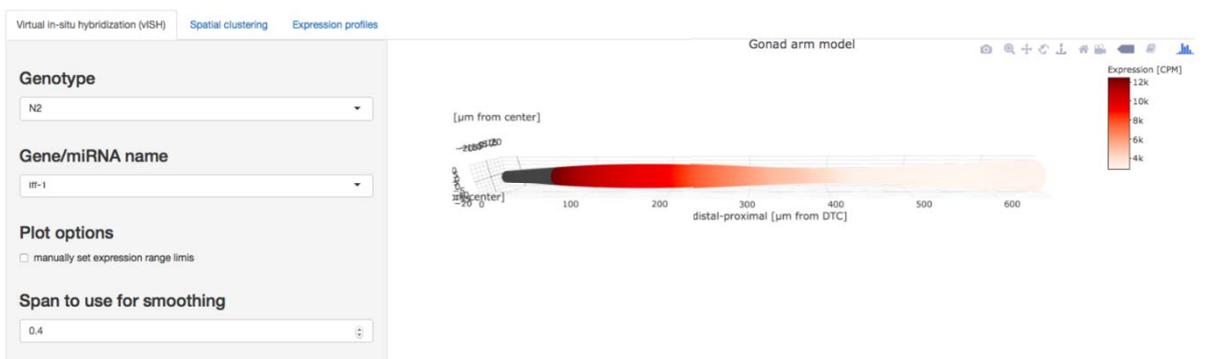


Figure 29 SPACEGERM: a user-friendly interface for exploring spatial expression across the germline in 2D and 3D. (A) Plotting options for each transcript detected in our data. As an example, the spatial expression of *iff-1* is shown for all biological and technical replicates of N2, LOESS \pm standard error (SE). **(B)** Global spatial gene expression can be investigated by clustering all detected genes according linear correlation (Pearson's, r) for all genotypes. μ : Mean. σ : standard deviation. NA: No data. **(C)** Result of clustering can be exported as an Excel file. **(D)** Virtual *in situ* hybridization (vISH) using reconstructed 3D germline model. As an example, the spatial expression of *iff-1* is shown for all biological and technical replicates of N2. (Diag et al., 2018)

Overall, we showed that both mRNAs and small RNAs (miRNAs, siRNAs and piRNAs) were specifically localized within the *C. elegans* gonad. For several cases we validated these patterns independently. We discovered many novel miRNA candidates, validating several of them using orthogonal methods. Additionally, we extended the 3' UTR annotation, again validating several extensions independently. Moreover, we showed, that differential 3' UTR usage is coordinated in the *C. elegans* gonad. Finally, we provided perturbation data by repeating the experiment in several mutant conditions, showing drastic changes between N2 wild type and *gld-2 gld-1* double mutant strains. Comparison to additional mutants will increase our understanding about germline development.

At this point, we note that more measurements are necessary such as subcellular localization, mRNA modifications, protein levels, and many more as we only quantified RNAs at steady state level with our method. Nevertheless, we believe that our study is a crucial starting point for understanding germline integrity and viability.

5 References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.

Arnold, A., Rahman, M.M., Lee, M.C., Muehlhaeusser, S., Katic, I., Gaidatzis, D., Hess, D., Scheckel, C., Wright, J.E., Stetak, A., et al. (2014). Functional characterization of *C. elegans* Y-box-binding proteins reveals tissue-specific functions and a critical role in the formation of polysomes. *Nucleic Acids Res.* 42, 13353–13369.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.

Bartel, D.P. (2018). Metazoan MicroRNAs. *Cell* 173, 20–51.

Batchelder, C., Dunn, M.A., Choy, B., Suh, Y., Cassie, C., Shim, E.Y., Shin, T.H., Mello, C., Seydoux, G., and Blackwell, T.K. (1999). Transcriptional repression by the *Caenorhabditis elegans* germ-line protein PIE-1. *Genes Dev.* 13, 202–212.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522, 363–367.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* 7, 1749–1756.

Besse, F., and Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat. Rev. Mol. Cell Biol.* 9, 971–980.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94.

Brenner, J.L., and Schedl, T. (2016). Germline Stem Cell Differentiation Entails Regional Control of Cell Fate Regulator GLD-1 in *Caenorhabditis elegans*. *Genetics* 202, 1085–1103.

- Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K.J., Ji, F., Schwarz, B.A., Kalocsay, M., Elledge, S.J., et al. (2018). Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* 172, 106–120.e21.
- Bukhari, S.I.A., Vasquez-Rifo, A., Gagné, D., Paquet, E.R., Zetka, M., Robert, C., Masson, J.-Y., and Simard, M.J. (2012). The microRNA pathway controls germ cell proliferation and differentiation in *C. elegans*. *Cell Res.* 22, 1034–1045.
- Bystrykh, L.V. (2012). Generalized DNA barcode design based on Hamming codes. *PLoS One* 7, e36852.
- Chen, F.W., and Ioannou, Y.A. (1999). Ribosomal proteins in cell proliferation and apoptosis. *Int Rev Immunol* 18, 429–448.
- Ciulli Mattioli, C., Rom, A., Franke, V., Imami, K., Arrey, G., Terne, M., Woehler, A., Akalin, A., Ulitsky, I., and Chekulaeva, M. (2018). Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. *Nucleic Acids Res.*
- Corsi, A.K., Wightman, B., and Chalfie, M. (2015). A Transparent Window into Biology: A Primer on *Caenorhabditis elegans*. *Genetics* 200, 387–407.
- Crittenden, S.L., Leonhard, K.A., Byrd, D.T., and Kimble, J. (2006). Cellular analyses of the mitotic region in the *Caenorhabditis elegans* adult germ line. *Mol. Biol. Cell* 17, 3051–3061.
- Dallaire, A., Frédérick, P.-M., and Simard, M.J. (2018). Somatic and Germline MicroRNAs Form Distinct Silencing Complexes to Regulate Their Target mRNAs Differently. *Dev. Cell* 47, 239–247.e4.
- Diag, A., Schilling, M., Klironomos, F., Ayoub, S., and Rajewsky, N. (2018). Spatiotemporal m(i)RNA Architecture and 3' UTR Regulation in the *C. elegans* Germline. *Dev. Cell* 47.
- Ding, X.C., Slack, F.J., and Grosshans, H. (2008). The let-7 microRNA interfaces extensively with the translation machinery to regulate cell differentiation. *Cell Cycle* 7, 3083–3090.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

- Dotd, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* *1*, 895–905.
- Ebbing, A., Vértesy, Á., Betist, M.C., Spanjaard, B., Junker, J.P., Berezikov, E., van Oudenaarden, A., and Korswagen, H.C. (2018). Spatial Transcriptomics of *C. elegans* Males and Hermaphrodites Identifies Sex-Specific Differences in Gene Expression Patterns. *Dev. Cell* *47*, 801–813.e6.
- Evsikov, A.V., Graber, J.H., Brockman, J.M., Hampl, A., Holbrook, A.E., Singh, P., Eppig, J.J., Solter, D., and Knowles, B.B. (2006). Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes Dev.* *20*, 2713–2727.
- Fox, P.M., Vought, V.E., Hanazawa, M., Lee, M.-H., Maine, E.M., and Schedl, T. (2011). Cyclin E and CDK-2 regulate proliferative cell fate and cell cycle progression in the *C. elegans* germline. *Development* *138*, 2223–2234.
- Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* *40*, 37–52.
- Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the *Drosophila* PRD gene set as part of a gene network. *Cell* *47*, 735–746.
- Gartner, A., Boag, P.R., and Blackwell, T.K. (2008). Germline survival and apoptosis. *WormBook* 1–20.
- Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell* *158*, 1281–1292.
- Gonda, T.J., and Ramsay, R.G. (2015). Directly targeting transcriptional dysregulation in cancer. *Nat. Rev. Cancer* *15*, 686–694.
- Grosswendt, S., Filipchuk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site

interactions by different types of ligation reactions. *Mol. Cell* 54, 1042–1054.

Hansen, D., and Schedl, T. (2013). Stem cell proliferation versus meiotic fate decision in *Caenorhabditis elegans*. *Adv. Exp. Med. Biol.* 757, 71–99.

Hansen, D., Wilson-Berry, L., Dang, T., and Schedl, T. (2004a). Control of the proliferation versus meiotic development decision in the *C. elegans* germline through regulation of GLD-1 protein accumulation. *Development* 131, 93–104.

Hansen, D., Hubbard, E.J.A., and Schedl, T. (2004b). Multi-pathway control of the proliferation versus meiotic development decision in the *Caenorhabditis elegans* germline. *Dev. Biol.* 268, 342–357.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* 17, 77.

Hirsh, D., Oppenheim, D., and Klass, M. (1976). Development of the reproductive system of *Caenorhabditis elegans*. *Dev. Biol.* 49, 200–219.

http://www.wormbase.org/species/c_elegans/gene/WBGene00003784

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.

Hubbard, E.J.A. (2007). *Caenorhabditis elegans* germ line: a model for stem cell biology. *Dev. Dyn.* 236, 3343–3357.

Hubbard, E.J., and Greenstein, D. (2000). The *Caenorhabditis elegans* gonad: a test tube for cell and developmental biology. *Dev. Dyn.* 218, 2–22.

- Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* 469, 97–101.
- Jansen, R.P. (2001). mRNA localization: message on the move. *Nat. Rev. Mol. Cell Biol.* 2, 247–256.
- Jens, M., and Rajewsky, N. (2015). Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* 16, 113–126.
- Jungkamp, A.-C., Stoeckius, M., Mecnas, D., Grün, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* 44, 828–840.
- Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. *Cell* 159, 662–675.
- Kadyk, L.C., and Kimble, J. (1998). Genetic regulation of entry into meiosis in *Caenorhabditis elegans*. *Development* 125, 1803–1813.
- Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. (2004). Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J.* 23, 616–626.
- Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* 10, 857–860.
- Kelly, W.G., Schaner, C.E., Dernburg, A.F., Lee, M.-H., Kim, S.K., Villeneuve, A.M., and Reinke, V. (2002). X-chromosome silencing in the germline of *C. elegans*. *Development* 129, 479–492.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Kim, K.W., Wilson, T.L., and Kimble, J. (2010). GLD-2/RNP-8 cytoplasmic poly(A) polymerase is a broad-spectrum regulator of the oogenesis program. *Proc. Natl. Acad. Sci. USA* 107,

17445–17450.

Koller, M., and Stahel, W.A. (2011). Sharpening Wald-type inference in robust regression for small samples. *Comput. Stat. Data Anal.* 55, 2504–2515.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.

Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.-A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M., et al. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 33, 878–889.

Lai, E.C., Wiel, C., and Rubin, G.M. (2004). Complementary miRNA pairs suggest a regulatory role for miRNA:miRNA duplexes. *RNA* 10, 171–175.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* 131, 174–187.

Lee, M.-H., and Schedl, T. (2010). *C. elegans* star proteins, GLD-1 and ASD-2, regulate specific RNA targets to control development. *Adv. Exp. Med. Biol.* 693, 106–122.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* (80-.). 343, 1360–1363.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A.,

- Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838.
- Maciejowski, J., Ugel, N., Mishra, B., Isopi, M., and Hubbard, E.J.A. (2006). Quantitative analysis of germline mitosis in adult *C. elegans*. *Dev. Biol.* 292, 142–151.
- Mangone, M., Manoharan, A.P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S.D., Mis, E., Zegar, C., Gutwein, M.R., Khivansara, V., et al. (2010). The landscape of *C. elegans* 3'UTRs. *Science* (80-.). 329, 432–435.
- Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell* 136, 719–730.
- Martin, K.C., and Zukin, R.S. (2006). RNA trafficking and local protein synthesis in dendrites: an overview. *J. Neurosci.* 26, 7131–7134.
- Mayr, C. (2017). Regulation by 3'-Untranslated Regions. *Annu. Rev. Genet.* 51, 171–194.
- Mayr, C. (2018). What are 3' utrs doing? *Cold Spring Harb. Perspect. Biol.*
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673–684.
- Medioni, C., Mowry, K., and Besse, F. (2012). Principles and roles of mRNA localization in animal development. *Development* 139, 3263–3276.
- Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr. Biol.* 18, 1476–1482.
- Millonigg, S., Minasaki, R., Nusch, M., Novak, J., and Eckmann, C.R. (2014). GLD-4-mediated translational activation regulates the size of the proliferative germ cell pool in the adult *C. elegans* germ line. *PLoS Genet.* 10, e1004647.
- Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* 3, e215.

- Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* *43*, 854–859.
- Nousch, M., and Eckmann, C.R. (2013). Translational control in the *Caenorhabditis elegans* germ line. *Adv. Exp. Med. Biol.* *757*, 205–247.
- Nousch, M., Minasaki, R., and Eckmann, C.R. (2017). Polyadenylation is the key aspect of GLD-2 function in *C. elegans*. *RNA* *23*, 1180–1187.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.
- Rebagliati, M.R., Weeks, D.L., Harvey, R.P., and Melton, D.A. (1985). Identification and cloning of localized maternal RNAs from *Xenopus* eggs. *Cell* *42*, 769–777.
- Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., et al. (2000). A global profile of germline gene expression in *C. elegans*. *Mol. Cell* *6*, 605–616.
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* *127*, 1193–1207.
- Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A variety of dicer substrates in human and *C. elegans*. *Cell* *159*, 1153–1167.
- Sambrook, J., and Russell, D.W. (2006). Rapid Amplification of 3' cDNA Ends (3'-RACE). *CSH Protoc* *2006*.
- Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* (80-.). *320*, 1643–1647.
- Schmid, M., Küchler, B., and Eckmann, C.R. (2009). Two conserved regulatory cytoplasmic poly(A) polymerases, GLD-4 and GLD-2, regulate meiotic progression in *C. elegans*. *Genes Dev.* *23*, 824–836.

- Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. *Development* *124*, 2191–2201.
- Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Priess, J.R., and Fire, A. (1996). Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature* *382*, 713–716.
- Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* *17*, 761–772.
- Sherrard, R., Luehr, S., Holzkamp, H., McJunkin, K., Memar, N., and Conradt, B. (2017). miRNAs cooperate in apoptosis regulation during *C. elegans* development. *Genes Dev.* *31*, 209–222.
- Shkumatava, A., Stark, A., Sive, H., and Bartel, D.P. (2009). Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev.* *23*, 466–481.
- Smit, A.F.A., Hubley, R., and Green, P. (2013-2015). RepeatMasker. Retrieved from <http://www.repeatmasker.org>.
- Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl. Acad. Sci. USA* *103*, 2746–2751.
- Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J., et al. (2011). A spatial and temporal map of *C. elegans* gene expression. *Genome Res.* *21*, 325–341.
- Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* (80-.). *353*, 78–82.
- Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.-P., Friedländer, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nat. Methods* *6*, 745–751.

- Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., Herzog, M., Selbach, M., and Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in *Caenorhabditis elegans* uncovers a novel mRNA clearance mechanism. *EMBO J.* *33*, 1751–1766.
- Strome, S., Kelly, W.G., Ercan, S., and Lieb, J.D. (2014). Regulation of the X chromosomes in *Caenorhabditis elegans*. *Cold Spring Harb. Perspect. Biol.* *6*.
- Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* *100*, 64–119.
- Tabuchi, T.M., Deplancke, B., Osato, N., Zhu, L.J., Barrasa, M.I., Harrison, M.M., Horvitz, H.R., Walhout, A.J.M., and Hagstrom, K.A. (2011). Chromosome-biased binding and gene regulation by the *Caenorhabditis elegans* DRM complex. *PLoS Genet.* *7*, e1002074.
- Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* *31*, 2032–2034.
- Tenenhaus, C., Subramaniam, K., Dunn, M.A., and Seydoux, G. (2001). PIE-1 is a bifunctional protein that regulates maternal and zygotic gene expression in the embryonic germ line of *Caenorhabditis elegans*. *Genes Dev.* *15*, 1031–1040.
- Tzur, Y.B., Winter, E., Gao, J., Hashimshony, T., Yanai, I., and Colaiácovo, M.P. (2018). Spatiotemporal Gene Expression Analysis of the *Caenorhabditis elegans* Germline Uncovers a Syncytial Expression Switch. *Genetics* *210*, 587–605.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* *40*, e115.
- Wang, X., Zhao, Y., Wong, K., Ehlers, P., Kohara, Y., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hansen, D. (2009). Identification of genes expressed in the hermaphrodite germ line of *C. elegans* using SAGE. *BMC Genomics* *10*, 213.
- Weick, E.-M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. *Development* *141*, 3458–3471.
- West, S.M., Mecnas, D., Gutwein, M., Aristizábal-Corrales, D., Piano, F., and Gunsalus, K.C.

(2018). Developmental dynamics of gene expression and alternative polyadenylation in the *Caenorhabditis elegans* germline. *Genome Biol.* 19, 8.

Wolke, U., Jezuit, E.A., and Priess, J.R. (2007). Actin-dependent cytoplasmic streaming in *C. elegans* oogenesis. *Development* 134, 2227–2236.

Xu, X., Xiong, X., and Sun, Y. (2016). The role of ribosomal proteins in the regulation of cell proliferation, tumorigenesis, and genomic integrity. *Sci China Life Sci* 59, 656–672.

Zhang, Q., Shalaby, N.A., and Buszczak, M. (2014). Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. *Science* (80-.). 343, 298–301.

Appendix

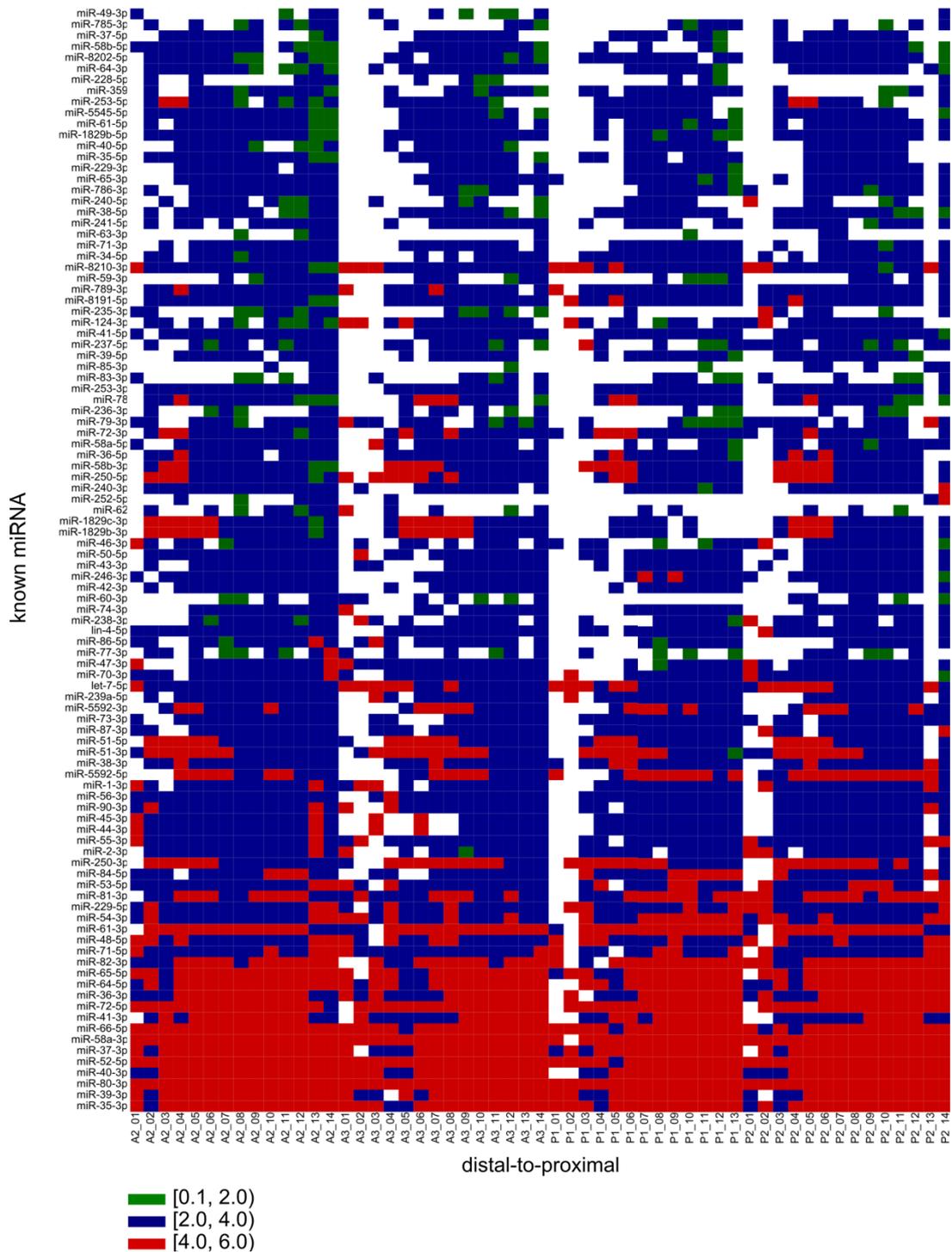


Figure 30. Known miRNAs are quantified in space and time along the gonad. Heatmap of the top 100 expressed known mature miRNAs ($\log_{10}[1+\text{CPM}]$) detected in this study. Each row represents one miRNA and each column one slice of one replicate (A2_01-14, A3_01-14, P1_01-13, P2_01-14), ordered from distal-to-proximal.

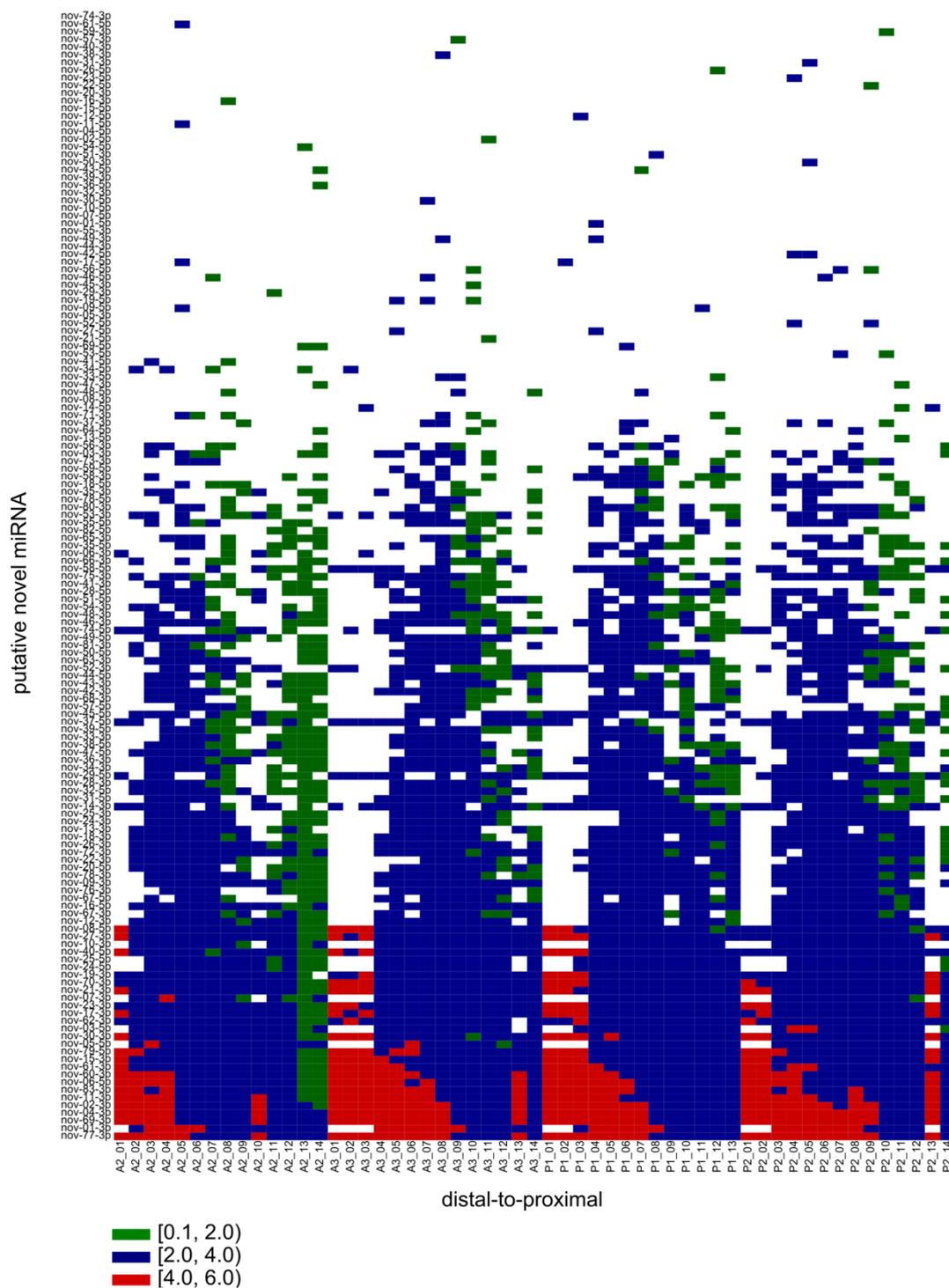


Figure 31. Novel miRNAs are quantified in space and time along the gonad. Heatmap of all putative novel miRNAs ($\log_{10}[1+\text{CPM}]$) detected in this study. Each row represents one miRNA and each column one slice of one replicate (A2_01-14, A3_01-14, P1_01-13, P2_01-14), ordered from distal-to-proximal.

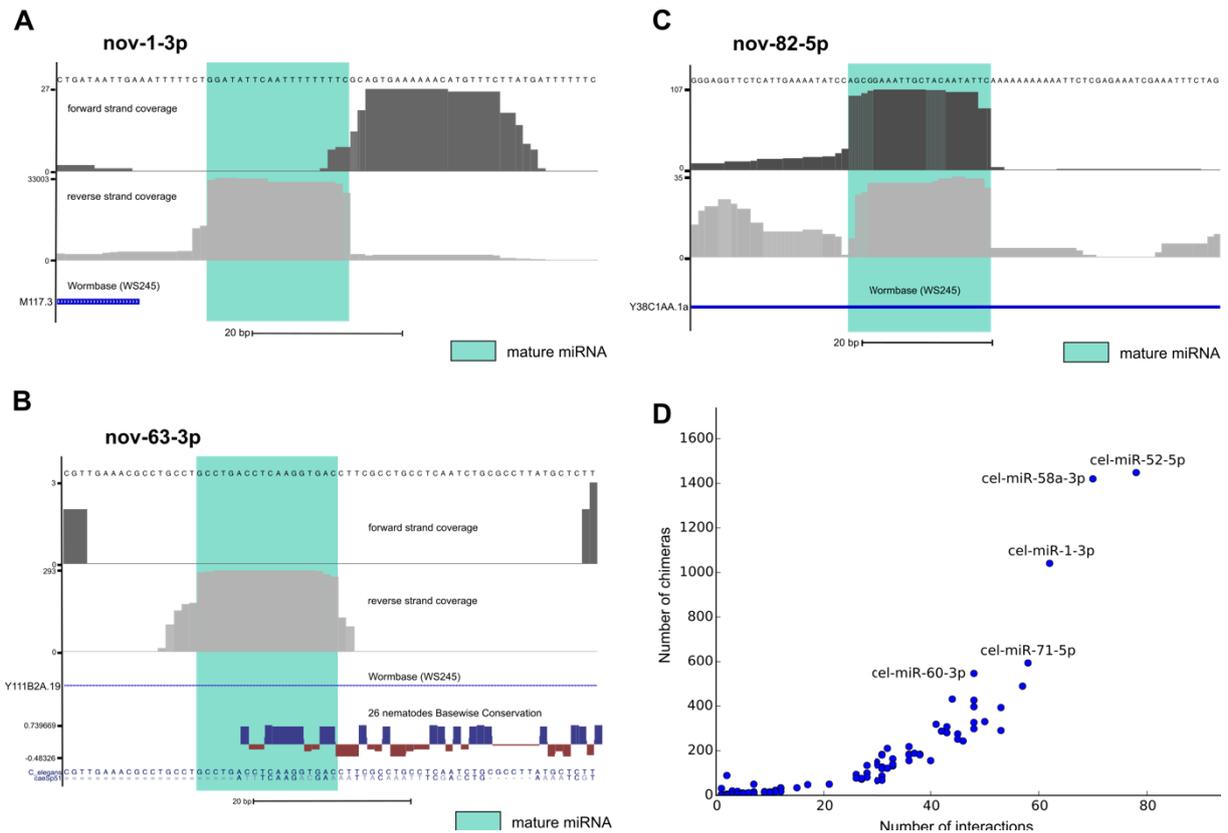


Figure 32. Novel miRNA predictions exhibit miRNA-like features. (A) Genome browser track showing read coverage of predicted miRNA candidate, nov-1-3p **(B)** nov-64-3p and **(C)** nov-82-5p with reads stacking up mostly on the mature sequence (light blue) at aligned 5' positions. Forward strand coverage is indicated in dark grey and reverse strand coverage is indicated in light grey. **(D)** Number of miRNA:mRNA chimeras for the novel miRNA nov-72-3p. (Diag et al., 2018)

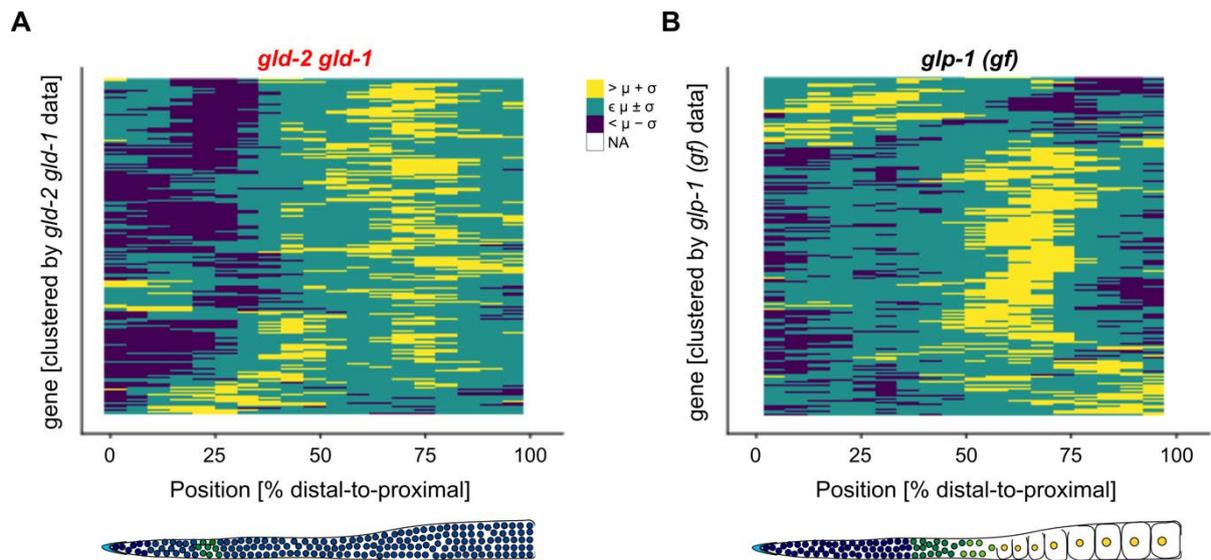


Figure 33. *gld-2 gld-1* double mutant and *glp-1 (gf)* mutant display mRNA localization. (A) Hierarchical clustering of germline specific genes by linear correlation ($1 - \text{Pearson's } r$) for *gld-2 gld-1* double mutant and **(B)** for *glp-1 (gf)* mutant. μ : Mean. σ : Standard deviation. NA: No data. (Diag et al., 2018)

Selbstständigkeitserklärung

Hiermit erkläre ich, Asija Diag, dass ich die vorliegende Dissertation eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Asija Diag

Berlin, 7. März 2019