Bioinformat-Eggs: An Educational Primer for Use with "LIN-41 and OMA Ribonucleoprotein Complexes Mediate a Translational Repression-to-Activation Switch Controlling Oocyte Meiotic Maturation and the Oocyte-to-Embryo Transition in *Caenorhabditis elegans*"

Deborah Thurtle-Schmidt* and Te-Wen Lo†,1
*Biology Department, Davidson College, North Carolina 28035 and †Biology Department, Ithaca College, New York 14850
ORCID ID: 0000-0002-1231-5531 (D.T.-S.)

ABSTRACT High-throughput sequencing and bioinformatic techniques have enhanced classical genetic analysis and are essential methods for geneticists. Tsukamoto and colleagues use numerous genomic and bioinformatics methods to explore the role of ribonucleoprotein complexes in regulating oocyte meiotic maturation, which is the transition between diakinesis and metaphase of meiosis I. This primer provides guidance for both educators and students as they read "LIN-41 and OMA Ribonucleoprotein Complexes Mediate a Translational Repression-to-Activation Switch Controlling Oocyte Meiotic Maturation and the Oocyte-to-Embryo Transition in *Caenorhabditis elegans*." The primer provides background information on the utility of the *C. elegans* germ line as a model for meiotic regulation, and further describes methods of bioinformatic analysis used to study translational and post-translational gene regulation. Additionally, the primer provides discussion questions and an active learning exercise designed to enhance student learning of critical genetic concepts.

KEYWORDS oocyte meiotic maturation; high-throughput bioinformatic analysis; education; *C. elegans* germ line; ribonucleoprotein complex

TABLE OF CONTENTS

doi: https://doi.org/10.1534/genetics.118.301139
Manuscript received May 15, 2018; accepted for publication May 15, 2018.
[1]Corresponding author: Ithaca College, Biology Department, 953 Danby Road, Ithaca, NY 14850. E-mail: twlo@ithaca.edu

IN many animal species, the number of oocytes in females is limited. As a result, oocyte production and maturation are great investments, both with regard to time and resources. The process of meiosis is thus tightly coordinated with developmental signals. Unfortunately, these tightly coordinated mechanisms are poorly understood. Tsukamoto *et al.* (2017), use *Caenorhabditis elegans* to uncover genetic mechanisms that regulate the spatial and temporal regulation of oocyte meiotic maturation, which is defined by the transition between diakinesis and metaphase of meiosis I (Kim *et al.* 2013).

The C. elegans germ line: a model for oocyte development

Since the late 1970s, *C. elegans* has proven to be a powerful model organism for studying numerous biological processes. For an extensive introduction to *C. elegans* see Corsi *et al.* (2015). The *C. elegans* hermaphrodite germ line is ideally suited for studying meiosis and oocyte development for a number of reasons. (1) *C. elegans* are transparent; individual cells are readily visible in the germ line as they progress from mitosis to meiosis and then gametogenesis (spermatogenesis and oogenesis), moving from the distal end of the germ line to the proximal end near the vulva. (2) Nuclear morphology is easily visualized and distinguished using **DAPI** (definitions of bolded key terms can be found in Table 1) staining during meiotic prophase (**pachytene**, **diplotene**, and **diakinesis**). (3) The specific signals and response to those signals that regulate meiotic maturation in *C. elegans* have been well defined, providing a wealth of information to build upon.

As in many sexually reproducing animals, *C. elegans* oocytes arrest in diakinesis of prophase I until oocytes receive the proper signals, as illustrated by Tsukamoto and colleagues in figure 1A. At the end of the gonadal arm is the **spermatheca**, where sperm are stored. The sperm secrete the major sperm protein (MSP), which is both a hormone that triggers oocyte meiotic maturation and a major cytoskeletal element required for amoeboid locomotion of nematode sperm [reviewed in Ward *et al.* (1981) and Ellis and Stanfield (2014)]. Yes, that is right, nematode sperm are amoeboid and do not have flagella like human sperm. For more information about nematode sperm motility (Smith 2014) and a cool oocyte video see Greenstein, 2005 Oocytes adjacent to the spermatheca are considered to be in the 1 position [see figure 1A of Tsukamoto *et al.* (2017)], where they progress through meiotic maturation. The **sheath cells**, which are part of the **somatic gonad**, act as MSP sensors, enabling oocytes to complete meiosis when sperm are present. However; in the absence of sperm, the sheath cells inhibit meiotic maturation. Oocyte maturation is characterized by multiple visual indicators: (1) nuclear envelope breakdown that begins shortly before ovulation and (2) oocyte shape changing from square to spherical after nuclear envelope breakdown [as illustrated in figure 1B of Tsukamoto *et al.* (2017)]. Protein signaling molecules regulate oocyte maturation. One of these protein signals is active maturation-promoting factor (MPF), a protein complex consisting of two subunits, Cdk1 and cyclin B, which triggers nuclear maturation and promotes chromosome condensation and meiotic spindle assembly. During oocyte meiotic maturation, cytoplasmic maturation also occurs, including the accumulation and reorganization of organelles and ribonucleoproteins (RNPs) in the cytoplasm, rearrangement of the cytoplasm, and changes in protein translation. Both cytoplasmic and nuclear maturation ensure that the embryo possesses the necessary genomic and cytoplasmic entities, such as the mitochondria, for proper development.

Regulatory regions of mRNA

As is the case for organismal development, oocyte development requires regulated gene expression. Because full-grown oocytes are transcriptionally quiescent, post-transcriptional regulation mediates oocyte maturation. RNA transcripts are subjected to various regulatory mechanisms to precisely time each mRNA's translation. The oocyte is endowed with many mRNAs that are only translated later in the embryo; these mRNAs must be preserved and stored until fertilization. The 5′- and 3′-UTRs (untranslated regions) of mRNA often contain specific nucleotide sequences that RNA-binding proteins recognize and bind to, promoting or inhibiting translation. Addition of polyadenosine monophosphates [poly(A)s] at the 3′ end of mRNAs is critical for the efficient translation and longevity of the mRNA molecules. Such 3′ polyadenylation usually occurs in the nucleus, but in some specialized cells such as germ cells and early embryonic cells, there are also dedicated poly(A) polymerases in the cytoplasm. Cytoplasmic poly(A) polymerases, such as GLD-2 in *C. elegans*, extend the poly(A) tail of mature mRNAs, which contributes to precise translational control and post-transcriptional gene regulation in oocytes and the early embryo.

Table 1 Glossary of terms

adaptors: short unique DNA sequences that are added to DNA fragments so that the fragments can attach to the flowcell to be sequenced
cDNA: otherwise known as complementary DNA synthesized from a single stranded RNA template by using a reverse transcriptase
DAPI: a fluorescent dye that labels the DNA in the nucleus
diakinesis: (sub-stage of prophase 1 in meiosis) chromosomes are at their most condensed, homologous chromosomes are still connected via chiamata
diplotene: (sub-stage of prophase 1 in meiosis) the synaptonemal complex disassembles and chiasmata (the physical representation of crossover events) becomes visible
fluorophore: fluorescent chemical compound that will emit light when excited
high-throughput sequencing: refers to sequencing techniques in which the whole genome is sequenced, not just one gene or a portion of a gene
mass-spectrometry: digests proteins into small fragments and based on the molecular weight determines the sequence of those short stretches of amino acids
microfluidic device: equipment that utilizes very small amounts of liquid on a microchip to perform laboratory experiments
model organism: non-human organism used for studies that will provide insights into the same biological functions in other organisms such as humans; often chosen for study due to their simplicity, tractability, cost, and quick reproductive time
oligos: short DNA fragments (often single-stranded)
pachytene: (sub-stage of prophase 1 in meiosis) chromosomes shorten and thicken and the synaptonemal complex forms
sheath cells: somatic cells that surround the gonad arm
somatic gonad: all of the non-reproductive tissues of the *C. elegans* gonad arm
spermatheca: somatic structure where the sperm are stored
synaptonemal complex: a multi-protein complex that forms between pairs of homologous chromosomes during meiosis that is required for proper recombination and chromosome segregation

LIN-41 and OMA RNA-binding proteins regulate oocyte maturation

In *C. elegans*, as in all organisms thus far examined, MPF does what its name suggests: it promotes continued meiosis or oocyte maturation. How does it do so? MPF consists of two subunits, a catalytic subunit (referred to as CDK-1 in *C. elegans*) and a regulatory subunit (cyclin B) (Dunphy *et al.* 1988; Gautier *et al.* 1988, 1990; Lohka *et al.* 1988). When the CDK-1 subunit is phosphorylated, MPF is inactive. CDC-25.3, a protein phosphatase, removes these inhibitory phosphates on CDK-1 (Kumagai and Dunphy 1991), creating an active MPF. In *C. elegans*, *cdc-25.3* mRNA associates with both LIN-41 and the OMA proteins (OMA-1 and OMA-2) (Spike *et al.* 2014a,b). LIN-41 and the OMA proteins both inhibit CDK-1 activation in part via 3′-UTR-dependent translational repression of *cdc-25.3* mRNA (Spike *et al.* 2014a) (see Figure 1). However, when mutated, LIN-41 and the OMA proteins have opposing phenotypes. When *lin-41* is deleted, pachytene-stage oocytes aberrantly activate CDK-1, cellularize prematurely, disassemble the **synaptonemal complex**, enter M phase, and assemble spindles, resulting in sterile animals (Spike *et al.* 2014a). When *oma-1* and *oma-2* are deleted, CDK-1 is not activated and the oocytes fail to undergo meiotic maturation resulting in sterile animals (Detwiler *et al.* 2001). Thus, the authors were interested in investigating this apparent conundrum: how is it that the OMA proteins and LIN-41 confer opposing phenotypes, but are known to associate in the same protein complex, both regulating *cdc-25.3* mRNA?

Unpacking the Work

To investigate the conundrum described above, Tsukamoto and colleagues used a combination of biochemical and genomic techniques to identify both the proteins and the mRNAs that associate with the OMA and LIN-41 complexes. The authors hypothesized that other mRNAs regulated by LIN-41 and the OMA proteins would illuminate the opposing phenotypes observed in the *lin-41* and *oma-1*; *oma-2* mutants.

Identification of LIN-41-associated proteins

In a previous study, this same group of researchers identified LIN-41 as a protein that interacts with OMA-1 and OMA-2 (Spike *et al.* 2014b). Using a similar strategy, the authors were now interested in identifying proteins that associate with LIN-41. The scientists used **immunopurification** (IP) to isolate LIN-41 along with any associated proteins. They were careful to select conditions that ensured that associated proteins copurified with LIN-41. As visualized by their gel using a sensitive protein stain (figure 3C in Tsukamoto *et al.* 2017), a large band the size of LIN-41 (124 kDa) was visible as well as many additional bands, revealing the many unknown proteins that copurified with LIN-41. To determine the identity of the proteins present in each band, the authors used **mass spectrometry,** which identifies short stretches of amino acids (peptides) and "maps" these peptide sequences to the translated regions of the genome to determine the proteins present. The proteins found to co-immunopurify with LIN-41 included other germline RNA-binding proteins such as GLD-1, GLD-2, MEX-3, SPN-4, and POS-1. These data suggested that LIN-41 acts with OMA-associated RNPs in the germ line to control oogenesis.

Unbiased detection of mRNAs in LIN-41 and OMA complexes

LIN-41 is part of RNP complexes, meaning that in addition to associating with other proteins, the complex also contains RNA, specifically mRNA. The authors were interested in identifying the mRNAs that associate specifically with LIN-41, the

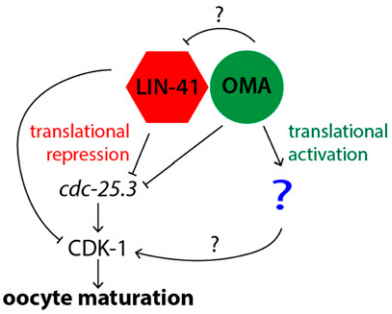

Figure 1 Simplified pathway highlighting the role of LIN-41 and OMA proteins in oocyte maturation adapted from Huelgas-Morales and Greenstein (2017). Question marks indicate relationships that are still currently unknown. Blue question mark represents the components that are identified by Tsukamoto *et al.* (2017).

OMA-proteins, or both complexes. Similar to using mass spectrometry for protein identification, the authors utilized an unbiased technique, termed **high-throughput sequencing**, to identify all the mRNAs associated with LIN-41 and OMA-1.

High-throughput sequencing techniques have revolutionized the field of genomics by allowing millions of short DNA fragments to be sequenced at once. First, the nucleic acid of interest was isolated. Tsukamoto and colleagues were interested in isolating all the mRNAs associated with LIN-41 and OMA-1. To prepare the mRNA for sequencing, a library of mRNA was generated (see Figure 2). First, the mRNA was converted to cDNA and then fragmented to $<$ 500 bp. Then, **adaptors**, which are short pieces of DNA, were ligated to the cDNA fragments. The adaptor sequences are important for two reasons: (1) the adaptors are used to hybridize all the pieces of DNA to the chip—a **microfluidic device**—for sequencing, and (2) DNA polymerase requires a primer with a 3′-hydroxyl group to synthesize the complementary strand and initiate sequencing. Thus, by adding the general adaptor sequences, any piece of DNA can be sequenced with a general primer.

Most high-throughput sequencing relies on a process called "sequencing by synthesis," which utilizes the same principles as PCR. DNA polymerase reads the single-stranded DNA and adds the complementary nucleotide. Each nucleotide—A, T, G, and C—has as a different colored **fluorophore** attached and, after each base is added, the sequencing chip is imaged to determine the base added at each location. To help magnify the signal of the fluorophores for each fragment being sequenced, prior to sequencing, each piece of DNA is amplified on the chip through PCR so that at each location on the chip there is a cluster of all the same fragments of DNA in one spot. Each returned sequence is termed a sequencing "read." For a typical high-throughput sequencing run, there can be $>$ 150 million reads or fragments sequenced, with each mRNA represented by many reads. Each sequencing read is mapped to its corresponding location in the genome using a computer algorithm. Those places where more reads map are called "peaks" of sequencing coverage and they represent the sequences (in this case mRNAs) that are enriched in that sample (Figure 2, step 9). This exon-specific sequencing coverage is nicely illustrated in figure 5C of Tsukamoto *et al.* (2017). To compare between samples and between transcripts, the data must be "normalized," which takes into account both the length of the gene the reads map to and how many total reads were sequenced for that sample. This normalization is called fragments per kilobase of transcript per million mapped reads (FPKM) and results in quantifying the abundance of every transcript in the *C. elegans* genome from a specific condition. Thus, the expression of every gene is measured through high-throughput sequencing and can be compared in different conditions, such as those transcripts that associate with LIN-41 *vs.* those associated with OMA-1.

Utilizing bioinformatics to identify enriched mRNAs and motifs

High-throughput sequencing results in very large amounts of data, necessitating computer programming algorithms to quantify and compare the results. The field of bioinformatics is this intersection of computer science and biological data analysis.

Principal component analysis: Since the authors sequenced mRNA from multiple sources, the researchers wanted to determine what factors best differentiate the data sets. Imagine for a moment that you have three batches of different types of cookies and that you want to determine what differences are the most important for differentiating the cookies. Do they contain chocolate chips? Or nuts? Are they made with a cookie cutter? And so on. Utilizing a statistical method termed **principal component analysis** (PCA), all of the variance (or difference) between the cookies is quantified and the factors that describe most of the variance between cookies is determined without considering the identity of the factor. Each variable in this analysis is called a principal component (PC). The first PC (PC1) explains the most variance, the second PC (PC2) explains the second greatest amount of variance, etc. When mathematically generating the PCs for the data sets, the identity of what could be contributing to the factor (in the case of cookies, chocolate chips, nuts, cookies made with a cookie cutter, etc.) is not considered. By plotting the data sets (in this case cookies) with their respective values for PC1 *vs.* PC2, the factors that the PCs have quantified can be determined by overlaying the known information about the data sets (or cookies, *e.g.*, chocolate chip cookie, peanut butter cookie, cut-out cookies, etc.). If the chocolate chip cookies and peanut butter cookies group together on PC1 separate from cut-out cookies, you could surmise that PC1 and thus most of the variance between these types of cookies is due to whether the cookies are made with a cookie cutter or not (Figure 3). By also graphing PC2, it becomes apparent that the chocolate chip cookies separate from the peanut butter cookies and the cut-out cookies, so the second greatest amount of variance is likely due to whether the cookies contain chocolate or not (Figure 3). Additionally, cookies

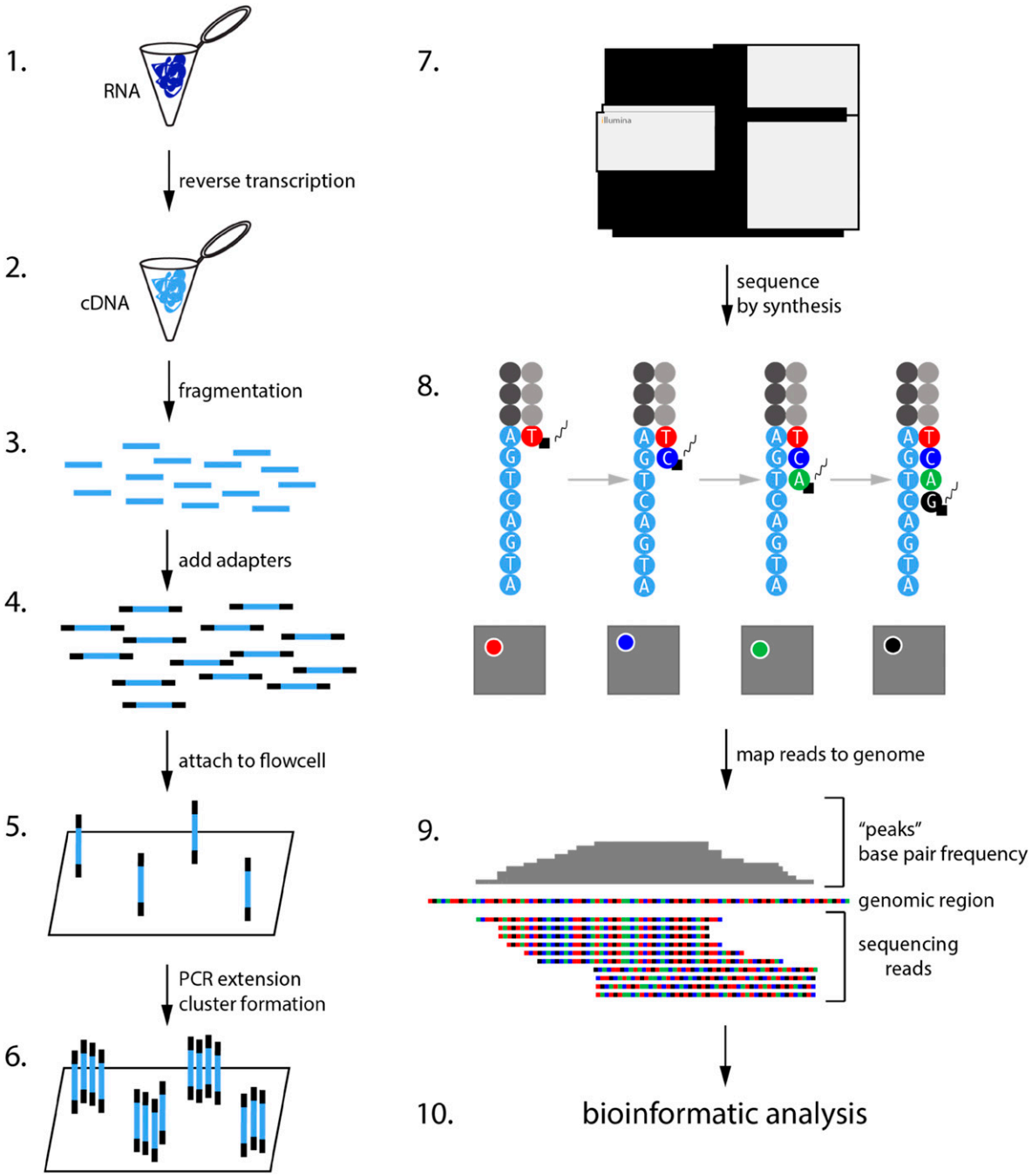

Figure 2 Overview of RNA-sequencing protocol. (1) RNA sample is prepared. (2) Reverse transcription of RNA to obtain cDNA. (3) cDNA is fragmented. (4) Sequencing library is generated by adding adapters to cDNA fragments. (5) cDNA fragments are attached to the flow cell via adapters. (6) cDNA fragments are PCR amplified using the adaptor as a primer-binding site to form clusters. (7) cDNA fragments are sequenced using (in this case) an Illumina sequencing machine. (8) Fragments are sequenced by synthesis and imaged on the flow cell. (9) Sequenced "reads" are mapped to the genome. (10) Bioinformatic analysis is performed.

of the same type group together for both PC1 and PC2, so there is not much variance within the same type of cookie.

PCA is ideally suited for use when there are many data sets with a lot of possible differences between them and the researcher wants an unbiased way to reduce the complexity of the differences to the two most important differences. Tsukamoto *et al.* (2017) compared the FPKM values for each gene in the *C. elegans* genome of eight total data sets: two biological replicates of the mRNAs associated with LIN-41, two biological replicates of the mRNAs associated with OMA-1, previously published results of mRNAs associated with OMA-1 (Spike *et al.* 2014b), and three control RNA-sequencing (RNA-seq) data sets of all the mRNAs present in the lysates of the worms before IP with LIN-41 or OMA-1. After performing

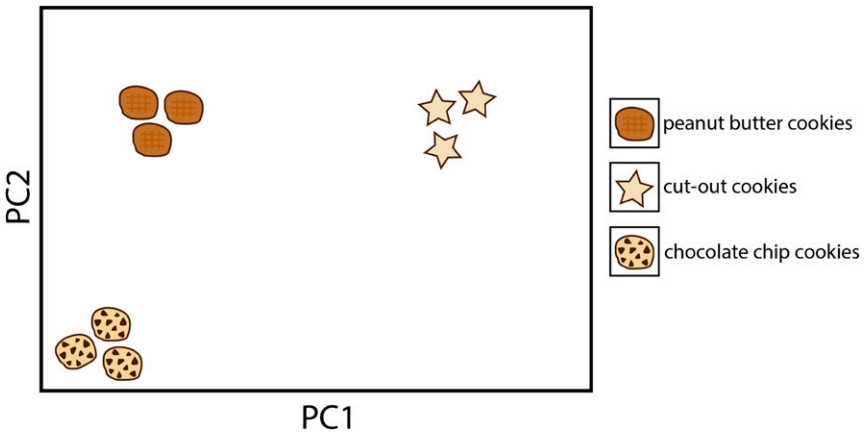

Figure 3 Principal component (PC) analysis (PCA) of cookies. Example of a PCA for three different types of cookies: chocolate chip cookies, peanut butter cookies, and cookies made with a cookie cutter (cut-out cookies). The chocolate chip cookies and peanut butter cookies have similar values for PC1, whereas the peanut butter cookies and cut-out cookies have similar values for PC2. For each type of cookie, there were three individual cookies of that type included in the PCA (nine cookies in total).

the PCA and plotting PC1 *vs.* PC2 [figure 4A of Tsukamoto *et al.* (2017)], the samples and biological replicates were color-coded to convey the known information. Interpreting the PCA results, many features of the data sets become apparent: (1) replicates cluster very closely, indicating that IP enrichment of mRNA was a very robust technique; (2) differences in mRNAs that associated with LIN-41 *vs.* OMA-1 accounted for most of the variance between the data sets (PC1, 52% of the variance); and (3) the lysate control accounted for the second most variance between the data sets. Thus, the IP enrichment of mRNA was a reliable technique to identify mRNAs associated with LIN-41 and OMA-1.

Identifying differential enrichment: Once the reliability of the data sets was verified, the next question the authors investigated was: what mRNAs associate specifically with OMA-1, specifically with LIN-41, or both OMA-1 and LIN-41? For each gene, the enrichment of that transcript in the experimental sample (the fold change) was determined by calculating the ratio of the FPKM for that gene in the experimental sample (LIN-41 or OMA-1 IP) over the amount of that transcript (FPKM) in the total lysate (negative control). Often, the ratio is $\log_2$ transformed; this can help to easily spot which genes are enriched (a $\log_2$ transformed positive value) *vs.* those that are depleted (a $\log_2$ transformed negative value). By plotting the OMA-1 enrichment *vs.* the LIN-41 enrichment for each transcript, three categories of transcripts became apparent [Figure 4, B and D in Tsukamoto *et al.* (2017)]: (1) transcripts specifically associated with LIN-41 (colored orange); (2) transcripts associated with both OMA-1 and LIN-41, including known targets of both (pink) such as *cdc-25.3*; and (3) transcripts specifically associated with OMA-1 (blue). Since RNA-seq data quantifies the enrichment of transcripts for every gene in the genome, the gray dots represent those transcripts that were not specifically enriched in any condition. For the 40 transcripts showing the most enrichment in one of the two IP samples, Tsukamoto and colleagues used a heat map to show the FPKM value in each condition represented as a color on the scale from blue (low FPKM value) to red (high FPKM value).

As described above, the *C. elegans* hermaphroditic gonad contains both oocytes and sperm, and it makes sperm before switching fate to make oocytes. The RNA sequenced was harvested from whole worms, so the authors were interested in determining if the transcripts specifically associated with either LIN-41 or OMA-1 were more enriched in the oogenic or the spermatogenic gonad. A previous study (Ortiz *et al.* 2014) used RNA-seq to identify RNAs that are specific to the spermatogenic or oogenic gonad. Tsukamoto and colleagues used **volcano plots** to visualize whether transcripts are more abundant in the oogenic or spermatogenic gonad. Volcano plots are a type of specialized scatter plot in which the magnitude of a change (in this case comparing spermatogenic expression *vs.* oogenic expression for each transcript) is plotted on the *x*-axis and the significance associated with that fold change (the *P*-value from a statistical test) is plotted on the *y*-axis. This type of plot makes it easy to identify significant data points in a large data set. In this case, each transcript (a single point on the plot) was colored based on its enrichment in the OMA-1 IP [figure 4E in Tsukamoto *et al.* (2017)] or LIN-41 IP [(Figure 4F in Tsukamoto *et al.* (2017)]. This data analysis allows the comparison of multiple related data sets to glean additional information about the transcripts and where they are expressed in the *C. elegans* gonad.

Identifying common features of transcripts: RNA-binding proteins, such as LIN-41 and OMA-1 and OMA-2, are proteins that recognize and bind to specific sequences in the regulatory regions of mRNAs. A previous study found through an *in vitro* biochemical assay that OMA-1 associates with the three nucleotide motif UA(A/U) in the 5′- and 3′-UTRs of mRNAs (Kaymak and Ryder 2013). Thus, the authors were interested in determining (1) how many of the transcripts enriched in the three different populations contain this UA (A/U) motif and (2) how many UA(A/U) motifs were present per transcript. A **violin plot**, similar to a box plot, nicely displays both of these aspects of motif distribution: the *y*-axis represents the number of UA(A/U) motifs present in a single transcript, taking into account the length of the 5′- and 3′-UTRs, and the width of the plot represents how many transcripts have that number of UA(A/U) motifs. Another study identified the binding motif for LIN-41 (Ray *et al.* 2009) as illustrated by a **sequence logo**, or a graphical representation

of a consensus motif. The *y*-axis represents the frequency of that base at that position on a $\log_2$ scale called bits. A frequency of 2 bits means that a particular nucleotide is always found at that position in all sequences used to construct the consensus sequence logo (or in evolutionary terms that position is highly conserved). Positions with more than one base stacked on top of each other show the bases that are most often found at that position. Thus, in a single graphical representation, conservation information is easily displayed. Figure 5C of Tsukamoto *et al.* (2017) shows instances of this motif in a LIN-41-associated transcript, *spn-4*. Rather than a single FPKM value for the entire transcript, this representation of the genomic locus of *spn-4* reveals the distribution of sequencing reads (coverage) for the entire transcript. In this display, there are a high number of sequencing reads covering the exons, but introns are not sequenced as is expected for sequencing mRNAs. Additionally, there is consistent coverage across the entire transcript, which shows that the intact mRNA associates with LIN-41. It is worth noting that the authors chose the method of displaying their data that best conveys what they believe to be the most important points for the reader, though there are many other ways to display the same genomic data.

Testing hypotheses generated by bioinformatic analysis

Once the authors identified mRNAs associating with LIN-41, OMA-1, or both, they utilized more traditional genetic techniques to investigate how LIN-41 and the OMA proteins regulate the translation and expression of a subset of these transcripts. For transcripts found to associate with LIN-41 or OMA-1, the authors inserted fluorescent reporters to tag those genes at their endogenous loci in the genome by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 genome editing [see Boateng and Allen (2018) or Thurtle-Schmidt and Lo (2018) for a review of CRISPR/Cas9 genome editing]. By tagging the endogenous locus, researchers visualized the spatial and temporal protein expression in live worms. Tsukamoto and colleagues then observed the protein expression of the LIN-41- and OMA-1-associated transcripts in different genetic backgrounds where LIN-41 and the OMA proteins had been mutated or knocked-down by RNA interference (RNAi). For a review of how RNAi knocks down the expression of a target gene in *C. elegans* see Meneely and Bloom (2013).

From the mass spectrometry data, Tsukamoto and colleagues found that LIN-41 and the OMA proteins associate with the poly(A) polymerase GLD-2. Extension of the poly(A) tail by GLD-2 in the cytoplasm promotes the translation of mRNAs (Nousch *et al.* 2014). Thus, the authors hypothesized that the LIN-41- and OMA-associated transcripts may also be regulated by GLD-2. To test this hypothesis, the authors analyzed another high-throughput sequencing data set from a method termed poly(A)-test RNA-seq (PAT-seq) (Harrison *et al.* 2015). PAT-seq utilizes the same high-throughput sequencing technology as described above for RNA-seq; however, the 3′ end of the mRNA is specifically isolated when preparing the nucleic acid for sequencing and the resulting cDNA is sequenced specifically from the 3′ end. This preparation results in quantification of poly(A) tail length in addition to the abundance of the transcript. By comparing the length of the poly(A) tail for every transcript in wild-type *gld-2* worms and *gld-2* null mutants, transcripts requiring GLD-2 for proper poly(A) tail length were identified. Tsukamoto and colleagues then determined whether any of their LIN-41/OMA-1-associated mRNAs had GLD-2-dependent poly(A) tail length. Their figure 14B shows each transcript specifically associated with LIN-14 (orange vertical lines) and those selectively associated with OMA-1 (blue vertical lines) ordered by poly(A) tail length change in the *gld-2* mutant. By quantifying this distribution of transcripts, the authors found that a significant portion of the LIN-41-associated transcripts are regulated by *gld-2*. Thus, *gld-2* is likely important for the translation regulation toggle between LIN-41 and the OMA proteins.

Connections to Genetics Concepts

Both the techniques and the scientific questions being investigated in this article allow students to apply core genetics concepts to interpreting primary data, a key goal of the *Vision and Change* report (Brewer and Smith 2011). Gene regulation occurs at all steps of the central dogma and this article emphasizes those points of regulation many times overlooked by students: regulation of mRNA to protein. By including this article and primer in an upper-division genetics course, students can integrate mRNA regulation into the broader context of gene expression. Additionally, this article and primer emphasizes how meiosis must be regulated to ensure proper development of the next generation. A critical concept emphasized in this primer is how the use of high-throughput sequencing and bioinformatics has enhanced classical genetic analysis. In 2006, the first next-generation sequencer was launched, which allowed scientists to sequence 1 Gb of data in a single run (Reuter *et al.* 2015). It is worth noting that though the experiments in this paper were not possible 10 years ago, these analyses are now common in genetic studies. In addition, the advent of "big data" has resulted in sophisticated statistical analyses to identify signals among the noise. Thus, this primer and paper provide an excellent background to introduce geneticists to genomic and bioinformatic methods essential for research today.

Approach to Classroom Use

Active Learning Exercise

Instructors are encouraged to provide this primer article to students concurrently with Tsukamoto *et al.* (2017). Given the diversity of experimental approaches in this paper, we suggest using the Jigsaw method (Clarke 1994) to familiarize everyone with the necessary techniques to understand the

experiments carried out in the paper. The Jigsaw method is a cooperative classroom method that requires students to work together to succeed. As the name suggests, the classroom is divided into groups and the assignment is divided into pieces such that individuals must assemble to complete the "jigsaw puzzle" assignment.

The individuals in the class should first be divided into "home" groups of four. Within each group, students will each be assigned a topic in which they will be the "expert." We suggest the following "expert" topics:

1. Gene manipulation (RNAi, CRISPR and gene-tagging)
2. Protein experiments (immunofluorescence, western blotting, and IPs)
3. Whole genome-sequencing methods (RNA-seq and PAT-seq)
4. Bioinformatics

The groups then reorganize so that each expert on the same topic forms a single group to work together to learn about the topic. Then, everyone returns to the original home group and each expert presents on their topic. It is the job of each expert to educate members of their home group about the topic they were assigned. While "becoming the expert" about the method they were assigned, students should focus on key features that are relevant for understanding this particular paper. We have included the following questions to serve as guides. Alternatively, these questions can be used simply as classroom discussion questions.

1. Gene manipulation (RNAi, CRISPR and gene-tagging)
   a. How are RNAi and CRISPR used to interfere with gene function?
   b. What are the key differences between RNAi and CRISPR?
   c. What are the advantages of using CRISPR to endogenously tag genes?

2. Protein experiments (immunofluorescence, western blotting, and IPs)
   a. What is the difference between a primary antibody and a secondary antibody?
   b. What are the advantages of using an anti-FLAG or anti-GFP antibody?
   c. Both immunofluorescence and western blots were used in this paper to visualize proteins. What are the advantages of each of these methods? What are the major differences?
   d. What types of interactions are you trying to detect with an IP?
   e. Why are IPs treated with RNase?

3. Whole genome-sequencing methods (RNA-seq and PAT-seq)
   a. What methods did Tsukamoto *et al.* use to collect RNP samples?
   b. Tsukamoto *et al.* performed two technical replicates. What is the difference between a technical replicate and a biological replicate? Is it important to perform both?
   c. What is PAT-seq and how does it differ from RNA-seq?
   d. What were the authors hoping to learn via the PAT-seq experiments?
4. Bioinformatics
   a. What are the steps that need to be taken to map reads to the genome?
   b. What is the purpose of mapping the reads to the genome?
   c. What gene features must be taken into consideration when mapping cDNA reads?
   d. Why is it important to normalize your data?
   e. Describe the different methods the authors used to represent their data.
   f. Why are sequence logo analyses valuable?

Additional questions for discussion

1. Many sexually reproducing animals arrest their oocytes in the diplotene or diakinesis stage of meiotic prophase I until additional signals are present. Why might this be an advantageous strategy?
2. The authors used CRISPR/Cas9 genome editing techniques to insert fluorescent protein tags at various locations throughout the genome. Why did they insert these fluorescent tags? Why did they choose the locations that they did?
3. What negative controls were used for mass spectrometry experiments? What was the rationale for each of these control conditions?
4. The IP experiments utilized two different antibodies sequentially. What was the purpose of each antibody and what is the advantage of using the two different antibodies as opposed to just one of the two?
5. Genes contain coding regions and regulatory regions. What is the difference between these two regions and how can regulatory regions be altered to affect gene expression? Draw a schematic of a gene and an mRNA from that gene with each of the regulatory regions labeled.
6. The authors used numerous computational methods to analyze their data. What are the advantages and disadvantages of doing so? How would you decide if computational methods are appropriate for your analyses?
7. To what extent do the choices authors make in data presentation reflect their biases and to what extent might these bias readers? Under what circumstances might this be a problem?

Acknowledgments

The authors thank David Greenstein, Tatsuya Tsukamoto, Emily Siniscalco, and Elizabeth De Stasio for helpful

suggestions and proofreading, and the Spring 2018 BIOL 22700 class (Ithaca College) for piloting the classroom activities during preparation of this manuscript. Funding was provided by the Life Sciences Research Foundation and the Gordon and Betty Moore Foundation to D.T.-S. T.-W.L. is funded by National Institutes of Health grant RGM-122001A.

Literature Cited

Boateng, R., and A. K. Allen, 2018 New role for an old protein: an educational primer for use with "the identification of a novel mutant allele of topoisomerase II in Caenorhabditis elegans reveals a unique role in chromosome segregation during spermatogenesis". Genetics 208: 79–88. https://doi.org/10.1534/genetics.117.300482

Brewer, C., and D. Smith (Editors) 2011 *Vision and change in undergraduate biology education: a call to action*. American Association for the Advancement of Science, Washington, DC.

Clarke, J., 1994 Pieces of the puzzle: the jigsaw method, pp. 34–50 in *Handbook of Cooperative Learning Methods*, edited by S. Sharan. Greenwood Press, Westport, CT.

Corsi, K., B. Wightman, and M. Chalfie, 2015 A transparent window into biology: a primer on *Caenorhabditis elegans*. Genetics 200: 387–407 (erratum: Genetics 201: 339). https://doi.org/10.1534/genetics.115.176099

Detwiler, M. R., M. Reuben, X. Li, E. Rogers, and R. Lin, 2001 Two zinc finger proteins, OMA-1 and OMA-2, are redundantly required for oocyte maturation in *C. elegans*. Dev. Cell 1: 187–199. https://doi.org/10.1016/S1534-5807(01)00026-0

Dunphy, W. G., L. Brizuela, D. Beach, and J. Newport, 1988 The Xenopus cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis. Cell 54: 423–431. https://doi.org/10.1016/0092-8674(88)90205-X

Ellis, R. E., and G. M. Stanfield, 2014 The regulation of spermatogenesis and sperm function in nematodes. Semin. Cell Dev. Biol. 29: 17–30. https://doi.org/10.1016/j.semcdb.2014.04.005

Gautier, J., C. Norbury, M. Lohka, P. Nurse, and J. Maller, 1988 Purified maturation-promoting factor contains the product of a Xenopus homolog of the fission yeast cell cycle control gene cdc2+. Cell 54: 433–439. https://doi.org/10.1016/0092-8674(88)90206-1

Gautier, J., J. Minshull, M. Lohka, M. Glotzer, T. Hunt *et al.*, 1990 Cyclin is a component of maturation-promoting factor from Xenopus. Cell 60: 487–494. https://doi.org/10.1016/0092-8674(90)90599-A

Greenstein, D., 2005 Control of oocyte meiotic maturation and fertilization, WormBook, ed. The *C. elegans* Research Community, WormBook. https://doi.org/10.1895/wormbook.1.53.1, http://www.wormbook.org

Harrison, P. F., D. R. Powell, J. L. Clancy, T. Preiss, P. R. Boag *et al.*, 2015 PAT-seq : a method to study the integration of 3′ -UTR dynamics with gene expression in the eukaryotic transcriptome. RNA 21: 1502–1510. https://doi.org/10.1261/rna.048355.114

Huelgas-Morales, G., and D. Greenstein, 2017 Control of oocyte meiotic maturation in *C. elegans*. Semin. Cell Dev. Biol. 2017 Dec 26. pii: S1084-9521(17)30051-4. https://doi.org/10.1016/j.semcdb.2017.12.005.

Kaymak, E., and S. P. Ryder, 2013 RNA recognition by the *Caenorhabditis elegans* oocyte maturation determinant OMA-1. J. Biol. Chem. 288: 30463–30472. https://doi.org/10.1074/jbc.M113.496547

Kim, S., C. Spike, and D. Greenstein, 2013 Control of oocyte growth and meiotic maturation in *Caenorhabditis elegans*. Adv. Exp. Med. Biol. 757: 277–320. https://doi.org/10.1007/978-1-4614-4015-4_10

Kumagai, A., and W. G. Dunphy, 1991 The cdc25 protein controls tyrosine dephosphorylation of the cdc2 protein in a cell-free system. Cell 64: 903–914. https://doi.org/10.1016/0092-8674(91)90315-P

Lohka, M. J., M. K. Hayes, and J. L. Maller, 1988 Purification of maturation-promoting factor, an intracellular regulator of early mitotic events. Proc. Natl. Acad. Sci. USA 85: 3009–3013. https://doi.org/10.1073/pnas.85.9.3009

Meneely, P. M., and J. C. Bloom, 2013 SMG-ly knocking out gene expression in specific cells: an educational primer for use with "a novel strategy for cell-autonomous gene knockdown in *Caenorhabditis elegans* defines a cell-specific function for the G-protein subunit GOA-1". Genetics 195: 1201–1207. https://doi.org/10.1534/genetics.113.157701

Nousch, M., A. Yeroslaviz, B. Habermann, and C. R. Eckmann, 2014 The cytoplasmic poly(A) polymerases GLD-2 and GLD-4 promote general gene expression via distinct mechanisms. Nucleic Acids Res. 42: 11622–11633. https://doi.org/10.1093/nar/gku838

Ortiz, M. A., D. Noble, E. P. Sorokin, and J. Kimble, 2014 A new dataset of spermatogenic *vs.* oogenic transcriptomes in the nematode Caenorhabditis elegans. G3 (Bethesda) 4: 1765–1772. https://doi.org/10.1534/g3.114.012351

Ray, D., H. Kazan, E. T. Chan, L. P. Castillo, S. Chaudhry *et al.*, 2009 Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat. Biotechnol. 27: 667–670. https://doi.org/10.1038/nbt.1550

Reuter, J. A., D. V. Spacek, and M. P. Snyder, 2015 High-throughput sequencing technologies. Mol. Cell 58: 586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Spike, C. A., D. Coetzee, C. Eichten, X. Wang, D. Hansen *et al.*, 2014a The TRIM-NHL protein LIN-41 and the OMA RNA-binding proteins antagonistically control the prophase-to-metaphase transition and growth of *Caenorhabditis elegans* oocytes. Genetics 198: 1535–1558. https://doi.org/10.1534/genetics.114.168831

Spike, C. A., D. Coetzee, Y. Nishi, T. Guven-Ozkan, M. Oldenbroek *et al.*, 2014b Translational control of the oogenic program by components of OMA ribonucleoprotein particles in *Caenorhabditis elegans*. Genetics 198: 1513–1533. https://doi.org/10.1534/genetics.114.168823

Thurtle-Schmidt, D. M., and T. W. Lo, 2018 Molecular biology at the cutting edge: a review on CRISPR/CAS9 gene editing for undergraduates. Biochem. Mol. Biol. Educ. 46: 195–205. https://doi.org/10.1002/bmb.21108

Tsukamoto, T., M. D. Gearhart, C. A. Spike, G. Huelgas-Morales, M. Mews *et al.*, 2017 LIN-41 and OMA ribonucleoprotein complexes mediate a translational repression-to-activation switch controlling oocyte meiotic maturation and the oocyte-to-embryo transition in *Caenorhabditis elegans*. Genetics 206: 2007–2039. https://doi.org/10.1534/genetics.117.203174

Ward, S., Y. Argon, and G. A. Nelson, 1981 Sperm morphogenesis in wild-type and fertilization-defective mutants of *Caenorhabditis elegans*. J. Cell Biol. 91: 26–44. https://doi.org/10.1083/jcb.91.1.26

Communicating editor: E. De Stasio