# Computational modeling offers new insight into *Drosophila* germ granule development

Michael Valentino,[1] Bianca M. Ortega,[1] Bianca Ulrich,[1] Dominique A. Doyle,[1] Edward D. Farnum,[1] David A. Joiner,[1] Elizabeth R. Gavis,[2] and Matthew G. Niepielko[1,2,*]

[1]New Jersey Center for Science, Technology and Mathematics, Kean University, Union, New Jersey and [2]Department of Molecular Biology, Princeton University, Princeton, New Jersey

ABSTRACT  The packaging of specific mRNAs into ribonucleoprotein granules called germ granules is required for germline proliferation and maintenance. During *Drosophila* germ granule development, mRNAs such as *nanos* (*nos*) and *polar granule component* (*pgc*) localize to germ granules through a stochastic seeding and self-recruitment process that generates homotypic clusters: aggregates containing multiple copies of a specific transcript. Germ granules vary in mRNA composition with respect to the different transcripts that they contain and their quantity. However, what influences germ granule mRNA composition during development is unclear. To gain insight into how germ granule mRNA heterogeneity arises, we created a computational model that simulates granule development. Although the model includes known mechanisms that were converted into mathematical representations, additional unreported mechanisms proved to be essential for modeling germ granule formation. The model was validated by predicting defects caused by changes in mRNA and protein abundance. Broader application of the model was demonstrated by quantifying *nos* and *pgc* localization efficacies and the contribution that an element within the *nos* 3′ untranslated region has on clustering. For the first time, a mathematical representation of *Drosophila* germ granule formation is described, offering quantitative insight into how mRNA compositions arise while providing a new tool for guiding future studies.

SIGNIFICANCE  The organization of proteins and mRNAs into membrane-less organelles called biomolecular condensates is required for diverse biological processes. In *Drosophila,* biomolecular condensates called germ granules assemble during oogenesis and are essential for reproduction. During their development, thousands of these condensates form that vary in mRNA composition with respect to the different transcripts that they contain and their quantity. Thus, *Drosophila* germ granule development provides an excellent system to explore how mRNA composition is regulated in biomolecular condensates. By combining computational modeling with biological experiments, we identified the rules influencing *Drosophila* germ granule mRNA composition while offering quantitative insight into the assembly process. Our findings may shed light on the rules that govern mRNA compositions of biomolecular condensates in other systems.

## INTRODUCTION

The development and maintenance of the germline, the set of highly specialized cells responsible for passing on genetic material to the following generation, is essential for animal reproduction (1,2). Recent discoveries revealed that both germline function and maintenance require the formation of highly conserved ribonucleoprotein (RNP) granules called germ granules (1,3–6). Germ granules are biomolecular condensates that contain proteins and mRNAs that are important for germline differentiation, proliferation,

and maintaining primordial germ cell fate through post-transcriptional gene regulation (1–8).

In *Drosophila*, germ granules are components of a highly specialized cytoplasm located at the posterior of the oocyte called the germ plasm (9,10). Germ plasm and germ granule assembly are initiated through the local production of Oskar (Osk) protein at the posterior of the oocyte and the recruitment of additional germ granule proteins, such as the conserved helicase Vasa (Vas) and Tudor (Tud), the founding member of the Tudor domain family (11–14). Among the proteins in the germ granule protein ensemble, Osk is the only one that is both necessary and sufficient for germ plasm formation, making it the primary organizer of germ plasm and germ granule formation (13). Transcripts that comprise the mRNA component of germ granules, such as

*nanos* (*nos*) and *polar granule component* (*pgc*), are maternally synthesized in support cells called nurse cells located at the anterior of the oocyte. Germ granule mRNAs are then deposited into the oocyte where they diffuse as RNPs containing single transcripts throughout the oocyte cytoplasm. These single-transcript RNPs are stochastically incorporated into granules where they become enriched by forming homotypic clusters, mRNA aggregates that contain multiple copies of transcripts from the same gene (referred to as mRNA type) (15,16). Homotypic clusters form through a stochastic seeding and self-recruitment process within a germ granule protein ensemble where the ability of the mRNA to associate with the granule requires the 3′ untranslated region (3′ UTR) but self-recruitment does not (16–18). Each granule can contain clusters of different mRNA types and the process of homotypic cluster formation generates heterogeneity with respect to mRNA types and the number of each type that reside within each granule (15,16). Germ granules continuously increase in number and grow throughout oogenesis stages ∼9–14 and in the early embryo, until Osk production ceases due to the degradation of *osk* mRNA (13,16,19). Upon fertilization, the posteriorly localized germ granules induce the formation of pole cells, primordial germ cells, which develop at the posterior pole of the embryo (11,20). Developing pole cells inherit germ granules, including their mRNA constituents; thus, germ granule mRNA heterogeneity serves as a mechanism to facilitate the simultaneous and collective segregation of many mRNA types into the primordial germ cells (15). Such maternal mRNA inheritance supplies the transcriptionally silent pole cells with mRNAs that direct the production of proteins essential to germline development, viability, and function (1,2). The importance of germ granules is highlighted by defects, including sterility, that occur in the absence of germ granules (14,21,22).

With the advancement of single-molecule fluorescent in situ hybridization (smFISH) and super-resolution microscopy, precise quantification of germ granule mRNA content has been instrumental in deciphering the germ granule assembly process (6,15–18). To study germ granule development, germ granule mRNA composition is quantified in various ways using smFISH data: 1) the number of transcripts found within a homotypic cluster (i.e., cluster size); 2) the distribution of homotypic cluster sizes found in the germ plasm; 3) the frequency at which different types of homotypic clusters populate the same granule (referred to as colocalization); and 4) the relationship between the sizes of homotypic clusters of different mRNA types that reside within the same granule (measured using Pearson's correlation coefficient) (15,16). Visual representation of the landscape of germ granule mRNA content within the germ granules is accomplished using the Granule Census, which transforms smFISH data from 3D confocal images into a 2D quantified matrix (16). Despite our ability to detect and quantify germ granule mRNA characteristics, such as

cluster size, colocalization, and correlation, questions regarding how these germ granule features emerge remain largely unexplored. Our goal is to gain insight into how germ granule mRNA composition arises using computational modeling.

Computational modeling has been used to investigate the formation of biomolecular condensates by focusing on spontaneous processes and thermodynamic parameters that are necessary to form liquid-like biomolecular condensates (23–26). *Drosophila* germ granules, however, have distinct biophysical properties. Specifically, the germ granule protein ensemble has intermediate or gel-like properties given the partial ability of Osk and Tud to exchange with the cytoplasm (18). Despite the proteins' limited and variable mobility properties, germ granules also display solid-like physical properties since mRNAs remain stable as organized homotypic clusters (16), and fusion events that are observed in liquid-like condensates (27) are not seen among homotypic clusters (16). Similarly, recent work has shown that *Xenopus* L-bodies also have a dynamic protein-containing phase and a more stable RNA phase (28). Overall, *Drosophila* germ granules are exceptionally stable, as further demonstrated by their ability to be tracked moving directionally on microtubules for long-term retention at the posterior of the oocyte (29). Moreover, *Drosophila* germ granules develop over approximately 19 h (20), a longer timescale than expected for formation of liquid-like droplets (27,30,31). To advance our understanding regarding how *Drosophila* germ granules develop given their unique biophysical properties, we developed a new rules-based computational model that calculates the probabilities for granules to be seeded and homotypic clusters to gain and/or lose transcripts using ordinary differential equations (ODEs).

In addition to confirming previously reported developmental processes, the model supports the existence of unreported mechanisms that help determine germ granule mRNA content. The model-supported mechanisms include a clustering factor, a quantifiable effect that, together with gene expression levels, regulates the sizes of homotypic clusters for any particular mRNA type. Modeling also suggests that the ability of the germ granule protein ensemble to be seeded by a transcript increases over time and that seeding competition emerges as a granule develops. Comparing modeling data with biological data, we demonstrate that the computational model simulates the formation of *Drosophila* germ granules across 19 h of developmental time with a high degree of accuracy. To increase the utility of the model, it is designed with the flexibility to interrogate different germ granule mRNAs and provide predictive power by allowing users to adjust parameters to mimic genetic perturbations. Demonstrating the robustness of the model's predictive abilities, simulations with adjusted parameters reproduced germ granule defects that occur in four genetic backgrounds, including changes in Osk production and

*nos* and/or *pgc* mRNA levels. We also show how the model quantifies clustering ability and how this property can be assigned to a specific region within a 3′ UTR by integrating biological data into the model. Such flexibilities can be broadly applied to future studies to separate the contributions that gene expression and clustering factors make in regulating total mRNA localization. For the first time in the *Drosophila* system, the germ granule assembly process is transformed into a mathematical representation that offers new mechanistic insight, delivers a continuous visualization of germ granule formation, and provides a quantitative tool for guiding future research.

## MATERIAL AND METHODS

### Mathematical representation of *Drosophila* germ granule assembly

The general process of germ granule formation, which includes seeding and self-recruitment to form homotypic clusters, has been described previously and is summarized in Fig. 1 (15,16). The model was constructed using 13 parameters that are justified based on the essential components needed to develop *nos* and *pgc* homotypic clusters within germ granules, the seeding and self-recruitment process, and theoretic behaviors that were necessary to simulate biological data (Figs. S1 *A* and 2). A developmental time window of 19 h was chosen based on the onset of Osk translation and is divided into previously described stages: late stage 9 to 10B, stage 11, stage 12, stage 13, stage 14, and the first 1 h of embryonic development (13,20). The model uses *nos* and *pgc* data as reference mRNAs given the availability of quantified smFISH data, and it was designed to output modeled granule data

using the Granule Census (15–17,32). To replicate biological data, the computational model uses a hybrid rules-based approach where the decision for a granule to be seeded, gain, or lose a transcript is determined by probabilities calculated by ODEs that are updated on a 1-min time step. The choice to calculate probabilities using ODEs is based on the hypothesis that increased mRNA concentration within the germ granule may increase the probability of mRNAs to self-recognize and self-sort (18).

The model can be viewed as two components that represent mRNA 1 (*nos*) and mRNA 2 (*pgc*). Both *nos* and *pgc* follow the same methods, but hyperparameters that are unique to each are tweaked to give the model robustness and user flexibility. The general approach to modeling is outlined as follows:

I. Update probabilities using ODEs
II. Probabilities decide if a granule is seeded, gains, or loses on each time step

In the model, the probabilities for seeding, gaining, and losing transcripts are kept for every granule at any given time step. Seeding by a given mRNA type can only occur when a granule has 0 molecules of that mRNA type. Transcript seeding is based on a seeding probability and, if a seeding event occurs, the granule will gain one transcript. Once seeded, there are four possibilities for what happens to a granule on every time step with respect to the number of transcripts in its homotypic clusters. The probabilities to gain and lose transcripts are both used on every time step to determine the net change of the granule for a particular mRNA type. Out of the four possible outcomes, two of them result in a net change of 0 (Fig. S1 *B*).

The probability for a granule to be seeded by a transcript, *s*, which is only used when a germ granule has no transcript of a given type, is given by Eq. 1:

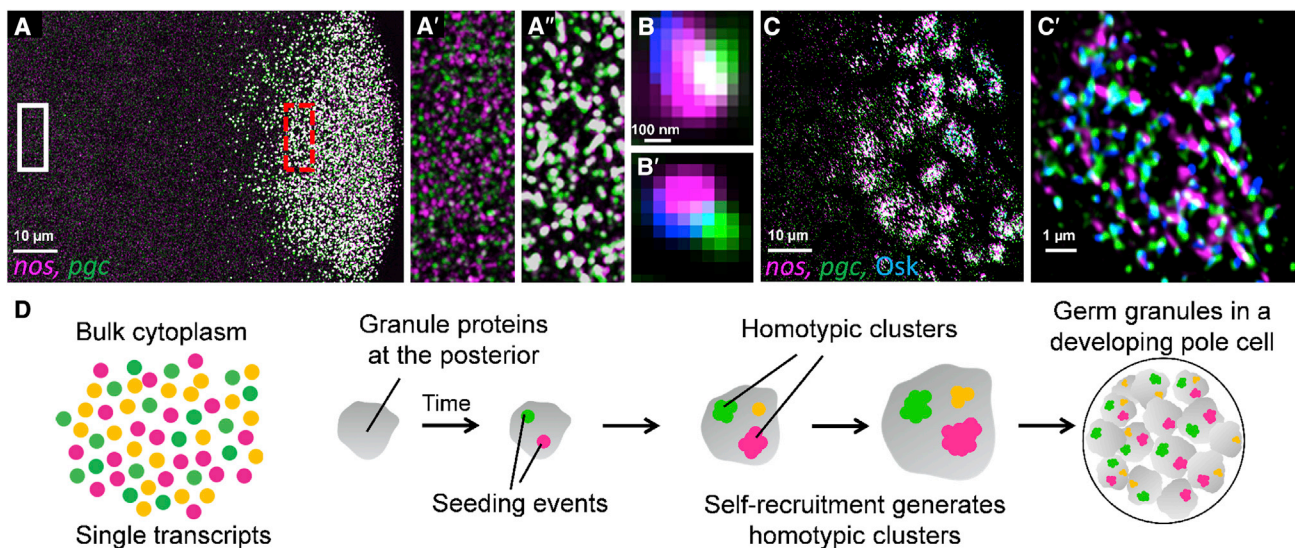$$\frac{ds}{dt} = k\left(1 - \frac{p_g}{p_t}\right)(1 - s). \tag{1}$$



FIGURE 1 Summary of *Drosophila* germ granule formation. (*A*) Confocal maximum projection of the posterior region of an early *Drosophila* embryo (posterior germ plasm is to the right). *nos* (*magenta*) and *pgc* (*green*) mRNAs were detected by smFISH. The white box denotes unlocalized single transcripts of *nos* and *pgc* in the bulk cytoplasm. The broken red box represents mRNAs that reside within germ granules in the germ plasm. (*A′*) Enlarged view of unlocalized single transcripts in the white boxed region in (*A*). (*A″*) Enlarged view of the germ plasm localized *nos* and *pgc* mRNAs that reside within germ granules in the red boxed region in (*A*). (*B, B′*) LIGHTNING super-resolution image of individual germ granules marked with Osk-GFP (*blue*), *pgc* (*green*), and *nos* (*magenta*) shows mRNA-specific subdivisions within a granule. (*C, C′*) Maximum projection super-resolution images of germ granules organizing around nuclei at the posterior of the embryo as pole cells begin to form. (*C′*) is an enlarged image of a single developing pole cell from (*C*). (*D*) Schematic summary of the germ granule formation processes including seeding, recruitment, and the incorporation of densely packed germ granules into developing pole cells.
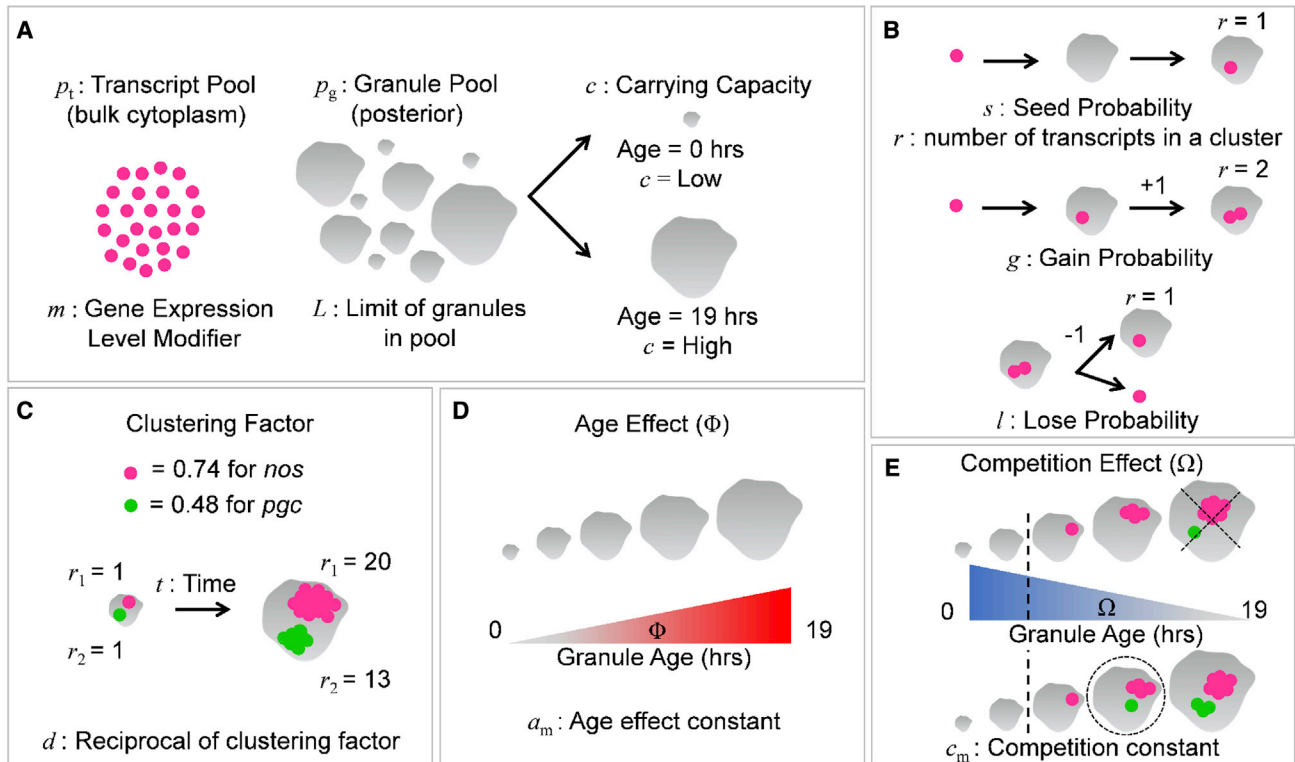
FIGURE 2   Key principles and parameters that are included in the computational model. (*A*) Parameters representing the transcript pool and gene expression for one mRNA type, the granule pool, which increases over time until its upper limit, and the carrying capacity, which increases with an individual granule's age. (*B*) Parameter used to describe the probability for a granule to be seeded by a transcript, the number of transcripts that are in a homotypic cluster, and the parameter for the probability of losing a transcript. (*C*) The clustering factor value for *nos* and *pgc*; increasing the clustering factor increases each mRNA type's clustering ability. Each mRNA's clustering factor is represented as a reciprocal, *d,* in the model. (*D*) The age effect increases a granule's probability to be seeded as it increases in age and is impacted by the age effect constant. (*E*) The competition effect generates a penalty for a granule's ability to be seeded and this penalty increases as a granule's age increases and is impacted by the competition effect constant. The competition effect is only applied to granules that contain a different mRNA than the one trying to seed as marked by the broken dotted line. Older granules that tend to have larger clusters will be least likely to accept a new seed (*broken x*), while younger granules may still seed due to a lesser penalty (*broken circle*).

Here, $k = 0.01$ is a constant, $p_g$ and $p_t$ are the granule pool and transcript pool for a particular mRNA type, respectively. The change in a granule's probability to be seeded should be proportional to the number of granules in the pool and the number transcripts in the pool. In addition, since these are probabilities, the seeding probability is bounded logistically above by 1. The transcript pool is in fact, on any given time step, always larger than the granule pool. This means that the probability to be seeded is always monotonically increasing (33). However, this probability is a base value and, overall, a granule's probability to be seeded on any given time step is a product of multiple parameters, and the age and competition effects, which are further described later.

Once seeded, a granule's ability to gain a transcript is represented by Eq. 2:

$$\frac{dg}{dt} = \begin{cases} 2k\left(\frac{cm}{d} - r\right)(1 - g) & \frac{cm}{d} - r > 0 \\ 2k\left(\frac{cm}{d} - r\right)g & \frac{cm}{d} - r \leq 0 \end{cases}. \quad (2)$$

Here, $k = 0.1$ is a constant, c is the carrying capacity of the granule, $r$ is the number of transcripts the granule currently has in its cluster for a given mRNA type, and $g$ is the current probability to gain. Of note, $c$ and $r$ are both functions of time. Finally, the variables $m$ and $d$ are special constants that are paired with each mRNA type in the model. Gene expression

(transcript levels) for both mRNA types are represented by a modifier, $m$, and accepts values from 0 to 1, where 0 is no expression and 1 is wild-type transcript expression level. Variable $d$ is the reciprocal of what we refer to as the clustering factor, a parameter that regulates each mRNA type's ability to participate in homotypic clustering and is further explained later. The probability to gain, denoted by $g$, is used on every time step of the model to potentially give a granule one new transcript after it is seeded. Each granule has an independent soft cap carrying capacity, $c$, that controls how large homotypic clusters can become. The soft cap allows for granules to recruit and contain mRNA clusters that are larger than their capacity, to some extent. Osk has the ability to leave the granule and exchange with the surrounding cytoplasm without mRNAs (18), suggesting the presence of support proteins within the granule that can stabilize clusters. Thus, a soft cap for carrying capacity was chosen to compensate for Osk's ability to leave the granule while compensating for other proteins that may help facilitate clustering to a lesser extent. However, the probability for a granule to have clusters that are larger than its carrying capacity becomes lower the larger this difference is. Since it is undesirable for the values to go far beyond their soft limit, the conditional for the piecewise differential equation is dependent on the sign of the term $cm/d$.

Since the model deals with probabilities, the probability to gain transcripts must begin decreasing before the granule actually reaches its carrying capacity (32). Otherwise, if the probability to gain decreases only after a granule hit its carrying capacity, it would still potentially gain many transcripts beyond its limit before the probability became low

enough. The final term for increasing the probability in the top equation is $(1–g)$, which gives a logistic cap to 1. The bottom equation is for when probability is decreasing, it is simply multiplied by its current probability $g$ so the value never goes below 0. Overall, the trajectory of the probability to gain is bounded between the values 0 and 1 and follows a parabolic structure.

There is also a probability for a granule to lose a transcript for each mRNA type, but this value is altered by a simple pair of arithmetic expressions. If the probability to lose decreases, we use Eq. 3:

$$l_{i+1} = kl_i. \tag{3}$$

Here, $l$ of $i$ is the current probability to lose, and $k = 0.1$ is a constant. This is equivalent to an exponential decay. If the probability to lose increases, we instead use Eq. 4:

$$l_{i+1} = f\left(1, \frac{l_i^2 + 2l_i}{1 + l_i}\right). \tag{4}$$

In Eq. 4, all variables are the same as Eq. 3. While Eq. 4 achieves an exponential-like increase for all values between 0 and 0.62, once the value exceeds 0.62, it is automatically mapped to 1 via the minimum function, $f$. Thus, there is an assumption in the model that a small number of germ granules are probabilistically unfit based on their tendency to lose transcripts.

During oogenesis, the germ plasm region and the number of germ granules increases over developmental time (15). Therefore, an additional ODE governs the growth of the granule pool and is given by Eq. 5:

$$\frac{dp_g}{dt} = kp_g\left(\frac{1 - p_g}{L}\right). \tag{5}$$

In Eq. 5, $p_g$ is the granule pool, $k = 0.35$ is a constant, and $L$ is the upper limit of this pool. The upper limit of the pool is adjustable in the model, and the default value is 16,000 granules.

The resulting probabilities from the ODE update are used in a simple yet extremely effective rules-based approach. Firstly, the seeding event probability is only used when a granule has 0 transcripts of that mRNA type in a cluster. On a time step where this criterion is true, this granule will have a chance equal to its seed probability to recruit an initial molecule of that mRNA type. If a granule already has this mRNA type in the granule, it has a chance to gain and lose a transcript depending on its probability for both events. Note that a granule may gain and lose a transcript on the same time step, resulting in a net gain of 0 (Fig. S1 B) (33).

Considering that all granules have an age associated with them from when they first formed, our model imposes the condition that older granules have larger carrying capacities. Specifically, the oldest granules in the model will have the largest carrying capacity, and the youngest granules will have the smallest carrying capacity. The design for carrying capacity dynamics is based on the comparison of Osk intensity between stage 10 and 13 oocytes and the correlation to homotypic cluster sizes (see Discussion) (16). The carrying capacity of all granules is monotonically increasing, although on an individual level granules can lose carrying capacity on any given time step, due to the ability for Osk to exchange with the surrounding cytoplasm (18). However, on an individual level, granules are overall growing larger over time (16), and there are no instances where a granule completely degrades. The carrying capacities for granule population follows an evolving weak exponential decay and was designed to be comparable with the biological distribution of Osk intensity in the embryo (Figs S2, A and B). To generate a model that represents Osk nucleation and growth, we calculated that ~16,000 granules form by factoring in that there are up to ~3 million *nos* transcripts (15), a ~4% *nos* localization rate (15,34), an average of 11 *nos* transcripts per granule in the embryo, and up to ~68% of Osk-gfp granules contain *nos* (16), (3,000,000 × 0.04/11)/ 0.68 = ~16,000 granules. By combining 19 h of developmental time, the

number for granules that form, and using the intensity distributions of Osk-gfp, we generated a model for Osk nucleation and carrying capacity growth (Video S1).

Cluster sizes are susceptible to changes in transcript levels (16). Therefore, it is important that the expression levels of *nos* and *pgc* can be independently controlled in the model. For *nos* and *pgc*, two independent transcript pools are generated by the same function of the model time, shown in Eq. 6:

$$p_t(t) = a + b(t + c)^4. \tag{6}$$

Here, $t$ is time (hours), and $a = 19,200$, $b = 21.34$, $c = 0.3833$, $p_t$ is the transcript pool for one mRNA type at a given time $t$. During oogenesis, mRNAs are transcribed in nurse cells and are continuously deposited into the oocyte through the process of nurse cell dumping (20), thus the mRNA pools must be represented as growing pools over model time. The coefficient and constants were chosen so that ~4% of the total amount of *nos* transcripts are found in germ granules at any given time step and ~3 million transcripts of *nos* transcripts are produced, based on biological observations (15,34). In the late oocyte (stages 13 and 14), *nos* and *pgc* transcripts are expressed at comparable levels (Fig. S3 A), thus the same values were used for both mRNA types in the model. How the mRNA pool changes over time and responds to different modifiers is shown in Fig. S3 B. Although, *nos* and *pgc* are expressed at equivalent levels in the oocyte, *nos* produces larger clusters than *pgc* on average (Fig. S3 A) and (15). To achieve cluster size tunability in the presence of equal expression levels, we introduced a parameter, the clustering factor, which is associated with each mRNA type and helps to influence of recruitment of a specific mRNA type into a cluster (the reciprocal of the clustering factor is represented as variable $d$ in Eq. 2). The larger the clustering factor, the better an mRNA type can be recruited into a cluster. To introduce variability into the model, this clustering factor can fluctuate based on random sampling from a truncated normal distribution where the average clustering factor for *nos* is set at 0.74 and for *pgc* is set at 0.48. The truncated distribution ensures that *pgc* does not form larger clusters than *nos* on average. Clustering factor has a direct impact on cluster size for each mRNA type, thus this factor is responsible for generating the observed difference in *nos* and *pgc* cluster sizes regardless of their equal transcript expression levels. In the model, ~4% of the total number of *nos* transcripts is found in granules, while ~2% of *pgc* transcripts accumulate in granules due to the reduced clustering factor. How changing the clustering factor value for *nos* or *pgc* affects the overall accuracy of the wild-type model is shown in Fig. S3 C.

Colocalization between *nos* and *pgc* increases over developmental time (16). To achieve this behavior in the model, we introduce an age effect, $\Phi$, which imposes a seeding penalty on granules based on age, defined as time from formation of the granule. The younger the granule, the higher the penalty it receives for its probability to be seeded and, therefore, older granules have a higher probability to be seeded. The penalty follows a sigmoid function, offset vertically by 0.5. The age effect is expressed as Eq. 7:

$$\Phi(x) = \frac{1}{1 + e^{-x*a_m}} - k. \tag{7}$$

In Eq. 7, $x$ is a vector of linearly spaced values from 0 to 1 and $a_m = 2.4$ and $k = 0.5$ are constants. It is a shifted and vertically scaled sigmoid function. As mentioned, this allows younger granules to be penalized more in their ability to be seeded.

Although introducing the age effect generated a desired increase in colocalization rate, the modeled Granule Census produced a phenomenon where the largest clusters of *nos* were colocalizing with single transcripts of *pgc* and vice versa. This phenomenon rarely occurs biologically (16) but occurred with a high frequency in our model (Fig. S3 D). To fix the model, we introduce a new competition effect, $\Omega$ (Eq. 8). The competition effect imposes a penalty for a granule to be seeded if it is already seeded by

another mRNA type. Contrary to the age effect, granules only receive a competition penalty when they are seeded by another mRNA type and the competition penalty increases with granule age (see Discussion). Germ granules that contain no mRNAs are not affected by the seeding competition penalty. The equation for the competition effect is a decaying exponential given by:

$$\Omega(x) = e^{-x*Cm} - k. \tag{8}$$

In Eq. 8, $x$ is a vector of linearly spaced values from 0.0 to 1.0, and $c_m = 2.3$ and $k = 0.16$ are constants. Both age effect and competition effect are used in conjunction with each other to produce results that are expected from biological data (Fig. S3 E) (see Discussion). Schematics that summarize each of the computational model's key parameters and behaviors are shown in Fig. 2.

To solve the ODEs, we used Euler's method, since the computational cost for using higher-order methods is high, given that this calculation is run on each time step of the model for every single granule. In the early stages, the number of granules is relatively small, but as the number grows as developmental time increases, the computational cost can increase significantly (35). The choice of using Euler's method is not without weighing the cost of accuracy. Since the model only deals with probabilities, which have a domain of (0,1) and the ODEs are variations of exponentials, our ODEs have the benefit of being smooth and bounded by small numbers. Thus, there is no significant difference when compared with higher-order methods (36). Initial conditions for new granules and equation constants were manually chosen and refined based on how well modeled data fit expected biological data and the Granule Census. How the model responds to different chosen values for respective initial conditions and constants are shown in Fig. S4. Technical replicates (n) of simulated modeled data are reported in the figure legends.

## smFISH data collection and microscopy

smFISH was carried out as described previously (15,16,37). In summary, ovaries were dissected from yeast-fed females in PBS, and 0–1-h old embryos were collected on apple juice agar plates with yeast paste. Tissues were fixed for 30 min in 4% paraformaldehyde in PBS. Prior to imaging, samples were mounted in Prolong Diamond (Thermo Fisher Scientific, Waltham, MA, USA) and allowed to cure for 3 days (15,37). Confocal imaging for quantification was performed as described in detail using a Leica SP5 confocal microscope (16) and super-resolution images were captured on a Leica STELLARIS 5 confocal microscope with LIGHTNING. Biological replicates (n) are reported in the figure legends.

## mRNA particle identification and quantification

The detection and quantification of single transcripts and homotypic cluster (referred to as mRNA particles) were carried out using a custom MATLAB (The MathWorks, Natick, MA) program that has been described previously in detail (15,38). In summary, the germ plasm area was first established with a user-defined polygon that was applied to the entire z series (15 confocal slices). We then identified mRNA particles in the germ plasm by setting an intensity threshold based on the average intensity of candidate particles in the bulk cytoplasm (RNA pools), which contain only a single transcript (15). By normalizing the intensity of germ plasm mRNA particles to the average intensity of single transcript particles, we were able to quantify the absolute number of transcripts in a homotypic cluster within each germ granule. As previously described, nos and pgc clusters were determined to reside within the same germ granule (referred to as colocalized) using a custom MATLAB program that selects colocalized pairs based on the following criteria: 1) two clusters must be within a $z$ distance of two slices for confocal images and 2) a colocalized cluster pair must also be within a distance limit of 200 nm in $x$-$y$ (15,16). Once nos and pgc clusters were

identified, quantified, and organized in colocalized pairs, the data were used to produce the Granule Census (16). The Granule Census was produced by organizing and allocating the occurrence of each observed cluster size to the proper row and column of a matrix where the row represents the number of pgc mRNAs and the column represents the number of nos mRNAs found in a cluster. Values for each matrix element were represented as a jet color scale from blue to red (16). The Granule Census was modified to use $x$, $y$, and $z$ distance thresholds to determine whether nos and pgc clusters reside in the same granule, whereas the original Granule Census program relied on marking granules with Osk-GFP (16). The modified program allowed for the generation of Granule Censuses without introducing additional Osk into various genetic backgrounds. For the super-resolution image in Fig. 1, Osk-GFP was used to mark the protein ensemble. All confocal images presented are maximum projections that were filtered by a balanced circular difference-of-Gaussian with a center radius size of 1.2 pixels and surround size of 2.2 pixels (15).

## Quantification of mRNA levels

Stage 13/14 oocytes were dissected from yeast-fed females. Following tissue homogenization, RNA was extracted using RNeasy kit (Qiagen, Hilden, Germany, cat no. 74104) following the standard protocol. cDNA synthesis was performed using QuantiTect Reverse Transcription kit (Qiagen, Hilden, Germany, cat no. 205311). Gene expression assay was performed using TaqMan Gene Expression Assays (Thermo Fisher Scientific, Waltham, MA, USA), nos (cat no. 4351372), rpl7 (cat no. 4351372), and a custom assay was designed for pgc (context sequence, TGGAACATCGTGAATGCACTTTTGA). TaqMan assay master mix (cat no. 4369514) was used for all assays, which were preformed using the Bio-Rad CFX96 Real-Time System. For all qPCR experiments, three technical replicates for each of three biological replicates were performed. Each biological replicate included >15 stage13/14 oocytes from multiple females. To calculate gene expression, we first generated standard curves by amplifying nos, rpl7, pgc using standard PCR and the primer sets: nos (fwd-CCACTGTGTCCACCAATCTCG, rev-TTTGGGGCACAGCACTCGGTTAAAG), rpl7 (fwd-TCCGCGCCGAGAAGTACCAGAATG, rev-CGCAGCATGTTGATGGTGGCCTTGTT, and pgc (fwd-GTCATCGCGGATAGATGGAGAT, rev-AAACAATGCGAGTTTTCACGA). Next, we calculated template copy number based on the amplicon length and concentration and completed an eight-step serial dilution. Each dilution was used to generate Ct values for a given template copy number (Fig. S7). Absolute transcript quantities for nos and pgc were normalized to internal rpl7 transcript values.

## Statistical analyses

All non-Germ Granule Census graphs were created using R statistical programming and the ggplot2 package in R studio (39–41). Pearson's r was calculated using R statistical programming with the cor() function. To determine the significance between average cluster sizes, correlation, colocalization, and slopes, $p$ values were calculated using a two-sample $t$-test between modeled and biological data with the t.test() function (39). Standard error of the mean is represented by $\pm$ for all replicates. Percent overlap values between density plots were calculated using the overlapping R package (42,43). To visualize expected cluster size targets in Fig. 4 B, corner boundary coordinates were generated by pairing the four different combinations of upper and lower limits of average and maximum cluster sizes that were observed in biological germ plasms. To identify outliers produced by the model in reference to biological data, we performed the following analysis: 1) germ plasm's average cluster size ($x$ axis) was paired with its max cluster size ($y$ axis) from biological replicates and from randomly generated modeled germ granules. Each of these pairs was plotted and analyzed in a stage- and mRNA-specific manner. 2) The centroid for stage-specific biological data was calculated using Euclidian distances. 3) The 1.5× (IQR)
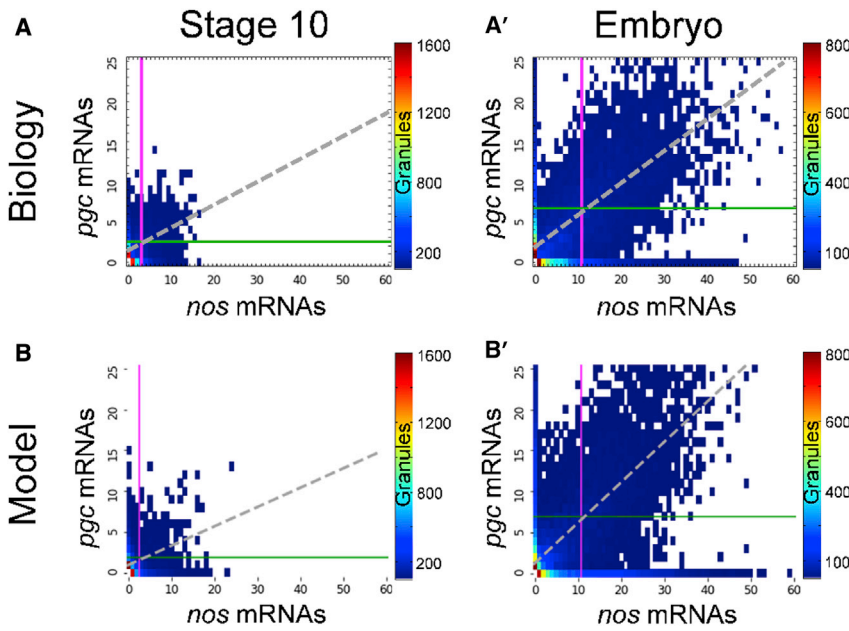
FIGURE 3 Modeling germ granule formation produces Granule Censuses that are comparable with biology. (*A, A'*) Granule Census produced from analyzing biological germ granules at stage 10 and in the early embryo. (*B, B'*) Granule Census generated based on modeling germ granule formation at stage 10 and in the early embryo. For all censuses, the average *nos* cluster size (*magenta vertical lines*) and average *pgc* cluster size (*green horizontal line*) increase from stage 10 to early embryo. The relationship between the sizes of *nos* and *pgc* mRNAs that reside in the same germ granule can be visualized using a line of best fit for colocalized *nos* and *pgc* clusters (*broken gray line*). The heatmaps indicate the number of granules with each observed mRNA composition, $n = 11$ biological germ plasms for both stages and $n = 10$ random simulations for modeled data.

to the third quartile rule was applied to the distances of modeled data from the biological centroid to identify outliers. Outliers were quantified and the frequency with which the model produces outliers on average was determined by replicating the outlier analysis 10 times using 10 different sets of randomly generated modeled germ granules for each analysis. To conduct the *k*-means and entropy analysis, we pooled biological and
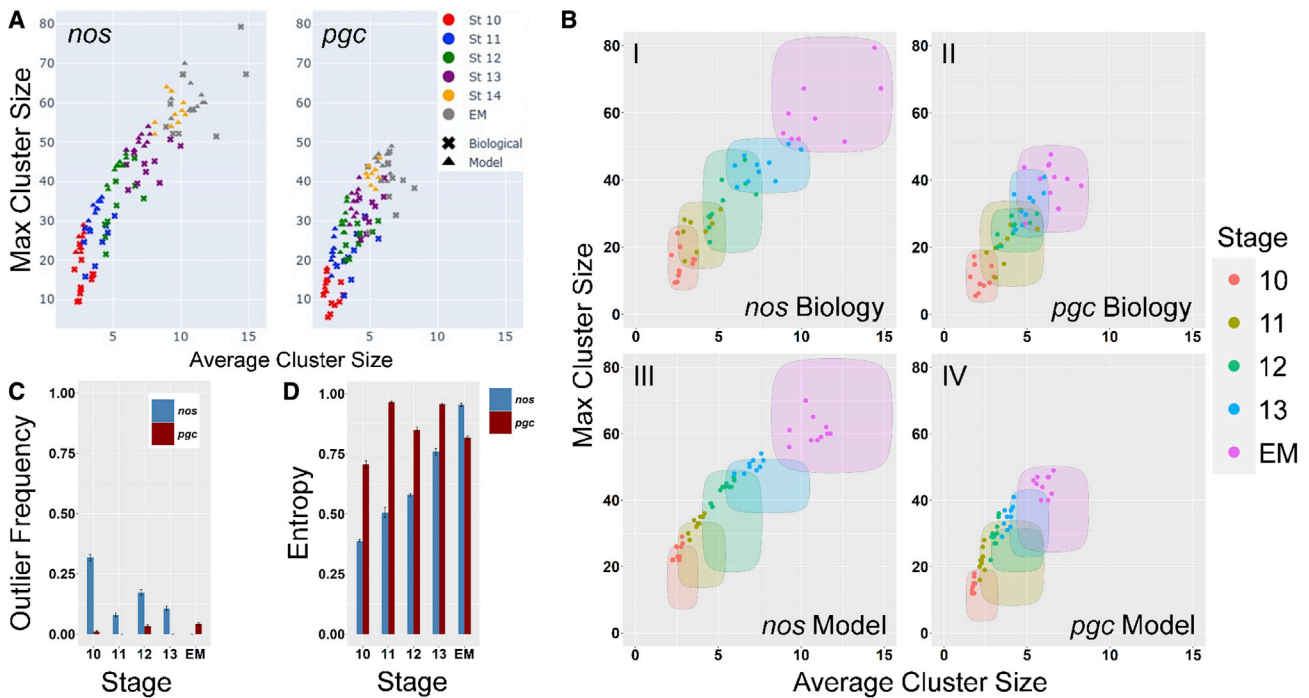


FIGURE 4 Computational modeling of homotypic cluster formation captures expected cluster sizes for both *nos* and *pgc* across developmental stages. (*A*) The maximum cluster size plotted against the average cluster size for *nos* and *pgc* from modeled and biological germ plasms. Plotted data reveal that the modeled homotypic clusters qualitatively aggregate with biological data on the graph at all stages for both *nos* and *pgc*. (*B*) Maximum cluster size plotted against average cluster size. Shaded regions indicate expected clusters sizes at a given developmental time for biological (top row, panels I and II, $n > 8$ biology germ plasms for each stage) and model data (bottom row, panels III and IV, $n = 10$ random simulations). (*C*) Analysis of outliers that are produced by the computational model as measured from the centroid generated from biological data. (*D*) Entropy values from *k*-means analyses of biological and modeled maximum and average clusters sizes for *nos* and *pgc*. Entropy scale is 0 to 1, where 0 represents the absence of any mixing of biological and modeling data and 1 for an even mixture of modeling and biological data from the *k*-means analysis. In both (*C*) and (*D*), *nos* data are blue and to the left of *pgc* data (*red*) for each stage. Error bars represent mean ± SE from 10 replicated analyses.

modeled cluster size pairs (average clusters size on the $x$ axis paired with the max clusters size on the $y$ axis) and set $k = 2$. Next, we tracked which elements corresponded to biological and modeled data and calculated the entropy in the data sets that were produced from $k$-means results. An entropy score of 0 represents complete separation of biological data from modeled data while a score of 1 represents an equal mixing of biological and modeling data. The average entropy between modeled and biological data was calculated by replicating $k$-means 10 times using 10 different sets of randomly generated modeled germ granules for each analysis in a stage- and mRNA-specific manner. Additional details about statistical analyses can be found in the figure legends and the Results section.

To measure how well the model produces biological germ granule mRNA compositions overall, we developed an accuracy score that takes into consideration how well the model predicts the six measurable mRNA germ granule compositions. The scoring system is on a scale from 0.0 (least accurate) to 1.0 (most accurate). To calculate the overall accuracy score, we first identified the centroids for each of the six measurements from replicated biological data (*nos* average cluster size, *pgc* average cluster size, *nos* maximum cluster size, *pgc* maximum cluster size, colocalization rate, and correlation) in a stage-specific manner. Next, the root-mean-square deviation (RMSD) for each measurement was calculated using biological centroid values as the expected results and 10 randomly modeled data points as the predicted values in a stage-specific manner. RMSD values were scaled using the mean. To produce a single scoring value that represents the model's overall performance, we calculated the Euclidean norm of individual stage-specific vector that contained all six scaled RMSD values. In principle, the most inaccurate model has a vector that contains six scaled RMSD values that are each equal to 1 and a Euclidean norm that is equal to the square root of 6, ~2.4495. Stage-specific Euclidean norm values were divided by 2.4495 and then subtracted from 1.0. This results in a scale in which 1.0 represents perfect accuracy, and 0.0 is the least accurate. Using a similar approach, the overall accuracy of individual parameters, such as colocalization, can also be scored across all stages.

## Fly strains

The $y^1$, $w^{67c23}$ strain (Bloomington *Drosophila* Stock Center 6599) was used as the wild-type. Females heterozygous for the $nos^{BNX}$ (16), $pgc^{\Delta 1}$ (44), and $osk^{A87}$ (45) were used to create $1\times$ *nos*, $1\times$ *pgc*, $1\times$ *nos* and $1\times$ *pgc*, and $1\times$ *osk*. The *osk-gfp* transgene (fTRG_1394) was a gift from H. Jambor (46). The *nos*Δ*3* transgene was created by deleting nucleotides 185–403 from the *nos* 3′ UTR in a 4.3-kb genomic *nos* rescue fragment (22). The *nos* sequences were cloned into the pattB vector and inserted into the attP40 landing site by phiC31-mediated recombination (47). One copy of the transgene was introduced into $nos^{BNX}$ homozygous females to create the $1\times$ *nos*Δ*3* strain.

## RESULTS

### Computational modeling produces biology-like germ granule landscapes

Our current understanding of *Drosophila* germ granule assembly, including a seeding and self-recruitment process that forms homotypic clusters, has been described previously (16) and is summarized in Fig. 1. To gain quantitative insight into the germ granule assembly process, we developed a new rules-based computational model that uses ODEs to calculate the probabilities for granules to be seeded and homotypic clusters to gain and/or lose a transcript (Materials and methods and Fig. 2). The model's key parameters

were identified based on the essential components needed to develop *nos* and *pgc* homotypic clusters (Figs. 1 *D*, 2 *A*, and 2B and Materials and methods). Using the key parameters, the model incorporates the following principles that dictate the composition of germ granules and are justified from previously published and/or new biological data: 1) based on the log-normal distribution of *nos* cluster size (15), as homotypic cluster size increases in the model, clusters have a higher probability to gain a transcript; 2) given that Osk protein levels tend to increase within a granule over developmental time (16), the carrying capacity of germ granules generally increases over developmental time in the model; 3) modeled germ granules' carrying capacity follows the distribution of Osk protein found in germ granules (Fig. S2, *A* and *B*); 4) live imaging revealed that mRNA is stable in large homotypic clusters (16). Therefore, cluster stability increases with homotypic cluster size in the model; 5) in the late oocyte (stages 13 and 14), *nos* and *pgc* are expressed at equivalent levels, yet *nos* produces larger clusters than *pgc* on average (Fig. S3 *A*). To fine-tune cluster size in conjunction with expression levels, the model assigns a value called the "clustering factor" that influences the recruitment of a specific mRNA type into a cluster (Fig. 2 *C*).

In addition to experimentally justified parameters and principles of germ granule development, we added theoretical behaviors to the model (Fig. 2 *D* and *E* and Material and methods): 1) as granules increase in age, they gain a higher probability to be seeded by a transcript, we reference this behavior as "age effect" and 2) as an mRNA type tries to seed a granule that has already been seeded by a different mRNA type, the probability for seeding is reduced by applying a penalty to the "age effect." We reference this behavior as "seeding competition" and this penalty increases as the age of individual granules increases. Once a granule is seeded by a particular mRNA type, the model assumes that there are four possible outcomes that can occur with respect to homotypic cluster growth on each time step (Figs. S1 *B* and S1 *C*).

To begin testing our model, we analyzed the size distribution of modeled *nos* homotypic clusters that contain four or more transcripts in the early embryo and found that it resembles the expected log-normal distribution as previously described for biological data (Fig. S2 *C*) (15). Next, we tested the model by visualizing the germ granule landscape that it produces by outputting modeled data as a quantified matrix called the Granule Census. In the Granule Census, each element of the matrix represents a unique combination of *nos* homotypic cluster size (*x* axis) and *pgc* homotypic cluster size (*y* axis), and a colormap corresponds the number of granules with that size combination. Granules that do not contain at least one *nos* or *pgc* transcript are not included in the matrix (16). The Granule Census has been employed to depict the transformation that occurs in the germ plasm from stage 10 to stage 13 as homotypic clusters grow (16). Comparing

biological data and modeling data, we find that the computational model successfully reproduces these dynamics and is visualized in the Granule Census for all stages. Specifically, modeling data capture a gradual approximately threefold increase in the average size of *pgc* homotypic clusters and the approximately fourfold increase in the average size of *nos* homotypic clusters that occurs biologically from stage 10 to early embryo (Fig. 3). Furthermore, the model also detects the shift that occurs in the relationship between the sizes of *nos* and *pgc* mRNAs that reside in the same germ granule as visualized using a line of best fit (Figs. 3 and S5). Using the model, we predict the germ granule landscape for the latest stage in oogenesis, stage 14 (Fig. S5), which is difficult to accurately analyze biologically due to the presence of a fully developed eggshell that interferes with probe penetration and imaging (16). Representation of quantified biological germ granule data is limited to fixed samples and a static Granule Census for each stage. Using the computational model, continuous visualization of germ granule formation is made possible by using a dynamic Granule Census, providing a unique monitoring prospective that cannot be accomplished using current live imaging methods (Video S2). Together, our computational model captures expected homotypic growth dynamics that can be visualized continuously using an animated Granule Census across 19 h of germ granule development.

## The computational model captures expected *nos* and *pgc* homotypic clusters sizes

The full distribution of homotypic cluster sizes identified within an entire germ plasm can be summarized by identifying the homotypic cluster with the largest number of transcripts (max cluster size) and plotting it against the average number of transcripts found in a homotypic cluster (average cluster size) for *nos* and *pgc* (16). This dimension reduction allows for the distribution of homotypic clusters to be represented as a single point on a graph for easier comparisons (16). To test the accuracy that our model reproduces biological observed germ granules, we analyzed germ granule data from 10 randomly generated modeled germ plasms from stage 10 to early embryo. For each modeled germ plasm, we summarized the data by calculating the average and maximum cluster sizes for *nos* and *pgc* at all stages. Plotting the summarized modeled data together with biological data revealed that both data sets qualitatively organize together in a stage-specific manner. Furthermore, our model predicts the sizes of homotypic clusters that are expected at stage 14 (Fig. 4 A).

To visualize the precision of our model's homotypic clustering ability, we first marked an area of the graph where homotypic cluster sizes are expected to be plotted based on biological data (referred to as targets) and plotted modeled data on top of the targets (Fig. 4 B). In total, 86% of the modeled *nos* (43 out of 50) and 84% of modeled *pgc* (42 out of 50) cluster points plotted in their expected regions.

Most importantly, 100% of modeled homotypic clusters sizes were plotted in the expected region both *nos* and *pgc* in the embryo, the end point of germ plasm development (Fig. 4 B).

Next, we measured the frequency with which the model produced homotypic cluster outliers in reference to biological data (Fig. 4 C and Material and methods). Outliers produced by the model were largely restricted to *nos* at stage 10 at an average frequency of $0.32 \pm 0.01$. Homotypic cluster outliers were not observed for *nos* and were rarely detected for *pgc* ($0.04 \pm 0.01$) in the embryo (Fig. 4 C). To quantify how well modeling data resemble biological data, we performed a *k*-means analysis and calculated entropy (see Material and methods). The *k*-means and entropy analysis determined that modeled data and biological data could not be completely separated at any stage for either mRNA type. Specifically, *nos* at stage 10 had the lowest entropy at $0.38 \pm 0.01$ (least similar), while the highest entropy was measured for *nos* in the embryo at $0.95 \pm 0.01$ (most similar). As for *pgc*, entropy was measured between 0.70 and $0.97 \pm 0.01$ across all stages (Fig. 4 D). The mixing of modeled and biological data across all stages demonstrates the model's ability to replicate the development of germ granules is comparable with biological data.

The most important goal for the computational model was to be able to reproduce the final state of germ granules in the germ plasm, just prior to when they induce pole cell formation and segregate into the pole cells (Fig. 1 C). Therefore, we further tested the precision of germ granule modeling by focusing on data from the early embryo. Specifically, we compared the distributions of mRNA content in granules between modeled and biological data by measuring the overlap between density plots. The comparison revealed considerable overlap between modeled and biological data, 88% for *nos* and 90% for *pgc*. These overlap values were comparable with differences in control overlaps between biological replicates, 84% for *nos* and 85% for *pgc* (Fig. S6). In addition, the average homotypic cluster sizes in the embryo were not significantly different between modeled and biological data for both *nos* ($10.7 \pm 0.3$ vs $11.0 \pm 0.62$, $p = 0.68$) and *pgc* ($6.0 \pm 0.13$ vs 6.4, $p = 0.21$) (Fig. 3, A′ and B′). Based on the results from our comprehensive homotypic cluster analysis, we conclude that the computational model captures expected homotypic cluster growth across 19 h of developmental time with a high degree of accuracy and confidence.

## Colocalization and correlation dynamics between *nos* and *pgc* homotypic clusters are recapitulated in modeled germ granules

The rate at which *nos* and *pgc* populate the same granule, referred to as colocalization (co-loc), is dynamic from stage 10 (average co-loc $34 \pm 2.4\%$) to stage 13 (average co-loc $47 \pm 1.4\%$) (16). To compare how well our model captures colocalization dynamics, we modeled the formation of germ

granules in 10 germ plasms and measured the colocalization rates for *nos* and *pgc*. Similar to previously published biological data, modeled colocalization rates were dynamic from stage 10 (average co-loc 34 ± 0.2%) to stage 13 (47.5 ± 0.4%). Our results also show that the modeling data recapitulate the average colocalization rate in the early embryo (biology = 54 ± 1.7% and modeling = 52 ± 0.5%). For all developmental stages, our results demonstrate that modeled colocalization rates between *nos* and *pgc* are within the expected biological range and averages are not significantly different for all stages ($p > 0.21$) (Fig. 5 A).

The relationship between the sizes of colocalized *nos* and *pgc* homotypic clusters is also dynamic over developmental time. First, at stage 10, the sizes of *nos* and *pgc* colocalized clusters are moderately correlated (Pearson's r = 0.4) and, by the early embryo, colocalized *nos* and *pgc* homotypic clusters are strongly correlated (Pearson's r = 0.7) (16). In agreement with the biological data, we found that colocalized *nos* and *pgc* homotypic cluster size was moderately correlated at stage 10 (average Pearson's r = 0.34 ± 0.05) and strongly correlated in the early embryo (average Pearson's r = 0.71 ± 0.01) (Fig. 5 B). Like biological data, modeling results also produced a moderately correlated relationship between colocalized *nos* and *pgc* homotypic cluster size (average Pearson's r = 0.31 ± 0.02), which increased to a strong correlation in the early embryo (0.71 ± 0.01) (Fig. 5 B). Correlation values between biological and modeled data were not significantly different for all stages ($p > 0.07$) except for stage 12 ($p = 0.02$). However,

this difference was small (0.65 ± 0.01 vs 62 ± 0.01) and fell within the expected biological range. We conclude that, in addition to capturing homotypic cluster size, the computational model produced accurate representations of germ granule mRNA composition across all developmental stages as measured by colocalization rates and correlation between *nos* and *pgc* cluster sizes within shared granules (Fig. 5, A and B).

## Dynamic colocalization rate and correlation is generated by variation in the probability for granules to be seeded

Both colocalization rate and correlation increase over developmental time (16) (Fig. 5). To achieve these granule properties, we expanded the computational model to incorporate two additional unreported mechanisms that function through the germ granule protein ensemble. First, we added a mechanism that allows for an increase in the probability that the granule protein ensemble is seeded as it continues to develop, referred to as "age effect" (see Materials and methods and Discussion). When the age effect is not included in the model, "no age effect model," colocalization rates between *nos* and *pgc* decrease over time, starting at 69.5 ± 0.2% at stage 10 and ending at 58.8 ± 0.2% in the embryo, which is opposite to what occurs biologically (Fig. 5, A and C). In addition, correlation between colocalized *nos* and *pgc* was static at 0.56 ± 0.001 from stage 10 to 0.58 ± 0.004 in the early embryo (Fig. 5 D). The second
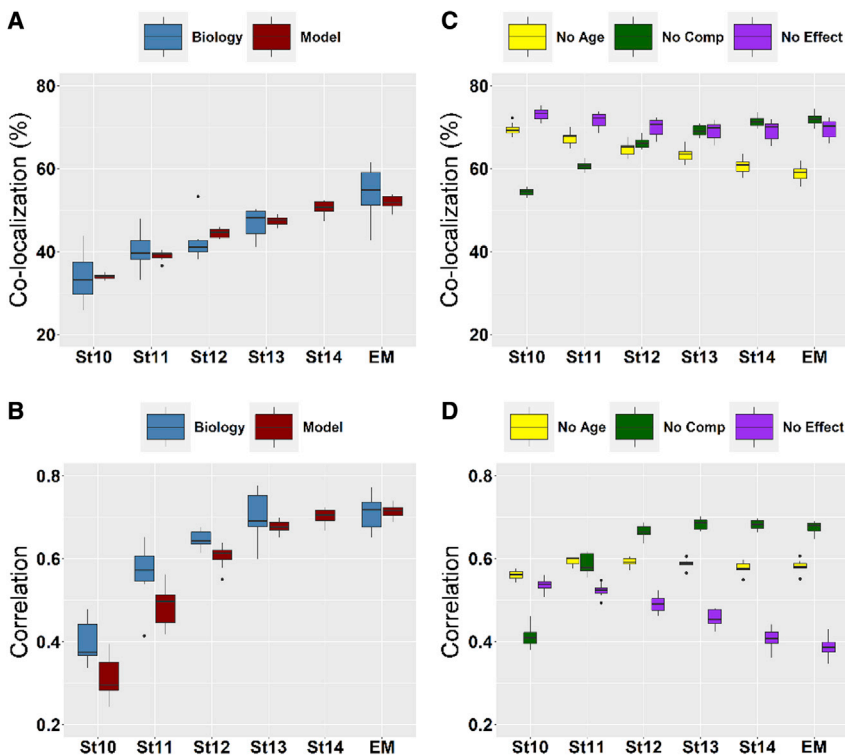


FIGURE 5 Colocalization and correlation dynamics between *nos* and *pgc* are captured by the computational model. (A) Colocalization rates between *nos* and *pgc* homotypic clusters from both biology and modeled data for all stages. (B) Correlation between the sizes of colocalized *nos* and *pgc* homotypic clusters, as measured by Pearson's r, from biology and modeled data for all stages. The model generates a colocalization rate and correlation value that is expected for stage 14 and cannot be extrapolated from biological data due to experimental limitations. In (A) and (B), biological data are blue and represented to the left of modeled data (*red*) for each stage. (C) Colocalization rates between *nos* and *pgc* homotypic clusters in computational models where there is no age effect, no competition effect, or neither effect (No Effect). (D) Correlation coefficient, as measured by Pearson's r, calculated between colocalized *nos* and *pgc* homotypic clusters in computational models where there is no age effect, no competition effect, or neither effect (No Effect). In (C) and (D), no age effect is yellow, no competition effect is green, and no effect is purple, $n > 8$ biology germ plasms for each stage and $n = 10$ random simulations for modeled data.

previously unreported granule behavior incorporated into the model is referred to as the "seeding competition effect" (see Materials and methods and Discussion). When the seeding competition effect is eliminated from the model, referred to as the "no competition model," colocalization rates increase from stage 10 to embryo. However, the values were higher than observed in biological data, starting at 54 ± 0.28% at stage 10 and ending at 72 ± 0.5% in the embryo (Fig. 5 *C*). Furthermore, this model does not produce a comparable Granule Census (Fig. S3 *D*).

To compare the performance and the overall accuracy of the four different models, wild-type, no age, no competition, and no effect, we developed an accuracy score that measures how well the model generates germ granules with biological mRNA compositions (see Materials and methods). Using the accuracy score, we determined that the wild-type computational model increases its overall accuracy over time, starting at 0.70 (stage 10) and ending at 0.91 (embryo). The no age effect model ranged from 0.27 to 0.89, no competition 0.54 to 0.83, and no effect ranged from 0.18 to 0.75 (Fig. 6). Together, our modeling data demonstrate that the model's overall performance decreases without the age and competition effects and data generated do not resemble the biological data (Figs. 5 and 6). The combination of age and competition effects assigns germ granule protein ensembles with varying and dynamic probabilities to accept a seed transcript. Thus, the model reveals that the probability that an mRNA will seed a granule varies highly among germ granule protein ensembles (Video S3). These modeling results support the presence of mechanisms

that affect mRNA's ability to seed during germ granule formation and provide insight into how the protein ensemble affects germ granule mRNA composition.

## Effects from genetic perturbations are predicted using computational modeling

Germ granule mRNA composition is affected by the expression levels of germ granule mRNAs and Osk protein (16). To test our model and validate its performance, we conducted smFISH experiments using oocytes with reduced *nos*, *pgc*, *nos*, *pgc*, or *osk* expression levels and compared biological results to modeling results (Fig. 7). First, we determined expression level using qPCR and determined that 1× *nos* stage 13/14 oocytes had 63 ± 6.7% of the wild-type *nos* level, while 1× *pgc* oocytes had 73 ± 10.5% of the wild-type *pgc* level (Fig. S7, *A–D*). 1× *osk* oocytes have been previously been reported to express 60% of wild-type Osk levels (48). Consistent with previous findings, reducing the levels of *nos* and/or *pgc* reduced their respective homotypic cluster sizes, while reducing Osk levels affected both cluster types (Fig. 7, *A–D'*) (16). Next, we predicted the Granule Census for each genetic background by configuring the model's parameters to match the experimental expression levels: 1× *nos* (RNA pool 1 = 0.63), 1× *pgc* (RNA pool 2 = 0.73), and 1× *nos* and 1× *pgc* (RNA pool 1 = 0.63 and RNA pool 2 = 0.73), and 1× *osk* (carrying capacity = 0.6 and granule number = 0.6). In all genetic backgrounds tested, the Granule Census was remarkably similar between biological data and modeling counterparts (Fig. 7). Specifically, average *nos* and *pgc* cluster sizes produced by the perturbed models were comparable with the expected biological values. Average colocalization rates generated by each perturbed model also fell within or near the expected ranges for their biological experiment counterpart (Figs. 7 and S7 *E*). Changes to germ granule mRNA composition can be quantified by measuring the slope of the line of best fit from plotted sizes of colocalized homotypic clusters (16). The perturbed models captured the change in slope direction that agrees with previously published data (Figs. 7 and S7). Together, these data demonstrate the accuracy of the model in the context of non-wild-type parameters and show the model's robustness by confidently capturing perturbations to germ granule mRNA compositions by only adjusting the parameters affected in the biological experiments.
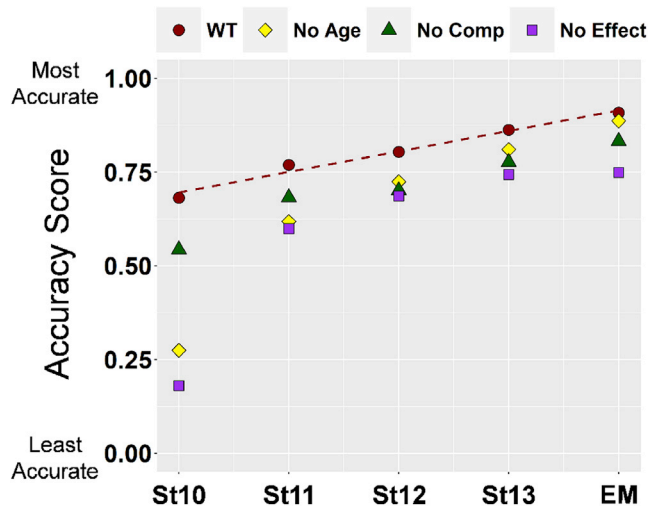


FIGURE 6 Quantification of the model's overall performance and accuracy. (*A*) The model's accuracy score (*y* axis) for each of the five stages (*x* axis). Red circles represent the wild-type biological model, yellow diamonds represent data from a model that lacks the age effect, green triangles represent data from a model that lacks the competition effect, and purple squares are from a model with no age and no competition effect. The broken red line is the line of best fit for the wild-type model and shows how the model improves its accuracy over time.

## Validation and quantification of clustering factor

The presence of "clustering elements," specific regions found within the 3′ UTR of *nos* and *pgc* that regulate homotypic clustering, have been identified using reporter assay experiments (17). This strategy identified a *nos* clustering element within nucleotides 185 to 403 of the *nos* 3′ UTR (designated as the +3 element) (17,49). To further validate
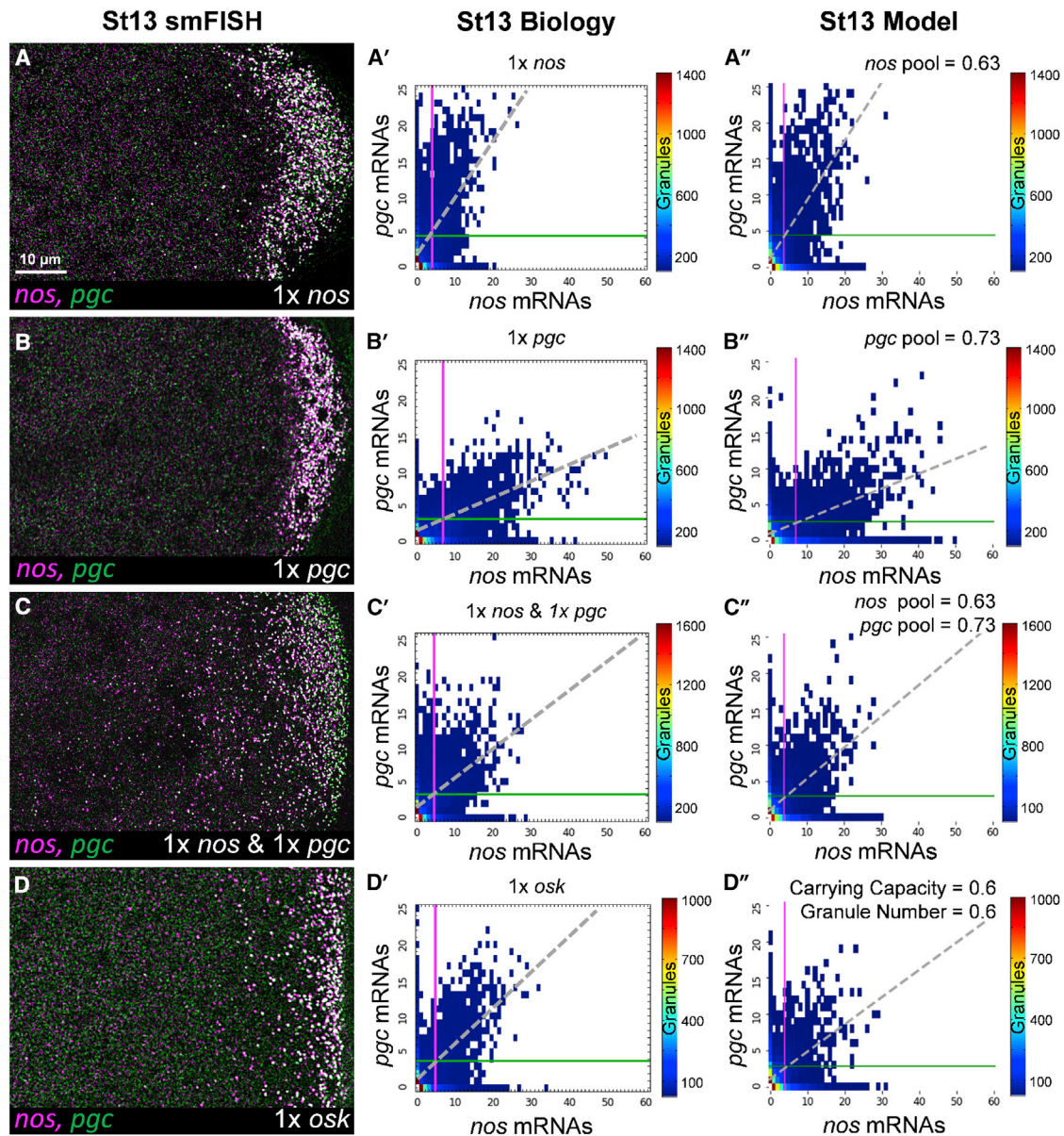
FIGURE 7 The computational model reproduces effects caused by genetic perturbations. (*A*–*D*) Stage 13 oocytes from the indicated genetic backgrounds, with *nos* (*magenta*) and *pgc* (*green*) mRNAs are detected using smFISH: (*A*) reduced *nos* expression (1× *nos*, *n* = 5); (*B*) reduced *pgc* expression (1× *pgc*, *n* = 5); (*C*) reduced *nos* and *pgc* expression (1× *nos* and 1× *pgc*, *n* = 9); and (*D*) reduced Osk expression (1× *osk*, *n* = 6). All images are confocal maximum projections with the posterior germ plasm oriented to the right. (*A′* – *D′*) Granule Census generated from biological germ plasms for each genetic background. (*A″*–*D″*) Granule Census generated from modeled germ granules for each genetic background, *n* = 10 randomly generated model simulations. Parameters used to predict genetic effects are listed within each census. For all censuses, the average *nos* cluster size is indicated with a magenta vertical line, while average *pgc* cluster size is indicated with a green horizontal line. The relationship between the sizes of *nos* and *pgc* mRNAs that reside in the same germ granule is visualized using a line of best fit for colocalized *nos* and *pgc* clusters (broken gray line). The heatmaps indicate the number of granules with each observed mRNA composition.

the function of this element in a more native context, we generated *nos* RNA null mutant flies carrying one copy of a genomic *nos* rescue transgene with this region deleted (1× *nos*Δ*3*) (Fig. 8 *A*). The expression level of 1× *nos*Δ*3* mRNA is nearly equivalent to two copies of *nos* in wild-type stage 13/14 oocytes (91 ± 8.6%). In contrast, in 1× *nos* ovaries, the level of *nos* is 63 ± 6.7% of the wild-

type *nos* level, demonstrating that the single copy of the *nos*Δ*3* rescue transgene is robustly expressed (Fig. S7 *D*). Despite having comparable expression levels with wild-type, 1× *nos*Δ*3* produced *nos* homotypic clusters that were on average 57% smaller than wild-type (3.24 ± 0.15 vs 7.60 ± 0.42, *p* < 0.001) and, despite having ~30% more *nos* transcripts than 1× *nos*, 1× *nos*Δ*3* produced
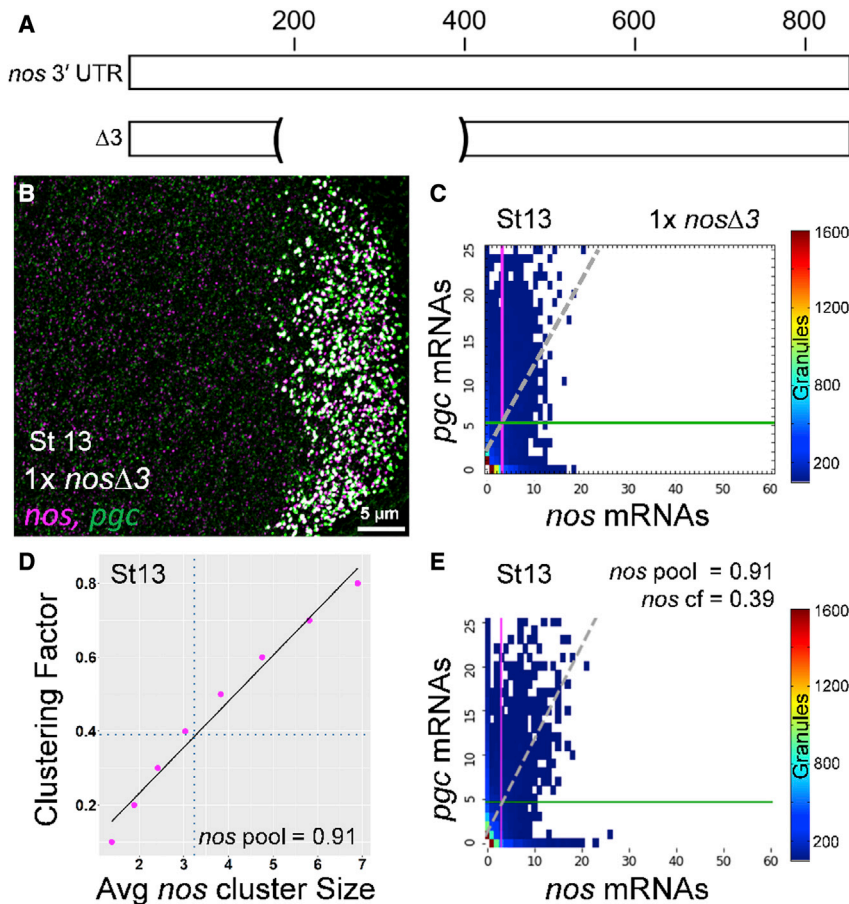
FIGURE 8 Quantification of homotypic clustering efficacy from a specific 3′ UTR. (*A*) Schematic of the wild-type (wt) *nos* 3′ UTR and the *nosΔ3* mutant. (*B*) Maximum projection of a stage 13 1× *nosΔ3* oocyte with *nos* (*magenta*) and *pgc* (*green*) detected by smFISH, posterior germ plasm is to the right. (*C*) Granule Census reveals that 1× *nosΔ3* produces an average *nos* cluster size of 3.24 transcripts (*vertical magenta line*, $n = 9$ germ plasms). The relationship between colocalized *nos* and *pgc* clusters is visualized with a line of best fit (*broken gray line*). (*D*) Standard curve produced by the computational model with the *nos* mRNA pool fixed at 0.91 while the clustering factor increases from 0.10 to 0.80 ($y = 0.1245x - 0.0173$, $R^2 = 0.98$). Fitting the average *nos* cluster size of 3.24 transcripts to the curve, we calculate that the *nos* +3 element has a clustering factor of 0.39 ($n = 10$ random simulations). (*E*) Granule Census produced using the computational model where the *nos* transcript pool was set to 0.91 and the clustering factor (cf) for *nos* was set to 0.39 captures expected changes in the germ plasm landscape including an average *nos* cluster size of 3.00 and a shift in the slope to 1.05.

average cluster sizes that were also smaller than 1× *nos* (4.07, $p < 0.005$) (Figs. S7 *E* and 8). The decrease in *nos* cluster size caused by deleting the *nos* +3 element is further apparent from the slope of the best fit line produced by co-localized *nos* and *pgc* clusters in stage 13 oocytes. In 1× *nosΔ3*, the average slope is $1.00 \pm 0.04$, which is a significant shift from the slopes for wild-type ($0.49 \pm 0.03$) and 1× *nos* ($0.75 \pm 0.07$) (16), $p < 0.002$ (Figs. S7 *E* and 8 *C*).

Next, we analyzed the impact that deleting the *nos* +3 element has on seeding by measuring colocalization rates with *pgc*. Colocalization of 1× *nosΔ3* mRNA with *pgc* was $42.85 \pm 1.88\%$, which is not significantly different from the colocalization rate that is expected due to a decrease in *nos* cluster size (16), as determined for 1× *nos* ($42.77 \pm 1.1\%$, $p = 0.58$). These data demonstrate that deleting the *nos* +3 element affects *nos* homotypic clustering without affecting the ability to seed and decouples expression level as the sole regulator of cluster sizes at the transcript level. Together, our experimental findings confirm the presence of clustering elements in the context of a genomic *nos* transgene and validate the inclusion in the model of a clustering factor parameter that can fine-tune homotypic clusters sizes in conjunction with expression levels.

Next, we aimed to quantify the clustering efficacy of the *nos* clustering element while factoring in 1× *nosΔ3* mRNA levels in stage 13 oocytes. First, we generated a standard curve ($y = 0.1245x - 0.0173$, $R^2 = 0.98$) between the clustering factor and average cluster size by modeling stage 13 oocytes with the gene expression reduced to 91% of wild-type (Figs. S3 *B* and 8 *D*). By fitting the average *nos* cluster size of 3.24, which was determined from biological 1× *nosΔ3* data to the modeled standard curve, we determined that 1× *nosΔ3* had a clustering factor of 0.39, which is a 47% reduction when compared with the wild-type *nos* clustering factor of 0.74. Furthermore, deleting the +3 element causes *nos* to have a clustering factor less than *pgc*'s clustering factor of 0.48. To test the combined effect of *nos* expression level and clustering factor, we set the model's *nos* mRNA pool parameter to 0.91 and the *nos* clustering factor to 0.39. The resulting modeled Granule Census was comparable with the expected biological Granule Census (Fig. 8 *E*). Specifically, the modeled data captured an average *nos* cluster size of $3.0 \pm 0.01$, which was not significantly different from 1× *nosΔ3* ($p = 0.16$) (Figs. S7 *E* and 8 *E*). In addition, the model produced slope values and co-localization rates that fell within the expected biological data range (Fig. S7 *E*). Previous studies have relied on

qualitative methods to gauge the contribution of specific *nos* 3′ UTRs to localization (49). Here, our combined modeling and biological results demonstrate a novel quantitative method, the clustering factor, that can measure the impact that a specific region of a transcript has on localization.

## DISCUSSION

Despite the essential roles that various classes of RNP granules play in the reproduction, stress responses, and nervous system, investigating how RNP granules form has been challenging for numerous class-specific reasons (50). In *Drosophila*, germ granules are densely packed into a confined space within the developing germ plasm. Thousands of these membrane-less electron dense germ granules form, each of which differs in the quantity and the types of mRNAs that they contain (15). With the sheer number of granules that exist and the mRNA heterogeneity observed among individual germ granules, several experimental approaches, such as smFISH, customized 3D computational analyses, and super-resolution microscopy have been combined to provide important insights into granule formation and composition (15,16). Although these strategies have provided important insights, most of what is currently known about the germ granule assembly process comes from data collected from only a handful of mRNA types and focuses mainly on the mRNA portion of germ granules (15–18,32). In other systems, such as *C. elegans*, in vitro studies of P granules have been instrumental in understanding the role of granule proteins in regulating the overall RNP architecture (51–53). In *Drosophila*, in vitro studies of germ granules have thus far been unsuccessful, limiting our understanding of the role that the germ granule protein ensemble plays in dictating germ granule mRNA composition and heterogeneity.

To help overcome experimental and technical limitations, we developed a computational model to explore known and discover unknown mechanisms that enable germ granules to attain their mRNA compositions, including the contribution of the protein ensemble. When the model was built using only known parameters, many granule characteristics and their dynamics did not agree with the biologically observed data for *nos* or *pgc*, including the Granule Census, colocalization rates, and correlation (Figs. 5, 6, and S3 *D*). Thus, we reasoned that other, as yet unknown, mechanisms must be at work during the granule formation process. To capture known germ granule mRNA characteristics, we created an enhanced computational model that incorporates additional mechanisms that help shape the mRNA content of germ granules, two of which function through the granule protein ensemble (Fig. 5).

Modeling germ granule assembly required the addition of a parameter that controls the efficacy of homotypic clustering in an mRNA type-specific manner. We name this quantifiable ability the "clustering factor." Without the addition of a clustering factor, homotypic cluster size regulation is limited to mRNA expression levels. In the case of *nos* and *pgc*, we found that expression levels cannot fully explain why *nos* forms larger clusters on average than *pgc* in the oocyte (Fig. S3 *A*). With the addition of a clustering factor, homotypic cluster sizes can be regulated even when mRNAs are expressed at similar levels. In the model, this single parameter ultimately controls the differences between *nos* and *pgc* cluster sizes. Biologically, this clustering factor could be achieved by a combination of factors, including features within 3′ UTRs. Indeed, by altering the clustering factor parameter, our model predicted the expected results from experiments, with *nos* RNA lacking a known clustering element (*nosΔ3*) (Fig. 8). Other factors that may contribute to an mRNA's clustering factor could be affinity for proteins, such as Osk, Vas, or Tud, which could affect mRNA dwell time. Consistent with this idea, *nos* and *pgc* mRNAs have been shown to bind to the Lotus domain of Osk (54). Regardless of the biological mechanisms controlling the clustering ability of mRNAs, we provide modeling and biological evidence supporting a clustering factor effect as a means for germ granules to fine-tune homotypic cluster size and generate varying cluster sizes between mRNA types in conjunction with mRNA expression levels (Fig. 8). Surprisingly, 3′ UTRs do not control the ability of two different RNA types to sort out from each other and self-recruit to form homotypic clusters (18). However, they do contain elements that influence cluster size (17). Combining results from those studies with our biological and in silico results, we propose a model where mRNA association with the germ granule protein ensemble requires sequences in the 3′ UTR that can generate a clustering factor to regulate mRNA abundance within a granule, while self-sorting is governed by unknown mechanisms that are independent from clustering factor 3′ UTR sequences. An additional 59 germ plasm mRNAs appear to localize to the germ plasm in a manner similar to *nos* and *pgc* (9), but only three of these have been analyzed for clustering elements (17). Thus, we expect that our model will have broader application in future germ granule mRNA studies by determining clustering factor values in biological experiments that investigate 3′ UTR clustering elements.

The biological mechanisms regulating how colocalization rate and correlation increase over time have not been explored. Using modeling, we were able to simulate the biologically observed increasing trends in colocalization rate and correlation between *nos* and *pgc* by assigning an age effect to individual granule protein ensembles, a parameter that dictates the probability for a granule ensemble to be seeded over time (Figs. S3 *E* and 5). These results support a mechanism mediated by the protein granule ensemble that impacts colocalization and correlations between different types of homotypic clusters. We reason that the age effect represents protein ensemble maturation that occurs downstream of *osk* translation and is created by the

time needed to accumulate additional Osk and other proteins within nascent germ granule protein ensembles. Specifically, measurements of Osk accumulation in granules over time showed that the average amount of Osk protein in a granule more than doubles from stage 10 to stage 13 (16), with up to five times more Osk present in granules in the early embryo (Fig. S2 *A*). We reason that, as more Osk is incorporated into a granule over time, the probability for a granule to capture and hold on to a seed transcript increases. Thus, Osk could be acting as a "sticker," which has been modeled in liquid-like droplets and acts to increase mRNA dwell time when abundant, allowing for more efficient seeding of a transcript in the condensate (55). The physical size growth of granules may also have a role in this effect given that germ granules can grow up to 1 μm with an average of ∼300 nm in diameter (56). Larger granules may simply have a larger surface area to capture a seed transcript.

Following *osk* localization and translation, additional proteins, such as Vas, Tud, and Aubergine (Aub), become incorporated into granule protein ensembles (11,12,14,57,58). Failure to properly accumulate such proteins dramatically decreases the posterior accumulation of *nos* and/or *pgc*, demonstrating that there are indeed essential maturation steps in the formation of a germ granule protein ensemble downstream of *osk* translation (21,48,59,60). In RNP condensates that form via liquid-liquid phase separation, the partitioning of RNA into the condensate scaffold is thought to occur through networks of thermodynamically favorable, multivalent RNA-RNA, RNA-protein, and protein-protein interactions. Such multivalent RNA-protein interactions may be enabled by multiple protein binding sites in the RNA as well as by proteins with intrinsically disordered regions (IDRs) that can bind multiple RNAs (25,61–63). Thus, the age effect may represent a biophysical mechanism that depends on the downstream accumulation of multivalent proteins and/or IDR containing proteins within nascent germ granule ensembles, which increases valency for seed transcripts as the granule matures. To account for any of several biological mechanisms or combination of mechanisms that could contribute to a granule's probability to be seeded over time, we simply implemented an increase in the probability to be seeded based on the developmental age of individual protein ensembles. Future studies aim to identify the underlying mechanisms that generate the age effect by exploring the roles that Vas, Tud, Aub, and IDR-containing proteins have in ensemble maturation, transcript seeding, and homotypic clustering.

Despite the addition of a clustering factor and a seeding age effect, the modeled Granule Census, colocalization, and correlation values were not comparable with the values obtained from biological measurements (Figs. S3 *D* and 5). To address these issues, we theorized that there may be seeding competition that arises within the granule protein ensemble and implemented a competition effect in the model that is applied to granule protein ensemble when it already contains mRNA that is different from the incoming seed mRNA. The resulting model generated a biologically comparable Granule Census and dynamic colocalizations rates and correlation values between *nos* and *pgc* that were analogous to biological data (Figs. 3 and 5). We reason that competition can be achieved biologically through physical space limitations or that granule proteins, such as Osk, are shared by multiple mRNA types for seeding and/or clustering, limiting their capacity to accommodate additional mRNA types when larger clusters are present. Since older granules tend to have larger mRNA clusters (15), we simplified the competition effect by designing a larger penalty based on a granule's age. By using the age of germ granule protein ensembles as the basis for increasing the competition penalty, we avoid introducing or suggesting any specific competition between a set of mRNA types. Rather, we suggest that seeding competition arises through limitations in the availability of shared granule proteins within the germ granule protein ensemble that are already involved in the clustering of other mRNA types currently present. In other types of RNP granules, such as stress granules and P-bodies, competing protein-RNA interaction networks control multiphase organization (62). Here, our modeling results demonstrate that a competition effect essentially limits the number of homotypic clusters that can form within individual *Drosophila* germ granules. Thus, competition may have a conserved role in shaping RNP compositions in different classes of biomolecular condensates with varying biophysical properties.

The competition effect in the model is triggered when a single mRNA type is present in a granule and, realistically, at least 59 additional mRNA types could potentially form clusters in germ granules (9). However, it is currently unknown how many different homotypic clusters a single granule can contain. Information regarding how many unique mRNAs can cluster within the same germ granule should be the focus of future biological experiments and incorporated into the model to better quantify and understand how competition emerges. Nevertheless, the application of age and competition effects ultimately assigns dynamic and varying probabilities to be seeded among thousands of individual germ granules protein ensembles (Video S3). This heterogeneous landscape was essential to simulate the formation of germ granules with key biological features, such as colocalization and correlation between *nos* and *pgc* (Fig. 5). Thus, our modeling demonstrates that the germ granule protein ensemble has a significant influence on germ granule mRNA composition, which is consistent with studies showing the importance of protein-based condensation mechanisms in the assembly of RNA-rich P granules in *C. elegans* (52,64).

Although our model accurately captures wild-type dynamics (Figs. 3–6), we recognize its limitations. We note

that the presented modeling data are less variable than biological data. In the model, data are collected from the same germ plasm developing over all stages, whereas biological data are collected from independent samples across all stages. Furthermore, the model collects cluster size data precisely at the end of each developmental stage, where biologically timing of stage is achieved qualitatively. These differences, in addition to natural biological variation, likely result in the model producing data that have less variation and slightly higher maximum cluster sizes in some stages. Modeling data also report more single molecules in the Granule Census than biological data (Figs. 3 and S5). We reason that this difference is likely due to limitations in smFISH detection and thresholding that are necessary to quantify germ plasm data (16). In the model, the same germ granule mRNA compositions can be accomplished in multiple ways by balancing mRNA expression levels with the clustering factor. However, incorrect balancing of mRNA expression levels with clustering factor results in an erroneous representation of the mRNA localization process as measured by the percentage of total transcripts that localize to germ granules (Fig. S8). Thus, the model's predictive power and solutions are limited to incorporating experimentally determined mRNA expression level data before predicting or adjusting for the mRNA's clustering factor value. Regardless of the limitations, the advantages to the model include the ability to independently control parameters for different mRNA types and the protein ensemble, allowing users to explore and identify mRNA and protein-specific effects on the overall composition of germ granules. In testing these controls, our model successfully predicted germ plasm defects caused by reduced Osk levels and expression levels of *nos* and/or *pgc* (Fig. 7). To separate the combined contributions of clustering factor and expression levels have on total mRNA localization, we demonstrate the model's broader capabilities by introducing a new metric, the clustering factor, that can be used to score the clustering ability of a transcript (Fig. 8).

Similarly to *Drosophila*, germ plasm formation in zebrafish is also initiated through a master protein organizer, resulting in a germ plasm containing homotypic RNPs of *nanos*3 and other RNAs (65,66); and, in *Xenopus* oocytes, enrichment of vegetally localizing mRNAs in the L-body requires specific RNA sequence features (28), resembling the requirement for *nos* 3′ UTR elements. Given the similar characteristics among different RNA granules, our model's principles and parameters, such as clustering factor, age effect, and competition effect, should be explored in other systems to identify conserved mechanisms that influence mRNA localization and the mRNA compositions of biomolecular condensates. For the first time, we present a mathematical representation of the *Drosophila* germ granule assembly process that confirms previously reported development mechanisms, offers new insight into how germ granules attain their mRNA compositions that can be explored in other systems, and provides a tool that can be integrated into biological experiments to support future studies.

## SUPPORTING MATERIAL

## AUTHOR CONTRIBUTIONS

Conceptualization, M.V., E.D.F., D.A.J., and M.G.N.; methodology, M.V., B.M.O., B.U., D.A.D., E.R.G., and M.G.N.; software, M.V., B.M.O., and M.G.N.; investigation, M.V., B.M.O., B.U., D.A.D., and M.G.N.; writing – original draft, M.V., E.D.F., D.A.J., and M.G.N.; writing – review & editing, E.R.G. and M.G.N.; supervision, E.D.F., D.A.J., E.R.G., and M.G.N.; funding acquisition, B.M.O., E.R.G., and M.G.N.

## ACKNOWLEDGMENTS

## REFERENCES

1. Sengupta, M. S., and P. R. Boag. 2012. Germ granules and the control of mRNA translation. *IUBMB Life.* 64:586–594. https://doi.org/10.1002/iub.1039.

2. Cinalli, R. M., P. Rangan, and R. Lehmann. 2008. Germ cells are forever. *Cell.* 132:559–562. https://doi.org/10.1016/j.cell.2008.02.003.

3. Ewen-Campen, B., E. E. Schwager, and C. G. Extavour. 2010. The molecular machinery of germ line specification. *Mol. Reprod. Dev.* 77:3–18. https://doi.org/10.1002/mrd.21091.

4. Voronina, E., G. Seydoux, …, I. Nagamori. 2011. RNA granules in germ cells. *Cold Spring Harb Perspect. Biol.* 3:a002774. https://doi.org/10.1101/cshperspect.a002774.

5. Strome, S., and R. Lehmann. 2007. Germ versus soma decisions: lessons from flies and worms. *Science.* 316:392–393. https://doi.org/10.1126/science.1140846.

6. Trcek, T., and R. Lehmann. 2019. Germ granules in *Drosophila*. *Traffic.* 20:650–660. https://doi.org/10.1111/tra.12674.

7. Gao, M., and A. L. Arkov. 2013. Next generation organelles: structure and role of germ granules in the germline. *Mol. Reprod. Dev.* 80:610–623. https://doi.org/10.1002/mrd.22115.

8. Seydoux, G., and R. E. Braun. 2006. Pathway to totipotency: lessons from germ cells. *Cell.* 127:891–904. https://doi.org/10.1016/j.cell.2006.11.016.

9. Rangan, P., M. DeGennaro, …, R. Lehmann. 2009. Temporal and spatial control of germ-plasm RNAs. *Curr. Biol.* 19:72–77. https://doi.org/10.1016/j.cub.2008.11.066.

10. Lecuyer, E., H. Yoshida, …, H. M. Krause. 2007. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell.* 131:174–187. https://doi.org/10.1016/j.cell.2007.08.003.

11. Mahowald, A. P. 2001. Assembly of the *Drosophila* germ plasm. *Int. Rev. Cytol.* 203:187–213.

12. Breitwieser, W., F. H. Markussen, H. Horstmann, and A. Ephrussi. 1996. Oskar protein interaction with Vasa represents an essential step in polar granule assembly. *Genes Dev.* 10:2179–2188.

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

14. Boswell, R. E., and A. P. Mahowald. 1985. tudor, a gene required for assembly of the germ plasm in *Drosophila* melanogaster. *Cell.* 43:97–104.

15. Little, S. C., K. S. Sinsimer, …, E. R. Gavis. 2015. Independent and coordinate trafficking of single *Drosophila* germ plasm mRNAs. *Nat. Cell Biol.* 17:558–568. https://doi.org/10.1038/ncb3143.

16. Niepielko, M. G., W. V. I. Eagle, and E. R. Gavis. 2018. Stochastic seeding coupled with mRNA self-recruitment generates heterogeneous *Drosophila* germ granules. *Curr. Biol.* 28:1872–1881.e3. https://doi.org/10.1016/j.cub.2018.04.037.

17. Eagle, W. V. I., D. K. Yeboah-Kordieh, M. G. Niepielko, and E. R. Gavis. 2018. Distinct cis-acting elements mediate targeting and clustering of *Drosophila* polar granule mRNAs. *Development.* 145. dev164657. https://doi.org/10.1242/dev.164657.

18. Trcek, T., T. E. Douglas, …, E. Rothenberg, R. Lehmann. 2020. Sequence-independent self-assembly of germ granule mRNAs into homotypic clusters. *Mol. Cell.* 78:941–950.e12. https://doi.org/10.1016/j.molcel.2020.05.008.

19. Eichler, C. E., A. C. Hakes, …, E. R. Gavis. 2020. Compartmentalized oskar degradation in the germ plasm safeguards germline development. *Elife.* 9:e49988. https://doi.org/10.7554/eLife.49988.

20. Spardling, A. C. 1993. Developmental Genetics of Oogenesis. Cold Spring Harbor Laboratory Press, pp. 1–70.

21. Thomson, T., and P. Lasko. 2004. *Drosophila* tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. *Genesis.* 40:164–170. https://doi.org/10.1002/gene.20079.

22. Gavis, E. R., and R. Lehmann. 1992. Localization of nanos RNA controls embryonic polarity. *Cell.* 71:301–313.

23. Sanchez-Burgos, I., J. R. Espinosa, …, R. Collepardo-Guevara. 2021. Valency and binding affinity variations can regulate the multilayered organization of protein condensates with many components. *Biomolecules.* 11:278. https://doi.org/10.3390/biom11020278.

24. Benayad, Z., S. von Bulow, …, G. Hummer. 2021. Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theor. Comput.* 17:525–537. https://doi.org/10.1021/acs.jctc.0c01064.

25. Choi, J. M., F. Dar, and R. V. Pappu. 2019. LASSI: a lattice model for simulating phase transitions of multivalent proteins. *PLoS Comput. Biol.* 15:e1007028. https://doi.org/10.1371/journal.pcbi.1007028.

26. Boeynaems, S., A. S. Holehouse, …, A. D. Gitler. 2019. Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U S A.* 116:7889–7898. https://doi.org/10.1073/pnas.1821038116.

27. Brangwynne, C. P., C. R. Eckmann, …, A. A. Hyman. 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science.* 324:1729–1732. https://doi.org/10.1126/science.1172046.

28. Neil, C. R., S. P. Jeschonek, …, K. L. Mowry. 2021. L-bodies are RNA-protein condensates driving RNA localization in Xenopus oocytes. *Mol. Biol. Cell.* 32:ar37. https://doi.org/10.1091/mbc.E21-03-0146-T.

29. Sinsimer, K. S., J. J. Lee, …, E. R. Gavis. 2013. Germ plasm anchoring is a dynamic state that requires persistent trafficking. *Cell Rep.* 5:1169–1177. https://doi.org/10.1016/j.celrep.2013.10.045.

30. Brangwynne, C. P. 2013. Phase transitions and size scaling of membrane-less organelles. *J. Cell Biol.* 203:875–881. https://doi.org/10.1083/jcb.201308087.

31. Brangwynne, C. P., T. J. Mitchison, and A. A. Hyman. 2011. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U S A.* 108:4334–4339. https://doi.org/10.1073/pnas.1017150108.

32. Trcek, T., M. Grosch, …, R. Lehmann. 2015. *Drosophila* germ granules are structured and contain homotypic mRNA clusters. *Nat. Commun.* 6:7962. https://doi.org/10.1038/ncomms8962.

33. Setnes, M., R. Babuska, and H. B. Verbruggen. 1998. Rule-based modeling: precision and transparency. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* 28:165–169. https://doi.org/10.1109/5326.661100.

34. Bergsten, S. E., and E. R. Gavis. 1999. Role for mRNA localization in translational activation but not spatial restriction of nanos RNA. *Development.* 126:659–669.

35. Zwillinger, D. 1997. Handbook of Differential Equations. Academic Press.

36. Hale, J. K. 1971. Analytic Theory of Differential Equations in Functional Differential Equations. Springer, pp. 9–22.

37. Abbaszadeh, E. K., and E. R. Gavis. 2016. Fixed and live visualization of RNAs in *Drosophila* oocytes and embryos. *Methods.* 98:34–41. https://doi.org/10.1016/j.ymeth.2016.01.018.

38. Little, S. C., G. Tkacik, …, T. Gregor. 2011. The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS Biol.* 9:e1000596. https://doi.org/10.1371/journal.pbio.1000596.

39. R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.

40. Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

41. RStudio Team. 2020. RStudio: Integrated Development for R. RStudio. PBC, Boston, MA http://www.rstudio.com/.

42. Pastore, M. 2018. Verlapping: a {R} package for estimating overlapping in empirical distributions. *J. Open Source Softw.* 3:1023.

43. Pastore, M., and A. Calcagni. 2019. Measuring distribution similarities between samples: a distribution-free overlapping index. *Front Psychol.* 10:1089. https://doi.org/10.3389/fpsyg.2019.01089.

44. Hanyu-Nakamura, K., H. Sonobe-Nojima, …, A. Nakamura. 2008. *Drosophila* Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature.* 451:730–733. https://doi.org/10.1038/nature06498.

45. Lehmann, R., and C. Nusslein-Volhard. 1991. The maternal gene nanos has a central role in posterior pattern formation of the *Drosophila* embryo. *Development.* 112:679–691.

46. Sarov, M., C. Barz, …, F. Schnorrer. 2016. A genome-wide resource for the analysis of protein localisation in *Drosophila*. *Elife.* 5:e12068. https://doi.org/10.7554/eLife.12068.

47. Bischof, J., M. Bjorklund, …, K. Basler. 2013. A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*. *Development.* 140:2434–2442. https://doi.org/10.1242/dev.088757.

48. Becalska, A. N., Y. R. Kim, …, E. R. Gavis. 2011. Aubergine is a component of a nanos mRNA localization complex. *Dev. Biol.* 349:46–52. https://doi.org/10.1016/j.ydbio.2010.10.002.

49. Gavis, E. R., D. Curtis, and R. Lehmann. 1996. Identification of cis-acting sequences that control nanos RNA localization. *Dev. Biol.* 176:36–50. https://doi.org/10.1006/dbio.1996.9996.

50. Zhang, X., and J. Mahamid. 2020. Addressing the challenge of in situ structural studies of RNP granules in light of emerging opportunities. *Curr. Opin. Struct. Biol.* 65:149–158. https://doi.org/10.1016/j.sbi.2020.06.012.

51. Seydoux, G. 2018. The P granules of *C. elegans*: a genetic model for the study of RNA-protein condensates. *J. Mol. Biol.* 430:4702–4710. https://doi.org/10.1016/j.jmb.2018.08.007.

52. Folkmann, A. W., A. Putnam, …, G. Seydoux. 2021. Regulation of biomolecular condensates by interfacial protein clusters. *Science.* 373:1218–1224. https://doi.org/10.1126/science.abg7071.

53. Putnam, A., M. Cassani, …, G. Seydoux. 2019. A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos. *Nat. Struct. Mol. Biol.* 26:220–226. https://doi.org/10.1038/s41594-019-0193-2.

54. Yang, N., Z. Yu, …, R. M. Xu. 2015. Structure of *Drosophila* Oskar reveals a novel RNA binding protein. *Proc. Natl. Acad. Sci. U S A.* 112:11541–11546. https://doi.org/10.1073/pnas.1515568112.

55. Ranganathan, S., and E. I. Shakhnovich. 2020. Dynamic metastable long-living droplets formed by sticker-spacer proteins. *Elife.* 9:e56159. https://doi.org/10.7554/eLife.56159.

56. Illmensee, K., and A. P. Mahowald. 1974. Transplantation of posterior polar plasm in *Drosophila*. Induction of germ cells at the anterior pole of the egg. *Proc. Natl. Acad. Sci. U S A.* 71:1016–1020. https://doi.org/10.1073/pnas.71.4.1016.

57. Kirino, Y., A. Vourekas, …, Z. Mourelatos. 2010. Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA.* 16:70–78. https://doi.org/10.1261/rna.1869710.

58. Vo, H. D. L., S. J. Wahiduzzaman, …, A. L. Arkov. 2019. Protein components of ribonucleoprotein granules from *Drosophila* germ cells oligomerize and show distinct spatial organization during germline development. *Sci. Rep.* 9:19190. https://doi.org/10.1038/s41598-019-55747-x.

59. Dufourt, J., G. Bontonou, …, M. Simonelig. 2017. piRNAs and Aubergine cooperate with Wispy poly(A) polymerase to stabilize mRNAs in the germ plasm. *Nat. Commun.* 8:1305. https://doi.org/10.1038/s41467-017-01431-5.

60. Styhler, S., A. Nakamura, …, P. Lasko. 1998. Vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development.* 125:1569–1578.

61. Shimobayashi, S. F., P. Ronceray, …, C. P. Brangwynne. 2021. Nucleation landscape of biomolecular condensates. *Nature.* 599:503–506. https://doi.org/10.1038/s41586-021-03905-5.

62. Sanders, D. W., N. Kedersha, …, C. P. Brangwynne. 2020. Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell.* 181:306–324.e8. https://doi.org/10.1016/j.cell.2020.03.050.

63. Riback, J. A., L. Zhu, …, C. P. Brangwynne. 2020. Composition-dependent thermodynamics of intracellular phase separation. *Nature.* 581:209–214. https://doi.org/10.1038/s41586-020-2256-2.

64. Schmidt, H., A. Putnam, …, G. Seydoux. 2021. Protein-based condensation mechanisms drive the assembly of RNA-rich P granules. *Elife.* 10:e63698. https://doi.org/10.7554/eLife.63698.

65. Bontems, F., A. Stein, …, R. Dosch. 2009. Bucky ball organizes germ plasm assembly in zebrafish. *Curr. Biol.* 19:414–422. https://doi.org/10.1016/j.cub.2009.01.038.

66. Eno, C., C. L. Hansen, and F. Pelegri. 2019. Aggregation, segregation, and dispersal of homotypic germ plasm RNPs in the early zebrafish embryo. *Dev. Dyn.* 248:306–318. https://doi.org/10.1002/dvdy.18.