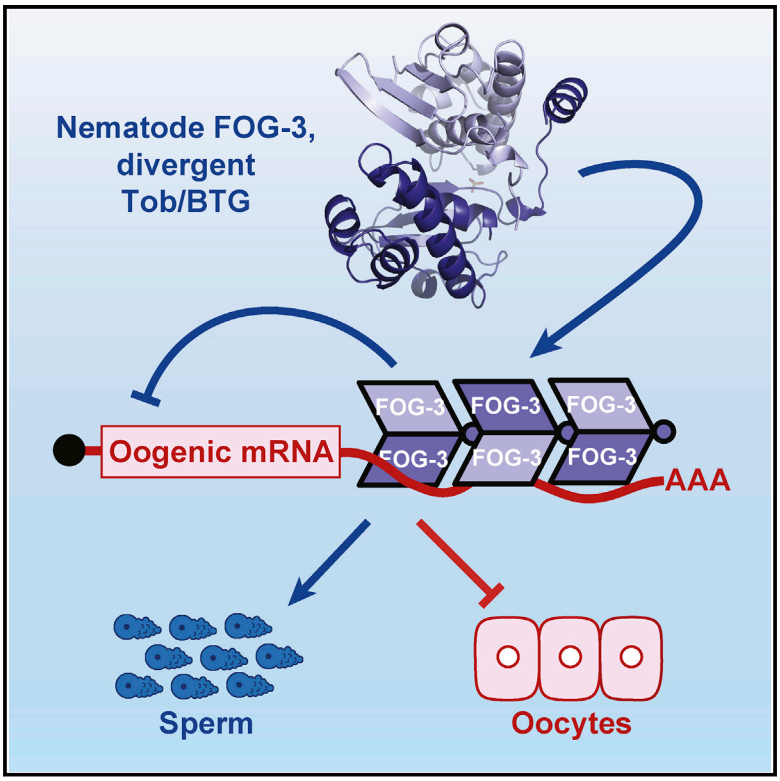
Report

Cell Reports

An RNA-Binding Multimer Specifies Nematode Sperm Fate

Graphical Abstract



Authors

Scott T. Aoki, Douglas F. Porter, Aman Prasad, Marvin Wickens, Craig A. Bingman, Judith Kimble

Correspondence

jekimble@wisc.edu

In Brief

The mechanism of the sperm or oocyte fate decision has been elusive. Aoki et al. report that nematode FOG-3, a Tob/BTG protein driving sperm fate, has evolved from monomeric to multimeric form with acquisition of a divergent Tob/BTG mechanism for mRNA repression.

Highlights

- FOG-3 crystal structure reveals sites of missense mutations
- FOG-3 assembles into dimers that can multimerize
- FOG-3 binds directly to mRNAs in the oogenic program
- FOG-3 recruited to a reporter mRNA represses its expression

Data and Software Availability

5TD6
GSE76521

Aoki et al., 2018, Cell Reports *23*, 3769–3775
June 26, 2018
https://doi.org/10.1016/j.celrep.2018.05.095

CellPress

Cell Reports
Report

An RNA-Binding Multimer Specifies Nematode Sperm Fate

Scott T. Aoki,[1] Douglas F. Porter,[1,3] Aman Prasad,[1] Marvin Wickens,[1] Craig A. Bingman,[1] and Judith Kimble[1,2,4,*]
[1]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
[2]Howard Hughes Medical Institute, University of Wisconsin-Madison, Madison, WI 53706, USA
[3]Present address: Department of Dermatology, Stanford University School of Medicine, Stanford, CA 94305, USA
[4]Lead Contact
*Correspondence: jekimble@wisc.edu
https://doi.org/10.1016/j.celrep.2018.05.095

SUMMARY

FOG-3 is a master regulator of sperm fate in *Caenorhabditis elegans* and homologous to Tob/BTG proteins, which in mammals are monomeric adaptors that recruit enzymes to RNA binding proteins. Here, we determine the FOG-3 crystal structure and *in vitro* demonstrate that FOG-3 forms dimers that can multimerize. The FOG-3 multimeric structure has a basic surface potential, suggestive of binding nucleic acid. Consistent with that prediction, FOG-3 binds directly to nearly 1,000 RNAs in nematode spermatogenic germ cells. Most binding is to the 3′ UTR, and most targets (94%) are oogenic mRNAs, even though assayed in spermatogenic cells. When tethered to a reporter mRNA, FOG-3 represses its expression. Together these findings elucidate the molecular mechanism of sperm fate specification and reveal the evolution of a protein from monomeric to multimeric form with acquisition of a distinct mode of mRNA repression.

INTRODUCTION

Gene discovery often paves the way to a molecular understanding of mysterious biological phenomena. But how do we learn molecular functions of newly identified genes? A common method takes advantage of amino acid comparisons to find homologs or at least protein domains that can provide clues. The major challenge is to test the predictions of those clues and tease out the function biochemically. Our focus is molecular regulation of the sperm or oocyte cell fate decision, which has remained elusive despite analyses over decades. Here, we report studies of a protein harboring a broadly conserved domain whose biochemical analysis reveals a striking case of protein evolution that could not have been predicted from sequence comparison.

Metazoan germ cells differentiate as either sperm or oocyte, depending on organismal sex. Molecular regulation of germline sexual fate relies in part on sex-determining signals from somatic tissues and in part on regulators responding to those signals in germ cells (Ellis and Schedl, 2007; Murray et al., 2010). In the nematode *Caenorhabditis elegans*, a network of intrinsic germline fate regulators drives expression of two key proteins that execute the sperm or oocyte fate decision (Ellis and Schedl, 2007; Kimble and Crittenden, 2007). Those two key terminal regulators are FOG-1 and FOG-3 (Figure 1A), named for their loss-of-function Fog (feminization of germline) phenotype (Barton and Kimble, 1990; Ellis and Kimble, 1995). FOG-1 is a *C. elegans* ortholog of the cytoplasmic polyadenylation element RNA binding protein (CPEB) (Jin et al., 2001; Luitjens et al., 2000), and FOG-3 harbors a protein domain that places it in the Tob/BTG family (Chen et al., 2000; Ellis and Kimble, 1995). Mammalian Tob/BTGs interact with RNA binding proteins and recruit enzymes to modify mRNAs and repress their expression (Hosoda et al., 2011; Ogami et al., 2014; Yu et al., 2016). Most relevant to this work, mammalian Tob/BTG binds CPEB proteins, and they recruit the CCR4-Not complex to shorten poly(A) tails and repress mRNA translation of CPEB target mRNAs (Hosoda et al., 2011). Similar to its mammalian counterparts, nematode FOG-3/Tob binds FOG-1/CPEB, and both proteins associate with a common set of 76 mRNAs, with 90% belonging to the oogenesis program (Noble et al., 2016). Because FOG-1 and FOG-3 specify the sperm fate, the inferred function was repression of oogenic RNAs.

We began this work with the idea that the FOG-3 mechanism of post-transcriptional control would echo that of its mammalian relatives, but we instead found a distinct mechanism that advances understanding of sperm fate specification and highlights the potential for undiscovered protein multimers in biology.

RESULTS

FOG-3 Is a Divergent Member of the Tob/BTG Family

Like canonical Tob/BTG family members, the FOG-3 primary sequence possesses a predicted N-terminal Tob/BTG domain and a disordered C-terminal region (Figures 1B and S1) (Chen et al., 2000). Comparison of FOG-3 amino acid sequences from several *Caenorhabditis* species revealed a further nematode-specific conservation that extends ~20 amino acids past the predicted Tob/BTG fold. We used recombinant *C. elegans* FOG-3 to identify a single domain (amino acids 1–137) that spans the canonical Tob/BTG fold plus this nematode-specific extension (see Supplemental Experimental Procedures and Figures S2A and S2B).

We pursued the FOG-3 crystal structure to gain insight into this putative Tob/BTG protein. FOG-3 crystals were obtained

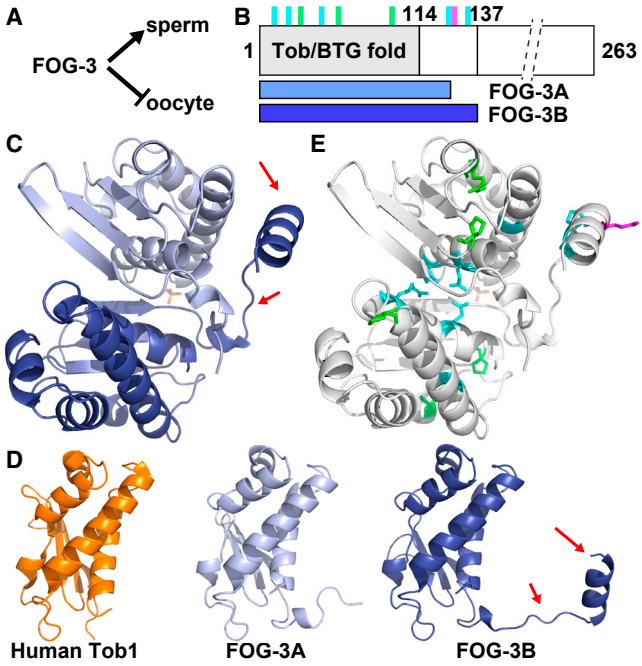


Figure 1. FOG-3 Is a Divergent Tob/BTG Protein

(A) FOG-3 is a terminal regulator of the sperm fate, and hence essential for sperm fate specification in both hermaphrodite larvae and males.

(B) Linear diagram of FOG-3 protein. Predicted Tob/BTG fold is N-terminal (gray). C terminus is abbreviated (dashed lines). Vertical lines mark missense mutation sites (also see Figure S1): green, conserved in all orthologs; cyan, conserved in nematode orthologs; magenta, missense mutation generated in this study. Horizontal bars below show extents of subunits in crystal dimer, termed FOG-3A (light blue) and FOG-3B (dark blue).

(C) Asymmetric unit of the FOG-3 crystal structure (PDB: 5TD6). FOG-3A (chain A, light blue) and FOG-3B (chain B, dark blue). Arrows highlight the nematode-specific linker-helix extension.

(D) FOG-3 and human Tob/BTG structures. RMSD of human Tob1 (orange, PDB: 2Z15) compared with FOG-3A and FOG-3B was 1.062 Å and 1.086 Å, respectively. Arrows highlight the linker-helix extension.

(E) Location of FOG-3 missense mutants in the crystal structure. Color scheme matches that in (B).

with C-terminally truncated recombinant protein that had changes to two non-conserved amino acids (1–137; H47N C117A). These crystals provided a full dataset to 2.03 Å (Research Collaboratory for Structural Bioinformatics [RCSB] PDB: 5TD6; Table S1). Phase information was acquired using a human Tob structure (see Supplemental Experimental Procedures). Each asymmetric unit contained two copies of FOG-3; we dub chain A as FOG-3A and chain B as FOG-3B (Figures 1B and 1C). For FOG-3A, we could model residues 1–123, and for FOG-3B we could model nearly the entire peptide chain (1–136) (Figures 1B and 1C). The structural alignment was excellent between FOG-3A, FOG-3B, and a previously determined human Tob1 structure (root-mean-square deviation [RMSD] 1.062–1.084 Å; Figure 1D) (Horiuchi et al., 2009), confirming the predicted Tob/BTG fold in FOG-3. However, unlike other Tob/BTG structures, the FOG-3 structure included a linker-helix extension past the classic Tob/BTG fold (Figures 1C and 1D).

Details of the FOG-3 crystal structure supported the idea that the two FOG-3 subunits in the asymmetric unit (ASU) represent a bona fide dimer. The buried surface area between the two FOG-3 subunits was large (1,034.2 Å), favorable for assembly ($\Delta G = -7.2$ kcal/mol), and the interface intricate (Figures S2D and S2E). The linker-helix extension of FOG-3B folded around FOG-3A (Figure 1C), making several hydrogen bonds (Figures S2D and S2E). The Tob/BTG folds of FOG-3A and FOG-3B also contacted each other at their N-terminal helices, with hydrogen bonds and arginine planar stacking between conserved residues (Figures S2G and S2H). To ask whether this potential FOG-3 dimer may have biological significance, we analyzed the sites of the missense mutations (Chen et al., 2000). All mutations abolish sperm fate specification and hence render the protein non-functional (Ellis and Kimble, 1995). The FOG-3 structure included all eight missense sites (Figures 1B, 1E, S1, and S2F). Three mutations (P21L, R56Q, P94S) alter residues conserved across Tob/BTG folds, and five others change residues conserved only in FOG-3 and nematode paralogs (Figure S1). The three Tob/BTG fold mutations include two prolines located between helices and an arginine making a hydrogen bond characteristic of Tob/BTG folds (Figures S2G and S2I) (Horiuchi et al., 2009; Yang et al., 2008). Because of their locations, contacts, and conservation, we speculate that these residues facilitate protein folding. Five other mutations change residues conserved only in nematodes. These include three (E7K, R14K, G33K) in the Tob/BTG fold (Figures S2G and S2H) and two (P125L, A132T) in the nematode-specific linker-helix extension (Figures S2G and S2J). At the potential dimer interface of the Tob/BTG folds, two missense residues (E7K, R14K) map to the interface itself (Figures S2G and S2H) and the third (G33K) maps to a central helix, where a bulky lysine residue could disrupt dimerization via steric hindrance (Figures S2G and S2I). The two mutations outside the Tob fold (P125L, A132T) map to the internal face of the link-helix extension in FOG-3B (Figures S2G and S2J). Thus, the FOG-3 crystal structure and sites of *fog-3* mutations provide evidence that FOG-3 has a Tob/BTG fold but is divergent with potential to dimerize.

FOG-3 Dimerization and Higher-Order Assembly

The mutations in the linker-helix suggest that this extension is crucial for FOG-3 function. The asymmetrical interaction between the linker-helix of one subunit and the Tob/BTG fold of the other seemed an unusual strategy for dimerization (Figure 1C). We wondered whether the linker-helix might also mediate an interaction between dimers. To explore this idea, we extended the crystal symmetry to visualize FOG-3 dimer-dimer interactions in the structure and found the linker-helix extension tucked neatly into a cleft of the adjacent FOG-3 dimer (Figure 2A). Each dimer was rotated 180° relative to its neighbor in a continuous pattern to form a polymeric-like assembly within the crystal (Figure 2B). Moreover, features of the structure suggested that the dimer-dimer interface may be authentic: its surface area is 1,112.5 Å (Figure S3A), a value similar to that between subunits in the dimer, and the predicted interface included 10 hydrogen bonds and four salt bridges (Figures S3A–S3C), with many residues conserved among FOG-3 orthologs (Figure S1).

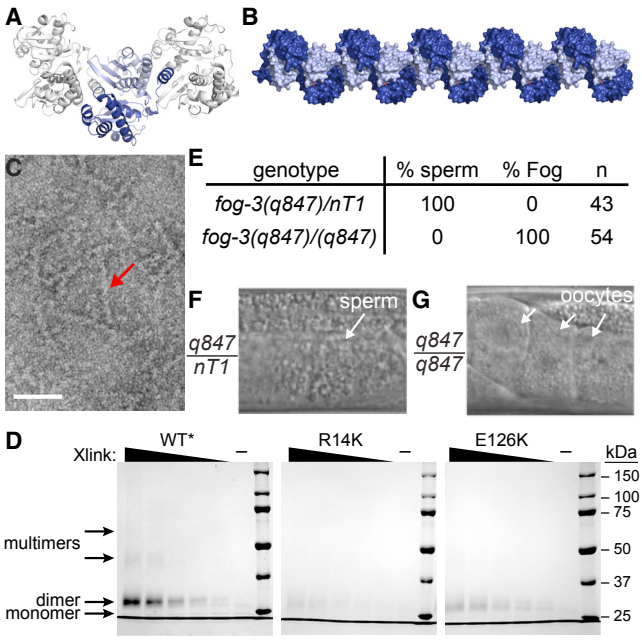


Figure 2. FOG-3 Forms Higher-Order, Multi-dimer Assemblies

(A) Crystal packing of the FOG-3 dimer. FOG-3A and FOG-3B are represented in light and dark blue, respectively.

(B) Model for multimerization of FOG-3 dimers, as observed in the crystal. Image was generated by extending the crystal symmetry. Subunits are colored as in (A).

(C) Negative-stain electron microscopy of recombinant FOG-3 1–137. Note the presence of long rods (arrow). Scale bar, 50 nm.

(D) Coomassie-stained gel of modified FOG-3 recombinant protein incubated with increasing amounts of BS3 crosslinker. "–" represents no BS3 included. Different exposure shown in Figure S3I.

(E–G) Mutation of a key residue, E126K, transforms the germline from spermatogenic to oogenic. The E126K allele, *fog-3(q847)*, was maintained over a balancer (*nT1*). (E) Heterozygous or homozygous adult males were analyzed for sperm or oocytes. (F and G) Representative DIC images of adult males, heterozygous (F) or homozygous (G) for the E126K mutation.

See Figures S3L–S3N for similar results in hermaphrodites.

We sought to test FOG-3 dimerization and higher-order multimerization *in vitro*. On a sizing column, recombinant FOG-3 eluted distinctly at high versus low concentrations (Figure S2A), suggesting dimerization at higher protein concentration. The column gave no hint of larger multimers (Figure S2A), but they might be unstable. We turned next to negative-stain electron microscopy (EM), which revealed rods when recombinant FOG-3 was assayed at high concentration (Figure 2C). The FOG-3 rods were consistent with multimerization, but their formation was sporadic, making EM an unreliable assay to compare wild-type versus mutant protein. We finally turned to biochemistry and used the crystal structure to insert lysine substitutions at non-conserved residues to facilitate chemical crosslinking (Figure S3D–S3F). The lysine variant could be purified (Figures S3G and S3H) and robustly crosslinked as a dimer when incubated (Figures 2D and S3I). At a higher crosslinker concentration, larger species formed, which we attribute to multi-dimer assemblies (Figures 2D and S3I). We next made a lysine variant that also harbored a missense mutation (R14K) (Chen et al., 2000), which is predicted from the structure to impede dimerization (Figures S2F–S2H). Indeed, dimerization was greatly reduced with this R14K mutant and multimerization was abolished (Figures 2D and S3I). Therefore, R14K disrupts dimerization, as predicted, and dimerization is required for higher-ordered assembly.

Classical genetics failed to generate a mutation that would disrupt the dimer-dimer interaction. To make our own, we identified a glutamate (E126) on the solvent-exposed surface of the linker-helix that contacts the neighboring dimer (Figures S3J and S3K). We reasoned that changing E126 to a positively charged amino acid (E126K) should disrupt the dimer-dimer interface specifically. Indeed, an E126K lysine variant still formed dimers, but not higher-order, multimeric species (Figures 2D and S3I). To ask whether dimer-dimer contact was critical for sperm fate specification *in vivo*, we introduced E126K in the endogenous *fog-3* gene (see Supplemental Experimental Procedures). Animals homozygous for either of two independently generated E126K alleles failed to make sperm and instead had a fully penetrant Fog phenotype in both males (Figures 2E–2G) and hermaphrodites (Figures S3L–S3N). The E126K mutation may affect FOG-3 in ways other than multimerization, but the simplest explanation is that FOG-3 functions *in vivo* as a multimer of dimers to promote sperm fate.

FOG-3 Binds RNA Directly to 3′ UTRs of Oogenic-Associated Transcripts

The electrostatic surface potential of the FOG-3 multimer is highly basic (Figures S4A and S4B) and prompted us to ask whether FOG-3 binds directly to RNA. In nematode spermatogenic germ cells, FOG-3 immunoprecipitated radiolabeled RNA after UV crosslinking (Figures S4C–S4F), a treatment creating covalent bonds between protein and RNA (Huppertz et al., 2014). Using *in vivo* crosslinking and immunoprecipitation (iCLIP) (Huppertz et al., 2014), we identified RNAs crosslinked to FOG-3 (Figures S4G–S4I; Tables S2 and S3) and sites of FOG-3 binding within those RNAs. After normalization to a negative control (see Supplemental Experimental Procedures), FOG-3 enriched for 955 mRNA targets and 38 non-coding RNAs (Figure 3A). Remarkably, ~94% of FOG-3-bound mRNAs belonged to the oogenesis program (Figures 3B and S4J), despite immunoprecipitation from sperm-fated germ cells. The FOG-3 targets identified by iCLIP overlapped with those identified by microarray (Figures S4J and S4K; $p < 1 \times 10^{-208}$), but iCLIP significantly increased both number of targets as well as their enrichment for oogenic RNAs (Figures S4J and S4K). Because iCLIP is more stringent than microarray methods (Wang et al., 2009), we suggest that iCLIP improved the signal-to-noise ratio of RNAs immunoprecipitating with FOG-3. We conclude that FOG-3 binds directly to RNAs that belong largely to the oogenesis program.

The vast majority of FOG-3 binding sites mapped to 3′ UTRs (Figures 3C and S4L–S4N), implying that FOG-3 binding is largely restricted. A multimer is expected to leave an extensive footprint. Consistent with this idea, 624 of 955 protein-coding genes (65.3%) had two or more sequence peaks in their 3′ UTRs (Figures 3D, 3E, and S4O). This pattern is reminiscent of multi-site RNA binding proteins, like HuR (Lebedeva et al., 2011). Gaps between peaks might signify authentic binding absences from breaks in FOG-3 multimerization or binding by

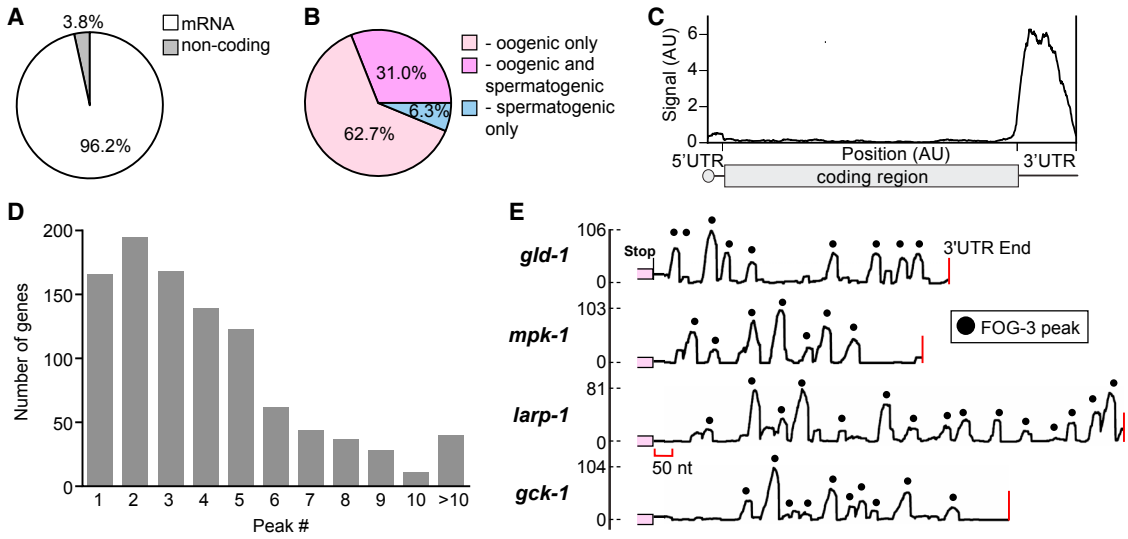


Figure 3. FOG-3 in Spermatogenic Germ Cells Binds Directly to 3′ UTRs of Oogenic mRNAs

(A) FOG-3 iCLIP enriches for mRNAs.

(B) Most FOG-3-bound mRNAs belong to the oogenesis program, which includes RNA expressed only in oogenic germlines (light pink) and in both oogenic and spermatogenic germlines (dark pink), as categorized previously (Noble et al., 2016).

(C) Distribution of FOG-3 iCLIP sequence reads within mRNAs. Transcript lengths are normalized so that 5′ UTRs, coding sequences (CDS), and 3′ UTRs (50, 1,000, and 200 nt, respectively) are reported as arbitrary units (AU).

(D) Many FOG-3 targets possess multiple binding peaks in their 3′ UTRs, as shown in (E).

(E) Examples of FOG-3 binding peaks across 3′ UTRs. x axis, 3′ UTR with coding region in pink and 3′ end marked by red line; y axis, number of mapped reads. Peaks are marked by black dots; their heights correspond to number of mapped reads.

independent dimers. These gaps may also be sites of preferred enzymatic digestion during iCLIP (see Supplemental Experimental Procedures). Regardless, we conclude that FOG-3 binds the 3′ UTRs of its target mRNAs and that most of its targets belong to the oogenic program, despite being immunoprecipitated from spermatogenic cells.

Tethered FOG-3 Represses Expression of a Reporter mRNA

Genetically, FOG-3 is a master regulator of sperm fate (Ellis and Kimble, 1995) and FOG-3 binds many oogenic mRNAs directly in spermatogenic germ cells (this work). The simplest model is that FOG-3 specifies the sperm fate by repression of oogenic mRNAs. One cannot remove FOG-3 to test this idea because germ cells make oocytes rather than sperm without FOG-3. As an alternative approach, we turned to a protein-mRNA tethering reporter assay used in other systems (Coller and Wickens, 2002) and *C. elegans* (Wedeles et al., 2013), which takes advantage of λN22 peptide binding to boxB RNA hairpins (Baron-Benhamou et al., 2004) (Figure 4A). Endogenous FOG-3 was engineered to include a C-terminal 3xFLAG epitope tag, with or without λN22 (see Supplemental Experimental Procedures). The reporter mRNA expressed GFP-tagged histone and carried three boxB hairpins in its 3′ UTR (Figure 4A). Both engineered FOG-3 proteins promoted the sperm fate and were expressed as expected (Figures 4B and 4F). Thus, our modifications did not affect FOG-3 function or expression.

We compared fluorescence in germ cells expressing engineered FOG-3 and the reporter. GFP was easily detected throughout the male germline when FOG-3::3xFLAG was not fused to λN22 (Figures 4B–4E). However, when FOG-3::3xFLAG was fused to λN22, GFP fluorescence decreased in the region of FOG-3 expression (Figures 4F–4I). Therefore, FOG-3 represses GFP expression when tethered to a reporter mRNA.

DISCUSSION

Our findings support a model for the molecular mechanism of sperm fate regulation (Figure 4J). We propose that the functional form of FOG-3 is multimeric, that FOG-3 multimers bind directly to the 3′ UTRs of target mRNAs, and that FOG-3 is a broad-spectrum repressor of the oogenesis program. Because FOG-3 and its orthologs specify sperm fate in both *C. elegans* sexes and related dioecious species (Chen et al., 2001), this mechanism is likely fundamental to nematode sperm fate specification. Protein expression appears sufficient for such function; FOG-3 is readily observed by immunoblot and its staining is granular (Noble et al., 2016), implying that FOG-3 may be further concentrated within the cytoplasm for multimerization. Our results challenge the idea that FOG-3 regulates RNAs via the mechanism elucidated for mammalian Tob/BTG proteins, which function as monomeric adapters. Although FOG-3 might also function as a monomeric adaptor, such a canonical Tob/BTG mechanism is unlikely for its role in sperm fate. Terminal regulators of mammalian germ cell fate are not yet known, so we cannot exclude the possibility that the biological role of Tob/BTG proteins may have been conserved in germ cells.

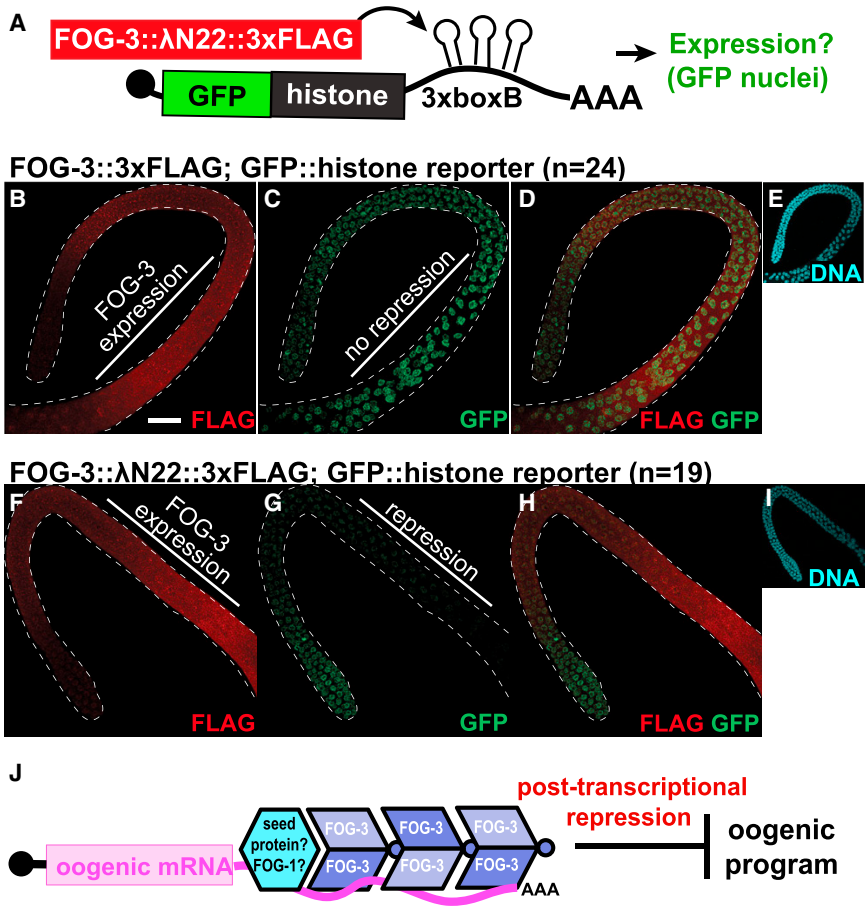


Figure 4. FOG-3 Represses mRNA Reporter Expression When Tethered in Nematodes

(A) Summary of the protein-mRNA tethering assay. With λN22, FOG-3 can bind and regulate the protein expression of a GFP::histone reporter transcript containing 3xboxB hairpins in its 3′ UTR. (B–I) Maximum intensity projections from confocal images of representative male adult germlines expressing both modified FOG-3 and the GFP reporter. FOG-3::3xFLAG (B–E) or FOG-3::λN22::3xFLAG (F–I) with the GFP::histone reporter were imaged by fluorescent confocal microscopy for (B and F) FLAG, (C and G) GFP, (D and H) FLAG and GFP overlay, and (E and I) DNA (DAPI). DNA images reduced in size 2.5-fold. Dashed white line outlines male germlines. Scale bar, 20 μm.

(J) Model of the FOG-3 molecular mechanism. Multimers of FOG-3 dimers bind 3′ UTRs and promote the sperm fate by repressing mRNAs in the oogenic program. FOG-3 may find its targets by interacting with a distinct sequence-specific seed protein, one candidate being FOG-1 (see Discussion).

The restricted footprint of FOG-3 to 3′ UTRs raises questions about the regulation of FOG-3 binding. FOG-3 may itself provide RNA binding specificity, although no sequence element could be detected beneath its iCLIP peaks. More likely, another sequence-specific RNA-binding protein seeds FOG-3 assembly (Figure 4J). FOG-1/CPEB is a strong candidate for such a seed protein, because it is also drives sperm fate specification (Barton and Kimble, 1990), binds to FOG-3, and associates with oogenic mRNAs (Noble et al., 2016). The mechanism of repression remains a challenge for the future. The FOG-3 multimer might recruit effector complexes to modify the mRNA, similar to mammalian Tob/BTG. If true, details are likely different because residues corresponding to the deadenylase binding interface of human Tob (Horiuchi et al., 2009) are not conserved in FOG-3 and no missense mutations mapped to that potential interface. Other possibilities include competition with an activator or moving mRNAs to sites of repression within the cell.

Fundamental aspects of nematode FOG-3 function diverge from mammalian Tob/BTG. Although mRNA repression is shared, FOG-3 functions as a multimer and binds RNA directly, whereas mammalian Tob/BTG proteins function as monomeric adaptors that link sequence-specific RNA binding proteins to enzymes (Hosoda et al., 2011; Ogami et al., 2014; Yu et al., 2016). Consistent with this divergence, FOG-3 possesses a nematode-specific linker-helix extension that mediates dimer-dimer interactions, and those dimer-dimer interactions are integral to sperm fate specification. This strategy of FOG-3 assembly is reminiscent of certain viral RNA binding proteins that package RNA viral genomes using a core domain plus a C-terminal linker-helix or linker-β sheet extension to drive multimerization (Harrison, 2017). In an analogous but distinct strategy, yeast RNA-binding protein Rim4 assembles into amyloid-like fibers to repress translation of mRNAs required for gametogenesis (Berchowitz et al., 2015). Collaborative assembly with RNA is common among viral assembly proteins and other nucleic acid binding polymers (Ghosal and Löwe, 2015). FOG-3 may similarly require an mRNA scaffold for higher-order assembly, given that FOG-3 dimers could not multimerize into higher-order assemblies at low protein concentration.

Our work on FOG-3 highlights the concept that evolution can usurp a well-conserved domain to form multimers with only modest changes to its primary sequence and structural fold. Multimerization appears to be a unique feature of FOG-3-related Tob/BTG proteins. The critical residues at the dimer interface are conserved in nematodes, but not in mammalian Tob/BTG proteins, and mammalian homologs show no evidence of multimerization (Horiuchi et al., 2009; Yang et al., 2008). Because only a few Tob/BTG proteins have been characterized biochemically, FOG-3 might yet have a vertebrate counterpart. However, we favor instead the idea that an existing protein fold was adapted during evolution to transform a protein that acts as a monomer into a multimeric repressor with RNA binding properties. The transformation from monomer to multimer required evolution of inter-subunit interacting

surfaces to form dimers and evolution of inter-dimer interacting surfaces to permit multimerization. In the case of FOG-3, the inter-dimer interface was created by adding a linker-helix extension that fits into the cleft of its neighbor, adding a binding surface to one side of the dimer. This extension provides directionality for multimer assembly. Therefore, FOG-3 provides an elegant example of protein evolution, in which a broadly conserved protein domain is redeployed to acquire a distinct mode of mRNA repression.

EXPERIMENTAL PROCEDURES

Crystallization and Structure Determination

Crystallization conditions were screened with sitting drop trays set up using the Mosquito (TTP Labtech, Cambridge, MA, USA). We obtained crystals using recombinant FOG-3 (1–137 H48N C117A) with an intact histidine tag and incubating our trays at 4°C. Coordinates, reflection data, and further experimental details are available at RCSB (PDB: 5TD6) and in the Supplemental Experimental Procedures.

Negative-Stain Electron Microscopy

EM was performed with recombinant FOG-3 (1–137 H48N C117A) as previously described (Bozzola and Russell, 1999). See Supplemental Experimental Procedures.

Molecular Genetics

Worm Maintenance

C. elegans were maintained as described previously (Brenner, 1974). Strains used were JA1515: weSi2 II; unc-119 III; JK2739: hT2[qIs48](I;III)/lin-6(e1466) dpy-5(e61)I; JK4871: fog-3(q520) I; qSi41[fog-3::3xFLAG] II; JK5437: fog-3(q847) I/hT2[qIs48](I;III); JK5439: fog-3(q849) I/hT2[qIs48](I;III); JK5921: qSi375 [mex-5 promoter::EGFP::linker::his-58::3xboxB::tbb-2 3′ UTR] II; JK5942: fog-3(q873[fog-3::3xFLAG]) I; qSi375 II; JK5943: fog-3(q874[fog-3::λN22::3xFLAG]) I; qSi375 II; and N2 Bristol. Strains are available at the *Caenorhabditis* Genetics Center (https://cbs.umn.edu/cgc/home) or upon request. See Supplemental Experimental Procedures for details on generation of the *fog-3* alleles and *in vivo* fluorescent reporter.

Fertility and Fog Phenotype

Heterozygous and homozygous *fog-3(q847)*, or *fog-3(q849)*, were singled onto plates as L4 larvae. After 3 and 4 days, worms were scored for the presence of L1 larvae and Fog phenotype.

Imaging

Live worms were imaged by differential interference contrast (DIC) microscopy with 0.1 mM levamisole (Sigma, St. Louis, MO, USA). For fluorescent imaging, germlines were extruded, fixed, permeabilized with Triton-X (0.5%), and stained as previously described (Crittenden et al., 2017). See Supplemental Experimental Procedures for further details.

iCLIP

In vivo crosslinking and immunoprecipitation (iCLIP) was carried out essentially as described (Huppertz et al., 2014), with modifications to worm growth, crosslinking, lysis, RNase digestion, and data analysis described in the Supplemental Experimental Procedures. All reads within a gene had their position randomized to empirically determine a cluster p value. A Benjamini-Hochberg (BH) correction was applied (1% false discovery rate [FDR]). Only overlapping clusters called independently as significant in at least two replicates were retained. Fisher's exact test was used to compare our results with previous FOG-3 RIP results (Noble et al., 2016).

DATA AND SOFTWARE AVAILABILITY

The accession number for the structure reported in this paper is PDB: 5TD6. The accession number for the raw sequence files of all replicates reported in this paper is GEO: GSE76521 (https://www.ncbi.nlm.nih.gov/geo/).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and three tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.05.095.

ACKNOWLEDGMENTS

The authors thank J. Claycomb and J. Ahringer for worm strains; R. Massey for electron microscopy image acquisition; M. Cox for equipment; A. Helsley-Marchbanks for help preparing the manuscript; L. Vanderploeg for help with the figures; and members of the Kimble and Wickens labs for helpful discussions. Use of the Advanced Photon Source was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. GM/CA@APS has been funded in whole or in part with federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289-01A1). S.T.A. was supported by NIH grants F32HD071692 and K99HD081208. C.A.B. was supported by NIH grants GM094584, GM094622, and GM098248. M.W. was supported by NIH grant GM50942. J.K. is an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

S.T.A. designed and performed experiments, analyzed the data, and wrote the paper; D.F.P. and A.P. designed and performed experiments and analyzed the data; C.A.B. acquired and analyzed data; M.W. analyzed the data and drafted the manuscript; and J.K. analyzed the data and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 6, 2018
Revised: April 26, 2018
Accepted: May 30, 2018
Published: June 26, 2018

REFERENCES

Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E., and Hentze, M.W. (2004). Using the lambdaN peptide to tether proteins to RNAs. Methods Mol. Biol. *257*, 135–154.

Barton, M.K., and Kimble, J. (1990). *fog-1*, a regulatory gene required for specification of spermatogenesis in the germ line of *Caenorhabditis elegans*. Genetics *125*, 29–39.

Berchowitz, L.E., Kabachinski, G., Walker, M.R., Carlile, T.M., Gilbert, W.V., Schwartz, T.U., and Amon, A. (2015). Regulated formation of an amyloid-like translational repressor governs gametogenesis. Cell *163*, 406–418.

Bozzola, J.J., and Russell, L.D. (1999). Electron Microscopy: Principles and Techniques for Biologists (Jones and Bartlett Publishers).

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

Chen, P.-J., Singal, A., Kimble, J., and Ellis, R.E. (2000). A novel member of the tob family of proteins controls sexual fate in *Caenorhabditis elegans* germ cells. Dev. Biol. *217*, 77–90.

Chen, P.-J., Cho, S., Jin, S.-W., and Ellis, R.E. (2001). Specification of germ cell fates by FOG-3 has been conserved during nematode evolution. Genetics *158*, 1513–1525.

Coller, J., and Wickens, M. (2002). Tethered function assays using 3′ untranslated regions. Methods *26*, 142–150.

Crittenden, S.L., Seidel, H.S., and Kimble, J. (2017). Analysis of the C. elegans germline stem cell pool. Methods Mol. Biol. *1463*, 1–33.

Ellis, R.E., and Kimble, J. (1995). The *fog-3* gene and regulation of cell fate in the germ line of *Caenorhabditis elegans*. Genetics *139*, 561–577.

Ellis, R., and Schedl, T. (2007). Sex determination in the germ line. WormBook, The C. elegans Research Community. http://www.wormbook.org/chapters/www_sexgermline.2/sexgermline.html.

Ghosal, D., and Löwe, J. (2015). Collaborative protein filaments. EMBO J. *34*, 2312–2320.

Harrison, S.C. (2017). Protein tentacles. J. Struct. Biol. *200*, 244–247.

Horiuchi, M., Takeuchi, K., Noda, N., Muroya, N., Suzuki, T., Nakamura, T., Kawamura-Tsuzuku, J., Takahasi, K., Yamamoto, T., and Inagaki, F. (2009). Structural basis for the antiproliferative activity of the Tob-hCaf1 complex. J. Biol. Chem. *284*, 13244–13255.

Hosoda, N., Funakoshi, Y., Hirasawa, M., Yamagishi, R., Asano, Y., Miyagawa, R., Ogami, K., Tsujimoto, M., and Hoshino, S. (2011). Anti-proliferative protein Tob negatively regulates CPEB3 target by recruiting Caf1 deadenylase. EMBO J. *30*, 1311–1323.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274–287.

Jin, S.-W., Kimble, J., and Ellis, R.E. (2001). Regulation of cell fate in *Caenorhabditis elegans* by a novel cytoplasmic polyadenylation element binding protein. Dev. Biol. *229*, 537–553.

Kimble, J., and Crittenden, S.L. (2007). Controls of germline stem cells, entry into meiosis, and the sperm/oocyte decision in *Caenorhabditis elegans*. Annu. Rev. Cell Dev. Biol. *23*, 405–433.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol. Cell *43*, 340–352.

Luitjens, C., Gallegos, M., Kraemer, B., Kimble, J., and Wickens, M. (2000). CPEB proteins control two key steps in spermatogenesis in *C. elegans*. Genes Dev. *14*, 2596–2609.

Murray, S.M., Yang, S.Y., and Van Doren, M. (2010). Germ cell sex determination: a collaboration between soma and germline. Curr. Opin. Cell Biol. *22*, 722–729.

Noble, D.C., Aoki, S.T., Ortiz, M.A., Kim, K.W., Verheyden, J.M., and Kimble, J. (2016). Genomic analyses of sperm fate regulator targets reveal a common set of oogenic mRNAs in *Caenorhabditis elegans*. Genetics *202*, 221–234.

Ogami, K., Hosoda, N., Funakoshi, Y., and Hoshino, S. (2014). Antiproliferative protein Tob directly regulates c-myc proto-oncogene expression through cytoplasmic polyadenylation element-binding protein CPEB. Oncogene *33*, 55–64.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Wedeles, C.J., Wu, M.Z., and Claycomb, J.M. (2013). Protection of germline gene expression by the *C. elegans* Argonaute CSR-1. Dev. Cell *27*, 664–671.

Yang, X., Morita, M., Wang, H., Suzuki, T., Yang, W., Luo, Y., Zhao, C., Yu, Y., Bartlam, M., Yamamoto, T., and Rao, Z. (2008). Crystal structures of human BTG2 and mouse TIS21 involved in suppression of CAF1 deadenylase activity. Nucleic Acids Res. *36*, 6872–6881.

Yu, C., Ji, S.Y., Sha, Q.Q., Dang, Y., Zhou, J.J., Zhang, Y.L., Liu, Y., Wang, Z.W., Hu, B., Sun, Q.Y., et al. (2016). BTG4 is a meiotic cell cycle-coupled maternal-zygotic-transition licensing factor in oocytes. Nat. Struct. Mol. Biol. *23*, 387–394.

Cell Reports, Volume *23*

Supplemental Information

An RNA-Binding Multimer Specifies Nematode Sperm Fate

Scott T. Aoki, Douglas F. Porter, Aman Prasad, Marvin Wickens, Craig A. Bingman, and Judith Kimble

Supplemental Information

Supplemental Experimental Procedures

Biochemistry and crystallography

Protein expression and purification
Full length FOG-3 1-263 was amplified from *C. elegans* N2 cDNA with primers that included a six-histidine tag, stop codon and 12 nucleotides suitable for annealing with ligation independent cloning (LIC) (Aslanidis and de Jong, 1990). Mixed stage N2 cDNA was generated by reverse transcription (SuperScript® II Reverse Transcriptase, Thermo Fisher; Waltham, MA) with oligo-dT (Invitrogen; Carlsbad, CA). The FOG-3 PCR product was cloned into a pET21a (EMD Millipore; Billerica, MA) bacterial expression plasmid by LIC. From this plasmid, histidine-tagged FOG-3 1-238 was amplified by PCR and inserted into pET21a using LIC. For FOG-3 1-137, an *E. coli* codon-optimized sequence for histidine-tagged FOG-3 1-137 was ordered as a gBlock (Integrated DNA Technologies (IDT); Coralville, IA) and cloned into pET21a by LIC. All further changes to the pET21a histidine-tagged FOG-3 1-137 were by Gibson assembly cloning (Gibson, 2009). Residues 1-137 gave good yields, but solubility and stability remained issues at high concentration. We improved solubility by retaining the histidine tag and mutating non-conserved residues to amino acids in related FOG-3 orthologs (H47N and C117A) (**Figure S1**). For chemical crosslinking, we modified this optimized, histidine-tagged FOG-3 1-137 expression plasmid to include mutations (R22K, L64K, I112K, R82K) and extended the C-terminus three amino acids (138-140) to permit histidine tag cleavage (not used). This lysine substitution FOG-3 plasmid was later modified to include mutations R14K or E126K. Final plasmids used for protein expression listed (below).

Expression plasmids were transformed into Rosetta™2(DE3) cells (EMD Millipore; Billerica, MA) and grown in LB (MP Biomedicals; Santa Ana, CA) for 5 hours at 37°C until A600 = ~0.8. The culture was then induced with 0.1 mM IPTG (MP Biomedicals; Santa Ana, CA) and grown at 16°C for 16-20 hours prior to collection, centrifugation, washing and freezing in liquid nitrogen. Cells were defrosted on ice and reconstituted in lysis buffer (20 mM $NaPO_4$ pH 7.4, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol) with cOmplete protease inhibitors (Roche; Indianapolis, IN). Cells were lysed with a French Press, centrifuged (3,220 *x g* and 10,000 *x g*) to remove unlysed cells and precipitate, and incubated with Nickel-NTA beads (Thermo Fisher; Waltham, MA) for 2 hours at 4°C with rocking. Beads were washed with lysis buffer and eluted with an imidazole step gradient (imidazole at 20, 40, 60, 80, 100, 250 mM) in elution buffer (20 mM $NaPO_4$ pH 7.4, 300 mM NaCl, 5 mM β-mercaptoethanol). Protein used for biochemical experiments was dialyzed in FOG-3 buffer (20 mM HEPES pH 7.0, 50 mM NaCl, 0.5 mM (tris(2-carboxyethyl)phosphine) (TCEP)), while protein for crystallization was dialyzed in crystallization buffer (20 mM HEPES pH 7.0, 100 mM $MgS0_4$, 0.5 mM TCEP). Samples were concentrated with Amicon Ultra-4 3000 MW concentrators (EMD Millipore; Billerica, MA) and run on a Superdex 200 (GE Healthcare; Pittsburgh, PA). Recombinant protein was again concentrated with Amicon Ultra-4 3000 MW concentrators and protein concentration estimated by A280.

Plasmids
pJK2068: pET21a vector backbone, *C. elegans* FOG-3 1-238::six histidines
pJK2069: pET21a vector backbone, *E. coli* codon-optimized *C. elegans* FOG-3 1-137 H47N C117A::six histidines
pJK2070: pET21a vector backbone, *E. coli* codon-optimized *C. elegans* FOG-3 1-140 H47N C117A, lysine substitutions (R22K, I112K, L64K, R82K)::six histidines
pJK2071: pET21a vector backbone, *E. coli* codon-optimized *C. elegans* FOG-3 1-140 R14K H47N C117A, lysine substitutions (R22K, I112K, L64K, R82K)::six histidines
pJK2072: pET21a vector backbone, *E. coli* codon-optimized *C. elegans* FOG-3 1-140 H47N C117A E126K, lysine substitutions (R22K, I112K, L64K, R82K)::six histidines
pJK1910: pDD162 vector backbone, CRISPR-Cas9 sgRNA targeting *fog*-3 (caatcagtccccgagtacg)
pJK1925: pDD162 vector backbone, CRISPR-Cas9 sgRNA targeting *fog-3* (ggttctgaccacgtactcg)

Crystallization, data collection and refinement
Initial crystallization trials proved unfruitful. Thinking that a cofactor was missing, we performed a thermal folding assay (Ericsson et al., 2006) with various additives (see below). Magnesium and sulfate improved protein thermostability (**Figure S2C**) so we reasoned that they might aid stability during crystallization. Crystallization conditions were screened with sitting drop trays set up using the Mosquito (TTP Labtech; Cambridge, MA). We obtained crystals using recombinant FOG-3 (1-137 H47N C117A) with an intact histidine tag and incubating our

trays at 4°C. After 3 weeks, rhomboid crystals were observed in conditions A (0.1 M sodium citrate pH 5.6, 10% (vol/vol) isopropanol, 10% (wt/vol) PEG 4000) and B (0.1 M magnesium acetate, 0.1 M sodium citrate pH 5.6, 8% (wt/vol) PEG 10000). UV scanning with a UVEX-M (280 nm excitation, 350 nm emission; JANSi; Seattle, WA) identified these to be protein crystals. Both conditions were reproducible. We were able to collect complete datasets from the crystals grown directly from the condition B screening trays. Phasing was accomplished with molecular replacement using Phaser (McCoy et al., 2007) and a human Tob homolog (PDB ID: 2Z15) as a starting model. Model building and refinement were done in Phenix (Adams et al., 2010) and Coot (Emsley and Cowtan, 2004). Water molecules were first modeled by Phenix before being checked manually. Three densities were too big to be water molecules. We could model one of the densities with sulfate. Two densities were observed in the solvent-accessible area adjacent to residues 52-56 in both copies in the ASU. Density is observed at FoFc contour levels past 6 σ. We attempted modeling of acetate (too small) and citrate (too large), both molecules that were present in the crystallization conditions, but the fit was unsatisfactory. Thus, the final uploaded model does not account for these two large densities. Analyses of protein assemblies, dimer interactions and free energy estimations were done in PISA (Krissinel and Henrick, 2007). Structural images were generated in Pymol (The PyMOL Molecular Graphics System, Version 2.0,1 Schrödinger, LLC).

The FOG-3 missense mutations were modeled in Coot (Emsley and Cowtan, 2004). Their disruption of FOG-3 folding, dimerization or multimerization was inferred based on mutant residue disruption of hydrogen bonds, salt bridges and hydrophobic packing (steric hindrance) in the structure. The lysine substitution sites were chosen based on whether the locations were close enough for intra-dimer (R22K, I112K) and dimer-dimer (L64K, R82K) BS3 crosslinking, sequence variability (a lack of sequence conservation, **Figure S1**) and their expected tolerance for a lysine mutation. R22 was conserved in FOG-3 paralogs, but lysine was easily modeled. Two mutations changed hydrophobic residues, but lysines at these positions were found in other FOG-3 orthologs and could be modeled.

FOG-3 protease cleavage

Comparison of FOG-3 sequences from several *Caenorhabditid* species reveals further nematode-specific conservation that extends ~20 amino acids past the predicted Tob/BTG fold. We used recombinant *C. elegans* FOG-3 protein and proteases to identify a single domain spanning the canonical Tob/BTG fold and a nematode-specific extension. Recombinant FOG-3 1-238 with a C-terminal histidine tag was incubated with trypsin (Sigma-Aldrich; St. Louis, MO) and elastase (Sigma-Aldrich; St. Louis, MO) at room temperature prior to SDS-PAGE. The gel was stained with Coomassie to visualize cleavage products. The proteases generated ~15 kDa protected fragments (**Figure S2B**). Samples were also cleaved with trypsin or elastase for 45 minutes at room temperature (~20°C) and submitted for in-solution mass spectrometry (University of Wisconsin-Madison Biotechnology Center). The mass spectrometry fragment that most closely matched the SDS-PAGE band mapped to residues 1-135 for trypsin and 1-142 for elastase. Both fragments included the predicted Tob/BTG fold plus the nematode-specific extension (**Figure 1B**). FOG-3 1-137 exhibited a broad elution peak at higher versus lower concentrations (**Figure S2A**) that could be attributed to different dimer versus monomer states.

Protein folding assay

The protein folding assay followed published protocols (Ericsson et al., 2006). Briefly, recombinant FOG-3 1-137 with histidine tag was incubated with 90x concentrated SYPRO orange (5000x stock, Invitrogen; Carlsbad, CA) in FOG-3 buffer. 18 µl of the protein-dye mix was mixed with 2 µl Additive Screen (Hampton Research; Aliso Viejo, CA) and heated in a 7500 Real-Time PCR System thermocycler (Applied Biosystems; Foster City, CA) at 0.1°C/s from 20°C to 70°C while monitoring A405. SYPRO orange dye bound to unfolded protein. Thus, FOG-3 unfolded at a certain temperature, allowing dye binding and increasing A405 absorbance. The additive was judged as enhancing thermostability based upon the shift in the melting curve to the right, or requiring higher temperatures for signal. This assay was performed twice with similar results.

Negative-stain electron microscopy

Samples were negative stained with Nano-W (Nanoprobes; Yaphank, NY) using the two-step method. A 2 µl droplet of samples was placed on a Pioloform (T. Pella) coated 300 mesh Cu Thin-Bar grid (EMS; Hatfield, PA), coating side down. The excess was wicked with filter paper and allowed to barely dry. A 2 µl droplet of Nano-W was applied, wicked again with clean new filter paper, and allowed to dry. The sample was viewed on a Philips CM120 transmission electron microscope at 80 kV and documented with a SIS (Olympus / Soft Imaging Systems; Münster, Germany) MegaView III digital camera.

Protein crosslinking

Bis[sulfosuccinimidyl] suberate (BS3) (Thermo Scientific; Waltham, MA) was diluted in crosslinking buffer (20 mM HEPES pH 7.0, 150 mM NaCl) and added to recombinant protein for a final concentration of 0.5 mg/ml protein and 0.5, 0.25, 0.125, 0.0625, 0.03125 mM BS3 crosslinker. Buffer alone was added as a negative control. After 30 minutes at room temperature (~20°C), the reaction was quenched with 1M Tris pH 6.8 (50 mM final concentration). Samples were analyzed by SDS-PAGE and coomassie staining. Experiments were performed twice in these conditions with similar results.

***C. elegans* E126K alleles**

CRISPR-Cas9 gene editing of the endogenous *fog-3* gene was achieved using a *dpy-10* roller co-injection strategy (Arribere et al., 2014). Briefly, an sgRNA construct containing the U6 promoter and sgRNA scaffold from pDD162 (Dickinson et al., 2013) along with the targeting sequences caatcagtccccgagtacg (pJK1910) and ggttctgaccacgtactcg (pJK1925) was cloned into the *Xma*I site of pUC19 using one step isothermal DNA assembly. The repair template was a ssDNA oligo (Integrated DNA Technologies (IDT); Coralville, IA) that inserted an E126K mutation and removed an *Ava*I restriction site. See Table (below) for CRISPR-Cas9 target sequences and repair oligos used. Injections were performed in young N2 hermaphrodite *C. elegans,* using either 1) *fog-3* sgRNA plasmids, *dpy-10* sgRNA plasmid, *fog-3* E126K repair template, and Cas9 plasmid as described (Arribere et al., 2014), or 2) *fog-3* crRNA, *dpy-10* crRNA, *fog-3* E126K repair template and recombinant Cas9 protein as described (Paix et al., 2015) (see Table below for reagent sequences), and F1 rollers were screened for the desired mutation by PCR and *Ava*I digest. Alleles were recovered from separate injected animals and therefore represent independent editing events. We verified the *fog-3* mutations by Sanger sequencing. Homozygous mutants had a Fog phenotype and thus could only produce oocytes. These worms were outcrossed twice with N2 before crossing with JK2739 containing balancer *hT2[qIs48](I;III)*.

CRISPR-Cas9 guide RNAs and repair oligos

Name	Type	Strain targeted	Sequence
CRISPR-Cas9 guide RNAs:			
fog-3 5 crRNA	CRISPR-Cas9 RNA	N2	target sequence: taatactgggaaattaaaaa
fog-3 crRNA 8	CRISPR-Cas9 RNA	N2	target sequence: cggttctgaccacgtactcg
his-58 tbb-2 crRNA 1	CRISPR-Cas9 RNA	JA1515	target sequence: aggatcttgcatTTACTTGC
single stranded DNA repair oligos:			
fog-3 E126K repair	ssDNA repair oligo	N2	ataaaaatactttaaatttcatttttccagctaccaatcagtccccAagtaTgtTgtcCgaaccgctgcaatccgcgcggagccttgctcgaatcttgg
fog-3 FLAG3x repair	ssDNA repair oligo	N2	attcaagcatcaacgaccaaatgagatattctccccgtGGAGGATCCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGttttaatttcccagtattagaatctcaattatcataccgt
fog-3 lambda repair	ssDNA repair oligo	N2	tcaacgaccaaatgagatattctccccgtATGGACGCCCAAACCCGCCGCCGCGAGCGCCGTGCCGAGAAGCAAGCCCAATGGAAGGCCGCCAACGGAGGATCCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGttttaatttcccagtattagaatctcaatt
his-58 tbb-2 boxB repair	ssDNA repair oligo	JA1515	CAA GGC CGT CAC CAA GTA CAC TTC TAG CAA GTA AAT GCA AGA TCC AAC TAC TAA ACT GAT TCC TGG GCC CTG AAG AAG GGC CCC TCG ACT AAG TCC AAC TAC TAA ACT GGG CCC TGA AGA AGG GCC CAT ATA GGG CCC TGA AGA AGG GCC CTA TCG AGG ATA TTA TCT CGA CTT TCA AGC ATT CCC TTC TTC TCT ATC AC

Reporter design and application

The goal was to develop a reporter that could take advantage of the lambda/boxb mRNA tethering system (Baron-Benhamou et al., 2004) and a germline fluorescent expression reporter. However, the published boxb-containing reporters and our own MosSCI-generated reporters could not be detected, most likely due to weak signal in the

germline region expressing FOG-3. We therefore modified a worm strain with a robust germline fluorescent reporter. Using CRISPR-Cas9 gene editing (Paix et al., 2015), three boxB hairpins (3xboxB) were inserted into *weSi2* (JA1515) (Zeiser et al., 2011), a transgene reporter expressing a GFP-tagged histone under the ubiquitous *mex-5* germline promoter and the *tbb-2* 3'UTR. 3xboxB was a modified design based on previous mammalian 3'UTR boxB reporters (Wang et al., 2011). Inserts were screened by a phenotype-based co-injection marker (Arribere et al., 2014) and by PCR for potential inserts. Candidates were homozygosed and insertions sequenced for correct repair. After outcrossing, we confirmed that the allele-containing strains still expressed nuclear GFP throughout the adult germline. For the FOG-3 tagged alleles, either 3xFLAG or λN22::3xFLAG was inserted into *fog-3* at its C-terminus, one amino acid (R262) away from the end. Insertions were screened and sequenced as described for the reporter. Worms with correct insertions were homozygous fertile and could be maintained as cross-fertile male-hermaphrodite lines, strong evidence for fully functional FOG-3. These worms were crossed into the reporter and imaged as described. See accompanying Table (above) for a summary of the CRISPR-Cas9 target sequences and repair oligos used.

Imaging

For fluorescent imaging, germlines were extruded, fixed and permeabilized with Triton-X (0.5%) as previously described (Crittenden et al., 2017). Germlines were incubated with primary antibodies to FLAG (M2® (mouse), Sigma; St. Louis, MO) and GFP (Rabbit anti-GFP, Invitrogen; Carlsbad, CA) overnight, stained with fluorophore-labeled secondary antibodies (Alexa 555 Donkey anti-Mouse, Alexa 488 Goat anti-Rabbit; Invitrogen; Carlsbad, CA) and DAPI (Invitrogen; Carlsbad, CA), washed, and mounted in Vectashield (Vector Laboratories; Burlingame, CA). Germlines were imaged by confocal microscopy on a Leica SP8 scanning laser confocal microscope, taking 1 µm slices in sequence. Maximum intensity stack projections were generated with ImageJ (Schindelin et al., 2015) and brightness adjusted with Photoshop (Adobe; San Jose, CA). All images were equally treated in ImageJ and Photoshop, with the exception of the reported DAPI images. Imaging experiments were repeated at least twice with similar results.

iCLIP

In vivo crosslinking and immunoprecipitation (iCLIP) was carried out essentially as described (Huppertz et al., 2014), with modifications to worm growth, crosslinking, lysis, RNase digestion, and data analysis as described.

Nematode culture and UV crosslinking for iCLIP

L1 larvae from *C. elegans* strain JK4871 were obtained by bleaching and synchronizing via standard methods (Stiernagle, 2006). Larvae were plated onto 10 cm OP50 plates (~50,000 per plate) and propagated at 20°C for ~40-46 hours until most worms were at the early L4 stage when FOG-3 expression is greatest. Worms were washed with M9 (42.3 mM Na_2HPO_4, 22 mM KH_2PO_4, 85.6 mM NaCl, 1 mM $MgSO_4$), pooled into groups of 250,000 living worms, placed on a 10 cm NGM agarose plate and liquid removed. Animals were irradiated two times sequentially at 254 nm with 0.9999 J/cm^2 in a XL-1000 UV Crosslinker (Spectrolinker, Thomas Scientific; Swedesboro, NJ). Non-crosslinked samples were incubated at room temperature as a negative control for the radiolabeled gel (**Figure S4E**). For the iCLIP negative control, we performed the pulldown of crosslinked JK4871 worm lysate with beads alone (no antibody). Worms were rinsed from plates with cold M9, washed once, and transferred to a 2 mL Eppendorf tube. The pellet was washed again in freezing buffer (50 mM Tris pH 7.5, 150 mM NaCl, 10% (wt/vol) glycerol, 0.05% (vol/vol) Tween 20) and frozen with liquid nitrogen. Pellets were stored at -80ºC until use.

Lysis and RNA digestion

C. elegans pellets were thawed by adding ice cold lysis buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1% Pierce NP-40, 0.1% SDS, 0.5% sodium deoxycholate, Roche cOmplete EDTA-free Protease Inhibitor Cocktail, Ambion ANTI-RNase) and incubated for 20 minutes at 4ºC with rocking. The thawed pellets were centrifuged at 1,000 x *g*, 4ºC for 1 minute and washed 3 times with ice cold lysis buffer. Lysis buffer was added to the pellet along with a 5 mm Retsch stainless steel ball (Verder Scientific; Newtown, PA). Lysis was performed in the cold room using a Retsch 400 MM mill mixer (Verder Scientific; Newtown, PA). Lysis was completed after three 10-minute cycles at a setting of 30 Hz, with four-minute freeze-thaws after the first and second cycles. Freeze-thaws were performed by immersion in liquid nitrogen for 1 minute, then returning to liquid state by immersion in room temperature water for 4 minutes. Worm lysis was confirmed by observing a small aliquot of final lysate on a dissection scope. The lysate was cleared by centrifugation for 15 minutes at 16,100 x *g*, 4ºC. Protein concentration of the cleared lysate was determined with the Direct Detect spectrometer (EMD Millipore; Billerica, MA). Our pellets containing 250,000 worms yielded ~12 mg/mL of total protein, and we used 10 mg total protein per biological replicate. Double RNase

digestion of protein-RNA complexes was performed as previously described (Spitzer et al., 2014). For the first digestion, which occurred immediately after lysis and just prior to immunoprecipitation, guanosine specific RNase T1 (Thermo Fisher; Waltham, MA) was added to the cleared lysate at a final concentration of 1 Unit/µL. The sample was incubated in a Thermomixer for 15 minutes at 22ºC, 1,100 rpm and then cooled on ice for 5 minutes.

Protein G Dynabeads (Life Technologies; Waltham, MA) were aliquoted to a fresh RNase-free roundbottom tube (USA Scientific; Ocala, FL). The tube was placed on a Dynal magnet (Invitrogen; Carlsbad, CA), the existing buffer was removed, and M2 FLAG antibody (Sigma-Aldrich; St. Louis, MO) added at 20 µg antibody to 3 mg Protein G Dynabeads in PBS-T (PBS pH 7.2 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, 1.8 mM KH_2PO4), 0.02% Tween-20). The beads plus antibody solution was incubated at room temperature on a rotator for 45 minutes. The tube was again placed on the magnet, the antibody solution removed, and the cleared lysate added. Immunoprecipitation was carried out overnight at 4°C. As a negative control, we performed the pulldown of crosslinked JK4871 worm lysate with beads alone (no antibody).

Following immunoprecipitation, the beads were washed as described (Huppertz et al., 2014), with minor modifications. We performed washes in the cold room (~4ºC) with two wash buffers: a high-salt wash buffer (50 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA, 1% Pierce NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and PNK buffer (20 mM Tris-HCl pH 7.5, 10 mM MgCl2, 0.2% Tween 20). The second RNase T1 digestion was then performed on the washed beads at a final concentration of 100 Units/µL in PNK buffer. Samples were incubated in a Thermomixer for 15 minutes at 22ºC shaking at 1,100 rpm, cooled on ice for 5 minutes, and then processed through the remaining iCLIP protocol as described (Huppertz et al., 2014). We confirmed that FOG-3::3xFLAG crosslinked to RNA by visualizing 5' radioactively labeled RNA bound to the FOG-3 protein when antibody was present on the beads (**Figure S4E**). We confirmed immunoprecipitation of FOG-3 from experimental versus negative control samples by immunoblot with an M2 FLAG antibody (**Figure S4F**).

Single-end sequencing was performed on an Illumina HiSeq 2000 (University of Wisconsin Biotechnology Center; Madison, WI). The cDNA library of each replicate was prepared with a unique "Rclip" reverse transcription primer (as in Huppertz et al., 2014), which contained a partially randomized sequence (i.e., a "barcode"). The constant portion of the barcode enabled each read to be identified by replicate and allowed for replicate multiplexing. The randomized portion of the barcode allowed for computational filtering of artifacts from individual reads caused by PCR amplification of the cDNAs, such as read duplication. After high-throughput sequencing, the barcode sequence preceded the cDNA sequence and thus could be easily identified and removed prior to read mapping.

iCLIP sequence analysis

Reads (**Table S2**, Tab 1, column B) were aligned to the WS235 genome using STAR (Dobin et al., 2013) and previously described parameters (Kassuhn et al., 2016), except for the parameter --alignEndsType Local (mismatches at the ends of reads are tolerated). Multi-mapping reads were removed, and high-confidence mappings were selected as those with alignment scores of at least 20 (**Table S2**, Tab 1, column C). PCR duplicates were collapsed to unique reads (**Table S2**, Tab 1, column E) using the method described in Weyn-Vanhentenryck et al. (2014). Reads were assigned to genes using HTSeq (Anders et al., 2015). CIMS (crosslinking induced mutation sites) and CITS (crosslink induced truncation sites) analyses were performed as described (Weyn-Vanhentenryck et al., 2014), except we did not require CIMS to reproduce between replicates, and are included in **Table S3**, tab 3. For peak analysis, "clusters" were defined as regions of overlapping reads. Using the reads indicated in **Table S2**, tab 1, column E, all reads within a gene had their position randomized 1000 times to empirically determine a cluster p value as the odds of having a cluster with the given maximum read depth from randomized read positions. This is similar to the local Poisson method (Zisoulis et al., 2010), as the Poisson approximates of read scrambling. A Benjamini-Hochberg (BH) correction for multiple hypothesis testing was then applied at 1% FDR, resulting in the cluster numbers in **Table S2**, tab 2, column F. Finally, only overlapping clusters called independently as significant in at least 2 of the 4 replicates were retained as reproducible clusters, resulting in the cluster numbers in **Table S2**, tab 2, column G. Final clusters for FOG-3 and control samples are given in **Table S3**. While this is a simple method that does not account for background RNA abundance, it resulted in only 6 clusters for the negative control samples, suggesting it is effective at removing background in our datasets. We define peaks as all maxima at least 5 reads deep and at least 5% of the highest peak in the given gene; we counted neighboring peaks as distinct only if signal dropped to 50% or less of the lower peak maxima. Our definition of peaks differs from our definition of clusters, which are regions of continuous read coverage that pass the 1% FDR threshold. Clusters extend until iCLIP coverage drops to zero, thereby containing any number of distinct signal concentrations, and motivating our separate definitions of peaks and clusters.

To compare our results with previous FOG-3 RIP results (Noble et al., 2016), we calculated overlap with the top 722 FOG-3 targets, and evaluated significance by Fisher's exact test. To determine whether FOG-3 targets were associated with oogenesis, spermatogenesis, or mitosis, we used the method described previously (Noble et al., 2016), with significance evaluated by Fisher's exact test. Figures depicting iCLIP results (**Figure S4M,N**) were generated using Matplotlib (Hunter, 2007) and Python scripts available at https://github.com/dfporter.

We could not find enriched sequence motifs at FOG-3 binding sties. Instead, we sought motifs enriched near the binding sites and found enrichment of a CUCAC motif (**Figure S4P,** p value < 1.8×10^{-229}). CUCA is part of the GLD-1/STAR signature motif (Ryder et al., 2004). GLD-1 regulates germline sex determination, but it can promote either the sperm or oocyte fate (Francis et al., 1995).

Raw sequence files of all replicates are available through Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/; GEO accession GSE76521).

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

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

Arribere, J.A., Bell, R.T., Fu, B.X., Artiles, K.L., Hartman, P.S., and Fire, A.Z. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. Genetics *198*, 837-846.

Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res *18*, 6069-6074.

Dickinson, D.J., Ward, J.D., Reiner, D.J., and Goldstein, B. (2013). Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. Nat Methods *10*, 1028-1034.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Emsley, P., and Cowtan, K. (2004). *Coot*: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem *357*, 289-298.

Francis, R., Barton, M.K., Kimble, J., and Schedl, T. (1995). *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. Genetics *139*, 579-606.

Gibson, D.G. (2009). Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res *37*, 6984-6990.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Computing In Science & Engineering *9*, 90-95.

Kassuhn, W., Ohler, U., and Drewe, P. (2016). Cseq-Simulator: A data simulator for CLIP-Seq experiments. Pac Symp Biocomput *21*, 433-444.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J Mol Biol *372*, 774-797.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. J Appl Crystallogr *40*, 658-674.

Paix, A., Folkmann, A., Rasoloson, D., and Seydoux, G. (2015). High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. Genetics *201*, 47-54.

Ryder, S.P., Frater, L.A., Abramovitz, D.L., Goodwin, E.B., and Williamson, J.R. (2004). RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1. Nat Struct Mol Biol *11*, 20-28.

Schindelin, J., Rueden, C.T., Hiner, M.C., and Eliceiri, K.W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. Mol Reprod Dev *82*, 518-529.

Spitzer, J., Hafner, M., Landthaler, M., Ascano, M., Farazi, T., Wardle, G., Nusbaum, J., Khorshid, M., Burger, L., Zavolan, M., *et al.* (2014). PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. Methods Enzymol *539*, 113-161.

Stiernagle, T. (2006). Maintenance of *C. elegans*. In WormBook, T.*C.e.*R. Community, ed. (WormBook).

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., *et al.* (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120-124.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q.*, et al.* (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep *6*, 1139-1152.
Zeiser, E., Frøkjær-Jensen, C., Jorgensen, E., and Ahringer, J. (2011). MosSCI and gateway compatible plasmid toolkit for constitutive and inducible expression of transgenes in the *C. elegans* germline. PLoS One *6*, e20082.
Zisoulis, D.G., Lovci, M.T., Wilbert, M.L., Hutt, K.R., Liang, T.Y., Pasquinelli, A.E., and Yeo, G.W. (2010). Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. Nat Struct Mol Biol *17*, 173-179.

Figure S1

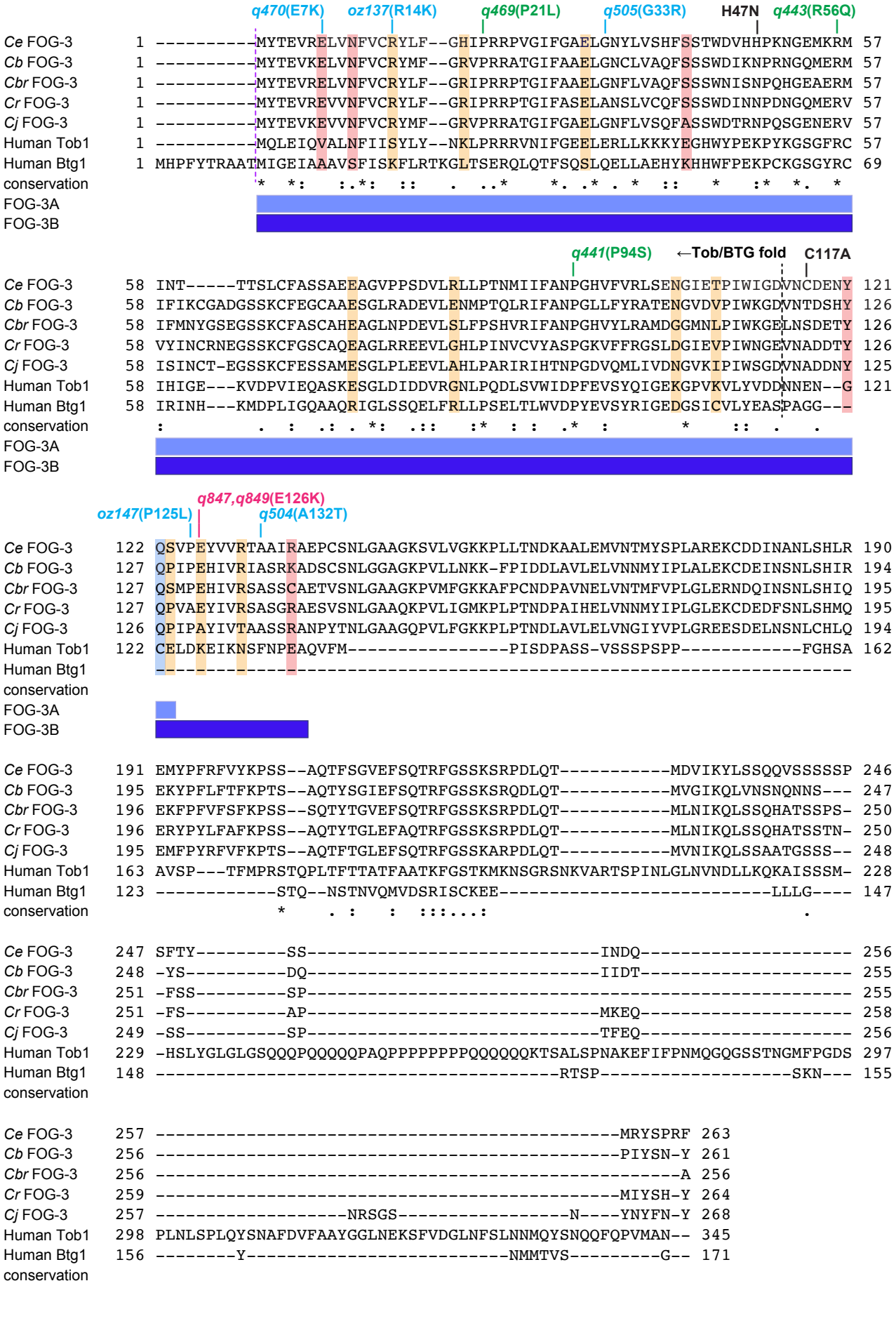
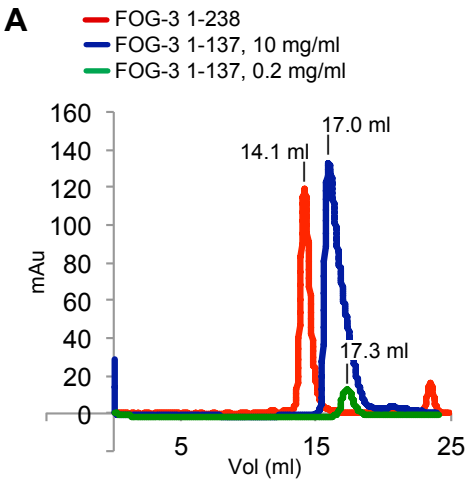
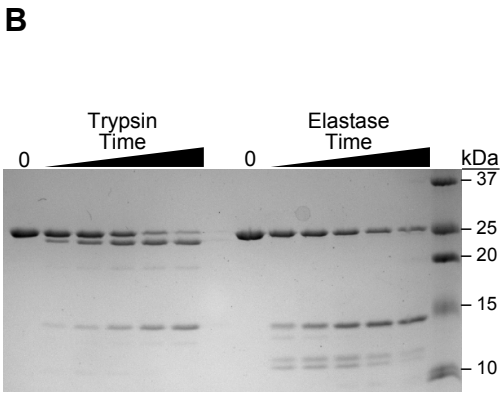
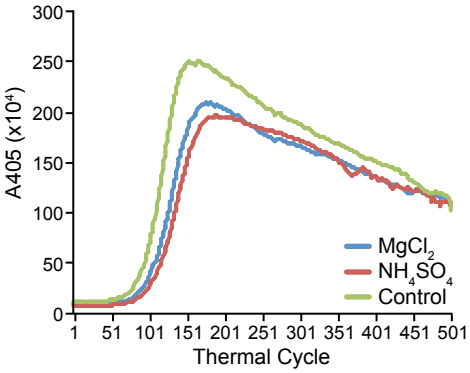

q470(E7K)
oz137(R14K)
q469(P21L)
q505(G33R)
H47N
q443(R56Q)
Ce FOG-3 1 ----------MYTEVRELVNFVCRYLF--GHIPRRPVGIFGAELGNYLVSHFSSTWDVHHPKNGEMKRM 57
Cb FOG-3 1 ----------MYTEVKELVNFVCRYMF--GRVPRRATGIFAAELGNCLVAQFSSSWDIKNPRNGQMERM 57
Cbr FOG-3 1 ----------MYTEVRELVNFVCRYLF--GRIPRRPTGIFAAELGNFLVAQFSSSWNISNPQHGEAERM 57
Cr FOG-3 1 ----------MYTEVREVVNFVCRYLF--GRIPRRPTGIFASELANSLVCQFSSSWDINNPDNGQMERV 57
Cj FOG-3 1 ----------MYTEVKEVVNFVCRYMF--GRVPRRATGIFGAELGNFLVSQFASSWDTRNPQSGENERV 57
Human Tob1 1 ----------MQLEIQVALNFIISYLY--NKLPRRRVNIFGEELERLLKKKYEGHWYPEKPYKGSGFRC 57
Human Btg1 1 MHPFYTRAATMIGEIAAAVSFISKFLRTKGLTSERQLQTFSQSLQELLAEHYKHHWFPEKPCKGSGYRC 69
conservation
FOG-3A
FOG-3B
q441(P94S)
←Tob/BTG fold
C117A
Ce FOG-3 58 INT-----TTSLCFASSAEEAGVPPSDVLRLLPTNMIIFANPGHVFVRLSENGIETPIWIGDVNCDENY 121
Cb FOG-3 58 IFIKCGADGSSKCFEGCAAESGLRADEVLENMPTQLRIFANPGLLFYRATENGVDVPIWKGDVNTDSHY 126
Cbr FOG-3 58 IFMNYGSEGSSKCFASCAHEAGLNPDEVLSLFPSHVRIFANPGHVYLRAMDGGMNLPIWKGELNSDETY 126
Cr FOG-3 58 VYINCRNEGSSKCFGSCAQEAGLRREEVLGHLPINVCVYASPGKVFFRGSLDGIEVPIWNGEVNADDTY 126
Cj FOG-3 58 ISINCT-EGSSKCFESSAMESGLPLEEVLAHLPARIRIHTNPGDVQMLIVDNGVKIPIWSGDVNADDNY 125
Human Tob1 58 IHIGE---KVDPVIEQASKESGLDIDDVRGNLPQDLSVWIDPFEVSYQIGEKGPVKVLYVDDNNEN--G 121
Human Btg1 58 IRINH---KMDPLIGQAAQRIGLSSQELFRLLPSELTLWVDPYEVSYRIGEDGSICVLYEASPAGG---
conservation
FOG-3A
FOG-3B
q847,q849(E126K)
oz147(P125L)
q504(A132T)
Ce FOG-3 122 QSVPEYVVRTAAIRAEPCSNLGAAGKSVLVGKKPLLTNDKAALEMVNTMYSPLAREKCDDINANLSHLR 190
Cb FOG-3 127 QPIPEHIVRIASRKADSCSNLGGAGKPVLLNKK-FPIDDLAVLELVNNMYIPLALEKCDEINSNLSHIR 194
Cbr FOG-3 127 QSMPEHIVRSASSCAETVSNLGAAGKPVMFGKKAFPCNDPAVNELVNTMFVPLGLERNDQINSNLSHIQ 195
Cr FOG-3 127 QPVAEYIVRSASGRAESVSNLGAAQKPVLIGMKPLPTNDPAIHELVNNMYIPLGLEKCDEDFSNLSHMQ 195
Cj FOG-3 126 QPIPAYIVTAASSRANPYTNLGAAGQPVLFGKKPLPTNDLAVLELVNGIYVPLGREESDELNSNLCHLQ 194
Human Tob1 122 CELDKEIKNSFNPEAQVFM----------PISDPASS-VSSSPSPP----------FGHSA 162
Human Btg1 ---
conservation
FOG-3A
FOG-3B
Ce FOG-3 191 EMYPFRFVYKPSS--AQTFSGVEFSQTRFGSSKSRPDLQT----------MDVIKYLSSQQVSSSSSP 246
Cb FOG-3 195 EKYPFLFTFKPTS--AQTYSGIEFSQTRFGSSKSRQDLQT----------MVGIKQLVNSNQNNS--- 247
Cbr FOG-3 196 EKFPFVFSFKPSS--SQTYTGVEFSQTRFGSSKSRPDLQT----------MLNIKQLSSQHATSSPS- 250
Cr FOG-3 196 ERYPYLFAFKPSS--AQTYTGLEFAQTRFGSSKSRPDLQT----------MLNIKQLSSQHATSSTN- 250
Cj FOG-3 195 EMFPYRFVFKPTS--AQTFTGLEFSQTRFGSSKARPDLQT----------MVNIKQLSSAATGSSS-- 248
Human Tob1 163 AVSP---TFMPRSTQPLTFTTATFAATKFGSTKMKNSGRSNKVARTSPINLGLNVNDLLKQKAISSSM- 228
Human Btg1 123 ----------STQ--NSTNVQMVDSRISCKEE----------LLLG---- 147
conservation
Ce FOG-3 247 SFTY---------SS---------INDQ---------- 256
Cb FOG-3 248 -YS----------DQ---------IIDT---------- 255
Cbr FOG-3 251 -FSS---------SP---------- 255
Cr FOG-3 251 -FS----------AP---------MKEQ---------- 258
Cj FOG-3 249 -SS----------SP---------TFEQ---------- 256
Human Tob1 229 -HSLYGLGLGSQQQPQQQQQQPAQPPPPPPPPQQQQQQQKTSALSPNAKEFIFPNMQGQGSSTNGMFPGDS 297
Human Btg1 148 ----------RTSP----------SKN--- 155
conservation
Ce FOG-3 257 ----------MRYSPRF 263
Cb FOG-3 256 ----------PIYSN-Y 261
Cbr FOG-3 256 ----------A 256
Cr FOG-3 259 ----------MIYSH-Y 264
Cj FOG-3 257 ----------NRSGS----------N----YNYFN-Y 268
Human Tob1 298 PLNLSPLQYSNAFDVFAAYGGLNEKSFVDGLNFSLNNMQYSNQQFQPVMAN-- 345
Human Btg1 156 --------Y----------NMMTVS---------G-- 171
conservation

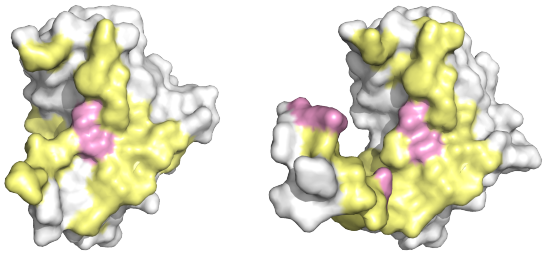
Figure S1. Alignment of FOG-3 ortholog sequences, related to Figure 1. Amino acid sequence alignment of FOG-3 orthologs, including human Tob and BTG proteins. Nematode orthologs include FOG-3 from *C. elegans* (*Ce*), *C. briggsae* (*Cb*), *C. brenneri* (*Cbr*), *C. remanei* (*Cr*), and *C. japonica* (*Cj*). Alignment by T-Coffee (Di Tommaso et al., 2011). Conservation noted by identity (*) plus high (:) or moderate (.) similarity. Missense alleles are labeled with their amino acid changes (Chen et al., 2000; this work); allele labels are colored to mark conservation among most metazoan orthologs (green), conservation among most nematode orthologs (blue) and a mutation generated in this study (magenta). Boundary of the canonical Tob/BTG fold is marked with a dashed line; extents of dimer subunits are shown below, including subunit A (light blue) and subunit B (dark blue). Amino acids highlighted indicate dimer subunit-subunit contacts (red), dimer-dimer contacts (orange), and both (blue); these contacts include both hydrogen bonds and salt bridges.

Figure S2









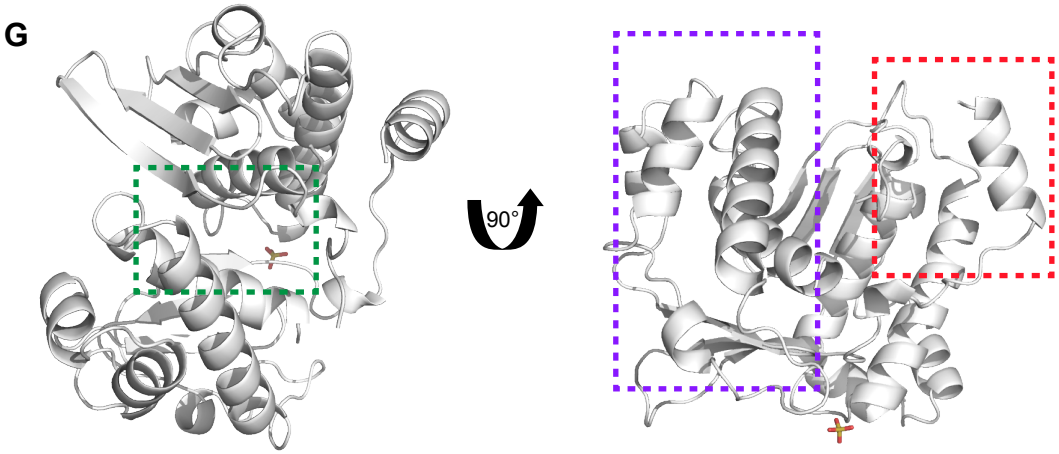
E

Dimer subunit: H-bonds

A	B
E7	N10
N10	E7
S41	R135
Y121	Q122

F

Mutation	Conservation	Site in structure	Proposed defect
P21L	All Tob/BTG folds	Tob/BTG fold	Tob/BTG Folding
R56Q	All Tob/BTG folds	Tob/BTG fold	Tob/BTG Folding
P94S	All Tob/BTG folds	Tob/BTG fold	Tob/BTG Folding
E7K	FOG-3 Tob/BTG fold	Dimer interface	Dimerization
G33K	FOG-3 Tob/BTG fold	Dimer interface	Dimerization
R14K	FOG-3 Tob/BTG fold	Dimer interface	Dimer/polymer
P125L	FOG-3 extension	Linker-helix	Dimer/polymer
A132T	FOG-3 extension	Linker-helix	Dimer/polymer
E126K	FOG-3 extension	Linker-helix	Polymerization



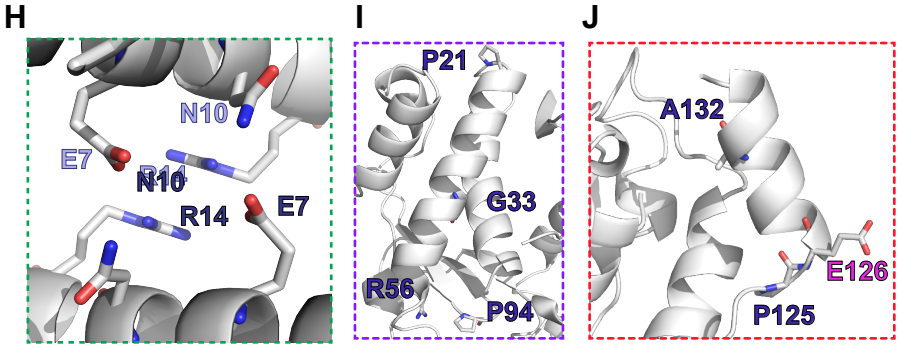
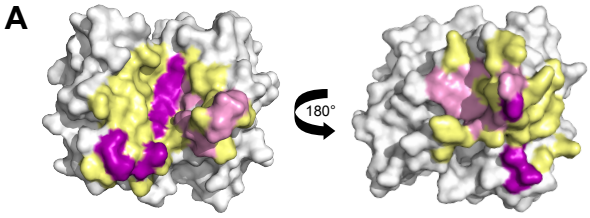


Figure S2. FOG-3 biochemical characterization, dimer subunit protein contacts and missense mutants, related to Figure 1. (**A**) Size exclusion chromatography elution profile of recombinant FOG-3 protein. Red, amino acids 1-238 with its histidine tag at 10 mg/ml; blue, amino acids 1-137 with histidine tag at 10 mg/ml; green, amino acids 1-137 with histidine tag at 0.2 mg/ml. Position of max peaks labeled. A280 milliabsorbance units, mAU. (**B**) Protease mapping the FOG-3 Tob/BTG-containing domain. FOG-3 1-238 was incubated with either trypsin or elastase, and samples were collected over time. Incubation with either protease produced a cleavage product of ~15 kDa. Protein incubated without protease labeled as "0." (**C**) Thermal-folding assay reveals domain stabilization with magnesium and sulfate. See Supplemental Experimental Procedures and main text for further details. (**D**) Contact surface between FOG-3 dimer subunits. Residues making hydrogen bonds (pink) and other contacting residues (yellow) based on distance. (**E**) Table of subunit-subunit hydrogen bonds. (**F**) Summary of FOG-3 missense mutations. (**G**) FOG-3 dimer with boxed regions enlarged in **H-J**. (**H-J**) Residues from FOG-3A are colored light blue and those from FOG-3B are dark blue. (**H**) Sites of *fog-3* missense mutants at the Tob/BTG domain dimer interface. (**I**) Sites of *fog-3* missense mutants conserved across FOG-3 orthologs and human Tob proteins. (**J**) Sites of *fog-3* missense mutants in linker-helix extension, including one generated in this study (magenta).



B

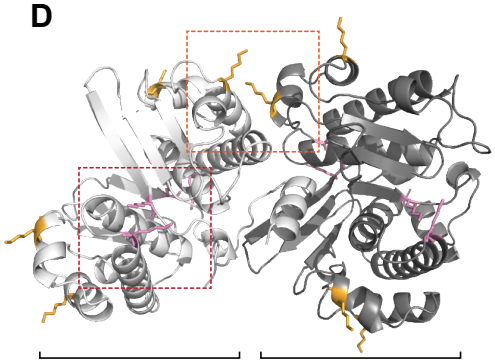
Dimer-dimer H-bonds

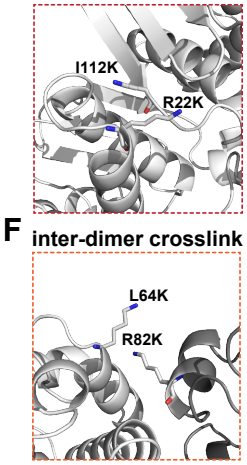
Dimer 1	Dimer 2
A:F11	B:R130
A:E103	A:S123
A:N104	A:D118
A:N104	A:Y121
A:N104	A:Q122
A:G105	A:Q122
A:G105	B:Q122
A:E107	B:Q122
A:E107	A:V124
A:T108	B:V124

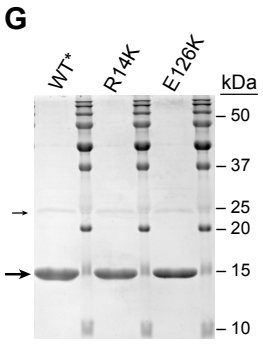
C

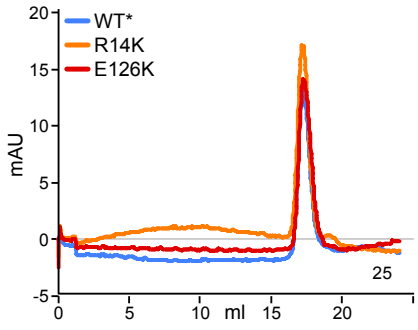
Dimer-dimer salt bridges

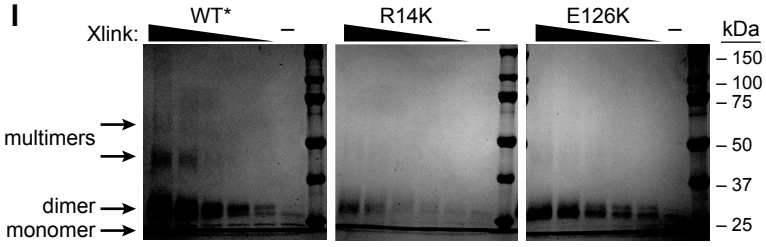
Dimer 1	Dimer 2
A:R14	B:E126
B:R14	B:E126
B:H19	A:E31
B:R82	A:E72

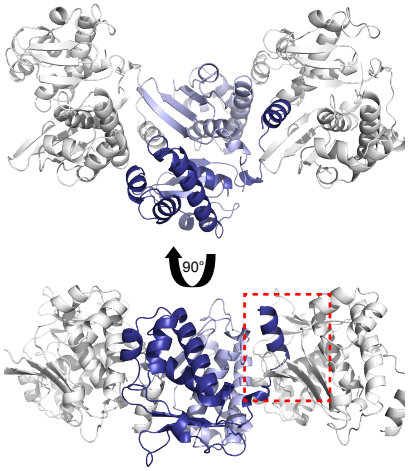


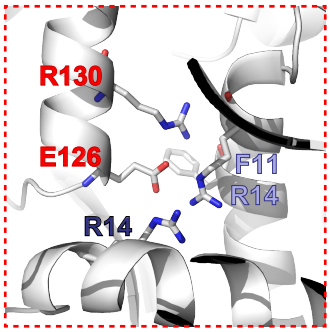






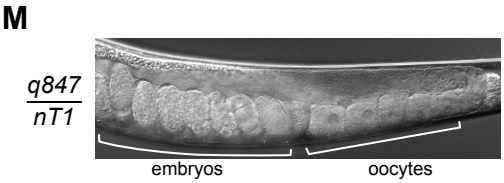






L

genotype	% fertile	% Fog	n
fog-3(q847)/nT1	100	0	105
fog-3(q847)/(q847)	0	100	106
fog-3(q849)/nT1	100	0	107
fog-3(q849)/(q849)	0	100	110



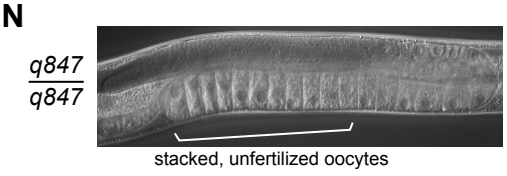
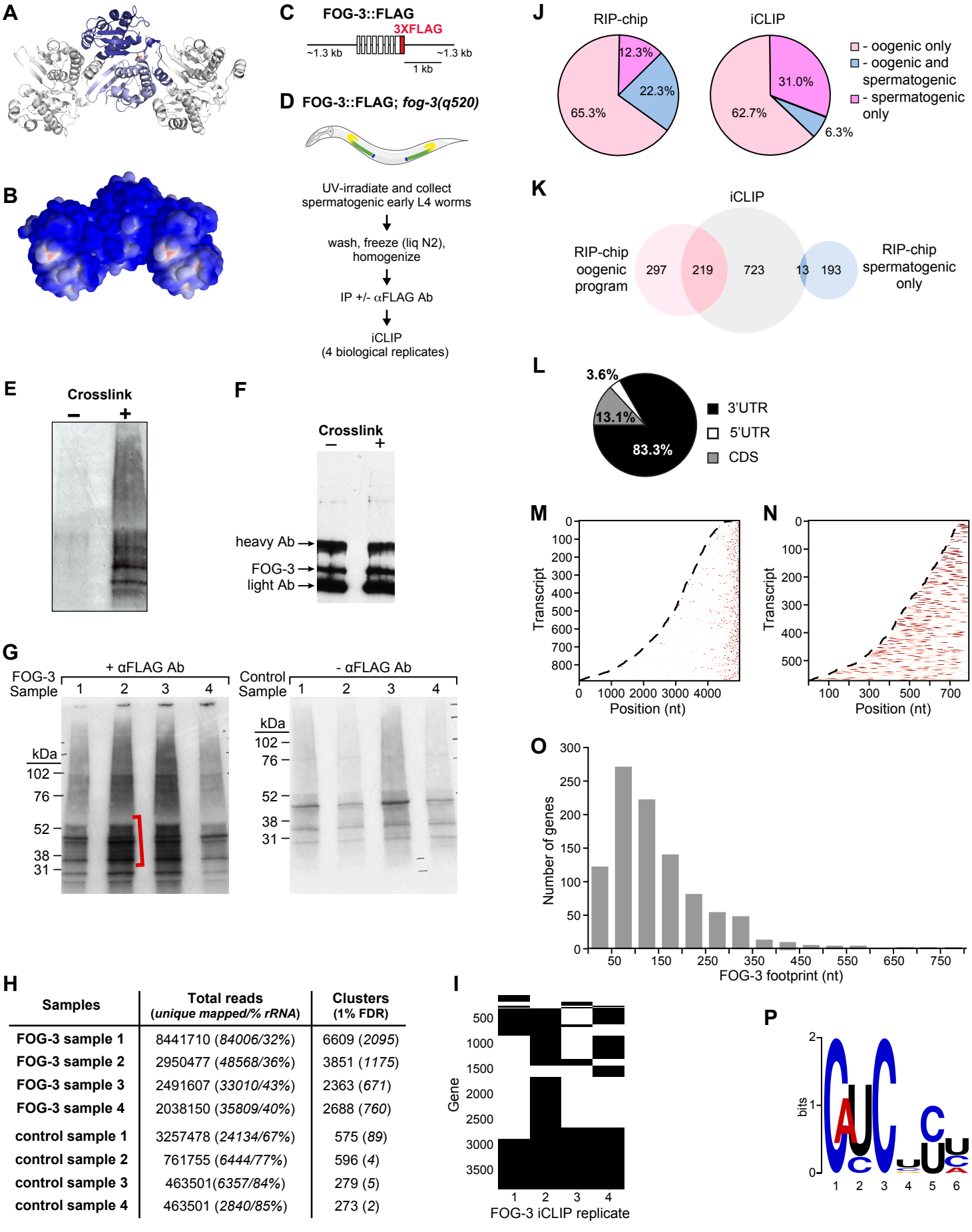


Figure S3. Dimer-dimer subunit protein contacts and supplemental biochemical analyses, related to Figure 2. (**A**) Interacting residues between FOG-3 dimers. Salt bridges (magenta), hydrogen bonds (pink), and contacting residues (yellow). (**B**) Table of predicted dimer-dimer hydrogen bonds. (**C**) Table of predicted dimer-dimer salt bridges. (**D-F**) Lysine substitution of recombinant FOG-3 generates residues permitting intra- and inter-dimer crosslinks. Location of lysine substitutions in the context of two FOG-3 dimers (**D**), which are enlarged in **E** (red box) and **F** (orange box). **(E)** Lysine substitutions R22K and I112K facilitate an intra-dimer crosslink. **(F)** Lysine substitutions L64K and R82K facilitate an inter-dimer crosslink. (**G**) Coomassie-stained gel of purified FOG-3 recombinant proteins. Left, "wild type" (WT*) FOG-3 (1-140, H47N C117A) with lysine substitutions; middle, missense mutant R14K predicted to abrogate dimer formation; right, missense mutant E126K predicted to abrogate polymer formation. All proteins (10 µg each) ran at ~15 kDa (large arrow). In addition, a minor ~25 kDa contaminant was observed (small arrow). (**H**) Size exclusion chromatography elution profile of WT*, R14K, and E126K recombinant FOG-3 (0.2 mg/ml). A280 milli-absorbance units, mAU. (**I**) Different exposure of **Figure 2D**. Coomassie-stained gel of modified FOG-3 recombinant protein incubated with increasing amounts of BS3 crosslinker. "–" represents no BS3 included. (**J**) Crystal packing of the FOG-3 dimer. FOG-3A and FOG-3B in the asymmetric unit are represented in light and dark blue, respectively. Each dimer buries the helix extension in FOG-3B into the adjacent dimer. The helix extension (red box) is enlarged in **K**. (**K**) Packing of the helix extension of one FOG-3 subunit into the adjacent dimer. Amino acids in subunits colored as in **J**, except for red residues from helix extension of the adjacent dimer. (**L**) Mutation of a key residue in the helix extension sexually transforms the germline (Fog phenotype). Two identical but independently-generated CRISPR-Cas9 alleles (*q847* and *q849*) mutated glutamate 126 to ~~a~~ lysine (E126K). Alleles were placed over a GFP-expressing balancer (*nT1*). Heterozygous (green) and homozygous (non-green) L4 worms were singled and analyzed 3 and 4 days later for fertility and the Fog phenotype. (**M-N**) Representative DIC images of hermaphrodite adults, heterozygous (**M**) or homozygous (**N**) for the E126K mutation. The embryos in heterozygotes demonstrate fertility, while oocytes stacking in homozygotes demonstrate sterility due to lack of sperm and hence lack of embryo production.

Figure S4



H

Samples	Total reads (*unique mapped/% rRNA*)	Clusters (1% FDR)
FOG-3 sample 1	8441710 (*84006/32%*)	6609 (*2095*)
FOG-3 sample 2	2950477 (*48568/36%*)	3851 (*1175*)
FOG-3 sample 3	2491607 (*33010/43%*)	2363 (*671*)
FOG-3 sample 4	2038150 (*35809/40%*)	2688 (*760*)
control sample 1	3257478 (*24134/67%*)	575 (*89*)
control sample 2	761755 (*6444/77%*)	596 (*4*)
control sample 3	463501(*6357/84%*)	279 (*5*)
control sample 4	463501 (*2840/85%*)	273 (*2*)

Figure S4. FOG-3 binds RNA *in vivo,* related to Figure 3 and Experimental Procedures. (**A**) Model of FOG-3 multimer composed of three dimers; subunits colored as in **Figure 1**. (**B**) Electrostatic surface potential of multimer modeled in **A**. Blue, basic; red, acidic. (**C**) Diagram of rescuing, epitope-tagged FOG-3::3xFLAG transgene, adapted from (Noble et al., 2016). (**D**) Outline of FOG-3 iCLIP protocol. See Supplemental Experimental Procedures for details. (**E-F**) FOG-3 crosslinks with RNA *in vivo*. Living worms expressing FOG-3::3xFLAG were UV-crosslinked (+) or mock treated (-), and then immunoprecipitated with αFLAG antibody. Bound sample 5' radiolabeled and run on SDS-PAGE. (**E**) Overnight exposure of radiolabeled samples run on SDS-PAGE. (**F**) Immunoblot of FOG-3::3xFLAG immunoprecipitation samples, visualized with αFLAG antibody. FOG-3 and heavy and light antibody chains labeled. (**G**) Gel analysis of samples used for iCLIP. Each sample was immunoprecipitated with (+) or without (-) αFLAG antibody, radiolabeled and run on the SDS-PAGE gel. The region above the expected size for FOG-3::3xFLAG (red bracket) was used for iCLIP processing and sequencing. Samples include four biological replicates and their paired controls. (**H**) FOG-3 iCLIP reads and cluster (regions of overlapping reads) statistics. Unique mapped reads (middle column) were determined by mapping with STAR, filtering out multimapping reads and low confidence alignments, and collapsing duplicate reads (see Supplemental Experimental Procedures). The fraction of unique mapped reads that mapped to rRNA is given as a percentage. The number of significant clusters at FDR 1% (right column) is highly dependent on FOG-3 purification. See Supplemental Experimental Procedures for further details. (**I**) Targets (black) identified for separate FOG-3 replicates. (**J**) Comparison of FOG-3 targets identified using FOG-3 iCLIP versus FOG-3 RIP-chip (microarray). Targets belonging to the oogenic program include mRNAs found only in oogenic germlines as well as mRNAs found in both oogenic and spermatogenic germlines, as described (Noble et al., 2016). (**K**) Venn diagram of mRNA target overlap between FOG-3 iCLIP (this study) and FOG-3 RIP-chip (FOG-3 IP with microarray analysis of associated RNAs) (Noble et al., 2016). (**L**) Percentages of FOG-3 iCLIP reads in mRNA regions. (**M-N**) FOG-3 binding sites are represented on a heat map, from no signal (white) to strong signal (red). Only genes with annotated 3'UTRs of at least 50 nt were included. (**M**) FOG-3 binding sites are at the 3' end of mRNAs, which are arranged by predicted nucleotide (nt) length from 5' to 3'. Dashed line marks the 5' end. Note prevalence of FOG-3 binding sites at the 3' ends of transcripts. (**N**) Binding sites occur throughout 3'UTRs, which are arranged by predicted 3'UTR nucleotide (nt) length from stop codon (dashed line) to 3' end. (**O**) FOG-3 iCLIP footprint on target transcripts. Coverage includes transcript regions above two reads deep. (**P**) Motif analysis of iCLIP clusters. Analysis and image generated by MEME (Bailey et al., 2015), except T was replaced by U.

Table S1. Data collection and refinement statistics, related to Figure 1.

	C. elegans FOG-3
PDB ID	5TD6
Wavelength (Å…)	0.9537
Resolution range (Å…)	29.22 - 2.034 (2.106 - 2.034)
Space group	P 31 2 1
Unit cell	64.992 64.992 133.57 90 90 120
Total reflections	147208 (14689)
Unique reflections	21617 (2091)
Multiplicity	6.8 (7.0)
Completeness (%)	99.79 (98.18)
Mean I/sigma(I)	24.12 (2.56)
Wilson B-factor	39.54
R-merge	0.0498 (0.7383)
R-meas	0.05399
CC1/2	1 (0.789)
CC*	1 (0.939)
Reflections used for R-free	1993
R-work	0.1769 (0.2255)
R-free	0.2285 (0.2886)
Number of non-hydrogen atoms	2210
macromolecules	2067
ligands	5
water	138

Protein residues	260
RMS(bonds)	0.009
RMS(angles)	1.10
Ramachandran favored (%)	98
Ramachandran allowed (%)	2
Ramachandran outliers (%)	0
Clashscore	3.89
Average B-factor	46.60
macromolecules	46.70
ligands	32.50
solvent	45.50

Statistics for the highest-resolution shell are shown in parentheses.