Title

In silico screening by AlphaFold2 program revealed the potential binding partners of nuage-localizing proteins and piRNA-related proteins

Authors

Kawaguchi Shinichi[1,+,*], Xu Xin[1,+], Soga Takashi[2], Yamaguchi Kenta[3], Kawasaki Ryuuya[3], Shimouchi Ryota[4], Date Susumu[2,4], Kai Toshie[1,*]

Affiliations

[1] Graduate School of Frontier Biosciences, Osaka University, Osaka 565-0871, Japan

[2] D3 Center, Osaka University, Osaka 567-0047, Japan

[3] NEC Solution Innovators, Ltd., Tokyo136-8627, Japan

[4] Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan

+ These authors contributed equally to this work

* Corresponding authors

E-mail: kawaguchi.shinichi.fbs@osaka-u.ac.jp (KS)

E-mail: kai.toshie.fbs@osaka-u.ac.jp (KT)

Key words

AlphaFold2, protein-protein interaction, in silico screening, piRNA, Nuage

Abstract

Protein-protein interactions are the fundamental features for understanding the molecular functions and regulations of proteins. Despite extensive databases, many interactions remain uncharacterized due to the intensive labor required for experimental validation. In this study, we utilized the AlphaFold2 program to predict interactions among proteins localized in the nuage, a germline-specific non-membrane organelle critical for piRNA biogenesis and RNA regulation. We screened 20 types of nuage proteins for 1:1 interactions and predicted dimer structures. Among those, five pairs represented novel interaction candidates. Three pairs, including Spn-E_Squ, were validated through co-immunoprecipitation in cultured cells and confirmed the interactions. Disruption of the salt bridges at the Spn-E_Squ interface verified their functional importance, underscoring the predictive model's accuracy. Our analysis was extended to include interactions between three representative nuage components, Vas, Squ, and Tej, and approximately 430 oogenesis-related proteins. Following this extended analysis, co-immunoprecipitation in S2 cells verified interactions for three pairs: Mei-W68_Squ, CSN3_Squ, and Pka-C1_Tej. Furthermore, the majority of *Drosophila* proteins, ~12,000, were screened for the interaction with Piwi protein, a central player in the piRNA pathway. Approximately 1.5% of the pairs, totaling 164 pairs, with a score above 0.6, were identified as potential binding partners. This in silico approach not only efficiently identifies potential interaction partners but

also significantly reduces the gap by facilitating the integration of bioinformatics and experimental biology.

Introduction

Around 10,000 to 20,000 different types of proteins are encoded in the genome of most organisms, catalyzing the vast majority of physico-chemical reactions in cells[1]. Many proteins have specialized functions and they are often regulated through protein-protein interactions, where the formation of protein complexes can activate, inhibit, or stabilize their partners. Furthermore, protein-protein interactions can recruit target proteins to specific locations where they will function or regulate the mobility of the protein complex[2]. Within cells, proteins are thought to exist in a crowded environment and frequently interact with other molecules[3]. Thus, characterizing protein-protein interactions is fundamental for understanding protein function and regulation. Large-scale analyses of protein-protein interactions have been carried out, including Tandem Affinity Purification coupled with Mass Spectrometry (TAP-MS) for the yeast proteome[4] and the comprehensive 2-hybrid screening for the Human Reference Interactome (HuRI)[5]. Despite these extensive studies, the overall protein-protein interactions are still not fully understood in many organisms.

The binding between proteins is significantly influenced by their three-dimensional (3D) structures. The characteristics of their interfaces, including hydrogen bonds, salt bridges, and hydrophobicity, determine the interactions[6]. Therefore, to analyze protein-protein interactions physically and chemically, information on the individual 3D structures of proteins is

necessary. The 3D structures of proteins have been determined through experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy[7]. However, these techniques demand considerable labor and time. The recently developed AlphaFold2 program can predict the 3D structure from its amino acid sequence with high accuracy[8]. AlphaFold2 requires sequence homology information to predict protein-protein interactions and the complex structure model. The reliability of these predictions is basically dependent on the strength of co-evolutionary signals[9]. This tool has not only been utilized in computational studies but has also become a valuable resource in experimental sciences for predicting protein complexes, as demonstrated with yeast protein complexes[10].

In this study, we attempted a rapid screening of the protein interactions using AlphaFold2 prediction, primarily focusing on components of nuage, a germline specific, non-membrane organelle that involves wide variety of proteins containing unique motifs and domains in *Drosophila melanogasater*[11]. Nuage is known to serve as the production and amplification site for small non-coding piRNA, which is bound to PIWI-family proteins. The piRNAs and the PIWI family proteins function to repress mobile genetic elements, or transposons, that disrupt the genomes through their active transpositions[12]. Not only proteins involved in piRNA production, but also translation repressor proteins including Me31B, Cup, and Trailer hitch (Tral), also localize in nuage[13]. Previous studies have shown that the

localization of several components in nuage depends on their partners in a hierarchical manner[14]. However, the interaction and organization among nuage components remain unclear.

By using AlphaFold2 predictions, we investigated 20 of the nuage-localizing or piRNA-related proteins for pairwise interactions. AlphaFold2 was initially trained to predict the structure of individual proteins[8]. Its application to complex prediction is an extrapolative use beyond its original intended scope, and its accuracy remains unverified. Even high-confidence predictions may not correspond to actual interactions, necessitating experimental validation to confirm whether predicted protein dimers truly bind. In this study, we confirmed the novel interactions of candidate pairs including Spindle-E (Spn-E)_Squash (Squ), by co-immunoprecipitation assay using cultured cells. In addition, a Squ mutant, which disrupts the salt bridges predicted at the interface with Spn-E, failed to interact with Spn-E, validating the accuracy of the predicted dimer structure. This screening was expanded for direct interacting pairs between piRNA-related proteins and proteins involved in oogenesis, as well as Piwi and other *Drosophila* proteins. This in silico approach not only streamlines the identification of interaction partners but also bridges the gap between bioinformatics predictions and experimental validation in biological research.

Result/Discussion

The nuage-localizing proteins and piRNA-related proteins used in the AlphaFold2 screening

Several dozen proteins engaged in piRNA production in germline cells exert their function by recruiting piRNA precursors and interacting with their partner proteins, forming non-membrane structure called a nuage[11,14]. Previous studies reported that many piRNA-related proteins localized to nuage and some proteins localized in mitochondria (Table 1). In addition, protein components of processing bodies and sponge bodies, which are involved in the translation, storage, degradation, and transportation of mRNAs—such as Me31B, Cup, and Tral—also localize to nuage[13] (Table 1). However, the details of how these proteins interact and organize themselves within the nuage remain unclear.

In this study, we used the AlphaFold2 program to screen for interactions among 20 proteins that are localized in the nuage and/or involved in piRNA production in *Drosophila* (Table 1). The monomeric structures of these 20 proteins, ranging in size from 20 kDa to 250 kDa, have already been predicted and are registered in databases[15]. This set includes both well-structured proteins and those that are largely disordered with numerous loops (Supplemental Fig. S1A). Of those, eight proteins feature one or more Tudor domains or extended Tudor (eTud) domains. The Tudor domain contains approximately 60 residues and folds into an antiparallel β-sheet with five strands forming a barrel-like fold, while the eTud domains include an additional

Oligonucleotide/oligosaccharide-Binding fold domain[16]. Both Tudor and eTud domains are known to bind predominantly to methylated lysine or arginine residues. In addition, five RNA helicases, such as Vasa (Vas) and the fly homolog of Tdrd9, Spn-E, which are essential for piRNA processing, are also included (Table 1). The Vas's C-terminal region is known to bind to the Lotus domain shared by two nuage components, Tejas (Tej) and Tapas. Spn-E is also recently shown to interact with Tej[17]. Among those 20 proteins, the Molecular Interaction Search Tool (MIST), a conventional database of protein-protein interactions, registers eight interacting pairs as direct binding, and 28 interactions which are direct or indirect (Table 1, Supplemental Fig. S1B, C)[18].

Screening for the protein-protein interactions by AlphaFold2

We used AlphaFold2 program to predict the direct protein-protein interaction and 3D structure of the complex. Assuming a 1:1 binding of 20 types of proteins, a total of 400 pairs of dimer predictions were calculated by a supercomputer. The prediction flow of AlphaFold2 consisted of two main parts[8]. Initially, a multiple sequence alignment was performed for each query protein and stored for the future use. Subsequently, the AlphaFold2 program predicted 3D dimer structures based on the co-evolution inferred from the multiple sequence alignments. For each dimer prediction, five different structure models with varying parameters were generated.

Among these, the model with the highest prediction confidence score (ranking confidence) was selected as the final prediction result. The ranking confidence is constituted by two evaluations, the overall structure (pTM) and an evaluation of the dimeric interface (ipTM), emphasizing the interface evaluation as represented by the following formula[9]:

$$\text{ranking confidence} = 0.8 \times \text{ipTM} + 0.2 \times \text{pTM}$$

These 3 values, ranking confidence, ipTM and pTM, for each prediction pairs were visualized in the separate heatmaps (Fig. 1A, Supplemental Table S1). In general, ranking confidence and ipTM values showed similar trends although a well-structured protein (e.g. Spn-E) tended to have a higher pTM value, which slightly elevated the ranking confidence. Based on this, in this study, we used the ranking confidence as an indicator of the protein-protein interaction. Each heterodimeric pair was calculated twice in the pairwise screening (e.g. proteins A_B and B_A), and the ranking confidences were plotted (Fig. 1B). The results showed that there was significant variance in the pairs with lower ranking confidences, while pairs with ranking confidences above 0.6 had relatively higher reproducibility. Consequently, we set a threshold of 0.6 and considered protein pairs with ranking confidences above 0.6 as likely complex-forming candidates. This approach identified 13 pairs; seven of these were already known to form complexes, confirming the effectiveness of AlphaFold2 in predicting complex formations (Table 2). The highest ranking confidence pair was the Zuc homodimer, possibly because AlphaFold2

had learned from Zuc homodimer's crystal structure registered in the database[19]. The structures of the 20 proteins used in this study have been analyzed to varying extents in previous studies (Supplemental Table S2). A complex of Vas and the Lotus domain of Osk has been reported [20], and based on this complex structure, the interaction between Vas and Tej Lotus domain was predicted with a high score. Although the conformational analyses of the RNA helicase domain and the eTud domain have been reported previously, many of those cover only a subset of the regions and unlikely to affect our predictions in this study.

The predicted 3D structures and the Predicted Aligned Error (PAE) plots for the 12 pairs, are shown in Fig. 1C. Consistent with a previous report using silkworm *Bombyx mori*[21], both Argonaute 3 (AGO3) and Aub, members of PIWI-family proteins sharing 50%-60% amino acid sequence similarity, were predicted to form dimers with Maelstrom (Mael) (Fig. 1C-i, ii, Table 2). AGO3 and Aub appeared well-folded protein except for their N-terminal flexible regions. In contrast, Mael protein was divided into three parts: N-terminal HMG domain, middle MAEL domain, and C-terminal disordered region[22] (Fig. 1C-i, ii). AlphaFold2 predicted the MAEL domain interacted with AGO3 and Aub.

Me31B, Tral, and Cup are recognized as RNA regulators localized to the nuage and/or sponge body, though they are not directly involved in the piRNA pathway. Previous studies have indicated that these proteins form complexes[13,23,24]. Me31B is a well-conserved RNA

helicase and showed the tightly-folded structure composed of two concatenated RecA helicase domains[25]. On the other hand, Tral and Cup were predicted largely disordered structure with some secondary structures (Fig. 1C-iii, iv). The predicted dimer structures of Me31B_Tral and Cup_Me31B showed scores of 0.74 and 0.68, respectively. (Table 2). Consistent with the previous study[24], AlphaFold2 predicted that the FDF motif of Tral, which contains a Phe-Asp-Phe sequence folded into two a-helixes from residue 405 to 537, was associated with Me31B. (Fig. 1C-iii). In addition, an α-helix and loop regions of Cup were predicted to make a contact with Me31B (Fig. 1C-iv). BoYb and Vret, both are eTud domain containing proteins[26] and their direct interaction has been suggested by the high retrieval rate for BoYb in the immunoprecipitant of Vret from the ovary[27]. The predicted structure revealed that both BoYb and Vret proteins consist of two domains, one at the N-terminal and the other at the C-terminal, connected by a flexible region. (Fig. 1C-v). Interactions were predicted between their N-terminal domains and between C-terminal domains, respectively. It has been reported that Tej, known as Tdrd5 in mammal, binds directly to Vas through its N-terminal Lotus domain[20] (Fig. 1C-vi) and to Spn-E through its loop region continuing the eTud domain[17] (Fig. 1C-vii). The predicted structures of Tej_Vas and Spn-E_Tej were consistent to their binding properties reported previously.

The remaining five pairs, previously unreported as directly interacting, were

considered novel binding pairs (Table 2, Fig. 1C-viii-xii). These interactions were experimentally examined using *Drosophila* S2 culture cells derived from embryonic somatic cells that lack germline-specific proteins. Previously, Squ was co-immunoprecipitated with Spn-E along with other nuage components from ovarian lysate[28], but whether this interaction was direct had not been examined. Co-immunoprecipitation assay in S2 cells, Myc-Spn-E was strongly detected in the precipitant of Flag-Squ by Western blotting, possibly supporting the direct interaction between Spn-E and Squ in the S2 cells devoid of germline proteins (Fig. 1D-i). Similarly, AlphaFold2 predicted a direct interaction between Aub and Vret, which was corroborated by co-immunoprecipitation assays (Fig. 1D-ii). The binding capabilities of another pair, BoYb-Shutdown (Shu), were also confirmed in S2 cells (Fig. 1D-iv). Three out of five candidate pairs confirmed interactions, validating the effectiveness of AlphaFold2 in identifying the binding partners. However, BoYb-Spn-E and Me31B-Vret did not show interaction in these assays (Fig. 1D-iii, v), possibly suggesting weak interactions that co-immunoprecipitation may have failed to detect. While co-immunoprecipitation is a widely used method, it may not always detect weak or transient interactions. Other validation methods, such as FRET or co-localization assay in culture cells, could offer further insights to support the results. It is also important to note that AlphaFold2's predictions are not definitive and may lead to false positives, particularly when analyzing a large number of interactions.

Evaluation of Spn-E and Squ interaction in culture cells and ovaries

Among the binding candidates, we focused on the predicted dimer structure of Spn-E and Squ pair. Spn-E is an evolutionarily conserved RNA helicase which is expressed in germline cells. It plays a crucial role in the piRNA production and transposon suppression in germline cells[28,29]. Similarly, Squ is also expressed in ovary and testis and involved in the piRNA production, although its molecular role is less defined[29,30]. While *squ* is conserved across *Drosophila* species (Supplemental Fig. 2A, B), vertebrate orthologs remain unidentified. Spn-E contains four domains: DEAD/DEAH helicase, Hel-C, HA2, and eTud domains (Fig. 2A). Its predicted 3D structure was well folded and contained few flexible regions (Fig. 1C-viii). In contrast, Squ was predicted to be largely disordered, consisting of three α-helices and two β-strands (Fig. 2A). The middle parts of Squ were in close contact with Spn-E, showing lower PAE values, suggestive of their interaction (Fig. 1C-viii, 2A). AlphaFold2 predicts the five structure models for each query using different initial model parameters (models 1-5) and ranking confidence is given to each model. As for Spn-E_Squ pair, the ranking confidence scores were ranging from 0.74 to 0.77. The 3D structures of Spn-E were very similar across all five models, superimposing almost perfectly (Fig. 2B). The middle region of Squ was consistently positioned relative to Spn-E, although the N- and C-terminal regions of Squ remained flexible (Fig. 2B).

The closer examination of the Spn-E_Squ dimer interface revealed a short α-helix of Squ (106th-116th residues) fitted into a groove on the Spn-E surface, while the anti-parallel β-sheet (140th-153rd) was also predicted to interact with Spn-E (Fig. 2A, C). Physico-chemical structural analysis using PDBePISA server (EMBL-EBI) identified salt bridges between Spn-E and Squ (Supplemental Table S3, S4)[31]. To validate these predicted interactions, we generated Squ mutants substituting each residue involved in the four salt bridges (E107, E109, R115, and K163) with alanine (Fig. 2D, Supplemental Fig. S2B) and assessed their interactions by co-immunoprecipitation in S2 cells expressing tagged proteins, Myc-Spn-E and Flag-Squ. The assay revealed that while the E107A single mutation did not affect the interaction, other single mutations mildly reduced the binding affinity of Squ to Spn-E (Supplemental Fig. S3A), Furthermore, the localization of GFP-tagged Squ and mKate2 (mK2)-tagged Spn-E were examined in S2 cells. When only Squ was expressed, it was dispersed in cytosol (Supplemental Fig. S3B). On the other hand, when only Spn-E was expressed, it localized in the nucleus as reported previously[17]. In the co-expression of Squ wildtype or single mutants, Spn-E was moved to the cytoplasm and form granules together with Squ, suggesting the interaction between them. Although the single mutants still could bind to Spn-E, Squ quadruple mutant (Squ^{4A}) completely lost the binding (Fig. 2E) and did not show the co-localization with Spn-E in S2 cells (Supplemental Fig. S3B). These results suggest that the salt bridges are important for the

interaction between Spn-E and Squ and support the accuracy of their dimer structure predicted by AlphaFold2.

While the RNA binding site of Spn-E has not been extensively studied, it is presumed to be near the helicase domain, similar to the Vas helicase-RNA complex[32]. In addition, Lin *et al* demonstrated that Hel-C domain of Spn-E interacted with the Tej's eSRS region, which recruits Spn-E to nuage[17], a site distinct from the predicted Squ binding sites (Fig. 2A). Interestingly, a tetramer complex of Spn-E_Squ_Tej_RNA predicted by the recently available AlphaFold3[33] placed the single strand RNA (ssRNA) near Spn-E's helicase domain (Fig. 2F), aligning with the ssRNA binding position found in Vas (Supplemental Fig. S3C). The predicted tetramer model suggests that Squ binding to Spn-E does not inhibit but may potentially regulate Spn-E's interaction with Tej or RNA by stabilizing the domain orientation of Spn-E (Fig. 2F).

In addition to the Spn-E_Squ_Tej complex, 1:1 dimer prediction described above further suggested potential trimers (Fig. 1; Supplemental Fig. S4). For example, Tej protein is predicted to bind both Vas and Spn-E, and AlfaFold3 indeed further predicted a Vas_Tej_Spn-E trimer, where Tej's Lotus and eTud domains interact with Vas and Spn-E, respectively. However, Lin *et al*. reported that Tej binds exclusively either with Vas or Spn-E, but not simultaneously, in *Drosophila* ovary[17], suggesting that the predicted trimers may be weak or transient. Similarly, the BoYb_Vret_Shu and the Me31B_Cup_Tral trimers remain hypothetical and require

experimental verification (Supplemental Fig. S4).

We investigated whether Spn-E also interacts with Squ within the *Drosophila* ovary. The antibody against Squ detected a specific band at the expected size by Western blotting in the heterozygous control ovarian lysate, which was absent in the transheterozygote mutant, $squ^{PP32/HE47}$ (Fig. 3A)[30]. Consistent with the previous report conducted with the transgenic line expressing HA-Squ[30], immunostaining of ovaries revealed the Squ's localization in nuage, which overlaps with endogenously-tagged Spn-E with mK2 (Fig. 3B). Spn-E was co-immunoprecipitated together with Squ from ovarian lysate, indicating the interaction between Squ and Spn-E (Fig. 3C). While the previous mass spectrometry analysis detected PIWI family proteins, Piwi, Aub, and AGO3, in Spn-E immunoprecipitates[28], these three proteins were not present in the immunoprecipitant of Squ (Fig. 3C), further supporting the direct interaction between Squ and Spn-E.

In this study, three novel protein-protein interactions were predicted and experimentally confirmed. AlphaFold2 also predicted the 3D structure of these complexes, providing insight into the important regions involved in complex formation. These predictions will provide fundamental information to elucidate nuage assembly. Nuage is thought to form by liquid-phase separation; however, direct protein-protein interactions likely occur within protein-dense nuage, facilitating RNA processing. Although the precise roles of individual interactions require further

study, characterization of protein-protein interactions within nuage will help clarify the mechanism of piRNA production.

Screening oogenesis-related proteins for interaction with nuage proteins.

Given the role of nuage for piRNA biogenesis and germline development, interactions between nuage-localized proteins and those involved in oogenesis were expected. We employed AlphaFold2 to predict these interactions using Vas, Squ, and Tej, the representative nuage components yet remain elusive, as baits. Of 430 proteins in oogenesis pathway[34], dimeric binding of 1,290 pairs was predicted (Supplemental Table S5), with 18 pairs showing dimer structures scoring above 0.6 (Table 3). Among those, co-immunoprecipitation in S2 cells confirmed interactions of three pairs, Mei-W68_Squ, CSN3_Squ, and Pka-C1_Tej (Fig. 4A, B, Table 3). The Mei-W68_Squ dimer, scoring 0.63, the binding site of Squ to MeiW68 was predicted at α-helixes in its middle region, which overlaps with the interacting site to Spn-E (Table 3, Fig. 4A-i, compare with Fig.1C-viii). Mei-W68 is a topoisomerase, known as Spo11 in many organisms, which is required for the formation of double strand breaks during meiosis[35]. Interestingly, Squ also plays a role in DNA damage response pathway and showed the genetic interaction with *chk2*, a meiotic checkpoint gene[30]. These results suggest that the binding of Squ to Mei-W68 may regulate the enzymatic activity of Mei-W68 in order to suppress the excessive

formation of double-strand breaks. Another confirmed pair was CSN3_Squ pair scoring 0.62 (Fig. 4A-ii, B-ii). CSN3, a component of COP9 signalosome which removes Nedd8 modifications from target proteins, is required for the self-renewal of the germline stem cells[36]. Pka-C1, a cAMP-dependent protein kinase involved in axis specification, rhythmic behavior and synaptic transmission[37] and predicted to bind with the N-terminal Lotus domain of Tej (Score 0.64, Fig. 4A-iii, B-iii), which is also known as binding site to Vas[20]. This suggests a potential competitive interaction between Pka-C1 and Vas for Tej. Although the success rate of confirmed interactions was low (3 out of 18) (Table 3, Supplemental Fig. S5), the results indicate that these protein pairs could interact within cells if co-expressed *in vivo*. The ranking confidence score reflects the reliability of AlphaFold2's predicted structure but does not always ensure accuracy. Therefore, we assessed complex affinity based on the predicted three-dimensional structures (Supplemental Table S6). Most dimers with high ranking confidence scores exhibited low Kd values indicative of high affinity, while some showed high Kd values indicating weak interactions (Supplemental Table S6). For example, the Baf_Vas complex had a high AlphaFold2 ranking confidence score (0.85) but a relatively high Kd value (1.1E-4 M), indicating low affinity. Consistently, Baf_Vas binding was not detected in Co-IP experiments (Supplemental Fig. S5C). Although accurate Kd prediction may be limited due to insufficient structural optimization, it could serve as a valuable secondary screening tool following

AlphaFold2 predictions.

Screening all *Drosophila* proteins for Piwi-interacting proteins

Given the crucial role of Piwi in piRNA biogenesis, heterochromatin formation, and germline stem cell (GSC) maintenance, we employed AlfaFold2 to screen all proteins in *Drosophila melanogaster* for potential Piwi interactions. Piwi, the founder member of the PIWI family proteins, is not only essential for binding piRNAs and regulating complementary mRNAs but also plays a critical role in GSC self-renewal[39]. Studies have shown that Piwi, lacking the N-terminal moiety containing the nuclear localization signal (NLS), still retains GSC self-renewal capabilities. Its function in GSC self-renewal is realized independently in the cytoplasm of GSC niche cells, separate from its role in transposon repression. The crystal structures of *Drosophila* Piwi and silkworm Siwi have been solved and revealed the organization of four domains (N, PAZ, MID, and PIWI)[40,41]. Recently, the ternary structure of piRNA, target RNA, and MILI, a mouse ortholog of Piwi, has been reported and the bound piRNA threaded through the channel between N-PAZ and MID–PIWI lobes (Supplemental Fig. S6A)[42].

To identify novel Piwi-binding proteins, we conducted a 1:1 interaction screening involving approximately 12,000 *Drosophila* proteins, excluding any proteins over 2,000 amino acid residues due to the computational limits. The ranking confidences by AlphaFold2 were

primarily low, with over 98% being below 0.6, suggesting a low likelihood of interaction between Piwi and the vast majority of the proteins (Fig. 5A). Approximately 1.5% of the pairs, totaling 164 pairs, scored above 0.6, was expected to contain the novel binding partners (Supplemental Table S7). Top 24 candidates with greater than 0.75 ranking confidence were listed in Table 4. This list contained many metabolic enzymes and three piRNA-related proteins, Asterix (Arx), Mael, and Hen1. The interactions between Mael and Piwi-family proteins have been already reported[21]. Arx, known as Gtsf1 in mammals and integral to Piwi–piRISC-mediated transcriptional silencing in nucleus[43], had high ranking confidences (0.83, Table 4). Despite its known three-dimensional structure determined by NMR spectroscopy[44], the Arx_Piwi complex structure remained elusive. AlphaFoldF2 predicted that while Arx lacked a compact domain, the majority of Arx protein associated around the PIWI domain, except for the flexible C-terminal region (130th-167th residues) (Fig. 5B-i). Three Arx paralogs in *Drosophila* (CG34283, CG32625, and CG14036) were also predicted to bind to Piwi with high ranking confidences, suggesting their interactions within the cells (Supplemental Fig. S6B). Although CG34283 is not expressed, CG32625 and CG14036 are moderately and highly expressed in ovary, respectively[37]. However, unlike *arx*, knockdown of each paralogous gene did not result in de-repression of a transposon, *mdg1*[43], suggesting that they may be pseudogenes or possess redundant roles.

Hen1 is a methyltransferase known to mediate methylation of the terminal 2' hydroxyl group of small interfering RNAs and piRNAs, thereby enhancing the stability of the small RNAs. Consistent with the previous report showing Hen1 binding to Piwi[43], the dimer structure of Hen1_Piwi was predicted with high ranking confidence, 0.77. This prediction further suggests that Hen1 is recruited to Piwi, thereby positioning it closer to the piRNA substrate (Fig. 5B-ii). Another potential interacting protein for Piwi was CG33703, a protein whose functions remains uncharacterized despite having 75 paralogs listed in *Drosophila* genome[37]. Together with three of these paralogs (CG33783, CG33647, and CG33644), CG33703 was predicted to form dimer with Piwi (ranking confidences 0.82) (Table 4, Supplemental Fig. S6C). The domain of unknown function, DUF1091[45], shared by these paralogs was predicted to associate with the PIWI-domain (Fig. 5B-iii). Although these proteins are generally not expressed under the normal conditions[37], their potential to bind Piwi suggests a regulatory role in the abnormal or stress conditions where CG33703 or its paralogs are expressed. In addition, we investigated two oogenesis-related proteins, Twinfilin (Twf, ranking confidence 0.64, Fig. 5B-iv) and Brainiac (Brn, ranking confidence 0.63, Fig. 5B-v), for their binding with Piwi through co-immunoprecipitation (Fig. 5C, Supplemental Table S7). While no binding was observed with Twf, significant binding was detected with Brn, which is involved in dorsal-ventral polarity determination in follicle cells[46].

This study identifies several potential protein interactions, but AlphaFold2 predictions require caution. Protein-protein interactions involve conformational changes and dependencies on ligands, ions, and cofactors, which AlphaFold2 does not consider, potentially reducing prediction accuracy. Notably, the presence of a high-scoring model in terms of structural complementarity does not guarantee that the interaction is biologically significant. The expression patterns of these candidate proteins within the organism are crucial for further validation of our findings. It is likely that these proteins interact when co-expressed in the same cellular context. Under typical growth conditions, these interactions might not occur; however, in stress or disease states where these proteins are upregulated, the likelihood of interaction increases, potentially implicating these interactions in the disruption of normal cellular functions and contributing to disease or tumorigenesis. Furthermore, in silico screening proves extremely valuable, especially when dealing with toxic bait proteins, as it allows us to narrow down the list of potential candidates and reduce the need for hazardous experimental procedures. Ultimately, establishing these potential interactions in vivo could significantly advance our understanding of protein functions under both normal and pathological conditions.

Materials and Methods

Antibodies

The anti-Squ antibody was generated as follows. His-tagged full-length Squ was expressed in *Escherichia coli* BL21(DE3) strain, with the plasmid that subcloned the *squ* coding region into pDEST17 vector (Thermo Fisher Scientific). His-Squ was solubilized with 6 M Urea in PBS, purified using Nickel Sepharose beads (GE healthcare) following the manufacturer's protocol, and subsequently used for immunization in rats. The antibodies used for Western blotting analysis were rat anti-Spn-E[17] (1:500), rat anti-Ago3[17] (1:200), guinea pig anti-Aub[47] (1:1000), mouse monoclonal anti-Piwi (G-1, sc-390946, Santa Cruz Biotechnology (United States)), and mouse monoclonal anti-α-Tubulin (DM1A, sc-32293, Santa Cruz Biotechnology). The secondary antibodies used in this study were HRP-conjugated goat anti-guinea pig (Dako, Cat.# P0141), HRP-conjugated goat anti-rat (Dako, Cat.# P0450), HRP-conjugated goat anti-mouse (BioRad, Cat.# 1706516) and HRP-conjugated goat anti-rabbit (BioRad, Cat.# 1706515). HRP-conjugated anti-DDDDK-tag antibody (MBL, Cat.#M185-7) and HRP-conjugated anti-Myc-tag antibody (MBL, Cat.#M192-7) were used to detect FLAG-tagged and Myc-tagged proteins, respectively.

AlphaFold2 prediction for the direct interacting protein pairs

Amino acid sequences for *Drosophila* proteins were obtained from Flybase[37]. For proteins annotated with multiple isoforms, only the longest isoform was selected. Proteins exceeding 2,000 residues were excluded due to computational limitations. AlphaFold v2.2 program was installed in the Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) at Cyber Media Center in Osaka University. All necessary protein sequence databases for AlphaFold2 were stored on an SSD device connected to the SQUID system.

The AlphaFold2 prediction process was divided into two steps: generation of the multiple sequence alignment (MSA) and the prediction of the 3D structure. The MSAs were computed on SQUID's CPU node and stored for reuse. The calculation of the MSA took on average 2-4 hours per protein, with the more homologs of the protein in query, the longer it took. For dimer structure prediction, two MSAs corresponding to the dimer pair were placed in the directory of msas/A and msas/B. The calculations were performed on the GPU node with the options of -t 2022-05-14 -m multimer -l 1 -p true. AlphaFold2 generates five structural models for each prediction. To speed up the prediction, five computations were assigned to five GPU units, even though the original AlphaFold2 program computes 5 models one at a time. Prediction of dimer structure took approximately 1-2 hours per pair on average, depending on protein size. Each user can compute 100~200 pairs of calculations per day, but since the

supercomputer is shared, job availability varies with overall demand.

The prediction confidence score (ranking confidence) was provided for each model and among 5 models, the highest ranking confidence was used as the prediction score for the corresponding dimer structure. PAE plots for dimer structures were drawn by extracting the data form pkl files generated by AlphaFold2. The list of protein pairs scoring above 0.6 and the corresponding PAE plots and PDB structures are available on Github (https://dme-research.github.io/AF2_2/).

AlphaFold3 prediction for the structure of the trimer complex

The structure of Spn-E_Squ_Tej complexed with RNA, 5'-CUGACUACCGAAGUACUACG-3', was predicted by the AlphaFold3 prediction server (https://alphafoldserver.com/)[33]. The trimer structures of Spn-E_Squ_Tej, Vas_Tej_Spn-E, BoYb_Vret_Shu, and Me31B_Cup_Tral were also predicted by AlphaFold3.

Analysis of protein 3D structure

The protein 3D structure was visualized using ChimeraX software[48]. The SpnE_Squ dimer interface was analyzed with the 'Protein interfaces, surfaces and assemblies' service (PISA) at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)[31].

Fly stocks

All stocks were maintained at 25□ with standard methods. Mutant alleles of *squ* (*squ*^{pp32} and *squ*^{HE47}) were used in this study[30]. The mK2-tagged Spn-E-mK2 knock-in fly was previously generated[17]. *y w* strain served as the control.

Western blotting

Ovaries were homogenized in the ice-cold PBS and denatured in the presence of SDS sample buffer at 95°C for 5 min. The samples were then subjected to SDS-PAGE and transferred to ClearTrans SP PVDF membrane (Wako). The primary and secondary antibodies described above were diluted in the Signal Enhancer reagent HIKARI (Nacalai Tesque). Chemiluminescence was induced by the Chemi-Lumi One reagent kit (Nacalai Tesque) and detected with ChemiDoc Touch (Bio-Rad). The bands were quantified using ImageJ[49] or Image Lab software (Bio-Rad).

Co-immunoprecipitation in S2 cells

Drosophila Schneider S2 cells were cultured at 28°C in Schneider's medium supplemented with 10% (v/v) fetal bovine serum and antibiotics (penicillin and streptomycin). Protein coding

regions were cloned into pENTR vector (Thermo Fisher Scientific) and then transferred into pAFW or pAMW destination vectors. S2 cells (0.2-2x10^6 cells/ml) were seeded in 12-well plates overnight and transfected using Hilymax (Dojindo Molecular Technologies, Japan). After 36-48 hours, S2 cells were resuspended in 360 μl of ice-cold PBS containing 0.02% Triton-X100 and 1x protease inhibitor cocktail (Roche), and sonicated (0.5 sec, 5 times). The resulted lysate was clarified by spinning at 15,000 xg for 15 min at 4°C. 300 μl of supernatant was incubated with 6 μl of prewashed anti-FLAG magnetic beads (MBL) or anti-Myc magnetic beads (Thermo Fisher Scientific) for 1.5 h at 4°C with gentle rotation. After incubation, the beads were washed three times with 800 μl of ice-cold PBS with 0.02% Triton-X100, denatured in SDS sample buffer and subjected to SDS-PAGE and Western blot. 1% of the total lysates were loaded as input samples.

Co-localization assay in S2 cells

Construction of GFP-tagged or mKate2-tagged proteins and transfection were conducted as described in the previous section. After 48 h of transfection, the cells were placed onto the concanavalin A-coated coverslips for 20 min, fixed with PBS containing 4% (w/v) paraformaldehyde for 15 min at room temperature, permeabilized with PBX [PBS containing 0.2% (v/v) TritonX-100] for 10 min twice, stained with DAPI (1:1000) and mounted with

Fluoro-Keeper Antifade Reagent (Nacalai Tesque). Images were taken by ZEISS LSM 900 with Airy Scan 2 using 63x oil NA 6.0 objectives and processed using ZEISS ZEN 3.0 and ImageJ[49].

Crosslinking immunoprecipitation (CL-IP)

As previously described[17], 100 ovaries from *y w* flies were dissected in ice-cold PBS and fixed in PBS containing 0.1% (w/v) paraformaldehyde for 20 min on ice, quenched in 125 mM glycine for 20 min, and then homogenized in CL-IP lysis buffer. The lysate was incubated at 4°C for 20 min and then sonicated. After centrifugation at maximum speed for 10 min at 4°C, the supernatant was collected and diluted with an equal volume of CL-IP wash buffer. 10 μl of pre-washed Dynabeads Protein G/A mixture (1:1) (Invitrogen) was added for pre-clearance at 4°C for 1 h. Anti-Squ antibody was added to the cleared supernatant with 1:500 dilution and incubated at 4°C overnight. The 20 μl of pre-washed Dynabeads Protein G/A 1:1 mixture beads (Invitrogen) were added for binding and incubated at 4°C for 3 h. After washed with CL-IP wash buffer for 3 times, beads were collected and 50 μl of CL-IP wash buffer containing SDS sample buffer was added. The beads were boiled at 95°C for 5 min and subjected for SDS-PAGE and Western blotting analysis.

Immunostaining of ovaries

As previously described[17,47], ovaries were dissected, fixed, permeabilized with PBX and immunostained. The primary and the secondary antibodies were anti-Squ antibody (in this study, 1:500) and Alexa Fluor 488-conjugated anti-rat IgG (Thermo Fisher Scientific, 1:200). Images were taken by ZEISS LSM 900 with Airy Scan 2 using 63X oil NA 1.4 objectives and processed by ZEISS ZEN 3.0 and ImageJ[49].

References

1. Galperin, M. Y. *et al.* COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2020).
2. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604–609 (2000).
3. Yu, I. *et al.* Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* **5**, e19274 (2016).
4. Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
5. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
6. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of Protein−Protein Interactions: What are the Preferred Ways For Proteins To Interact? *Chem. Rev.* **108**, 1225–1244 (2008).
7. Burley, S. K. *et al.* RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **51**, D488–D508 (2023).
8. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
9. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034 Preprint at https://doi.org/10.1101/2021.10.04.463034 (2021).
10. Humphreys, I. R. *et al.* Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).
11. Pek, J. W., Patil, V. S. & Kai, T. piRNA pathway and the potential processing site, the nuage, in the Drosophila germline. *Dev. Growth Differ.* **54**, 66–77 (2012).
12. Ross, R. J., Weiner, M. M. & Lin, H. PIWI proteins and PIWI–interacting RNAs in the soma. *Nature* **505**, 353–359 (2014).
13. McCambridge, A. *et al.* Comparative Proteomics Reveal Me31B's Interactome Dynamics, Expression Regulation, and Assembly Mechanism into Germ Granules during Drosophila Germline Development. *Sci. Rep.* **10**, 564 (2020).
14. Lim, A. K. & Kai, T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6714–6719 (2007).
15. Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing

structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).

16. Ren, R. *et al.* Structure and domain organization of Drosophila Tudor. *Cell Res.* **24**, 1146–1149 (2014).

17. Lin, Y., Suyama, R., Kawaguchi, S., Iki, T. & Kai, T. Tejas functions as a core component in nuage assembly and precursor processing in Drosophila piRNA biogenesis. *J. Cell Biol.* **222**, e202303125 (2023).

18. Hu, Y. *et al.* Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. *Nucleic Acids Res.* **46**, D567–D574 (2018).

19. Nishimasu, H. *et al.* Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284–287 (2012).

20. Jeske, M., Müller, C. W. & Ephrussi, A. The LOTUS domain is a conserved DEAD-box RNA helicase regulator essential for the recruitment of Vasa to the germ plasm and nuage. *Genes Dev.* **31**, 939–952 (2017).

21. Namba, Y. *et al.* Maelstrom functions in the production of Siwi-piRISC capable of regulating transposons in *Bombyx* germ cells. *iScience* **25**, 103914 (2022).

22. Matsumoto, N. *et al.* Crystal Structure and Activity of the Endoribonuclease Domain of the piRNA Pathway Factor Maelstrom. *Cell Rep.* **11**, 366–375 (2015).

23. Tritschler, F. *et al.* Structural Basis for the Mutually Exclusive Anchoring of P Body Components EDC3 and Tral to the DEAD Box Protein DDX6/Me31B. *Mol. Cell* **33**, 661–668 (2009).

24. Tritschler, F. *et al.* Similar Modes of Interaction Enable Trailer Hitch and EDC3 To Associate with DCP1 and Me31B in Distinct Protein Complexes. *Mol. Cell. Biol.* **28**, 6695–6708 (2008).

25. Peter, D. *et al.* Molecular basis for GIGYF–Me31B complex assembly in 4EHP-mediated translational repression. *Genes Dev.* **33**, 1355–1360 (2019).

26. Chen, C., Nott, T. J., Jin, J. & Pawson, T. Deciphering arginine methylation: Tudor tells the tale. *Nat. Rev. Mol. Cell Biol.* **12**, 629–642 (2011).

27. Handler, D. *et al.* A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30**, 3977–3993 (2011).

28. Andress, A. *et al.* Spindle-E cycling between nuage and cytoplasm is controlled by Qin and PIWI proteins. *J. Cell Biol.* **213**, 201–211 (2016).

29. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A Transcriptome-wide RNAi Screen in the *Drosophila* Ovary Reveals Factors of the Germline piRNA Pathway. *Mol. Cell* **50**, 749–761 (2013).

30. Pane, A., Wehr, K. & Schüpbach, T. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev. Cell* **12**, 851–862 (2007).

31. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

32. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural Basis for RNA Unwinding by the DEAD-Box Protein *Drosophila* Vasa. *Cell* **125**, 287–300 (2006).

33. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

34. Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

35. McKim, K. S. & Hayashi-Hagihara, A. mei-W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* **12**, 2932–2942 (1998).

36. Pan, L. *et al.* Protein competition switches the function of COP9 from self-renewal to differentiation. *Nature* **514**, 233–236 (2014).

37. Öztürk-Çolak, A. *et al.* FlyBase: updates to the Drosophila genes and genomes database. *Genetics* **227**, iyad211 (2024).

38. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).

39. Klenov, M. S. *et al.* Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18760–18765 (2011).

40. Matsumoto, N. *et al.* Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. *Cell* **167**, 484-497.e9 (2016).

41. Yamaguchi, S. *et al.* Crystal structure of Drosophila Piwi. *Nat. Commun.* **11**, 858 (2020).

42. Li, Z. *et al.* Mammalian PIWI–piRNA–target complexes reveal features for broad and efficient target silencing. *Nat. Struct. Mol. Biol.* 1–10 (2024) doi:10.1038/s41594-024-01287-6.

43. Ohtani, H. *et al.* DmGTSF1 is necessary for Piwi–piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. http://genesdev.cshlp.org doi:10.1101/gad.221515.113.

44. Ipsaro, J. J., O'Brien, P. A., Bhattacharya, S., Palmer, A. G. & Joshua-Tor, L.

Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons. *Cell Rep.* **34**, 108914 (2021).

45. Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**, D458–D460 (2021).

46. Goode, S., Melnick, M., Chou, T.-B. & Perrimon, N. The neurogenic genes egghead and brainiac define a novel signaling pathway essential for epithelial morphogenesis during Drosophila oogenesis. *Development* **122**, 3863–3879 (1996).

47. Lim, L.-X., Isshiki, W., Iki, T., Kawaguchi, S. & Kai, T. The Tudor Domain-Containing Protein, Kotsubu (CG9925), Localizes to the Nuage and Functions in piRNA Biogenesis in D. melanogaster. *Front. Mol. Biosci.* **9**, 818302 (2022).

48. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci. Publ. Protein Soc.* **30**, 70–82 (2021).

49. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

50. Patil, V. S., Anand, A., Chakrabarti, A. & Kai, T. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. *BMC Biol.* **12**, 61 (2014).

51. Webster, A. *et al.* Aub and Ago3 are recruited to nuage through two mechanisms to form a ping-pong complex assembled by Krimper. *Mol. Cell* **59**, 564–575 (2015).

52. Liu, L., Qi, H., Wang, J. & Lin, H. PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Dev. Camb. Engl.* **138**, 1863–1873 (2011).

53. Johnstone, O. *et al.* Belle is a Drosophila DEAD-box protein required for viability and in the germ line. *Dev. Biol.* **277**, 92–101 (2005).

54. Nguyen, T. T. M. *et al.* In vivo profiling of the Zucchini proximal proteome in the Drosophila ovary. *Dev. Camb. Engl.* **150**, dev201220 (2023).

55. Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R. & Brennecke, J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. *Mol. Cell* **47**, 954–969 (2012).

56. Patil, V. S. & Kai, T. Repression of Retroelements in *Drosophila* Germline via piRNA Pathway by the Tudor Domain Protein Tejas. *Curr. Biol.* **20**, 724–730 (2010).

Acknowledge

The prediction by AlphaFold2 was achieved through the use of large-scale computer systems, Supercomputer for Quest to Unsolved Interdisciplinary Datascience (SQUID) at the Cybermedia Center, Osaka University through the Research Proposal-based Use, Large-Scale High-Performance Computing Projects to K.S. (Cyber media center, Osaka University) and the High Performance Computing Infrastructure (HPCI) System Research Project (Project ID: hp240099 to K.S.). We thank Dr Trudi Schüpbach (Princeton University) for generous gifts of squ mutant flies. We also thank the FBS Core Facility in Osaka University for providing access to the LSM 900 and ChemiDoc Touch. We appreciate the insightful discussion and suggestions from all the members of KT's laboratory.

Author contributions

Conceptualization: S.K.;

Methodology: S.K., X.X., S.T., D.S.;

Software: S.K., S.T., Y.K., K.R., S.R., D.S.;

Validation: S.K.;

Formal analysis: S.K., X.X.;

Investigation: S.K., X.X.;

Resources: S.K., T.K.;

Data curation: S.K.;

Writing - original draft: S.K., X.X., T.K.;

Visualization: S.K., X.X.;

Supervision: S.K., T.K.;

Project administration: S.K., T.K.;

Funding acquisition: S.K., T.K.

Data availability statement

PDB files and PAE plots for the protein dimers whose ranking confidences were more than 0.6, were deposited and available at GitHub Pages (https://dme-research.github.io/AF2_2/)

Competing Interests Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding Statement

This work was supported by TAKEDA Bioscience Research Grant [J191503009 to K.T.]; Grant-in-Aid for Transformative Research Areas (A) [21H05275 to K.T.]; and Osaka University Institute for Datability Science "Transdisciplinary Research Project" [Na22990007 to K.T. and K.S.].

Figure legends

Figure 1: The 1:1 dimer structure prediction by AlphaFold2 for piRNA-related proteins

(A) Heatmaps of the prediction confidence scores (ranking confidence, Green), pTM values (Blue), and ipTM values (Red) provided by AlphaFold2. The 20 types of proteins are aligned from top to bottom and left to right in the same order. Boxes on diagonal line represent homodimers.

(B) Scatter plot of the ranking confidences. The scores from first and second predictions for each heterodimer pair are plotted on X and Y axis, respectively.

(Ci ~ xii) The predicted 3D structures (top panels) and the Predicted Aligned Error (PAE) plots (bottom panels) for each candidate heterodimers scoring above 0.6. The PAE plot displays the positional errors between all amino acid residue pairs, formatted in a matrix layout.

(D) Co-immunoprecipitation assays using tagged proteins to verify interactions between

specific pairs: Spn-E_Squ (i), Aub_Vret (ii), Spn-E_BoYb (iii), BoYb_Shu (iv), and Me31B_Vret (v). Single transfected cells expressing only Myc-tagged but not Flag-tagged proteins are used as negative controls for each set. Box and whisker plots show the intensity ratio between immunoprecipitated and input bands (n=3).

Figure 2: Interaction between Spn-E and Squ

(A) Schematic of Spn-E domain structures defined in SMART[45]. Boxes (α-helix: orange) and arrow (β-sheet: green) for Squ structure. The predicted interacting regions between Spn-E and Squ are indicated in gray boxes. Tej interaction site of Spn-E is also shown[17].

(B) The predicted five models of heterodimer of Spn-E (in gray) and Squ (in magenta). Spn-E molecules in all five models are superimposed.

(C) 3D structure of the Spn-E_Squ dimer colored by Spn-E domains as indicated in (A), with Squ in magenta. The enlarged image of the interface indicated by box is also shown.

(D) The predicted salt bridges at the interface, with Spn-E in gray and Squ in magenta. The residues forming salt bridges are depicted in stick model.

(E) Co-immunoprecipitation assay using S2 cell lysate to examine the interaction between Myc-Spn-E and Flag-Squ mutant (4A) whose salt bridge-forming residues are mutated to Ala. S2 cells expressing Myc-Spn-E alone is used as a control. The ratios of the band intensity (IP/input) are shown in a box and whisker plot (n=3).

(F) The heterotetramer model of Spn-E_RNA_Squ_Tej predicted by AlphaFold3. Spn-E is shown as a space filled model in gray, Squ in magenta, Tej in cyan, and RNA in yellow. The model on the left is rotated 180º in the Y axis to produce the image on the right.

Figure 3: Spn-E and Squ interact in *Drosophila* ovary

(A) Western blotting analysis using anti-Squ antibody reveals a specific band at the expected size (approximately 28 kDa) for endogenous Squ in *Drosophila* ovarian lysates of the heterozygous control. This band is absent in the transheterozygoute, $squ^{PP32/HE47}$.

(B) Immunostaining of *Drosophila* egg chambers with anti-Squ antibody and anti-mKate2 (mK2) antibody demonstrates colocalization of Squ and Spn-E-mK2 in nuage, a perinuclear granule in germline cells. The enlarged images of nuclei are shown in the panels below. Scale bars: 10 μm (top row), 2.5 μm (enlarged images)

(C) Immunoprecipitation of the endogenous Squ from ovarian lysate revealed the interaction with Spn-E protein. Proteins were detected by western blotting analysis using the specific antibody for each protein. The negative control was performed without anti-Squ antibody (beads only).

Figure 4: Squ- and Tej-interacting proteins predicted by AlphaFold2

(A i-iii) The predicted dimer structures (top) and Predicted Aligned Error (PAE) plots (bottom) of Mei-W68 in blue and Squ in magenta (i), CSN3 in green and Squ in magenta (ii), Pka-C1 in orange and Tej in cyan (iii). The PAE plot displays the positional errors between all amino acid residue pairs, formatted in a matrix layout.

(B i-iii) Co-immunoprecipitation assays using tagged proteins to verify interactions between specific pairs: Mei-W68_Squ (i), CSN3_Squ (ii), and Pka-C1_Tej (iii). Single transfected cells expressing only Myc-tagged but not Flag-tagged proteins are used as negative controls for each set. Box and whisker plots show the intensity ratio between immunoprecipitated and input bands (n=3).

Figure 5: Screening for Piwi-interacting proteins in *Drosophila* proteome

(A) Pie chart displaying the distribution of ranking confidences from the AlphaFold2 screening for Piwi-interacting proteins among those encoded by *Drosophila* genome.

(Bi~ v) The predicted dimer structure (top) and PAE plots (bottom) for the Piwi and the binding candidates in red: Arx (i), Hen1 (ii), CG33703 (iii), Twf (iv), and Brn (v). Piwi is shown in the same colors as Supplemental Fig. S5A.

(C) Co-immunoprecipitation assays using tagged proteins to verify interactions between Piwi and the binding candidates, Twf and Brn. Single transfected cells expressing only Flag-Piwi is used as negative control. Box and whisker plots show the intensity ratio between immunoprecipitated and input bands (n=3).

Table 1 The piRNA production-related proteins used in this study

Protein	Ortholog	Number of residues	Domain	Direct binding (MIST databse)	Localization	Reference,
Vas	DDX4	661	DEAD-Box, Hel-C	Aub	Nuage	[14]
Spn-E	Tdrd9	1434	DEAD-Box, Hel-C, HA2, eTud		Nuage	[28]
Tej	Tdrd5	559	Lotus, eTud		Nuage	[17]
Tapas	Tdrd7	1222	Lotus, eTud		Nuage	[50]
Qin	Rnf17	1857	RING, eTud		Nuage	[28]
Kots	Tdrd1	892	eTud		Nuage	[47]
Krimp	-	746	eTud		Nuage	[14]
Squ	-	241			Nuage	[30]
Mael	Mael	462	HMG, MAEL		Nuage	[14]
Aub	PIWIL2	866	N, PAZ, PIWI, MID	Vas, Papi, Me31B	Nuage	[14]
AGO3	PIWIL4	867	N, PAZ, PIWI, MID	Papi	Nuage	[51]
Papi	Tdrkh	576	eTud, KH	Aub, AGO3	Mitochondria	[52]
Vret	Tdrd1	691	eTud	BoYb	Nuage	[27]
Bel	DDX3	801	DEAD-Box		Nuage	[53]
Zuc	Pld6	253	PLD-like	Zuc	Mitochondria	[54]
Cup	Eif4enif1	1117		Me31B	Nuage	[13]
Tral	Lsm14	657	Lsm, FDF	Me31B	Nuage	[13]
Me31B	DDX6	459	DEAD-Box	Aub, Cup, Tral	Nuage	[13]
Shu	Fkbp6	455	PPIase		Nuage	[55]
BoYb	Tdrd12	1059	DEAD-Box, eTud	Vret	Nuage	[27]

Table 2 The screening for the interacting proteins (prediction confidence score, ranking confidence > 0.6)

Protein A_B 1st prediction	ranking confidence	**Protein B_A 2nd prediction**	ranking confidence	**Reference**	**Validation by Co-IP**
Zuc_Zuc	0.85	N/A	N/A	[19]	N/A
AGO3_Mael	0.78	Mael_AGO3	0.78	[21]	N/A
Aub_Mael	0.78	Mael_Aub	0.78	[21]	N/A
Spn-E_Squ	0.77	Squ_Spn-E	0.78	This study	++
Me31B_Tral	0.74	Tral_Me31B	0.72	[13]	N/A
Aub_Vret	0.72	Vret_Aub	0.72	This study	+
BoYb_Spn-E	0.69	Spn-E_BoYb	0.69	This study	-
Cup_Me31B	0.68	Me31B_Cup	0.70	[13]	N/A
Spn-E_Tej	0.65	Tej_Spn-E	0.65	[17]	N/A
BoYb_Vret	0.64	Vret_BoYb	0.65	[27]	N/A
BoYb_Shu	0.64	Shu_BoYb	0.56	This study	+
Me31B_Vret	0.64	Vret_Me31B	0.45	This study	-
Tej_Vas	0.61	Vas_Tej	0.62	[56]	N/A

Table 3. The binding candidates predicted by AlphaFold2

Protein_A	Protein_B	AlphaFold2 ranking confidence	Validation by Co-IP	Function of Protein_A
Vps25	Squ	0.71	No	a member of the ESCRT-II complex
Nup44A	Squ	0.65	No	a nuclear pore protein
Nclb	Squ	0.64	No	Chromatin-associated factor
Mei-W68	Squ	0.63	Bound	formation of double strand breaks
DNaseII	Squ	0.63	N/E	Deoxyribonuclease II
Spn-D	Squ	0.62	No	homologous recombinational DNA repair
CSN3	Squ	0.62	Bound	subunit of the COP9 signalosome
Jagn	Tej	0.72	No	located in the endoplasmic reticulum
Pka-C1	Tej	0.64	Bound	serine/threonine kinase
Rab7	Tej	0.62	No	vesicle trafficking regulation
Baf	Vas	0.85	No	chromatin organization
Mats	Vas	0.79	No	Coactivator of Warts (Wts) kinase
Abo	Vas	0.68	No	negative regulator of histone transcription genes
CathD	Vas	0.67	N/E	apoptosis and the defense response
Rab11	Vas	0.67	No	endomembrane traffiking
Vls	Vas	0.63	No	substrate recognition platform for cusl
Hsc70-4	Vas	0.62	No	protein folding
RhoL	Vas	0.61	N/E	maturation of hemocytes

N/E: Not examined. The expression plasmids were not constructed due to the technical reasons.

Table 4. Piwi-interacting proteins predicted by AlphaFold2 (Socre >= 0.75).

protein	length (residue)	ranking confidence	human ortholog	Gene Summary (Flybase)
CG34283	153	0.85	GTSF1	-
CG32625	144	0.84	GTSF1	-
Arx	167	0.83	GTSF1	It plays an essential role in piRNA-guided transcriptional silencing, interacting probably directly with the product of piwi.
CG33703	181	0.82	-	No phenotypic data is available.
GstE12	223	0.82	GSTT2B	Glutathione S transferase E12 (GstE12) encodes an enzyme involved in glutathione metabolism.
CAH4	279	0.81	CA6	Predicted to enable carbonate dehydratase activity. Predicted to be active in cytoplasm.
CG13192	323	0.81	GNB1L	Predicted to be involved in social behavior.
Mael	462	0.79	MAEL	Involved both in the piRNA and miRNA metabolic processes.
Adk3	366	0.78	ADK	Predicted to enable adenosine kinase activity.
Alg11	475	0.78	ALG11	Predicted to enable GDP-Man:Man3GlcNAc2-PP-Dol alpha-1,2-mannosyltransferase activity.
CG41378	228	0.78	IFI30	Predicted to enable oxidoreductase activity.
CG14036	93	0.77	GTSF1	Involved in copper ion homeostasis.
CG7966	486	0.77	SELENBP1	Predicted to enable methanethiol oxidase activity.
Hen1	391	0.77	HENMT1	Hen1 encodes a methyltransferase that methylates the terminal 2' hydroxyl group of small interfering RNAs and Piwi-interacting RNAs.
Rpp14b	112	0.77	RPP14	Predicted to enable ribonuclease P RNA binding activity.
CG33783	164	0.76	-	No phenotypic data is available.
AANATL4	224	0.75	-	Predicted to enable aralkylamine N-acetyltransferase activity.
CG14787	260	0.75	CDYL2	Is expressed in adult heart; embryonic Malpighian tubule; and embryonic main segment of Malpighian tubule.
CG33160	258	0.75	PRSS1	Predicted to enable serine-type endopeptidase activity.
CG3397	342	0.75	AKR7A2	Predicted to enable D-arabinose 1-dehydrogenase [NAD(P)+] activity.
CG4390	330	0.75	ESD	Enables serine hydrolase activity.
CG7142	334	0.75	KLK1	Predicted to enable serine-type endopeptidase activity.
JanA	135	0.75	PHPT1	JanA and janB regulate somatic sex differentiation.
Yip7	270	0.75	CTRB1	Enables serine hydrolase activity.

Supplemental information

Supplemental Figure S1: The Nuage Proteins Analyzed in This Study

Supplemental Figure S2: Comparative Analysis of Squ and Spn-E Orthologs in *Drosophila*

Supplemental Figure S3: Interaction and Localization Analysis of Spn-E and Squ in S2 Cells

Supplemental Figure S4: Trimer structures predicted by AlphaFold3

Supplemental Figure S5: Validation of Predicted Protein Interactions via Co-Immunoprecipitation from S2 cell lysate

Supplemental Figure S6: Structural Analyses of Piwi Complexes and Interactions

Supplemental Table S1: The prediction confidence scores (ranking confidences) for the pairwise dimer predictions by AlphaFold2 as shown in Fig. 1A

Supplemental Table S2: The experimentally determined 3D structure models for the nuage-localizing/piRNA-related proteins.

Supplemental Table S3: The salt-bridges and H-bonds found in the predicted interface between Spn-E and Squ dimer

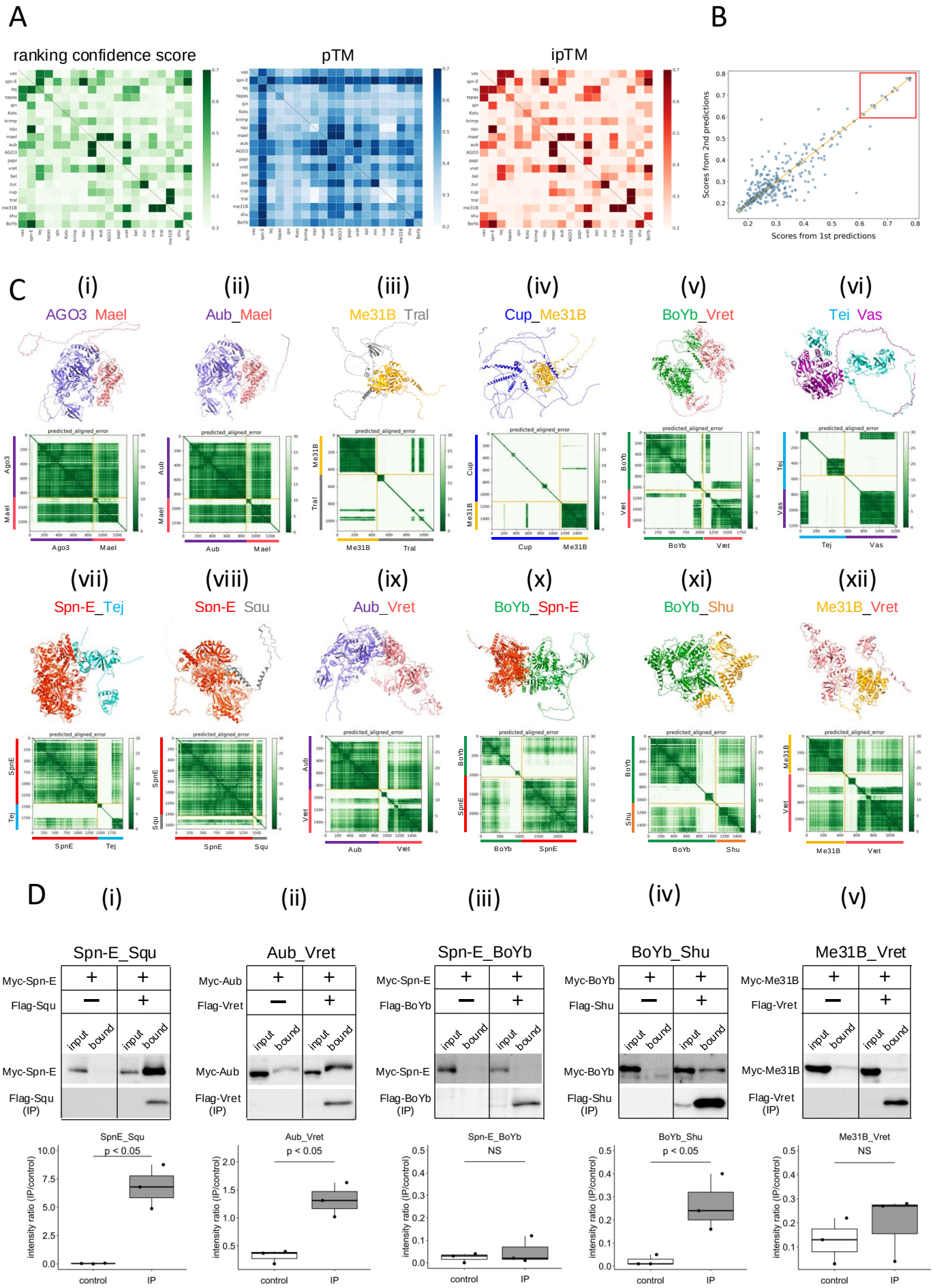
Supplemental Table S4: The hydrophobic residues found in the predicted interface between Spn-E and Squ proteins

Supplemental Table S5: The AlphaFold2 screening for the interacting pairs of Squ, Tej, Vas against 430 oogenesis-related proteins

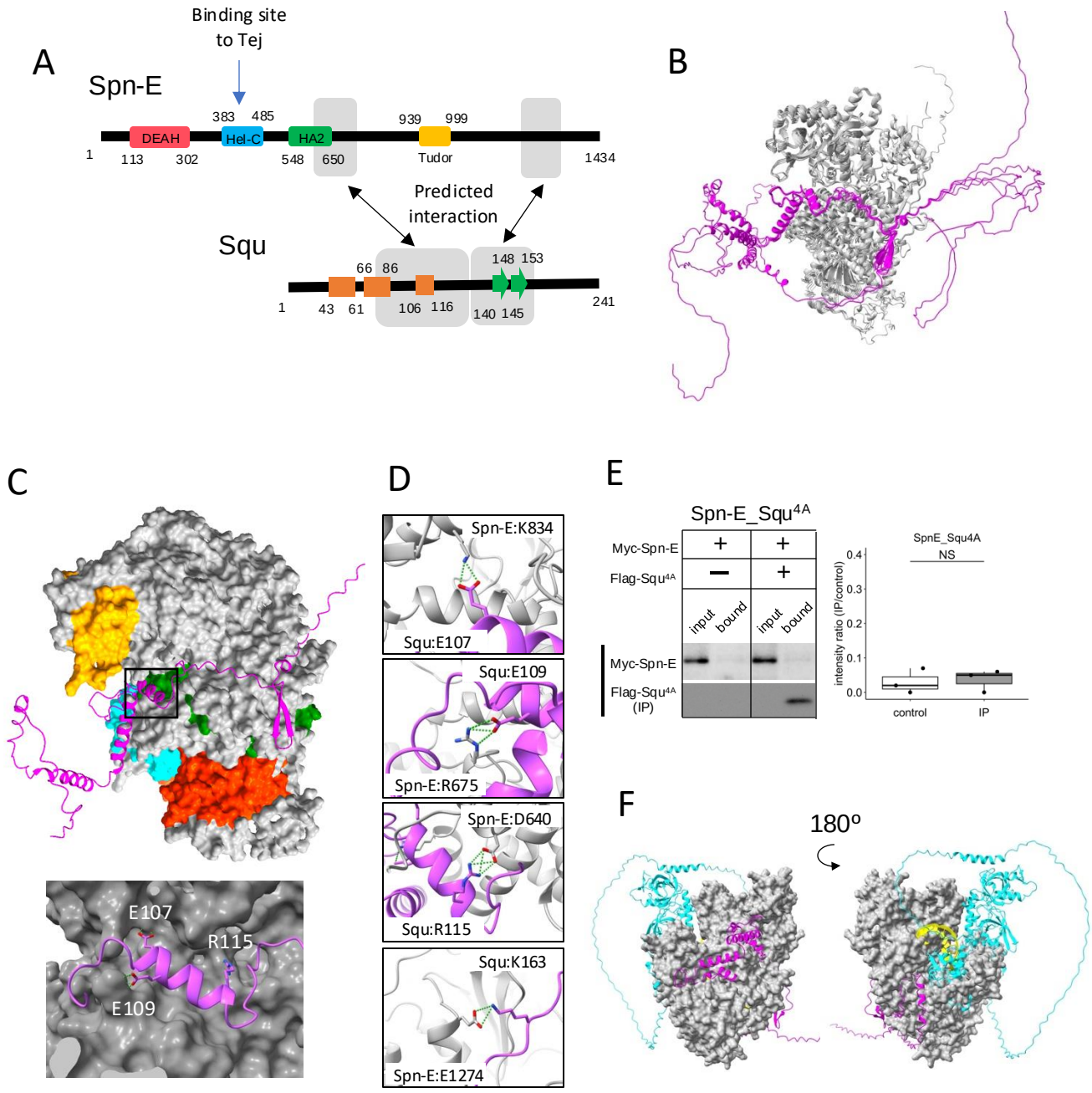
Supplemental Table S6: The computed binding affinity of the protein-protein complex on the basis of the three-dimensional structure predicted by AlphaFold2.

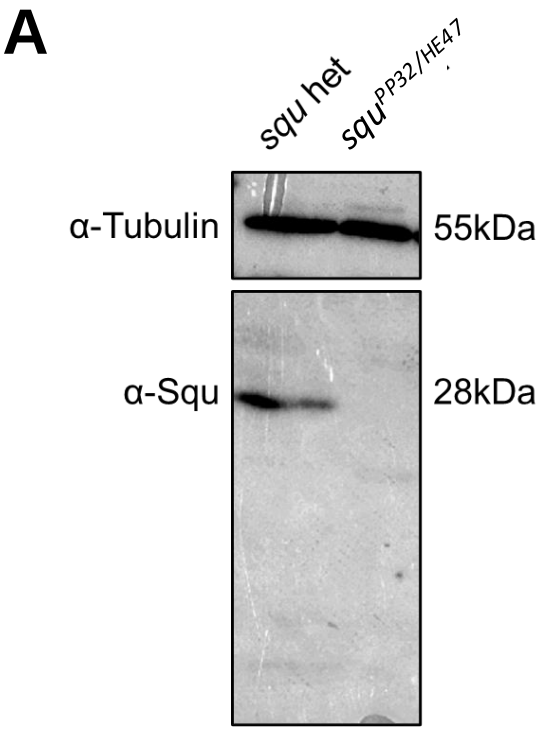
Supplemental Table S7: The AlphaFold2 screen for Piwi interacting pairs against 12,427 *Drosophila* proteins.

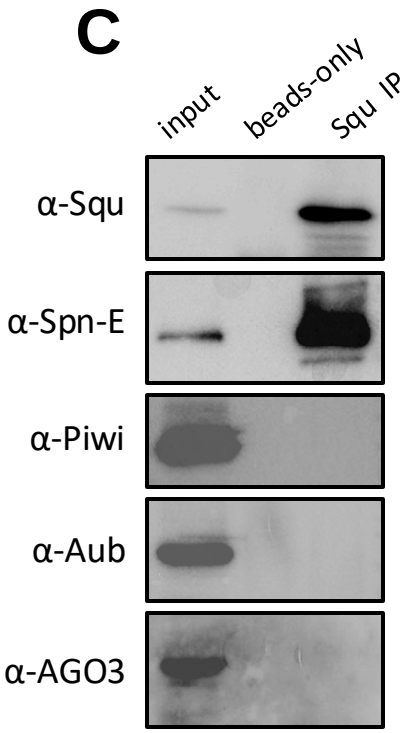
Kawaguchi, Figure 1

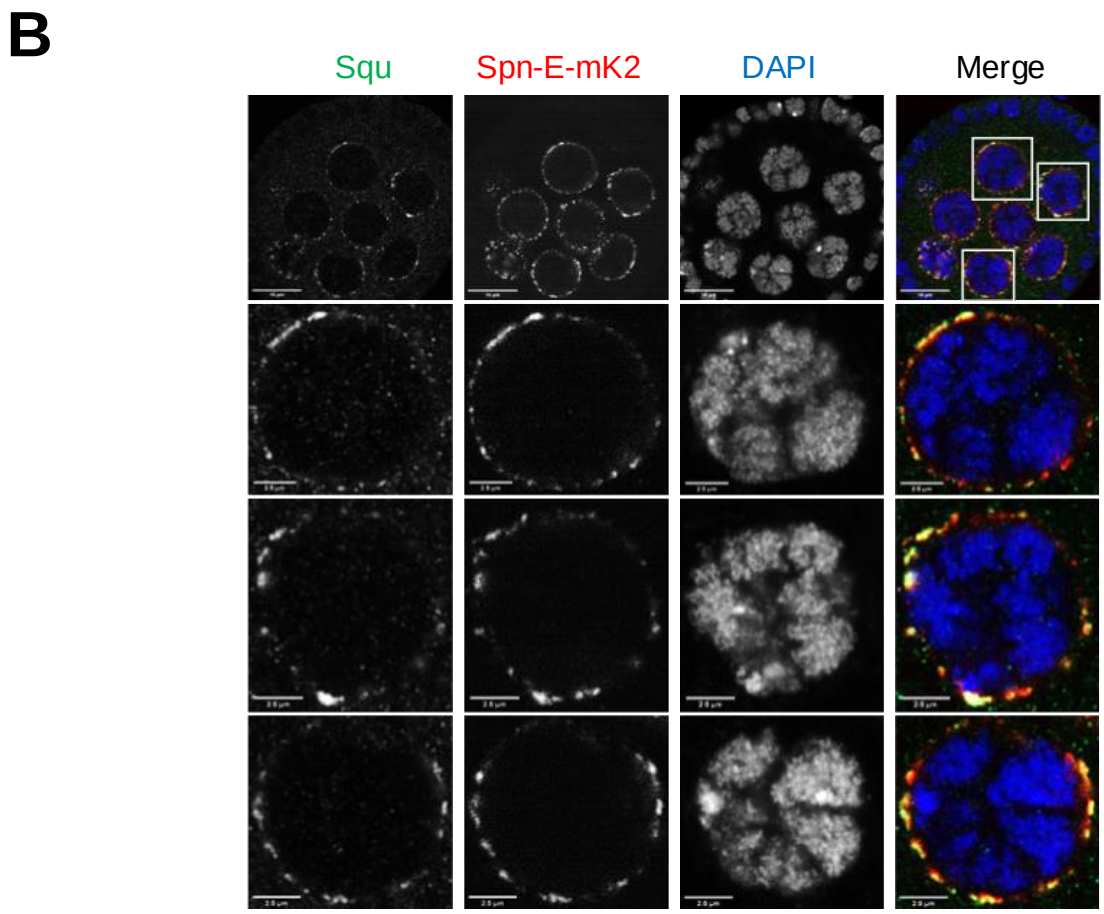


Kawaguchi, Figure 2




A
squ het
squ PP32/HE47
α-Tubulin
55kDa
α-Squ
28kDa

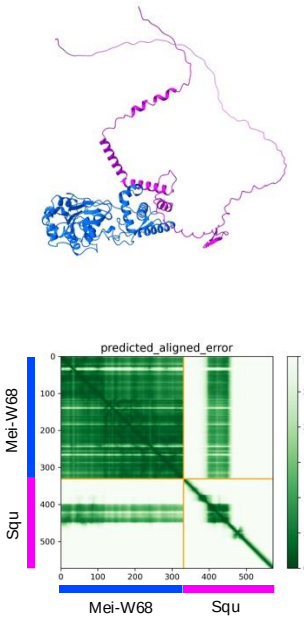

C
input
beads-only
Squ IP
α-Squ
α-Spn-E
α-Piwi
α-Aub
α-AGO3


B
Squ
Spn-E-mK2
DAPI
Merge

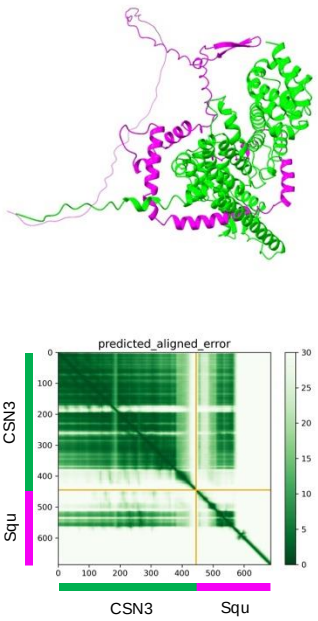
Kawaguchi, Figure 4

A

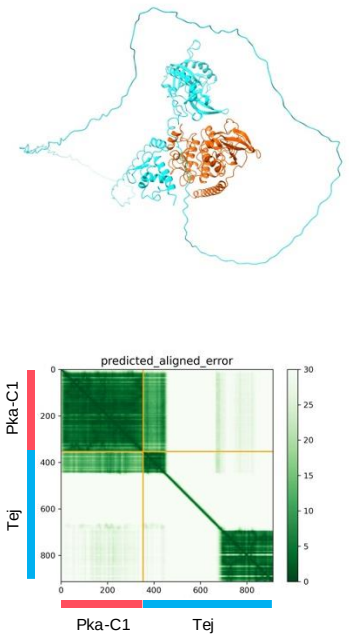
(i) Mei-W68_Squ



(ii) CSN3_Squ

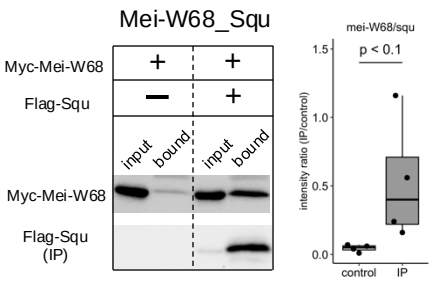


(iii) Pka-C1_Tej

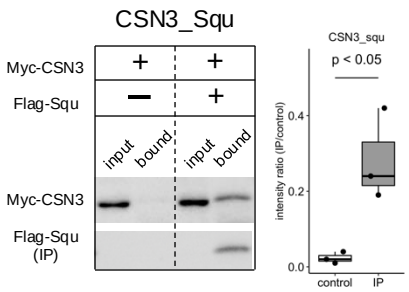


B

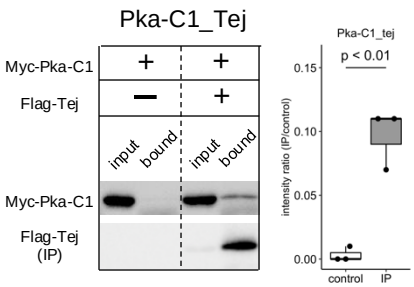
(i) Mei-W68_Squ



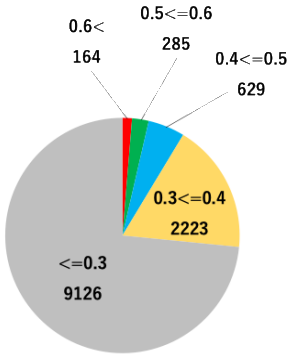
(ii) CSN3_Squ



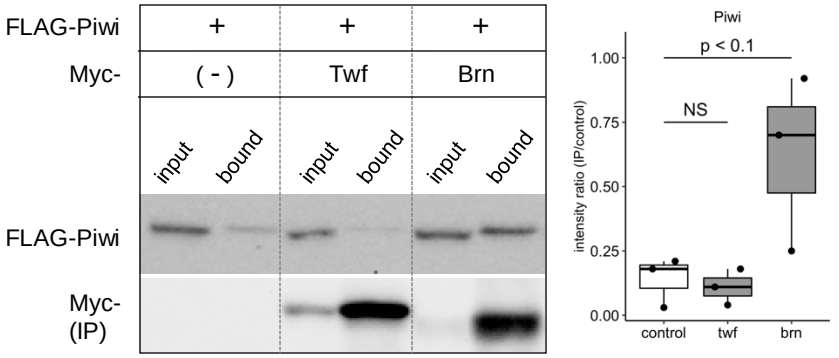
(iii) Pka-C1_Tej



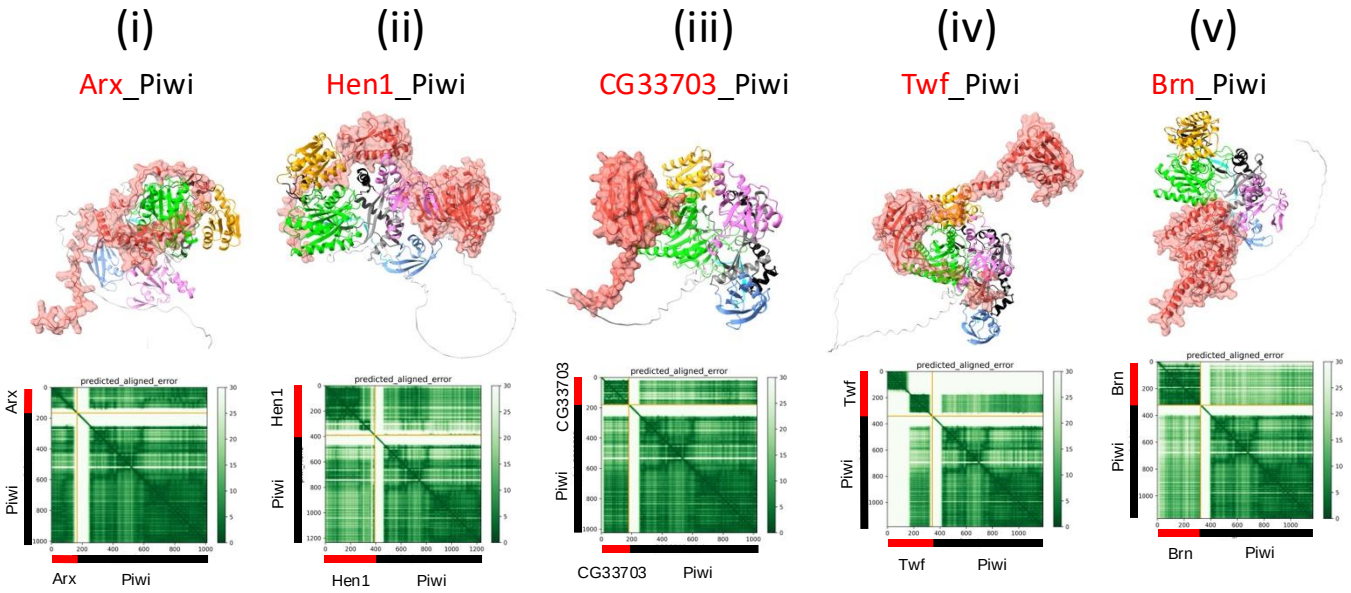
A


0.6<
164
0.5<=0.6
285
0.4<=0.5
629
0.3<=0.4
2223
<=0.3
9126

C


FLAG-Piwi
Myc-
+
+
+
(-)
Twf
Brn
input
bound
input
bound
input
bound
FLAG-Piwi
Myc-
(IP)
Piwi
p < 0.1
NS
intensity ratio (IP/control)
1.00
0.75
0.50
0.25
0.00
control
twf
brn

B


(i)
Arx_Piwi
(ii)
Hen1_Piwi
(iii)
CG33703_Piwi
(iv)
Twf_Piwi
(v)
Brn_Piwi
predicted_aligned_error
Arx
Piwi
Hen1
CG33703
Twf
Brn