Development, Growth & Differentiation
The Japanese Society of Developmental Biologists

*Develop. Growth Differ.* (2012) **54**, 32–43 doi: 10.1111/j.1440-169X.2011.01308.x

Review Article

# Tudor domain-containing proteins of *Drosophila melanogaster*

Muying Ying[1] and Dahua Chen[2]*

[1]*Department of Molecular Biology and Biochemistry, Basic Medical College of Nanchang University, Nanchang; and* [2]*State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China*

Tudor domain-containing proteins (Tudor proteins), which recognize and bind to methyl-arginine/lysine residues, play important roles in diverse epigenetics, gene expression and the regulation of various small RNAs. Using the complete set of 23 Tudor proteins from *Drosophila*, together with the available functional information, we propose a putative link for different types of Tudor domains (histone-binding, SMN and SND1) and the four functional groups of Tudor proteins (Group 1, binding the methyl-lysine/arginine of histone tails; Group 2, binding the methyl-RG/RA box of ligand; Group 3, binding the methyl-RG/RA box of microRNPs; and Group 4, binding the methyl-RG/RA box of PIWI proteins). Tudor domain types are distinguished by the nature of the sequence flanking the canonical Tudor domains. Sequence analysis indicates that Tudor domains experienced stepwise transit from one type to another during evolution. Tudor proteins of Group 4, collectively representing the great majority of Tudor proteins in *Drosophila*, are characterized by multiple Tudor domain repeats, which might be required for associating with several molecules of the same germ granule components. Tudor domain, a segment of approximately 60 amino acid residues, has been found in fungi, protozoa, unicellular eukaryota, plants and metazoa but not in the *Guillardia theta* nucleomorph. Similar frequencies of Tudor-containing genes (Tudor genes) among vertebrates and the frequent occurrence of orthologues among vertebrates, along with similar observations within arthropods suggest that Tudor genes are inherited largely vertically during evolution within different phylogenetic lineages.

**Key words:** evolutionary track, flanking sequence, functional group, Tudor domain, vertical inherit.

## Functional groups of Tudor domain-containing proteins

The Tudor domain and the Chromo, MBT, PWWP and Agenet-like domains constitute the ''Royal family'' of protein–protein interacting domains (Kim *et al.* 2006). As members of the Royal family, Tudor domain-containing proteins (Tudor proteins) have been implicated in diverse epigenetic functions, including methylation-dependent chromatin-remodeling, histone-binding, pre-RNA-processing, RNA-silencing and transposon silencing in ligands (Siomi *et al.* 2010a,b). Of Tudor proteins, Tdrd11/SND1, a well-understood Tudor protein, shows multiple biological functions, including participation in double-stranded RNA editing, pre-mRNA splicing (Selenko *et al.* 2001), microRNA-mediating gene silencing (Yoo *et al.* 2011) and piRNA biogenesis in germlines (Saito *et al.* 2010). Tdrd11/SND1 has been shown to interact with Piwi, Ago2, STAT5-6, Pim1, Myb, G3BP and AEG-1 (Leverson *et al.* 1998; Gao *et al.* 2010; Siomi *et al.* 2010a,b; Yoo *et al.* 2011), and functions in many cancers. AEG-1 and SND1 are components of RISC, overexpression of any one of them can lead to increased RISC activity that might contribute to hepatocarcinogenesis (Yoo *et al.* 2011). Upregulation of SND1 may occur at a very early stage in colon carcinogenesis and contribute to the posttranscriptional regulation of key players in colon cancer development, including anaphase-promoting complex (APC) and $\beta$-catenin (Tsuchiya *et al.* 2007).

On the basis of the available functional data, Tudor proteins appear to fall into four groups (Fig. 1): Group 1, binding methyl-lysine/arginine of histone tails, including Tdrd3, PHF1, PHF20, the JMJD family and TP53BP1. As a ''reader'' of H3R17me2a and H4R3me2a marks on histone tails, human Tdrd3 is recruited to an estrogen-responsive element in a

*Author to whom all correspondence should be addressed.
Email: chendh@ioz.ac.cn
Received 3 September 2011; revised 26 September 2011; accepted 28 September 2011.

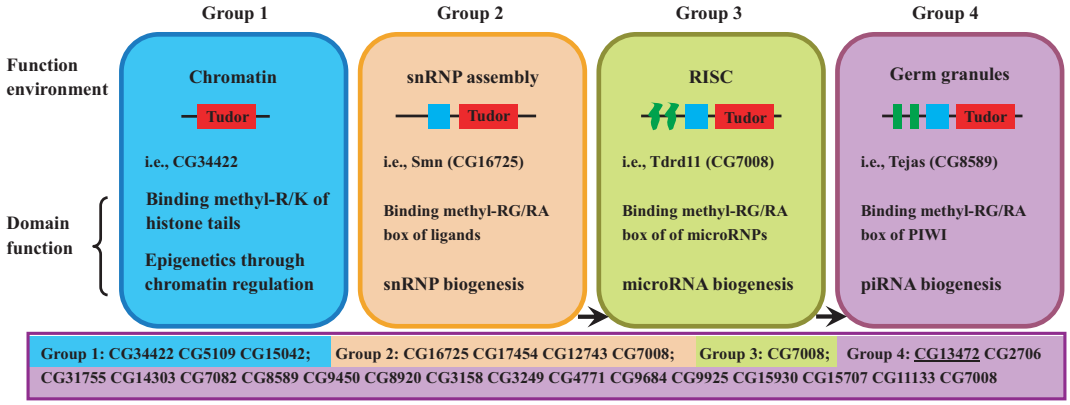


**Fig. 1.** Groups of Tudor domain types (chromatin regulation, snRNP biogenesis, microRNA biogenesis and piRNA biogenesis) in *Drosophila melanogaster* were indicated, together with the example with experimental data and the known or predicted functions. Black arrows indicated a potential evolutionary transition of Tudor domains from Group 2 (Smn) to Group 3 (Tdrd11) and to germline Tudor (Tejas). The structural organization of Tudor domain groups was compared within the colored boxes, with the Tudor domain colored in red, and the N-terminal extensions of a double $\beta$-strands and an $\alpha$-helix (known or predicted) in green and blue, respectively. Group 1: histone-binding-type Tudor domains, include CG34422, CG5109 and CG15042. All of them lack the unique structural extensions; Group 2: SMN-type Tudor domain, include CG7008, CG16725, CG17454 and CG12743. The last three were only provided with a single predicted $\alpha$-helix N-terminal to their canonical Tudor core domains; while CG7008 simultaneity appeared in Groups 2, 3 and 4 was provided with the unique structural extensions of a double $\beta$-strands and $\alpha$-helix N-terminal to their canonical Tudor core domains; Group 4: SND1-type Tudor domain, include CG13472, CG2706, CG31755, CG14303, CG7082, CG8589, CG9450, CG8920, CG3158, CG3249, CG4771, CG9684, CG9925, CG15930, CG15707, CG11133 and CG7008. Of them, except for CG13472 (underlined for easy identification), all the others were provided with the unique structural extensions of a double $\beta$-strands and an $\alpha$-helix N-terminal to their canonical Tudor core domains. For details, please refer to Figures S1 and S2. R, arginine; K, lysine; G, glycine; A, alanine.

Carm1-dependent manner, and promotes transcription by binding methyl-arginine marks on histone tails (Yang *et al.* 2010). Tudor domains of JMJD2A bind to methyl-lysine residues at histone H3 and H4 tails (especially H3K4me3, H3K9me3, H4K20me2 and H4K20me3), and interact with different histone tails by different binding modes (Ozboyaci *et al.* 2011). As a key transducer of the DNA damage checkpoint signal during mitosis, methylation-dependent direct binding of two tandem Tudor domains to histone H3 of TP53BP1 maintains the genome process (Huyen *et al.* 2004). Group 2 Tudor proteins bind to methyl-arginine of ligands and representative members include SMN and SMNDC1, which are found in pre-mRNA splicing factors and are required for efficient assembly of small nuclear ribonucleoproteins (snRNPs). Direct association of the Tudor domain of protein SMN with Sm proteins requires symmetric dimethylarginine (sDMA) modifications in the Sm target protein (Brahms *et al.* 2001). Splicing factor SMNDC1 bridges an interaction between the prespliceosome protein U2AF35 and tri-snRNP hPrp3 by association of the N-terminal Tudor domain of SMNDC1 with U2AF35 (Little & Jurica 2008). Group 3 is represented by the Tudor domain of SND1. SND1, a main component of the RNA-inducing silencing complex (RISC), binds to hyper-edited, double-stranded RNA and promotes its cleavage, and plays important roles in various cellular pathways (Scadden 2005). Also, SND1 can bind to methylated ligands or to the methylated 5′ cap of spliceosomal snRNAs to promote both the formation of *in vitro* spliceosome complexes and the first step of pre-mRNA splicing (Friberg *et al.* 2009). A recent study showed that the association of the Tudor domain of SND1 with Piwil1/Miwi in germ cells is dependent on sDMA modification of Piwil1/Miwi (Liu *et al.* 2010). Liu *et al.* superimposed the complex structure of SND1 and the R4me2s peptide on the complex structure of SND1 and the R14me2s peptide and found that they do not overlay well except for the methyl-arginine residues, wherein the wide binding groove of the unique structural extensions of double $\beta$-strands and an additional $\alpha$-helix N-terminal to SND1 canonical Tudor core domains provides plasticity that accommodates methyl-arginine peptides in different binding modes (Liu *et al.* 2010). Of Group 4, many Tudor proteins, including Tdrd1-9 and Tdrd11, have been identified in methylation-dependent association with PIWI proteins Ago3, Aub and Piwi (Siomi *et al.* 2010a,b). Tdrd1 promotes the formation of piRNP complexes and ensures the entry of correct transcripts into the normal piRNA pool by recognizing methyl-arginine residues at the N terminus of Mili (Reuter *et al.* 2009; Ishizu *et al.* 2011). Targeted mutation of Tdrd5 in mice

leads to male sterility because of postnatal spermatogenic defects. Tdrd5, Tdrd6 and Tdrd7 are essential for haploid spermatid development and promoting cytoplasmic RNP formation of the chromatoid body and mutation of any one of them causes male sterility (Tanaka *et al.* 2011; Yabuta *et al.* 2011). Furthermore, Tdrd7 suppresses LINE1 retrotransposons independently of piRNA biogenesis wherein Tdrd1 and Tdrd9 operate, suggesting that distinct Tdrd pathways against retrotransposons might exist in the male germline (Tanaka *et al.* 2011).

## Tudor domain-containing proteins in *Drosophila melanogaster*

With genetic tractability and a well-annotated genomic sequence, *Drosophila melanogaster* represents an excellent model system for studying a number of fascinating biological processes, such as stem cell regulation, germ cell meiosis and oocyte determination (Yang *et al.* 2007b,c). Using traditional genetic screening, Tudor protein was originally identified to be a component of the germ granule required for *Drosophila* oogenesis (Boswell & Mahowald 1985). Subsequently, Tudor proteins were found within a number of complexes, including spliceosomal complexes, RNA editing complexes, and RISC and piRNA complexes of the chromatoid body, and have been linked to many cellular processes, such as DNA repair, chromatin remodeling, self-renewing divisions of germlines, etc. With the development of classical genetic methods and a reverse genetic technique for analyzing gene function in *Drosophila melanogaster* (Haley *et al.* 2008; Wang *et al.* 2011), identification of novel genes involved in germline stem cell (GSC) regulation and germline development will become more tractable. Tudor protein in *D. melanogaster* is a component of two types of germ granules: nuage and polar granules of the germ plasm. The former has been implicated in the degradation of transposon mRNAs through the piRNA pathway (Nagao *et al.* 2010). The latter is necessary and sufficient for germ cell formation at posterior during early embryogenesis and Tudor protein is the earliest gene absolutely required for germ cell formation (Kirino *et al.* 2010).

On the basis of the known and predicted function information for Tudor proteins or the corresponding orthologues (Arkov & Ramos 2010; Jin *et al.* 2009; Safran *et al.* 2010; Siomi *et al.* 2010a,b; Handler *et al.* 2011), the complete set of 23 Tudor proteins in *D. melanogaster* can be identified as Tudor proteins expressed selectively in germ cells and those expressed broadly (Table 1). The number of putative members identified in groups 1–4 are 3, 4, 1 and 17,

**Table 1.** Domain architecture and interologue interacting partner of Tudor proteins in fruit fly

| Name | Domain architecture | Binding Piwi/partner |
|---|---|---|
| CG14303 | Tudor × 5 | Ago3, Piwi |
| CG7082 | KH × 2-Tudor | Ago3, Piwi |
| CG3249 | KH-Tudor | Siah2, Ubiquitin, Porin |
| CG8589 | Tudor | Ago3, Aub |
| CG9450 | Tudor × 10 | Ago3, Piwi |
| CG8920 | Tudor × 2 | Piwi |
| CG3158 | Dexdc-HELICc-HA2-Tudor-ZnF_C2H2 | Ago3, Piwi |
| CG12743 | Out-Tudor | ALG1, ALG14 |
| CG7008 | SNc-SNc-SNc-SNc-Tudor | Piwi, Ago2, STAT5-6 |
| *CG9925* | Tudor × 2 | Not known |
| *CG15707* | ZF_C3HC1-Tudor | Not known |
| CG13472 | DUF-UBA-Tudor | Piwi, FMRP, FXR1-2 |
| CG34422 | Tudor-RBB1NT-BRIGHT-Chromo | Not known |
| CG5109 | Tudor-PDH-PHD | Smad3, PRKD2 |
| CG17454 | Tudor | SmD1, SmD3, |
| CG16725 | Tudor | SNRPB, SNRPD1-3 |
| *CG2706* | DEAD-Tudor | Armitage |
| *CG4771* | Tudor × 2 | Piwi, Armitage, Yb, Zucchini |
| *CG11133* | Q motif-DEAD-Tudor | Not known |
| *CG31755* | Tudor-Q motif-Tudor | Not known |
| *CG9684* | MYND-tudor -tudor | Not known |
| *CG15042* | Tudor | Not known |
| *CG15930* | Tudor | Not known |

Interologue interacting partner of Tudor proteins are from the corresponding gene report or its orthologues. For details, please refer to Table S2. Tudor proteins indicated by blue letters are expressed in germ cells, and others are broadly expressed. Tudor proteins indicated by italic letters lack the corresponding orthologues from human.

respectively (Fig. 1). CG34422 in Group 1 members are orthologues of human ARID4B, which has two repression domains: a C-terminal domain that interacts with the mSin3A–HDAC complex, and an N-terminal Tudor domain that functions independently of mSin3A–HDAC (Fleischer *et al.* 2003). Studies of human ARID4B indicate that the Tudor domain could be responsible for targeting the mSin3A complex to DNA, stabilization of the complex via multiple interaction domains or recruitment of additional factors that might contribute to repression (Fleischer *et al.* 2003). *D. melanogaster* Pcl-Tudor contains an atypical, incomplete aromatic cage that does not interact with known Tudor domain ligands, and might engage in intra- or intermolecular interactions through an exposed hydrophobic surface patch, whereas human Pcl orthologues exhibit a complete aromatic cage and might recognize methyl-arginine/lysine residues (Friberg *et al.* 2010). Members of Group 2 involved in pre-mRNA splicing are represented by the *Drosophila* SMN complex proteins Gemin2-3 and Gemin5, which are components of U bodies, interact with sDMAs of Sm protein *in vivo* and play an important role in snRNP biogenesis (Cauchi *et al.* 2010). SMNDC1/CG17454 has been suggested to function as a core component of the spliceosome function in splicing regulators (Park *et al.* 2004). The *Drosophila* Otu/CG12743 transcript can be spliced to generate 98 and 104 kDa proteins, and expression of the 104 kDa protein including the Tudor domain is detected in the early stages of oogenesis (Steinhauer & Kalfayan 1992). Mutations affecting the amount of Otu104 or its activity lead to the development of tumorous egg chambers and cause defects in the dorsoventral polarity of the egg (Van Buskirk & Schüpbach 2002). Studies in zebrafish (*Danio rerio*) showed that, with the Otu domain involved in ubiquitination, Otu might link the ubiquitin signaling pathway to early oogenesis and maintain the totipotency of embryonic cells (Mo *et al.* 2005). Three classes of small RNAs, including small interfering RNAs (siRNAs), microRNAs and Piwi-interacting RNAs (piRNAs) have been shown to control stem cell regulation and GSC maintenance in *Drosophila* (Yang *et al.* 2007b,c, 2009). In Group 3, which is involved in RISC, SND1/TDRD11 (CG7008) is a well known and versatile Tudor protein that can participate in microRNA-mediated RNA silencing (Cauchi *et al.* 2010; Caudy & Hannon 2004), facilitate spliceosome assembly of pre-mRNA splicing (Yang *et al.* 2007a), associate with Piwil1/Miwi in germ cells (Liu *et al.* 2010). Group 4 members are involved in piRNA biogenesis. SpnE, Krimp and Tejas, which contain Tudor domains that recognize and bind to sDMAs on targets, are components of perinuclear nuage in the nurse cells of the fly egg chamber and function in the repression of retrotransposons in the germline. Krimp is crucial for the production of Su(Ste) and AT-chX piRNAs (Nagao *et al.* 2010). Aub and Spn-E bind to the Tudor domain at the C terminus of Tejas, which enables Tejas to contribute to the formation of a macromolecular complex in the perinuclear region and engages it in the production of germline piRNAs (Patil & Kai 2010). Germ plasm assembly requires binding of the Tudor domain to methylation-modified Aub, and the interaction between the methylation-modified PIWI proteins and Tudor proteins is evolutionarily conserved in germ cells (Kirino *et al.* 2010). fs(1)Yb/CG2706 acts via the PIWI- and Hedgehog-mediated signaling pathways that emanate from the same signaling cells to control germline cell division (King *et al.* 2001). In addition, the results of epistasis experiments indicate that vreteno/CG4771 acts upstream of or parallel with the stem cell maintenance signals piwi and pumilio, and might function in restricting or refining GSC maintenance signals from the somatic niche (Davis 2007).

## Evolutionary transition of Tudor domain types in *D. melanogaster*

On the basis of the notion that (i) SND1/CG7008 consists of four repeat staphylococcal nuclease-like domains (SNc1-SNc4) at the N terminus followed by SND1-type Tudor and the SND1-type Tudor domain is composed of a complete SNc-like domain interdigitated with a Tudor domain (Zheng *et al.* 2009). (ii) The entire Tudor domain and a bifurcated SNc domain are required for its association with Piwil1/Miwi in germ cells, whereas the canonical Tudor domain alone is insufficient for methyl-arginine ligand binding (Liu *et al.* 2010). (iii) A unique structural feature of germline cell Tudor domains is an additional two $\beta$-strands and an $\alpha$-helix N-terminal to the canonical Tudor core domain (Jin *et al.* 2009). We used *D. melanogaster* Tudor proteins to analyze the nature of sequences flanking the canonical Tudor core domains in order to examine the relationships of sequence difference and function diversity among Tudor domains. The results showed that (Figs 1, 2 and S1) (i) all members (CG34422, CG5109 and CG1504) of Group 1 with Tudor domains binding methyl-arginine/lysine residues of histone tails lack the unique structural extensions of double $\beta$-strands and an additional $\alpha$-helix N-terminal to their canonical Tudor core domains. We name them histone-binding-type Tudor domains for easy differentiation. (ii) Members of Group 2 with Tudor domains involved in snRNP biogenesis for pre-mRNA splicing are represented by the Tudor domain found in the SMN protein (CG16725), which only has a predicted

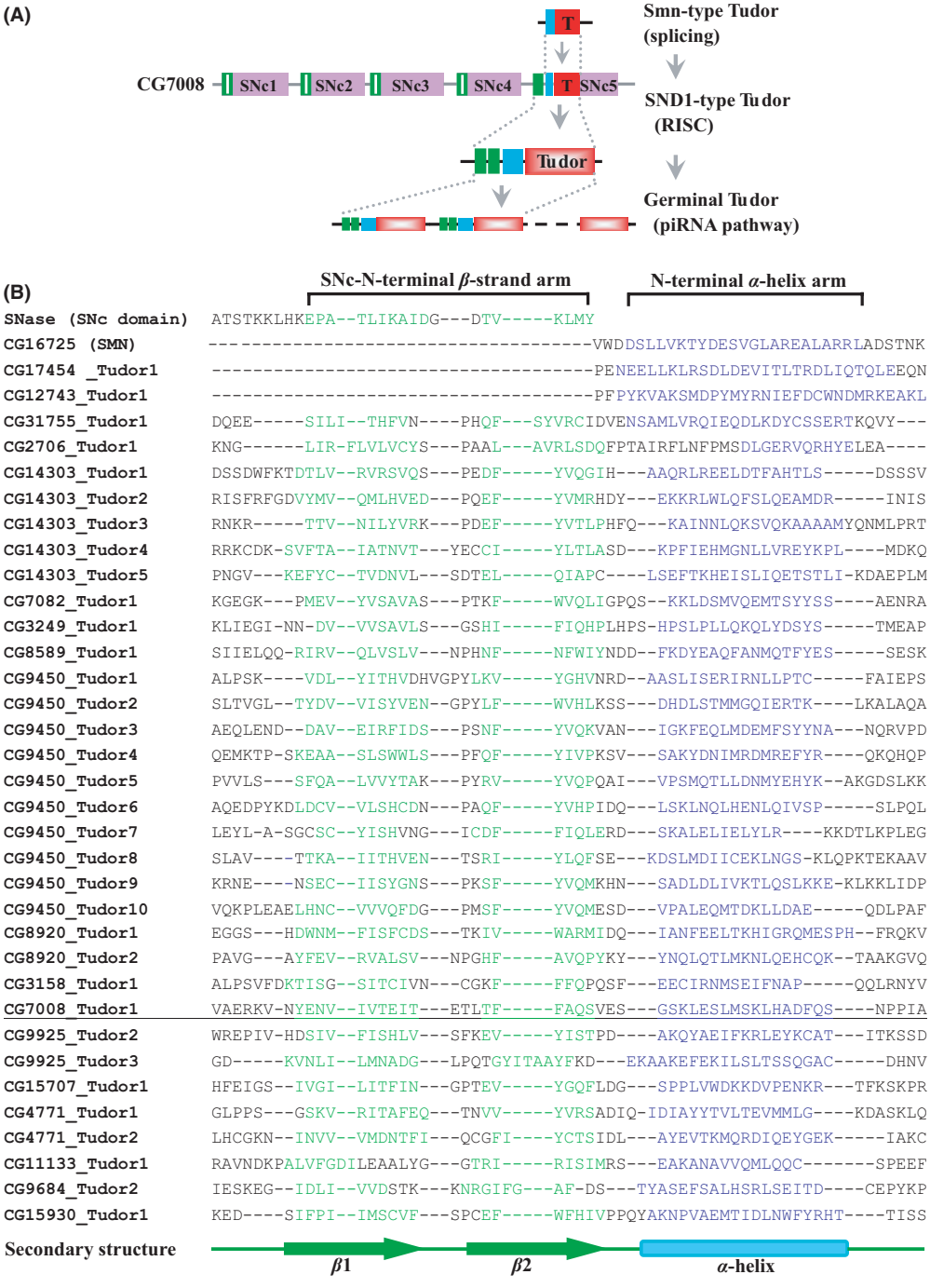


**Fig. 2.** Putative stepwise change in functional evolution of Tudor domains in *Drosophila melanogaster*. (A) Putative stepwise change of Tudor domain function. An SMN-type Tudor domain (red rectangle; T, Tudor), with a predicted N-terminal a-helix (light blue rectangle), was inserted into an SNc nuclease domain (SNc5 in SND1) downstream of its N-terminal double $\beta$-strands (green rectangle), which come into being SND1-type Tudor domain. Germline Tudor proteins are predicted to have the unique structural extensions of a double $\beta$-strands and an $\alpha$-helix N-terminal to their canonical Tudor core domains; these could have derived from an SND1-type Tudor or evolved independently. Arrows indicate a potential evolutionary track of these Tudor domain classes. (B) Multiple alignments of germline Tudor domain sequence flanking the canonical Tudor core domains, showing 50 amino acids N-terminal to the Tudor domain core. Secondary structure features (green letters: $\beta$-strands, blue letters: $\alpha$-helix) were predicted by NetSurfP or were colored according to experimentally-determined structural data. The SND1 (CG7008) sequence was indicated by underline for easy identification. Corresponding sequences of staphylococcal nuclease (SNase, ZP_04839945) of *Staphylococcus aureus* subsp. aureus str. CF-Marseille and *Drosophila melanogaster* SMN (CG16725) were shown in the first two lines of sequence alignments.

α-helix extension to the N terminus of the Tudor domain. In agreement with this, CG16725 CG17454 and CG12743 Tudor proteins in Group 2 involved in pre-mRNA splicing possess only a predicted α-helix extension to the N terminus of the Tudor domain. We name them SMN-type Tudor domains for easy differentiation. (iii) An SMN-type Tudor domain, with a predicted N-terminal α-helix, was inserted into an SNc nuclease domain (SNc5 in SND1) downstream of its N-terminal double β-strands, which created a composite SNc–Tudor structure (SND1-type Tudor domains) and were equipped with unique structural extensions of double β-strands and an additional α-helix N-terminal to their canonical Tudor core domains (Fig. 2). This notion is supported by the observation that the β-barrel core of the fifth SN domain can be superimposed well with the canonical Tudor domain of SND1 (Liu *et al.* 2010). An apparently stepwise accumulation of structure describes its evolutionary footprint of transition in Tudor domain function from chromatin remolding in the context of Tdrd3 and ARID4B to pre-mRNA splicing in the context of the SMN proteins, and to processing small noncoding RNAs, including siRNAs, microRNAs and piRNA, in the context of the SND1. (iv) Consistently, secondary structure analysis of 17 germline Tudor proteins in Group 4 suggested that all, except CG13472, possess the unique structural extensions of double β-strands and an α-helix N-terminal to their canonical Tudor core domains (Figs 2 and S1). Furthermore, the Tudor domain of SND1/CG7008 is the most ancient molecule with a unique N-terminal extension (Table S2). All of these observations support the notion that germline Tudor proteins could have originated from a precursor(s) of SND1-type Tudor domains; alternatively, the SND1-type Tudor domains and germline Tudor domains might have converged on the same N-terminal structural motif (Jin *et al.* 2009).

## Domain architecture of Tudor proteins in *D. melanogaster*

Acquiring such unique N-terminal extensions might have equipped Tudor domains with new intrinsic properties and new biological functions, especially when coupled with novel domains. Furthermore, as basic building blocks and basic functional units of proteins, protein functions have been tightly linked to protein domains (Ying *et al.* 2011). Therefore, we analyzed domain architectures of *Drosophila* Tudor proteins in Figure S2. Tudor proteins have been linked to chromatin remolding, snRNP biogenesis, microRNA biogenesis and piRNA biogenesis. Consistent with those functions, domain architecture analysis showed that Tudor domains are associated predominantly with various RNA-binding motifs (DEAD-box, DEXDc and KH-domain), RNA helicase-associated domains (HELICc, Q motif and HA2), chromatin-binding domains (Chromo and PHD finger) and DNA- and histone-binding domains (BRIGHT, RBB1NT, ARID and Chromo) in the same polypeptide (Fig. S2). Tudor domains in Group 1 are usually associated with DNA- and histone-binding domains, such as Chromo and PHD finger. For example, chromatin-remodeling gene CG34422 (orthologue of human Arid4a and Arid4b) contains a Tudor domain along with three histone- and DNA-binding domains of RBB1NT, ARID and Chromo. Deficiency of Arid4a and Arid4b alters epigenetic modifications with reduced trimethylation of histone H4K20 and H3K9 and reduced DNA methylation, and suppresses genomic imprinting defects in the PWS/AS domain (Wu *et al.* 2006). Without sequence preference in p270 DNA-binding activity, ARID family proteins might be involved in a wider range of DNA interactions. Pcl/CG5109 comprises a Tudor domain and tandem PHD fingers. These domains are known to recognize methyl-lysine/arginine residues and could contribute to targeting of Pcl-PRC2 (Friberg *et al.* 2010). The PHD fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9 (Mansfield *et al.* 2011). Multiple similar property domains in the same polypeptide might enhance plasticity and enable different ligands in different binding modes, and increased affinity of Tudor proteins to bind histone and DNA, or to recruit additional factors. One distinct feature of the germline Tudor proteins in Group 4 is that most of them are provided with multiple Tudor domain repeats, which might be required for associating with several molecules of the same germ granule components (Fig. S2). With multiple Tudor domain repeats and the combination of these with other functionally distinct domains, Tudor proteins might serve as scaffolds and function in a specialized manner in the piRNA pathways that are necessary for various cellular processes. For example, Tudor proteins of CG11133 and CG2706 include the DEAD-box domain of RNA helicases, which are known components of RNP germ granules required for germline development. Most of the additional domains of the germline Tudor proteins are zinc-binding domains (zfMYND, ZfC2H2 and ZfC3H1), which are stable motifs of a few residues ligating metal ions and might be more favored than others for stabilization of small domains in a reducing environment where disulfide bonds do not form readily. In addition, Otu and UBA domains of ubiquitination are found in the germline Tudor proteins and cross-talk might occur between the ubiquitin signaling pathway and the piRNA pathway. Tudor proteins in Group 4 and their

orthologues in other organisms, including mice, human and frog, have been localized to the germinal body-like structures of male germ cells, and/or have been identified in sDMA-dependent association with PIWI proteins (Golam Mostafa *et al.* 2009; Reuter *et al.* 2009; Siomi *et al.* 2010a,b; Tanaka *et al.* 2011; Yabuta *et al.* 2011).

## Tudor-containing gene distributions in different organisms

By stepwise accumulation of structure from histone-binding-type Tudor domains to SMN-type Tudor domains and to SND1-type Tudor domains, germline Tudor proteins finally obtained the unique N-terminal extension of double $\beta$-strands and an $\alpha$-helix preceding the N terminus of the Tudor core domain. We asked to what extent are the N-terminal extensions of Tudor domains related to different functional types of Tudor domains? Are the N-terminal extensions of Tudor domains evolutionarily conserved? To obtain a more complete picture of N-terminal extensions of Tudor domains, we first collected Tudor-containing genes (Tudor genes) from 18 species, which indicated that Tudor genes experienced evolutionary expansion accompanied by an increase in species complexity (Tables S1, S2). Similar frequencies of Tudor genes among vertebrates, including *Homo sapiens* (31) and *Pan troglodytes* (30) in primates, *Mus musculus* (29) and *Rattus norvegicus* (29) in rodents and *Gallus gallus* (30), *Danio rerio* (32) and *Xenopus tropicalis* (30) in other vertebrates, suggest that the common ancestor of vertebrates had finished evolutionary Tudor gene expansion before their divergence 485 million years ago (Mya) in the Paleozoic Era. With similar frequencies of Tudor genes among arthropods, including *D. melanogaster* (23), *Anopheles gambiae* (19) and *Apis mellifera* (18), the common ancestors of insects might have finished their evolutionary expansion of Tudor genes before their divergence 535 Mya in the Paleozoic Era (Fig. 3). With only one member in *Schizosaccharomyces pombe* and *Schizosaccharomyces japonicus yFS275*, the Tudor gene is notably absent in fungi. From protozoan (two in *Dictyostelium discoideum*) to unicellular protozoan (four in *Monosiga brevicollis*) and to multicellular organisms (11 in *Caenorhabditis elegans*), Tudor genes experienced marked evolutionary rates of expansion. As an endosymbiont, the *Guillardia theta* nucleomorph (as an endosymbiont, many genes may have been lost due to its condition, and it is kept only while it encodes something necessary for survival) lacks detectable Tudor genes (Table S1). With a small assignment in plant (*Oryza sativa* and *Arabidopsis thaliana*), Tudor proteins are encoded by six Tudor genes in *O. sativa* (genome of 28 236 genes) and 11 in *A. thaliana* (genome of 28 000 genes). Consistent with this observation, piRNAs are abundant in most metozoa but notably absent from plants and fungi (Grimson *et al.* 2008). The observations may suggest that Tudor genes and piRNAs have a common and essential function in the germline.

With high proportions of alternatively spliced transcripts in primates (human: 21/31, 67.7%; chimpanzee:

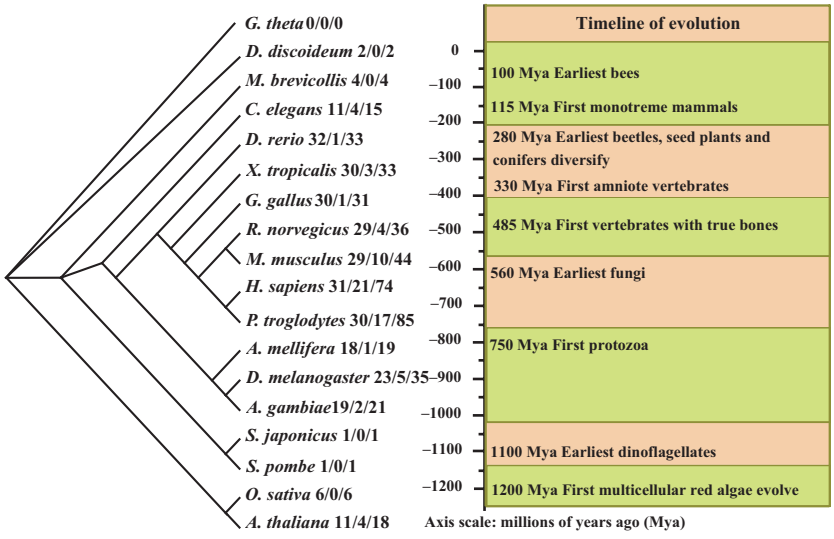


**Fig. 3.** Estimated numbers of Tudor-containing genes in different organisms. The numbers following the corresponding organism name represents the number of Tudor-containing genes, the number of alternative spliced genes and the corresponding mRNA numbers, respectively. For details, please refer to Tables S1 and S2.

17/30, 56.7%), 31 human and 30 chimpanzee Tudor genes encode 74 and 85 mRNAs of Tudor proteins, respectively (Tables S1, S2). Although the numbers of Tudor genes among vertebrates are almost equal, primate (human and chimpanzee) cells make extensive use of alternative splicing to generate more transcripts from a single gene than the number of genes in an entire genome. Earlier studies showed that duplicated genes have fewer alternative splicing isoforms than single-copy genes, and that recent duplicates usually lose alternative splicing isoforms, whereas the ancient duplicates could evolve new alternative splicing isoforms during evolutionary processes (Ying *et al.* 2009). We propose that the ancient Tudor proteins of primates experienced evolutionary expansion at the transcription level by alternative splicing. With high proportions of orthologues (ranging from 45.5% to 100%) in animals (Table S1), including protozoa, unicellular protozoa and metazoa, Tudor genes are largely inherited vertically and experienced strong selective pressure for conservation throughout evolution. Orthologous Tudor genes in metazoa largely occur, while the occurrence of orthologues in plants (*A. thaliana* and *O. sativa*) is quite small. Similar to the observations, Tudor genes and piRNAs are notably absent from plants (Tables S1, S2) (Grimson *et al.* 2008).

Among the 18 species mentioned above, 307 Tudor genes have been identified, which include 472 Tudor domains. We examined all the known or predicted structural features of the N terminus of these canonical Tudor core domains (Fig. S1 and Table S2) and the results showed that: (i) Tdrd3, ARID4A, ARID4B, PHF19, MTF2, PHF1, LBR, JMJD2A-C, ZGPAT, TP53BP1, PHF20 and PHF20L1, and their corresponding orthologues in the 18 species, lack the unique structural extensions of double $\beta$-strands and an additional $\alpha$-helix N-terminal to the canonical Tudor core domains, which are highlighted in bright yellow in Table S2. Of these Tudor proteins, Tudor domains in Tdrd3, Arid4a-4b, JMJD2A and TP53BP1 have been suggested to recognize and bind to methyl-arginine/lysine marks on histone tails. By binding methyl-arginine marks on histone tails, an intact Tudor domain of Tdrd3 is required for Tdrd3 promoting transcription (Yang *et al.* 2010). The chromatin remodeling genes Arid4a and Arid4b play suppressive roles in epigenetic alterations of leukemogenesis and link leukemia suppression and the epigenetic definition of histone modifications (Wu *et al.* 2008). JMJD2A and TP53BP1 have a similar double Tudor domain configuration but with distinct folds despite the sequence similarity. The unusual folds of these proteins are required for JMJD2A to recognize and to bind to methylated histones (Huang *et al.* 2006). Binding of the TP53BP1 Tudor domain to K382me2 of the non-histone peptide p53 might facilitate p53 accumulation at sites of DNA damage and promote DNA repair. (ii) SMNDC1, SMN2, SMN1, ALG13, Tdrd8 and Tdrd10, and their corresponding orthologues in the 18 species, have a single predicted $\alpha$-helix at the N terminus of the Tudor domain (highlighted in bright green in Table S2). Of these Tudor proteins, Tudor domains in SMNDC1 and SMN1-2 have been shown to be involved in RNP biogenesis for mRNA splicing (Little & Jurica 2008; Sun *et al.* 2005). (iii) Tdrd1, Tdrd2, Tdrd4-9, Tdrd11, AKAP1, SETDB1 and Tdrd12, and their corresponding orthologues in the 18 species, have structural extensions of double $\beta$-strands and an $\alpha$-helix N-terminal to their canonical Tudor core domains. Of these Tudor proteins, Tdrd1, Tdrd2, Tdrd4-9, Tdrd11 and AKAP1 or their orthologues in other organisms, have been shown to have an sDMA-dependent association with PIWI proteins (Reuter *et al.* 2009; Vasileva *et al.* 2009;Siomi *et al.* 2010a,b; Tanaka *et al.* 2011; Yabuta *et al.* 2011).

Finally, on the basis of the 472 known Tudor domain sequences from 17 species, we constructed a wheel of phylogenetic analysis of Tudor domains using neighbor-joining (NJ) for inferring the possible evolutionary clue of Tudor proteins and the N-terminal extension of these canonical Tudor core domains (Fig. S3). The subfamilies are marked on the branches on the basis of tree topologies and the subfamilies of human Tudor proteins. (i) Except for the *G. theta* nucleomorph lacking Tudor genes, putative orthologues of SND1 can be identified within all the 17 species. SND1 is the most ancient molecule of the Tudor protein family. (ii) Marked expansion of Tudor genes has been accompanied by increased complexity of the reproductive process from asexual reproduction (*D. discoideum,* two Tudor genes) to schizogony (*M. brevicollis*, four Tudor genes) and to sexual reproduction (*C. elegans*, 11 Tudor genes), which strongly suggests a common and essential function of Tudor proteins in the germline. (iii) Orthologues of SETDB1, LBR, ZGPAT and JMJD2A-C from arthropods lack a clearly identifiable Tudor domain. Pairwise sequence comparison of Tudor domain segments in orthologous pairs of SETDB1, LBR and ZGPAT indicated that there is, to some extent, similarity between orthologous pairs from vertebrates and arthropods, whereas the corresponding segment of SETDB1 from *D. melanogaster* but not *A. gambiae* or *A. mellifera* has lost the characteristic Tudor domain. Similarly, the Tudor domain of LBR is readily detectable in vertebrates and *A. mellifera* but not in *D. melanogaster* or *A. gambiae.* The corresponding segment of the Tudor domain of ZGPAT in arthropods has lost the characteristic Tudor domain.

Parsimoniously, loss of the Tudor domain from arthropod genes might originate mostly from amino acid mutation (SETDB1, LBR and ZGPAT) upon the selection and evolution of Tudor domains. Demethylase JMJD2A-C in arthropods consists of two different domains, JmjC and JmjN, whereas demethylase JMJD2A-C in vertebrates consists of four different domains JmjC, JmjN, two PHD and two tandem Tudor domains. The availability of the catalytic activity of human JmjC and JmjN domains requires the Tudor domains of JMJD2A to bind to methylated histone (Huang *et al.* 2006). The lack of the Tudor domain in arthropod JMJD2A suggests it is possible that when arthropod JMJD2As function as histone demethylases, other proteins might be required to function as the equivalent of Tudor domains in human JMJD2A. (iv) SETDB1, LBR ZGPAT and JMJD2A-C experienced taxonomic-independent evolutionary processes in vertebrates and arthropods, which might have resulted in gain- and loss-of-function of these Tudor proteins in vertebrates and arthropods. Further, a few Tudor proteins are taxonomically more restricted. For example, Tdrd10 is a primate-specific Tudor protein and Agenet, a Tudor-like domain, is a plant-specific Tudor protein (Table S2).

## Materials and methods

*Sequences*

Tudor protein sequences were retrieved from Genbank of the National Center for Biotechnology Information (NCBI), databases of University of California Santa Cruz (UCSC) and Ensemble. The species we investigated included human (*Homo sapiens*, Build 37.2), chimpanzee (*Pan troglodytes*, Build 2.1), mouse (*Mus musculus*, Build 37.2), rat (*Rattus norvegicus*, RGSC v3.4), chicken (*Gallus gallus*, Build 2.1), zebrafish (*Danio rerio*, Zv9), western clawed frog (*Xenopus tropicalis*, Build 1.1) fruit fly (*Drosophila melanogaster*, Release 5.30), African malaria mosquito (*Anopheles gambiae*, AgamP3.3), honey bee (*Apis mellifera*, Amel4.5), nematode (*Caenorhabditis elegans*, WS225), baker's yeast (*Saccharomyces cerevisiae*, Build 2.1), *Schizosaccharomyces japonicus* yFS275 (Build 1.1), social amoeba (*Dictyostelium discoideum*, Build 2.1), choanoflagellate (*Monosiga brevicollis*, v1.0), rice (*Oryza sativa*, RAP Build 3), thale cress (*Arabidopsis thaliana*, Build9.1), and a red algal nucleomorph (*Guillardia theta* nucleomorph, Build 1.0). In addition, WormBase, FlyBase, VectorBase, SGD, DictyBase and M. brevicollis v1.0 were also searched by the basic local alignment search tool (BLAST) (see Reference database). Sequences containing Tudor domain (Tudor in SMART, InterProScan, Pfam) were retained for further data analysis.

*Phylogenetic analysis*

Removing redundant and false positive sequences, 472 Tudor domain sequences from the 18 organisms were retained. Putative orthologues of Tudor proteins from the 18 species were identified by Reciprocal Best Blast Hits (Altschul *et al.* 1990). Conserved Tudor domains (the canonical Tudor core domain in histone-binding-type Tudor domains, a predicted N-terminal $\alpha$-helix plus the canonical Tudor core domain in SMN-type Tudor domains, and a predicted N-terminal extension of a double $\beta$-strands and an $\alpha$-helix plus the canonical Tudor core domain in SND1-type Tudor domains) were further analyzed by multi-sequence alignment using MUSCLE (MUltiple Sequence Comparison by Log-Expectation) and manual adjustment (Edgar 2004). Phylogenetic trees were constructed using the program Molecular Evolutionary Genetics Analysis (MEGA) package version 5 (Kumar *et al.* 2008). The evolutionary analysis was inferred using NJ method (Saitou & Nei 1987). To assess the reliability of the phylogenetic tree, bootstrap test (1000 replicates; random seed = 34 000) were conducted. The evolutionary distances were computed under the model of JTT (Jones-Taylor-Thornton) matrix-based method (Jones *et al.* 1992) and are in the units of the number of amino acid substitutions per site. All sites containing alignment gaps and missing-information were retained initially, excluding them as necessary using the pairwise-deletion option. Substitution patterns among lineages were allowed to vary among sites using gamma-distributed rates (shape parameter = 1.2).

## References

Altschul, S. F., G, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Arkov, A. L. & Ramos, A. 2010. Building RNA-protein granules: insight from the germline. *Trends Cell Biol.* **20**, 482–490.

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

Brahms, H., M, L., de Brabandere, V., Fischer, U. & Lührmann, R. 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**, 1531–1542.

Cauchi, R. J., S-P, L. & Liu, J. L. 2010. Drosophila SMN complex proteins Gemin2, Gemin3, and Gemin5 are components of U bodies. *Exp. Cell Res.* **316**, 2354–2364.

Caudy, A. A. & Hannon, G. 2004. Induction and biochemical purification of RNA-induced silencing complex from Drosophila S2 cells. *Methods Mol. Biol.* **265**, 59–72.

Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Fleischer, T. C., Y, U. & Ayer, D. E. 2003. Identification and characterization of three new components of the mSin3A corepressor complex. *Mol. Cell. Biol.* **23**, 3456–3467.

Friberg, A., C, L., Mourão, A. & Sattler, M. 2009. Structure and ligand binding of the extended Tudor domain of *D. melanogaster* Tudor-SN. *J. Mol. Biol.* **387**, 921–934.

Friberg, A., O, A., Klymenko, T., Müller, J. & Sattler, M. 2010. Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. *Protein Sci.* **19**, 1906–1916.

Gao, X., G, L., Shao, J., Su, C., Zhao, H., Saarikettu, J., Yao, X., Yao, Z., Silvennoinen, O. & Yang, J. 2010. Tudor-SN interacts with and co-localizes with G3BP in stress granules under stress conditions. *FEBS Lett.* **584**, 3525–3532.

Golam Mostafa, M., S, T., Hiyoshi, M., Kawasaki, H., Kubo, H., Matsumoto, K., Abe, S. & Takamune, K. 2009. Xtr, a plural tudor domain-containing protein, coexists with FRGY2 both in cytoplasmic mRNP particle and germ plasm in Xenopus embryo: its possible role in translational regulation of maternal mRNAs. *Dev. Growth Differ.* **51**, 595–605.

Grimson, A., S, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degnan, B. M., Rokhsar, D. S. & Bartel, D. P. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197.

Haley, B., H, D., Trang, V. & Levine, M. 2008. A simplified miRNA-based gene silencing method for *Drosophila melanogaster*. *Dev. Biol.* **321**, 482–490.

Handler, D., O, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R. & Brennecke, J. 2011. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* **30**, 3977–93.

Huang, Y., F, J., Bedford, M. T., Zhang, Y. & Xu, R. M. 2006. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748–751.

Huyen, Y., Z, O., Ditullio R. A. Jr, Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Sheston, E. A., Mellert, H. S., Stavridi, E. S. & Halazonetis, T. D. 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411.

Ishizu, H., N, A. & Siomi, H. 2011. Gatekeepers for Piwi-piRNA complexes to enter the nucleus. *Curr. Opin. Genet. Dev.* **21**, 484–490.

Jin, J., X, X., Chen, C., Park, J. G., Stark, C., James, D. A., Olhovsky, M., Linding, R., Mao, Y. & Pawson, T. 2009. Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.* **2**, ra76.

Jones, D. T., T, W. & Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.

Kim, J., D, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y. & Bedford, M. T. 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **7**, 397–403.

King, F. J., S, A., Cox, D. N. & Lin, H. 2001. Yb modulates the divisions of both germline and somatic stem cells through piwi- and hh-mediated mechanisms in the Drosophila ovary. *Mol. Cell* **7**, 497–508.

Kirino, Y., V, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T. A. & Mourelatos, Z. 2010. Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA* **16**, 70–78.

Kumar, S., N, M., Dudley, J. & Tamura, K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**, 299–306.

Leverson, J. D., K, P., Orrico, F. C., Rainio, E. M., Jalkanen, K. J., Dash, A. B., Eisenman, R. N. & Ness, S. A. 1998. Pim-1 kinase and p100 cooperate to enhance c-Myb activity. *Mol. Cell* **2**, 417–425.

Little, J. T. & Jurica, M. 2008. Splicing factor SPF30 bridges an interaction between the prespliceosome protein U2AF35 and tri-small nuclear ribonucleoprotein protein hPrp3. *J. Biol. Chem.* **283**, 8145–8152.

Liu, K., C, C., Guo, Y., Lam, R., Bian, C., Xu, C., Zhao, D. Y., Jin, J., MacKenzie, F., Pawson, T. & Min, J. 2010. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc. Natl Acad. Sci. USA* **107**, 18398–18403.

Mansfield, R. E., M, C., Kwan, A. H., Oliver, S. S., Garske, A. L., Davrazou, F., Denu, J. M., Kutateladze, T. G. & Mackay, J. P. 2011. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J. Biol. Chem.* **286**, 11779–11791.

Mo, S., S, P., Lv, D., Chen, Y., Zhou, W., Gong, W. & Zhu, Z. 2005. Zebrafish z-otu, a novel Otu and Tudor domain-containing gene, is expressed in early stages of oogenesis and embryogenesis. *Biochim. Biophys. Acta* **1732**, 1–7.

Nagao, A., M, T., Huang, H., Chen, D., Siomi, M. C. & Siomi, H. 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. *RNA* **16**, 2503–2515.

Ozboyaci, M., G, A., Erman, B. & Keskin, O. 2011. Molecular recognition of H3/H4 histone tails by the tudor domains of JMJD2A: a comparative molecular dynamics simulations study. *PLoS One* **6**, e14765.

Park, J. W., P, K., Celotto, A. M., Reenan, R. A. & Graveley, B. R. 2004. Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc. Natl Acad. Sci. USA* **11**, 15974–15979.

Patil, V. S. & Kai, T. 2010. Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein tejas. *Curr. Biol.* **20**, 724–730.

Reuter, M., C, S., Tanaka, T., Franz, T., Stark, A. & Pillai, R. S. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat. Struct. Mol. Biol.* **16**, 639–646.

Safran, M., D, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A. & Lancet, D. 2010. GeneCards Version 3: the human gene integrator. *Database (Oxford)* doi: 10.1093/database/baq020.

Saito, K., I, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H. & Siomi, M. C. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev.* **24**, 2493–2498.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Scadden, A. 2005. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat. Struct. Mol. Biol.* **12**, 489–496.

Selenko, P., S, R., Stier, G., Bühler, D., Fischer, U. & Sattler, M. 2001. SMN tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.* **8**, 27–31.

Siomi, M. C., M, T. & Siomi, H. 2010a. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev.* **24**, 636–646.

Siomi, M. C., M, T. & Siomi, H. 2010b. piRNA-mediated silencing in Drosophila germlines. *Semin. Cell Dev. Biol.* **21**, 754–759.

Steinhauer, W. R. & Kalfayan, L. 1992. A specific ovarian tumor protein isoform is required for efficient differentiation of germ cells in Drosophila oogenesis. *Genes Dev.* **6**, 233–243.

Sun, Y., G, M., Schwarzer, V., Schoenen, F., Fischer, U. & Wirth, B. 2005. Molecular and functional analysis of intragenic SMN1 mutations in patients with spinal muscular atrophy. *Hum. Mutat.* **25**, 64–71.

Tanaka, T., H, M., Vagin, V. V., Reuter, M., Hayashi, E., Mochizuki, A. L., Kitamura, K., Yamanaka, H., Kondoh, G., Okawa, K., Kuramochi-Miyagawa, S., Nakano, T., Sachidanandam, R., Hannon, G. J., Pillai, R. S., Nakatsuji, N. & Chuma, S. 2011. Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *Proc. Natl Acad. Sci. USA* **108**, 10579–84.

Tsuchiya, N., O, M., Nakashima, K., Ubagai, T., Sugimura, T. & Nakagama, H. 2007. SND1, a component of RNA-induced silencing complex, is up-regulated in human colon cancers and implicated in early stage colon carcinogenesis. *Cancer Res.* **67**, 9568–9576.

Van Buskirk, C. & Schüpbach, T. 2002. Half pint regulates alternative splice site selection in Drosophila. *Dev. Cell* **2**, 343–353.

Vasileva, A., T, D., Firooznia, A., Müller-Reichert, T. & Jessberger, R. 2009. Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr. Biol.* **19**, 630–639.

Wang, H., M, Y. & Chen, D. 2011. Effective gene silencing in Drosophila ovarian germline by artificial microRNAs. *Cell Res.* **21**, 700–703.

Wu, M. Y., E, K. & Beaudet, A. L. 2008. Identification of chromatin remodeling genes Arid4a and Arid4b as leukemia suppressor genes. *J. Natl Cancer Inst.* **100**, 1247–1259.

Wu, M. Y., T, T. & Beaudet, A. L. 2006. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. *Genes Dev.* **20**, 2859–2870.

Yabuta, Y., O, H., Abe, T., Kurimoto, K., Chuma, S. & Saitou, M. 2011. TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *J. Cell Biol.* **192**, 781–795.

Yang, J., V, T., Hong, J., Bu, T., Yao, Z., Jensen, O. N., Frilander, M. J. & Silvennoinen, O. 2007a. Transcriptional co-activator protein p100 interacts with snRNP proteins and facilitates the assembly of the spliceosome. *Nucleic Acids Res.* **35**, 4485–4494.

Yang, L., C, D., Duan, R., Xia, L., Wang, J., Qurashi, A., Jin, P. & Chen, D. 2007b. Argonaute 1 regulates the fate of germline stem cells in Drosophila. *Development* **134**, 4265–4272.

Yang, L., D, R., Chen, D., Wang, J., Chen, D. & Jin, P. 2007c. Fragile X mental retardation protein modulates the fate of germline stem cells in Drosophila. *Hum. Mol. Genet.* **16**, 1814–1820.

Yang, Y., L, Y., Espejo, A., Wu, J., Xu, W., Liang, S. & Bedford, M. T. 2010. TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol. Cell* **40**, 1016–1023.

Yang, Y., X, S., Xia, L., Wang, J., Wen, S., Jin, P. & Chen, D. 2009. The bantam microRNA is associated with drosophila fragile × mental retardation protein and regulates the fate of germline stem cells. *PLoS Genet.* **5**, e1000444.

Ying, M., H, X., Zhao, H., Wu, Y., Wan, F., Huang, C. & Jie, K. 2011. Comprehensively surveying structure and function of RING domains from *Drosophila melanogaster*. *PLoS One* **6**, e23863.

Ying, M., Z, Z., Wang, W. & Chen, D. 2009. Origin and evolution of ubiquitin-conjugating enzymes from *Guillardia theta* nucleomorph to hominoid. *Gene* **447**, 72–85.

Yoo, B. K., S, P., Gredler, R., Chen, D., Emdad, L., Bhutia, S., Pannell, L., Fisher, P. B. & Sarkar, D. 2011. Increased RNA-induced silencing complex (RISC) activity contributes to hepatocellular carcinoma. *Hepatology* **53**, 1538–1548.

Zamparini, A. L., Davis, M. Y., Malone, C. D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G. J. & Lehmann, R. 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* **138**, 4039–50.

Zheng, J., L, J., Liu, H., Li, J. & Chen, K. 2009. Sequence and structural analysis of 4SNc-Tudor domain protein from *Takifugu rubripes*. *Bioinformation* **4**, 127–131.

## Reference database

1. EBI (http://www.ebi.ac.uk/).
2. Ensemble (http://www.ensembl.org/index.html).
3. ExPASy (http://www.expasy.ch/).
4. FlyBase (http://flybase.org/).
5. Genecards (http://www.genecards.org).
6. NCBI (http://www.ncbi.nlm.nih.gov/).
7. NetSurfP (http://cbs.dtu.dk/services/).
8. PFAM (http://pfam.sanger.ac.uk/).
9. SMART (http://smart.embl-heidelberg.de/).
10. UCSC (http://genome.ucsc.edu/).
11. UniProt (http://www.uniprot.org/).

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** Analysis of sequences flanking the canonical Tudor domain core.

**Fig. S2.** Domain architectures of Tudor-domain proteins from *Drosophila melanogaster*.

**Fig. S3.** A wheel constructed on phylogenetic analysis of Tudor domains on the basis of 472 Tudor domain sequences from 17 organisms (*Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus, Danio rerio*, *Xenopus tropicalis*, *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Caenorhabditis elegans*, *Monosiga brevicollis*, *Dictyostelium discoideum*, *Arabidopsis thaliana*, *Oryza sativa*, *Schizosaccharomyces pombe*, *Schizosaccharomyces japonicus yFS275*).

**Table S1.** Estimated numbers of different species Tudor-containing genes and the corresponding mRNA numbers.

**Table S2.** Information related to Tudor-containing domain proteins from human, and its corresponding putative orthologues from *Pan troglodytes, Mus musculus, Rattus norvegicus, Danio rerio, Xenopus tropicalis, Drosophila melanogaster, Anopheles gambiae, Apis mellifera, Caenorhabditis elegans, Monosiga brevicollis, Dictyostelium discoideum, Arabidopsis thaliana, Oryza sativa, Schizosaccharomyces pombe, Schizosaccharomyces japonicus* yFS275.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.