BIOINFORMATICS DISCOVERY NOTE Vol. 26 no. 9 2010, pages 1140–1144
doi:10.1093/bioinformatics/btq122

Sequence analysis

Advance Access publication March 19, 2010

LOTUS, a new domain associated with small RNA pathways in the germline

Isabelle Callebaut[1,2,3,*] and Jean-Paul Mornon[1,2,3]

[1]IMPMC-UMR7590, Université Pierre et Marie Curie-Paris 6, Paris F-75005, [2]IMPMC-UMR7590, CNRS, Paris F-75016 and [3]IMPMC-UMR7590, Université Denis Diderot-Paris 7, Paris F-75013, France

Associate Editor: Alex Bateman

ABSTRACT

We describe here LOTUS, a hitherto uncharacterized small globular domain, which was identified using sensitive sequence profile analysis. The LOTUS domain is found in germline-specific proteins that are present in the nuage/polar granules of germ cells. TDRD5 and TDRD7, two mammalian members of the germline Tudor group, possess three copies of the LOTUS domain in their extreme N-termini. The Tudor domains of these proteins bind symmetric dimethyl arginines present on the germ cell-specific Piwi proteins, which form a particular clade of Argonaute proteins. Piwi proteins interact with a specific class of non-coding RNAs [piwi-interacting RNAs (piRNAs)] and play a key role in the repression (silencing) of transposons and possibly other germline-specific functions. A LOTUS domain is also present in the Oskar protein, a critical component of the pole plasm in the *Drosophila* oocyte, which is required for germ cell formation. LOTUS domains are found in various proteins from metazoans and plants, are often associated with RNA-specific modules and are likely to adopt a winged helix fold. This suggests a germline-specific role in the mRNA localization and/or translation or a specific function toward piRNAs.

Contact: isabelle.callebaut@impmc.upmc.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on February 1, 2010; revised on March 16, 2010; accepted on March 17, 2010

1 INTRODUCTION

The Tudor domain, the name of which originates from its identification in the *Drosophila* Tudor protein (Callebaut and Mornon, 1997; Ponting, 1997), is a small domain belonging to the large 'Royal Family' that also includes Chromo, MBT, PWWP and plant Agenet domains (Maurer-Stroh *et al.*, 2003). Tudor domains are specialized in the specific recognition of methylated lysine or methylated arginine sites, and are involved in various epigenetic functions, such as chromatin remodeling and RNA processing. Based on their sequences and functions, four groups of Tudor domains were recently distinguished (Jin *et al.*, 2009). The first group, including the JMDJ family of histone demethylases, as well as the related 53BP1 and Crb2 proteins, participates in the histone code deciphering, as their tandems of Tudor domains bind methylated lysines of histone tails (Botuyan *et al.*, 2006; Huang *et al.*, 2006). Binding of symmetrically dimethylated arginines (sDMAs) is a property shared by the two following groups that include proteins involved in the RNA metabolism (Côté and Richard, 2005; Friberg *et al.*, 2009). The second group includes the Tudor domain of the SMN protein (the product of the survival of motor neuron gene), which interacts with the spliceosomal Sm proteins and regulates the mRNA splicing machinery (Selenko *et al.*, 2001). The third group, typified by the Tudor domain of SND1 (also called Tudor-SN or p100), a component of the RNA-induced silencing complex, is involved in the processing of small non-coding RNAs (Caudy *et al.*, 2003). The fourth group contrasts with the others that include widely expressed proteins. It contains proteins highly enriched or selectively expressed in germ cells and involved in the formation of polar/germinal granules or nuage, those electron-dense organelles, abundant in RNA and proteins and characteristic of the germline in many organisms. These proteins are crucial for spermatogenesis and small RNA pathways (Chuma *et al.*, 2006; Vasileva *et al.*, 2009). Several recent studies showed that the Tudor domains of proteins of this group, called the germline Tudor proteins, bind sDMAs present on the germ cell-specific Piwi proteins. These last proteins form a particular clade of Argonaute proteins, interacting with a specific class of non-coding RNAs [piwi-interacting RNAs (piRNAs)]. piRNAs–Piwi complexes are likely to have diverse roles in the germline (Chen *et al.*, 2009; Kirino *et al.*, 2010; Nishida *et al.*, 2009; Vagin *et al.*, 2009b; Wang *et al.*, 2009). They function in gene silencing, probably through a mechanism analogous to RNA interference by inducing the cleavage of gene transcripts [transposon silencing (Saito *et al.*, 2006)] or directly at the transcriptional level (Yin and Lin, 2007). Moreover, they were shown in *Drosophila* to regulate mRNA stability and translation, a function which thereby contrasts with the only negative effect of small interfering RNAs (siRNAs) and micro-RNAs (miRNAs) (Li, 2007). Arginine methylation of Piwis and interaction with Tudors seems to be a conserved feature of the Piwi pathway that distinguishes it from the somatic Ago pathway [the other clade of Argonaute proteins that bind to miRNAs and small interfering RNA (siRNA) and are ubiquitously expressed]. Thus, this specific interaction could be required for Piwi recruitment to nuage granules (Vagin *et al.*, 2009a), and germline Tudor proteins might provide the platform necessary for assembly of numerous proteins that function as a complex in the Piwi pathway (Vagin *et al.*, 2009a). This pathway is essential, in *Drosophila*, for germ cell determination and maintenance of germline stem cells and, in mice, for postnatal spermatogenesis [reviewed in Chuma *et al.* (2009)].

*To whom correspondence should be addressed.

In a survey of Tudor domain-containing proteins, we focused on these germline Tudor proteins. We identified multiple copies of a new domain in some mammalian members (TDRD5 and TDRD7), which is also present in other proteins with different domain architectures. We thus describe here this short, novel domain, that we called LOTUS (after Limkain, Oskar and TUdor-containing proteins 5 and 7).

2 METHODS

The PSI-BLAST program (BLAST 2.2.22+, default parameters) (Altschul *et al.*, 1997) was used to search for similarities against the non-redundant database (nr database; January 2010, 10 381 056 sequences) at NCBI (http://www.ncbi.nlm.nih.gov/blast). The *E*-value threshold was 0.005 for inclusion of sequences into a profile. Sequences were clustered with BLASTCLUST followed by manual sorting. Low sequence identity and length coverage threshold values were used in order to define only a few numbers of clusters, from which representative sequences were used as queries for further PSI-BLAST iterations.

Hydrophobic cluster analysis (HCA) was used to predict the position of globular domains in protein sequences, as well as to refine and extend the sequence alignments, as reported in the PSI-BLAST results. Guidelines to the use of this methodology can be found in (Callebaut *et al.*, 1997; Eudes *et al.*, 2007). The Smart (http://smart.embl-heidelberg.de/; Letunic *et al.*, 2009) and Pfam (http://pfam.sanger.ac.uk/; Finn *et al.*, 2010) domain databases, as well as the CDD collection (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Marchler-Bauer *et al.*, 2009) were searched using HMMSearch and RPS-BLAST, respectively, using default threshold values (e.g. for Pfam search; *E*-value = 0.1). Higher *E*-values were also considered in order not to miss potential interesting similarities, which could be further explored.

Multiple alignment of the selected sequences was constructed using MAFFT (Katoh and Toh, 2008), and manually adjusted with guidance from HCA. This multiple alignment was considered to predict protein secondary structures using JPRED (http://www.compbio.dundee.ac.uk/www-jpred/) (Cole *et al.*, 2008) and PSI-PRED (http://bioinf4.cs.ucl.ac.uk:3000/psipred) (McGuffin *et al.*, 2000), and to perform fold recognition using Phyre (http://www.sbg.bio.ic.ac.uk/phyre/) (Bennett-Lovsey *et al.*, 2008) or other methods, such as HHPred (Söding *et al.*, 2005) and MUSTER (http://zhanglab.ccmb.med.umich.edu/MUSTER/) (Wu and Zhang, 2008). It was also used in additional searches using the HMMER3 package (Eddy, 2008) with default parameters, on the Mobyle system (http://mobyle.pasteur.fr) (Néron *et al.*, 2009).

3 RESULTS

Examination of the sequences of various members of the germline Tudor proteins (called TDRDs) shows that globular region(s) are present upstream of the repeated Tudor domains. Of note, in the TDRD5 and TDRD7 proteins, a long stretch of ~400 amino acids is present at the extreme N-terminus and should include globular domain(s), as predicted from the analysis of HCA plots (see Supplementary Material 1). Indeed, this region includes fragments containing approximately one-third of strong hydrophobic amino acids gathered into clusters typical of regular secondary structures. However, it does not match any known profiles in the domain databases. Clear hinge regions can be predicted on these plots, corresponding to proline-rich regions and/or regions containing less hydrophobic amino acids (arrows in Supplementary Material 1; TDRD7: approximately amino acids 100–160 and 315–335; TDRD5: approximately amino acids 100–130 and 225–255), and which are far less conserved when comparing the TDRD5 and TDRD7 sequences. This suggests the presence of three separate globular domains in TDRD5 and TDRD7 (see Supplementary Material 1; TDRD7 domain 1: amino acid 1 to approximately 100 or 115, domain 2: amino acids 160–315 and domain 3: amino acids 335–420; TDRD5 domain 1: amino acid 1 to approximately 100, domain 2: amino acids 130–225 and domain 3: amino acids 255–410). We then performed PSI-BLAST searches against the nr database, using the first predicted domain of TDRD7 as query (ending either at amino acid 100 or amino acid 115). The first iteration identified similarities of TDRD7 with the TDRD5 first domain, while subsequent iterations identified, with significant *E*-values ($E < 0.005$), similarities with the *Drosophila* Oskar protein (second iteration, $E = 7 \times 10^{-7}$) and with various other uncharacterized proteins, several of which possess RNA-binding domains (e.g. K homology (KH) domain-containing protein from *Cryptosporidium muris*, ninth iteration; $E = 0.003$). Multiple copies of this domain were also found in limkain b1, a human autoantigen whose function and partners have not been reported to date (Dunster *et al.*, 2005). This domain was also found in a hypothetical orphan protein (D1081.7) from *Caenorhabditis elegans*, which apparently does not possess Tudor domains, but is reported in the STRING database (Jensen *et al.*, 2009) to be co-expressed with the germ cell-expressed R06C7.1 protein, a member of the Argonaute family. In this way, by convergence (iteration 14), we collected ~300 sequences of proteins possessing this domain, called LOTUS after the name of some representative members (Limkain, Oskar and TUdor-containing proteins 5 and 7). The LOTUS domain is present in various proteins of metazoans (including echinoderms, arthropods and nematodes) and plants (see Fig. 1 for the architecture in domains of main representative members of the LOTUS family).

The amino acid conservation profile of the LOTUS domain allowed the more precise prediction of its C-terminal end, defined at amino acid 96 of human TDRD7 and amino acid 100 of human TDRD5. Moreover and interestingly, along the PSI-BLAST iterations, a triplication of the LOTUS domain was highlighted in the N-terminus of the TDRD5 and TDRD7 proteins, covering the three domains described above [e.g. match with mouse TDRD5 amino acids 1–99 (30% identity; E-value $= 5 \times 10^{-21}$), amino acids 293–367 (17% identity; E-value $= 4 \times 10^{-5}$) and amino acids 130–197 (17% identity; E-value $= 1 \times 10^{-3}$)]. These repeats were further assessed at the 2D level, using HCA (see Supplementary Material 1). A more precise analysis of the LOTUS repeats in these proteins suggests that domains 1 and 2 of TDRD7 are homologous to domains 1 and 3 of TDRD5, the third domain of TDRD7 and the second one of TDRD5 likely resulting from a later duplication event. Additional hits were identified by searching the nr database using PSI-BLAST and representative members of the LOTUS family as queries, as well as the second and third LOTUS domain of the TDRD7 protein. A few borderline similarities [*E*-values >0.001; for example, *Oryza sativa* RSSG8 (UniProt Q94CF9_ORYSA) or other uncharacterized proteins from plants] were also confirmed through reciprocal PSI-BLAST searches.

Multiple alignment of the LOTUS domain was made, on the basis of the PSI-BLAST alignments and manually refined with guidance from HCA, which combines comparison of the primary and secondary structures (Fig. 2; Supplementary Material 1). Additional searches were also performed using this multiple alignment with HMMSEARCH, and these led to recover a few new, but also uncharacterized proteins (e.g. LRR2_PLAFA, containing one LOTUS domain between amino acids 1265 and 1360; see

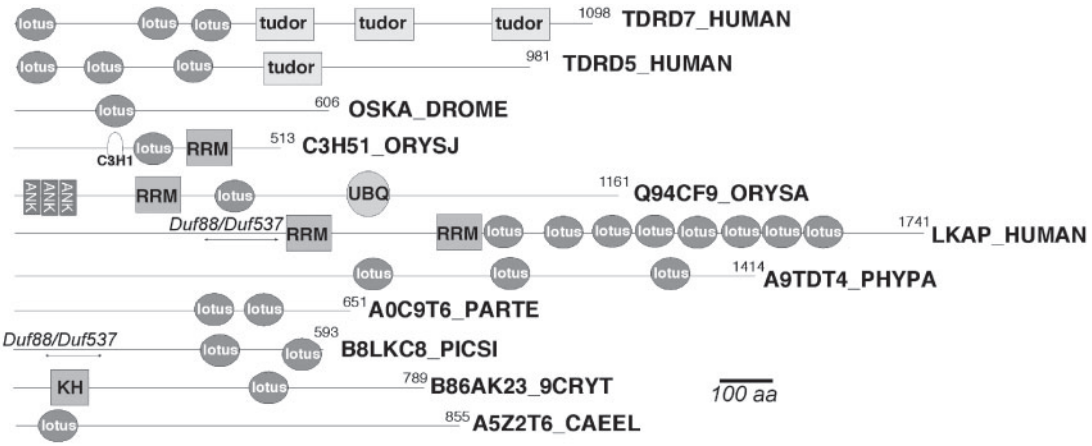

Fig. 1. Domain organization of some LOTUS proteins, as deduced from the SMART (Letunic *et al*., 2009), Pfam databases (Finn *et al*., 2010) and the present study. UniProt identifiers are indicated. Tudor, Tudor domain; RRM, RNA recognition motif; ANK, ankyrin; UBQ, ubiquitin homolog; Duf88 (PF01936) and Duf537 (PF04396) correspond to Pfam families with unknown function, which have been merged into a unique superfamily in the SUPFAM database (PNSF no. 066) (Pandit *et al*., 2004). HUMAN, *Homo sapiens*; DROME, *Drosophila melanogaster*; ORYSJ, *Oryza sativa japonica group*; ORYSA, *Oryza sativa*; PHYPA, *Physcomitrella patens* spp. *patens*; PARTE, *Paramecium tetraurelia*; PICSI, *Picea sitchensis*; 9CRYT, *Cryptosporidium muris RN66*; and CAEEL, *Caenorhabditis elegans*. A domain sharing similarities with the GDSL family of serine esterases/lipases is present C-terminal to the LOTUS domain in the Oskar protein, although not reported in the domain databases.

Supplementary Material 2 for a complete list). Conserved positions within the alignment are principally occupied by strong hydrophobic amino acids that should correspond to core-forming residues (Fig. 2). Outside these, only a few other positions are more conserved, occupied by proline or glycine residues (in loops) or by an aromatic residue (in helix α2). This one might play a role in the interaction of the LOTUS domain with its patrner(s). The LOTUS globular domain is predicted to contain three main helices (α1 to α3), followed by two β-strands (β1 and β2) (Fig. 2). Two additional small β-strands may be suspected on the basis of the HCA plot analysis (see β′ and β″ in Supplementary Material 1 and Fig. 2). Two additional hydrophobic clusters, which would be important for correct folding, are also present at the N-terminal and C-terminal limits of the LOTUS domain (β_N and α_C, e.g. in the C-terminus of human TDRD7 LOTUS domain a and b, as well in the in the C-terminus of *Drosophila* Oskar LOTUS domain, see Supplementary Material 1). However, these are not very conserved and the a_C cluster is even absent in some sequences, as for instance in the limkain b1 repeats or in the second LOTUS domain of the uncharacterized protein from *Pichia sitchensis* (B8LKC8) (Fig. 2). Fold recognition programs, such as Phyre, indicate marginal but interesting similarities with several families within the winged helix superfamily [SCOP a.4.5; e.g. with the ScpB winged helix domain (WHD, pdb 1t6s) (Kim *et al.*, 2006, 2008); *E*-value 5.2]. These similarities are supported by a perfect correspondence between secondary structures, suggesting that the LOTUS domain may actually adopt a WHD fold, also made of three consecutive α helices and a small two-stranded β-sheet. The additional small β′ strand predicted in the LOTUS domain may integrate this β-sheet, as observed in the ScpB WHD2 (see Supplementary Material 3). Of note is that the last helix (α4) found in the ScpB WHDs after the two-stranded β-sheet consists of a linker region that does not directly participate in the WHD core (see Supplementary Material 3). The variability or even absence of this helix (called a_C) in some LOTUS domains is thus consistent with its particular, non-core-forming role in the ScpB structure, and suggests that helix α_C may also play a linker role in proteins of the LOTUS family.

4 DISCUSSION

Various mammalian TDRD proteins have been reported to interact with Piwi proteins (Chen *et al.*, 2009; Vagin *et al.*, 2009b; Wang *et al.*, 2009). The LOTUS domain described here is, however, unique to the TDRD5 and TDRD7 proteins. Although having repeats of Tudor domains, the other TDRDs indeed differ in their extreme N-termini, which include either MYND domain (TDRD1), KH domains (TDRD2) or helicase domain (TDRD9). TDRD7 interacts more specifically with Miwi through its Tudor domains (Chen *et al.*, 2009; Vagin *et al.*, 2009b). In *Drosophila*, Aubergine (Aub) and AGO3, two members of the Piwi family, interact with the Tudor protein (Kirino *et al.*, 2010; Nishida *et al.*, 2009), which has apparently no LOTUS domain in its N-terminal fragment. We, however, show here that a LOTUS domain is found in the *Drosophila* Oskar protein, whose localized activity (through mRNA localization and subsequent synthesis) in the nuage at the posterior pole of the oocyte, termed germ or pole plasm, induces germline formation and correct patterning in the fly embryo (Ephrussi and Lehmann, 1992; Ephrussi *et al.*, 1991; Kim-Ha *et al.*, 1991; Lehmann and Nüsslein-Volhard, 1986). The localized Oskar protein has a critical function in the biogenesis of polar granules (Lehmann and Nusslein-Volhard, 1986) and recruits downstream components of the pole plasm, such as Vasa and Tudor proteins (Mahowald, 2001). It is interesting to note that, while many of the *Drosophila* germplasm components are conserved and play an essential role in mammalian germ cell formation as well (reviewed in Chuma *et al.*, 2009), the key factor in *Drosophila*, Oskar, has no known orthologs in any other species. The results reported here show that the LOTUS domain is a conserved feature of the protein complexes present in the Piwi-containing

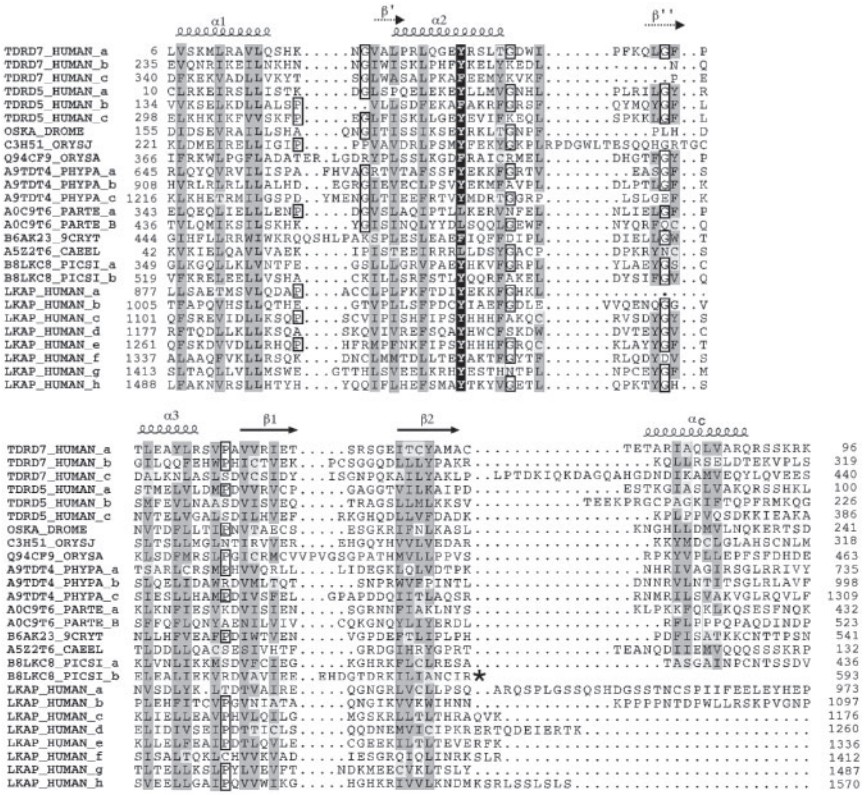

Fig. 2. Multiple sequence alignment of LOTUS domains from various proteins. Positions where hydrophobicity is conserved, which should participate in the protein core structure, are boxed and shaded gray. Other striking identities (involving glycine and proline residues) are boxed. Predicted secondary structures are reported up to the alignment. Note that the helix a_C is poorly conserved and does not exist in some of the sequences. The star indicates the protein end. The protein sequences are designated by their UniProt identifiers, and the limits (amino acid numbers) of the LOTUS domains in these sequences are reported. Note that the C-terminal end of the domain is difficult to predict with accuracy, due to the extreme variability in this region. The FASTA-formatted file of this alignment is given in the Supplementary Material 4. This figure was drawn using ESPript (http://espript.ibcp.fr/).

germline-specific granules in different organisms. LOTUS domains are indeed present, in mice, in germline TDRDs and in *Drosophila*, in the long and short isoforms of Oskar that directly interact with the Tudor protein. In that respect, TDRD5 and/or TDRD7 might represent, at least in part, the functional homolog of Oskar in mammals.

The exact molecular role of the LOTUS domain remains to be discovered. Its occurrence in proteins having domains associated with RNA metabolism (RNA recognition motif and KH domains) suggests that it might be involved in a RNA-binding function. This hypothesis is supported by the presence of several basic residues, as well as fold recognition programs, which indicate that the LOTUS domain may adopt a WHD fold. Possible RNA partners are piRNAs brought by the TDRD5/7–Piwi complexes or the numerous mRNA present in the nuage granules. The RNA-binding function might then be involved in the regulation of mRNA translation and/or localization or of a specific function in which piRNAs participate.

The identification of this new domain in the mouse germline Tudor proteins, TDRD5 and TDRD7, and in the *Drosophila* Oskar protein should help to understand the function of these proteins in the piRNA pathways and germline determination, as well as to uncover the function of as yet uncharacterized proteins in this context.

ACKNOWLEDGEMENTS

We thank the editor and the anonymous reviewers for their helpful comments and suggestions.

Funding: Institutional grants from the Centre National de la Recherche Scientifique (CNRS) and the Université Pierre et Marie Curie (UPMC); grant from the the Institut National du Cancer/Canceropôle Ile de France.

Conflict of interest: none declared.

REFERENCES

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Bennett-Lovsey,R.M. *et al.* (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, **70**, 611–625.

Botuyan,M.V. *et al.* (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell*, **127**, 1361–1373.

Callebaut,I. and Mornon,J.-P. (1997) The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in Drosophila melanogaster development. *Biochem. J.*, **321**, 125–132.

Callebaut,I. *et al.* (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci.*, **53**, 621–645.

Caudy,A.A. *et al.* (2003) A micrococcal nuclease homologue in RNAi effector complexes. *Nature*, **425**, 411–414.

Chen,C. *et al.* (2009) Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc. Natl Acad. Sci. USA*, **106**, 20336–20341.

Chuma,S. *et al.* (2006) Tdrd1/Mtr-1, a tudor-related gene, is essential for male germ-cell differentiation and nuage/germinal granule formation in mice. *Proc. Natl Acad. Sci. USA*, **103**, 15894–15899.

Chuma,S. *et al.* (2009) Ultrastructural characterization of spermatogenesis and its evolutionary conservation in the germline: germinal granules in mammals. *Mol. Cell Endocrinol.*, 17–23.

Cole,C. *et al.* (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.

Côté,J. and Richard,S. (2005) Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.*, **280**, 28476–28483.

Dunster,K. *et al.* (2005) Limkain b1, a novel human autoantigen localized to a subset of ABCD3 and PXF marked peroxisomes. *Clin. Exp. Immunol.*, **140**, 556–563.

Eddy,S.R. (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput. Biol.*, **4**, e1000069.

Ephrussi,A. and Lehmann,R. (1992) Induction of germ cell formation by oskar. *Nature*, **358**, 387–392.

Ephrussi,A. *et al.* (1991) Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*, **66**, 37–50.

Eudes,R. *et al.* (2007) A generalized analysis of hydrophobic and loop clusters within globular protein sequences. *BMC Struct. Biol.*, **7**, 2.

Finn,R. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.

Friberg,A. *et al.* (2009) Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN. *J. Mol. Biol.*, **387**, 921–934.

Huang,Y. *et al.* (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science*, **312**, 748–751.

Jensen,L.J. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.

Jin,J. *et al.* (2009) Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.*, **2**, ra76.

Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.*, **9**, 286–298.

Kim,J.S. *et al.* (2006) Crystal structure of ScpB from Chlorobium tepidum, a protein involved in chromosome partitioning. *Proteins*, **62**, 322–328.

Kim,J.S. *et al.* (2008) Crystal structure and domain characterization of ScpB from Mycobacterium tuberculosis. *Proteins*, **71**, 1553–1556.

Kim-Ha,J. *et al.* (1991) Oskar mRNA is localized to the posterior pole of the Drosophila oocyte. *Cell*, **66**, 23–35.

Kirino,Y. *et al.* (2010) Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA*, **16**, 70–78.

Lehmann,R. and Nüsslein-Volhard,C. (1986) Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. *Cell*, **47**, 141–152.

Letunic,I. *et al.* (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.*, **37**, D229–D232.

Li,H. (2007) piRNAs in the germ line. *Science*, **316**, 397.

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

Marchler-Bauer,A. *et al.* (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.*, **37**, D205–D210.

Maurer-Stroh,S. *et al.* (2003) The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.*, **28**, 69–74.

McGuffin,L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.

Néron,B. *et al.* (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005–3011.

Nishida,K.M. *et al.* (2009) Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *EMBO J.*, **28**, 3820–3831.

Pandit,S.B. *et al.* (2004) SUPFAM: a database of sequence superfamilies of protein domains. *BMC Bioinformatics*, **15**, 5–28.

Ponting,C.P. (1997) Tudor domains in proteins that interact with RNA. *Trends Biochem. Sci.*, **22**, 51–52.

Saito,K. *et al.* (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev.*, **20**, 2214–2222.

Selenko,P. *et al.* (2001) SMN tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.*, **8**, 27–31.

Söding,J. *et al.* (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.*, **33**, W244–W248.

Vagin,V.V. *et al.* (2009a) Arginine methylation as a molecular signature of the Piwi small RNA pathway. *Cell Cycle*, **8**, 4003–4004.

Vagin,V.V. *et al.* (2009b) Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.*, **23**, 1749–1762.

Vasileva,A. *et al.* (2009) Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. *Curr. Biol.*, **19**, 630–639.

Wang,J. *et al.* (2009) Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr. Biol.*, **19**, 640–644.

Wu,S. and Zhang,Y. (2008) MUSTER: improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins*, **72**, 547–556

Yin,H. and Lin,H. (2007) An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature*, **450**, 304–309.