Supporting Information

Chen et al. 10.1073/pnas.0911640106

SI Text

Sample Preparation for Mass Spectrometry. Silica tubing (200 μm, Polymicro Technologies) was sealed at one end by Kasil solution. After methenal wash, column (3 cm) was packed with SCX resin and rinsed with 10 mM KH_2PO_4 buffer pH 3.0. Samples of eluted immunoprecipitates were loaded onto column by air vessel at 600 psi. After washing with KH_2PO_4 buffer and H_2O, samples were incubated with 2 μl of 100 mM DTT in 10 mM NH_4HCO_3 pH 8.0 for 30 min and followed by a wash with H_2O. Proteins are digested with trypsin solution [2 mg/ml trypsin (Promega) in 100 mM Tris-HCl pH8.0 and 10 mM iodoacetamide] on column for 1 h at RT. Peptides were eluted using 200 mM NH_4HCO_3 and acidified with formic acid for mass spectrometry analysis.

Mass Spectrometry. All experiments were performed on a QSTAR Elite QqTOF mass spectrometer equipped with a nanospray III ion source (AB/MDS Sciex), coupled to an Eksigent 1D+ Nano LC. Samples were injected directly onto home-built "packed-tip" columns. An approximatley 4 μm tip was pulled at the end of a 25 cm piece of 360 μm od $\times$ 75 μm id fused silica tubing using a Model 2000 Micropipette Laser Puller (Sutton Instrument Co.). A methanol slurry of 3 μm ReprosilGold 120 C-18 stationary phase media was prepared in a 1.5 ml glass vial and then placed into a pressure vessel (pressure bomb); 120 mm of stationary phase was packed into the 360 μm od $\times$ 75 μm id fused silica pulled-tip using the pressure vessel set at 1500 psi.

Samples were run using a 90 min gradient from 5%–30% solvent B (Solvent A: water with 0.1% formic acid; Solvent B: acetonitrile with 0.1% formic acid) at a flow rate of 250 nl/min. An Information Dependent Acquisition method was set up with the MS survey mass range set between 250-1200 amu, and dependent MS/MS scans with mass range set between 45–1600 amu. Note the low-end mass range in the MS/MS was set to allow detection of diagnostic immonium fragment ions at 46 amu and 71 amu produced from peptides containing dimethyl-arginine modified amino acids (1).

LC-MS/MS raw data produced on the QSTAR Elite was searched using Mascot v2.2. The search parameters were set as follows: database: ENSEMBL Mus musculus revision 49; enzyme: Trypsin with 4 missed cleavages; fixed modifications: Carbamidomethyl C; variable modifictions: monomethyl R, dimethyl R, deamidation NQ, pyro-Glu QE, and oxidation M. Peptide tolerance was set to 70 ppm, MS/MS tolerance was set to 0.15 amu, and the instrument selected was ESI-Quad-TOF.

For estimation of relative amounts of proteins, total spectra counts for each protein identified were used. Data were clustered simultaneously for both protein ID (gene symbol) and IP sample names by average correlation-based distance calculation using the BioPython modules (biopython.org). The heat maps were generated using the ReportLab Python module (reportlab.org). The network diagrams were initially generated with the NAViGaTOR software (ophid.utoronto.ca/navigator), and redrawn in PDF format using the ReporLab module.

Cloning, Expression and Purification of Recombinant Tdrkh. DNA sequence corresponding to amino acids 327 to 420 of human Tdrkh was cloned into the pET-28a-MHL vector via ligase-independent cloning. Recombinant His-6-TEV-Tudor was expressed in a SGC-generated derivative strain of BL21 *E. coli* (Structural and Genomic Consortium, Toronto, Canada) with the pRARE plasmid for codon biased expression. Cells were grown in Terrific Broth (TB) in the presence of kanamycin and chloramphenicol at 37 °C to an optical density of approximately 2.5. Protein expression was induced with 1 mM isopropyl-1-thio-D-galactopuranoside and the cell cultures continued for approximatley 16 h at 15 °C. The cells were harvested via centrifugation and the resultant pellet stored at −80 °C before purification.

The cell pellet from a 2-liter culture was resuspended in 200 ml of lysis buffer consisting of 1 $\times$ PBS, 250 mM NaCl, 5% glycerol, and 0.1 μM phenylmethyl sulfonyl fluoride. The homogenized suspension was lysed via sonication and insoluble material removed via centrifugation. The clarified supernatant was passed >5 ml of Ni-NTA resin (Qiagen) that had been preequilibrated with 20 mM Tris-HCl pH 8.0, 250 mM NaCl. The resin was washed with 10 column volumes of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, and protein eluted in 15 ml of 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 250 mM imidazole. Further purity was achieved via gradient elution (0 to 1M NaCl) from a HiTrap Q HP anion exchange column (GE Healthcare).

Purified protein was then incubated with TEV protease (Invitrogen) overnight at 4 °C. The Tudor domain was isolated via passage over 1 ml of Ni-NTA resin (Qiagen) and was further purified via a Superdex 75 size exclusion column in 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM DTT. As assessed via SDS/PAGE, fractions containing pure recombinant protein were pooled and concentrated to 20 mg/ml in 20 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM DTT.

Crystallization and Structure Determination. Crystals of excellent diffraction quality were grown at 18 °C using the sitting drop method by mixing an equal volume of protein solution with reservoir solution containing 30% (wt/vol) polyethylene glycol 4000, 0.2 M ammonium acetate, 0.1 M sodium citrate pH 5.6. Crystals were then soaked in the crystallization solution supplemented with 25% ethylene glycol and frozen in liquid nitrogen.

Collection of x-ray diffraction data were also carried out at 100 K using the in-house Rigaku FRE rotating anode with MAR image plate. Intensities were integrated and scaled using HKL2000. The structure was determined by molecular replacement in PHASER (2) using the previously determined NMR structure of Tdrkh as the search model (PDB 2DIQ). Model building and refinement were carried out in COOT (3) with Refmac5 (4), respectively.

Docking simulations using modified GRG peptides were carried out in AutoDock 4.2 with AutoDockTool-1.5.4 (5). Flexibility of the ligand binding site residues was permitted and the docking carried out using the genetic algorithm. Default docking parameters were used. Clusters were scored and the lowest energy conformational clusters analyzed.

1. Rappsilber J, Friesen WJ, Paushkin S, Dreyfuss G, Mann M (2003) Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. *Anal Chem* 75:3107–3114.
2. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674.
3. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60(Pt 12 Pt 1):2126–2132.
4. Morris GM, et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.*
5. Morris GM, Huey R, Olson AJ (2008) Using AutoDock for ligand-receptor docking. *Curr Protoc Bioinformatics* Chapter 8:Unit 8 14.

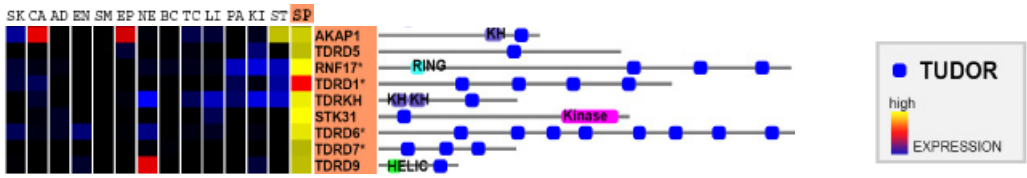

Fig. S1. Domain organization and expression of the germline Tudor proteins. Abbreviations are the following: skeletal muscle (SK), cardiac muscle (CA), adipocyte (AD), endothelium (EN), smooth muscle (SM), epithelium (EP), neuron (NE), B cell (BC), T cell (TC), liver (LI), pancreas (PA), kidney (KI), stem cell (ST) and spermatocyte (SP). Data were retrieved from human microarray repository of GEO. Domain architectures are to the right. Asterisks mark the gene encoding multiTudor domain proteins.

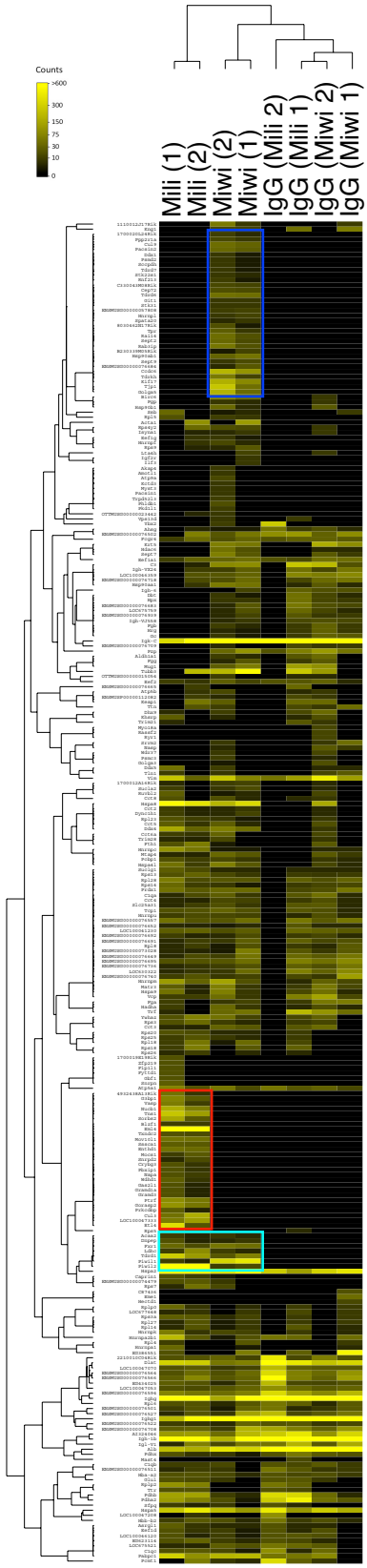

Fig. S2. Hierarchical clustering of proteins identified by mass spectrometric analysis of immunoprecipitations of Miwi, Mili, and IgG control from adult testis lysate. This is a higher resolution version of Fig. 1*A*.

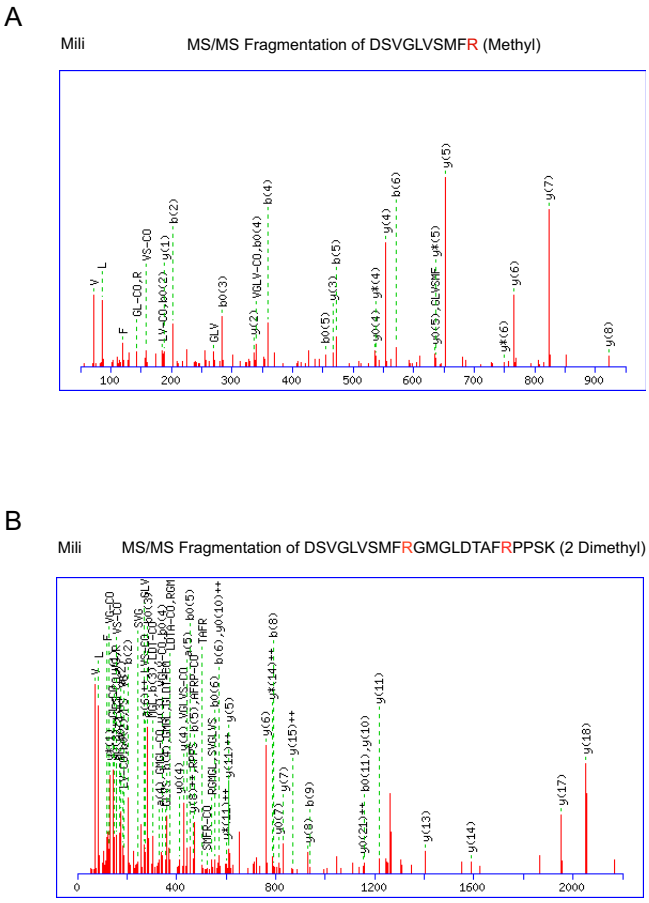

Fig. S3. Representative MS/MS spectra of methylarginine containing peptides. (*A*) R74 of Mili is monomethylated. (*B*) R74 and R83 of Mili are dimethylated.

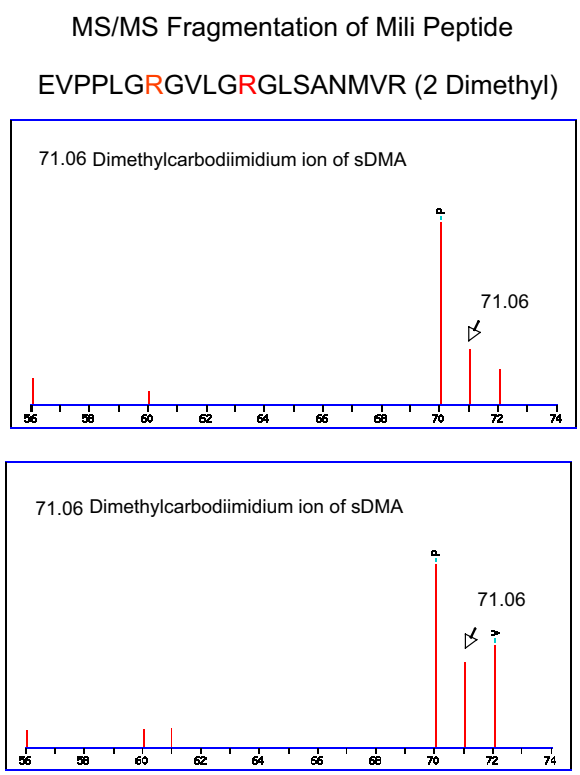

Fig. S4. Representative MS/MS spectra for sDMA peptides. Arrows indicate the identified dimethylcarbodiimidium ion of sDMA.

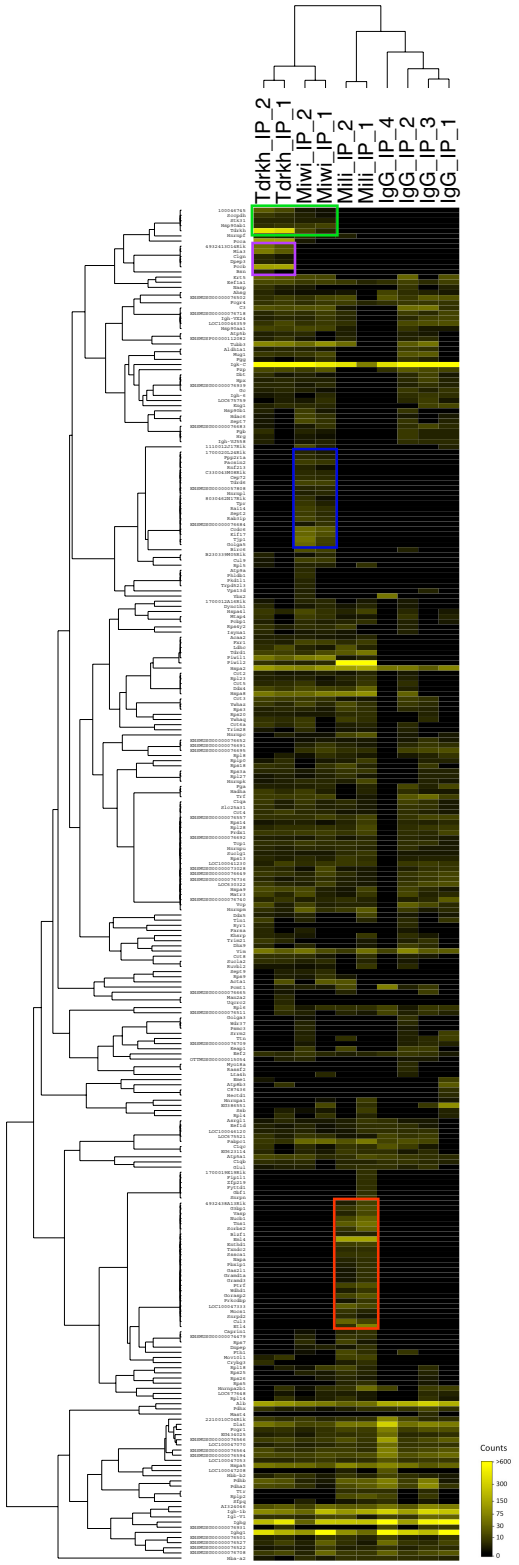

Fig. S5. Hierarchical clustering of proteins identified by mass spectrometric analysis of immunoprecipitations of Miwi, Mili, Tdrkh and IgG control from adult testis lysate. Proteins for which 3 or more peptides were identified are shown. Proteins that associate specifically with Tdrkh, Miwi or Mili are illustrated in purple, blue, and red boxes, respectively. The green box shows common proteins complexed with both Tdrkh and Miwi.

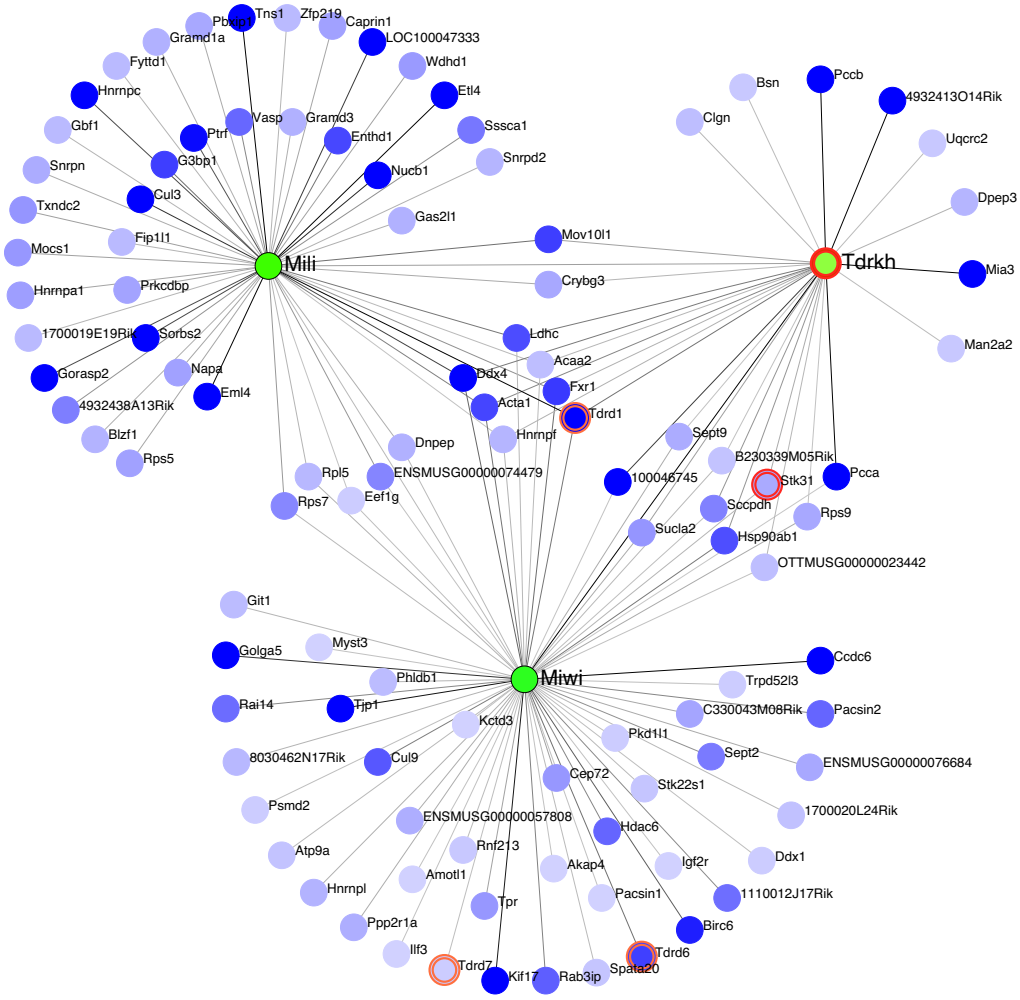

Fig. S6. Miwi, Mili, and Tdrkh protein interaction network. Baits for immunoprecipitation are illustrated in green. Interacting proteins are illustrated in blue; the peptide numbers identified by mass spectrometry correlate with the intensity of the color. The Tudor domain proteins are highlighted by red rings.

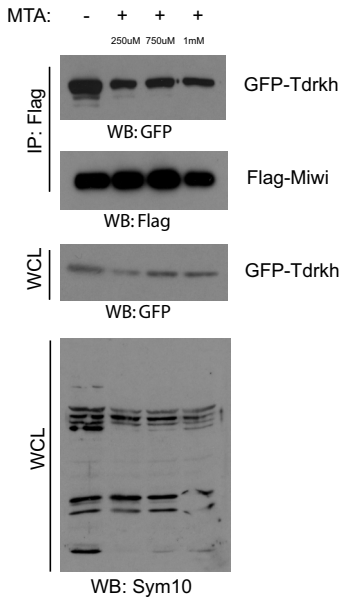

Fig. S7. Tdrkh-Miwi interaction is attenuated by methylation inhibitor treatment. 293T cells cotransfected with Flag-Miwi and GFP-Tdrkh were treated with either DMSO or various concentrations of methyltransferase inhibitor MTA and immuoprecipitated with anti-Flag agarose and immnuoblotted with anti-GFP or anti-Flag antibodies as indicated. Whole-cell lysates (WCL) were blotted with anti-GFP or Sym10 (Anti-dimethyl-arginine, symmetric) antibodies.

Table S1. Crystallographic and refinement statistics

Data Collection	
Wavelength (Å)	1.54
Space Group	$P2_12_12_1$
Unit Cell Parameters (Å)	a = 41.6, b = 45.1, c = 46.6
Resolution (Å)	40–1.74 (1.81–1.75)
Reflections	
Unique	8475 (887)
Completeness (%)	91.0 (99.8)
Redundancy	5.8 (5.3)
I/σ(I)	21.6 (14.6)
Rsym (I)[b] (%)	7.2 (17.2)
Refinement	
Resolution (Å)	40–1.74
Reflections	(1.81–1.75)
Number	8438(636)
Completeness (%)	90.9 (97.6)
Test Set (%)	5
R_{work}	0.214(0.257)
R_{free}	0.257 (0.237)
E.S.U. (Å)[c]	0.155
Contents of A.U.[d]	
Protein Molecules/Protein Atoms/Solvent Atoms	1/689/84
Mean B-Factors (Å^2)	22.2
Protein	21.3
Solvent	29.7
Ramachandran Plot (%)	
Preferred	97.6
Allowed	2.4
Outlier	0
RMSD[e] from Target Geometry	
Bond Lengths (Å)	0.019
Bond Angles (°)	1.633
PDB ID	3FDR

[a]Data for the highest resolution shell in parenthesis.
[b]$R_{sym}(I) = \Sigma_{hkl}\Sigma_i \mid I_i(hkl) - \langle I(hkl)\rangle \mid / \Sigma_{hkl}\ \Sigma_i \mid I_i\ (hkl)\mid$; for n independent reflections and I observations of a reflection; $\langle I(hkl)\rangle$-average intensity of the I observations.
[c]E.S.U., estimated overall coordinate error based on maximum likelihood.
[d]A.U., asymmetric unit.
[e]RMSD, root mean squared deviation.