ketu mutant mice uncover an essential meiotic function for the ancient, putative RNA helicase YTHDC2

Devanshi Jain[1], M. Rhyan Puno[2,4], Kathryn V. Anderson[3], and Scott Keeney[1,4] *

[1]Molecular Biology Program, [2]Structural Biology Program, [3]Developmental Biology Program, and [4]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065 USA

Running title: *Ythdc2* is essential for mouse meiosis

Keywords
Meiosis, spermatogonia, MEIOC, Bgcn, Bam, RNA binding proteins, gametogenesis

*Correspondence:

Scott Keeney
Howard Hughes Medical Institute
Molecular Biology Program
Memorial Sloan Kettering Cancer Center
1275 York Ave, Box 97
New York, NY 10065
(212) 639-5182
s-keeney@ski.mskcc.org

ABSTRACT

Mechanisms regulating mammalian meiotic progression are poorly understood. Here we identify mouse YTHDC2 as a critical component of this regulation. A screen yielded a mutant, "*ketu*", with male and female sterility caused by a *Ythdc2* missense mutation. Mutant germ cells enter meiosis but proceed prematurely to aberrant metaphase and apoptosis. *ketu* phenocopies mutants lacking MEIOC, a YTHDC2 partner. YTHDC2 is cytoplasmic and a YTH domain solution structure reveals conservation of a hydrophobic N^6-methyladenosine recognition pocket, consistent with roles in post-transcriptional regulation. Orthologs are present throughout metazoans, but are structurally diverged in nematodes and, more dramatically, in Drosophilidae, where Bgcn is a biochemically distinct descendant of a *Ythdc2* gene duplication. We also uncover similarity between MEIOC and Bam, a Bgcn partner unique to schizophoran flies. We propose that regulation of gene expression by the YTHDC2-MEIOC complex is an evolutionarily ancient strategy for controlling the germline transition from mitosis to meiosis.

INTRODUCTION

Sexual reproduction requires formation of gametes with half the genome complement of the parent organism. The specialized cell division of meiosis achieves this genome reduction by appending two rounds of chromosome segregation to one round of DNA replication (Page and Hawley, 2003). Homologous maternal and paternal chromosomes segregate in the first meiotic division, then sister centromeres separate in the second. Prior to the first division, homologous chromosomes pair and recombine to form temporary connections that stabilize the chromosomes on the metaphase I spindle (Page and Hawley, 2003; Hunter, 2007). Errors in these processes can cause gametogenic failure and infertility, or yield aneuploid gametes that in turn lead to miscarriage or birth defects in offspring (Hassold and Hunt, 2001; Sasaki et al., 2010).

In well-studied metazoan species, meiosis occurs specifically in a dedicated germ cell lineage after a period of limited expansion via mitotic "transit-amplifying" cell divisions (de Rooij, 2001; Davies and Fuller, 2008). The coordination of germline stem cell divisions with entry into meiosis and the subsequent progression of cells through meiotic divisions are tightly regulated (e.g., Griswold, 2016), but mechanisms underlying this regulation are not fully understood, particularly in mammals. And, more generally, the catalog of mammalian genes required for germ cell development, meiosis, and gametogenesis remains incomplete. In efforts to overcome this lack, we carried out a phenotype-based, random chemical mutagenesis screen to identify novel mouse meiotic mutants. One hit was a male-sterile mutant we named *rahu*, for "recombination-affected with hypogonadism from under-populated testes (Jain et al., 2017). This mutant is defective for the function of a rodent-specific DNA methyltransferase paralog, DNMT3C. Here, we describe a new mutant that we named *ketu*. Ketu is the partner of Rahu in Vedic mythology.

ketu is a missense mutation in *Ythdc2* (YTH-domain containing 2), which encodes a putative RNA helicase with a YT521-B homology (YTH) RNA-binding domain (Stoilov et al., 2002; Morohashi et al., 2011). *Ythdc2*^{ketu} homozygotes are both male- and female-sterile. In the testis, mutant germ cells carry out an abortive attempt at meiosis: they express hallmark meiotic proteins and initiate recombination, but fail to fully extinguish the spermatogonial mitotic division program, proceed prematurely to an aberrant metaphase-like state, and undergo apoptosis. This phenotype is similar to mutants lacking MEIOC, a meiosis-specific protein that was recently shown to be a binding partner of YTHDC2 and that has been proposed to regulate male and female meiosis by controlling the stability of various mRNAs (Abby et al., 2016; Soh et al., 2017). Our results thus reveal an essential role for YTHDC2 in the germlines of male and

female mice and show that YTHDC2 is an indispensable functional partner of MEIOC. Furthermore, phylogenetic studies demonstrate that the YTHDC2-MEIOC complex is an evolutionarily ancient factor, present in the last common ancestor (LCA) of Metazoa. Nevertheless, despite high conservation in most metazoans, we uncover unexpectedly complex evolutionary patterns for YTHDC2 and MEIOC family members in specific lineages, particularly nematodes and the Schizophora section of flies, which includes *Drosophila melanogaster*.

RESULTS

Isolation of the novel meiotic mutant *ketu* from a forward genetic screen

To discover new meiotic genes, we carried out a phenotype-based, random mutagenesis screen in mice (Jain et al., 2017). Mutagenesis was performed by treatment of male mice of the C57BL/6J strain (B6 hereafter) with the alkylating agent *N*-ethyl-*N*-nitrosourea (ENU). ENU introduces *de novo* mutations in the germline, predominantly single base substitutions (Hitotsumachi et al., 1985; Caspary and Anderson, 2006; Probst and Justice, 2010). To uncover recessive mutations causing defects in male meiosis, we followed a three-generation breeding scheme including outcrossing with females of the FVB/NJ strain (FVB hereafter) (Caspary and Anderson, 2006; Caspary, 2010; Jain et al., 2017) (**Figure 1A**).

Third-generation (G3) male offspring were screened for meiotic defects by immunostaining squash preparations of testis cells for SYCP3, a component of chromosome axes (Lammers et al., 1994; Zickler and Kleckner, 2015), and for γH2AX, a phosphorylated form of the histone variant H2AX that is generated in response to meiotic DNA double-strand breaks (Mahadevaiah et al., 2001) (**Figure 1B**). In normal meiosis, SYCP3-positive axial elements begin to form during the leptotene stage of meiotic prophase I; these elongate and begin to align with homologous chromosome axes to form the tripartite synaptonemal complex in the zygotene stage; the synaptonemal complex connects homologous chromosomes along their lengths in the pachytene stage; and then the synaptonemal complex begins to disassemble during the diplotene stage (**Figure 1B**). Double-strand break formation occurs principally during leptonema and zygonema, yielding strong γH2AX staining across nuclei, but this staining diminishes as recombination proceeds (**Figure 1B**). Recombination-independent γH2AX also appears in the sex body, a heterochromatic domain that encompasses the X and Y chromosomes and that is particularly evident in pachytene and diplotene cells (**Figure 1B**).

In an F1 founder line we named *ketu*, 5 of 26 G3 males screened (**Figure 1C**) contained SYCP3-positive spermatocytes displaying extreme meiotic defects, with no cells resembling normal meiotic prophase I stages (**Figure 1B,D**). For quantification, we divided mutant spermatocytes into classes on the basis of SYCP3 patterns (**Figure 1B**). Type I cells displayed few or none of the normal dots or lines of SYCP3 staining typical of early stages of axial element formation in leptonema; this was the most abundant class, accounting for 74 to 88% of cells (**Figure 1D**). Type II cells displayed prominent dots or short lines of SYCP3; these accounted for 12 to 25% of cells (**Figure 1D**). Type III cells had numerous longer lines of SYCP3, consistent with more advanced axial element formation; these were rare in the younger animals screened (<0.6%) but accumulated to slightly higher levels (3%) in the older G3 animal screened (**Figure 1B,D**). All three cell types had prominent aggregates of SYCP3 and pan-nuclear γH2AX staining (**Figure 1B**). These patterns are unlike those seen in mutants with typical meiotic recombination or synaptonemal complex defects, such as *Spo11*$^{-/-}$, *Dmc1*$^{-/-}$, or *Sycp1*$^{-/-}$ (Pittman et al., 1998; Yoshida et al., 1998; Baudat et al., 2000; Romanienko and Camerini-Otero, 2000; Barchi et al., 2005; de Vries et al., 2005), and suggests that the *ketu* mutation causes an earlier and more severe block to spermatogenesis.

ketu maps to a missense mutation in the *Ythdc2* gene

Because mutagenesis was carried out on B6 males, ENU-induced mutations should be linked to B6 variants for DNA sequences that differ between the B6 and FVB strains. Moreover, all *ketu*-homozygous G3 males should be homozygous for at least some of the same linked B6 variants (Caspary, 2010; Horner and Caspary, 2011). We therefore roughly mapped the *ketu* mutation by hybridizing genomic DNA from five G3 mutants to mouse SNP genotyping arrays and searching for genomic regions where all five mice shared homozygosity for B6 SNPs (Caspary, 2010; Jain et al., 2017). This yielded a 30.59-Mbp interval on chromosome 18, flanked by heterozygous SNPs *rs4138020* (Chr18:22594209) and *gnf18.051.412* (Chr18:53102987) (**Figure 2A**). Whole-exome sequencing of DNA from mutants then revealed that this interval contained a single un-annotated DNA sequence variant located in the *Ythdc2* coding sequence (**Figure 2B,C**).

This variant is an A to G nucleotide transition at position Chr18:44840277, resulting in a missense mutation in predicted exon 6 (**Figure 2B,C**). The mutation changes codon 327 (CAT, histidine) to CGT (arginine), altering a highly conserved residue adjacent to the DEVH box (described in more detail below).

Ythdc2 mRNA is expressed in adult testes as well as widely in other adult and embryonic tissues (**Figure 2B,D**), thus placing *Ythdc2* expression at an appropriate time to contribute to spermatogenesis. While this work was in progress, YTHDC2 protein was reported to interact *in vivo* with the meiosis-specific MEIOC protein, which is itself required for meiosis (Abby et al., 2016; Soh et al., 2017). Furthermore, a CRISPR/Cas9-induced frameshift mutation in exon 2 of *Ythdc2* (**Figure 2B,C**) failed to complement the *ketu* mutation (see below). We conclude that this ENU-induced point mutation disrupts *Ythdc2* function and is the cause of the *ketu* mutant phenotype in males.

Ythdc2^{ketu} causes male and female sterility from gametogenic failure

Ythdc2$^{ketu/+}$ heterozygotes had normal fertility and transmitted the mutation in a Mendelian ratio (30.8% *Ythdc2*$^{+/+}$, 46.6% *Ythdc2*$^{ketu/+}$, and 22.6% *Ythdc2*$^{ketu/ketu}$ from heterozygote × heterozygote crosses; n = 305 mice; p = 0.28, Fisher's exact test). No obvious somatic defects were observed in *Ythdc2*$^{ketu/ketu}$ mice. However, *Ythdc2*$^{ketu/ketu}$ homozygous males were sterile: none of the three animals tested sired progeny when bred with wild-type females. Mutant males showed a 76.4% reduction in testes-to-body-weight ratio compared to littermates (mean ratios were 0.14% for *Ythdc2*$^{ketu/ketu}$ and 0.58% for wild-type and heterozygous animals; $p < 0.01$, one-sided Student's t test; **Figure 3A**).

In histological sections of adult testes, seminiferous tubules from *Ythdc2*$^{ketu/ketu}$ males were greatly reduced in diameter and contained only Sertoli cells and early spermatogenic cells, with no post-meiotic germ cells (**Figure 3B**). To elucidate the timing of spermatogenic failure, we examined juveniles at 10 and 14 days *post partum* (*dpp*). Meiosis first begins in male mice during the second week after birth, with a population of germ cells proliferating mitotically and then entering meiosis in a semi-synchronous wave (Bellve et al., 1977; Griswold, 2016).

In wild type at 10 *dpp*, testes displayed a less advanced subset of tubules containing spermatogonia and Sertoli cells (La in **Figure 3C**), alongside more advanced tubules containing germ cells with morphological characteristics of pre-leptonema and leptonema (Ma in **Figure 3C**). At the same age, testis sections from *Ythdc2*$^{ketu/ketu}$ mice also had a mix of tubules at slightly different stages, but a few tubules (Ab in **Figure 3C**) contained germ cells with abnormal morphology in which the chromosomes were condensed and individualized (arrowheads in **Figure 3C**), reminiscent of metaphase rather than prophase. By 14 *dpp*, essentially all tubules in wild type had germ cells in meiotic prophase, but *Ythdc2*$^{ketu/ketu}$ mutants displayed a mix of tubule types: some tubules (Ab in **Figure 3C**) had normal looking germ cells along with cells with metaphase-like chromosomes, often containing cells with highly compacted, presumably

apoptotic nuclei as well; and some had only a single layer of cells (spermatogonia plus Sertoli cells) along the tubule perimeter (Ep in **Figure 3C,D**). The cells with metaphase-like chromosomes were sometimes alongside cells with a more normal pre-leptotene/leptotene morphology (arrowheads and arrows in **Figure 3C**, respectively). TUNEL staining confirmed a higher incidence of apoptosis at 10 *dpp* and 14 *dpp* in the *Ythdc2*$^{ketu/ketu}$ mutant compared to wild type (**Figure 3D**). Adult testes also contained cells with abnormally condensed metaphase-like chromosomes, as well as nearly empty tubules (**Figure 3B**).

We interpret the less populated tubules in juveniles and adults as those in which apoptosis has already eliminated aberrant cells. We conclude that *Ythdc2*$^{ketu/ketu}$ spermatogonia are able to proliferate mitotically but then transition to an aberrant state in which chromosomes condense prematurely, and are rapidly eliminated, which accounts for the hypogonadism, absence of postmeiotic cells in adults, and sterility.

To verify that the *Ythdc2* point mutation in the *ketu* line is causative for the spermatogenesis defect, we used CRISPR/Cas9 and a guide RNA targeted to exon 2 to generate an endonuclease-mediated (*em1*) allele containing a 5 bp deletion plus 1 bp insertion, resulting in a frameshift (**Figure 2B,C**). *Ythdc2*$^{em1/+}$ heterozygotes transmitted the mutation in a Mendelian ratio (30.7% *Ythdc2*$^{+/+}$, 48.5% *Ythdc2*$^{em1/+}$, and 20.8% *Ythdc2*$^{em1/em1}$; n = 101 mice; p = 0.62, Fisher's exact test). We expected that this mutation near the 5′ end of the gene would be a null or cause severe loss of function. Indeed, *Ythdc2*$^{em1/em1}$ homozygotes displayed hypogonadism and altered testis histology indistinguishable from *Ythdc2*$^{ketu/ketu}$ homozygotes, including the appearance of cells with abnormally condensed chromosomes (0.19% mean testes-to-body-weight ratio for *Ythdc2*$^{em1/em1}$ and 0.57% for wild-type and heterozygous animals; 66.8% reduction; $p < 0.01$, one-sided Student's t test) **(Figure 3A,E**). Moreover, *Ythdc2*$^{em1/em1}$ males were sterile (neither of the two animals tested sired progeny when bred with wild-type females for 9 weeks). *Ythdc2*$^{ketu/em1}$ compound heterozygotes were equally defective (0.20% mean testes-to-body-weight ratio for *Ythdc2*$^{ketu/em1}$ and 0.63% for wild-type and single-heterozygous animals; 67.5% reduction; $p < 0.01$, one-sided Student's t test) (**Figure 3A,F**) and sterile (two animals tested did not sire progeny when bred with wild-type females for 9 weeks). Thus, *ketu* is allelic to *Ythdc2*^{em1}.

Ythdc2$^{ketu/ketu}$ females were also sterile: no pregnancies were observed from crosses of 5 homozygous mutant females to wild-type males (bred for 6–19 weeks). Ovaries from *Ythdc2*$^{ketu/ketu}$ females were dramatically smaller compared to wild-type littermates, and no primary or developing follicles were visible (**Figure 3G**). Again, *Ythdc2*$^{em1/em1}$ homozygotes displayed the same phenotype as *Ythdc2*$^{ketu/ketu}$ females (**Figure 3G**) and a cross between one

homozygous mutant female to a wild-type male yielded no pregnancies (bred for 21 weeks). These findings reveal an essential function for YTHDC2 in both male and female gametogenesis.

Precocious meiotic progression in *Ythdc2*$^{ketu/ketu}$ spermatocytes

YTHDC2 and MEIOC coimmunoprecipitate from testis extracts and they bind an overlapping set of transcripts as assessed by RNA immunoprecipitation, suggesting that the two proteins function together in germ cells (Abby et al., 2016; Soh et al., 2017). This hypothesis predicts similar phenotypes for mutants defective for either gene. In mice homozygous for a targeted *Meioc* mutation, male and female germ cells enter meiosis at the correct developmental stage (i.e., in juvenile males around 10 *dpp*, and in fetal ovary around embryonic day 14.5), but show substantial meiotic defects including rapid progression to a metaphase-like state with condensed univalent chromosomes and monopolar spindles (Abby et al., 2016; Soh et al., 2017). Our initial findings from the screen showed comparable defects in *Ythdc2* mutants (**Figure 1B,D** and **3B,C,E**), so we evaluated this phenotypic similarity more closely.

To evaluate the molecular characteristics of cells containing prematurely condensed chromosomes, we stained testis sections to detect histone H3 phosphorylation on serine 10 (pH3), which appears at high levels on metaphase chromatin, and α-tubulin, a spindle component. At 14 *dpp*, *Ythdc2*$^{ketu/ketu}$ mutants showed pH3-positive cells near the tubule periphery, likely mitotic spermatogonia, with discrete foci that colocalized with pericentromeric heterochromatin (DAPI-bright regions) (arrows in **Figure 4A**, right); this staining was also present in wild type, and has been reported previously (Hendzel et al., 1997; Kimmins et al., 2007; Song et al., 2011) (**Figure 4A**, left). More importantly, mutants also showed bright staining in all cells that had abnormally condensed chromosomes (arrowheads in **Figure 4A**, right). At this age, no wild-type spermatocytes have progressed far enough to reach a strongly pH3-positive (metaphase I) stage (**Figure 4A**, left). In wild-type adults, spermatocytes at metaphase I showed the expected bipolar α-tubulin-containing spindles and well-aligned chromosomes at mid-spindle, with pericentromeric heterochromatin (DAPI-bright regions) oriented toward the poles (**Figure 4B**, left). In contrast, *Ythdc2*$^{ketu/ketu}$ mutants showed no spermatocytes with bipolar spindles; instead, spermatocytes containing a mass of condensed chromosomes had microtubules forming a single aster, consistent with the presence of monopolar spindles (**Figure 4B**, right).

In normal spermatogenesis, cyclin A2 (CCNA2) is expressed during mitotic cell cycles in spermatogonia, but is downregulated upon meiotic entry and is undetectable in SYCP3-positive

spermatocytes (Ravnik and Wolgemuth, 1999). *Meioc*$^{-/-}$ spermatocytes fail to properly extinguish CCNA2 expression, suggesting that inappropriate retention of the spermatogonial (mitotic) cell cycle machinery contributes to precocious metaphase and other meiotic defects (Soh et al., 2017). When testis sections from 14-*dpp* wild-type animals were immunostained for SYCP3 and CCNA2, the expected mutually exclusive localization pattern was observed: CCNA2-positive spermatogonia were located at tubule peripheries while SYCP3-positive spermatocytes occupied more lumenal positions (**Figure 4C**, left). In testis sections from *Ythdc2*$^{ketu/ketu}$ animals of the same age, however, most SYCP3-positive spermatocytes also contained detectable levels of nuclear CCNA2.

We conclude that *Ythdc2*-defective germ cells make an abortive attempt to enter meiosis but then progress precociously to a metaphase-like state and undergo apoptosis. These findings reveal substantial phenotypic similarities between *Ythdc2* and *Meioc* mutants, supporting the hypothesis that YTHDC2 and MEIOC function together to regulate germ cell development around the time of meiotic entry.

YTHDC2 localizes to the cytoplasm of prophase I spermatocytes

To determine the temporal and spatial distribution of YTHDC2 protein in the male germline, we immunostained testis sections from adult wild-type and mutant animals (**Figure 5 and Figure S1**). YTHDC2 staining was prominent in SYCP3-positive spermatocytes in wild type, whereas little to no staining was observed in spermatocytes from *Ythdc2*$^{em1/em1}$ littermates (**Figure 5A**), validating antibody specificity.

In contrast to wild-type littermates, *Ythdc2*$^{ketu/ketu}$ mutants displayed only background levels of YTHDC2 staining in SYCP3-positive cells (arrowheads in **Figure 5B**), indistinguishable from *Ythdc2*$^{em1/em1}$ mutants. The *Ythdc2*^{ketu} mutation may destabilize the protein, and/or YTHDC2 may be required (directly or indirectly) for its own expression as cells transition into meiosis.

In wild type, no YTHDC2 staining above background was observed in spermatogonia or Sertoli cells (e.g., exemplified by the stage I–III tubule in **Figure 5C**). Strong staining first became detectable in pre-leptotene spermatocytes (stage VIII tubule, **Figure 5C**), remained strong from leptonema through diplonema, then returned to background levels in postmeiotic cells (round spermatids) (**Figure 5C**). YTHDC2 appeared exclusively cytoplasmic throughout prophase I, with no detectable nuclear signal above background. Indistinguishable patterns were seen with an independent anti-YTHDC2 antibody (**Figure S1**), further confirming

specificity. Our observations agree with and extend prior reports of cytoplasmic YTHDC2 staining in meiotic prophase I spermatocytes (Abby et al., 2016; Soh et al., 2017).

Structure and domain architecture of YTHDC2

The predicted *Ythdc2* transcript encodes a putative RNA helicase of 1445 amino acids and 161 kDa (presented schematically in **Figure 6A**). Protein domains and amino acid sequence motifs characteristic of superfamily 2 DExH-box helicases are conserved in mouse YTHDC2 and its homologs throughout Eumetazoa (**Figure 6A,B**). More specifically, YTHDC2 is predicted to have the helicase core modules (DEXDc and HELICc domains, including matches to helicase conserved sequence motifs) and two C-terminal extensions (the helicase-associated 2 (HA2) and oligonucleotide binding (OB) domains) that are characteristic of the DEAH/RNA helicase A (RHA) helicase family (**Figure 6A,B, Table S1**) (Fairman-Williams et al., 2010). It is important to note, however, that although YTHDC2 has been shown to have RNA-stimulated ATPase activity (Morohashi et al., 2011) and is sometimes referred to as an RNA helicase (e.g.,Tanabe et al., 2016; Soh et al., 2017), direct demonstration of helicase activity has not yet been reported.

The *ketu* mutation (H327R) alters the amino acid located four residues N-terminal of the DEVH sequence (**Figure 6B**). The biochemical effect of this mutation remains to be determined, but we note that the affected histidine is invariant across likely YTHDC2 orthologs (**Figure 6B**).

On the basis of sequence alignments, human and mouse YTHDC2 cluster most closely with DEAH-box proteins DHX30, DHX9, and TDRD9 (**Figure 6C**), but YTHDC2 is decorated with several auxiliary domains not found in these other proteins, namely, an N-terminal R3H domain, an ankyrin repeat domain (ARD) inserted between the two helicase core domains and containing a pair of ankyrin repeats, and a C-terminal YTH domain (**Figure 6A** and **S2A**). R3H domains are implicated in nucleic acid binding and protein oligomerization (He et al., 2013; He and Yan, 2014). The ARD may be involved in protein-protein interactions (Li et al., 2006).

Most characterized YTH domains bind specifically to RNA containing N^6-methyladenosine (m^6A) (Dominissini et al., 2012; Schwartz et al., 2013; Wang et al., 2014; Xu et al., 2015). A crystal structure of the YTH domain of human YTHDC1 bound to 5′-GG(m^6A)CU revealed that the methyl group in m^6A is accommodated in a pocket composed of three aromatic/hydrophobic residues (Xu et al., 2014). These residues are present in the YTH domains of human and mouse YTHDC2 (triangles in **Figure 6D**). To evaluate this conservation in more detail, we examined an NMR structure of the YTH domain of human YTHDC2 (**Figure 6E**). The YTHDC2 YTH domain adopts an open α/β fold with a core composed of six β strands

surrounded by four alpha helices. This solution structure matched closely with the crystal structure of RNA-bound YTHDC1 (Cα r.m.s.d. = 2.27 Å) and, importantly, the conserved m^{6}A-binding residues aligned well (**Figure 6F**). Consistent with this structural conservation, the human YTHDC2 YTH domain was reported to bind m^{6}A-modified RNAs, albeit with substantially weaker affinity than YTH domains from other proteins (Xu et al., 2015).

The YTH domain of the *Schizosaccharomyces pombe* Mmi1 protein is an exception to the more widely found m^{6}A-binding specificity. Rather than binding m^{6}A, the Mmi1 YTH domain binds a specific RNA sequence called a "determinant of selective removal" (DSR: 5′-UNAAA/C) found in transcripts of meiotic and other genes (Harigaya et al., 2006; Yamashita et al., 2012; Chatterjee et al., 2016). Crystal structures of the *S. pombe* Mmi1 YTH domain alone and bound to RNA with a DSR motif implicated an RNA interaction site distinct from the protein surface by which other YTH domain proteins bind m^{6}A-modified RNA (Chatterjee et al., 2016; Wang et al., 2016). Although the human YTHDC2 structure superimposes well on that of Mmi1 (**Figure S2B**), several key residues involved in DSR binding are not conserved in mouse and human YTHDC2 (red boxes in **Figure 6D**). Specifically, RNA binding was shown to be compromised by mutation of Mmi1 Tyr-466, the side-chain hydroxyl of which forms a hydrogen bond with the N^1 atom of DSR nucleobase A_4 (Wang et al., 2016). This position is a proline in both human and mouse YTHDC2 (**Figure 6D** and **S2B**). We infer that it is unlikely that the YTH domain of YTHDC2 utilizes the DSR interaction surface to bind RNA.

Evolutionary path of ancestral *Ythdc2* and its divergent paralog, Drosophila *bgcn*

YTHDC2 homologs were found in metazoan species in a limited analysis of conservation (Soh et al., 2017). The closest homolog in *Drosophila melanogaster* is Bgcn (benign gonial cell neoplasm), which regulates germ cell differentiation via translational control of target mRNAs (Li et al., 2009b; Kim et al., 2010; Insco et al., 2012; Chen et al., 2014). Bgcn was thus proposed to be the fruit fly ortholog of YTHDC2 (Soh et al., 2017). However, Bgcn lacks YTH and R3H domains (**Figure 7A**) and is missing motifs critical for ATP binding and hydrolysis (Ohlstein et al., 2000) (**Figure 7B**). These differences led us to hypothesize either that the YTHDC2 family is highly plastic evolutionarily, or that YTHDC2 and Bgcn are not orthologs. To distinguish between these possibilities, we characterized the YTHDC2 phylogenetic distribution.

Likely YTHDC2 orthologs were readily found throughout the Eumetazoa subkingdom by BLAST and domain architecture searches (**Figure 7C**, **S3** and **Table S2**). The complete YTHDC2 domain architecture (R3H, helicase core domains with intact helicase motifs I and II and ARD insertion, HA2, OB, and YTH domains) was the most common form (**Figure 6B, 7C,**

S4). Orthologs of this type were present in Cnidaria, the deepest branching eumetazoans examined (*Hydra vulgaris* and the starlet sea anemone *Nematostella vectensis*) (**Figures 6B, 7C, S3, S4**), indicating that full-length YTHDC2 was present in the LCA of the Eumetazoa.

Homologs with the full domain architecture, except for the YTH domain, were observed in green plants, exemplified by *Arabidopsis thaliana* NIH (nuclear DEIH-box helicase) (Isono et al., 1999), and choanoflagellates, e.g., *Salpingoeca rosetta* (Genbank accession XP_004993066). No homologs with this domain architecture were found in other eukaryotic lineages, including fungi (see Methods). The function of the plant homologs has not been reported to our knowledge, but their existence suggests an even more ancient evolutionary origin for this protein family.

Most metazoan YTHDC2 orthologs are highly conserved; for example, the mouse and *H. vulgaris* proteins are 44.6% identical. Nonetheless, there were exceptions to this conservation, including apparent sporadic loss of the YTH domain in some species [e.g., platypus and Japanese gekko among vertebrates; and Annelida (e.g., the leech *Helobdella robusta*) and Echinodermata (the purple sea urchin *Strongylocentratus purpuratus*) among invertebrates] (**Figures 7C** and **S3**). Assuming these losses are not errors in genome assembly or annotation, these findings suggest that the mode of YTHDC2 binding to RNA can be evolutionarily plastic.

Even more diversity was observed in the superphylum Ecdysozoa, which includes nematodes and arthropods (**Figures 7C,D** and **S3**). Most of these species retain the full architecture with the YTH domain, e.g., the horseshoe crab *Limulus polyphemus*; the common house spider *Parasteatoda tepidariorum*; the deer tick *Ixodes scapularis*; and most insect lineages. However, the YTH domain was apparently lost in at least one crustacean (*Daphnia magna*) and more widely in nematodes and in dipteran and hymenopteran insects. We examined these latter exceptions in more detail.

Nematodes. All of the identified YTHDC2 homologs in the phylum Nematoda have a form distinct from other lineages: they retain the R3H, DExH RNA helicase (including intact helicase motifs I and II), HA2, and OB domains, but they lack a detectable YTH domain. More uniquely, they have a sequence between the DEXDc and HELICc helicase core domains that aligns poorly with the equivalent region (including the ARD) in other family members, and that is more divergent between nematode species than between most other species (**Figures 6B, 7A,D, S4 and S5**). For example, the mouse and zebra finch (*Taeniopygia guttata)* proteins are 82.3% identical across this region, whereas the *Caenorhabditis elegans* and *Caenorhabditis briggsae* proteins share only 49.7% identity (**Figure S5B**). If the ARD of YTHDC2 mediates

protein-protein interactions, the diverged structure of this region suggests that the protein-binding partners have also diverged substantially in nematodes.

Insects. Most insect lineages have YTHDC2 orthologs with the full architecture including the YTH domain and intact helicase motifs I and II (**Figures 6B, 7C,D, S3, and S4**). Examples include Coleoptera (e.g., the red flour beetle *Triboleum castaneum*), Lepidoptera (e.g., the silk moth *Bombyx mori*), and Hemiptera (true bugs, e.g., the bed bug *Cimex lectularius*). We conclude that the ancestral metazoan form was present in the LCA of insects.

However, likely orthologs in Hymenoptera (e.g., bees, wasps, ants) and Diptera (e.g., flies, mosquitoes) uniformly lack the YTH domain, suggesting this domain was lost in the LCA of these two clades. In all Hymenoptera and some Diptera examined (mosquitoes), this YTH-less form was the only close homolog of YTHDC2 (**Figure 7D**).

Remarkably, an even more diverged homolog was found specifically in members of the Schizophora section of true flies (**Figure 7D**). Every species that we examined in this clade has a YTHDC2 homolog annotated as an ortholog of *D. melanogaster* Bgcn (**Figure 7E**, **Table S2**). Each has the ARD diagnostic of the YTHDC2 family but, like Bgcn, lacks R3H and YTH domains and has Bgcn-like sequence alterations in helicase motifs I and II that are expected to preclude ATP binding and hydrolysis (Ohlstein et al., 2000) (**Figure 7A,B, and S4**). In addition, most schizophoran flies also have a second YTHDC2 homolog that lacks the YTH domain but that, unlike Bgcn, has intact R3H and helicase motifs I and II (e.g., the house fly *Musca domestica* and the tephritid fruit flies *Rhagoletis zephyria*, *Ceratitus capetata*, and *Bactrocera* species; **Figure 7B,D,E, and S4**). None of the Drosophila genomes examined had this more YTHDC2-like version (**Figure 7D,E**).

A straightforward interpretation is that the YTH domain was lost before the LCA of Hymenoptera and Diptera, then a gene duplication before the LCA of Schizophora was followed by substantial sequence diversification, creating the Bgcn subfamily. Most schizophoran species have both YTHDC2-like and Bgcn-like paralogs, but Drosophilidae retain only the Bgcn version. Thus, although Bgcn is the closest YTHDC2 homolog in *D. melanogaster*, Bgcn is a paralog of YTHDC2, not its ortholog.

Supporting this interpretation, a phylogram based on multiple sequence alignments placed one version from each of the non-Drosophila schizophoran species closer to YTHDC2 orthologs from other species, including mouse, and placed the other copy closer to *D. melanogaster* Bgcn (**Figure 7E**). However, Bgcn paralogs from Drosophila and non-Drosophila species formed two distinct clusters, and the more YTHDC2-like schizophoran paralogs formed

a cluster separated from other YTHDC2 members, including other insect proteins lacking a YTH domain (fuchsia branches in **Figure 7E**). Relative to YTHDC2 from non-schizophoran species, there is greater diversity between Bgcn family members within Schizophora and even within Drosophila (**Figure 7E**). Rapid evolution of Bgcn in Drosophila was noted previously, with evidence of positive selection in *D. melanogaster* and *D. simulans* (Civetta et al., 2006; Bauer DuMont et al., 2007) but not *D. ananassae* (Choi and Aquadro, 2014). Our findings extend this diversity to other schizophoran flies and place the evolutionary dynamics of Bgcn (and schizophoran YTHDC2) in striking contrast to the conservation of the ancestral YTHDC2 form in most other metazoan lineages.

Drosophila *bag of marbles* encodes a highly diverged homolog of MEIOC

Like YTHDC2, MEIOC is highly conserved, with likely orthologs throughout most lineages in Eumetazoa, including Cnidaria (Abby et al., 2016; Soh et al., 2017) (**Figure 8A**). The MEIOC C-terminus contains a domain with a putative coiled-coil motif (DUF4582 (domain of unknown function); pfam15189); this domain is necessary and sufficient for interaction with YTHDC2 (Abby et al., 2016). In Eumetazoa, DUF4582 appears to be unique to MEIOC orthologs and is the protein's most highly conserved feature.

In *D. melanogaster*, the product of the *bag of marbles* (*bam*) gene is a functional collaborator and direct binding partner of Bgcn (Gonczy et al., 1997; Lavoie et al., 1999; Ohlstein et al., 2000; Li et al., 2009b; Shen et al., 2009; Kim et al., 2010; Insco et al., 2012; Chen et al., 2014). By analogy with YTHDC2-MEIOC, it was thus proposed that Bam may be a functional analog of MEIOC (Soh et al., 2017). However, no sequence similarity between MEIOC and Bam has been detected. Moreover, the rapid diversification of Bgcn and in particular its biochemical divergence from the ancestral YTHDC2 form raised the possibility that Bgcn has acquired novel interaction partners, i.e., that Bam and MEIOC are not evolutionarily related. To address these issues, we examined the sequence and phylogenies of Bam and MEIOC.

BLAST searches using mouse or human MEIOC as queries against available dipteran genomes identified no clear homologs (expected (E) value threshold < 20), even though homologs containing DUF4582 were easily found in other insect lineages ($E < 10^{-35}$), including Hymenoptera (e.g., the bumblebee *Bombus terrestris*) (**Figure 8A**, **Table S2**). Conversely, searches using *D. melanogaster* Bam as the query identified homologs in schizophoran flies ($E < 10^{-3}$; **Figure 8A**, **Table S2**), establishing that Bam orthologs are coincident with the presence

of Bgcn-like proteins. However, these searches failed to find Bam homologs in any other species, including non-schizophoran Diptera.

Nonetheless, evidence of remote sequence similarity was observed when MEIOC orthologs from widely divergent species were compared directly with schizophoran Bam orthologs by multiple sequence alignment using COBALT (constrained basic alignment tool (Papadopoulos and Agarwala, 2007)) (**Figure 8B,C**) or PROMALS3D ((Pei and Grishin, 2014); data not shown). Bam orthologs are shorter, lacking an N-terminal extension present in MEIOC (**Figure 8B**). Short patches of sequence similarity with Bam were distributed across the central, nondescript region of MEIOC, but the region with greatest similarity spanned much of the DUF4582 domain (**Figure 8B**). Supporting the significance of this similarity, the C-terminus of Bam, including part of the conserved region, mediates the direct interaction with Bgcn (Li et al., 2009b), as DUF4582 does for MEIOC-YTHDC2 (Abby et al., 2016). Interestingly, however, the COILS prediction algorithm (Lupas et al., 1991) did not detect putative coiled-coil motifs in Bam, unlike MEIOC (data not shown). We conclude that *D. melanogaster* Bam is evolutionarily derived from MEIOC and has a functionally homologous version of the DUF4582 domain, albeit diverged enough that it is not readily recognized as such.

A neighbor-joining tree based on multiple sequence alignments divided schizophoran Bam-like proteins into two clusters representing, respectively, Drosophila and non-Drosophila species (**Figure 8A**). Furthermore, Bam-like proteins showed substantially more sequence diversity than MEIOC-like proteins, and there also was more diversity within Drosophila species than within the other schizophoran flies (**Figure 8A**). Thus, conservation patterns are correlated between MEIOC/Bam and YTHDC2/Bgcn: YTHDC2 and MEIOC are much more highly conserved than are Bgcn and Bam, and Bgcn and Bam display even greater sequence diversity among Drosophila than in other clades in Schizophora.

DISCUSSION

This study establishes an essential function for *Ythdc2* in the germlines of male and female mice, specifically at the stage when stem cells transition from mitotic to meiotic divisions. *Ythdc2* mutant spermatogonia are able to initiate at least part of the meiotic developmental program, making synaptonemal complex precursors and initiating recombination, but most cells then rapidly progress to a premature metaphase-like state and die by apoptosis. *Ythdc2* and *Meioc* mutants have highly similar meiotic phenotypes, supporting the hypothesis that these proteins function together to regulate germ cell differentiation in the mouse gonad (Abby et al., 2016; Soh et al., 2017). During the course of this work, we isolated a novel point-mutated allele of *Ythdc2* (*ketu*) that harbors a non-synonymous mutation of a conserved residue, illustrating the power of phenotype-based forward-genetic approaches for dissecting mammalian reproductive processes.

The YTHDC2 domain architecture, with its RNA interaction modules and putative RNA helicase domains, leads to the obvious hypothesis that YTHDC2 regulates gene expression posttranscriptionally via direct interaction with specific RNA targets. Specifically how this regulation is accomplished remains unclear, however. Two distinct models have been proposed in which the YTHDC2-MEIOC complex controls mRNA stability, either stabilizing the transcripts of meiotic genes (Abby et al., 2016) or destabilizing the transcripts of mitotic cell cycle genes (Soh et al., 2017). Conversely, it has been speculated that YTHDC2 promotes mRNA translation, based on indirect data concerning the effects of shRNA knockdown of *YTHDC2* expression in human colon cancer cell lines (Tanabe et al., 2016). We propose a further alternative that YTHDC2-MEIOC is a translational suppressor, based on analogy with Bgcn-Bam, which has well documented translational suppression functions in *D. melanogaster* germ cells (Li et al., 2009b; Shen et al., 2009; Kim et al., 2010; Insco et al., 2012; Li et al., 2013; Chen et al., 2014). The cytoplasmic localization of the YTHDC2-MEIOC complex (this study and (Abby et al., 2016; Soh et al., 2017)) is consistent with all of these possibilities.

RNA co-immunoprecipitation data suggest that YTHDC2 interacts with specific RNA targets *in vivo* (Abby et al., 2016; Soh et al., 2017), but the detailed list of putative targets has differed between studies and the molecular determinants of binding specificity remain unknown. Our structural analyses support the conclusion that the YTH domain mediates direct interaction with m^6A-containing RNA substrates, also suggested by direct binding studies *in vitro* (Xu et al., 2015). RNA-stimulated ATPase activity has been reported for purified YTHDC2 (Morohashi et al., 2011), but helicase activity has not yet been demonstrated, and what the ATPase function might be remains unclear. Additionally, the biochemical role of MEIOC within the complex is

unknown. Non-exclusive possibilities include modulating YTHDC2 RNA binding specificity, facilitating interactions with other protein partners, and/or modifying ATPase/helicase activities of YTHDC2.

Notwithstanding the intimate connection of YTHDC2 to MEIOC in the germline, it is likely that YTHDC2 has additional, independent functions, because it is expressed much more widely than the highly germline-specific MEIOC. Supporting this hypothesis, YTHDC2 has been implicated in hepatitis C virus replication and cell proliferation in cultured, transformed human cells not known to express MEIOC (Morohashi et al., 2011; Tanabe et al., 2014; Tanabe et al., 2016). Although we have not observed obvious somatic phenotypes in *Ythdc2*$^{ketu/ketu}$ or *Ythdc2*$^{em1/em1}$ homozygotes, or in *Ythdc2*$^{ketu/em1}$ compound heterozygotes, we cannot rule out cellular defects that do not yield gross pathology.

We demonstrate here that the YTHDC2 sequence is well conserved across most metazoan lineages, including the deeply branching Cnidaria. Hence, we conclude that the full-length YTHDC2 is the ancestral form, already present in the LCA of Metazoa. A related protein (possibly lacking the YTH domain) was likely present even earlier, before the LCA of green plants and metazoans. However, we also uncovered substantial structural diversity in the nematode, hymenopteran, and dipteran lineages within Ecdysozoa, and particularly in schizophoran flies. In the simplest cases (Hymenoptera and non-schizophoran Diptera), the structural variation consists principally of loss of the YTH domain. The YTH domain thus appears to be an evolutionarily and biochemically modular contributor to YTHDC2 function.

The nematode variant also lacks a YTH domain, but in addition, the region equivalent to the ARD is highly divergent relative to the ancestral sequence and even between different nematode species. It remains to be determined whether this diversification reflects positive selection for changing sequence (source unknown) or neutral selection (e.g., if the nematode protein has lost an interaction partner, relaxing constraint on the ARD sequence). The function of this protein in nematodes also remains unknown. The *C. elegans* ortholog of YTHDC2 is F52B5.3 (**Table S2**), and the MEIOC ortholog is Y39A1A.9 (Abby et al., 2016). Both proteins are poorly characterized and, to our knowledge, no phenotypes caused by mutation or RNAi have been observed (http://www.wormbase.org/).

More striking still, the dipteran YTH-less YTHDC2 family member in the LCA of the Schizophora appears to have been duplicated to form the Bgcn family, which also lost the R3H domain and accumulated the previously described alterations in ATPase motifs that preclude ATP binding and hydrolysis (Ohlstein et al., 2000). Our studies also revealed for the first time that the Bgcn partner Bam is a divergent homolog of MEIOC unique to Schizophora. Because

we have been unable to identify Bam/MEIOC homologs in non-schizophoran Diptera, it is currently unclear if *bam* arose from a *Meioc* gene duplication (analogous to the evolutionary trajectory of *bgcn*) or if it is simply a highly diverged *Meioc* ortholog. Given the presence of only Bam in non-Drosophila Schizophora, it would be interesting to know whether the YTHDC2 ortholog in these species can interact with Bam.

Previous researchers documented that both Bam and Bgcn are rapidly diversifying in Drosophila species (Civetta et al., 2006; Bauer DuMont et al., 2007; Choi and Aquadro, 2014). Our findings extend this property to schizophoran flies more generally and provide further context by showing that both Bgcn-like and, when present, YTHDC2-like proteins have experienced much greater sequence diversification within Schizophora than elsewhere, and that this is mirrored by more rapid sequence changes in Bam compared with the ancestral MEIOC in other lineages. The coincident occurrence — at or before the LCA of Schizophora — of the *Ythdc2/bgcn* gene duplication, the YTHDC2, Bgcn, and Bam diversification, and the Bgcn structural and biochemical changes makes it tempting to speculate that these were coordinate changes driven by a common set of selective pressures.

If so, what was (is) the source of these pressures? It was speculated that the diversification of Bam and Bgcn may be tied to infection with the alpha-proteobacteria *Wolbachia* (Bauer DuMont et al., 2007). *Wolbachia* is an endosymbiont in many species of insects and other arthropods that is transmitted from parent to offspring and manipulates host reproductive processes (Engelstadter and Hurst, 2009; Pietri et al., 2016). Interestingly, *Wolbachia* also infects a number of nematode species (Ferri et al., 2011), suggesting a possible link to the rapid diversification of YTHDC2 orthologs in that clade as well. However, direct evidence in any species for a link between Bgcn/Bam and *Wolbachia* remains elusive. Moreover, many arthropod species across diverse taxa are infected with *Wolbachia* (Engelstadter and Hurst, 2009; Pietri et al., 2016), but we find that the majority of these taxa have more evolutionarily stable YTHDC2 and MEIOC sequences. Thus, if Bgcn and Bam diversification can be attributed to host genomic conflicts with *Wolbachia*, it may reflect a mode of interaction between *Wolbachia* and the germline that is unique to schizophoran flies (and possibly nematodes).

The complex evolutionary relationships we document here raise the possibility that mammalian YTHDC2 and fly Bgcn have substantially different functions and mechanisms of action, especially given the striking biochemical changes private to the Bgcn subfamily. Moreover, the phenotypic outcomes in mutants differ, for example *bam* and *bgcn* mutant germ cells do not enter meiosis (McKearin and Spradling, 1990; Gonczy et al., 1997), whereas *Meioc*

and *Ythdc2* mutants do. Nonetheless, our findings, along with the recent characterization of *Meioc* mutants (Abby et al., 2016; Soh et al., 2017), also establish intriguing parallels with *bgcn* and *bam*; in both mouse and fruit fly these genes play critical roles in the switch from transit amplifying mitotic cell cycles into meiosis, in both male and female germlines. On the basis of this similarity, and because members of both the YTHDC2/Bgcn and MEIOC/Bam families appear to be nearly ubiquitous in metazoans, we propose that the YTHDC2-MEIOC complex has an evolutionarily ancient and conserved function as a regulator of germ cell fate and differentiation.

MATERIALS AND METHODS

Generation of *ketu* mutants and endonuclease-targeted *Ythdc2* mutations

All experiments conformed to regulatory standards and were approved by the Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Animal Care and Use Committee. Wild-type B6 and FVB mice were purchased from The Jackson Laboratory (Bar Harbor, Maine). Details of the ENU mutagenesis and breeding for screening purposes are provided elsewhere (Jain et al., 2017) (**Figure 1A**) and were similar to previously described methods (Caspary, 2010; Probst and Justice, 2010).

To screen for meiotic defects, spermatocyte squash preparation and immunostaining for SYCP3 and γH2AX (described below) were carried out using testes from pubertal G3 males that were ≥15 days *post partum* (*dpp*) or from adult G3 males (**Figure 1B**). At these ages, spermatocytes in the relevant stages of meiotic prophase I are abundant in normal mice (Bellve et al., 1977). Testes were dissected, frozen in liquid nitrogen, and stored at -80° until samples from ~24 G3 males had been collected, then immunocytology was carried out side by side for all animals from a given line. One testis per mouse was used for cytology and the second was reserved for DNA extraction.

Genotyping of *ketu* animals was done by PCR amplification using *ketu* F and *ketu* R primers (oligonucleotide primer sequences are provided in **Table S3**), followed by digestion of the amplified product with BstXI (NEB, Ipswich, Massachusetts). Wild-type mice contain a *Bst*XI restriction site that is mutated by the *ketu* mutation (A to G).

CRISPR/Cas9-mediated genome editing was done by the MSKCC Mouse Genetics Core Facility to generate *em* alleles. A guide RNA (target sequence 5′-AATAAAGGCTCTTTCCGTAC) was designed to target predicted exon 2 of *Ythdc2* (NCBI Gene ID: 240255 and Ensembl Gene ID: ENSMUSG00000034653) and used for editing as described (Romanienko et al., 2016). Using the T7 promoter in the pU6T7 plasmid, the gRNA was synthesized by *in vitro* transcription and polyadenylated, then 100 ng/μl of gRNA and 50 ng/μl of Cas9 mRNA were co-injected into the pronuclei of CBA × B6 F2 hybrid zygotes using conventional techniques (Hogan and Lacy, 1994). Founder mice were tested for the presence of mutated alleles by PCR amplification of exon 2 using *Ythdc2* F1 and *Ythdc2* R1primers, followed by T7 endonuclease I (NEB) digestion.

To determine the specific mutations in T7-positive *Ythdc2*^{em} founder mice, the targeted region was amplified by PCR of tail-tip DNA (*Ythdc2* F1 and *Ythdc2* R1 primers) and sequenced on the Illumina MiSeq platform (Illumina Inc, San Diego, California) at the MSKCC Integrated Genomics Operation. Reads were aligned to mouse genome assembly GRCm37/mm9 and

variants were identified using Genome Analysis Toolkit version 2.8-1-g932cd3a (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013). Variants with a minimum variant frequency of 0.01 were annotated using VarScan v2.3.7 software (Koboldt et al., 2012). *Ythdc2*^{em} founder males mosaic for frame-shift mutations were bred with B6 mice and potential heterozygote carriers were genotyped by PCR amplification of the targeted region (*Ythdc2* F2 and *Ythdc2* R2 primers), followed by Sanger sequencing (*Ythdc2* Seq1 primer) and analysis with CRISP-ID (Dehairs et al., 2016). A single founder male carrying the *em1* mutation was used to establish the *Ythdc2*^{em1} line, after two backcrosses to B6 mice. *Ythdc2*^{em1} heterozygote carriers were then interbred to generate homozygotes, or crossed to *ketu* mice to generate compound heterozygotes carrying both the *ketu* allele and a *Ythdc2*^{em1} allele. Genotyping of *Ythdc2*^{em1} animals was done by PCR amplification using *Ythdc2* F3 and *Ythdc2* R3 primers, followed by digestion of the amplified product with RsaI (NEB). The 5-bp deletion in *Ythdc2*^{em1} (**Figure 1D**) removes an *Rsa*I site that is present in wild-type mice.

Genetic mapping and exome sequencing

Genome assembly coordinates are from GRCm38/mm10 unless indicated otherwise. *ketu* was coarsely mapped by genome-wide microarray SNP genotyping (Illumina Mouse Medium Density Linkage Panel) using genomic DNA from testes or tail biopsies as described (Jain et al., 2017). Five G3 mutant mice obtained from the initial screen cross (a, b, c, d, e in **Figure 1C**), as well as the F1 founder, one B6 and one FVB control mice were genotyped. Microarray analysis was performed at the Genetic Analysis Facility, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada. For bioinformatics analysis, 777 SNPs were selected based on the following criteria: allelic variation in B6 and FVB, heterozygosity in F1 founder, and autosomal location.

We performed whole-exome sequencing on the same five mutant G3 mice analyzed by microarray SNP genotyping and DNA was prepared as for microarray analysis. Whole-exome sequencing was performed at the MSKCC Integrated Genomics Operation. A unique barcode was incorporated into the DNA library prepared from each mouse, followed by library amplification with 4 PCR cycles. Libraries were then quantified and pooled at equal concentrations into a single sample for exome capture. Exome capture was performed using SureSelectXT kit (Agilent Technologies, Santa Clara, California) and SureSelect Mouse All Exon baits (Agilent Technologies). Libraries were amplified post-capture with 6 PCR cycles and sequenced to generate approximately 80 million 100-bp paired-end reads. Read alignment, variant calling and variant annotation were done as described (Jain et al., 2017), with the

following two modifications. Reads with three or more mismatches and reads without a pair were discarded using SAMtools version 0.1.19 (Li et al., 2009a). Variants were filtered to only include those that 1) had a minimum sequencing depth of five reads, 2) were called as homozygous, and 3) were not known archived SNPs.

ENCODE data analysis

ENCODE long-RNA sequencing data (release 3) with the following GEO accession numbers were used: testis GSM900193, cortex GSM1000563, frontal lobe GSM1000562, cerebellum GSM1000567, ovary GSM900183, lung GSM900196, large intestine GSM900189, adrena gland GSM900188, colon GSM900198, stomach GSM900185, duodenum GSM900187, small intestine GSM900186, heart GSM900199, kidney GSM900194, liver GSM900195, mammary gland GSM900184, spleen GSM900197, thymus GSM900192, whole brain E14.5 GSM1000572, limb E14.5 GSM1000568, liver E14.5 GSM1000571. We acknowledge the ENCODE Consortium (Consortium, 2012) and the ENCODE production laboratory of Thomas Gingeras (Cold Spring Harbor Laboratory) for generating the datasets.

Histology

Histological analysis was conducted as described (Jain et al., 2017). Testes from adult or juvenile mice were fixed overnight in 4% paraformaldehyde (PFA) at 4°, or in Bouin's fixative for 4 to 5 hr at room temperature. Bouin's-fixed testes were washed in water for 1 hr at room temperature, followed by five 1-hr washes in 70% ethanol at 4°. Wild-type and mutant ovaries were fixed in 4% PFA, overnight at 4° and for 1 hr at room temperature, respectively. PFA-fixed tissues were washed twice for 5 min in water at room temperature. Fixed tissues were stored in 70% ethanol for up to 5 days prior to embedding, embedded in paraffin, and sectioned (5 µm). Periodic acid Schiff (PAS) staining, immunohistochemical TUNEL assay, and immunofluorescent staining were performed by the MSKCC Molecular Cytology Core Facility using the Autostainer XL (Leica Microsystems, Wetzlar, Germany) automated stainer for PAS with hematoxylin counterstain, and using the Discovery XT processor (Ventana Medical Systems, Oro Valley, Arizona) for TUNEL and immunofluorescent staining. For immunofluorescent staining, slides were incubated with primary antibody for 5 hr, followed by 60 min incubation with biotinylated goat anti-rabbit, horse anti-goat, or horse anti-mouse antibodies (1:200, Vector Labs, Burlingame, California). Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems) was used for detection, followed by incubation with Tyramide Alexa Fluor 488 or 594 (Invitrogen, Carlsbad, California). PAS-stained and TUNEL slides were digitized using

Pannoramic Flash 250 (3DHistech, Budapest, Hungary) with 20× objective. Images were produced and analyzed using the Pannoramic Viewer software (3DHistech). Higher magnification images of PAS-stained slides were produced using Axio Observer Z2 microscope (Carl Zeiss, Oberkochen, Germany) with 63× oil-immersion objective. Immunofluorescence images were produced using a TCS SP5 II confocal microscope (Leica Microsystems) with 40×/1.25 NA or 63×/1.4 NA oil-immersion objective.

Cytology

Spermatocyte squashes were prepared as described (Page et al., 1998), with modifications as indicated in (Jain et al., 2017) and slides were stored at -80°. Slides were thawed in 1× PBS for 5 min with gentle agitation and immunofluorescent staining was performed as described (Dowdle et al., 2013) using primary and appropriate Alexa Fluor secondary antibodies (1:100; Invitrogen). Primary antibody staining was done overnight at 4° and secondary antibody staining was done for 30 min at room temperature. All antibodies were diluted in blocking buffer. Stained slides were rinsed in water and mounted with coverslips using mounting medium (Vectashield, Vector Labs) containing 4′,6-diamidino-2-phenylindole (DAPI). Slides were stored at 4° for up to 5 days, and were imaged on a Marianas Workstation (Intelligent Imaging Innovations (Denver, Colorado); Zeiss Axio Observer inverted epifluorescent microscope with a complementary metal-oxide semiconductor camera) using a 63× oil-immersion objective.

Antibodies

Primary antibodies and dilutions used for cytology are as follows: mouse anti-SYCP3 (SCP-3 (D-1), 2 µg/ml, Santa Cruz (Dallas, Texas), sc-74569), rabbit anti-γH2AX (p-Histone H2A.X (ser 139), 0.13 µg/ml, Santa Cruz, sc-101696). Those used for histology are as follows: mouse anti-SYCP3 (SCP-3 (D-1), 1 µg/ml, Santa Cruz, sc-74569), goat anti-YTHDC2 (YTHDC2 (G-19), 5 µg/ml, Santa Cruz, sc-249370), rabbit anti-YTHDC2 (YTHDC2, 1 µg/ml, Bethyl Laboratories (Montgomery, Texas), A303-025A), rabbit anti-CCNA2 (anti-Cyclin A2 (Y193), 2.5 µg/ml, Abcam (Cambridge, Massachusetts), ab32386), mouse anti-α-Tubulin (anti-α-Tubulin, 2.5 µg/ml, Millipore (Billerica, Massachusetts), MABT205), rabbit anti-pH3 (anti-phospho-Histone H3 (Ser10), 1 µg/ml, Upstate (Millipore) , 06-570).

Phylogenetic analysis

Mouse YTHDC2 was used as a query in searches using BLASTP (version 2.6.1 (Altschul et al., 1997)) and CDART (conserved domain architecture retrieval tool (Geer et al., 2002)) using NCBI servers. Searches were performed iteratively, with new searches seeded with hits from initial searches. When multiple accessions were present in the same species, we chose the longest isoform available. Searches were further repeated in targeted mode (i.e., restricting the taxon ID in BLASTP) to examine specific lineages in more detail (e.g., Nematoda, Schizophora, Insecta, Crustacea). MegAlign Pro (DNASTAR Inc., Madison, Wisconsin) version 14.1.0 (118) was used to generate multiple sequence alignments with Clustal Omega (Sievers et al., 2011) or MUSCLE (Edgar, 2004) using default settings; to calculate alignment distances using the scoredist function (Sonnhammer and Hollich, 2005); and to output neighbor-joining trees using the BioNJ algorithm (Gascuel, 1997). COBALT (Papadopoulos and Agarwala, 2007) alignments were carried out on the NCBI server (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi). NCBI taxonomic cladograms were constructed using the PhyloT web tool (http://phylot.biobyte.de/). Trees were visualized using the interactive tree of life (ITOL) server (http://itol.embl.de/) (Letunic and Bork, 2016) or FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree).

From targeted BLASTP searches of the following taxa, no clear matches to the YTHDC2 architecture were observed: Amoebozoa, Fornicata, Euglenozoa, Alveolata, Apusozoa, Cryptophyta, Haptophyceae, Heterolobosea, Parabasalia, Rhizaria, Rhodophyta, Stramenopiles. In these lineages, the closest homologs found were RNA helicase-like proteins more similar in architecture to the DHX30 family, without R3H and YTH domains and lacking the ARD insertion between the helicase core domains.

We also did not find clear matches to the YTHDC2 architecture in fungi. The closest *Saccharomyces cerevisiae* homolog (YLR419W) lacks the diagnostic ARD insertion between the helicase core domains, has N-terminal UBA (ubiquitin-associated) and RWD domains rather than an R3H domain, and lacks a YTH domain. YLR419W thus more closely resembles human DHX57 (**Figure S2A**), which is indeed the top hit when YLR419W is used as the query in a BLASTP search against the human genome (GenBank accession AAH65278.1; 29% identity, E value 2×10^{-110}).

Protein analysis

Domain annotations were obtained from SMART (Letunic et al., 2015) and pfam database (Finn et al., 2016) searches. Atomic coordinates of NMR and crystal structures of YTH domains

were retrieved from the following Protein Data Bank (PDB) entries: 2YU6 (human YTHDC2 apo structure), 4R3I (RNA-bound human YTHDC1), 5DNO (RNA-bound *S. pombe* Mmi1), and 5H8A (*S. pombe* Mmi1 apo structure). Alignments of three-dimensional structures were performed using the cealign command (Vertrees, 2007) and the figures prepared using PyMol (Schrodinger, 2015). Protein accession numbers are listed in **Tables S1 and S2**.

Data availability

Reagents and mouse strains are available upon request.

ACKNOWLEDGMENTS

We thank Christopher Lima (MSKCC) for sharing resources and discussion on structural analysis. We thank Keeney lab members Luis Torres, Diana Y. Eng, and Jacquelyn Song for assistance with genotyping and mouse husbandry. We thank Ning Fan, Dmitry Yarilin, Yevgeniy Romin, and Mesruh Turkekul (MSKCC Cytology Core) for help with histological analyses and imaging. We thank Nathalie J. Lailler (MSKCC Integrated Genomics Operation) for variant analyses of whole-exome sequencing data and Peter Romanienko (MSKCC Mouse Genetics Core Facility) for designing the gRNA. We thank the Genetic Analysis Facility (Centre for Applied Genomics, Hospital for Sick Children, Toronto, ON, Canada) for microarray analysis; the MSKCC Integrated Genomics Operation for whole-exome sequencing; and the MSKCC Mouse Genetics Core Facility for CRISPR/Cas9-targeted mice. We are grateful to Gabriel Livera and Jeremy Wang for sharing unpublished information.

REFERENCES

Abby, E., Tourpin, S., Ribeiro, J., Daniel, K., Messiaen, S., Moison, D., Guerquin, J., Gaillard, J.C., Armengaud, J., Langa, F.*, et al.* (2016). Implementation of meiosis prophase I programme requires a conserved retinoid-independent stabilizer of meiotic transcripts. Nat Commun *7*, 10324.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Barchi, M., Mahadevaiah, S., Di Giacomo, M., Baudat, F., de Rooij, D.G., Burgoyne, P.S., Jasin, M., and Keeney, S. (2005). Surveillance of different recombination defects in mouse spermatocytes yields distinct responses despite elimination at an identical developmental stage. Mol Cell Biol *25*, 7203-7215.

Baudat, F., Manova, K., Yuen, J.P., Jasin, M., and Keeney, S. (2000). Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Mol Cell *6*, 989-998.

Bauer DuMont, V.L., Flores, H.A., Wright, M.H., and Aquadro, C.F. (2007). Recurrent positive selection at *bgcn*, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein bam. Mol Biol Evol *24*, 182-191.

Bellve, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M., and Dym, M. (1977). Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J Cell Biol *74*, 68-85.

Caspary, T. (2010). Phenotype-driven mouse ENU mutagenesis screens. Methods Enzymol *477*, 313-327.

Caspary, T., and Anderson, K.V. (2006). Uncovering the uncharacterized and unexpected: unbiased phenotype-driven screens in the mouse. Dev Dyn *235*, 2412-2423.

Chatterjee, D., Sanchez, A.M., Goldgur, Y., Shuman, S., and Schwer, B. (2016). Transcription of lncRNA prt, clustered prt RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of pho1 gene expression under phosphate-replete conditions in fission yeast. RNA *22*, 1011-1025.

Chen, D., Wu, C., Zhao, S., Geng, Q., Gao, Y., Li, X., Zhang, Y., and Wang, Z. (2014). Three RNA binding proteins form a complex to promote differentiation of germline stem cell lineage in *Drosophila*. PLoS Genet *10*, e1004797.

Choi, J.Y., and Aquadro, C.F. (2014). The coevolutionary period of *Wolbachia pipientis* infecting *Drosophila ananassae* and its impact on the evolution of the host germline stem cell regulating genes. Mol Biol Evol *31*, 2457-2471.

Civetta, A., Rajakumar, S.A., Brouwers, B., and Bacik, J.P. (2006). Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in Drosophila. Mol Biol Evol *23*, 655-662.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Davies, E.L., and Fuller, M.T. (2008). Regulation of self-renewal and differentiation in adult stem cell lineages: lessons from the *Drosophila* male germ line. Cold Spring Harb Symp Quant Biol *73*, 137-145.

de Rooij, D.G. (2001). Proliferation and differentiation of spermatogonial stem cells. Reproduction *121*, 347-354.

de Vries, F.A., de Boer, E., van den Bosch, M., Baarends, W.M., Ooms, M., Yuan, L., Liu, J.G., van Zeeland, A.A., Heyting, C., and Pastink, A. (2005). Mouse *Sycp1* functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. Genes Dev *19*, 1376-1389.

Dehairs, J., Talebi, A., Cherifi, Y., and Swinnen, J.V. (2016). CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. Sci Rep *6*, 28973.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M.*, et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet *43*, 491-498.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Dowdle, J.A., Mehta, M., Kass, E.M., Vuong, B.Q., Inagaki, A., Egli, D., Jasin, M., and Keeney, S. (2013). Mouse BAZ1A (ACF1) is dispensable for double-strand break repair but is essential for averting improper gene expression during spermatogenesis. PLoS Genet *in press*.

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

Engelstadter, J., and Hurst, G.D.D. (2009). The ecology and evolution of microbes that manipulate host reproduction. Annual Review of Ecology Evolution and Systematics *40*, 127-149.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr Opin Struct Biol *20*, 313-324.

Ferri, E., Bain, O., Barbuto, M., Martin, C., Lo, N., Uni, S., Landmann, F., Baccei, S.G., Guerrero, R., de Souza Lima, S.*, et al.* (2011). New insights into the evolution of *Wolbachia* infections in filarial nematodes inferred from a large range of screened species. PLoS One *6*, e20843.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A.*, et al.* (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res *44*, D279-285.

Friedemann, J., Grosse, F., and Zhang, S. (2005). Nuclear DNA helicase II (RNA helicase A) interacts with Werner syndrome helicase and stimulates its exonuclease activity. J Biol Chem *280*, 31303-31313.

Fu, Q., and Yuan, Y.A. (2013). Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). Nucleic Acids Res *41*, 3457-3470.

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol *14*, 685-695.

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: protein homology by domain architecture. Genome Res *12*, 1619-1623.

Gonczy, P., Matunis, E., and DiNardo, S. (1997). *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. Development *124*, 4361-4371.

Griswold, M.D. (2016). Spermatogenesis: The commitment to meiosis. Physiol Rev *96*, 1-17.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and 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*, 3977-3993.

Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A., and Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature *442*, 45-50.

Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. Nat Rev Genet *2*, 280-291.

He, G.J., and Yan, Y.B. (2014). Self-association of poly(A)-specific ribonuclease (PARN) triggered by the R3H domain. Biochim Biophys Acta *1844*, 2077-2085.

He, G.J., Zhang, A., Liu, W.F., and Yan, Y.B. (2013). Distinct roles of the R3H and RRM domains in poly(A)-specific ribonuclease structural integrity and catalysis. Biochim Biophys Acta *1834*, 1089-1098.

Hendzel, M.J., Wei, Y., Mancini, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones, D.P., and Allis, C.D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. Chromosoma *106*, 348-360.

Hitotsumachi, S., Carpenter, D.A., and Russell, W.L. (1985). Dose-repetition increases the mutagenic effectiveness of N-ethyl-N-nitrosourea in mouse spermatogonia. Proc Natl Acad Sci U S A *82*, 6619-6621.

Hogan, B., and Lacy, E. (1994). Manipulating the mouse embryo : a laboratory manual, 2nd edn (Plainview, N.Y., Cold Spring Harbor Laboratory Press).

Horner, V.L., and Caspary, T. (2011). Creating a "hopeful monster": mouse forward genetic screens. Methods Mol Biol *770*, 313-336.

Hunter, N. (2007). Meiotic recombination. In Molecular Genetics of Recombination, A. Aguilera, and R. Rothstein, eds. (Berlin Heidelberg, Springer-Verlag), pp. 381-442.

Insco, M.L., Bailey, A.S., Kim, J., Olivares, G.H., Wapinski, O.L., Tam, C.H., and Fuller, M.T. (2012). A self-limiting switch based on translational control regulates the transition from proliferation to differentiation in an adult stem cell lineage. Cell Stem Cell *11*, 689-700.

Isono, K., Yamamoto, H., Satoh, K., and Kobayashi, H. (1999). An *Arabidopsis* cDNA encoding a DNA-binding protein that is highly similar to the DEAH family of RNA/DNA helicase genes. Nucleic Acids Res *27*, 3728-3735.

Jain, D., Meydan, C., Lange, J., Claeys Bouuaert, C., Mason, C.E., Anderson, K.V., and Keeney, S. (2017). *rahu* is a mutant allele of *Dnmt3c*, encoding a DNA methyltransferase required for meiosis and transposon repression in the mouse male germline. bioRxiv.

Kim, J.Y., Lee, Y.C., and Kim, C. (2010). Direct inhibition of Pumilo activity by Bam and Bgcn in *Drosophila* germ line stem cell differentiation. J Biol Chem *285*, 4741-4746.

Kimmins, S., Crosio, C., Kotaja, N., Hirayama, J., Monaco, L., Hoog, C., van Duin, M., Gossen, J.A., and Sassone-Corsi, P. (2007). Differential functions of the Aurora-B and Aurora-C kinases in mammalian spermatogenesis. Mol Endocrinol *21*, 726-739.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res *22*, 568-576.

Lammers, J.H., Offenberg, H.H., van Aalderen, M., Vink, A.C., Dietrich, A.J., and Heyting, C. (1994). The gene encoding a major component of the lateral elements of synaptonemal complexes of the rat is related to X-linked lymphocyte-regulated genes. Mol Cell Biol *14*, 1137-1146.

Lavoie, C.A., Ohlstein, B., and McKearin, D.M. (1999). Localization and function of Bam protein require the *benign gonial cell neoplasm* gene product. Dev Biol *212*, 405-413.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res *44*, W242-245.

Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. Nucleic Acids Res *43*, D257-260.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, J., Mahajan, A., and Tsai, M.D. (2006). Ankyrin repeat: A unique motif mediating protein-protein interactions. Biochemistry *45*, 15168-15178.

Li, Y., Minor, N.T., Park, J.K., McKearin, D.M., and Maines, J.Z. (2009b). Bam and Bgcn antagonize *Nanos*-dependent germ-line stem cell maintenance. Proc Natl Acad Sci U S A *106*, 9304-9309.

Li, Y., Zhang, Q., Carreira-Rosario, A., Maines, J.Z., McKearin, D.M., and Buszczak, M. (2013). Mei-p26 cooperates with Bam, Bgcn and Sxl to promote early germline development in the *Drosophila* ovary. PLoS One *8*, e58301.

Lin, L., Li, Y., Pyo, H.M., Lu, X., Raman, S.N., Liu, Q., Brown, E.G., and Zhou, Y. (2012). Identification of RNA helicase A as a cellular factor that interacts with influenza A virus NS1 protein and its role in the virus life cycle. J Virol *86*, 1942-1954.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Mahadevaiah, S.K., Turner, J.M., Baudat, F., Rogakou, E.P., de Boer, P., Blanco-Rodriguez, J., Jasin, M., Keeney, S., Bonner, W.M., and Burgoyne, P.S. (2001). Recombinational DNA double-strand breaks in mice precede synapsis. Nat Genet *27*, 271-276.

McKearin, D.M., and Spradling, A.C. (1990). *bag-of-marbles*: a *Drosophila* gene required to initiate both male and female gametogenesis. Genes Dev *4*, 2242-2251.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Morohashi, K., Sahara, H., Watashi, K., Iwabata, K., Sunoki, T., Kuramochi, K., Takakusagi, K., Miyashita, H., Sato, N., Tanabe, A.*, et al.* (2011). Cyclosporin A associated helicase-like protein facilitates the association of hepatitis C virus RNA polymerase with its cellular cyclophilin B. PLoS One *6*, e18285.

Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The *Drosophila* cystoblast differentiation factor, *benign gonial cell neoplasm*, is related to DExH-box proteins and interacts genetically with *bag-of-marbles*. Genetics *155*, 1809-1819.

Page, J., Suja, J.A., Santos, J.L., and Rufas, J.S. (1998). Squash procedure for protein immunolocalization in meiotic cells. Chromosome Res *6*, 639-642.

Page, S.L., and Hawley, R.S. (2003). Chromosome choreography: the meiotic ballet. Science *301*, 785-789.

Papadopoulos, J.S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics *23*, 1073-1079.

Pei, J., and Grishin, N.V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol *1079*, 263-271.

Pietri, J.E., DeBruhl, H., and Sullivan, W. (2016). The rich somatic life of *Wolbachia*. Microbiologyopen *5*, 923-936.

Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U., and Pestova, T.V. (2008). Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. Cell *135*, 1237-1250.

Pittman, D.L., Cobb, J., Schimenti, K.J., Wilson, L.A., Cooper, D.M., Brignull, E., Handel, M.A., and Schimenti, J.C. (1998). Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for *Dmc1*, a germline-specific RecA homolog. Mol Cell *1*, 697-705.

Probst, F.J., and Justice, M.J. (2010). Mouse mutagenesis with the chemical supermutagen ENU. Methods Enzymol *477*, 297-312.

Ravnik, S.E., and Wolgemuth, D.J. (1999). Regulation of meiosis during mammalian spermatogenesis: the A-type cyclins and their associated cyclin-dependent kinases are differentially expressed in the germ-cell lineage. Dev Biol *207*, 408-418.

Romanienko, P.J., and Camerini-Otero, R.D. (2000). The mouse *Spo11* gene is required for meiotic chromosome synapsis. Molecular Cell *6*, 975-987.

Romanienko, P.J., Giacalone, J., Ingenito, J., Wang, Y., Isaka, M., Johnson, T., You, Y., and Mark, W.H. (2016). A vector with a single promoter for *in vitro* transcription and mammalian cell expression of CRISPR gRNAs. PLoS One *11*, e0148362.

Sasaki, M., Lange, J., and Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol *11*, 182-195.

Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G.*, et al.* (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409-1421.

Shen, R., Weng, C., Yu, J., and Xie, T. (2009). eIF4A controls germline stem cell self-renewal by directly inhibiting BAM function in the *Drosophila* ovary. Proc Natl Acad Sci U S A *106*, 11623-11628.

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T.*, et al.* (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell *17*, 775-787.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Soh, Y.Q.S., Mikedis, M.M., Kojima, M., Godfrey, A.K., de Rooij, D.G., and Page, D.C. (2017). *Meioc* maintains an extended meiotic prophase I in mice. PLoS Genet *13*, e1006704.

Song, N., Liu, J., An, S., Nishino, T., Hishikawa, Y., and Koji, T. (2011). Immunohistochemical analysis of histone H3 modifications in germ cells during mouse spermatogenesis. Acta Histochem Cytochem *44*, 183-190.

Sonnhammer, E.L., and Hollich, V. (2005). Scoredist: a simple and robust protein sequence distance estimator. BMC Bioinformatics *6*, 108.

Stoilov, P., Rafalska, I., and Stamm, S. (2002). YTH: a new domain in nuclear proteins. Trends Biochem Sci *27*, 495-497.

Tanabe, A., Konno, J., Tanikawa, K., and Sahara, H. (2014). Transcriptional machinery of TNF-alpha-inducible YTH domain containing 2 (*YTHDC2*) gene. Gene *535*, 24-32.

Tanabe, A., Tanikawa, K., Tsunetomi, M., Takai, K., Ikeda, H., Konno, J., Torigoe, T., Maeda, H., Kutomi, G., Okita, K.*, et al.* (2016). RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1alpha mRNA is translated. Cancer Lett *376*, 34-42.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J.*, et al.* (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics *43*, 11 10 11-33.

Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y., and Akman, S.A. (2005). The DEXH protein product of the *DHX36* gene is the major source

of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. J Biol Chem *280*, 38117-38120.

Vertrees, J. (2007). cealign: A structure alignment plugin for PyMol.

Wang, C., Zhu, Y., Bao, H., Jiang, Y., Xu, C., Wu, J., and Shi, Y. (2016). A novel RNA-binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. Nucleic Acids Res *44*, 969-982.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., *et al.* (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117-120.

Wenda, J.M., Homolka, D., Yang, Z., Spinelli, P., Sachidanandam, R., Pandey, R.R., and Pillai, R.S. (2017). Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function. Dev Cell *41*, 623-637 e629.

Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., and Min, J. (2015). Structural basis for the discriminative recognition of N^6-methyladenosine RNA by the human YT521-B homology domain family of proteins. J Biol Chem *290*, 24902-24913.

Xu, C., Wang, X., Liu, K., Roundtree, I.A., Tempel, W., Li, Y., Lu, Z., He, C., and Min, J. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. Nat Chem Biol *10*, 927-929.

Yamashita, A., Shichino, Y., Tanaka, H., Hiriart, E., Touat-Todeschini, L., Vavasseur, A., Ding, D.Q., Hiraoka, Y., Verdel, A., and Yamamoto, M. (2012). Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. Open Biol *2*, 120014.

Yoo, J.S., Takahasi, K., Ng, C.S., Ouda, R., Onomoto, K., Yoneyama, M., Lai, J.C., Lattmann, S., Nagamine, Y., Matsui, T.*, et al.* (2014). DHX36 enhances RIG-I signaling by facilitating PKR-mediated antiviral stress granule formation. PLoS Pathog *10*, e1004012.

Yoshida, K., Kondoh, G., Matsuda, Y., Habu, T., Nishimune, Y., and Morita, T. (1998). The mouse RecA-like gene *Dmc1* is required for homologous chromosome synapsis during meiosis. Mol Cell *1*, 707-718.

Zheng, H.J., Tsukahara, M., Liu, E., Ye, L., Xiong, H., Noguchi, S., Suzuki, K., and Ji, Z.S. (2015). The novel helicase helG (DHX30) is expressed during gastrulation in mice and has a structure similar to a human DExH box helicase. Stem Cells Dev *24*, 372-383.

Zickler, D., and Kleckner, N. (2015). Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb Perspect Biol *7*, pii: a016626.

FIGURE LEGENDS

Figure 1. Mice from the ENU-induced mutant line *ketu* have meiotic defects.
(A) Breeding scheme. Mutagenized males (B6) were crossed to females of a different strain (FVB) to produce founder (F1) males that were potential mutation carriers. Each F1 male was then crossed to wild-type FVB females. If an F1 male was a mutation carrier, half of his daughters (second generation, G2) should also be carriers, so the G2 daughters were crossed back to their F1 sire to generate third-generation (G3) offspring that were potentially homozygous. For a line carrying a single autosomal recessive mutation of interest, one eighth of G3 males were expected to be homozygous. Un-filled shapes represent animals that are wild-type for a mutation of interest, half-filled shapes are heterozygous carriers, and filled shapes are homozygotes. **(B)** Representative images of squashed spermatocyte preparations immunostained for SYCP3 and γH2AX. Mutant spermatocytes were classified as Types I, II, or III on the basis of SYCP3 patterns. Scale bar represents 20 μm. **(C)** Screen results for the *ketu* line. The F1 male was harem-bred to six G2 females, yielding 26 G3 males that displayed either a wild-type or *ketu* (mice a, b, c, d, e) phenotype. **(D)** Distribution of SYCP3-staining patterns in four G3 *ketu* mutants (a, b, c, d) and their phenotypically wild-type littermates (a′, b′, c′). Wild-type spermatocytes were classified as either early prophase-like (leptonema or zygonema) or late prophase-like (pachynema or diplonema). Spermatocytes from mutant mice were categorized as described in panel B. The number of SYCP3-positive spermatocytes counted from each animal is indicated and raw data are provided in **Figure 1D−Source Data 1**.

Figure 2. *ketu* mice harbor a point mutation in *Ythdc2*.
(A) SNP genotypes of five G3 *ketu* mutants (a, b, c, d, e; from **Figure 1C**) obtained using the Illumina Medium Density Linkage Panel. The single 30.59-Mbp region of B6 SNP homozygosity that is shared between mutants is highlighted in pink. **(B)** Top: Schematic of *Ythdc2* (as predicted by Ensembl release 89) showing the locations of the ENU-induced lesion and the gRNA used for CRISPR/Cas9-targeting. Bottom: The density of ENCODE long RNA-sequencing reads (release 3) from adult testis within a window spanning from 3,500 bp upstream to 200 bp downstream of *Ythdc2*. The vertical viewing range is 0–50; read densities exceeding this range are overlined in pink. **(C)** The *ketu* and CRISPR/Cas9-induced (*em1*) alleles of *Ythdc2*. **(D)** *Ythdc2* and *Meioc* expression level estimate (mean reads per kilobase per million mapped reads (RPKM) values provided by ENCODE (**Figure 2D−Source Data 1**)) in adult and embryonic tissues.

Figure 3. *ketu* and *em1* alleles of *Ythdc2* lead to gametogenic failure and fail to complement each other.

(A) The ratios of testes weight to body weight for 6- to 34-week-old mice (**Figure 3A−Source Data 1**). **(B)** PAS-stained sections of Bouin's-fixed testes from an 8-month-old *Ythdc2*$^{ketu/ketu}$ male and a wild-type littermate. **(C)** PFA-fixed, PAS-stained testis sections. In wild type, examples are indicated of a less advanced ("La") tubule harboring cells with spermatogonia-like morphology and Sertoli cells, and more advanced ("Ma") tubules harboring cells with morphological characteristics of pre-leptonema and meiotic prophase stages. In the mutant, examples are indicated of abnormal ("Ab") tubules containing cells with condensed and individualized chromosomes (arrowheads), and an emptier-looking ("Ep") tubule harboring cells with spermatogonia-like morphology and Sertoli cells. **(D)** TUNEL-stained testis sections. Black arrowheads point to TUNEL-positive cells (stained dark brown). **(E** and **F)** PFA-fixed, PAS-stained testis sections from an 8-week-old *Ythdc2*$^{em1/em1}$ male, a 7-week-old *Ythdc2*$^{ketu/em1}$ male, and their wild-type littermates. **(G)** PFA-fixed, PAS-stained ovary sections from a 6-week-old *Ythdc2*$^{ketu/ketu}$ female and a wild-type littermate, and a 9-week-old *Ythdc2*$^{em1/em1}$ female and a heterozygous littermate. In the higher magnification views of panels B, C, E, and F, arrowheads indicate cells with condensed and individualized chromosomes, arrows indicate cells with morphological characteristics of pre-leptonema and leptonema, and S and Spg indicate Sertoli cells and cells with morphological characteristics of spermatogonia, respectively. In panels B, C, E, and F, the scale bars represent 50 μm and 20 μm in the lower and higher magnification images, respectively. In panels D and G, the scale bars represent 50 μm and 300 μm, respectively.

Figure 4. *Ythdc2*$^{ketu/ketu}$ spermatocytes show precocious meiotic progression.

(A) Anti-pH3 immunofluorescence on testis sections from 14-*dpp Ythdc2*$^{ketu/ketu}$ and wild-type littermates. Arrows indicate cells with spermatogonia-like morphology and arrowheads point to cells with abnormally condensed chromosomes. **(B)** Anti-α-tubulin immunofluorescence on testis sections from 2-month-old *Ythdc2*$^{ketu/ketu}$ and wild-type littermates. **(C)** CCNA2 and SYCP3 immunofluorescence on testis sections from 14-*dpp Ythdc2*$^{ketu/ketu}$ and wild-type littermates. Arrows point to cells with spermatogonia-like morphology and arrowheads indicate SYCP3-positive spermatocytes. In panels A and B, the scale bars represent 5 μm. In panel C, the scale bars represents 50 μm and 15 μm in the lower (left) and higher (right) magnification images, respectively.

Figure 5. YTHDC2 localization in wild-type and mutant testes.

(A and **B)** YTHDC2 and SYCP3 immunofluorescence on testis sections from 2-month-old *Ythdc2*$^{em1/em1}$, *Ythdc2*$^{ketu/ketu}$, and their wild-type littermates. Arrowheads indicate SYCP3-positive spermatocytes. Scale bars represent 100 μm. **(C)** YTHDC2 and SYCP3 immunofluorescence on testis sections from an adult B6 male. Approximate seminiferous epithelial cycle stages (based on SYCP3 and DAPI staining patterns) are provided. S, Sertoli cell; Spg, spermatogonia; L, leptotene spermatocyte; Z, zygotene spermatocyte; P, pachytene spermatocyte; D, diplotene spermatocyte; RS, round spermatid, ES, elongating spermatid. Scale bar represents 15 μm.

Figure 6. YTHDC2 domain architecture and structure of its YTH domain.

(A) Schematic of mouse YTHDC2 domain structure (not to scale). Sequence motifs characteristic of superfamily 2 DExH-box helicases (I, Ia, II, III, IV, V, VI) within the helicase core domain are indicated, along with sequence logos from Clustal Omega alignments of 157 superfamily 2 DExH-box helicases. The height of each stack of residues reflects the degree of conservation, and the height of each amino acid symbol within each stack is proportional to the frequency of a residue at that position. Amino acids are colored according to their physico-chemical properties (hydrophilic (blue), neutral (green), and hydrophobic (black)). **(B)** Clustal Omega alignments of sequences around helicase motifs I and II from YTHDC2 proteins of the indicated species. The position of the *ketu* mutation is indicated. The residues are shaded based on the percentage that agree with the consensus. Sequence logos were generated from Clustal Omega alignments of YTHDC2 homologs from 201 species and are colored as in panel A. **(C)** Cladogram of Clustal Omega protein sequence alignments of mouse and human YTHDC2 paralogs. The tree was rooted using vaccinia virus NPH-II sequence and *D. melanogaster* Bgcn is included for comparison. DHX29 is an RNA helicase that promotes translation initiation on mRNAs with structured 5′ untranslated regions (Pisareva et al., 2008). DHX57 is an uncharacterized protein of unknown function. DHX36 has G-quadruplex unwinding activity for DNA and RNA and is involved in antiviral responses to dsRNA (Vaughn et al., 2005; Yoo et al., 2014). DHX9 (also known as RNA helicase A and DNA helicase II) has both DNA and RNA helicase activities and multiple cellular functions, including in genome stability and viral and cellular RNA metabolism (Friedemann et al., 2005; Lin et al., 2012). DHX30 is a poorly characterized protein required for cell viability in the developing mouse embryo (Zheng et al., 2015). And TDRD9 forms a complex with MIWI2 involved in piRNA-directed transposon silencing in the male germline (Shoji et al., 2009; Wenda et al., 2017). **(D)** Clustal Omega alignment of YTH domain sequences. Inverted triangles in auburn indicate residues that make

up the hydrophobic pocket that binds m^6A. Residues boxed in auburn are required for Mmi1 interaction with RNA (Wang et al., 2016). **(E)** NMR structure of the YTH domain from human YTHDC2 (PDB ID: 2YU6). **(F)** Structure of the m^6A binding pocket is conserved in YTHDC2. The solution structure of YTH domain from human YTHDC2 (green) is shown superimposed on the crystal structure of the RNA-bound YTH domain from human YTHDC1 (pink; PDB ID: 4R3I) (Xu et al., 2014). The m^6A nucleotide and the hydrophobic amino acid residues lining the binding pocket are shown at a higher magnification on the right. Protein accession numbers for sequence logos in panel A are in **Table S1**; all other protein accession numbers are in **Table S2**.

Figure 7. Distribution of YTHDC2 variants in Metazoa.

(A) Schematic of domain architectures (not to scale) of metazoan YTHDC2 orthologs and paralogs. **(B)** Clustal Omega alignment of sequences around helicase motifs I and II for YTHDC2-like and Bgcn-like proteins from schizophoran flies. Mouse YTHDC2 is shown for comparison. Bgcn proteins have amino acid changes that are incompatible with ATPase activity (Ohlstein et al., 2000). **(C)** Phylogenetic distribution of YTHDC2 in Metazoa. The tree is an unrooted cladogram of NCBI taxonomy for a non-exhaustive collection of 234 species in which at least one close YTHDC2 homolog was identified. Tree leaves are color coded according to YTHDC2 domain architecture. The same tree topology is reproduced in **Figure S3** with complete species names. **(D)** Distribution of YTHDC2 variants in Ecdysozoa. The rooted cladogram shows NCBI taxonomy for the ecdysozoan portion of the metazoan tree in panel C. Background shading and color-coding of tree leaves is the same as in panel C. **(E)** Phylogram for sequence alignments of complete YTHDC2 and Bgcn orthologs from the indicated species. Note that protein sequence distances are similar within the YTHDC2 and Bgcn subfamily trees, but species in the YTHDC2 tree span much greater evolutionary distances. Sequences were aligned with Clustal Omega and the unrooted neighbor-joining tree was constructed from distances calculated using the scoredist function (Sonnhammer and Hollich, 2005). Tree leaves are color coded by YTHDC2 protein domain architecture as in panel C. Protein accession numbers are in **Table S2**.

Figure 8. Bam shares distant sequence similarity with MEIOC.

(A) Phylograms based on sequence alignments of MEIOC or Bam orthologs. Sequences were aligned with Clustal Omega and the unrooted neighbor-joining trees were constructed from distances calculated using the scoredist function. Note that the two trees have the same scale,

but the MEIOC proteins are from species separated by much greater evolutionary distances. **(B, C)** Remote sequence similarity between MEIOC and Bam, concentrated across the conserved DUF4582 domain of MEIOC. Sequences were aligned using COBALT (Papadopoulos and Agarwala, 2007). Panel B shows a schematic of the full alignment, with thin gray lines indicating gaps and thick gray lines indicating amino acid sequence. Species are in the same order as panel C, which shows a zoomed-in view of the region of greatest contiguous sequence similarity. Residues aligned across all proteins with no gaps are colored in blue or red according to relative entropy, with red indicating more highly conserved (entropy threshold = 2 bits) (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi). Boundaries of DUF4582 are indicated relative to their annotation in the mouse protein. Protein accession numbers are in **Table S2**.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure S1. YTHDC2 staining with an independent anti-YTHDC2 antibody in wild-type and mutant testes.

YTHDC2 and SYCP3 immunofluorescence on testis sections from 2-month-old *Ythdc2*$^{ketu/ketu}$ and wild-type littermates. Figure labels are as in **Figure 5**. Scale bars represent 100 μm (A) or 20 μm (B). In this figure, the anti-YTHDC2 antibody from Santa Cruz Biotechnology was used (Abby et al., 2016). In **Figure 5**, the anti-YTHDC2 antibody from Bethyl Laboratories was used (Soh et al., 2017).

Supplementary Figure S2. Domain architecture of YTHDC2 and related DExH-box helicases.

(A) Schematics of domains in YTHDC2 and related DExH-box helicases (not to scale). Domain annotations were obtained from SMART (Letunic et al., 2015) and pfam database (Finn et al., 2016) searches, except TDRD9, for which domains were as previously defined (Handler et al., 2011). DHX9/RHA contains two copies of a double-stranded RNA binding domain (dsRBD) (Fu and Yuan, 2013). DHX57 has a ubiquitin-associated (UBA) domain and an RWD domain whose functions have yet to be demonstrated. TDRD9 has an RNA recognition motif (RRM) typically involved in interacting with nucleic acids, and a Tudor domain that binds dimethylated arginine (Handler et al., 2011). **(B)** Superposition of YTH domain structures of human YTHDC2 and *S. pombe* Mmi1 in either the apo state (left, PDB ID: 5H8A) or RNA-bound state (right, PDB ID: 5DNO). A closer view of the interaction of *S. pombe* Mmi1 Tyr-466 with the DSR nucleobase A_4 is also shown.

Supplementary Figure S3. Phylogenetic distribution of YTHDC2 in Metazoa.

The same tree topology is reproduced in **Figure 7C**.

Supplementary Figure S4. Full length alignment of YTHDC2 orthologs and paralogs.

Clustal Omega was used to align YTHDC2 or Bgcn sequences from the indicated species. Domains and helicase core motifs are indicated, defined according to the annotation for the mouse protein. Accession numbers are in **Table S2**.

Supplementary Figure S5. The YTHDC2 ortholog in nematodes has a distinct and highly diverse sequence in the location equivalent to the ARD insertion in other species.

Phylogram **(A)** and pairwise sequence identity matrix **(B)** for multiple sequence alignments of the regions between the DEXDc and HELICc domains of the YTHDC2 orthologs from the indicated nematode species (green shaded portion of tree), with mammalian and avian examples for comparison (coral shading). Because of the substantial divergence between these protein segments, sequences were aligned using MUSCLE and the VTML200 substitution matrix. The neighbor-joining tree was constructed with the vertebrate orthologs as the outgroup. Distances were calculated using the scoredist function. Accession numbers are in **Table S2**.

LIST OF SUPPLEMENTARY TABLES

Supplementary Table S1. Protein accession numbers for Figure 6A.

Supplementary Table S2. Protein accession numbers for Figures 6B–D, 7 and 8.

Supplementary Table S3. Genotyping primers.

Figure 1D−Source Data 1. Number of SYCP3-staining cells.

Figure 2D−Source Data 1. *Ythdc2* and *Meioc* RPKM values.

Figure 3A−Source Data 1. Testes and body weights.

Figure 1

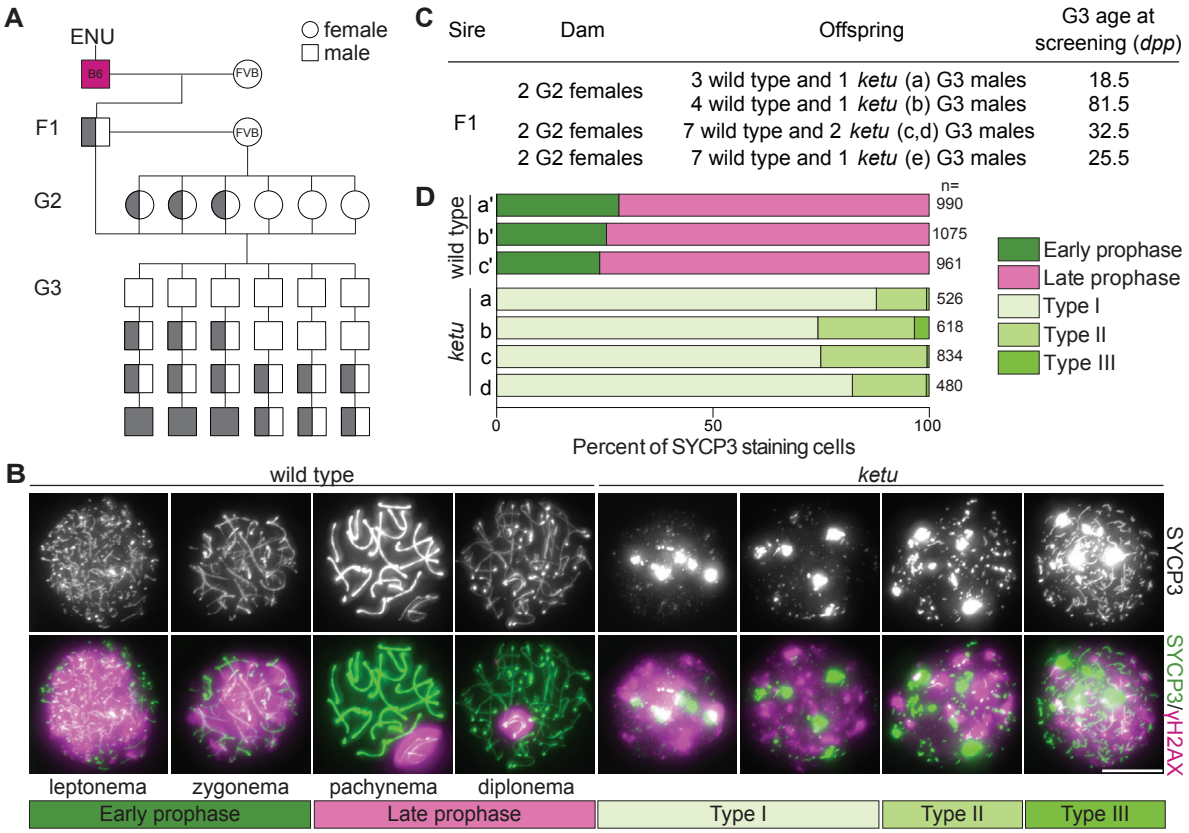


Figure 2

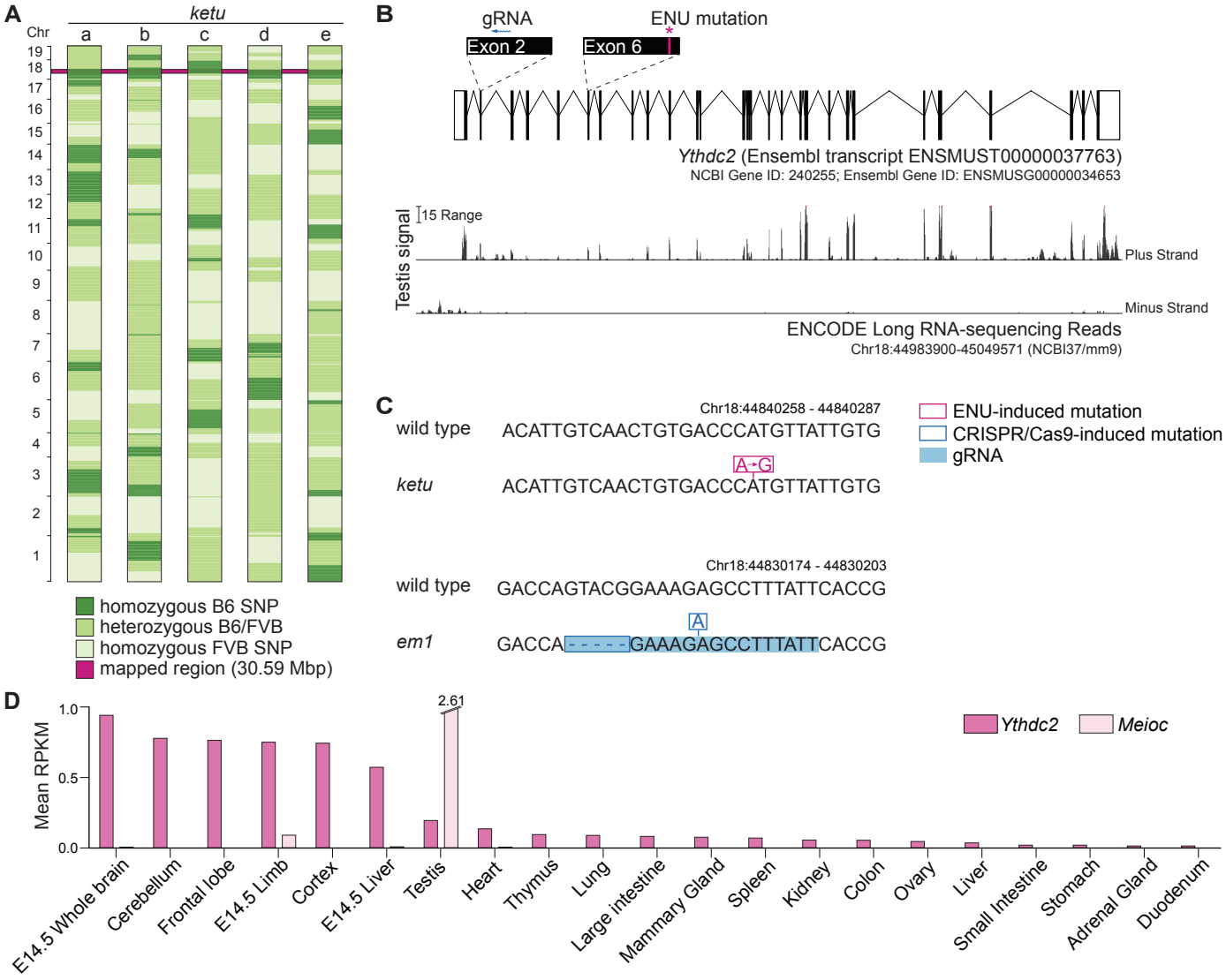


Figure 3

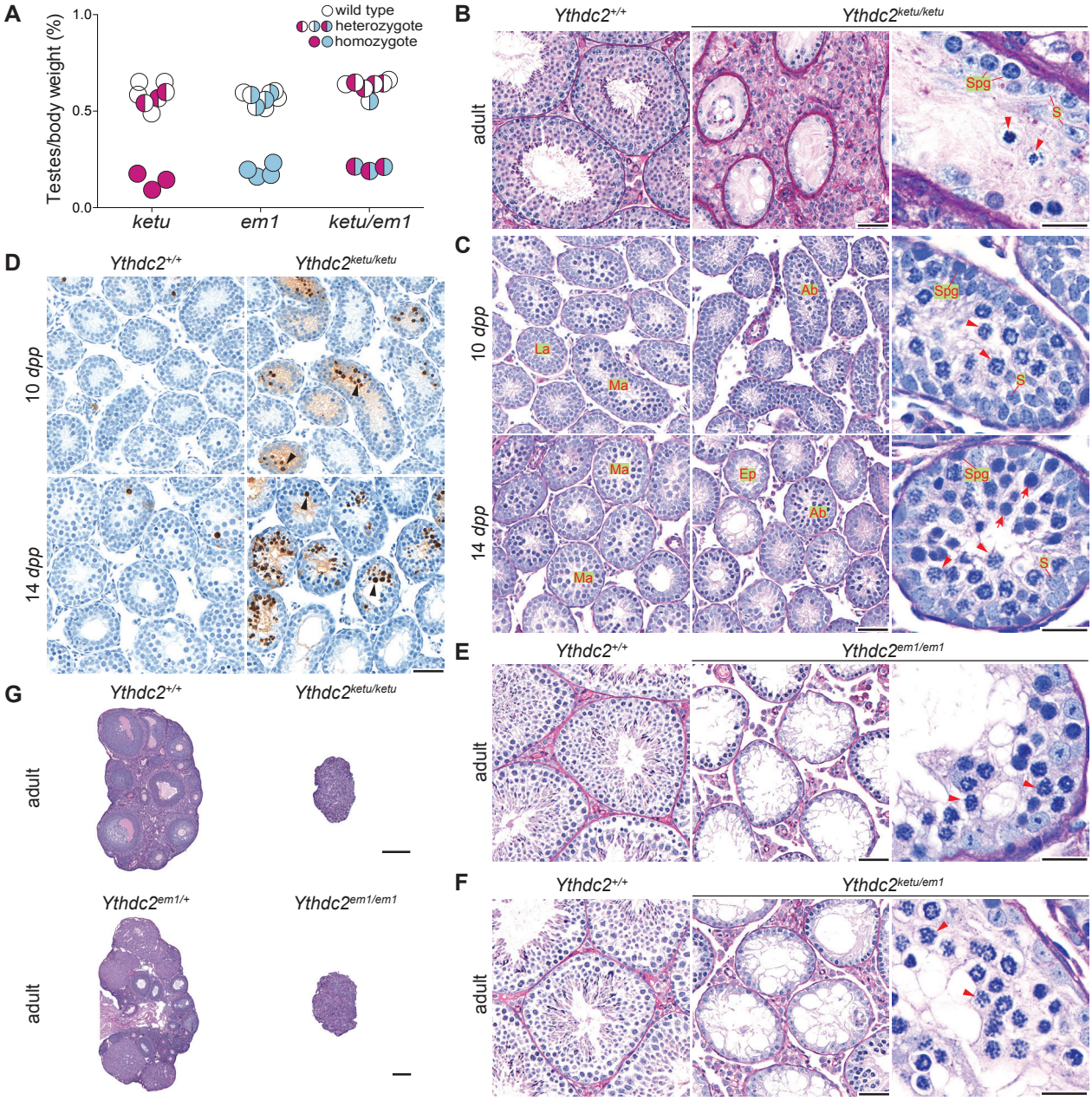


Figure 4

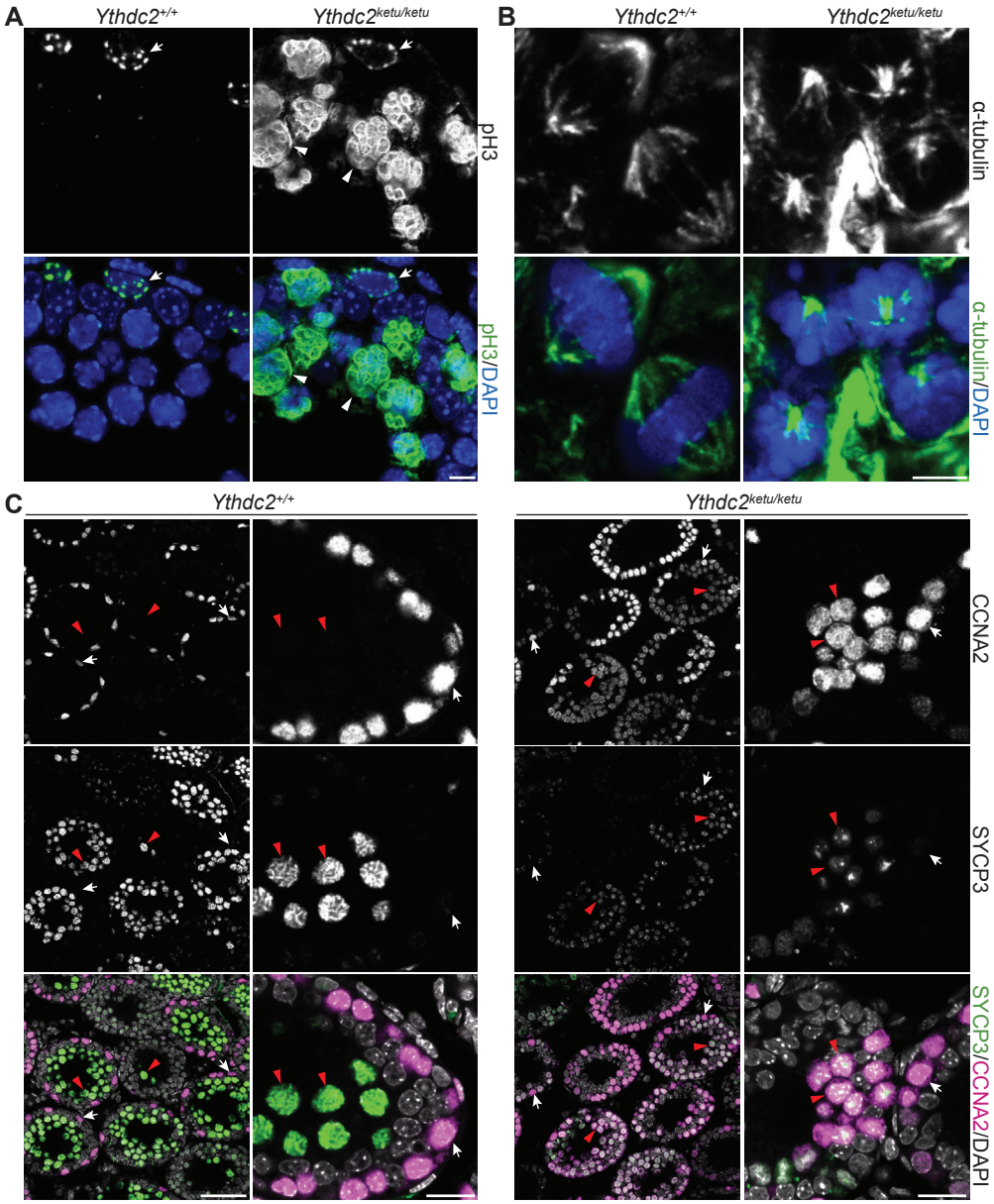


Figure 5

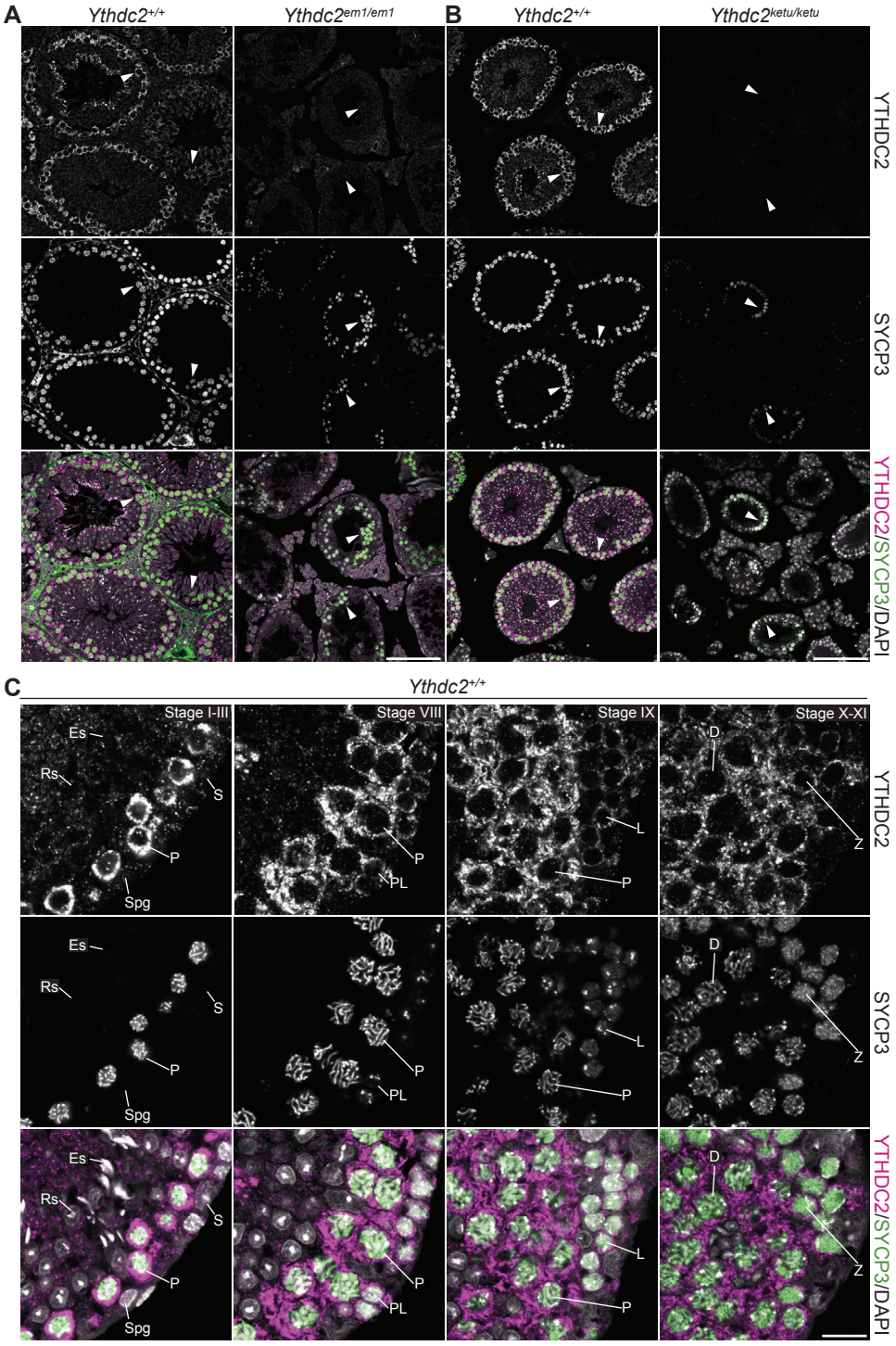


Figure 6

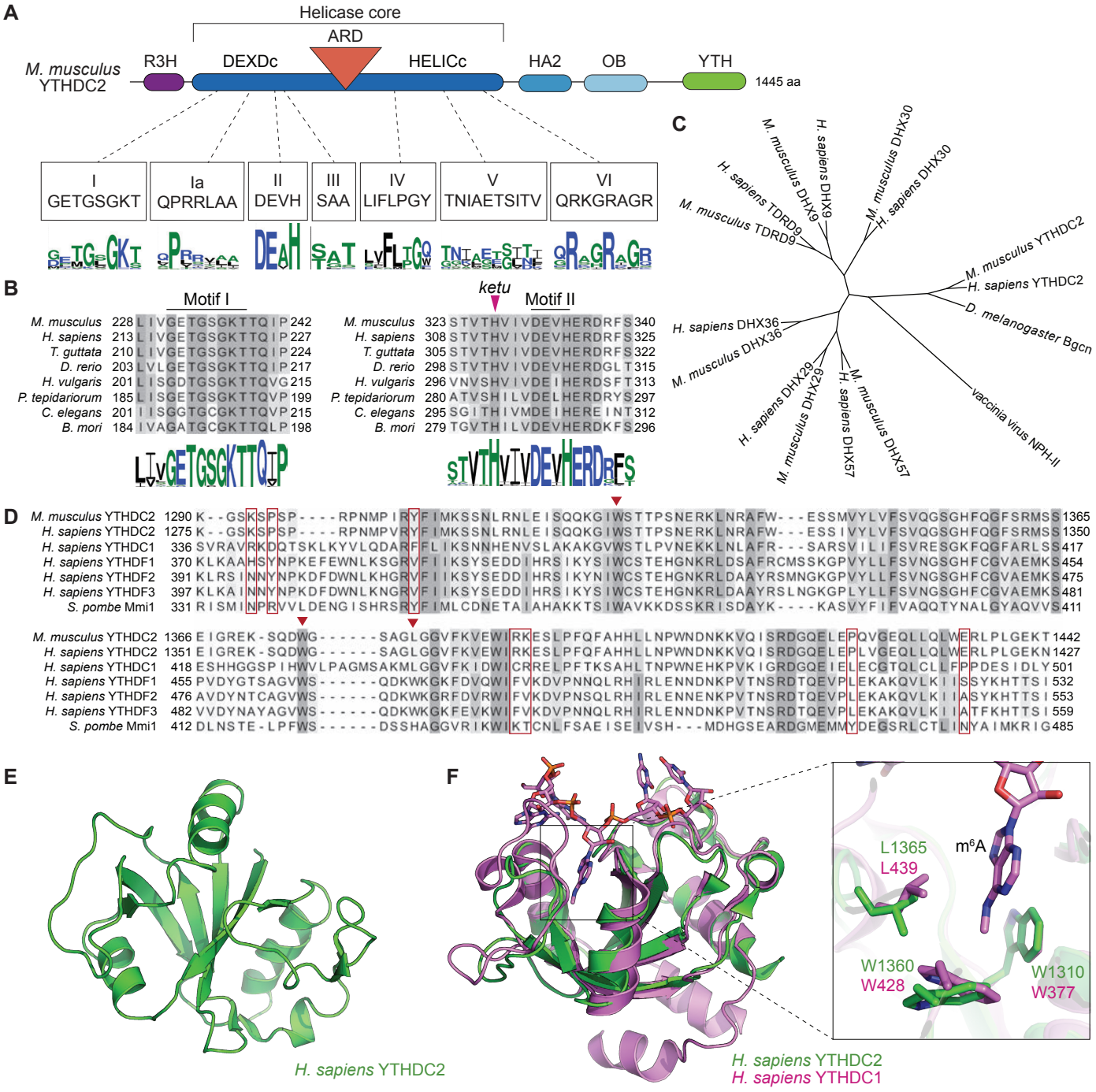


Figure 7

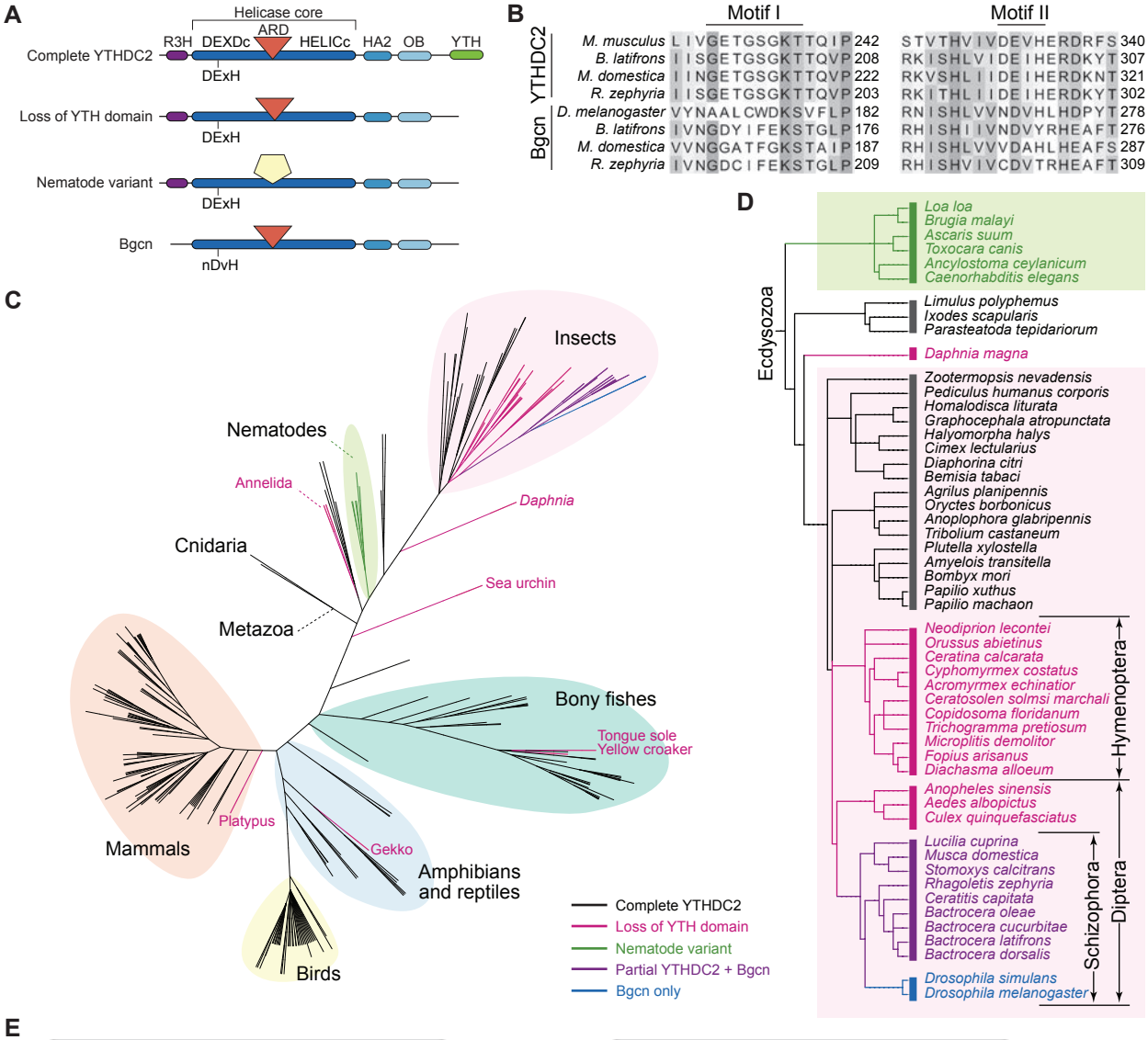
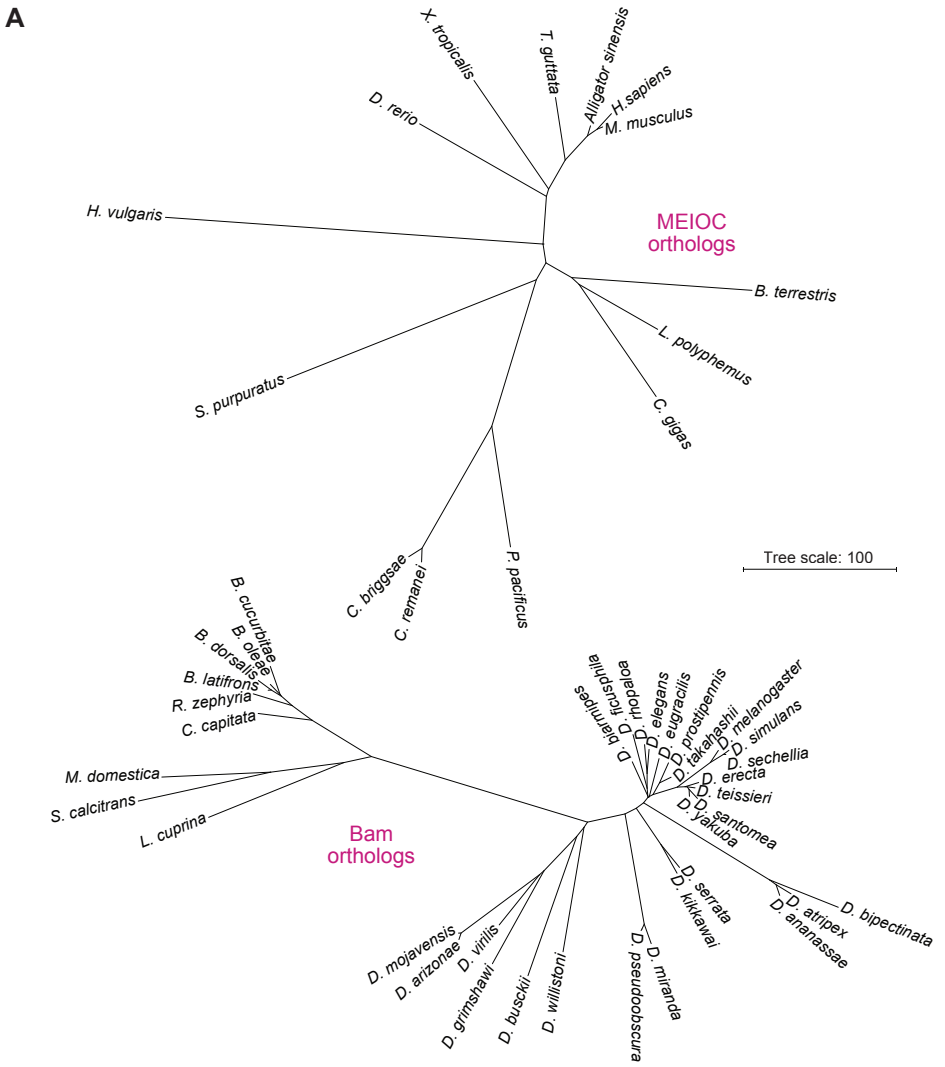
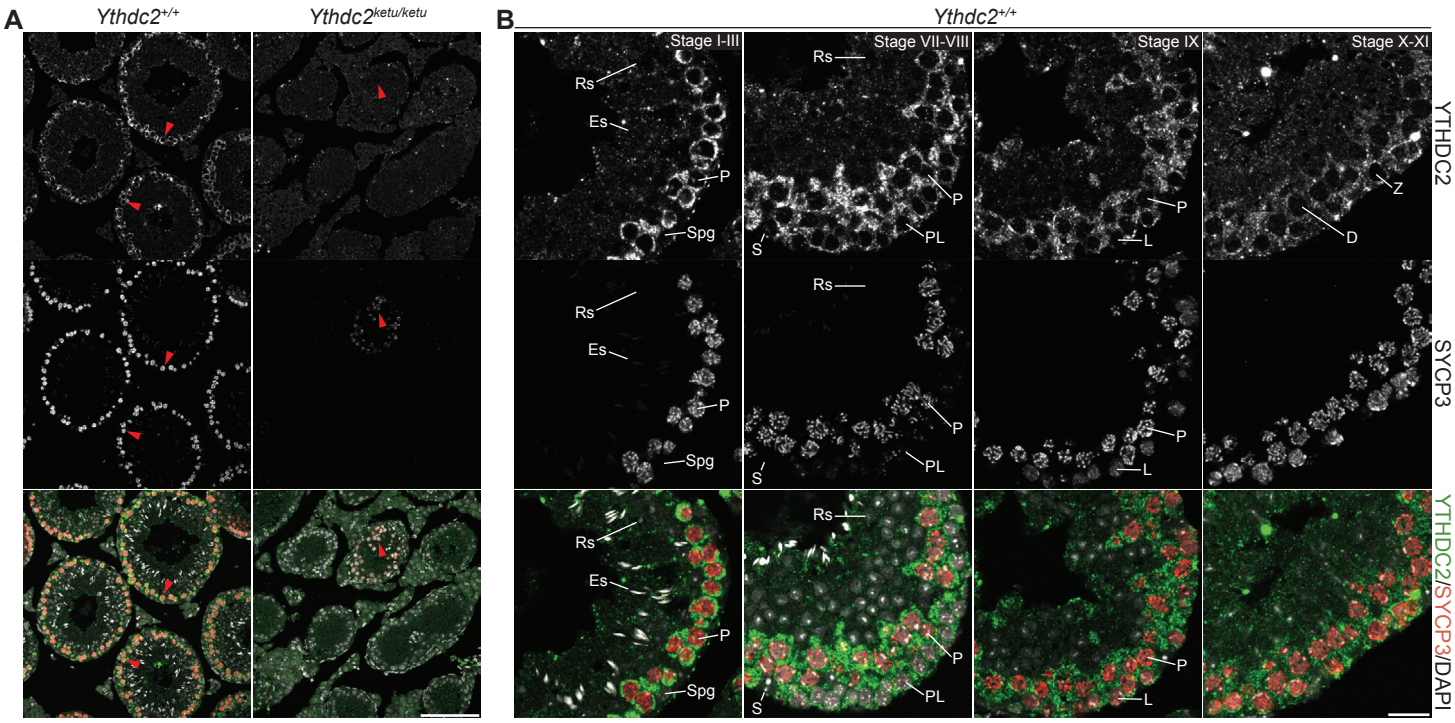


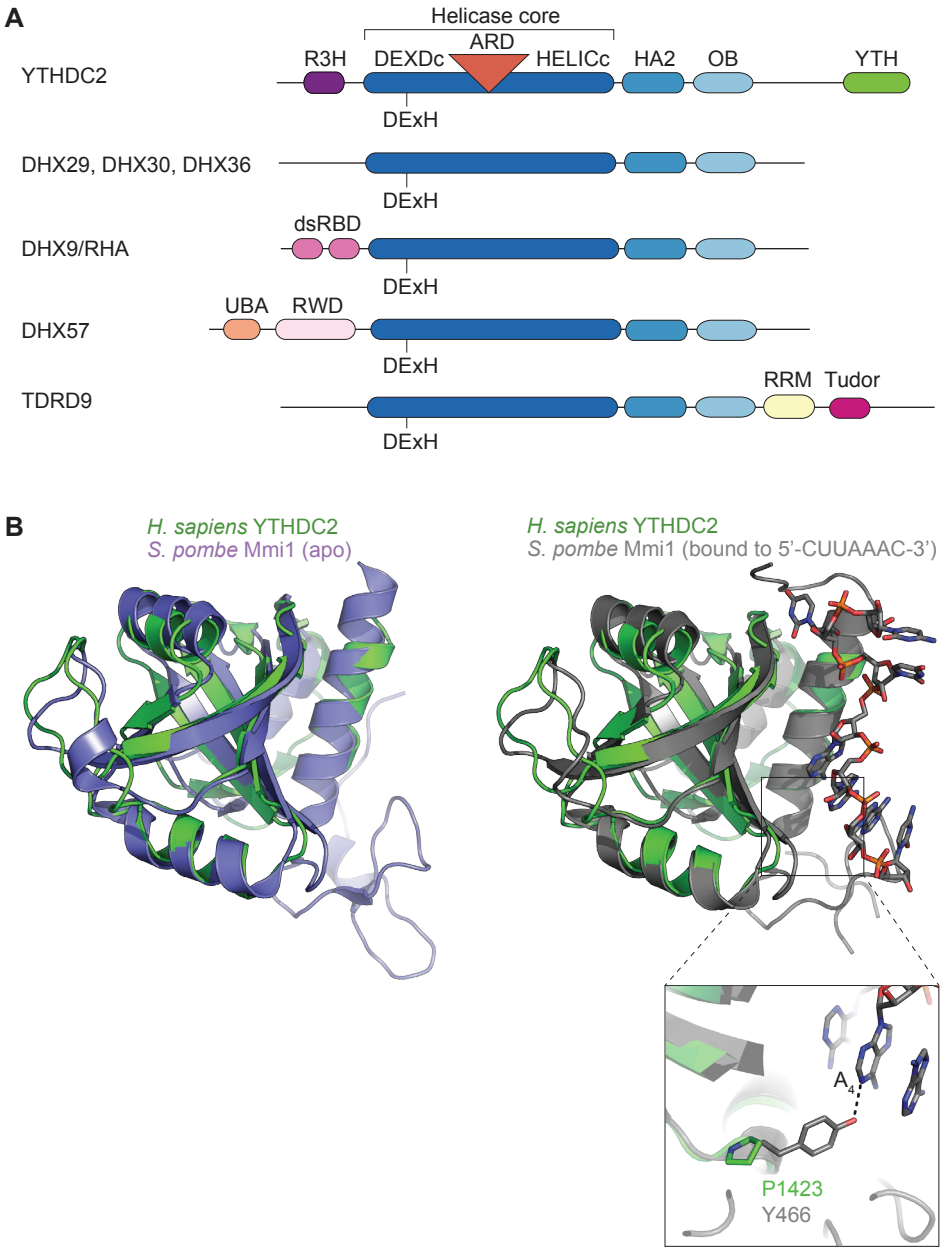
Figure 8



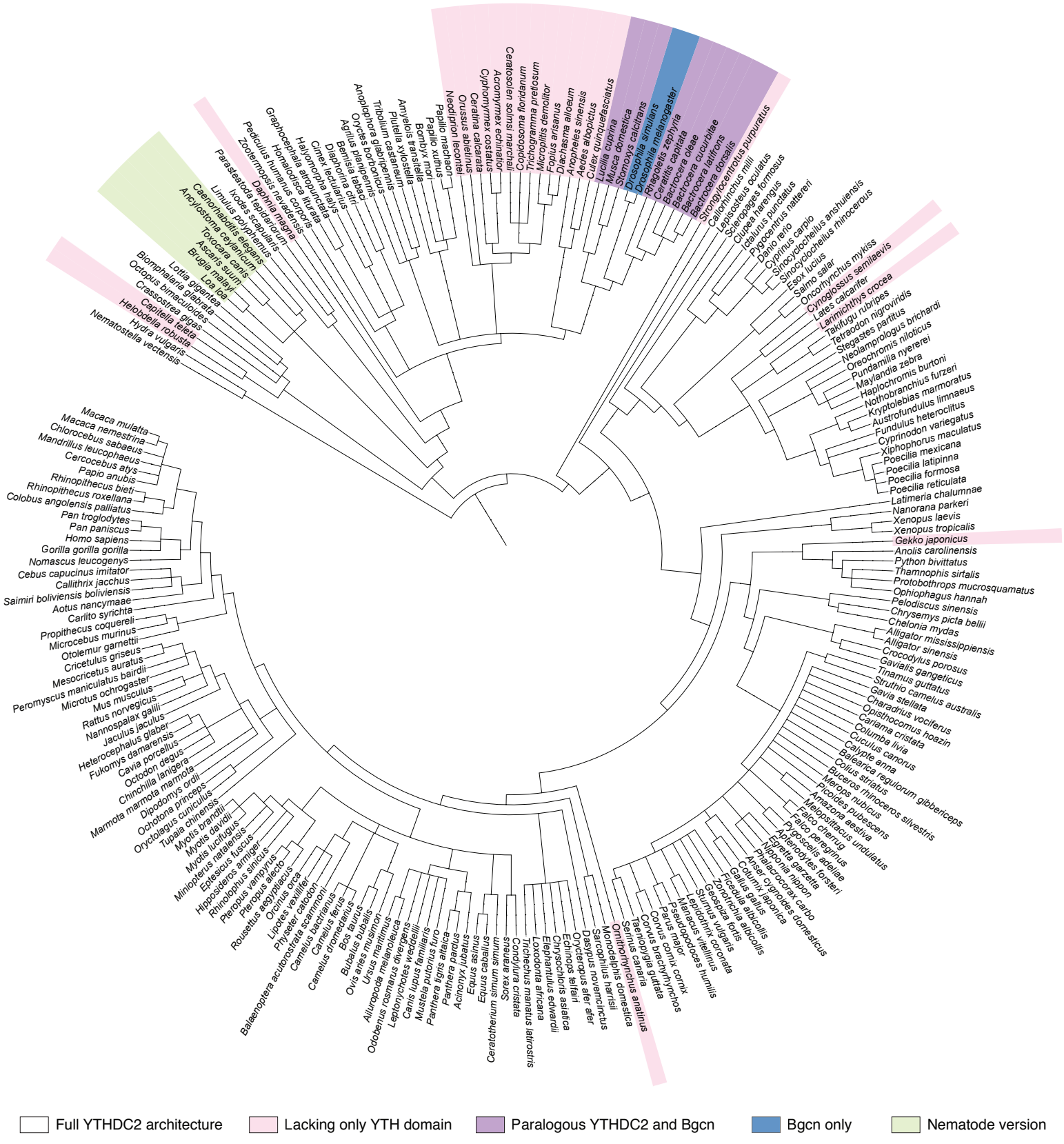
Supplementary Figure S1



Supplementary Figure S2

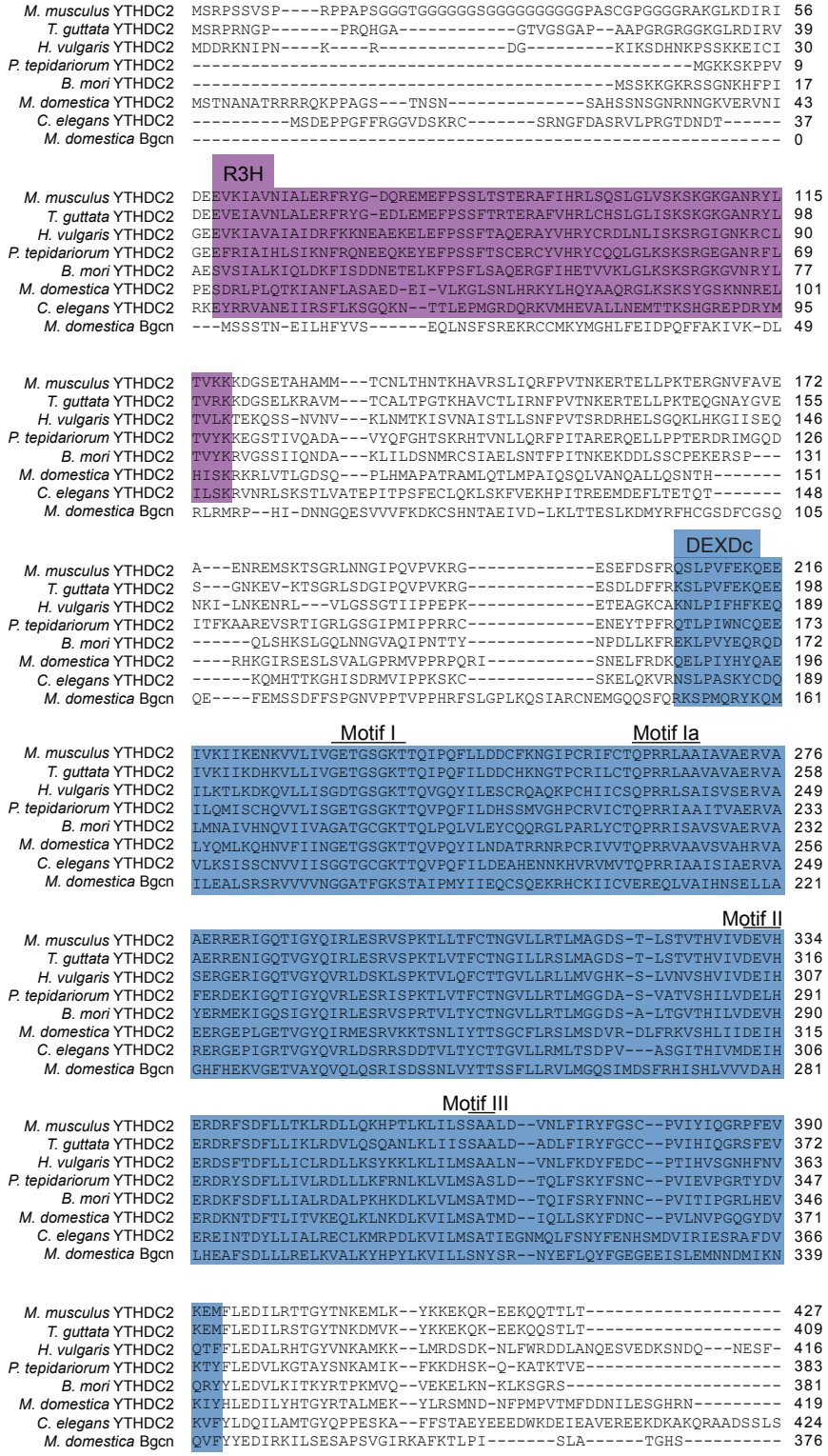


Supplementary Figure S3



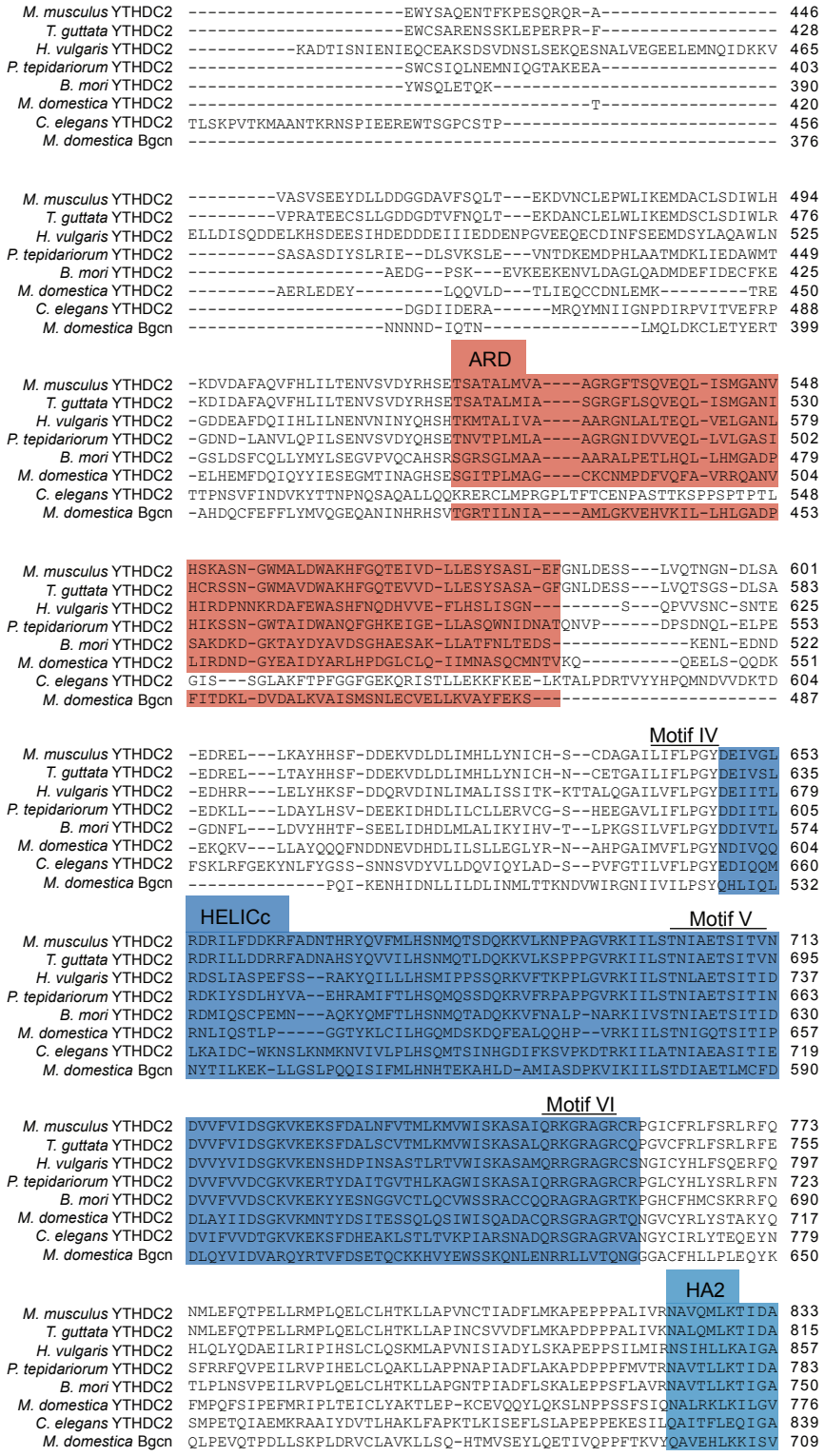
Supplementary Figure S4

Part 1 of 4



Supplementary Figure S4

Part 2 of 4



Supplementary Figure 3A

Part 3 of 4

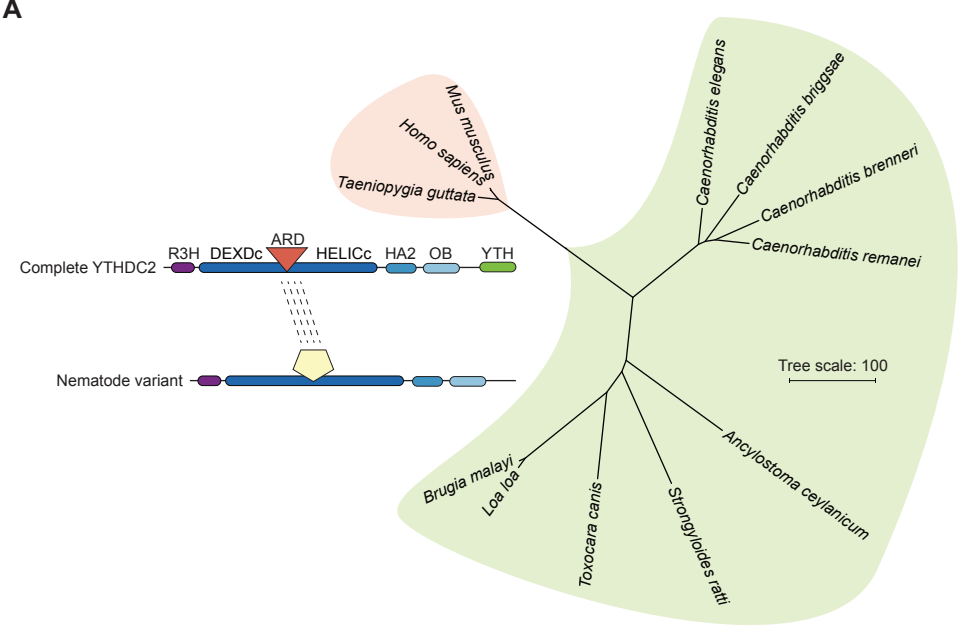
Supplementary Figure S4

Part 4 of 4

```
 M. musculus YTHDC2     EELPLASSWRSNNS--RKSTA-D-TEFADGSTTGERVLMKSPSPALHPPQKYKDRGILHP  1255
 T. guttata YTHDC2      EDHLVPSTWRSANT--RST-T-E-TEFSDSS-NAEKDLMRSARPILHQPKKYKRKNTLHS  1235
 H. vulgaris YTHDC2     --------KRSDKF--RYKSG-----------PHEGLRKTGNYEIAIP--DWRTRQTRGK  1273
 P. tepidariorum YTHDC2 YNSS-P-SHRSSGF--HNSNS-SYSVLASPCCPEDSPMMAGNSASNCYNVVYKNRGPQSP  1204
 B. mori YTHDC2         DSPN----WRMRVEEDQYRHDKYYPQYSQP-----------EYQNTFYPNYFRNGQGRPG  1154
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      SENPYASIVRNDNN--RRNHS-------D------------------PNQKSTERESRKP  1267
 M. domestica Bgcn      KLMMT-------LN------G----------------------GMLMAPCLYRSQFQTSA  1100

                                                                               YTH
 M. musculus YTHDC2     KRSTDDRSDQSSVKSTDSSSYPSPCASPSPPSSGKGS------KSPSPRPNMPIRYFIMK  1309
 T. guttata YTHDC2      KQTSDDRSDR-SVKSADSSSFPSPCASPSSPISGKAS------KSPSPRQNTPVRYFIMK  1288
 H. vulgaris YTHDC2     K----------VVADQDDSVLR---DEKGVQINTENE------AGCSEVKRRPASYFIMK  1314
 P. tepidariorum YTHDC2 MRQ----------------PYNSMQQMPA-----NME------YGHPTQTNSRVRYFIIK  1237
 B. mori YTHDC2         WRG-D------VMPQGFSPQYN----GPKSPMSPNGKVMMPNMEKYGEGAEGNAKQFMIK  1203
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      KEA---------VSHGNSRKQS----KPPKNHQTKTTSNFP-------SHNFPPPDLMRN  1307
 M. domestica Bgcn      ------------V----SKNMPTPNNVP--------------------VPQRSKQYFMVY  1124

 M. musculus YTHDC2     SSNLRN-LEI--SQQKG---------IWSTTPSNERKLN--RAFWESSMVYLVFSVQ---  1352
 T. guttata YTHDC2      SSNLQN-IDI--SQQKG---------IWSTTPSNERKLN--GAFWESSMVYLIFSVQ---  1331
 H. vulgaris YTHDC2     CNNDKN-MSI--SFERN---------IWATTRGNEKRLN--RAFNESDEVFLIFSVQ---  1357
 P. tepidariorum YTHDC2 ANTSRV-VEI--SMAKN---------IWAFTPTTERKLL--SAVKEGKEVILIFSIQ---  1280
 B. mori YTHDC2         VDELHA-LEA--AYSTG---------SYVFAQNIAKRLQ--KAKMEGHRVTVFFSCP---  1246
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      SHYNQN-MST--SHQSSSSNNHHYSHNYSYPHYNIQLPQNQYQSWTPQ--FDFHSMNTSY  1362
 M. domestica Bgcn      RKDLNDVVGVDMIMNTSSLRHS----EWFHLLNNLLDAK----TDCNRLTFIIFYTR---  1173

 M. musculus YTHDC2     -GSGHFQGFSRMSSEIGREKSQDWGSAGLGGVFKVEWIRKESLPFQFAHHLLNPWNDNKK  1411
 T. guttata YTHDC2      -GSGHFQGFARMGSSIGCEKSQDWGSAGFGGVFKVEWIRKESIPFQFAHHLLNPWNDNKK  1390
 H. vulgaris YTHDC2     -GSGHFQGVAKMTSEIGDRRCEDFGSLNLGGLFNIEWIHQEEIAFQYTQHLCNPWNDNKK  1416
 P. tepidariorum YTHDC2 -GSGHFQGFAHFTHHLSNQKPPEFSAANLGTCYQIEWMKKANIPFQSTRHLQNSWNENRK  1339
 B. mori YTHDC2         -SANKFVGCANLLDVS----------------GALEWISTQHVPYHMVRHSCGGARGC--  1287
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      GEHGH-NGNHQYMESSGIHGHIP----YQGGNFNTRGVSRGGFSYTSGNQ----------  1407
 M. domestica Bgcn      --------------------HPD----LLRSVYMVDWNKGYLELYP-----CFKFNINLH  1204

 M. musculus YTHDC2     VQISRDGQELEPQVGEQLLQLWERLPLGE--KTTSD------------------------  1445
 T. guttata YTHDC2      VQISRDGQELEPQVGEQLLQLWNQIPLE--------------------------------  1418
 H. vulgaris YTHDC2     VQISRDAQELETNVGKSLVEMWLHNSTLTKDLEQ----NEENLEQQDLNSPSLAKNNVVE  1472
 P. tepidariorum YTHDC2 VQISRDGQELEPNVGESLCKLISKCADVLGPEYTKVPRRQNAYPRENFDSNQLNRSQ---  1396
 B. mori YTHDC2         --VWREGELCGAG-GGALLAH--------------SRRRHARHPH---------------  1315
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      --------FRGRGLSRK---LYEH-----------NPRNE--------------------  1425
 M. domestica Bgcn      CVL---------------------------------------------------------  1207

 M. musculus YTHDC2     ------------------------------------------------------------  1445
 T. guttata YTHDC2      ------------------------------------------------------------  1418
 H. vulgaris YTHDC2     EAFYSDEVYSSNDPAYMDGYTLYSPPPFSPQYMPEFQHIPPFGHPQVFQQPYPMYAPPGG  1532
 P. tepidariorum YTHDC2 -A-----HWEQNT---AASSTWYDPAPGSNQ-----------------------------  1418
 B. mori YTHDC2         ------------------------------------------------------------  1315
 M. domestica YTHDC2    ------------------------------------------------------------  1167
 C. elegans YTHDC2      ------------------------------------------------------------  1425
 M. domestica Bgcn      ------------------------------------------------------------  1207

 M. musculus YTHDC2     -----------------------------------------  1445
 T. guttata YTHDC2      -----------------------------------------  1418
 H. vulgaris YTHDC2     QIYRHQDYYIPQYPPPRRNTAHDYYRDEGWDPRNSRGNYRQ  1573
 P. tepidariorum YTHDC2 ------KY---------------------WSPPNS------  1426
 B. mori YTHDC2         -----------------------------------------  1315
 M. domestica YTHDC2    -----------------------------------------  1167
 C. elegans YTHDC2      -----------------------------------------  1425
 M. domestica Bgcn      -----------------------------------------  1207
```

Supplementary Figure 35

A



B

Pairwise % identity	*H. sapiens*	*T. guttata*	*C. elegans*	*C. briggsae*	*C. remanei*	*C. brenneri*	*A. ceylan.*	*S. ratti*	*T. canis*	*L. loa*	*B. malayi*
M. musculus	98.0	82.3	22.4	18.4	17.0	21.8	16.3	21.1	19.0	14.3	13.6
H. sapiens		83.7	22.4	19.0	17.7	21.8	16.3	21.8	18.4	14.3	13.6
T. guttata			22.4	17.7	16.3	17.7	17.0	21.1	18.4	14.3	12.2
C. elegans				49.7	59.9	53.1	21.8	23.1	19.7	19.7	21.1
C. briggsae					55.1	49.0	18.4	18.4	18.4	17.7	17.7
C. remanei						59.2	23.1	16.3	19.0	18.4	19.0
C. brenneri							20.4	19.0	19.7	21.1	21.8
A. ceylanicum								27.2	28.6	27.9	27.2
S. ratti									32.0	27.9	28.6
T. canis										35.4	36.7
L. loa											90.5