

# NKAPL facilitates transcription pause-release and bridges elongation to initiation during meiosis exit

Received: 4 July 2023

Accepted: 16 December 2024

Published online: 17 January 2025



Zhenlong Kang<sup>1,12</sup>, Chen Xu<sup>1,12</sup>, Shuai Lu<sup>1,2,3,12</sup>, Jie Gong<sup>1,12</sup>, Ruoyu Yan<sup>1,12</sup>, Gan Luo<sup>1,12</sup>, Yuanyuan Wang<sup>1</sup>, Qing He<sup>1</sup>, Yifei Wu<sup>1</sup>, Yitong Yan<sup>4</sup>, Baomei Qian<sup>5</sup>, Shenglin Han<sup>1</sup>, Zhiwen Bu<sup>1</sup>, Jinwen Zhang<sup>1</sup>, Xian Xia<sup>6</sup>, Liang Chen<sup>7</sup>, Zhibin Hu<sup>1,2</sup>, Mingyan Lin<sup>4</sup>✉, Zheng Sun<sup>8</sup>✉, Yayun Gu<sup>1,9,10,11</sup>✉ & Lan Ye<sup>1,10,11</sup>✉

Transcription elongation, especially RNA polymerase II (Pol II) pause-release, is less studied than transcription initiation in regulating gene expression during meiosis. It is also unclear how transcription elongation interplays with transcription initiation. Here, we show that depletion of NKAPL, a testis-specific protein distantly related to RNA splicing factors, causes male infertility in mice by blocking the meiotic exit and downregulating haploid genes. NKAPL binds to promoter-associated nascent transcripts and co-localizes with DNA-RNA hybrid R-loop structures at GAA-rich loci to enhance R-loop formation and facilitate Pol II pause-release. NKAPL depletion prolongs Pol II pauses and stalls the SOX30/HDAC3 transcription initiation complex on the chromatin. Genetic variants in *NKAPL* are associated with azoospermia in humans, while mice carrying an *NKAPL* frameshift mutation (M349fs) show defective meiotic exit and transcriptomic changes similar to NKAPL depletion. These findings identify NKAPL as an R-loop-recognizing factor that regulates transcription elongation, which coordinates the meiotic-to-postmeiotic transcriptome switch in alliance with the SOX30/HDAC3-mediated transcription initiation.

Gene transcription by RNA polymerase II (Pol II) in mammals is a delicately regulated process involving transcription initiation, Pol II pausing, elongation, and termination<sup>1,2</sup>. The transition of Pol II from a paused state to productive elongation is a critical rate-limiting step in transcriptional regulation. Thousands of Pol II-transcribed genes in

metazoan contain a promoter-proximal paused site in the region 20–100 nt downstream of their transcription start sites (TSS)<sup>3,4</sup>. Release of paused Pol II into productive elongation requires hyperphosphorylation of the C-terminal repeat domain (CTD) of Pol II at serine-2<sup>5,6</sup>. Positive elongation factor b (P-TEFb), comprising cyclin-

<sup>1</sup>State Key Laboratory of Reproductive Medicine and Offspring Health, Nanjing Medical University, Nanjing, China. <sup>2</sup>Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China. <sup>3</sup>Changzhou Maternity and Child Health Care Hospital, Changzhou Medical Center, Nanjing Medical University, Nanjing, China. <sup>4</sup>Department of Neurobiology, School of Basic Medical Science, Nanjing Medical University, Nanjing, People's Republic of China. <sup>5</sup>Reproductive and Genetic Hospital, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China. <sup>6</sup>Jiangsu Key Laboratory of Neurodegeneration, Department of Pharmacology, Nanjing Medical University, Nanjing, China. <sup>7</sup>RNA Institute, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China. <sup>8</sup>Department of Medicine, Baylor College of Medicine, Houston, TX, USA. <sup>9</sup>Reproductive Genetic Center, The Affiliated Suzhou Hospital of Nanjing Medical University, Gusu School, Suzhou, Jiangsu, China. <sup>10</sup>Innovation Center of Suzhou Nanjing Medical University, Suzhou, Jiangsu, China. <sup>11</sup>National Center of Technology Innovation for Biopharmaceuticals, Suzhou, Jiangsu, China. <sup>12</sup>These authors contributed equally: Zhenlong Kang, Chen Xu, Shuai Lu, Jie Gong, Ruoyu Yan, Gan Luo. ✉e-mail: [linmingyan@njmu.edu.cn](mailto:linmingyan@njmu.edu.cn); [zheng.sun@bcm.edu](mailto:zheng.sun@bcm.edu); [yayungu@njmu.edu.cn](mailto:yayungu@njmu.edu.cn); [lanye@njmu.edu.cn](mailto:lanye@njmu.edu.cn)

dependent kinase 9 (CDK9) and Cyclin T1, mediates CTD serine-2 phosphorylation and promotes the initial recruitment of other elongation factors, including Pol II-associated factor 1 (PAF1)<sup>7–9</sup>. Components of elongation polymerase complexes undergo dynamic phosphorylation and dephosphorylation, and promoter-proximal stalled Pol II produces significant levels of short nascent RNAs. These nascent RNAs can re-anneal with their template DNA strand and form three-stranded RNA/DNA hybrid structures (R-loops), leading to the tethering of Pol II to the chromatin<sup>10,11</sup>. R-loops are prevalent over G/C-rich promoters and terminator regions<sup>12–15</sup>, and R-loop induction is linked with Pol II pausing at gene promoters<sup>16</sup>, suggesting that R-loops may regulate gene expression.

Serine and arginine-rich (SR) family proteins constitute a family of pre-mRNA splicing factors<sup>17</sup> with additional roles in RNA metabolism and transcriptional elongation<sup>18,19</sup>. SR and SR-like proteins contain SR domains enriched in arginine and serine amino acids, but SR-like proteins lack an RNA-recognition motif (RRM) that modulates interactions with RNA. Biochemical evidence suggests that SR proteins associate with the Pol II CTD<sup>20,21</sup>, which is necessary for the recruitment of SR proteins to sites of transcription. In vivo depletion of the SR protein ASF/SF2 causes genome instability, while ASF/SF2 recruitment to nascent transcripts by Pol II prevents the formation of mutagenic R-loop structures<sup>22</sup>. Another SR protein SRSF2 is recruited to active gene promoters to release paused Pol II<sup>18</sup>. Despite these advances, the role of SR-like proteins in transcription and its biological significance remains unclear.

The SR-like protein NKAPL is a retrotransposed homolog of NKAP. *Nkapl* is a testis-specifically expressed autosomal gene that lacks introns, while *Nkap* is an X-linked, ubiquitously-expressed gene that contains eleven exons and ten introns. Therefore, *Nkapl* could be an autosomal retroposed copy of *Nkap* that originates from the reverse transcription of a processed transcript by integration into the genome. Structurally different from SR family proteins with one or two RRMs, NKAPL and NKAP lack RRM. Instead, NKAPL and NKAP are composed of an N-terminal RS domain, repetitive basic sequences (the basic domain), and a C-terminal DUF926 domain of unknown functions<sup>23,24</sup>. A key feature of the basic domain is highly repetitive, and it contributes to a punctuate pattern of NKAP localization in HeLa cells. NKAP was originally identified as a possible regulator of NF- $\kappa$ B activation<sup>25</sup> and now is found to modulate cell type-specific transcription through HDAC3-dependent and -independent pathways<sup>23,26–28</sup>, demonstrating its role in transcriptional regulatory cascades is highly context-dependent<sup>26,29</sup>. NKAP was also found to play additional roles in mRNA splicing in HEK 293 T cells and mitotic progression<sup>24,30</sup>.

Meiotic exit in spermatogenesis requires reprogramming from meiotic to post-meiotic haploid gene programs. Accumulating evidence suggested distinct transcriptional regulatory mechanisms between somatic and germ cells, with a specific regulatory strategy in spermatogenesis<sup>31</sup>. Many distinct cis-acting regulatory elements are highly restricted to male germ cells. In mammals, testis-specific isoforms of the general transcription machinery components and TBP-like factors have been identified, and these specialized complexes correlate with the efficient and massive transcriptional activity at the beginning of spermiogenesis. Germ cells during late spermatogenesis present a robust accumulation of Pol II machinery components, with 100-fold higher levels of transcripts encoding TATA-binding proteins than somatic cells<sup>32</sup>. It is unclear how such a large transcription complex initiates at promoters and travels dynamically through the gene body during transcription at the meiotic exit. A previous study reported NKAPL is required for male fertility<sup>33</sup>, and most of germ cells in *Nkapl* deficient tubules were arrested at the pachytene stage. Here, we elucidate a role for NKAPL in meiotic exit, as genetic deletion of *Nkapl* in mice results in an arrest at the late stage of meiosis or early stage of round spermatids, which resembles testis-specific *Hdac3* knockouts<sup>34</sup> or *Sox30* null testes<sup>35–38</sup>. High-resolution profiles of

germline R-loops in this study combined with recently published R-loops in human and mouse cell types revealed that R-loops preferentially occur at GAA repeat-containing genomic loci. NKAPL binds prominently to promoter-associated RNA transcripts with GAA repeats, where it forms an optimal environment to drive R-loop formation and facilitate the release of paused Pol II. Our data revealed the role of NKAPL in R-loop formation and transcription elongation, with implications for human diseases because missense mutations in NKAPL in humans are associated with male infertility.

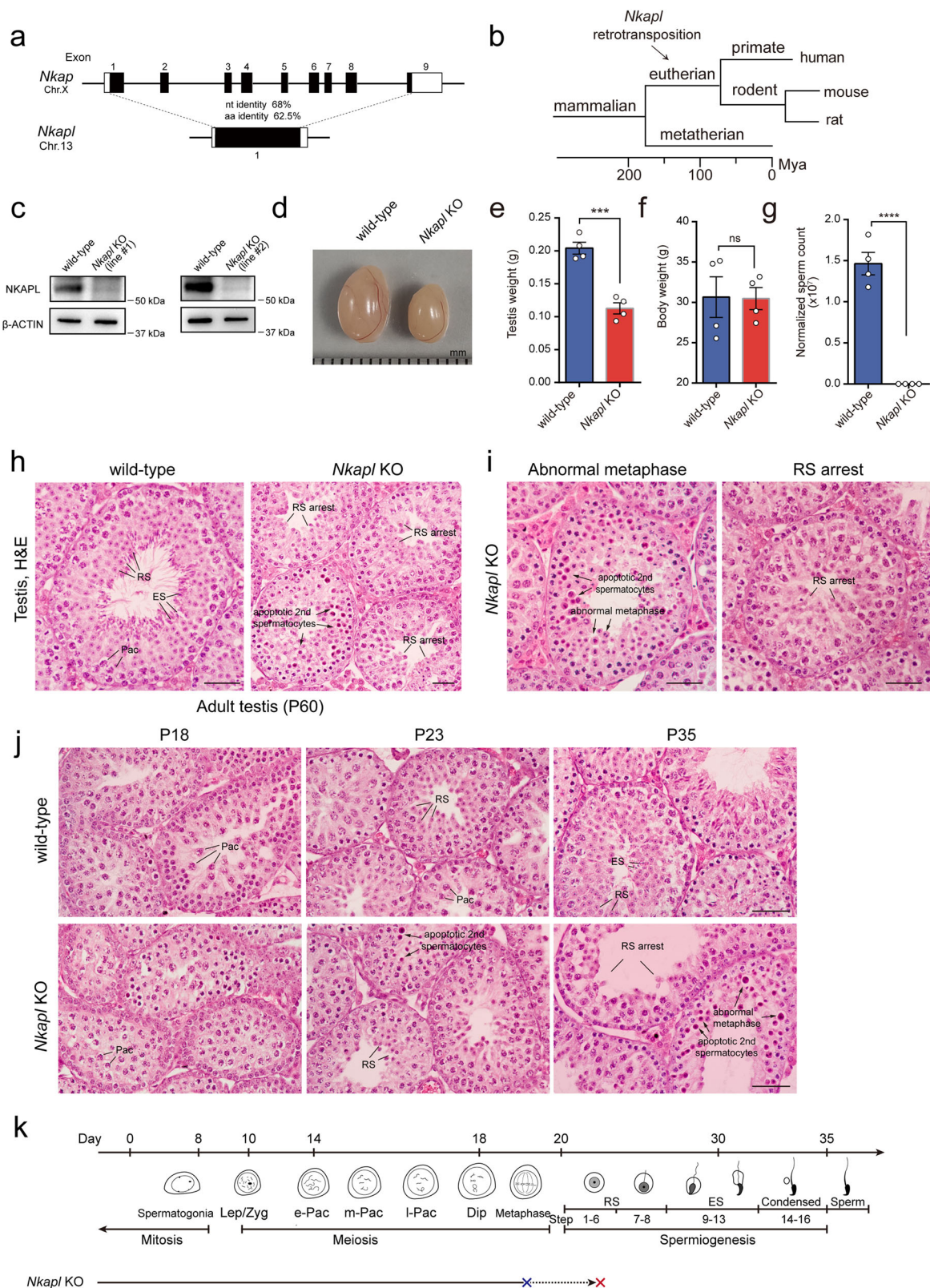
## Results

### *Nkapl* KO males exhibit infertility and defects in the meiotic exit

Mouse *Nkapl* is an autosomal gene that lacks introns and shares a conserved DUF926 domain with its X-linked progenitor *Nkap* at the C-terminus (Fig. 1a). Comparison of gene structure implies that *Nkapl* is a retrotransposed gene originating from its X-linked progenitor *Nkap*, and this retrotransposition event occurred before the divergence of the eutherians and metatherian (Fig. 1b). *Nkap* is a ubiquitously expressed gene with mRNA transcripts in various mouse tissues, but *Nkapl* is specifically restricted to the testis (Supplementary Fig. 1a), consistent with a previous report<sup>33</sup>. *Nkapl* mRNA was significantly increased in testis at postnatal day 18 (P18) and reached the highest level at P21 concurrent with the late stages of spermatocytes and early RS (Supplementary Fig. 1b). We generated three NKAPL polyclonal antibodies and found one against the RS and basic domains of murine NKAPL works for immunoprecipitation and western blot analysis. NKAPL protein migrated with a molecular mass of ~56 kDa (Supplementary Fig. 1c), and this band was observed only in testes (Supplementary Fig. 1c). Like its transcript, NKAPL protein was abundant in mouse testes at P18–P21 when later stages of spermatocytes and early RS were enriched, and started to decline by P28 (Supplementary Fig. 1d). To explore the physiological function of NKAPL in spermatogenesis, we generated *Nkapl* knockout mice using CRISPR/Cas9 genome editing technology (Supplementary Fig. 1e). Two small guide RNAs (gRNAs) were designed to target the coding region of *Nkapl*, and two independent alleles were obtained: *Nkapl-line 1* contained a 50-bp deletion in the exon 1, and *Nkapl-line 2* contained a single bp insertion immediately after the ATGT base of the start codon of *Nkapl* gene and a 13-bp deletion (Supplementary Fig. 1e, f). Both mutations in *Nkapl-line 1* and *Nkapl-line 2* produce premature termination codons. Western blot analysis of testes lysates further confirmed the absence of NKAPL protein in both homozygous mutants (Fig. 1c). Males from both *Nkapl-line 1* and *Nkapl-line 2* grew into adulthood with similar body size and appeared to be healthy. We examined three individual males from each *Nkapl-line* and found that all were sterile. Knockout of *Nkapl* resulted in markedly reduced testis size (Fig. 1d). The testes from 8-week-old *Nkapl* knockouts weighed 45% less than wild-type (Fig. 1e,f). Mature sperm were absent in the cauda epididymis of both *Nkapl* knockouts (Fig. 1g). Thus, we refer hereafter to both mutant lines as *Nkapl* KO mice.

Histological analysis showed wild-type seminiferous tubules contained a full spectrum of germ cells (Fig. 1h), whereas *Nkapl* KO tubules were narrow and lacked elongating spermatids and mature spermatozoa (Fig. 1h). At the epithelial stage XII, *Nkapl* KO spermatocytes showed heavily eosin-stained nuclei, with chromatin either structurally loose or very condensed (Fig. 1i). Aberrant secondary spermatocytes-like cells and metaphase spermatocytes with atypical nuclei were frequently observed (Fig. 1i). This represents a defect in the late stages of spermatocytes in the *Nkapl* KO mice. The remaining spermatocytes in the *Nkapl* KO tubules progressed through meiosis, with over 70% of tubules arrested at the round spermatid stage (Fig. 1i). This phenotype is different from the testicular defect of reported *Nkapl* knockout mice, which shows a complete meiotic arrest at the pachytene stage<sup>33</sup>.

To determine the onset time of defects in *Nkapl* knockout, we collected younger animals and examined the first wave of



spermatogenesis. Histological analysis of testes at P18 revealed that germ cell types were the same in the *Nkapl* KO and wild-type testes, and both wild-type and *Nkapl* KO germ cells have advanced to the pachytene stage of meiosis (Fig. 1j). However, *Nkapl* KO tubules at P23 contained many heavily eosin-stained cells at stage XII (Fig. 1j), confirming a similar defect to adult *Nkapl* KO at later stages of meiosis. Moreover, P35 testes showed the same defects in meiotic exit

observed in adult *Nkapl* KO mice (Fig. 1j). Consistent with the testis histology, no spermatozoa were present in epididymides from adult *Nkapl* KO mice. Instead, large numbers of degenerating germ cells were observed (Supplementary Fig. 1g). Collectively, *Nkapl* knockout males are infertile and exhibit defects in meiotic exit, as *Nkapl* null germ cells were arrested either at the late stages of spermatocytes or at the round spermatid stage (Fig. 1k).



**Fig. 1 | *Nkapl* KO males show infertility and defects during meiotic exit.**

**a** Schematic representation of mouse *Nkap* and *Nkapl* genes architecture. Exons were shown in black. **b** Evolution of the retrogene *Nkapl* by phylogenetic sequence analysis. **c** Western blot analysis of NKAPL protein in P21 wild-type and *Nkapl* KO testes. Experiments were performed in biological triplicates. **d** Significant size reduction in 8-week-old *Nkapl* KO males. **e–g** Testis weights (**e**), body weights (**f**) and epididymal sperm counts (**g**) of wild-type ( $n = 4$ ) and *Nkapl* KO ( $n = 4$ ) at 8-week-old. Data are presented as mean  $\pm$  SD. \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ , ns: not significant, two-tailed unpaired Student's *t* test. **h** Histological analysis of testes from 8-week-old wild type and *Nkapl* KO mice. Scale bars, 20  $\mu$ m. **i** Enlarged images of the *Nkapl* KO tubule. Apoptotic late stages of spermatocytes in stage XII tubules

of *Nkapl* KO are marked by arrows, and black lines indicate round spermatids. Scale bar, 20  $\mu$ m. **j** Histological analysis of juvenile *Nkapl* KO testes also revealed defects in meiotic exit in the first wave of spermatogenesis. Mouse testes from *Nkapl* KO mice at P18, P23 and P35 were collected. Biological duplicates were prepared for each time point during the first wave of spermatogenesis. Scale bars, 20  $\mu$ m. **k** A diagram representing arrested stages of germ cell development in *Nkapl* KO mice. Blue and red crosses on lines indicate the earliest and ultimate time point of spermatogenic arrest, respectively. Lep: Leptotene; Zyg: Zygotene; e-Pac: early-Pachytene; m-Pac: middle-Pachytene; l-Pac: late-Pachytene; Dip: Diplotene; RS: round spermatids; ES: elongating spermatids.

**NKAPL is specifically required for meiotic exit and transition to round spermatids**

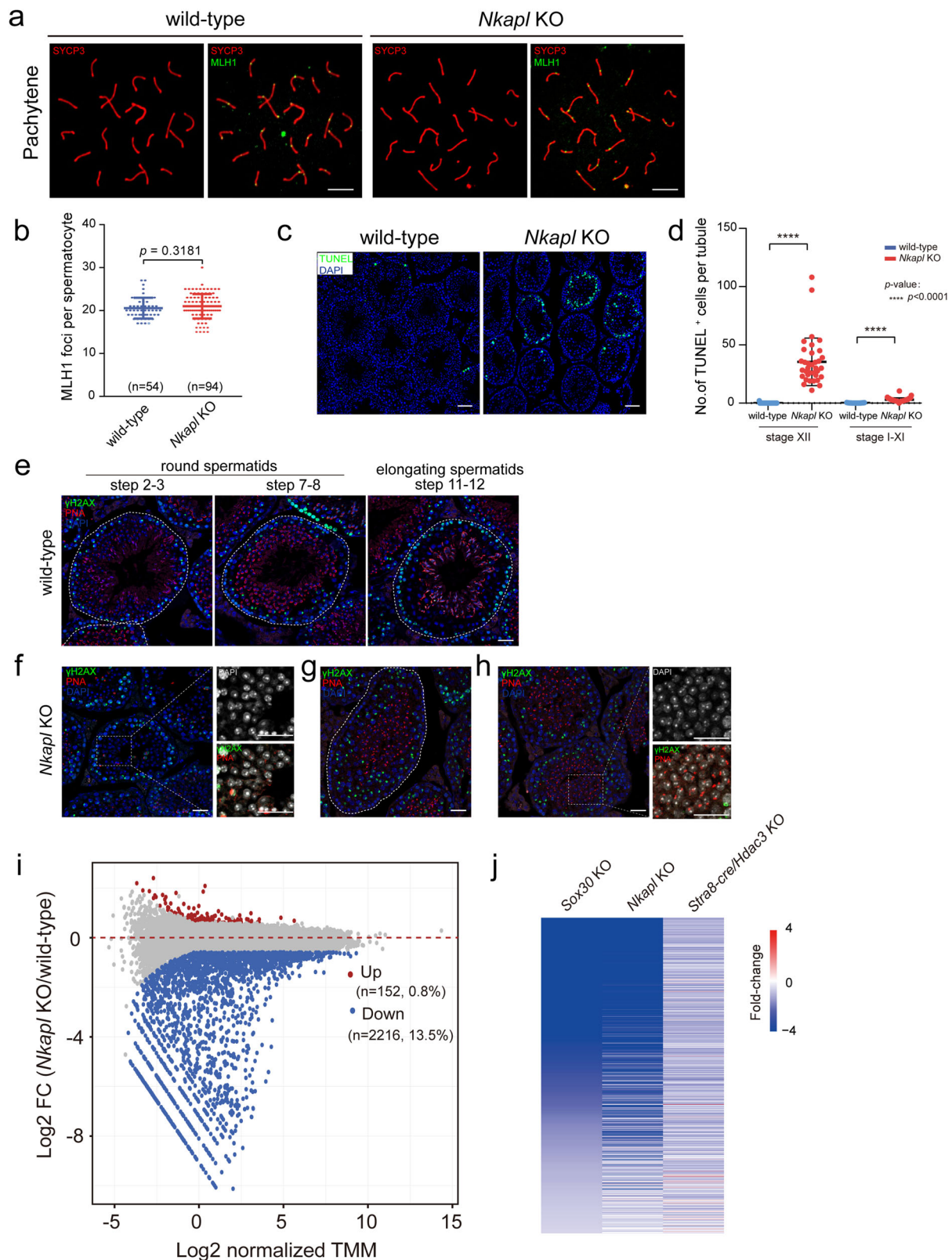
The key events in meiosis, such as chromosomal synapsis (SYCP1 and SYCP3; Supplementary Fig. 2a), DSB generation and meiotic recombination ( $\gamma$ H2AX; Supplementary Fig. 2b) and DSB repair (DMC1; Supplementary Fig. 2c) were not affected by the deletion of *Nkapl*, as evidenced by fully synapsed chromosomes and the distinguished XY body found at the pachytene stage both in wild-type and *Nkapl* KO. At the mid-late pachytene stage, MLH1 foci represent sites of crossovers, which was apparent in *Nkapl* KO mice similar to that in the wild-type (Fig. 2a, b). These results demonstrated that *Nkapl* KO spermatocytes advanced to the mid-late pachytene stage. Consistent with histological observations, many late stages of spermatocytes and secondary spermatocytes-like cells in *Nkapl* KO were undergoing apoptosis (Fig. 2c), as confirmed by an increase in stage XII spermatocytes positive for the TdT-mediated dUTP nick end labeling (TUNEL) assay (Fig. 2d). Apoptotic late-stage spermatocytes were observed in spermatid-arrested mutants including *Miw1* knockouts<sup>39</sup>. To further verify the role of NKAPL during meiotic exit, we examined the spindle dynamics and chromosome alignment since the above histological analysis revealed the presence of aberrant metaphase-like spermatocytes. We performed immunofluorescence analysis of the spindle marker  $\alpha$ -TUBULIN and pHH3 (phosphorylated histone H3 at Ser10), which identifies condensed M-phase chromosomes on the chromatin. Spindle disorganization and chromosome misalignment were frequently detected in metaphase I stage cells depleted of *Nkapl* (Supplementary Fig. 3a,b). The increase of metaphase cells with chromosome misalignment was further confirmed by immunofluorescence analyzes with antibodies for  $\alpha$ -TUBULIN and the centromere protein CREST (Supplementary Fig. 3c,d). To further determine whether the defect at the metaphase stage is caused by the absence of NKAPL, we isolated pachytene spermatocytes from adult wild-type and *Nkapl* null mice. In vitro cultured pachytene spermatocytes progress through meiotic prophase into the metaphase I stage upon treatment with the phosphatase inhibitor okadaic acid. We found that treatment of *Nkapl*-depleted spermatocytes with okadaic acid induced the progression into the metaphase I, as assessed by nuclear spread analysis with synaptonemal complex components SYCP1 and SYCP3 (Supplementary Fig. 3e). The percentage of diakinesis/metaphase I spermatocytes after OA treatment was high in the *Nkapl* knockout to a similar degree as the wild-type (Supplementary Fig. 3f). Despite their normal entry into metaphase I, immunostaining of spread nuclei of spermatocytes showed that *Nkapl*-depleted metaphase I cells often contained abnormal number of CREST foci, in contrast to the majority of wild-type metaphase I cells had 40 CREST foci (Supplementary Fig. 3g). The frequency of metaphase I cells with either less than 40 CREST foci or more than 40 CREST foci was elevated in the *Nkapl* knockout compared to the wild-type (Supplementary Fig. 3h), indicating that improper segregation of homologous chromosomes or premature segregation of sister chromatids during metaphase I. Similarly, the frequency of metaphase I cells with less than 40 CREST foci was increased in juvenile *Nkapl* knockouts at P17 (Supplementary Fig. 3g,h), although germ cells of *Nkapl* KO

progressed to pachytene spermatocytes without apparent defects (Supplementary Fig. 3i,j). NKAPL is a retrotransposed homolog of NKAP, which regulates chromosome alignment and mitotic progression through anchoring CENP-E to kinetochores<sup>30</sup>. These results suggest NKAPL has a similar role to its paralog NKAP in regulating chromosome alignment and segregation.

We noted that the majority of *Nkapl*-deleted spermatocytes could progress further to round spermatids in thus far histological studies, despite the initial defects at the metaphase stage with the elimination of aberrant late stages of spermatocytes. To determine the precise steps of round spermatid arrest, we immunostained the testes tubules with the acrosome marker PNA (peanut agglutinin) and a meiotic recombination protein  $\gamma$ H2AX. Post-meiotic spermatid development encompasses 16 steps (steps 1–16), and spermatids at various steps, including round spermatids at steps 7–8, were observed in wild-type testis tubules (Fig. 2e). However, most round spermatids in *Nkapl* KO contained 1–2 dotted PNA-positive proacrosomic vesicles, suggesting that they were arrested at an early step 2–3 (Fig. 2f). We noticed some spermatids developed beyond step 2–3, with an occasional step 7–8 round spermatids being visible (Fig. 2g). However, a large population of these round spermatids displayed a fragmentation of the chromocenter and defective acrosome formation (Fig. 2h). This represents an arrest at the early round spermatid stage in the testis, which occurred later than the defects during the late stages of meiosis. These observations reveal that NKAPL is required for meiotic exit, the development of round spermatids and the transition from round to elongating spermatids.

**NKAPL knockout downregulates haploid genes, similar to SOX30 or HDAC3 depletion**

To explore the molecular targets of NKAPL in germ cell development, we performed transcriptome analysis to address whether the phenotype in *Nkapl* KO testes accompanied alterations in gene expression. Because the majority of *Nkapl* KO tubules were arrested at the round spermatid stage, we prepared RNA-seq libraries from triplicate wild-type and *Nkapl* KO testes at P21, when round spermatids first appeared, and the cellular composition of testis remained similar. In contrast to the results in NKAP-depleted somatic cells<sup>40</sup>, NKAPL knockout resulted in downregulation of most genes, with significantly 152 and 2,216 genes up- and downregulated ( $p < 0.05$ , fold change  $\geq 1.5$ ), respectively, in *Nkapl* KO compared to wild-type (Fig. 2i). Gene ontology analysis indicated that haploid spermatid development-associated genes including those involved in the processes of acrosomal formation, haploid differentiation, sperm chromatin condensation and flagellated sperm motility were downregulated in *Nkapl* KO testes (Supplementary Fig. 4a). The downregulation of representative haploid genes was further confirmed by RT-qPCR (Supplementary Fig. 4b). To further confirm the above gene expression changes, we isolated pachytene spermatocytes (PS) and round spermatids (RS) from wild-type and *Nkapl* KO mice to analyze stage-specific transcriptomic alteration (Supplementary Fig. 4c). Consistent with the transcriptomic alteration in *Nkapl* KO testes, RNA-seq analyzes of PS and RS population from wild-type and *Nkapl* KO revealed



that NKAPL depletion causes a pronounced downregulation of transcripts, and the gene expression profiles were highly similar between *Nkapl* KO testes and *Nkapl* KO cells (Supplementary Fig. 4d,e). The downregulation of these haploid development-associated genes in *Nkapl*-depleted PS and RS was further confirmed by RT-qPCR (Supplementary Fig. 4d,e). These results indicate that most genes in *Nkapl* knockout exhibit downregulation but not upregulation.

Such downregulation of haploid spermatid genes was similar to the transcriptomic changes due to lacking either SOX30 or HDAC3 in germ cells. Like *Nkapl* knockout, testis-specific *Hdac3* or global *Sox30* knockout results in meiotic exit defects<sup>34</sup>, with germ cell arrest either at the late stages of spermatocytes or at the round spermatid stage. Therefore, we analyzed RNA-seq results from *Nkapl* KO testis and compared them with that from the germ cell-specific *Hdac3* KO (*Stra8*-

**Fig. 2 | *Nkapl* KO leads to defects in the meiotic-to-postmeiotic transition and predominant gene downregulation.** **a, b** MLH1 foci and the quantification of MLH1 foci in the wild-type and *Nkapl* KO spermatocytes at the pachytene stage. The number of MLH1 foci was comparable between wild type and *Nkapl* KO. Pachytene spermatocytes examined: wild-type,  $n = 54$ ; *Nkapl* KO,  $n = 94$ . Data are presented as mean  $\pm$  SD, two-tailed Student's  $t$  test. Scale bars, 10  $\mu$ m. **c** TUNEL assay of testicular sections from wild-type and *Nkapl* KO at 8-week-old. **d** Quantification of TUNEL-positive cells in adult wild-type and *Nkapl* KO. Many TUNEL-positive cells were observed in stage XII tubules. Tubules examined: wild-type,  $n = 296$ ; *Nkapl* KO,  $n = 401$ . Data are presented as mean  $\pm$  SD. \*\*\*\* $P < 0.0001$ , two-tailed unpaired Student's  $t$  test. Experiments were performed with biological triplicates. **e** Frozen sections from adult wild-type testes were immunolabelled with PNA (an acrosome marker, red) and  $\gamma$ H2AX (green). The differentiation of round spermatids into mature sperm encompasses 16 steps, which are indicated by PNA. All the

developmental steps were observed in the wild-type, and round spermatids at step 2-3, step 7-8 and step 11-12 were selectively shown. Biological duplicates were prepared. **f** Round spermatids of *Nkapl* KO mice were mainly arrested at step 2-3. Experiments were performed with biological duplicates. **g** Some spermatids in *Nkapl* KO developed beyond step 2-3 from biological duplicates. Scale bars, 20  $\mu$ m. **h** Some spermatids in *Nkapl* KO displayed fragmented chromocenter and defective acrosome formation. Areas within the rectangles were enlarged in the right panel. **i** Scatter plot of RNA-seq data showing differentially expressed genes (fold change  $\geq 1.5$  up (red) or down (blue) and  $p < 0.05$ ) in P21 wild-type and *Nkapl* KO testes using DESeq2, which employs a negative binomial distribution model along with the Wald test method. Experiments were performed in biological triplicates ( $n = 3$  for each genotype). **j** Downregulated genes in *Sox30* depleted testes (fold change  $\geq 1.5$ ,  $p < 0.01$ ) were selected DESeq2 and used to produce a heatmap depicting their expression profile in *Stras8-cre/Hdac3* KO and *Nkapl* KO testes.

*cre/Hdac3* KO) and *Sox30* KO mice. The majority of downregulated genes in *Sox30* KO also displayed similar downregulation upon depletion of either NKAPL or HDAC3 (Fig. 2j). The overlapped downregulated genes among *Nkapl* KO, *Stras8-cre/Hdac3* KO, and *Sox30* KO showed a strong enrichment for spermatid development-associated processes. Generally, *Nkapl* knockout in testis leads to more downregulation of genes than *Stras8-cre/Hdac3* KO or *Sox30* KO (Supplementary Fig. 4f).

### NKAPL co-localizes with SOX30/HDAC3 and regulates SOX30/HDAC3-DNA interaction dynamics

The transcriptomic analysis suggests that SOX30, NKAPL, and HDAC3 may act in the same signaling pathway. Previous studies show that NKAP acts as a transcriptional repressor in T cells, and it associates with DNA, although it's likely that NKAP binds DNA indirectly as it lacks any previously characterized DNA-binding domain<sup>23,24</sup>. To determine whether NKAPL indeed binds to genomic DNA in the male germline, we mapped its genome-wide binding by performing NKAPL ChIP-seq in biological duplicates from testes at P21 when NKAPL is abundantly expressed and when the late-stage spermatocytes and early RS are highly represented in testes. NKAPL bound to 1056 sites, and 65% of NKAPL resided in promoters and 5' UTR of annotated genes (Fig. 3a, b). HDAC3 and SOX30 binding strongly co-localized on a genome-wide scale in mouse testes<sup>34</sup>, and the location and intensity of NKAPL binding highly correlate with SOX30 and HDAC3 (Fig. 3c). Analysis of best-scored NKAPL ChIP-seq peaks revealed the Sox consensus sequence A(A/C)AATGGCGGCC as the top enriched motif (Fig. 3d). This consensus DNA sequence is also high for GC composition (Fig. 3d). We noticed that about two-thirds of the SOX30 binding sites were occupied by NKAPL. NKAPL binding motif shared a similarity with the identified SOX30 motif, but their sequences are not the same. Consistent with the notion that NKAPL lacks a well-defined DNA-binding domain, these results suggest the indirect and transient chromatin-binding of NKAPL likely occurs adjacent to SOX30 sites.

Looking more closely at the distribution of SOX30 and NKAPL binding sites, we observed that NKAPL binds to a region located 3' downstream of SOX30 sites and exhibited maximum occupancy downstream of transcription start sites (TSSs) (Fig. 3e). What mediates the cotranscriptional recruitment of NKAPL? Both NKAP and NKAPL associate with HDAC3 when co-transfected into HEK293T cells<sup>23,33</sup>. Our immunoprecipitation assay in HEK293T cells co-transfected with FLAG-tagged HDAC3 and HA-tagged NKAPL showed a weak association between exogenous expressed HDAC3 and NKAPL with long exposure time (Fig. 3f). HA-tagged truncation constructs of NKAPL with the individual deletion of its RS, the basic domain, or DUF926 domain as well as NKAPL C-terminal DUF926 were further over-expressed with HDAC3 (Fig. 3f). Interestingly, the NKAPL-HDAC3 interaction was enhanced upon the deletion of the NKAPL basic domain, and the NKAPL C-terminal DUF926 alone was able to associate

with HDAC3 (Fig. 3f). Nevertheless, the association of NKAPL with HDAC3 is not captured in vivo, as endogenous NKAPL does not pull down HDAC3 by immunoprecipitation of NKAPL in mouse testes (Fig. 3g). The NKAPL basic domain is highly repetitive with a potential to form nuclear bodies, and the enhanced NKAPL-HDAC3 association upon the deletion of the basic domain likely reflects the NKAPL-HDAC3 association is dynamic, which might be hard to detect due to their transient nature. Given the fact that NKAPL lacks a well-defined DNA-binding domain, the association of NKAPL with DNA likely reflects its residence within the chromatin proximity of the transcription initiation complex rather than in direct contact with DNA.

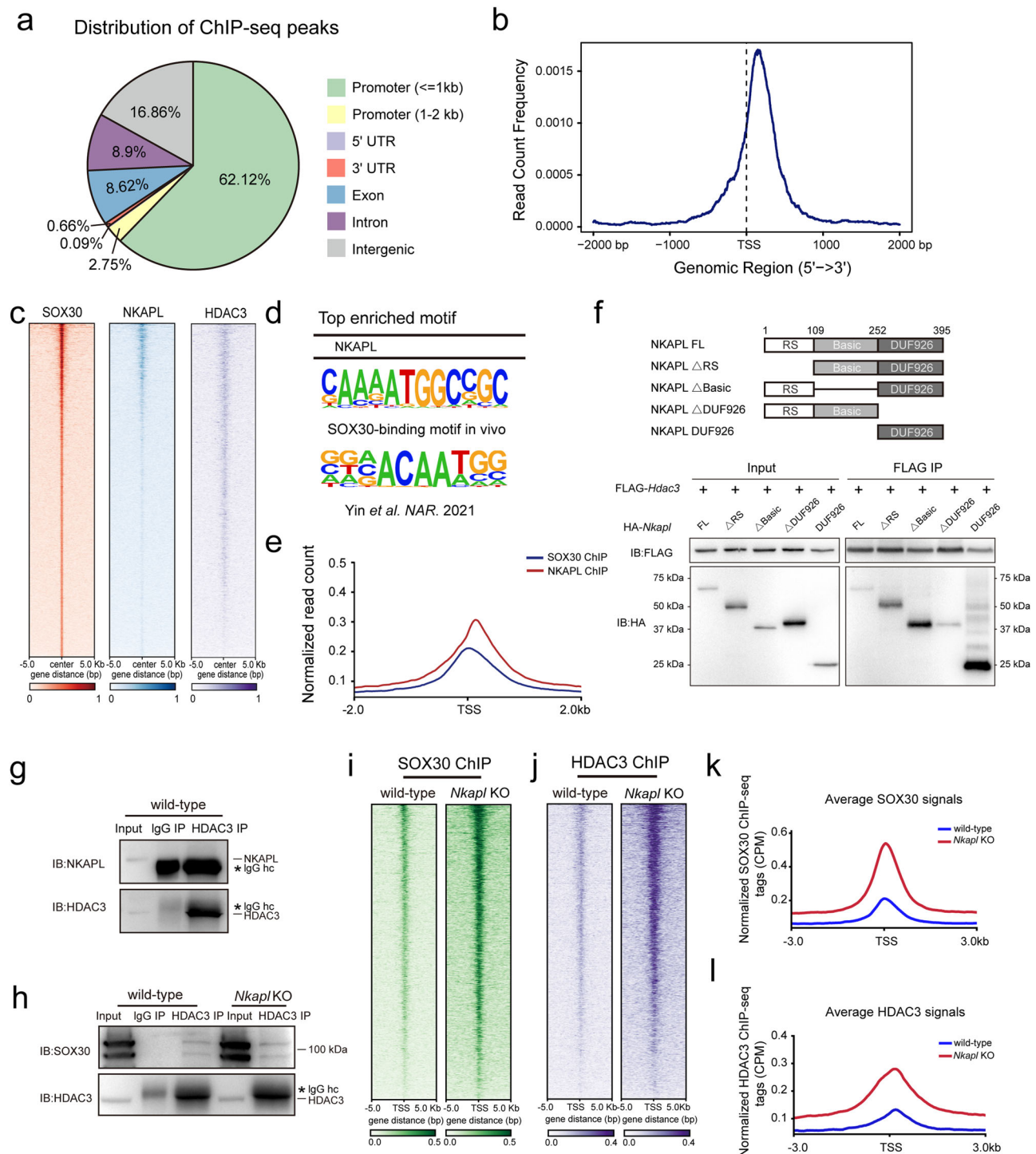
How does the loss of NKAPL produce a similar gene expression alteration with *Sox30* knockouts? To explore the role of NKAPL in the SOX30 complex, we address how NKAPL knockout affects the interaction between endogenous SOX30 and HDAC3 in mouse testes. Immunoprecipitation analysis of wild-type and *Nkapl* KO testes at P21 indicates that this physical interaction between SOX30 and HDAC3 remains intact without NKAPL (Fig. 3h). Next, we examined the genomic binding of the SOX30/HDAC3 protein complex in the absence of NKAPL. SOX30 ChIP-seq and HDAC3 ChIP-seq were performed in wild-type and *Nkapl* KO testes at P21. SOX30 binding at target binding loci in testes was significantly higher in testes from mice lacking *Nkapl* than in wild-type mice (Fig. 3i, k). Similarly, the DNA binding affinity of HDAC3 at its genomic target sites was higher in *Nkapl* null testes than in control (Fig. 3j, l). These results demonstrated that genetic ablation of *Nkapl* strengthens the association between HDAC3/SOX30 and DNA, associated with stalled transcription of haploid spermatid genes at the meiotic exit.

### NKAPL facilitates Pol II pause-release at promoters

Transcription is a highly controlled process that consists of five stages: the pre-initiation complex formation, initiation, pause/release, elongation, and termination. Loss of NKAPL led to accumulated SOX30 and HDAC3 ChIP-seq signals at many sites when compared to the wild-type, implying that the genomic binding of SOX30 and HDAC3 at target sites in testis is stronger. Gene transcription occurs as a discontinuous process described as transcriptional bursting. Accurate and efficient transcription requires the intermittent and dynamic recruitment and release of transcription factors and Pol II at promoters. NKAPL may coordinate high-level transcription by facilitating the assembly of a dynamic transcription initiation complex, since *Nkapl* depletion enhanced the binding between SOX30/HDAC3 complex and DNA at TSSs. NKAP protein, the paralog of NKAPL, forms extensive nuclear bodies in HeLa cells<sup>24</sup>. NKAPL also has large intrinsically disordered regions (IDRs) at its N-terminus before a structured DUF926 domain (Fig. 4a), suggesting that NKAPL may have a potential to form networks of weak protein-protein interactions.

Upon the formation of the pre-initiation complex by general transcription factors and subsequent recruitment of Pol II with an





unmodified CTD, initiation commences with the opening of double-stranded DNA. If transcription pre-initiation at sites of SOX30 occupancy is halted, the subsequent recruitment of RNA Pol II might be compromised. We then performed ChIP-seq experiments for total RNA Pol II in wild-type and *Nkapl* KO testes at P21. Contrary to our prediction, *Nkapl* knockout in mouse testes resulted in an increase of RNA Pol II occupancy at promoters of SOX30 binding sites (Fig. 4b). *Nkapl* knockout also resulted in an increased level of RNA Pol II enrichment at promoters of downregulated genes upon *Nkapl* knockout (Supplementary Fig. 5a). Next, we conducted a combined analysis of NKAPL ChIP-seq and RNA-seq datasets to examine the correlation between NKAPL binding and gene expression alterations. Integrating the DNA-binding sites of NKAPL and differentially expressed genes upon NKAPL depletion revealed a

significant enrichment of NKAPL binding around transcription start sites of downregulated genes (Supplementary Fig. 5b).

To confirm *Nkapl* knockout causes increased Pol II pausing, we calculated the traveling ratio (TR), which represents the relative ratio of Pol II read density in the promoter-proximal region and the gene body. Loss of NKAPL leads to increased traveling ratios of total RNA Pol II (Fig. 4c), indicating that Pol II is paused on the chromatin. An increase in the traveling ratio of RNA Pol II was similarly observed at the PS stage (Fig. 4d). The traveling ratio of RNA Pol II in NKAPL-depleted RS was also increased compared to wild-type cells (Fig. 4e). Moreover, the NKAPL binding signals generally overlap with the promoter-proximal Pol II peaks (Fig. 4f). Thus, NKAPL promotes transcription elongation by facilitating the release of paused Pol II at promoters.

**Fig. 3 | NKAPL co-localizes with HDAC3/SOX30 on the genome and regulates the HDAC3/SOX30-DNA interaction.** **a** Annotations of NKAPL ChIP-seq peaks across different genomic regions in P21 wild-type testes. ChIP experiments were performed in biologically duplicates. **b** Calculated NKAPL ChIP-seq tags' densities on UCSC mm10 RefSeq gene bodies ( $n = 20,460$ ), and the genomic regions from -2 kb to +2 kb surrounding the TSS of genes were shown. **c** Heat map of NKAPL, SOX30 and HDAC3 ChIP-seq signals in P21 wild type testes depicting their co-localization at many binding sites. Regions from -5 kb to +5 kb surrounding the center of SOX30 sites were plotted in heatmap views. **d** The top-ranked motif in the binding sites of NKAPL with  $p$  values using Homer motif search. Sequences within  $\pm 200$  bp from the centers of all the binding sites were used for de novo motif analysis.  $p$  values =  $1E^{-16}$ . The top enriched motif at SOX30 binding sites in mouse testis was also shown. Hypergeometric Distribution Test,  $p$  values =  $1E^{-132}$ . **e** Position and distance between SOX30 and NKAPL average ChIP-seq signals. Average ChIP-seq signals of SOX30 and NKAPL, represented by counts per million, from -2 kb to +2 kb

surrounding the TSS of genes were shown. **f** FLAG tagged HDAC3 was co-expressed with full-length HA-NKAPL or HA tagged NKAPL mutants with deleted fragments. Immunoprecipitation assay was performed with anti-FLAG antibody before western blot analysis with HA and FLAG antibodies. Representative images from biological duplicates were shown. **g** Testes protein lysates were immunoprecipitated either with HDAC3 or normal IgG antibodies followed by immunoblot analysis. **h** Co-immunoprecipitation of HDAC3 from wild-type and *Nkapl* KO testes at P21 and western blot with SOX30 antibodies. Protein lysates were prepared at P21.  $n = 3$  for each genotype. Experiments were performed with biological triplicates. **i, j** Heat map of SOX30 (**i**) and HDAC3 (**j**) signals in *Nkapl* null testes at P21 from -5 kb to +5 kb surrounding the center of SOX30 sites. **k** Average SOX30 ChIP-seq signals across  $\pm 3$  kb flanking TSS in *Nkapl* knockout versus their wild-types. CPM, counts per million. **l** Average HDAC3 ChIP-seq signals across  $\pm 3$  kb flanking TSS in *Nkapl* knockout versus their wild-types.

### NKAPL binds RNAs containing a tandemly GAA repeat at promoter regions

Transcription elongation is coupled to pre-mRNA processing. Since NKAPL contains an arginine/serine-rich domain (RS domain) with RNA-binding capacity, we sought to profile the RNA targets of NKAPL on chromatin using eCLIP-seq (enhanced crosslinking immunoprecipitation coupled with sequencing) (Fig. 5a). Biological replicate eCLIP-seq libraries of NKAPL in wild-type testes at P21 were prepared, and the two replicates were highly correlated via gene RPKMs (Supplementary Fig. 5c). eCLIP-seq analysis yielded 26,248 and 25,392 NKAPL peaks (with filters for FDR < 0.05, fold change  $\geq 2$ ) in two replicates, respectively, with 5379 protein-coding genes were bound in both replicates (Supplementary Fig. 5d). Over 99.5% of total NKAPL eCLIP peaks were located within protein-coding transcripts (Fig. 5b). Notably, we found a very large proportion (46%) of NKAPL eCLIP peaks was located within promoter regions ( $\pm 2$  kb from the TSS) (Fig. 5c), representing 52% of the NKAPL RNA targets. We next analyzed reads density across the mRNA transcript length. Metagene and boxplot analysis revealed a significant enrichment of NKAPL eCLIP reads over the 5' regions of genes (Fig. 5d), with the highest signals at the promoter region around 40 nt downstream of the TSS (Fig. 5e), which are known as hotspots for Pol II pause release. NKAPL also binds strongly to mRNA transcripts that were downregulated upon NKAPL depletion (Supplementary Fig. 5e).

Within all the identified NKAPL eCLIP peaks, motif analysis identified a simple sequence repeat, a sequence consisting largely of a tandem repeat of "GAA"n (Fig. 5f). The same top motif was repeatedly identified within the top 50% of NKAPL eCLIP peaks, as well as NKAPL binding sites located in promoter regions (Fig. 5f). The G-rich RNAs have the potential to associate with the template DNA strand to promote R-loop formation. Interestingly, this tandem sequence "GAA"n repeat is the top motif enriched in R-loop peaks in the plant *Arabidopsis* genome<sup>12</sup>, as shown in (Fig. 5g). The second top motif is its complementary sequence featuring CT repeats. In the human genome, promoter R-loop signal is positively associated with higher G content and G/C skew<sup>13</sup>. G-rich sequences also underlie mapped R-loops detected by R-ChIP based on the use of a catalytically dead RNase H1<sup>16</sup>. These results demonstrated that NKAPL binds strongly to the promoter-proximal regions, particularly at specific loci containing "GAA" repeated sequences.

### NKAPL co-localizes with R-loops at GAA-rich loci and promotes R-loop formation

Enriched in promoters, R-loops are highly dynamic and play active roles in genome organization and gene expression regulation<sup>13,16,41</sup>. To test whether NKAPL binds to R-loop hotspots and contributes to R-loop formation, we study the correlation between R-loops location and NKAPL binding peaks. To profile genome-wide R-loops, we employed ssDRIP-seq (single-strand DNA ligation-based library preparation after DNA:RNA hybrid immunoprecipitation by S9.6 and sequencing) in testes at P21 (Supplementary Fig. 6a)<sup>12</sup>. Similar to the

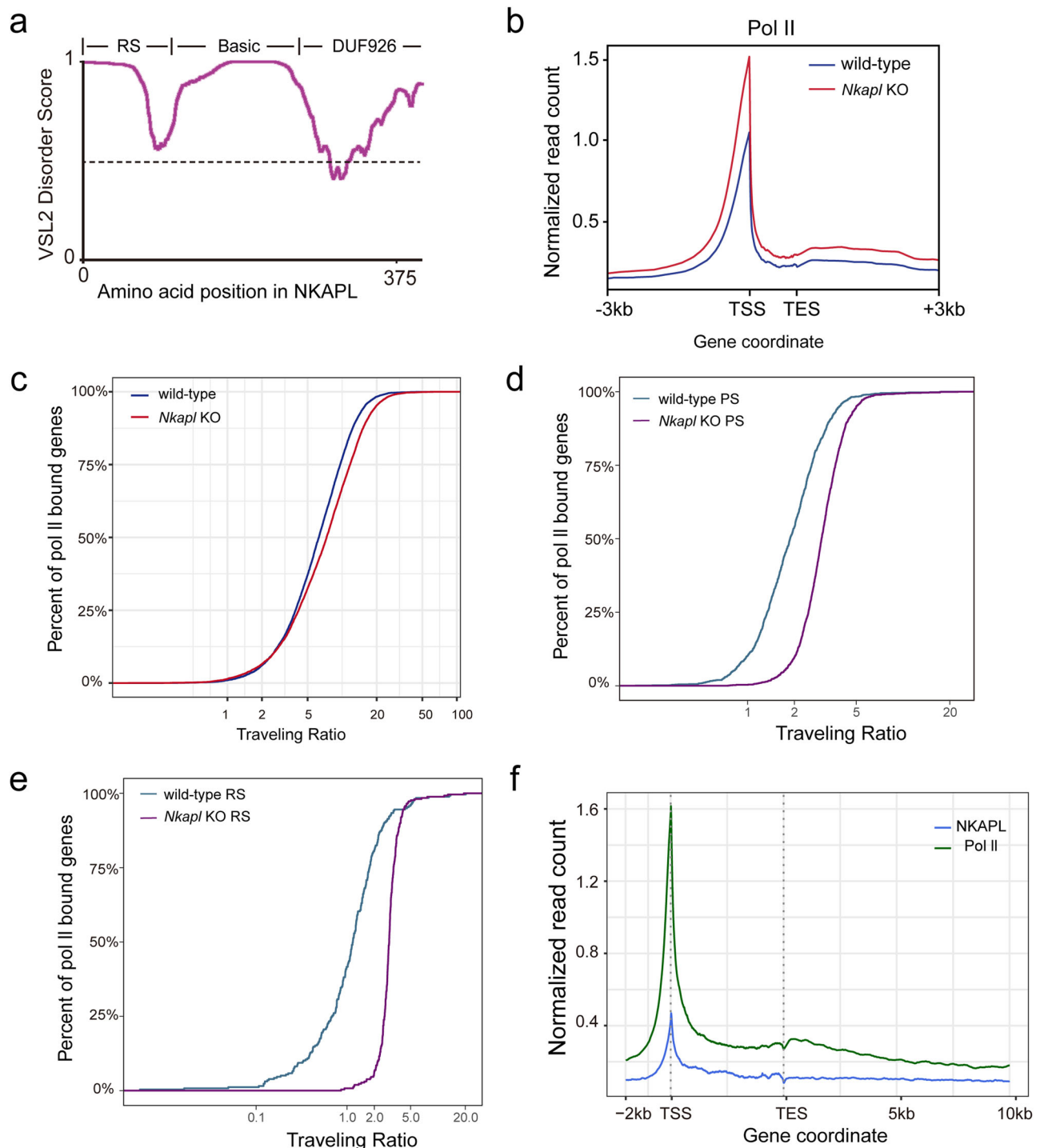
DRIP-seq (DNA:RNA immunoprecipitation followed by cDNA conversion coupled to high-throughput sequencing)<sup>13</sup>, ssDRIP-seq was also strand-specific, and biological ssDRIP-seq replicates were in strong agreement with each other (Supplementary Fig. 6b). Importantly, treatment of genomic DNA with RNase H1 abolished the R-loop signal, validating the assay (Fig. 6a). Our strand-specific analysis identified 52,764 peaks for R-loops in P21 wild-type testes. The majority of R-loop peaks in mouse testis range from 100 bp to 500 bp (Supplementary Fig. 6c), which is in close agreement with the size of R-loops in human and mouse cells<sup>13,16</sup>.

We next analyzed the localization of R-loops across the mouse testis genome. Most ssDRIP-seq signals were significantly enriched in gene promoter regions (Fig. 6b), in agreement with previous studies<sup>13,16</sup>. Promoter R-loop signals increased around the transcription start site and spanned the promoter-proximal regions (Fig. 6c). The top enriched motif within the best-scored R-loop peaks in mouse testis is a simple sequence repeat of "GAA"n (Fig. 6d), in line with a previous study in plants<sup>12</sup>. To test whether the mammalian genome shares the conserved sequence to favor R-loop formation, we searched for motifs using MEME based on previously published R-loop datasets in human cells<sup>13</sup>. Motif analysis revealed the presence of a tandem GAA repeat, which is identified as the top enriched motif in human R-loops (Fig. 6d). The second top motif is its complementary sequence featuring CTT repeats (Fig. 6d). The same tandemly GAA repeat and its complementary CTT sequence were enriched as the top within R-loop peaks captured by the R-ChIP approach in HEK293T cells (Fig. 6d). These data suggest that R-loop formation preferentially occurs at conserved genic hotspots in higher eukaryotic genomes.

Notably, this R-loop promoting GAA repeat sequence was preferentially recognized by NKAPL in mouse testis (Fig. 5f). We then compared NKAPL binding sites by eCLIP-seq with germline R-loop profiles in the mouse genome. Analysis of NKAPL eCLIP-seq identified that NKAPL binds mRNA transcripts corresponding to 5379 genes, of which 4317 contained R-loops (Supplementary Fig. 6d). A significant enrichment of R-loop signals was observed around NKAPL binding peaks (Fig. 6e). Mapping the ssDRIP-seq reads to NKAPL binding sites, we observed a maximal occupancy of germline R-loop signals at the center of NKAPL binding peaks (Fig. 6f). We note that R-loop reads at NKAPL target genes were significantly higher than those genes without NKAPL binding peaks (Fig. 6g), with a pronounced difference around the transcription start site, suggesting that NKAPL binding positively promotes promoter R-loop formation.

To further evaluate a potential role for NKAPL in R-loop formation, we investigated the R-loop profiles in *Nkapl* knockouts. Similar to the average R-loop peaks in wild-type samples, most of the R-loops in *Nkapl* knockouts were produced at 100 to 500 base pairs long. However, in *Nkapl* knockouts, R-loop signals were clearly below the wild-type (Fig. 6h). The steady-state levels of R-loops could be the result of an equilibrium between active DNA: RNA hybrid formation during



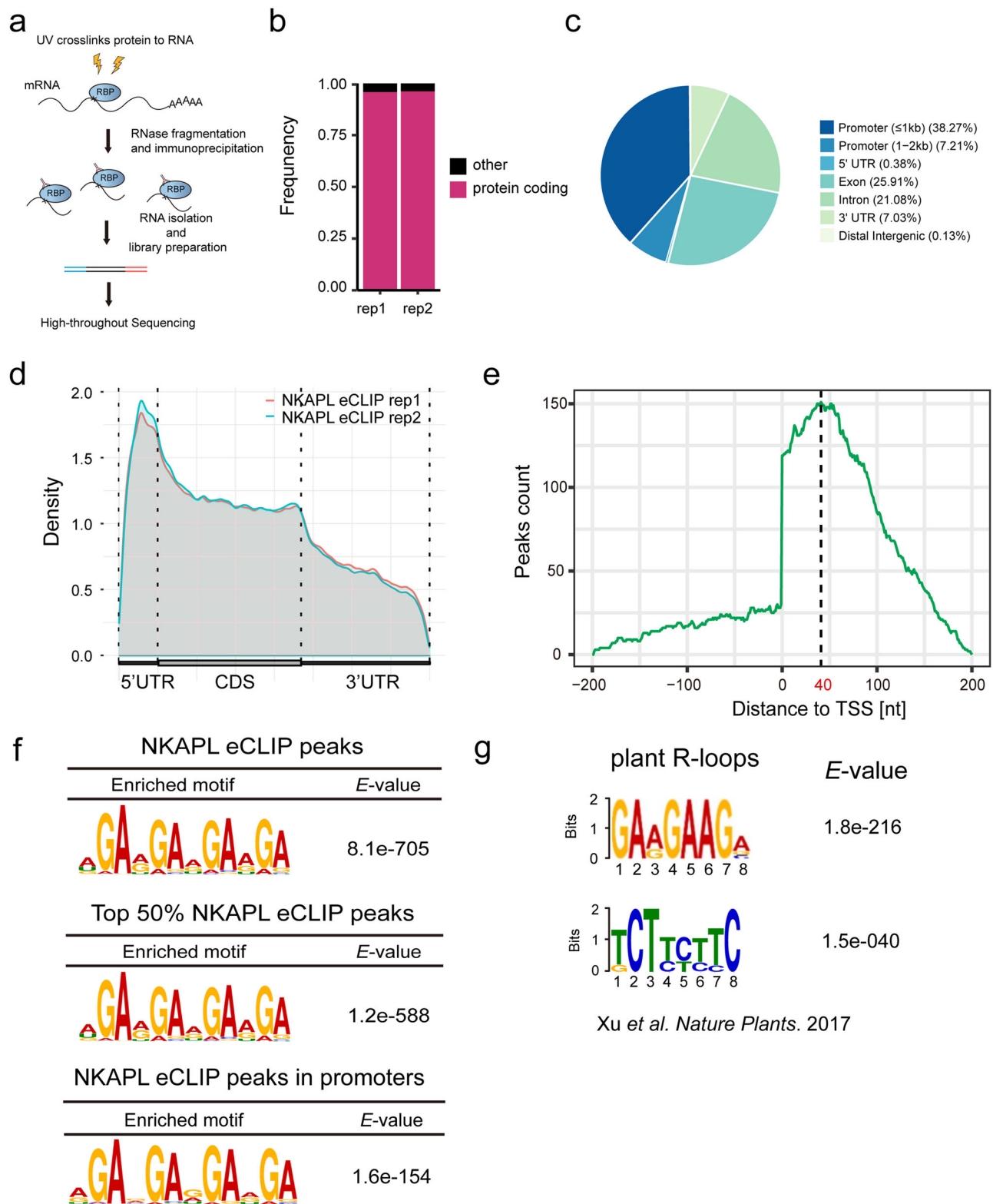


**Fig. 4 | NKAPL facilitates the release of paused Pol II at promoters.** **a** Predictions of intrinsic disorder of NKAPL as calculated by the VSL2 algorithm (<http://www.pondr.com>). **b** Pol II ChIP experiments with anti-total Pol II antibodies were performed in wild-type and *Nkapl* KO testes at P21. The average occupancy of total Pol II I along the length of genes occupied by SOX30. **c** Comparison of Pol II traveling ratios (the ratio between gene body and promoter-proximal polymerase) between wild-type and *Nkapl* KO testes. Pol II traveling ratio is defined as Pol II read density ratio between the promoter-proximal region (-80 bp to +250 bp around the transcription start site) and gene body (250 bp downstream

of the TSS to the transcription end site). A higher traveling ratio value indicates a higher degree of pausing. The coverage of each region was calculated by featureCounts with the parameter '--fracOverlap 0.5'. The statistics by Mann-Whitney test was performed for the comparison of Pol II traveling ratios between wild-type and *Nkapl* KO testes. The statistics and plots were completed by custom scripts using R language. **d–e** Calculation of Pol II traveling ratios (the ratio between gene body and promoter-proximal polymerase) in PS (**d**) and RS populations (**e**) from wild-type and *Nkapl* KO. **f** Comparison of the binding signals of NKAPL and Pol II across genes.

transcription and removal by nucleases such as RNase H<sup>42,43</sup>. As expected, RNase H treatment is able to efficiently remove R-loop levels in mouse testis (Fig. 6i). In the absence of NKAPL, the average R-loop signals were drastically reduced at promoter regions (Fig. 6i). To

further determine NKAPL protein co-localizes with R-loops at GAA-rich loci, recombinant NKAPL-mGFP fusion protein was purified and incubated it with a Cy3-labeled R-loop substrate containing DNA-RNA hybrid and displaced ssDNA. Fluorescent microscopy revealed that



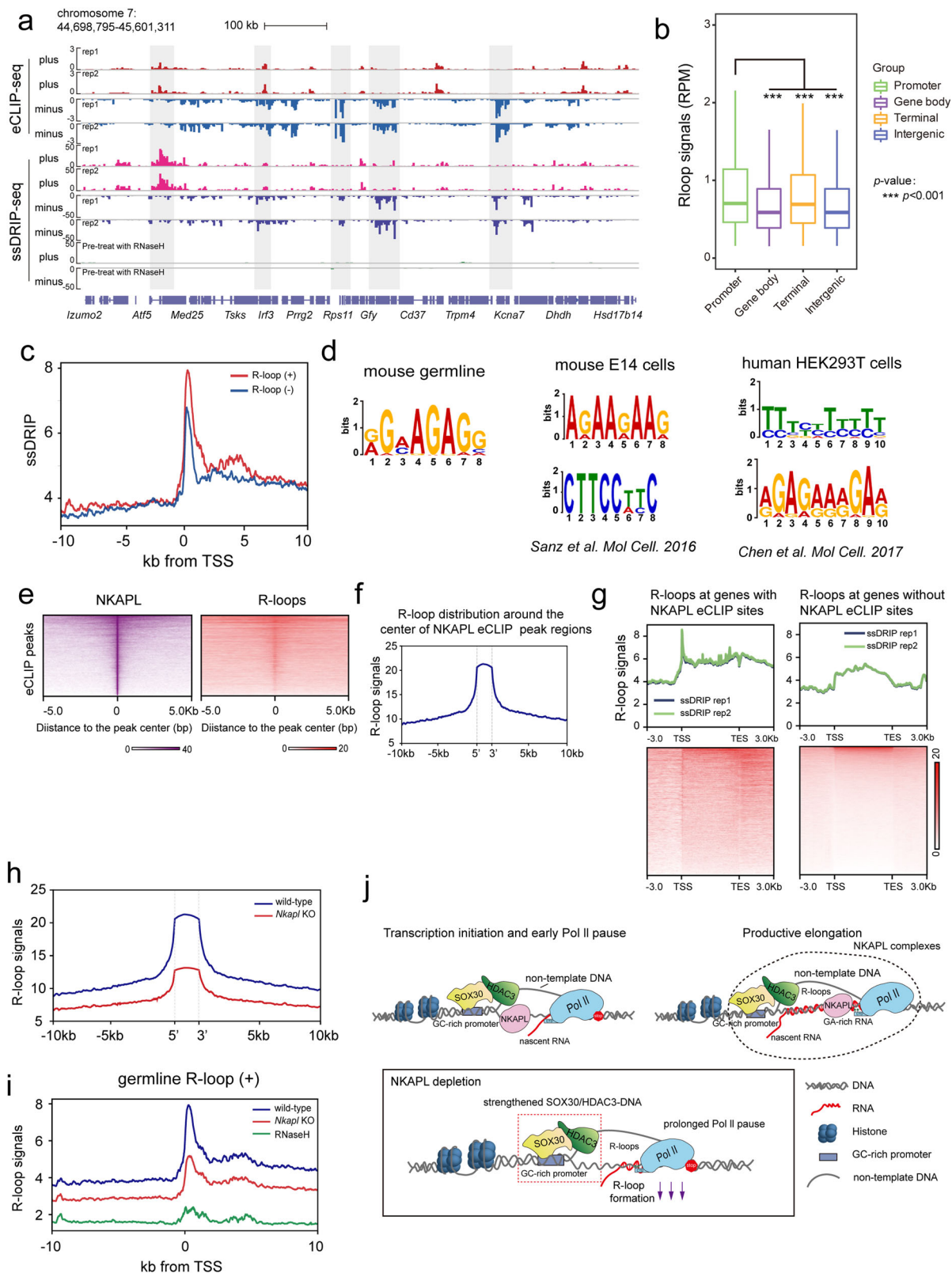
**Fig. 5 | NKAPL binds to promoter-associated RNAs. a** Schematic diagram of eCLIP-seq method. **b** RNA targets with NKAPL eCLIP peaks. **c** Genomic distribution of NKAPL-binding sites identified by eCLIP-seq. **d** Distribution of replicate NKAPL eCLIP peaks along the length of mRNA transcripts. **e** Averaged NKAPL eCLIP-seq

signals across mRNA. Signals of  $\pm 200$  bp flanking the TSS were shown. **f** NKAPL-binding motifs identified by MEME from all the NKAPL binding peaks, the top 50% of NKAPL eCLIP peaks, and NKAPL eCLIP peaks within promoter regions. **g** This GA-rich motif was enriched in R-loop peaks in the plant *Arabidopsis* in a previous study.

NKAPL-GFP incorporated this Cy3-labeled R-loop substrate in vitro (Supplementary Fig. 7). Thus, R-loops preferentially form at specific GAA-rich genomic loci, and the occupancy of NKAPL on the nearby genomic regions contributes to the R-loops formation or stability.

#### NKAPL DUF926 domain mutations are associated with azoospermia in humans

To explore the role of NKAPL dysfunctions in human spermatogenesis, mutations in the *NKAPL* gene were screened on genomic DNA



samples from 620 patients diagnosed with non-obstructive azoospermia (NOA). The sperm analysis of the semen was conducted in biological duplicates for each individual enrolled in the infertile group. The most common genetic causes of male infertility, including chromosomal abnormalities and Y chromosome microdeletions, were removed. We excluded patients with cryptorchidism, vas deferens obstruction, testicular inflammation or endocrine disorders

from our study. Sanger sequencing of the *NKAPL* coding region was conducted on these 620 patients with NOA. The genetic variants were screened based on the criteria that their minor allele frequency (MAF) was less than 0.0005 in the gnomAD database, and their potential functional consequences were predicted using the Combined Annotation Dependent Depletion (CADD) tool with a score greater than 15 (Fig. 7a, b).



**Fig. 6 | NKAPL co-localizes with R-loops at GA-rich sites and promotes R-loop formation.** **a** Representative genomic regions showing ssDRIP-seq signals in P21 wild-type testes. **b** Distribution of ssDRIP-seq signals for biological duplicates in different genomic regions, promoter,  $\pm 2$  kb of TSS; terminal regions,  $\pm 2$  kb of TTS. Bounds of box represent the interquartile range (IQR) from the 25th to the 75th percentile with the median as the horizontal line, while whiskers extend to the most extreme data points (outliers excluded) within  $1.5 \times \text{IQR}$ . \*\*\* $p < 0.001$ . **c** ssDRIP-seq signals within  $\pm 10$  kb of TSS. Red: forward strand signal; Blue: reverse strand signal. **d** The top enriched motif identified by MEME from germline R-loops in mouse testes as well as recently published R-loops in human and mouse cells. **e** NKAPL eCLIP binding signals at targeted loci generally co-localized with corresponding R-loop signals. **f** R-loop signal profiles within  $-10$  kb/  $+10$  kb of NKAPL eCLIP binding sites. **g** R-loop signals by ssDRIP-seq at NKAPL target genes versus those genes without NKAPL eCLIP peaks. **h** Relative R-loop signals at NKAPL eCLIP binding sites in wild-type and *Nkapl* KO testes at P21. **i** Comparison of R-loop signal occupancy in wild-

type and *Nkapl* KO on genes across  $\pm 10$  kb genomic region flanking TSS. **j** Model for NKAPL function in the release of paused Pol II and R-loop formation. Transcription initiation complex containing SOX30/HDAC3 assembles at promoters and recruits Pol II. NKAPL interacts with the SOX30/HDAC3 initiation complex dynamically and transiently, and occupies at promoter regions indirectly. Initiation occurs with the opening of double-stranded DNA, and a short nascent RNA is synthesized before the transcription machinery pauses promoter proximally. Promoter-proximal stalled Pol II produces more levels of short RNAs, which re-anneal with its template DNA strand to form three-stranded RNA/DNA hybrid structures (R-loops). NKAPL binds promoter-associated nascent RNAs, and co-localizes with R-loops at GA-rich loci, where it interacts with RNA-DNA hybrid structures to enable R-loop formation and efficient Pol II elongation. *Nkapl* knockout causes prolonged Pol II pause and a pronounced reduction of R-loops, resulting in a stalled initiation complex containing SOX30/HDAC3.

NKAPL is composed of an N-terminal RS domain, the basic domain, and a C-terminal DUF926 domain of unknown functions (Fig. 7b). Four unique (MAF = 0 in gnomAD populations) heterozygous variants were identified in four patients with NOA but absent in 2713 fertile controls (Fig. 7a, b). Among these patients, three carried missense mutations (c.844 G > A; c.896 C > G; c.1040 G > A), while one carried a frameshift mutation (c.1046\_1047TGdel, 2 bp; p.Met349fs). The deletion of TG nucleotides would be expected to cause a frameshift and thus premature termination of NKAPL, resulting in a truncated NKAPL protein comprising only the DUF926 domain half. Notably, all these mutations were located in the DNA region that encodes the C-terminal DUF926 domain of NKAPL (Fig. 7b, c).

#### *Nkapl*<sup>M349fs</sup> mice exhibit defects in meiotic exit, resembling *Nkapl* KO

Among the NKAPL mutations with significantly higher prevalence in azoospermic men, a frameshift mutation (c.1046-1047 delTG) in the DUF926 domain of NKAPL (*NKAPL*<sup>M349fs</sup>) is predicted to impair NKAPL protein function (Fig. 7a, b). To evaluate whether this frameshift mutation can cause male infertility, we generated *Nkapl*<sup>M349fs</sup> mutation in mice harboring a deletion of TG nucleotides in the DUF926 domain (Supplementary Fig. 8a, b). The heterozygous *Nkapl*<sup>+/M349fs</sup> mice were viable and appeared to be healthy (Supplementary Fig. 8c). This frameshift mutation produced a truncated NKAPL protein comprising only the DUF926 domain half (Fig. 7d). Interbreeding of heterozygous *Nkapl*<sup>+/M349fs</sup> mice produced a normal Mendelian ratio of *Nkapl*<sup>+/+</sup>, *Nkapl*<sup>+/M349fs</sup>, and *Nkapl*<sup>M349fs/M349fs</sup> offsprings (25.7% *Nkapl*<sup>+/+</sup>, 53.1% *Nkapl*<sup>+/M349fs</sup>, and 21.2% *Nkapl*<sup>M349fs/M349fs</sup>). *Nkapl*<sup>M349fs/M349fs</sup> homozygotes exhibited highly atrophied testes with a remarkable reduction in testis weight at 8 weeks old (Fig. 7e, f). The *Nkapl*<sup>M349fs</sup> homozygotes were infertile, and mature sperm were absent in the cauda epididymis (Fig. 7g and Supplementary Fig. 8e). This phenotype is similar to *Nkapl* KO males. Interestingly, the sperm counts of *Nkapl*<sup>+/M349fs</sup> heterozygotes were 10% less than those of wild-type males (Fig. 7g). Nevertheless, heterozygous *Nkapl*<sup>+/M349fs</sup> males were fertile, with normal litter size (Supplementary Fig. 8d).

Histological analysis of *Nkapl*<sup>M349fs</sup> homozygotes revealed defects in meiotic exit with an arrest at the late stage of meiosis or early stage of round spermatids (Fig. 7h), similar to *Nkapl* knockouts. Abnormal late stages of meiotic spermatocytes, including aberrant metaphase and secondary spermatocytes, were frequently observed in seminiferous tubules at stage XII of *Nkapl*<sup>M349fs</sup> homozygotes (Fig. 7h). Moreover, the remaining germ cells those complete meiosis in *Nkapl*<sup>M349fs</sup> homozygotes were uniformly arrest at post-meiotic round spermatid stage (Fig. 7h). Thus, *Nkapl*<sup>M349fs</sup> homozygotes exhibited the same meiotic exit phenotype with *Nkapl* KO mice.

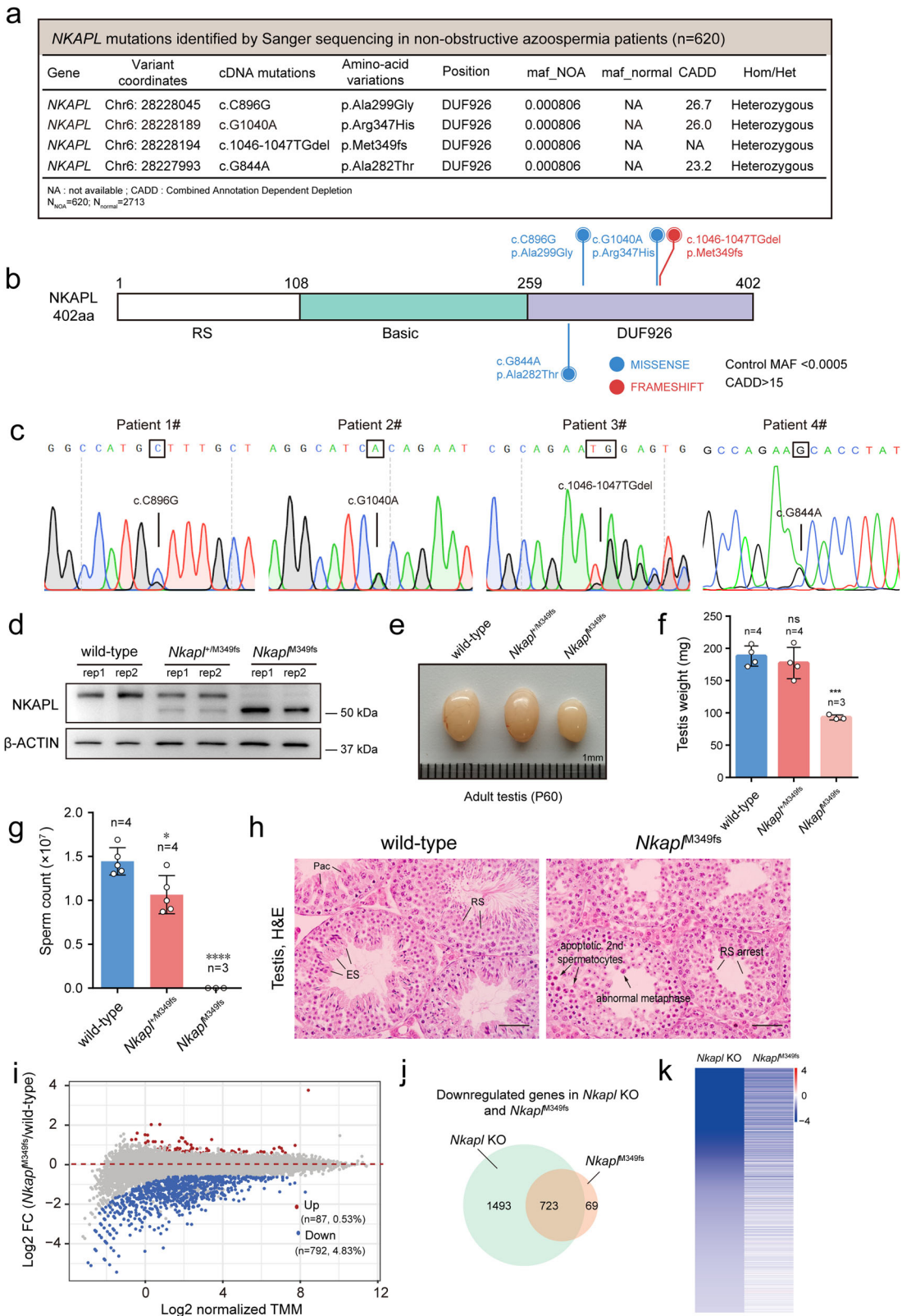
To address whether the observed phenotypes in *Nkapl*<sup>M349fs</sup> homozygotes accompanied alteration in gene expression, we sequenced the testis transcriptomes from P21 *Nkapl*<sup>M349fs</sup>

homozygotes, *Nkapl*<sup>+/M349fs</sup> heterozygotes, and wild-type controls. RNA-seq analysis revealed that 87 and 792 genes were significantly up- and downregulated ( $p < 0.05$ , fold change  $\geq 1.5$ ), respectively, in *Nkapl*<sup>M349fs</sup> compared to the wild-type control (Fig. 7i), indicating that most of the gene expression changes are downregulation of transcripts. Genes downregulated in *Nkapl*<sup>M349fs</sup> were associated with categories related to haploid development, such as spermatid differentiation, flagellated sperm motility and acrosomal formation (Supplementary Fig. 8f). We confirmed decreased transcript levels for these haploid genes using qRT-PCR (Supplementary Fig. 8g). We note that the genes downregulated in *Nkapl*<sup>M349fs</sup> homozygotes was smaller than those downregulated in *Nkapl* knockouts. Closer inspection of the RNA-seq data generated from *Nkapl* knockout and *Nkapl*<sup>M349fs</sup> homozygotes revealed a more notable reduction of spermatid development genes in *Nkapl* knockouts, compared to that gene expression changes in *Nkapl*<sup>M349fs</sup> homozygotes (Fig. 7j, k). These results suggest that the truncated NKAPL protein retained some functions. Nonetheless, these results support the NKAPL loss-of-function as the primary pathogenic role for M349fs. Collectively, these data indicate that the DUF926 domain of NKAPL is required for meiotic exit, and mutations in the DUF926 play a causal role in human azoospermia.

## Discussion

Meiosis is a unique stage in haploid gamete formation in sexually reproducing organisms. During male meiosis, X and Y chromosomes are specifically silenced in a process called meiotic sex chromosome inactivation (MSCI). It has been widely hypothesized that the transcriptional silencing of the X chromosome during meiosis is the evolutionary force that drives the retroposition of X-linked genes to autosomes<sup>44,45</sup>. *Nkapl* represents a functional retrotransposed gene of its X-linked progenitor *Nkap*, which is silenced during male meiosis. Similar to many X-originated retrogenes, *Nkapl* exhibits a testis-specific expression pattern and is abundantly expressed during the late stage of meiosis and early stage of round spermatids. In contrast to its progenitor NKAP's role in transcriptional repression, NKAPL depletion caused a dominant downregulation of transcripts, especially post-meiotic haploid genes. NKAPL depletion also results in an increased level of paused Pol II at promoter-proximal regions of its target genes. This increased accumulation of paused Pol II, along with the evidence that NKAPL binding generally co-localized with promoter-proximal Pol II peaks, suggest that NKAPL acts as a positive elongation regulator during meiotic exit.

It has been assumed that X-originated retrogenes have evolved essential functions to accomplish the special needs of germ cells. To avoid interference between extensive chromatin reorganization and gene transcription, we proposed that germ cells form a highly specialized transcriptional strategy that ensures an efficient and accurate transition into the post-meiotic haploid program at the meiotic exit. In this context, we provide insights into a meiotic exit-specific



transcriptional cascade to better understand how these components cooperate and their functional interdependence. SOX30 protein contains the conserved HMG domain with a high DNA-binding affinity, and early in vitro studies showed that SOX30 is able to recognize and bind DNA substrates<sup>46</sup>. In vivo, SOX30 associates with HDAC3, which promotes transcriptome reprogramming in a deacetylase-independent manner at the meiotic exit<sup>34</sup>. SOX30 is required for the recruitment of

HDAC3 to sites of transcription. Our data suggest that the assembly of the transcription initiation complex is not dependent on NKAPL because the SOX30-HDAC3 interaction remains intact in NKAPL-depleted testis. On the contrary, NKAPL depletion substantially strengthens the binding between SOX30/HDAC3 to DNA. This stalled transcription initiation complex containing SOX30/HDAC3 is likely a consequence of prolonged Pol II pause at promoter-proximal sites in

**Fig. 7 | Mutations in *NKAPL* contribute to human non-obstructive azoospermia.**

**a** *NKAPL* mutations identified in four patients with azoospermia. Three heterozygous missense mutations and a frameshift mutation were identified: heterozygous missense mutations and a frameshift mutation (c.1046\_1047T>Gdel [p.M349fs, referred to as *NKAPL*<sup>M349fs</sup>]). **b** *NKAPL* protein has an N-terminal arginine/serine-rich (RS) domain, repetitive basic sequences (the basic domain) and a C-terminal DUF926 domain of unknown functions. The deletion of TG base in *NKAPL* gene converts arginine 355 to a stop codon (R355\*), thus causes a premature termination. **c** Sanger sequencing confirmed three heterozygous missense mutations and a frameshift mutation in azoospermia patients. **d** Western blot showing the presence of wild-type and truncated *NKAPL* proteins in testes from wild-type, *Nkapl*<sup>+/M349fs</sup> heterozygotes and *Nkapl*<sup>M349fs</sup> homozygotes at P21. Experiments were performed with biological duplicates. **e**, **f** Testicular atrophy in *Nkapl*<sup>M349fs</sup> mice at 8-week-old. Testis sizes (**e**), testis weights (**f**) of *Nkapl*<sup>+/+</sup>, *Nkapl*<sup>+/M349fs</sup> and *Nkapl*<sup>M349fs</sup> mice.  $n = 4$  for *Nkapl*<sup>+/+</sup> and *Nkapl*<sup>+/M349fs</sup>,  $n = 3$  for *Nkapl*<sup>M349fs</sup>. Data are presented as

mean  $\pm$  SD.  $p < 0.001$ , two-tailed unpaired Student's *t* test. **g** Absence of sperm in *Nkapl*<sup>M349fs</sup> males.  $n = 3$ , 8-week-old males. Data are presented as mean  $\pm$  SD.

\*  $p < 0.05$ , \*\*\*\*  $p < 0.0001$ , two-tailed unpaired Student's *t* test. **h** Histological analysis of *Nkapl*<sup>+/+</sup> and *Nkapl*<sup>M349fs</sup> testes at 8 week-old. Apoptotic late stages of spermatocytes in stage XII tubules of *Nkapl*<sup>M349fs</sup> are marked by black arrows. *Nkapl*<sup>M349fs</sup> tubules were arrested either at late stages of spermatocytes (black arrows) or at the round spermatid stage (black lines). Scale bars, 20  $\mu$ m. **i** RNA-seq analysis of differentially expressed genes in *Nkapl*<sup>M349fs</sup> homozygotes compared with their wild-types at P21. Blue dots represent significantly downregulated transcripts, and red dots indicate upregulated transcripts (fold change  $\geq 1.5$ ,  $p < 0.05$ ). DESeq2 with the Wald test method was used. **j** Venn diagram depicting the overlap between downregulated genes in *Nkapl* KO and *Nkapl*<sup>M349fs</sup> homozygotes. **k** Downregulated genes in *Nkapl* KO (fold change  $\geq 1.5$ ,  $p < 0.05$ ) with DESeq2 were used to produce a heatmap depicting their corresponding expression levels in *Nkapl*<sup>M349fs</sup> homozygotes.

the absence of *NKAPL*, indicating a role of *NKAPL* as an essential factor in driving the progression of the SOX30/HDAC3-mediated transcription initiation machinery (Fig. 6j). Pol II pausing is a late step during the initiation stage of transcription. Transcription initiation and elongation is thought to be interconnected, but how this occurs is not clear. Our data reveals that *NKAPL* is an essential elongation factor that facilitates the transition from transcription initiation into elongation. Our study suggests that a defect in Pol II elongation can result in a longer-lived transcription initiation complex and an increase in the genomic occupancy of transcription factors at promoters.

What controls the genomic recruitment of *NKAPL* in the testis? During ChIP-seq analyzes, *NKAPL* co-localizes with SOX30/HDAC3. Since *NKAPL* lacks a well-defined DNA-binding domain and its paralog *NKAP* was found to interact with HDAC3, we originally thought that *NKAPL* might bind to the genome through its association with HDAC3. However, we were not able to detect *NKAPL* from endogenous HDAC3 pull-down or HDAC3 from endogenous *NKAPL* pull-down in the testis. In HEK293T cells overexpressing the related proteins, we did observe enhanced *NKAPL*-HDAC3 interaction after the deletion of the *NKAPL* basic domain, an intrinsically disordered region implicated in forming dynamic protein interactions. We think that *NKAPL*-DNA interaction in the testis is likely indirect. Consistent with this idea, *NKAPL* ChIP-seq peaks are shifted downstream of TSS compared to the SOX30 ChIP-seq peaks, supporting the involvement of R-loops that is known to participate in forming networks of weak protein-protein interactions<sup>47</sup>.

RNA-DNA interactions exist at gene promoters and regulatory DNA elements under physiological conditions. Within the R-loops, the G-rich nascent RNA anneals back to the template DNA strand, and thereby, the non-template DNA strand is left in a single-stranded configuration within the R-loop region. Significant progress has been made to improve the resolution of R-loop detection methods to map their specific locations at a genome-wide scale<sup>13,16,41</sup>, and a recent study employed Cryo-EM technology to reveal the biochemical recognition basis of S9.6 to RNA:DNA hybrids<sup>48</sup>. Unlike previously thought, R-loop formation at the transcription site is a conserved and dynamic feature of mammalian chromatin, and DNA replication coordinates with genome topological states to regulate R-loops<sup>49</sup>. Using a strand-specific method ssDRIP-seq<sup>12</sup>, we profiled the landscape of R-loops in the mouse germline and revealed its remarkable features during meiotic exit. Our data show that similar to the R-loops in somatic cells<sup>13,16</sup>, germline R-loops are observed at a large portion of the mouse genome. R-loops are highly enriched at promoter regions in the genome of mouse male germ cells, strongly correlated with GAA-rich sequences. Indeed, a recent study using CUT & Tag-seq reported that R-loops preferentially occur at promoters in meiotic spermatocytes<sup>50</sup>. Sequences containing G-clusters were shown to initiate R-loops more efficiently than random sequences<sup>51</sup>. Biochemical studies with purified DNA and prokaryotic RNA polymerases also show G-richness increases the annealing of the nascent RNA to the template DNA. This reflects the

local thermodynamic stability of the RNA-DNA hybrid is a key factor for R-loop initiation. G-clusters on the non-template DNA strand favor R-loop formation as well. These G-clustered regions adopt a DNA conformation that is favorable for DNA breathing and DNA-protein interactions<sup>52</sup>, and they also provide a permissive DNA conformation for the elongation of the R-loop once it has initiated<sup>16</sup>. These observations indicate that the G-rich nascent transcript contributes to the R-loop initiation by increasing the stability of the RNA-DNA hybrid, and the upstream G-clusters on the non-template DNA facilitate R-loop elongation. Our data suggest that, in addition to the G-clusters, the GAA-rich motifs are also R-loop hotspots in mouse germ cells, which is consistent with plants and mouse somatic cells.

Proteins with a capacity for the nascent RNA and/or the non-template DNA strand also have a key role in modulating R-loop dynamics, in parallel with a significant function of genomic nucleic acid sequence and feature in initiation and elongation of R-loops. The chromatin state of R-loops results from the balance between its formation and removal, which is tightly controlled by regulatory factors. R-loop reader protein GADD45A binds R-loops<sup>53</sup>, and the RNA helicase DHX9 and human capping enzyme promotes the formation of non-pathological R-loops by RNA polymerase II<sup>10,54</sup>. However, excessive R-loops are pathological and mediate stalling and collapse of replication forks as a source of genome instability. Factors such as Senataxin<sup>55,56</sup> and RNase H1<sup>42,43</sup> act to resolve the R-loops. The conserved eukaryotic THO complex prevents R-loop formation and transcription-associated genome instability<sup>57,58</sup>. The DNA repair protein BRCA2 also suppresses R-loop accumulation<sup>59</sup>. Our data suggest that *NKAPL* is an essential factor for R-loop formation during active transcription to ensure the progression of the transcription machinery. Our eCLIP-seq revealed *NKAPL* binds prominently to promoter-associated nascent RNAs, although it could have some false-positive binding since the eCLIP-seq was performed only in wild-type testes. Motif analysis identified the top enriched motif within the *NKAPL* binding sites contained the repeated sequence “GAA”<sub>n</sub>, the same feature to promote R-loop formation. The recruitment of *NKAPL* at these genomic loci promotes R-loop formation, as depletion of *NKAPL* substantially reduces the formation of the R-loop structure. Repeat sequences have the potential to form secondary structures and sequester RNA-binding proteins<sup>60</sup>. Thus, the GAA repeated sequence might facilitate the recruitment of *NKAPL* to R-loop hotspots. We noticed the motifs identified from *NKAPL* ChIP-seq and eCLIP-seq analyzes are different, reflecting the possibility that the binding of *NKAPL* with nascent RNAs occurs after its transient and indirect interaction with DNA elements. *NKAPL* consists of low-complexity amino acid sequences with an ability to promote networks of weak protein-protein interactions. Our study suggests that, in addition to the aforementioned thermodynamic stability mechanism, the preference of R-loops for GAA sequence repeats is also attributable to GAA-recognizing proteins such as *NKAPL*. *NKAPL* binds nascent RNA



with a GAA-rich sequence and co-localizes with R-loop hotspots, where NKAPL forms an optimal environment to promote R-loop formation (Fig. 6j), possibly by stabilizing DNA-RNA hybrids structure or reducing the access of DNA-RNA hybrids to R-loop enzymes that remove the RNA chain.

Recent studies demonstrated an interesting link between R-loop structure and transcriptional elongation. DRB is a selective inhibitor of CDK9, the kinase of positive elongation factor b (P-TEFb) required for transcription elongation. Treatment of human cells with 5,6-dichloro-1- $\beta$ -D-ribofuranosyl-benzimidazole (DRB) inhibits mRNA transcription by enhancing Pol II pausing and reducing R-loop formation at promoter regions<sup>13</sup>. Our results also reveal that NKAPL deficiency increases Pol II pausing and causes a drastic drop in R-loops. These results are in line with the fact that R-loop formation occurs through transcribed genic regions and is generally associated with active transcription and efficient transcription elongation. Therefore, NKAPL acts as R-loop promoting factor to facilitate Pol II pause-release into productive elongation. It is also possible that NKAPL could provide a functional environment that incorporates proteins and substrates for hyperphosphorylation of the CTD and efficient R-loop elongation. Having said this, excessive or unscheduled R-loops under pathological conditions can cause transcription stress, DNA damage, and genomic instability<sup>61</sup>. For example, mutations of elongation factor TFIIS result in slower elongation rates and an accumulation of R-loop structures<sup>62</sup>. These results suggest that R-loops can positively or negatively affect transcription elongation, but the precise mechanism for the difference remains unclear. Sequencing analyzes reveal substantially more widespread transcription and higher expression activity in meiotic spermatocytes and postmeiotic spermatids compared with other germ cells or somatic tissues<sup>63</sup>, potentially reflecting specific functional requirements. Meiosis is a unique stage in haploid gamete formation in sexually reproducing organisms. This process ensures genetic diversity through meiotic recombination, which is initiated at programmed hundreds of DNA double-strand breaks (DSBs) generated by SPO11<sup>64</sup>. To minimize the interference between local transcription and meiotic chromosome events, meiotic germ cells need to employ a complicated strategy for R-loop regulation. Consistent with this view, defective meiotic progression and SPO11-mediated DSBs persistence were observed in mice deficient for the R-loop helicase Senataxin<sup>65,66</sup>, which resolves RNA:DNA hybrids forming at DSBs. Despite the ubiquitously expression pattern of Senataxin in various tissues, depletion of Senataxin in vivo causes an obvious phenotype only in mouse testis, probably due to the increased susceptibility of meiotic spermatocytes to R-loop dysregulation<sup>67</sup>. This raised the possibility that germ cells might employ a specific and noncanonical R-loop regulation to achieve robust transcription activity and chromatin reorganizations during meiotic exit. Future studies are needed to elucidate the biochemical basis for the function of NKAPL in promoting R-loop formation.

## Methods

### Human participants

This human study was approved by the Ethical Review Board of the First Affiliated Hospital of Nanjing Medical University. All participants signed informed consent before participating in this research. A total of 620 patients with non-obstructive azoospermia and 2713 fertile males were recruited at the Clinical Center of Reproductive Medicine in Nanjing for this population study. All control males had fathered at least one healthy child. All the participants signed the informed content document. Semen sperm analysis was performed in biological replicates to confirm the absence of sperm, and patients with a medical history of cryptorchidism, testicular trauma, vasectomy procedure, orchitis, obstruction of the vas deferens, testis inflammation, or endocrine disorders were excluded. These azoospermia patients were further selected with the absence of the most common genetic causes of infertility, including chromosomal abnormalities and Y

chromosomal microdeletions. Whole blood samples were obtained from enrolled participants as a source of genomic DNA for Sanger sequencing analysis, and experimental protocols on human subjects were approved by the Ethics Committee of Nanjing Medical University. The *NKAPL* coding exon was amplified by PCR using forward primer (Forward 5'-CCGTTTGTAAGTACAGGAAGC-3') and reverse primer (Reverse 5'-TCTTTCTGGGCTTCTTTTGGTCC-3'). PCR products were purified and sequenced according to the manufacturer's instructions.

### Animals

All animals were maintained and used according to the guidelines of the Institutional Animal Care and Use Committee of Nanjing Medical University.

### *Nkapl* knockout mice

*Nkapl* is a retrotransposed gene that lacks introns. We targeted the endogenous *Nkapl* locus with two guide RNAs: 5'-AGATCGGGATACAGGCGACA-3' and 5'-GAGAACGACCGCTACCGCTG-3'. Cas9 mRNA and purified sgRNAs were introduced into C57BL/6J mouse zygotes by electroporation. Mutant analysis of the founders was performed using genomic DNA extraction, PCR, and Sanger sequencing. Among 8 founders, one *Nkapl* mutant female had a 50-bp deletion in exon 1, which produces premature termination. A stable *Nkapl* mutant mouse line harboring a 50-bp deletion (Line #1) was established by breeding the *Nkapl* mutant female with wild-type C57BL/6J males and confirmed by sequencing. Additionally, a female founder harbors a single bp insertion after the ATGT base of the start codon, followed by a 13-bp deletion (Line #2). Germline transmission of this mutation was validated by PCR and Sanger sequencing (Supplementary Fig. 1e, f). This mutation also created premature termination codons. Both mutations in Line #1 and Line #2 result in the loss of NKAPL protein. Therefore, Line #1 and Line #2 are collectively referred to as *Nkapl* KO mice. *Hdac3*<sup>fl/m</sup> mice were generously provided by Mitchell A Lazar at the University of Pennsylvania<sup>68</sup>, and this floxed *Hdac3* allele was generated using homologous recombination in embryonic stem cells. *Hdac3*<sup>fl/m</sup> mice were crossed with *Stra8-cre* mice<sup>69</sup> to generate germ cell-specific *Hdac3* knockout mice (*Stra8-cre/Hdac3* KO). *Sox30* KO mice were generated through CRISPR/Cas9 with two sgRNAs targeting the exon 2, which contains the DNA binding HMG-box domain.

### Generation of *Nkapl*<sup>M349fs</sup> mice

CRISPR/Cas9-mediated genome editing was used to generate *Nkapl*<sup>M349fs</sup> (M349 frameshift mutation). Guide RNAs (sgRNA) 5'-TAGCAGGCATCGCCGAATGG-3' for *Nkapl*<sup>M349fs</sup> was cloned into the pUC57 expression vector, in vitro transcribed, and purified as previously described<sup>70</sup>. The synthesized single-stranded oligodeoxynucleotide (ssODN) contains the M349fs (ATG to A):

5'AAGCGTATCCACGAAGAGGTGAAATTGGGTTGACAAGTGAAGAGATTGCCTCATTTGAATGTTTACGTTACGTCATGAGTGGTAGCAGCATCGCCGAA( $\Delta$ TG)GAGGCTGTAAAGACTGCGTAAAGAGAACCAGATTACAGTGCTGATGAGAAACGAGCCCTTGATCCTTTAAACCAAGAAGAGAGGCGGAAGAGAGAAAAATA-3'. Cas9 mRNA, sgRNA and ssODN were then microinjected into C57BL/6J mouse zygotes by electroporation. Embryos were allowed to recover and implanted into pseudo-pregnant females. Genomic DNA extraction, PCR, and Sanger sequencing were employed to screen mutations in the founders. Among 17 *Nkapl*<sup>M349fs</sup> founders, the genotypes were: 2 *Nkapl*<sup>M349fs/FS</sup> (1 male, 1 female) and 2 *Nkapl*<sup>M349fs/M349fs</sup> (1 male, 1 female). The *Nkapl*<sup>M349fs/FS</sup> male founders were infertile. Founders *Nkapl*<sup>M349fs/M349fs</sup> and *Nkapl*<sup>M349fs/FS</sup> females were mated with wild-type C57BL/6J males for breeding heterozygous *Nkapl*<sup>M349fs</sup> mice. Offsprings harboring the *Nkapl* M349 frameshift mutation (PCR primers 5'-CCATGCTCTGCTTCTCTGGTGAA-3' and 5'-CTCTTCCGCTCTCTTCTTGGT-3') were confirmed by PCR and Sanger sequencing (Supplementary Fig. 8a).

### Histological, chromosome spread, and TUNEL analyzes

Testes and epididymides were freshly fixed in Bouin's solution (Sigma, SLBJ3855V) for 24–48 hrs, dehydrated in graded ethanol (50%, 70%, 95%, 100%), and embedded in paraffin overnight at 55 °C. Sections were stained with hematoxylin and eosin. TUNEL staining was performed on frozen sections with the TUNEL BrightGreen Apoptosis Detection Kit (Vazyme, A112-02) according to the manufacturer's instructions. Nuclear spread analysis was performed as previously described<sup>71</sup>. The following primary antibodies were used for immunofluorescence: anti-SYCP3 (1:100, Abcam, ab97672), anti-SYCP1 (1:100, Abcam, ab15090), anti-DMC1 (1:100, Santa Cruz, sc-22768), anti-MLH1 (1:50, BD Biosciences, 550838) and  $\gamma$ H2AX (1:800, Millipore, 16-202 A). Spread confocal images were visualized on a Zeiss confocal microscope.

### Immunoprecipitation and immunoblotting

Tissues were rinsed with PBS and lysed in cold RIPA buffer supplemented with protease inhibitor cocktail tablets (Roche, 4693124001). Homogenized lysates were rotated at 4 °C for 1 hr, and centrifuged at 15,000 g for 40 mins at 4 °C. The protein concentration of the collected supernatant was determined by Bicinchoninic Acid (BCA) assay. 20  $\mu$ g proteins for each lane were separated by sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE). For immunoprecipitation, tissue mass (80 mg) were dounced 30 strokes or  $1 \times 10^7$  HEK293T cells were lysed in 500  $\mu$ l lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 0.5% sodium dodecylsulphate, 1 mM DTT, 1 $\times$ protease inhibitor) and left on ice for 10 mins. Lysates were further diluted with 500  $\mu$ l of dilution buffer (20 mM Tris pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM DTT, 1 $\times$ protease inhibitor). Lysates were centrifuged, and the supernatant was collected. The supernatant was pre-cleared with washed protein A/G beads (Millipore, 16-156) for 2 hrs, then incubated with 10  $\mu$ g indicated antibodies anti-HDAC3 (Abcam, ab7030), anti-SOX30 (ABclonal, A11759), anti-HA (CST, 3724S), anti-FLAG (MBL, M180-3) overnight at 4 °C with rotation. 25  $\mu$ l washed protein A beads were loaded into the mixture and incubated for 3 hrs at 4 °C with rotation. Bead complexes were pelleted and washed 5 times with wash buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA). The protein complex was eluted off the beads into 50  $\mu$ l 2 $\times$ loading buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH 6.8, 10% b-mercaptoethanol, 0.02% bromophenol blue).

### Immunofluorescence assay

Tissues were fixed in 4% paraformaldehyde (PFA) at 4 °C overnight. Fixation tissues were dehydrated with ethanol and cleaned in xylene. Then tissues were embedded in paraffin and cut into 5  $\mu$ m sections. Paraffin-embedded sections were rehydrated gradually in ethanol (100%, 95%, 70%, 50%) and distilled water. Slides were treated in 10 mM sodium citrate at 95 °C for 15 mins and washed in PBS twice after cooling to room temperature. Slides were blocked with the buffer (10% FBS, 1% BSA, 1% Triton X-100, 0.05% Tween-20 in PBS) at room temperature for 1 hr and incubated with primary antibodies diluted with blocking buffer at 4 °C overnight. Primary antibodies are as follows:  $\gamma$ H2AX (1:500, Millipore, 16-202 A); CREST (1:50, CREST, Antibodies Incorporated, 15-234); Phospho-Histone H3 (Ser10) (1:200, CST, 9701S);  $\alpha$ -Tubulin (1:200, Sigma, F2168). After three washes for 5 mins in PBS buffer with Tween (0.2% Tween-20 in PBS), slides were incubated at 37 °C for 1 hr in the dark with the secondary antibodies, and PNA (FITC-conjugated peanut agglutinin, Vector Labs, CL-1073-1). After secondary antibody incubation, slides were washed three times with PBS buffer with Tween and mounted with DAPI (Sigma, F6057). All the staining slides were visualized on a confocal microscope (Carl Zeiss, LSM700).

### Spermatogenic cell isolation

Pachytene spermatocytes and round spermatids were isolated from wild-type and *Nkapl* KO testes. Cells from 5 wild-type mice or 6 KO

mice were pooled for one biological replicate. Testes tissues were digested with collagenase buffer (0.25 mg/ml collagenase, 5  $\mu$ g/ml DNaseI, 0.5 mg/ml BSA in HBSS-with  $\text{Ca}^{2+}/\text{Mg}^{2+}$ ) at 37 °C for 5 mins. Cell suspension was subsequently digested with Trypsin buffer (1 mg/ml Trypsin, 5  $\mu$ g/ml DNase I in HBSS-without  $\text{Ca}^{2+}/\text{Mg}^{2+}$ ) at 37 °C for 10 mins to prepare single-cell suspensions. The cell suspension was then slowly filtered through a 70  $\mu$ m cell strainer to obtain single cells and to minimize the number of somatic cells. The cell pellets were re-suspended in elutriation buffer (1 $\times$ KREBS, 1 mM EDTA, 0.1% BSA) with 5  $\mu$ g/ml DNase I. Centrifugal elutriation technique (Beckman, J26S-XP) was used to isolate pachytene spermatocytes and round spermatids. Single-cell suspensions were further loaded in the sample chamber to collect stage-specific germ cell populations. Their distinct cell types were evaluated based on their morphological characteristics, cell diameters and DAPI staining pattern under a light microscope. In this study, the average purity of pachytene spermatocytes from wild-type and *Nkapl* KO was 81.9% and 77.6%, respectively. The average purity of round spermatids from each genotype was 96.7% and 95.1%, respectively.

### Culture of spermatocytes and okadaic acid treatment

Pachytene spermatocytes were isolated from adult and P17 wild-type and *Nkapl* KO testes with centrifugal elutriation technique (Beckman, J26S-XP). Cells were then cultured in MEM $\alpha$  containing 5% streptomycin, 7.5% penicillin, 0.29% DL-lactic acid sodium salt, 0.59% Hepes and 5% fetal bovine serum in six-well plates in 5%  $\text{CO}_2$  at 32 °C. After 24 hrs of culture, the experimental group cells were treated with 4  $\mu$ M okadaic acid, and the control group was treated with an equivalent volume of ethanol. After 5 hrs of okadaic acid (OA) treatment, cells were collected for nuclear spread analysis. Cells were treated in hypotonic buffer for 5 mins and incubated with anti-SYCP3 (1:100, Abcam, ab97672), anti-SYCP1 (1:100, Abcam, ab15090), and anti-CREST (1:100, Antibodies Incorporated, 15-234) antibodies. Spread confocal images were visualized on a confocal microscope (Carl Zeiss, LSM900).

### Quantitative RT-PCR Assay and RNA-seq

Testes samples were collected from wild-type and *Nkapl* KO mice (biologically triplicates for each genotype) at postnatal day 21. Total RNA was extracted from the samples using TRIzol reagent (Invitrogen). The concentration and purity of RNA were determined by absorbance at 260/280 nm. 1  $\mu$ g of total RNA was reverse transcribed using a PrimeScript<sup>TM</sup>RT Master Mix (TaKaRa). The cDNA was diluted by 5–6 times, and 1  $\mu$ l cDNA was used for each reaction using SYBR Green Premix Ex Taq II (RR820A, TaKaRa). A standard 20  $\mu$ l reaction volume contained 200 nmol/l of forward and reverse primers, 1  $\mu$ l cDNA and 10  $\mu$ l of SYBR Green Mix. PCR amplification was performed using StepOnePlus<sup>TM</sup> Real-Time PCR System (Applied Biosystems) with the following conditions: 95 °C for 30 s; 95 °C for 10 s, 60 °C for 30 s, 40 cycles. All reactions were performed in triplicate, and gene expression was normalized to the housekeeping gene *36B4* (*Arbp*). All the primers were listed in Supplementary Table 1. For RNA-seq, strand-specific libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation kit (Illumina) according to the manufacturer's instructions before submitting to the Illumina NovaSeq 6000 system. RNA-seq library preparation and sequencing were performed at Omics Core of Bio-Med Big Data Center, Chinese Academy of Sciences (Shanghai, China). Clean data were obtained by trimming the adapter sequence and removing sequences with low quality. Clean reads were mapped to the mouse genome (mm10) with HISAT2 (v.2.1.0; Johns Hopkins University) with a GTF file download from the UCSC database (University of California, Santa Cruz). The aligned reads of genes were counted using HTSeq followed by DESeq2 normalization to evaluate gene expression as normalized counts per million. Significant differentially expressed genes were identified as those with a *p*-value above the threshold (*p* < 0.05, fold-change  $\geq$  1.5) using DESeq2 software.

### ChIP-seq and data analysis

Fresh testes were collected from wild-type and *Nkapl* KO mice at postnatal day 21, ground into powder, and crosslinked by 1% formaldehyde, followed by quenching with glycine solution. Chromatin was fragmented by sonication in ChIP SDS lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, pH 8.0, 1% SDS) using the Covaris-S220 sonicator. Crosslinked chromatin was incubated with indicated antibodies (anti-NKAPL, Abclonal, E10124; anti-HDAC3, Abcam, ab7030; anti-SOX30, Abclonal, A11759; anti-RNA Pol II, Millipore, 05-623) in ChIP dilution buffer (50 mM HEPES pH 7.5, 155 mM NaCl, 1.1% Triton X-100, 0.11% Na-deoxycholate, 1 mM EDTA) with protease inhibitors overnight. Crosslinking was reversed overnight at 65 °C, and DNA was extracted by using phenol/chloroform/isoamyl alcohol. Precipitated DNA was amplified for deep-sequencing. For ChIP-seq, DNA was amplified according to ChIP Sequencing Sample Preparation Guide provided by Illumina using adapters and primers. Reads were aligned to the mouse reference genome mm10 with BOWTIE software (bowtie2 version 2.3.4.3), followed by the removal of those reads that with low mapping quality and duplicated through sample amplification or sequencing. Peak calling was carried out by using MACS2 (version 2.1.2, with options '-mfold 5, 50-p 0.0001'), on ChIP file against the input file. Genome-wide normalized signal coverage tracks were created by bamCoverage in deepTools (version 3.3.0) and visualized in the Integrative Genomics Viewer (IGV version 2.5.0). Peaks were annotated to the genomic region and the nearest genes within 2 kb of TSS using Bioconductor package ChIPSeeker (version 1.16.1). Peaks overlapping by at least 1 nt with unique gene model promoters ( $\pm$  2 kb of each unique gene model Transcription Starting Site) were considered as promoter located. De novo motif searches of ChIP-seq peaks were performed using Homer (version v4.11.1) with default parameters<sup>72</sup>.

### ssDRIP-seq library construction, sequencing, and data processing

ssDRIP experiments were performed on testis tissues using a standard ssDRIP protocol<sup>12,73</sup>. Briefly, nuclei were isolated from testis tissues before SDS/proteinase K treatment at 37 °C for 4–6 hrs, followed by genomic DNA extraction with the phenol:chloroform:isoamyl alcohol method. The precipitated DNA pellet was washed, air dried, resuspended in TE buffer, and quantified by the Qubit dsDNA HS Assay kit (Invitrogen, Q32854). 10  $\mu$ g gDNA was digested with a combination of restriction enzymes of MseI, DdeI, AluI, and MboI at 37 °C overnight, and the negative control was treated with RNase H (New England Biolabs, M0297S). For DNA:RNA hybrids immunoprecipitation (DRIP), 5  $\mu$ g gDNA was resuspended in 500  $\mu$ l TE buffer. 450  $\mu$ l gDNA mixture was incubated with 10  $\mu$ g S9.6 antibody for DNA:RNA immunoprecipitation, and the remaining 50  $\mu$ l was used as input DNA. Protein G dynabeads were pre-washed for three times with 1 $\times$ DRIP binding buffer (100 mM sodium phosphate pH 7.0, 1.4 M NaCl, 0.5% Triton X-100) and then added to the DNA-antibody mixtures with gentle shaking at 4 °C for 3–4 hours. DNA was eluted with DRIP elution buffer, and then sheared to 250 bp in size by sonication using the Covaris-S220. These DNA samples were denatured at 95 °C for 2 mins and then incubated on ice for 2 mins to obtain ssDNA fragments. ssDRIP-seq libraries were constructed by using the VAHTS ssDNA Library Prep Kit for Illumina (Vazyme, ND620-02), following instructions from the manufacturer. The quality of each library was evaluated with an Agilent Bioanalyzer, and deep sequencing was performed on Illumina NovaSeq (2 $\times$ 150). ssDRIP data were analyzed following a standard ssDRIP-Seq pipeline<sup>73</sup>. Adapters and low mapping quality bases of raw reads were removed by Cutadapt (v3.6) after quality control. Clean reads were aligned to the mouse reference genome mm10 with BOWTIE software (Bowtie2 version 2.4.5), then the duplicate reads marked by Picard Mark Duplicates tools (v2.27.4) were dropped. The total mapped reads were divided into forward and reverse strands. We carried out peak calling using MACS2 (version 2.2.7.1) only on ssDRIP samples firstly, and these peaks

were merged by Deseq2 to find significant peaks through samples in replicates against the input. Bigwig format files were created by bamCoverage tool in deepTools (version 3.5.1) with library size normalized. Motifs were searched by MEME (v5.0.5) with custom parameters. Peaks annotation referred to GTF file (vM23) from Gencode.

### eCLIP-seq and data analysis

Enhanced crosslinking immunoprecipitation coupled with sequencing (eCLIP-seq) was performed essentially as described<sup>74,75</sup>. Testes were collected from wild-type mice at P21, and the subsequent PBS suspension containing seminiferous tubules was crosslinked by UV-irradiation three times at 254 nm (400 mJ/cm<sup>2</sup>) and lysed in eCLIP lysis buffer with RNase I (LifeTech) and Turbo DNase (LifeTech). NKAPL-RNA complexes were immunoprecipitated with 10  $\mu$ g anti-NKAPL antibodies (Abclonal, E10124) and protein A dynabeads. 2% of lysates were used as input and processed in parallel with immunoprecipitated samples. NKAPL-RNA complexes were dephosphorylated by FastAP (LifeTech), and ligated with 3' RNA adapter using T4 RNA Ligase (NEB). NKAPL-RNA complexes were separated on SDS-PAGE gel, transferred to a nitrocellulose membrane, and the membrane between 50 and 125 kDa was collected for RNA extraction with acid phenol/chloroform/isoamyl alcohol. After purification, RNAs were reverse transcribed with AffinityScript (Agilent) followed by treatment with ExoSAP-IT (Affymetrix) to remove excess oligonucleotides. The DNA adapter was ligated to the 3' end of cDNA with T4 RNA Ligase (NEB). Samples were cleaned up with MyOne Silane Dynabeads (Thermo-fisher) and subject to quantitative PCR to determine the appropriate number of PCR cycles. Libraries were amplified with Q5 Polymerase (NEB). Amplified PCR products were separated on 3% low-melting temperature agarose gel (Seakem GTG LMP), and size-selected libraries were purified with a MinElute gel extraction kit (Qiagen). eCLIP data were analyzed using the ENCODE standard eCLIP pipeline (ENCODE Project Consortium, 2012). The adapters and 3' adapter-dimers were trimmed with cutadapt (V1.14), and PCR duplicates were removed with a custom python script. The reads aligned to the repeat elements from UCSC with STAR (V2.7.6a) were filtered out before mapping to mm10 reference genome (GENCODE vM23). CLIPPER was used to call peak regions as previously described<sup>74</sup>. Significantly enriched peaks were identified using the following parameters: IDR (irreproducible discovery rate)  $\leq$  0.01, FDR  $<$  0.05, fold enrichment  $\geq$  2. MEME was used to enrich the binding motifs among the peaks<sup>72</sup>.

### Protein expression and purification

Full-length *Nkapl* gene and mEGFP were amplified by PCR from mouse testis cDNA and the mEGFP-plasmid, respectively. cDNAs encoding NKAPL and mEGFP were cloned into the T5 PET28A expression vector through homologous recombination. The base vector was engineered to include a 3' 6 $\times$ histidine tag. The expression constructs (NKAPL-mEGFP-PET28A and mEGFP-PET28A) were confirmed by sequencing and transformed into *E. coli* BL21 (DE3) cells. Protein expression was induced with 0.5 mM IPTG at 25 °C overnight. Harvested cells were resuspended in 15 ml lysis buffer (50 mM Tris-HCl pH 8.0, 750 mM NaCl, 2 mM MgCl<sub>2</sub>, 10% Glycerol, 0.1% Triton X-100, 1 M DTT, 1 mg/ml lysozyme and 1 mM PMSF) supplemented with 10 mM imidazole and sonicated (out power 20%, 25 cycles of 10 s on, 15 s off). The lysates were centrifuged at 4000 g for 30 mins at 4 °C, and the supernatant was incubated with 500  $\mu$ l Ni-NTA agarose beads (30410, Qiagen) that had been pre-equilibrated with 5 volumes of the binding buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 2 mM MgCl<sub>2</sub>, 10% Glycerol, 0.1% Triton X-100, and 10 mM imidazole). After 3 hrs rotation, Ni-NTA beads-protein complexes were extensively washed with the washing buffer contained different concentrations of imidazole (20 mM, 30 mM, 40 mM, 60 mM and 80 mM). Proteins were eluted with washing buffer containing 500 mM imidazole and concentrated using centrifugal filter units (MWCO 3 K) (Thermo Scientific). Proteins were



dialyzed against the buffer containing 50 mM Tris-HCl pH 7.5, 1 M NaCl, and 1 mM DTT at 4 °C and stored at -80 °C.

### Preparation of R-loop substrates and in vitro incubation of NKAPL and R-loop

The indicated ssRNA oligo was synthesized and labeled with Cy3 at the 5' end. For the R-loop substrate preparation, the Cy3-labeled ssRNA oligo was annealed with two complementary DNA oligonucleotides by heating them at 95 °C for 5 mins and cooling gradually to 16 °C in 50 µl annealing buffer (50 mM Tris-HCl pH 7.5, 150 mM KCl, and 0.1 mM EDTA). The annealing created R-loop substrate was incubated with NKAPL recombinant protein at the indicated concentrations. After incubation for 30 mins, images were captured with Zeiss confocal microscope (Carl Zeiss, LSM900). The oligonucleotide sequences used in this study are as follows:

RNA oligo: GACGAAGAGGAGGGGAGGAAGCGGAGCUGGACGG AGAAC

R-loop-DNA-F:

CTGCCTCCTCCAGCTCCTCCTCCTCCAGCAGTTCTCCGTCCAG  
CTCCGCTTCTCCCTCCTCTCTCGTCCACCTCCAGCTCCAGCTCCG  
CCGTCTCGGA

R-loop-DNA-R:

TCCGAGACGGCGGAGCTGGAGCTGGAGGTGAGTACCGATTGAT  
ATAGAACGATAAGTAGAATAAGAGGTTGCTGGAGGAGGAGGAGCTG  
GAGGAGGCAG

### Quantification and statistical analysis

All data are reported as mean ± SD unless otherwise noted in the figure legends. Significance was tested by using the 2-tailed unpaired Student's t-test (\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ) using Prism 7.0 (GraphPad Software, La Jolla, CA, USA).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw data fastq files and processed BigWig files for all ChIP-seq, RNA-seq, eCLIP-seq, and ssDRIP-seq have been deposited in the Gene Expression Omnibus with accession numbers [GSE232398](#), [GSE232415](#), [GSE232416](#), and [GSE232417](#). Source data are provided with this paper.

### References

- Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. cell Biol.* **18**, 263–273 (2017).
- Noe Gonzalez, M., Blears, D. & Svejstrup, J. Q. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. cell Biol.* **22**, 3–21 (2021).
- Rahl, P. B. et al. c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
- Mayer, A. et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
- Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. cell Biol.* **16**, 167–177 (2015).
- Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).
- Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
- Yu, M. et al. RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).
- Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
- Chakraborty, P., Huang, J. T. J. & Hiom, K. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* **9**, 4346 (2018).
- Edwards, D. S. et al. BRD4 prevents R-loop formation and transcription-replication conflicts by ensuring efficient transcription elongation. *Cell Rep* **32** (2020).
- Xu, W. et al. The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat. Plants* **3**, 704–714 (2017).
- Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
- Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
- Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
- Chen, L. et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745 (2017).
- Lin, S. & Fu, X. D. SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.* **623**, 107–122 (2007).
- Ji, X. et al. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* **153**, 855–868 (2013).
- Long, J. C. & Cáceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical J.* **417**, 15–27 (2009).
- Misteli, T. & Spector, D. L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* **3**, 697–705 (1999).
- Sapra, A. K. et al. SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol. Cell* **34**, 179–190 (2009).
- Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
- Pajeroski, A. G., Nguyen, C., Aghajanian, H., Shapiro, M. J. & Shapiro, V. S. NKAP is a transcriptional repressor of notch signaling and is required for T cell development. *Immunity* **30**, 696–707 (2009).
- Burgute, B. D. et al. NKAP is a novel RS-related protein that interacts with RNA and RNA binding proteins. *Nucleic acids Res.* **42**, 3177–3193 (2014).
- Chen, D. Y. et al. Identification of a nuclear protein that promotes NF-kappa B activation. *Biochem Biophys. Res Commun.* **310**, 720–724 (2003).
- Shapiro, M. J., Lehrke, M. J., Chung, J. Y., Romero Arocha, S. & Shapiro, V. S. NKAP Must Associate with HDAC3 to Regulate Hematopoietic Stem Cell Maintenance and Survival. *J. Immunol.* **202**, 2287–2295 (2019).
- Pajeroski, A. G. et al. Adult hematopoietic stem cells require NKAP for maintenance and survival. *Blood* **116**, 2684–2693 (2010).
- Dash, B. et al. The Interaction between NKAP and HDAC3 Is Critical for T Cell Maturation. *Immunohorizons* **3**, 352–367 (2019).
- Hsu, F. C., Pajeroski, A. G., Nelson-Holte, M., Sundsbak, R. & Shapiro, V. S. NKAP is required for T cell maturation and acquisition of functional competency. *J. Exp. Med.* **208**, 1291–1304 (2011).
- Li, T. et al. SUMOylated NKAP is essential for chromosome alignment by anchoring CENP-E to kinetochores. *Nat. Commun.* **7**, 12969 (2016).
- Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–2178 (2002).

32. Schmidt, E. E. & Schibler, U. High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids. *Development* **121**, 2373–2383 (1995).
33. Okuda, H. et al. A novel transcriptional factor Nkapl is a germ cell-specific suppressor of Notch signaling and is indispensable for spermatogenesis. *Plos One* **10**, e0124293 (2015).
34. Yin, H. Q. et al. HDAC3 controls male fertility through enzyme-independent transcriptional regulation at the meiotic exit of spermatogenesis. *Nucleic acids Res.* **49**, 5106–5123 (2021).
35. Bai, S. et al. Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. *Development* **145**, dev164855 (2018).
36. Feng, C. A. et al. SOX30 is required for male fertility in mice. *Sci. Rep.* **7**, 17619 (2017).
37. Zhang, D. et al. The transcription factor SOX30 is a key regulator of mouse spermiogenesis. *Development* **145**, dev164723 (2018).
38. Chen, Y. et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res* **28**, 879–896 (2018).
39. Hsieh, C. L., Xia, J. & Lin, H. F. MIWI prevents aneuploidy during meiosis by cleaving excess satellite RNA. *Embo. J.* **39**, e103614 (2020).
40. Shapiro, V. S., Pajeroski, A., Nguyen, C., Aghajanian, H. & Shapiro, M. J. NKAP, a novel modulator of Notch signaling, is required for T cell development. *J. Immunol.* **182**, 696–707 (2009).
41. Wahba, L., Costantino, L., Tan, F. J., Zimmer, A. & Koshland, D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Gene Dev.* **30**, 1327–1338 (2016).
42. Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA: DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
43. Keller, W. & Crouch, R. Degradation of DNA RNA hybrids by ribonuclease H and DNA polymerases of cellular and viral origin. *Proc. Natl Acad. Sci. USA* **69**, 3360–3364 (1972).
44. Wang, P. J. X chromosomes, retrogenes and their role in male reproduction. *Trends Endocrinol. Metab.* **15**, 79–83 (2004).
45. Turner, J. M. Meiotic Silencing in Mammals. *Annu Rev. Genet* **49**, 395–412 (2015).
46. Osaki, E. et al. Identification of a novel Sry-related gene and its germ cell-specific expression. *Nucleic acids Res.* **27**, 2503–2510 (1999).
47. Dettori, L. G. et al. A tale of loops and tails: the role of intrinsically disordered protein regions in R-loop recognition and phase separation. *Front. Mol. Biosci.* **8**, 691694 (2021).
48. Li, Q. et al. Cryo-EM structure of R-loop monoclonal antibody S9.6 in recognizing RNA: DNA hybrids. *J. Genet Genomics* **49**, 677–680 (2022).
49. Li, Q. et al. DNA polymerase epsilon harmonizes topological states and R-loops formation to maintain genome integrity in Arabidopsis. *Nat. Commun.* **14**, 7763 (2023).
50. Jiang, Y. et al. Genome-wide map of R-loops reveals its interplay with transcription and genome integrity during germ cell meiosis. *Journal of advanced research* (2022).
51. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **4**, 442–451 (2003).
52. Tsai, A. G. et al. Conformational variants of duplex DNA correlated with cytosine-rich chromosomal fragile sites. *J. Biol. Chem.* **284**, 7157–7164 (2009).
53. Arab, K. et al. GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat. Genet.* **51**, 217–223 (2019).
54. Kaneko, S., Chu, C., Shatkin, A. J. & Manley, J. L. Human capping enzyme promotes formation of transcriptional R loops in vitro. *Proc. Natl Acad. Sci. USA* **104**, 17620–17625 (2007).
55. Cohen, S. et al. Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat. Commun.* **9**, 533 (2018).
56. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
57. Gonzalez-Aguilera, C. et al. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol. Biol. cell* **19**, 4310–4318 (2008).
58. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
59. Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
60. Wang, X. et al. C9orf72 and triplet repeat disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms. *RNA* **25**, 935–947 (2019).
61. Zardoni, L. et al. Elongating RNA polymerase II and RNA:DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. *Nucleic acids Res.* **49**, 12769–12784 (2021).
62. Zatreanu, D. et al. Elongation Factor TFIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. *Mol. Cell* **76**, 57–69 e59 (2019).
63. Soumillon, M. et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013).
64. Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).
65. Becherel, O. J. et al. Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *Plos Genet* **9**, e1003435 (2013).
66. Fujiwara, Y. et al. New allele of mouse DNA/RNA helicase senataxin causes meiotic arrest and infertility. *Reproduction* **166**, 437–450 (2023).
67. Liu, C. et al. Dual roles of R-loops in the formation and processing of programmed DNA double-strand breaks during meiosis. *Cell Biosci* **13** (2023).
68. Mullican, S. E. et al. Histone deacetylase 3 is an epigenomic brake in macrophage alternative activation. *Genes Dev.* **25**, 2480–2488 (2011).
69. Sadate-Ngatchou, P. I., Payne, C. J., Dearth, A. T. & Braun, R. E. Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. *Genesis* **46**, 738–742 (2008).
70. Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
71. Peters, A. H., Plug, A. W., van Vugt, M. J. & de Boer, P. A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res* **5**, 66–68 (1997).
72. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
73. Xu, W. et al. Quantitative, convenient, and efficient genome-wide R-loop profiling by ssDRIP-seq in multiple organisms. *Methods Mol. Biol.* **2528**, 445–464 (2022).
74. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).

75. Xu, Q. et al. Enhanced crosslinking immunoprecipitation (eCLIP) method for efficient identification of protein-bound RNA in Mouse testis. *J. Vis. Exp.* (2019).

## Acknowledgements

We thank Jeremy Wang from University of Pennsylvania for critical reading of the manuscript; Qin Li and Qianwen Sun from Tsinghua University for technique support on ssDRIP-seq assay. This work was supported by the National Key Research and Development Program of China (2022YFC2703500), National Natural Science Foundation of China Grant (32070843, 82371617), and Excellent Foundation of Jiang-shu Scientific Committee (BK20211532) to L.Y. and National Natural Science Foundation of China (31530047) to Z.B.H. Z.S. was supported by NIH (ES034768, AG069966).

## Author contributions

Z.K. and C.X. contributed equally and did most of the experiments in this study. Z.K., C.X., and J.G. generated *Nkapl* KO and *Nkapl*<sup>M349fs</sup> mice. Z.K., C.X., J.G., G.L., and Q.H. performed histology, immunofluorescence, nuclear spread assay, and analyzed phenotypes of *Nkapl* KO and *Nkapl*<sup>M349fs</sup> mice. Z.K. and C.X. prepared ChIP-seq libraries. C.X. performed most of immunoprecipitation assays, pulled down R-loops and performed ssDRIP-seq library construction. Z.K. performed eCLIP experiments. J.G., B.Q. and S.H. extracted RNA. R.Y., Y.W., J.Z., X.X., L.C., and M.L. analyzed RNA-seq, ChIP-seq, eCLIP-seq, and ssDRIP-seq data. S.L., Y.W., Y.G., and Z.H. performed *NKAPL* mutation analysis from patients with azoospermia. L.Y., Y.G., and Z.S. conceived the study, prepared and wrote the manuscript with input from everyone.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-55579-y>.

**Correspondence** and requests for materials should be addressed to Mingyan Lin, Zheng Sun, Yayun Gu or Lan Ye.

**Peer review information** *Nature Communications* thanks Chunsheng Han, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025