RESEARCH ARTICLE SUMMARY

REPRODUCTION

The landscape of RNA binding proteins in mammalian spermatogenesis

Yang Li†, Yuanyuan Wang†, Yue-Qiu Tan†, Qiuling Yue†, Yueshuai Guo†, Ruoyu Yan†, Lanlan Meng†, Huicong Zhai, Lingxiu Tong, Zihan Yuan, Wu Li, Cuicui Wang, Shenglin Han, Sen Ren, Yitong Yan, Weixu Wang, Lei Gao, Chen Tan, Tongyao Hu, Hao Zhang, Liya Liu, Pinglan Yang, Wanyin Jiang, Yiting Ye, Huanhuan Tan, Yanfeng Wang, Chenyu Lu, Xin Li, Jie Xie, Gege Yuan, Yiqiang Cui, Bin Shen, Cheng Wang, Yichun Guan, Wei Li, Qinghua Shi, Ge Lin, Ting Ni, Zheng Sun, Lan Ye, Anastasios Vourekas, Xuejiang Guo*, Mingyan Lin*, Ke Zheng*

INTRODUCTION: The mammalian testis transcribes the vast majority of the genome and thus harbors one of the most complex tissue transcriptomes. However, we lack a systematic understanding of how the expanding world of RNA binding proteins (RBPs) contributes to the dynamic regulation of the vast number of RNAs involved in spermatogenesis. In particular, there is an emerging research interest in the role of nondomain elements (short motifs found outside annotated domains) of RBPs, which have not been adequately explored with animal models thus far. Although malfunction of individual RBPs underlies the origin of many diseases, genome-wide association of such deficiencies with specific diseases including male infertility—a growing major reproductive problem worldwide—remains elusive.

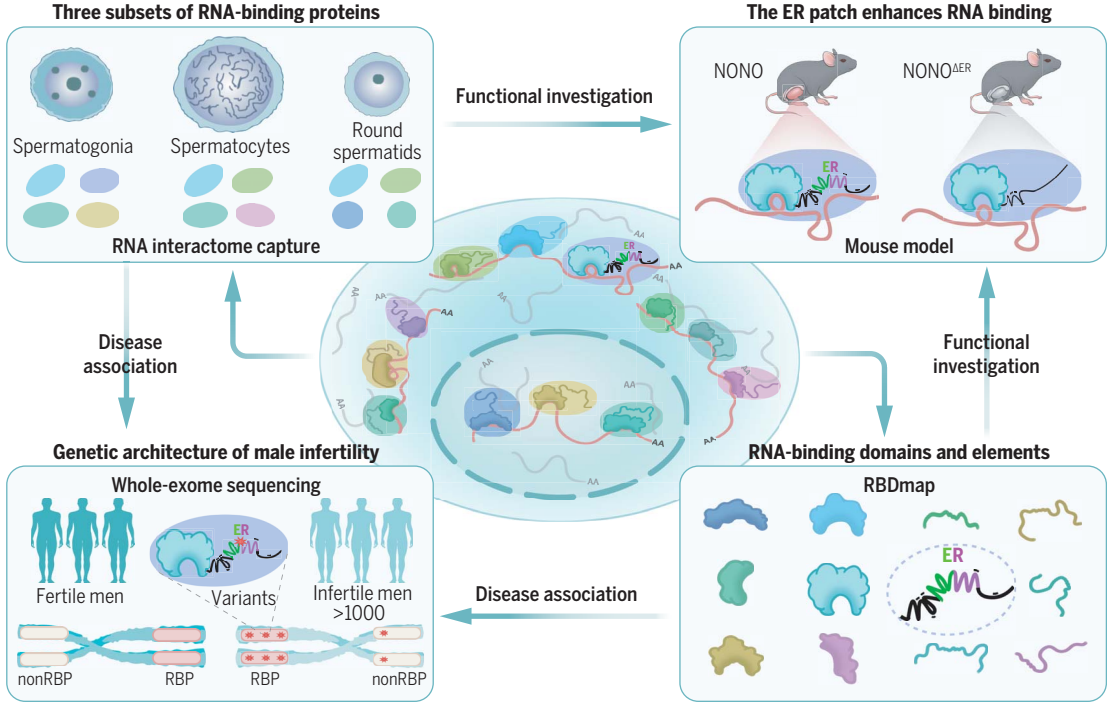
RATIONALE: To systematically identify RBPs in mouse male germ cells (mMGCs) at different stages, we isolated spermatogonia, pachytene spermatocytes, and round spermatids and performed RNA interactome capture. We further trapped testicular RNA–cross-linked peptides using RBDmap to screen for RNA binding domains and nondomain elements, followed by bioinformatic and experimental characterization of their prevalence, property, and function using orthogonal approaches. By integrating our large whole-exome sequencing profiles of male infertile cohort with over 1000 patients having nonobstructive azoospermia or oligozoospermia, we investigated the genetic contribution of RBPs to male infertility, shedding light on the clinical relevance of nondomain elements.

RESULTS: We established a comprehensive atlas of mMGC RBPs, including those with previously unknown RNA-related activities specialized for meiotic and postmeiotic functions. We showed that mMGC RBPs v
involved in multilayered dynamics cove
proteomic expression patterns, proteotranscriptomic discordancy, relative RNA binding activity, and ribonucleoprotein concordant behavior. We discovered a polyampholytic nondomain element, termed glutamic acid–arginine (ER) patch, that is positionally coupled with the coiled coils (CCs) of RBP, features evolutionary covariation of residue pairs, and enhances RNA binding of its host RBPs. We demonstrated that the ER patch in an RBP named NONO (non-POU domain-containing octamer-binding protein) regulates a mitosis-to-meiosis transition of male germ cells at single-cell resolution. Whole-exome sequencing analysis revealed a landscape of male infertility–associated variants in RBPs, including those in CCs and ER patches.

CONCLUSION: Our findings provide a resource for the germline RBPs relevant to male infertility and demonstrate the functional importance of a nondomain element in RNA binding and spermatogenesis, highlighting the potentially broad value of this resource for decoding the genetic and molecular basis of male fertility. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: guo_xuejiang@njmu.edu.cn (X.G.); linmingyan@njmu.edu.cn (M.L.); kezheng@njmu.edu.cn (K.Z.)
†These authors contributed equally to this work.
Cite this article as Y. Li *et al.*, *Science* **386**, eadj8172 (2024). DOI: 10.1126/science.adj8172

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adj8172



The landscape of RNA binding proteins in mammalian spermatogenesis and human infertility. RNA binding proteins (RBPs) dynamically interact with RNAs through various RNA binding domains and nondomain elements during spermatogenesis. The glutamic acid–arginine (ER) patch, a nondomain element existing in coiled coils and functioning as an RNA binding enhancer, promotes NONO-RNA interactions and secures proper male germ cells. RBPs, including their ER patch variants, are an important component of the genetic architecture of male infertility.

RESEARCH ARTICLE

REPRODUCTION

The landscape of RNA binding proteins in mammalian spermatogenesis

Yang Li[1]†, Yuanyuan Wang[1,2]†, Yue-Qiu Tan[3,4]†, Qiuling Yue[1,5]†, Yueshuai Guo[1]†, Ruoyu Yan[1,6]†, Lanlan Meng[3,4]†, Huicong Zhai[1], Lingxiu Tong[1], Zihan Yuan[1], Wu Li[1], Cuicui Wang[1], Shenglin Han[1], Sen Ren[1], Yitong Yan[2], Weixu Wang[7], Lei Gao[1], Chen Tan[3], Tongyao Hu[3], Hao Zhang[1], Liya Liu[2], Pinglan Yang[1], Wanyin Jiang[1], Yiting Ye[1], Huanhuan Tan[1], Yanfeng Wang[1], Chenyu Lu[1], Xin Li[1], Jie Xie[1], Gege Yuan[1], Yiqiang Cui[1], Bin Shen[1], Cheng Wang[1,8], Yichun Guan[9], Wei Li[10], Qinghua Shi[11], Ge Lin[3,4], Ting Ni[12], Zheng Sun[13], Lan Ye[1], Anastasios Vourekas[14], Xuejiang Guo[1]*, Mingyan Lin[2,15,16]*, Ke Zheng[1]*

Despite continuous expansion of the RNA binding protein (RBP) world, there is a lack of systematic understanding of RBPs in the mammalian testis, which harbors one of the most complex tissue transcriptomes. We adapted RNA interactome capture to mouse male germ cells, building an RBP atlas characterized by multiple layers of dynamics along spermatogenesis. Trapping of RNA–cross-linked peptides showed that the glutamic acid–arginine (ER) patch, a residue-coevolved polyampholytic element present in coiled coils, enhances RNA binding of its host RBPs. Deletion of this element in NONO (non-POU domain-containing octamer-binding protein) led to a defective mitosis-to-meiosis transition due to compromised NONO-RNA interactions. Whole-exome sequencing of over 1000 infertile men revealed a prominent role of RBPs in the human genetic architecture of male infertility and identified risk ER patch variants.

Mammalian spermatogenesis spans three prime phases: mitosis, meiosis, and spermiogenesis, during which germ cells differentiate in turn as spermatogonia, spermatocytes, and spermatids, ultimately producing spermatozoa. During this multistep, dynamic process, germ cells undergo drastic morphological alterations accompanied by a cascade of molecular events encompassing chromatin remodeling (*1*), transcriptional transition (*2*), and posttranscriptional regulation (*3*), that give rise to the remarkably complex testicular transcriptome (*4*, *5*). The testis expresses an abundance of RNA binding proteins (RBPs) (*6*), which are involved in processing, transporting, and modulating RNA, as well as coordinating transcription (*7*). Testis-specific RBPs orchestrate the storage and translational activation of spermiogenic mRNAs that are transcribed several days before they are needed because of the spermiogenic nucleus undergoing compaction and transcriptional silencing (*8*, *9*). Despite our growing knowledge of individual RBPs in spermatogenesis, a comprehensive atlas of authentic RBPs is still lacking.

RNA interactome capture (RIC) is a high-throughput RBP-profiling approach (*10*, *11*) that has advanced our understanding of structural and functional features of RBPs through its application in various cell lines and somatic tissues (*12*–*17*). However, capturing the RNA binding proteome (RBPome) in specific types of primary cells derived from a tissue is technically challenging. In the mammalian testis, germ cell–specific RIC is valuable for mining RBPs and studying their spermatogenic dynamics.

RBDmap and related methodologies allow systems-level detection of RNA–cross-linked peptides (*18*, *19*). These methods verify well-defined canonical RNA binding domains (RBDs) and reveal noncanonical RBDs and RNA binding nondomain elements (short motifs found outside annotated domains) especially from intrinsically disordered regions (IDRs). IDRs drive protein-protein and protein-RNA interactions independently or in cooperation with other globular RBDs (*12*, *20*). Although IDR elements are increasingly recognized to widely function in cellular processes and diseases, they are largely undefined, and only a few studies have used animal models to pinpoint their roles (*21*, *22*).

Human mutations in RBP-encoding genes are critical determinants for both Mendelian and somatic pathogenesis (*23*). Intersecting the known RBP repertoire with public disease association data showed heterogeneity of RBP mutation properties among different diseases (*23*). Male infertility is a growing major reproductive problem worldwide, and its etiology remains unknown in about 40% of patients, with genetic factors being recognized as crucial contributors (*24*). By far, scaled genomic sequencing of male infertile cohort has been largely restricted to patients with Western ancestry (*25*, *26*). Despite many efforts to decipher its monogenetic causes (*27*, *28*), our understanding of the genetic architecture underlying male infertility is rather limited (*29*), lagging behind that of other diseases.

In this study, we performed RIC-based proteome-wide mapping of endogenous RBPs in mouse male germ cells (mMGCs) and further delved into a systematic profiling of RBDs and RNA binding nondomain elements. Our approach led to the discovery of a polyampholytic nondomain element named glutamic acid–arginine (ER) patch, that is widely present in coiled-coil motifs of RBPs and promotes RNA binding activity of its host RBPs and male germ cell development. Intersecting these datasets with our whole-exome sequencing profiles from a large cohort of infertile men created an opportunity to explore the genetics of male infertility with a focus on and full coverage of RBPs.

Results

mMGC RIC identifies highly specialized, germline-specific RBPs

To systematically identify RBPs in mMGCs, we used STA-PUT velocity sedimentation to purify sufficient numbers of spermatogonia (SG), pachytene spermatocytes (pSC), and round spermatids (rST), which were subjected to a modified RIC procedure separately (Fig. 1A and fig. S1A). Several known RBPs were validated by Western blot assays, confirming that they were captured only when cross-linked using ultraviolet (UV) light (254 nm) in RIC (fig. S1B). Liquid chromatography–tandem mass spectrometry (LC-MS/MS) identified

[1]State Key Laboratory of Reproductive Medicine and Offspring Health, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [2]Department of Neurobiology, School of Basic Medical Science, Nanjing Medical University, Nanjing 211166, China. [3]Institute of Reproductive and Stem Cell Engineering, NHC Key Laboratory of Human Stem Cell and Reproductive Engineering, School of Basic Medical Science, Central South University, Changsha 410083, China. [4]Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha 410008, China. [5]Department of Andrology, Nanjing Drum Tower Hospital, the Affiliated Hospital of Nanjing University, Nanjing 210008, China. [6]College of Life Sciences, Northwest A&F University, Yangling 712100, China. [7]Institute of Computational Biology, Helmholtz Center Munich, Munich 85764, Germany. [8]Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China. [9]Center for Reproductive Medicine, the Third Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China. [10]Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China. [11]Division of Reproduction and Genetics, First Affiliated Hospital of USC, Hefei National Laboratory for Physical Sciences at Microscale, School of Basic Medical Sciences, Division of Life Sciences and Medicine, Biomedical Sciences and Health Laboratory of Anhui Province, University of Science and Technology of China, Hefei 230027, China. [12]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Human Phenome Institute, Shanghai Engineering Research Center of Industrial Microorganisms, School of Life Sciences and Huashan Hospital, Fudan University, Shanghai 200438, China. [13]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA. [14]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [15]Changzhou Medical Center, The Affiliated Changzhou Second People's Hospital of Nanjing Medical University, Changzhou 213000, China. [16]Division of Birth Cohort Study, Fujian Maternity and Child Health Hospital, Fuzhou 350014, China.
*Corresponding author. Email: guo_xuejiang@njmu.edu.cn (X.G.); linmingyan@njmu.edu.cn (M.L.); kezheng@njmu.edu.cn (K.Z.)
†These authors contributed equally to this work.

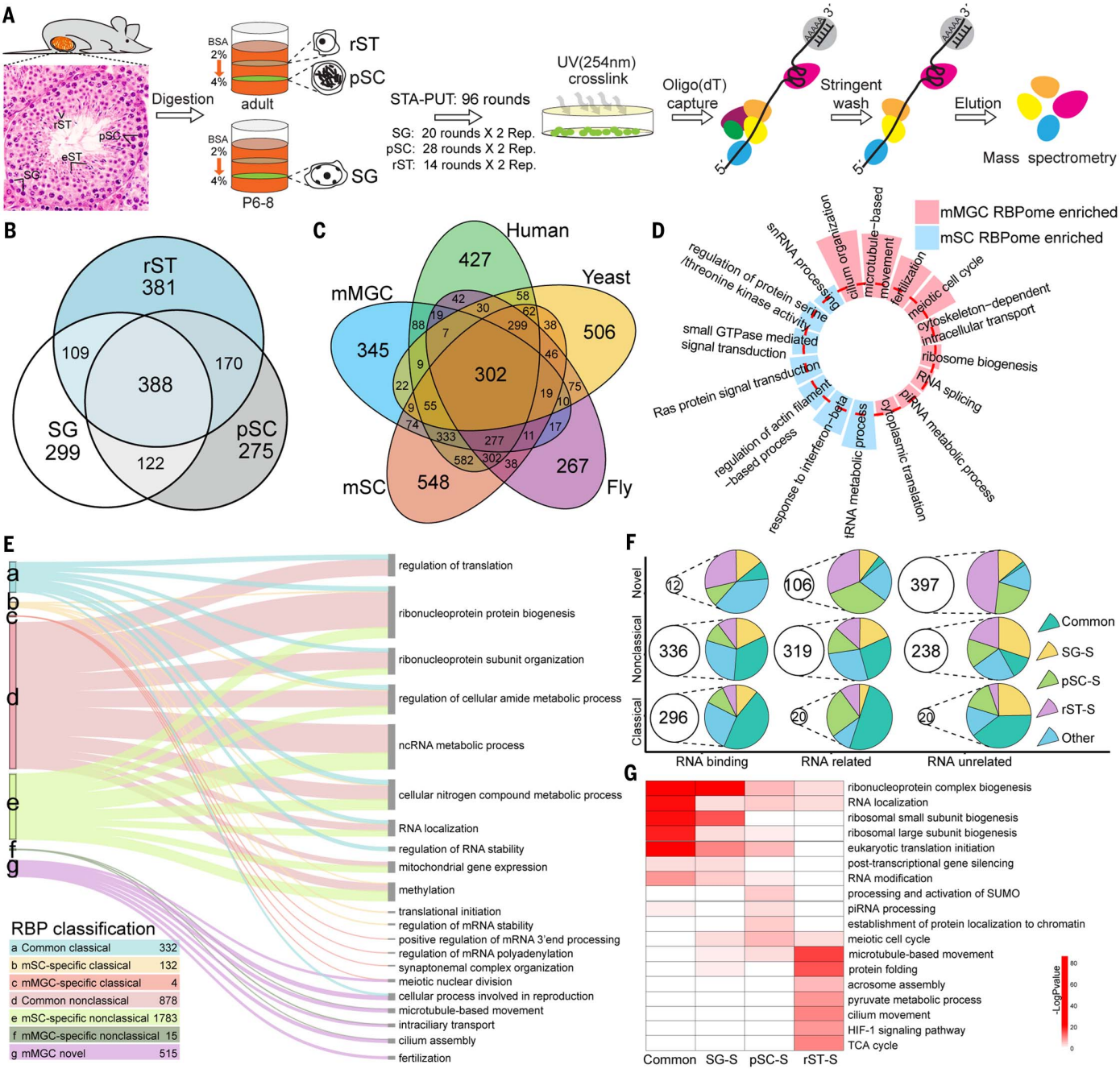


Fig. 1. RIC-based identification and characterization of mMGC RBPs. (**A**) Schematic of RIC for three types of germ cells (SG, pSC, and rST) purified from fresh mouse testes. SG, spermatogonia; pSC, pachytene spermatocyte; rST, round spermatid. STA-PUT yielded specific populations of SG with >85% purity, rST with >90% purity, and pSC with >85% purity. For each cell type, numbers of RIC replicates (Rep.) and STA-PUT rounds per replicate are indicated. eST, elongated spermatid; BSA, bovine serum albumin. (**B**) Venn diagram shows intersection of the three subsets of mMGC RBPs. (**C**) Venn diagram shows intersection of RBPomes from different species and origins. Proteins without mouse homologs are not included. mMGC (mouse male germ cell): RBPome in this study; mSC (mouse somatic cell), human, yeast, and fly: other reported RBPomes. (**D**) Circular bar plot shows the enriched GO term in mMGC RBPome and mSC RBPome. The bar heights are equal to the −log10 (*P* value) of each GO term. Red dashed ring denotes −log10 (*P* value = 0.05). snRNA, small nuclear RNA; GTPase, guanosine triphosphatase; piRNA, Piwi-interacting RNA. (**E**) Sankey plot shows the enriched biological processes GO terms of each RBP subset when intersecting mMGC RBPome with mSC RBPome. Inset table: RBP subsets on the basis of comparison between mMGC and mSC. Common: RBPs shared by mMGC and mSC. Numbers of proteins in each subset are indicated. ncRNA, noncoding RNA. (**F**) Balloon plot displays the protein categories of mMGC RBPome according to functional annotations (GO database) and RBP types. Protein number of each category is indicated; each category consists of five subsets. SG-S: RBPs specifically captured in SG; pSC-S: RBPs specifically captured in pSC; rST-S: RBPs specifically captured in rST; Common: RBPs shared by SG, pSC, and rST; Other: RBPs shared by two types of germ cells. (**G**) Heatmap shows the enriched GO and KEGG (Kyoto Encyclopedia of Genes and Genomes) terms across four RBP subsets generated from (F). SUMO, small ubiquitin-related modifier; TCA, tricarboxylic acid.

highly reproducible amounts of proteins between two biological replicates of RIC (fig. S1C). Under a stringent enrichment criterion of cross-link (CL)/non-cross-link (noCL) ≥ 10 for intensity-based absolute quantitation (iBAQ) in MS (fig. S1C), we defined 918 proteins from SG, 955 from pSC, and 1048 from rST (Fig. 1B and table S1), resulting in a total of 1744 nonredundant proteins referred to as "mMGC RBPome/RBPs," of which 388 proteins are present in all three cell types, and 299, 275, and 381 are specific to SG, pSC, and rST, respectively (Fig. 1B). As expected, Gene Ontology (GO) analysis of identified proteins displays their predominant involvement in RNA-related cellular processes and molecular functions (fig. S1D), and mMGC RBPome exhibits physicochemical properties similar to other somatic RBPomes (fig. S1E) (*10*), supporting the validity of our approach.

Taking into account only orthologous genes across species, 302 genes are common to our mMGC RBPome and other published RBPomes [data collected by (*30*)], whereas hundreds of proteins are dataset-specific (Fig. 1C), including 345 genes particular to our mMGC RBPome. At the protein level, comparative analysis of mMGC RBPome and combined mouse somatic cell (mSC) RBPomes (*30*) revealed distinct processes enriched in mMGC RBPome or mSC RBPomes (Fig. 1D). Excluding proteins with annotated RBDs (*10*, *30*) and those common to both mMGC RBPome and mSC RBPomes defined a group of 515 "mMGC novel" candidate RBPs, specifically enriched in spermatogenesis-associated processes such as reproduction, meiosis, and fertilization (Fig. 1, D and E, and table S1), and phenotypically related to male infertility and abnormal spermatogenesis (fig. S1F). A large proportion (n = 397) of these candidate RBPs lack RNA relevance (Fig. 1F, fig. S1G, and table S1). Relative to a minority (18.7%) of SG-specific RBPs, more (31.6%) of pSC-specific and even half (49.9%) of rST-specific RBPs were novel and not previously related to RNA (Fig. 1, F and G, and table S1).

To confirm that the identified candidate RBPs are bona fide factors that act on RNA, we applied the Protein-RNA Complex Capture (2C) technique (*31*, *32*), using single-cell suspension of whole adult testis and 293T cells (fig. S2A). UV cross-linking to capture ribonucleoprotein (RNP) in 2C was shown to be effective by silver staining (fig. S2A), and 16 mMGC RBPs were verified by Western blot (fig. S2, B to E). PDHA2 and TEX30, two 2C-validated novel RBPs, were orthogonally verified by cross-linking and immunoprecipitation (CLIP)–autoradiography (figs. S3A and S4A), enhanced CLIP sequencing (eCLIP-seq) (fig. S3, B to D; fig. S4, B to D; and table S2), and electrophoretic mobility shift assay (EMSA) (figs. S3E and S4E) and were both testis-specific and required for spermatogenesis (fig. S3, F to K, and fig. S4, F to L); these experiments are detailed in the supplementary text. Collectively, our results suggest that RIC faithfully captures testis-specific RBPs, which have likely evolved specialized features adapted to the complex meiotic and postmeiotic mechanisms of RNA regulation.

Temporal expression and RNA binding dynamics of mMGC RBPs

To delineate expression patterns of RBPs during spermatogenesis, we leveraged our previously published datasets of tandem mass tag (TMT)–based quantitative proteomes of whole-cell extract of four types of mMGCs [SG, pSC, rST, and elongated spermatids (eST)] (*33*) (fig. S5A) and clustered all proteins on the basis of their temporal expression correlation (Fig. 2A and table S3). Classical and nonclassical RBPs constitute the majority (80 to 94%) of RBPs that are primarily expressed in spermatogonia (Clusters 1 to 5), whereas factors that first appear in meiotic or postmeiotic stages are enriched (41 to 59%) for novel RBPs (Clusters 6 to 10) (Fig. 2B). Functional analysis revealed distinct roles of RBPs across clusters (Fig. 2C), in concert with the distinctive nature of the cellular processes in the four mMGC types. Protein-protein interaction (PPI) modules within several clusters (Clusters 2, 6, and 9) hint that multiple RBPs may jointly engage in certain pathways (fig. S5B and supplementary text).

Next, we tracked mRNA-protein concordance and discordance along spermatogenesis through integration of the TMT-MS data with RNA-seq data (*34*) (fig. S5A and table S4). Novel RBPs showed lower concordance than classical and nonclassical RBPs and than DNA binding proteins (DBPs) (fig. S5C). Moreover, rST-specific RBPs showed the lowest concordance among all subsets of mMGC RBPs and nonRBPs (fig. S5C). We then grouped genes with distinct mRNA-protein discordance patterns across four stages (SG, pSC, rST, and eST) into four clusters (Clusters A to D) (fig. S5D). Cluster B was enriched for known translation-delayed genes and chromatoid body genes (*35*, *36*) (fig. S5E and table S4). From the proportions of RBPs relative to nonRBPs, the Cluster B–type mRNA-protein discordance of RBPs is likely a consequence of widespread translational delay regulation (fig. S5F).

A recent study assessed relative RNA binding activity of RBPs by dividing RNA association degree (RIC) by protein abundance of whole-cell extract (WCE) and demonstrated that proteins in direct contact with RNA have a high RIC/WCE ratio, which was leveraged to model concordant behavior of RBPs as RNP subunits (*37*). Following this lead, we calculated relative RNA binding activity of 1218 RBPs in spermatogenesis (Fig. 2D and table S5). Classical RBPs showed consistently strong RNA binding, whereas novel RBPs exhibited overall lower but more volatile RNA binding that was inactive in SG but relatively active in pSC and rST (Fig. 2D). The latter could be explained by increased access to target RNAs (*38*) for novel RBPs because transcriptome evolves extensively with developmental transitions, and these RBPs primarily localize at germ cell–specialized entities where RNAs might be selectively associated and compartmentalized, especially those in rST (Fig. 2E and fig. S6, A and B).

We then classified our mMGC RBPs of different stages into three groups (Fig. 2F, table S5, and supplementary text): Strong [log2 (RIC/WCE) ≥ 2, log2(WCE) > 0], Weak [log2 (RIC/WCE) < −2], and Moderate [log2(RIC/WCE) < 2 and ≥ −2]. To test whether relative RNA binding activity analysis could reflect RNP assembly, MIWI-ELAVL1– associated RBPs (*9*) were examined, and these proteins exhibited a convergent rise of relative RNA binding activity across stages (pink module; fig. S6, C to E), a phenomenon compatible with the MIWI-ELAVL1 complex formation in rST (*9*). As expected, housekeeping ribosome proteins stably act as Strong RBPs across three stages (green module; fig. S6, F to H). We showed three additional modules of RNP assembly (supplementary text): Ribosome biogenesis–associated proteins (orange module; Fig. 2, F and G, and fig. S6I), mainly composed of nonclassical RBPs, emerged as a Strong module specifically in SG; conversely, some mRNA splicing and transport factors (red module; Fig. 2, F and G, and fig. S6J), mainly classical RBPs, were more active in pSC and rST than in SG; cilium movement and acrosome assembly proteins, mainly novel RBPs, exhibited a marked increase of relative RNA binding activity from SG to pSC and rST (blue module; Fig. 2, F and G, and fig. S6K). These examples suggest stage specificity of the RNP network in spermatogenesis.

RBDmap identifies testis-enriched RNA binding domains

According to annotations of RBDs from InterPro database, we found that canonical RBDs, such as the RNA recognition motif (RRM) and K homology (KH) domains, were overrepresented in mMGC-mSC common classical RBPs (fig. S7A and table S6). WD40 and Tudor domains frequently emerged in mMGC-mSC common nonclassical RBPs (fig. S7A). Domains enriched in mMGC novel RBPs generally lacked evidence for RNA binding, such as PKINASE and ANK2 domains (fig. S7A and supplementary text). These two domains were confirmed for their RNA binding capacity by 2C assay (fig. S7B). We also observed distinct sequence patterns: RS, polyQ, and polyG (*39*) were enriched in IDRs of classical and nonclassical RBPs, whereas the GPP motif was specifically presented in IDRs of mMGC novel RBPs (fig. S7C).

To experimentally profile RBDs, we conducted RBDmap (*18*, *19*) using adult and postnatal day 28 mouse testes separately. LC-MS/MS identified a total of 721 nonredundant peptides

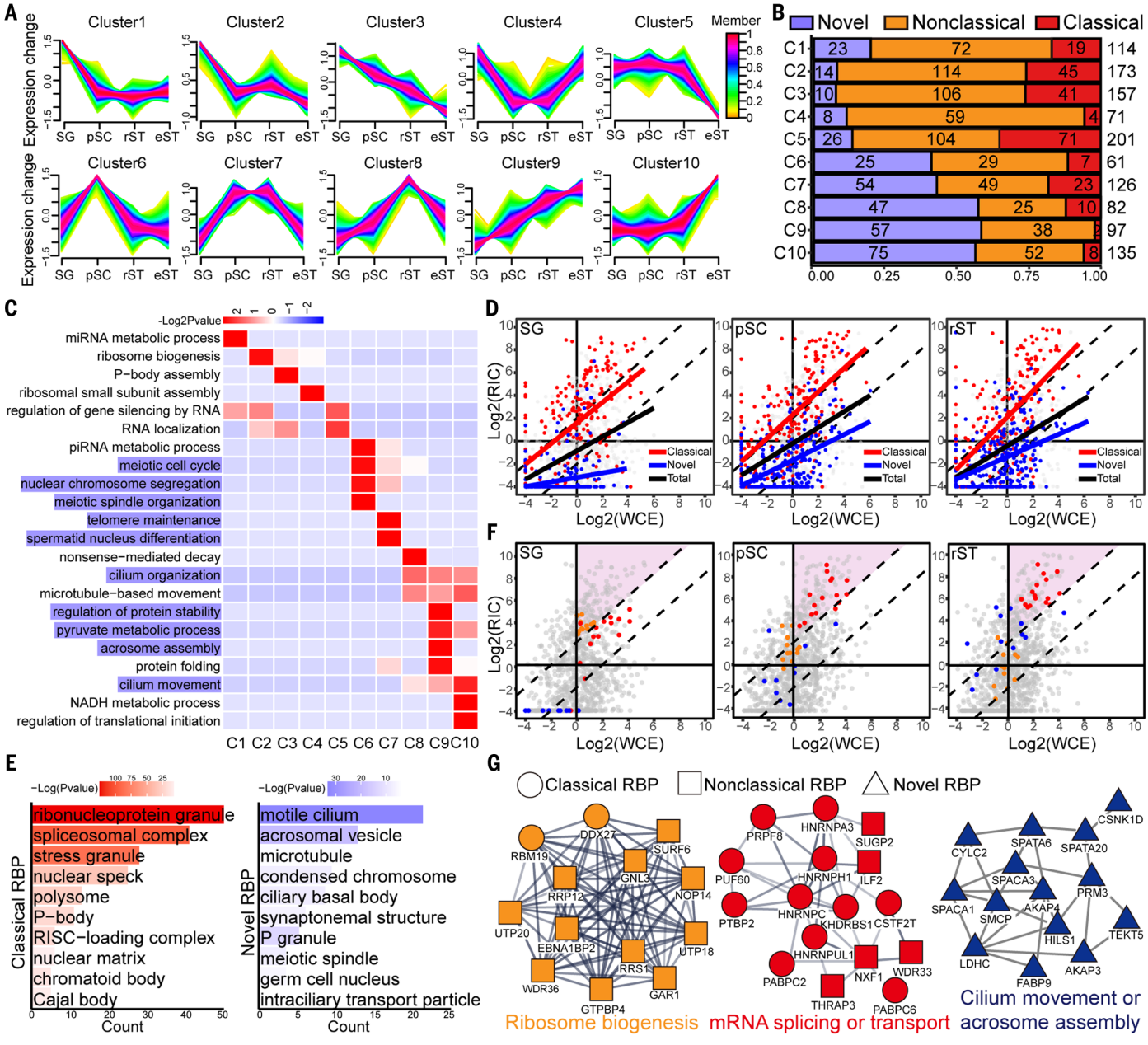


Fig. 2. Multihierarchical temporal dynamics of mMGC RBPs. (**A**) Ten distinct temporal patterns of protein abundance along spermatogenesis. Membership values are color-encoded with red shades denoting high membership values of proteins, whose temporal abundances are highly correlated with the cluster core, and green shades denoting low membership values of proteins. (**B**) The proportions and counts of each kind of RBP in each cluster are shown. (**C**) Heatmap shows the significantly enriched GO terms among 10 distinct temporal RBP clusters. GO terms with purple background harbor more than half of novel RBPs. (**D**) Scatter plots show dynamics of relative RNA binding activity of RBPs measured by log2(RIC/WCE). Red and blue dots denote classical RBPs and novel RBPs, respectively. Red, blue, and black lines show the linear relationship between RIC and WCE to model the relative RNA binding activity for classical RBPs, novel RBPs, and total RBPs, respectively. (**E**) Bar plot shows the subcellular localization of classical RBPs and novel RBPs generated by GO cell component analysis. Top 10 GO cell-component terms based on gene counts are shown. P-body, processing body; P granule, perinuclear granule. (**F**) Scatter plots show dynamics of relative RNA binding activity of RBPs measured by log2(RIC/WCE). Orange, red, and blue dots illustrate three representative cases of potentially coordinated RBPs with different extents of interstage dynamics. The pink triangle regions define the Strong RBPs with log2(RIC/WCE) ≥ 2 and log2(WCE) > 0 at given stages. (**G**) Network plots show protein-protein interaction (PPI) clusters of RBPs with colors corresponding to the scatter plots from (F). Classical, nonclassical, and novel RBPs are shown as circles, squares, and triangles, respectively.

(hereafter named RBDpeps) of RNA-bound regions of 393 RBPs under a Moderate criterion (Fig. 3A, fig. S8A, table S7, and supplementary text). A total of 511 RBDpeps overlapped with documented domains (domain RBDpeps), and 210 RBDpeps did not (nondomain RBDpeps) (Fig. 3B). The specificity of these RBDpeps was proven by their considerable overlapping with RRM domains of classical proteins such as DAZL (fig. S8, B and C, and supplementary text). As expected, both domain and nondomain RBDpeps were enriched with glycine (G) and arginine (R)

amino acids (fig. S8D), which are known hotspots of RNA interaction (*20*). Our RBDmap identified 89 RBPs shared with mouse HL-1 cells, in which 64 RBPs harbored overlapping RBDpeps (fig. S8E, left). Moreover, 307 and 155 domain RBDpeps (90% in total) mapped to classical and nonclassical RBPs, respectively (fig. S8E, right). Examples of these 462 RBDpeps included classical RBPs (YTHDF1 and PUM1) and nonclassical RBPs of the Tudor family (fig. S8, F and G; fig. S9, A and B; and supplementary text). We also identified 49 nonredundant domain RBDpeps mapped to 37 novel RBPs, of which 15 are testis-specific proteins (fig. S8E, right, and table S7). Five core histones and two linker histones were identified in both our mMGC RBPome (table S1) and testis RBDmap (table S7). Representative examples are testis-specific histone variants, such as HIST1H1T and H2AFB1 (Fig. 3, C and D, and fig. S10A). Additionally, six RBDpeps were found in the protamine-P2 domain of PRM2 and two in the TP2 domain of TNP2 (Fig. 3D). Additional instances of RBDpeps overlaying domains of novel RBPs were illustrated (fig. S10, B and C, and supplementary text). Four of these testis-specific RBDs were 2C-validated (fig. S10D), confirming their bona fide RNA binding capacity.

RNA binding elements identified in nondomain areas

Based on their physicochemical properties, motifs within the 210 nondomain RBDpeps (Fig. 3B) could be clustered into three major archetypes (*40*): polar tracts, polyelectrolytes, and polyampholytes (Fig. 3E and supplementary text). One polyampholytic RNA binding element identified herein is the ER patch that features a mixed-charged sequence composed of E and R amino acids (Fig. 3E and table S8). Furthermore, nondomain RBDpeps are much more enriched in alpha helices forming coiled coils (CCs) (n = 23; all CCs are annotated on the basis of the Pfam database here and after) than in disordered regions and low-complexity regions [Fisher's exact test; CCs: odds ratio (OR) = 9.62, $P = 1.1 \times 10^{-4}$; disordered regions: OR = 2.43, $P = 1.1 \times 10^{-3}$; low-complexity regions: OR = 2.46, $P = 9.8 \times 10^{-4}$] (fig. S11A). All the ER hits (n = 4) are evolutionarily conserved and located in CCs of NONO, SFPQ (splicing factor proline- and glutamine-rich), and SAFB (scaffold attachment factor B) (Pfam database), as well as PSPC1 (paraspeckle component 1) [suggested by (*41*)] (Fig. 3F). An additional conserved ER patch located in the CC of RBP (RBP-CC) was also identified in two somatic RBDmap datasets (*18*, *42*) (Fig. 3F). Because these five ER-containing CCs overlapped with RBDpeps, we speculated that they themselves have RNA binding capacity. We tested two of them by 2C assay and found that both indeed bound RNA when overexpressed in 293T cells (fig. S11B). We then used EMSA to further examine NONO CC and found that it can bind RNA substrates with various sequences, preferentially polyA and G-rich RNAs (fig. S11C). Collectively, RBDmap identified these RNA binding CCs harboring ER patch elements.

The ER patch is a frequent element coevolving in local in RBP-CCs

We explored the prevalence of the ER patch at a genome-wide level, especially those in CCs, because CCs were more widespread in mMGC RBPs than in nonRBPs and DBPs (Fig. 3G). Consensus sequence analysis further revealed a specific enrichment of the ER patch in mMGC RBP-CCs but not in nonRBP-CCs (fig. S12A). We extended the pool of ER patch by searching for any 10-mer peptide containing at least four Glu and four Arg (4E4R) or three Glu and three Arg (3E3R) throughout mMGC RBPome, DBPome, and nonRBPome (table S9). Both CC-3E3R and CC-4E4R were more frequently present in RBPs than in DBPs and nonRBPs (fig. S12B). The in silico–identified ER patches (4E4R) were classified into three groups (fig. S12B and table S9): The first group has 21 ER patches in nonCCs of 19 RBPs; the second group has 92 ER patches in CCs of 74 nonRBPs; the third group has 51 ER patches localized in CCs of 34 RBPs, in which four of these ER-containing CCs were capable of binding RNA (fig. S11B). Some RBPs restricted to somatic RBPomes also contain 4E4R patches (table S9), suggesting a potentially wider spread of the ER patch in RBPomes.

To infer potential structural and functional interaction between ER patches and other elements of the same proteins, we conducted sequence covariation analysis commonly used to predict inter-residue contacts and distance or functional relevance (*43*–*45*). We found no long-distance contacts. Instead, inter-residue interdependence was confined within ER patches (Fig. 3H). Amino acids in the ER patch of RBP-CC (RBP-CC-ER) were predicted by EVcouplings (*46*) to be evolutionarily more covariated than the other two groups (t test; RBP-CC-4E4R versus RBP-nonCC-4E4R: P = 0.046; RBP-CC-4E4R versus nonRBP-CC-4E4R: P = 0.027) (Fig. 3, H and I, and fig. S12C), suggesting that a mutual interdependent relation (*45*) of E and R pairing within the RBP-CC-ER is critical for maintaining function. For example, residues of LUC7L3 were intensively covariated within amino acids 238 to 297, where ER patches are located (Fig. 3H). Similar local evolutionary covariation was also observed in the ER patches of INCENP, UPF3B, and NONO (Fig. 3H). These results hint at the functional relevance of the ER patch from a coevolutionary perspective.

The ER patch can enhance RNA binding of its host RBP

Next, we asked whether the ER patch has a role for its host protein in RNA binding. To address this, we deleted the sequence EEERRRRE in NONO ($NONO^{\Delta ER}$) and conducted eCLIP-seq to compare RNA targets of NONO and $NONO^{\Delta ER}$ proteins in 3T3-L1 cells. $NONO^{\Delta ER}$ showed a decrease in the number of RNA binding peaks and genes and also a slight change in target RNA profiles (Fig. 4, A and B, and fig. S13, A to C). We also observed a global decrease in binding intensity for ~75% of target genes (Fig. 4C). To directly visualize radiolabeled protein-bound RNA levels, we used conventional CLIP and found that the RNA binding capacity of $NONO^{\Delta ER}$ was much weaker (Fig. 4D). This result was reinforced by in cellulo quantitative-2C (q-2C) assays comparing overexpressed NONO and $NONO^{\Delta ER}$ and in vitro q-2C comparing purified recombinant NONO (rNONO) and $NONO^{\Delta ER}$ ($rNONO^{\Delta ER}$) proteins as well as messenger RNP (mRNP) capture assay in 293T cells (Fig. 4, E and F, and fig. S13, D and E). Similar impairment of RNA binding capacity was also found by conventional CLIP for $PSPC1^{\Delta ER}$ (fig. S13F), indicating that the ER patch enhances RNA binding of these two RBPs.

To further assess whether E or R amino acid contributes to enhanced RNA binding, we produced a panel of NONO-CC-ER mutants with either substitution or deletion of certain amino acids on the ER patch (Fig. 4E). $NONO^{AA}$, $NONO^{EA}$, and $NONO^{RA}$ [substitution of all E and/or all R by alanine (A)] all exhibited impaired RNA binding (Fig. 4F and fig. S13D). Nonetheless, $NONO^{RE}$ (replacement of EEERRRRE by ERER-ERER) did not alter RNA binding capacity of NONO (Fig. 4F and fig. S13D). These data demonstrate that both E and R amino acids in the ER patch are functionally indispensable.

Phase separation drives intermolecular interaction–mediated NONO recruitment for granule assembly (*47*), for which we reasoned that the diminished RNA binding of the NONO mutants might be coupled with a negative effect on their phase-separated condensate properties. We therefore examined these green fluorescent protein (GFP)–fused NONO proteins in 293T cells. In cells expressing high levels of GFP-NONO, ER deletion or substitution induced a conversion of NONO from a droplet state to a fiber-like state (Fig. 4, G and H). In cells expressing GFP-NONO at low levels as endogenous NONO, the foci of the mutant proteins remained in a droplet state and these foci were examined for fluorescence recovery after photobleaching (FRAP) (Fig. 4I). We found that the ER-mutated NONO—except $NONO^{RE}$—exhibited a concordant reduction of the mobility of NONO foci (Fig. 4, I to K). Nonetheless, the protein interactome of NONO versus $NONO^{\Delta ER}$ in 293T cells and in mouse testes (see next paragraph) showed that ER patch deletion in NONO did not affect its major partners SFPQ and PSPC1 (fig. S14, A to D, and table S10) nor its colocalization with PSPC1 (fig. S14E).

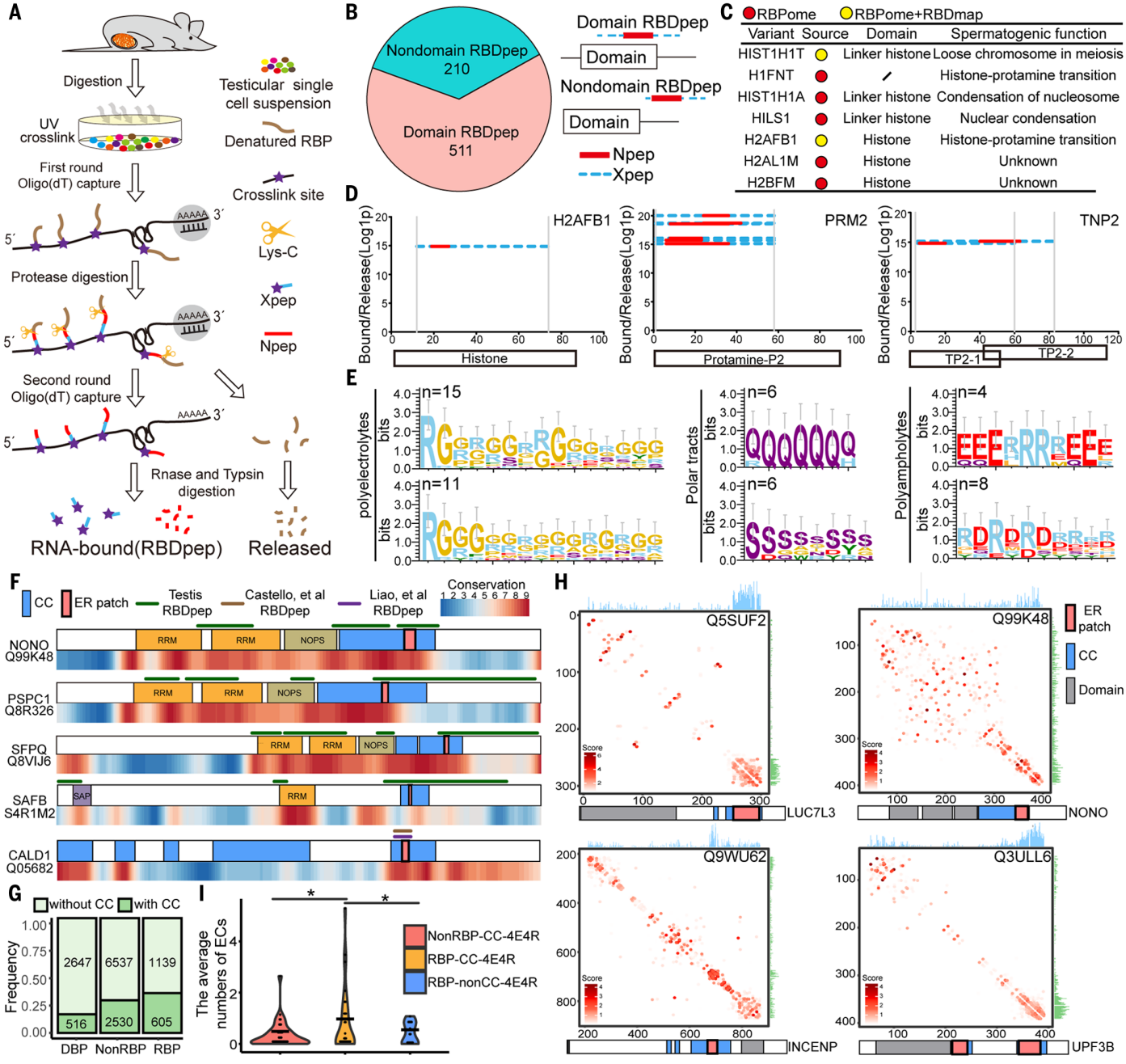


Fig. 3. RBDmap-based discovery of RBDs and RNA binding nondomain elements. (**A**) Schematic of RBDmap using fresh mouse testicular cells. Trypsin digestion of RNA-bound peptides generated two kinds of subpeptides: Npep and Xpep, which combined together are termed RBDpep. (**B**) Pie plot shows the proportion of domain and nondomain RBDpep in the testis RBDmap dataset. Xpep (blue dashed line) and Npep (red line) constitute RBDpep. Domain RBDpeps are defined overlapping with Pfam domain. Nondomain RBDpeps are defined as no amino acid overlapping with any domain. (**C**) List shows the testis-specific histone variants identified in mMGC RBPome and/or testis RBDmap dataset. (**D**) Graphs show all RBDpeps mapping to domain regions of H2AFB1, PRM2, and TNP2 proteins. The *x* axis shows RBDpeps (linear sequences) mapped to Pfam domains at relative positions of RBPs; *y* axis shows the enrichment (log1p) of RNA bound over released fraction for RBDpeps. Blue dashed line: Xpep; red line: Npep. (**E**) Sequence motifs within nondomain RBDpeps. Multiple sequence alignment of short patches within nondomain RBDpeps used Clustal Omega tool. The sequence logos (motif) were extracted from aligned nondomain fragments. All the motifs are classified to three archetypes according to the physiochemical characteristics of amino acids: polyelectrolytes, polar tracts, and polyampholytes. (**F**) Schematic representation of RBDpeps mapping to domain and nondomain areas of five ER-containing RBPs. All these ER patches identified by RBDmap are conserved and within coiled-coil motifs. The locations of our testis RBDpeps (green lines) and other somatic RBDpeps (brown and purple lines) are indicated. Coiled coils (CCs) and ER patches are shown by blue and pink rectangles, respectively. Conservation score and domain constitution are shown in the bottom color bars. Heatmap shows the evolutionary conservation scores of each residue on proteins from ConSurf software. The color of the heatmap indicates high (dark red) or low (blue) evolutionary conservation scores of each residue. (**G**) Bar plot shows the percentage of proteins with at least one CC motif in total proteins classified by its nucleic acid–binding characteristics. Protein list of DBP was generated from CIS-BP database. RBP: mMGC RBPs; nonRBP: proteins identified by TMT-MS but not by mMGC RIC. (**H**) EVcoupling's contact maps of several RBPs with coevolved

CC-ER patches. The coevolution scores are color-encoded with a continuous palette of red color. The bar plot on the top *x* axis (blue) displays the sums of the coevolution scores of each residue with the rest within each RBP, and the bar plot on the right *y* axis (green) displays the number of strongly evolutionarily coupled residue pairs (ECs) that each residue forms with the rest. (**I**) Box plot shows the average numbers of ECs within 4E4R patches in different protein groups. *$P < 0.05$, by *t* test.

The ER patch in RBP-CCs is functional during spermatogenesis

Nono and *Safb* are the two RBDmap-identified RBP genes that not only harbor ER patch sequences but also have knockout-based evidence demonstrating their physiological requirement in the testis. *Nono* knockout mice exhibited a decrease in body weight, testis weight, and sperm counts (*48*). *Safb* knockout mice exhibited male infertility with postnatal germ cell deficiency (*49*). To further assess the biological relevance of the ER patch, we created two knock-in mouse mutants with the ER patch precisely deleted from CCs of *Nono* and *Safb* genes (Fig. 5A and fig. S15A). ER patch–truncated RNA and protein in both $Nono^{\Delta ER/Y}$ and $Safb^{\Delta ER/\Delta ER}$ mice were expressed at similar levels compared with wild-type littermates (Fig. 5B and fig. S15, B and C). $Nono^{\Delta ER/Y}$ mice exhibited reduced body weights, sperm counts, and sperm motility, as well as increased numbers of apoptotic germ cells (Fig. 5C). Overall, in several measurements focusing on the testis, $Nono^{\Delta ER/Y}$ mice exhibited phenotypic defects resembling—albeit quantitatively less severe than—*Nono* knockout (*48*) (supplementary text). *Safb* knock-in mice also exhibited defects in germ cell development, with a milder phenotype than the *Safb* knockout (*49*). Whereas $Safb^{\Delta ER/\Delta ER}$ mice exhibited normal testis weight, sperm count, and sperm motility, there was an increase in the number of apoptotic germ cells and abnormal elongated spermatids in the testis (fig. S15, D and E). $SAFB^{\Delta ER}$ showed a trend for reduced RNA binding capacity in 293T cells, which, however, did not reach statistical significance (fig. S15F), possibly because SAFB harbors additional ER residues flanking the deleted ER patch (supplementary text). These results indicate that the ER patch should be indispensable for its host RBPs to fully function in spermatogenesis.

Coimmunostaining NONO with SOX9 (Sertoli cell marker) or with LIN28A [SG marker (*50*)] showed that NONO protein is predominantly expressed in Sertoli cells and spermatogonia, and its signal was detectable in spermatocytes (Fig. 5D and fig. S16). $Nono^{\Delta ER/Y}$ mice displayed increased numbers of Sertoli cells (differentiating to mitotic arrest in adulthood) and spermatogonia but decreased spermatocytes and spermatids (Fig. 5E and fig. S17, A to C). The partial loss of leptotene spermatocytes was observed at stage IX to X seminiferous tubules in $Nono^{\Delta ER/Y}$ mice (fig. S17D). Spermatocyte chromosome spread analysis showed a normal progression of meiotic prophase after leptotene stage (fig. S17E). These results suggest an abnormal expansion of Sertoli cell and spermatogonia populations and a subsequent defect in cell transition into meiotic leptotene spermatocytes.

The ER patch enhances endogenous NONO-RNA interactions and regulates mitotic transcripts ensuring germ cell transition into meiosis

Repressed RNA binding activity of NONO protein was also found in $Nono^{\Delta ER/Y}$ testes (Fig. 5F). We then conducted eCLIP-seq to assess transcriptome-wide alteration of RNA binding profiles of $NONO^{\Delta ER}$ proteins in the testis. Although $NONO^{\Delta ER}$ shared most of the bound targets with wild-type NONO in the testis (Fig. 5G and fig. S18, A to C), its binding capacity was compromised in that ~70% of bound targets showed decreased binding intensity in $NONO^{\Delta ER}$ testes (Fig. 5H and table S11), which is consistent with the aforementioned findings in cell lines (Fig. 4, A to C), corroborating the role of the NONO ER patch as an RNA binding enhancer.

To further investigate whether the diminished RNA binding capacity of $NONO^{\Delta ER}$ alters cellular gene expression across the spermatogenic lineage, we performed single-cell RNA sequencing (scRNA-seq) on testicular cells of adult $Nono^{+/Y}$ and $Nono^{\Delta ER/Y}$ mice. Cell types identified in our study corresponded well with previous testicular scRNA-seq datasets (*51*, *52*) (Fig. 5I and fig. S18, D and E). The higher ratio of mitotic germ cells (type A spermatogonia, type B spermatogonia, and preleptotene spermatocyte) to meiotic ones (leptotene, zygotene, pachytene, and diplotene spermatocyte) in $Nono^{\Delta ER/Y}$ versus $Nono^{+/Y}$ single testicular cells (fig. S18D) is in keeping with our cytological data (Fig. 5E and fig. S17, B to D), confirming the defect at the mitosis-to-meiotic transition in $Nono^{\Delta ER/Y}$ mice.

To ascertain whether and how NONO binds and regulates RNAs that are responsible for this transition (*53*), we intersected scRNA-seq datasets with eCLIP-seq profiles to dissect RBP-mediated cell state–specific RNA regulation (*54*). NONO CLIP targets are predominantly expressed in mitotic germ cells and Sertoli cells (Fig. 5J, top), the two cell types in which NONO is mainly expressed (Fig. 5D and fig. S16), supporting the specificity of our CLIP data. Furthermore, most CLIP targets are up-regulated specifically in these two cell types in $Nono^{\Delta ER/Y}$ mice (Fig. 5J, bottom, and table S11). These results suggest the cell type–specific target RNA binding and regulation mediated by the NONO ER patch in the testis.

Given that NONO tends to bind and up-regulate RNAs of transcription factors at mitotic phase in $Nono^{\Delta ER/Y}$ germ cells (Fig. 5J and fig. S18A), we assessed the transcriptome alterations between $Nono^{\Delta ER/Y}$ and $Nono^{+/Y}$ globally. We defined two gene sets with their expression levels enriched separately in the mitotic and meiotic phases on the basis of our wild-type scRNA-seq data (table S12). NONO bound 40.9% (247/604) mitosis-enriched genes and 2.1% (4/191) meiosis-enriched genes (fig. S19A), indicating that NONO regulates mitosis-enriched genes more directly. $Nono^{\Delta ER/Y}$ mitotic germ cells, but not meiotic germ cells, showed overall up-regulation of mitosis-enriched genes and overall down-regulation of meiosis-enriched genes compared with $Nono^{+/Y}$ (Fig. 5K and table S12). Among six representative mitosis-enriched genes illustrated by trajectory analysis were five NONO direct targets, and deletion of the NONO ER patch increased their expression (fig. S19, B and C, and supplementary text). Immunofluorescence analysis further revealed that the protein levels of two NONO targets were increased (fig. S19D). These results point to a role of the NONO ER patch in programming the mitosis-to-meiosis transition at the RNA level.

Homozygous RBP variants prioritized for human infertility

We performed a large whole-exome sequencing study of male infertility, including samples from 516 infertile men with idiopathic nonobstructive azoospermia (NOA) and 530 infertile men with oligozoospermia that passed our quality control procedures (table S13). An additional 300 fertile men with normal childbearing history were enrolled as controls (table S13). Population genomics analysis confirmed Asian origin of our cohort and no notable difference in population structure between patients with male infertility and controls (fig. S20). We first screened for loss of function (LOF) and damaging nonsynonymous (DNS) variants, and then applied a bioinformatic pipeline to identify potential pathogenic genes and prioritize the genetic components (types of genes, variants, and elements) more relevant to the pathogenesis of infertility (fig. S21). Both homozygous LOF (Fisher's exact test; OR = 2.52, $P = 1.2 \times 10^{-4}$) and DNS (Fisher's exact test; OR = 1.24, $P = 3 \times 10^{-5}$) variants were significantly enriched in patients (Fig. 6A), indicating that homozygous variants are more accountable for the pathogenesis of infertility.

Further, we observed significant overrepresentation of homozygous DNS in RBPs in patients (Fisher's exact test; OR = 1.62, $P = 0.0484$), suggesting a prominent role of RBPs in the genetic architecture of male infertility (Fig. 6A),

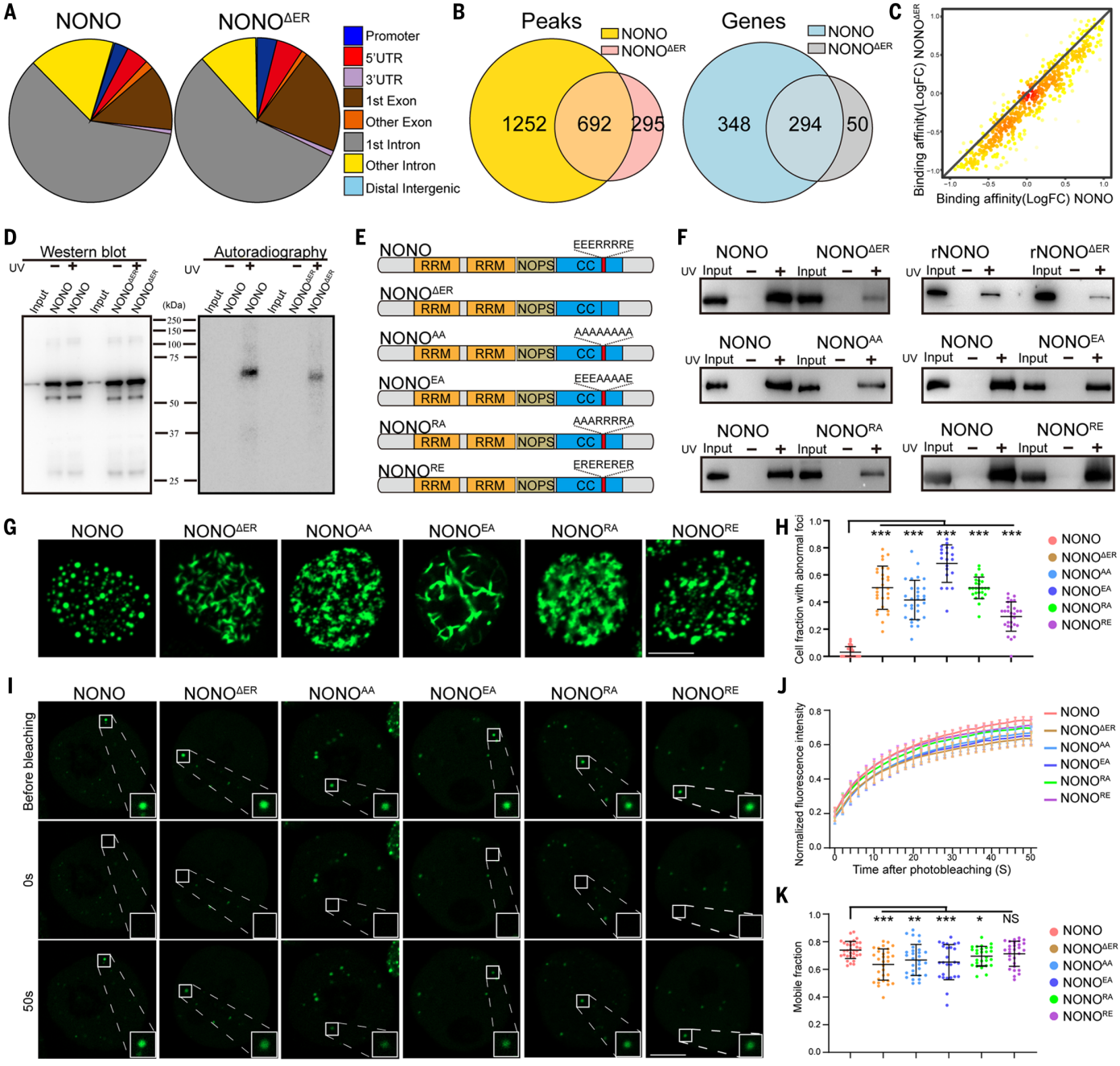


Fig. 4. The NONO ER patch functions in enhancing RNA binding activity. (**A**) Pie plots show the genomic distribution of all reproducible eCLIP-seq peaks for hemagglutinin (HA)-NONO-bound RNA and HA-NONO$^{\Delta ER}$-bound RNA in 3T3-L1 cells. (**B**) Venn diagrams show the overlaps of enriched eCLIP-seq binding peaks (left) and target genes (right) between NONO and NONO$^{\Delta ER}$ in 3T3-L1 cells. (**C**) Scatter plot compares the relative intensity of each target gene in NONO and NONO$^{\Delta ER}$ eCLIP-seq in 3T3-L1 cells. Each point shows log fold change (FC) between immunoprecipitation (IP) and input in NONO$^{\Delta ER}$ (*y* axis) and NONO (*x* axis) of one gene. (**D**) The RNA binding capacity of NONO and NONO$^{\Delta ER}$ protein was measured by radioactive CLIP assays. NONO- and NONO$^{\Delta ER}$-bound RNAs were radiolabeled by ^{32}P. Non-cross-linked samples served as negative controls. (**E**) Schematic diagram of Pfam domains and nondomain areas of wild-type and mutant NONO proteins. ER patches in the CC of NONO protein are shown as red rectangles. RRM, RNA recognition motif; NOPS, NONA/para-speckle. (**F**) Quantitative 2C (q-2C) assays compare the RNA binding capacity of wild-type and mutant NONO proteins. Total RNPs from 293T cells transfected with HA-NONO, HA-NONO$^{\Delta ER}$, HA-NONO^{RA}, HA-NONO^{EA}, HA-NONO^{AA}, and HA-NONO^{RE} were isolated by silica-based matrices after UV cross-link (+) or not (−). Total extracts (1% of input) and purified mRNPs were detected by Western blot with anti-HA antibody. In vitro q-2C assays also compare the RNA binding capacity of wild-type and mutant NONO proteins. Purified recombinant NONO (rNONO) and NONO$^{\Delta ER}$ (rNONO$^{\Delta ER}$) proteins were UV cross-linked (+) or not (−) with total RNA extracted from 293T cells. Protein-RNA complexes were isolated by silica-based matrices. One percent of input proteins and purified RNPs were detected by Western blot with anti-NONO antibody. Each experiment was repeated three times, and the quantitative measurement of each assay is shown in fig. S13D. (**G** and **H**) Representative images of foci morphology of overexpressed wild-type and mutant NONO proteins in 293T cells are shown in (G). The percentage of the cells with fiber-like foci is shown in (H). Only cells with dozens of bright foci, either orbicular or fiber-like, reflective of high

expression levels of NONO, were considered. ***$P < 0.001$, by t test. Scale bar in (G), 10 μm. (**I** to **K**) FRAP experiments in 293T cells. Only small foci, reflective of low expression levels of NONO, were picked to do FRAP. (I) Representative images show fluorescence recovery of wild-type and mutant NONO droplets before and after photobleaching a spot of predefined size within the nuclear region outlined in rectangle. (J) The FRAP fluorescence recovery curves record the normalized fluorescence intensity of wild-type and mutant NONO droplets 0 to 50 s after photobleaching. Mean with 95% confidence interval is shown. (K) Quantification of the mobile fraction of wild-type and mutant NONO droplets at 50 s after photobleaching. Each droplet was picked from one cell. $n = 33$ for NONO, $n = 31$ for $NONO^{\Delta ER}$, $n = 30$ for $NONO^{AA}$, $n = 26$ for $NONO^{EA}$, $n = 27$ for $NONO^{RA}$, $n = 29$ for $NONO^{RE}$; mean ± SD is shown; NS, not significant ($P > 0.05$), *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, by t test. Scale bar in (I), 10 μm.

although there was no significant difference in the odds of having homozygous LOF in RBPs between patients and controls (Fisher's exact test; OR = 2.24, $P = 0.3590$) (Fig. 6A). Using a stringent criterion (fig. S21), we compiled a list of candidate RBPs with homozygous variants exclusively in patients, leading to 22 LOF variants (candidate LOF) in 18 genes and 137 DNS variants (candidate DNS) in 105 genes, with their expression specificities, categorical and functional annotations, and patient origin in terms of infertility subtypes plotted (Fig. 6B, table S14, and supplementary text). Sanger sequencing confirmed 156 out of 159 total variants (~98%) (table S14). This genetic landscape of RBP variants associated with male infertility led to four observations (Fig. 6B): (i) Homozygous LOF variants were frequent in novel RBPs; (ii) homozygous DNS variants were overrepresented in nonclassical RBPs that are mostly ubiquitously expressed; (iii) variants were enriched in genes involved in four pathways (supplementary text), which were also overrepresented in mMGC RBPome relative to mSC RBPome (Fig. 1D); and (iv) variants were in an array of genes with known (above moderate-level evidence) and unknown (limited or no evidence) pathogenicity for human infertility, according to the online IMIGC (International Male Infertility Genomics Consortium) database (http://www.imigc.org/data-sharing/).

RBPs with heterozygous damaging variants tend to be mutation-intolerant

Sporadic studies have revealed pathogenic effects of individual heterozygous RBP variants in human infertility (*55*). At the genome-wide scale, we next examined whether heterozygous variants could contribute to the pathogenesis by prioritizing human disease genes on the basis of fitness consequences of disrupting variants, notably pLI (the probability of being LOF-intolerant) and z scores (the probability of being missense-intolerant) (*56*). We found that candidate genes (RBPs and nonRBPs) with patient-exclusive heterozygous LOF and DNS variants are more prone to be mutation-intolerant than genes with heterozygous LOF and DNS variants in controls (Wilcoxon test; $P = 2.21 \times 10^{-8}$ for RBPs with LOF; $P = 8.95 \times 10^{-19}$ for nonRBPs with LOF; $P = 3.27 \times 10^{-5}$ for RBPs with DNS; $P = 2.03 \times 10^{-8}$ for nonRBPs with DNS) (fig. S22, A and B). Moreover, candidate RBPs are more prone to be mutation-intolerant than candidate nonRBPs (Wilcoxon test; $P = 5.53 \times 10^{-6}$ for LOF; $P = 7.75 \times 10^{-8}$ for DNS) (fig. S22, A and B). These observations suggest that heterozygous damaging variants in RBPs are more likely to have a pathogenic impact for human infertility. In addition, candidate RBPs with heterozygous LOF or with heterozygous DNS are enriched for RNA splicing, translation, and cytoskeleton-associated functions (fig. S22, C and D, and table S15), which is in concert with mMGC RBPome (Fig. 1D) and homozygous candidate gene analysis (Fig. 6B), reflecting a convergence of the biological and etiological roles of RBPs underlying infertility.

Homozygous DNS variants identified in RBP-CCs and RBP-CC-ERs

Widespread DNS variants in nondomain areas including IDRs were shown in our (Fig. 6B and supplementary text) and others' (*23*) work, although they were not significantly overrepresented in patients (Fisher's exact test; OR = 0.96, $P = 0.3767$) (Fig. 6A). Nevertheless, we found a significant enrichment of DNS variants within RBP-CCs in patients (Fisher's exact test; OR = 1.24, $P = 0.0313$) (Fig. 6A). We did not detect significant enrichment of DNS variants in ER patches (3E3R) among patients (Fisher's exact test; OR = 1.45, $P = 0.1387$) (Fig. 6A), perhaps owing to the sparsity of ER patches genome-wide. All but one homozygous DNS variant in ER patches were found in infertile men (Fig. 6B, table S14, and supplementary text). All these patient-specific variants were located in the second CC-contained ER (3E3R) patch of UPF3B, including one highly recurrent homozygous variant (c.G1073A:p.R358H) in eight patients and one nonrecurrent homozygous variant (c.C1135T:p.R379C in one patient) (fig. S23, A and B, and table S14). UPF3B is known as a nonsense-mediated decay (NMD)–related and ubiquitously expressed nonclassical RBP (*57*). No report directly links UPF3B with male infertility, yet several clues hint at its functional role in spermatogenesis (*58*). The ER patch in the first CC of UPF3B is essential for its RNA affinity (*59*) and in its second CC was evolutionarily covaried (Fig. 3H). The recurrent homozygous variant in its second CC was validated for its impact on RNA binding (fig. S23C).

Discussion

We adapted RIC to three typical types of mMGCs (mitotic, meiotic, and postmeiotic), leading to the establishment of an atlas of mMGC RBPs, including proteins with previously unknown RNA-related activities that are specialized for meiotic and postmeiotic functions. This implies that meiotic and postmeiotic germ cells may provide an environment for proteins that harbor noncanonical or unknown RBDs to moonlight as RBPs, a trove for future studies. Some DBPs and chromatin-associated factors were identified as novel RBPs by our mMGC RIC. The concept of DNA and RNA binding proteins (DRBPs) has raised much interest because DRBPs could coordinate gene expression at multiple levels more efficiently (*60*, *61*). The highly ordered and dynamic chromatin-associated events during spermatogenesis may require DBPs or RBPs to evolve a dual role as DRBPs (*62*). The possibility cannot be excluded that residual DNA and its cross-linked DBPs could be captured by RNA-DNA hybrids during the polyA pulldown even with deoxyribonuclease treatment (*61*). Some structural proteins are present in the mMGC interactome; they could be components or structural constituents of large RNP complexes, and their interactions with RNAs could be indirect. In addition to mMGC RBPs that bind polyA RNAs, those exclusively binding non-polyA RNAs (such as many noncoding RNAs) could be missed by RIC (*63*). We then showed that mMGC RBPs were involved in multilayered dynamics covering proteomic expression patterns, proteotranscriptomic discordancy, relative RNA binding activity, and RNP concordant behavior. Our results highlighted developmental transition of expression and RNA binding activity of RBPs and illustrated RNP modules with distinct dynamic patterns of concordant behavior in spermatogenesis. With future availability of the global germline PPI network, we hope to create a comprehensive atlas of RNP dynamics during spermatogenesis. Our study presents the initial effort to systematically understand RBPs during mammalian spermatogenesis.

We used RBDmap to screen for testicular RBDs and RNA binding nondomain elements in order to look into RNA binding mechanisms, and discovered that the ER patch is a polyampholytic element present in RBP-CC and contains coevolutionarily coupled residue pairs. A recent perspective showed that CC abounds in human RBPome, implicating its potential role in RNA regulation (*64*). The positional coupling of ER patches and CCs could be

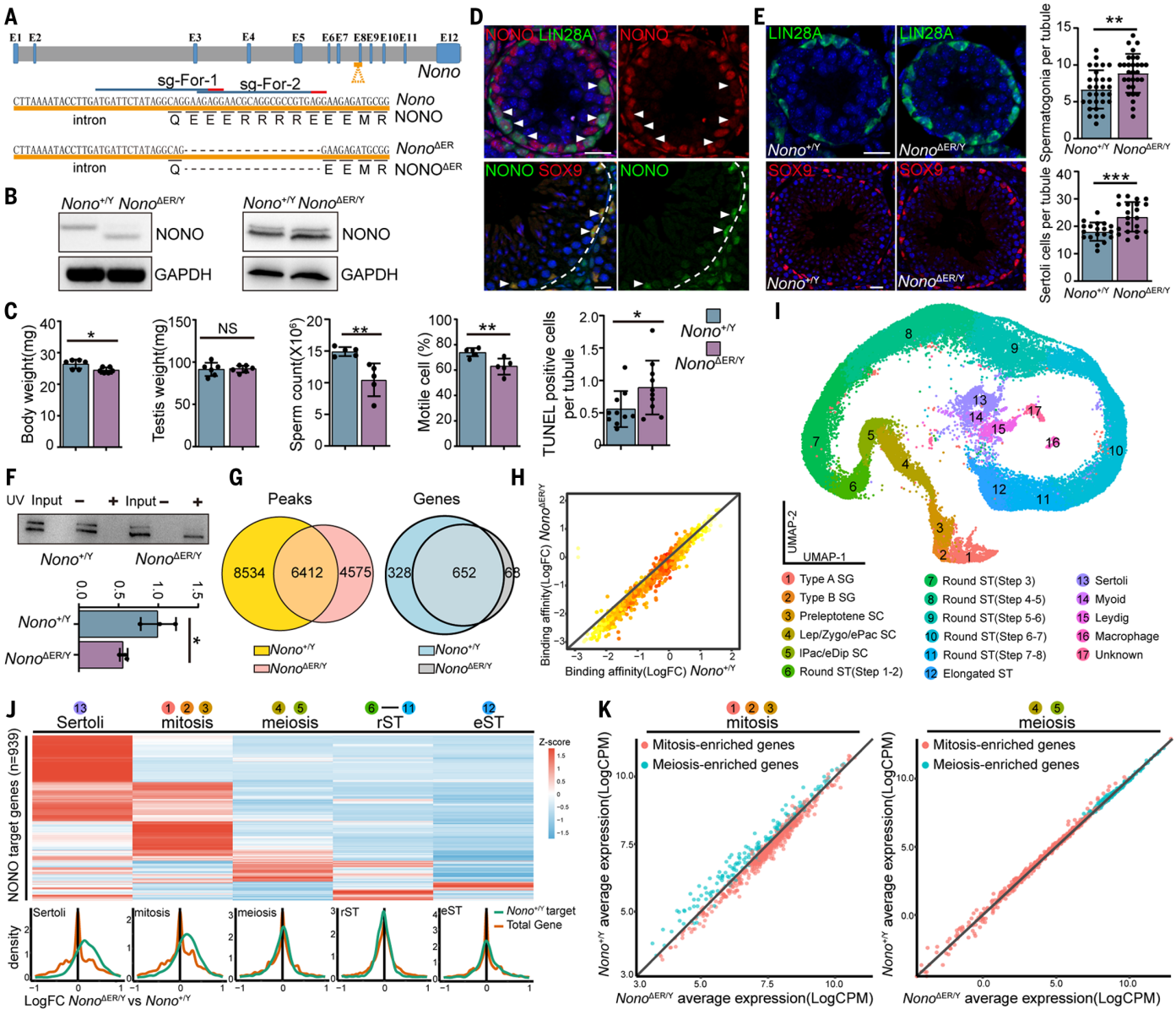


Fig. 5. The NONO ER patch mediates cell type–dependent RNA binding and regulation in mitosis-to-meiosis transition. (**A**) Schematic diagram shows the generation of mice harboring a *Nono*$^{\Delta ER}$ X-linked allele using the CRISPR-Cas9 system. Blue sequences: guide RNA; red sequences: protospacer-adjacent motif (PAM). (**B**) Reverse transcriptase polymerase chain reaction (RT-PCR) (left) and Western blot (right) validation of NONO expression in *Nono*$^{\Delta ER/Y}$ testis. (**C**) Bar plots compare the body weight, testis weight, sperm count, motile sperm, and apoptotic cells between *Nono*$^{+/Y}$ and *Nono*$^{\Delta ER/Y}$ mice. (**D**) Coimmunofluorescence of NONO with the spermatogonia-specific protein LIN28A in 2-week-old testis sections (top) and with the Sertoli cell marker SOX9 in adult testis sections (bottom). White arrowheads point to NONO-positive spermatogonia and Sertoli cells, respectively. Scale bars, 20 μm. (**E**) Immunofluorescence analysis of LIN28A-positive cells in 2-week-old *Nono*$^{\Delta ER/Y}$ versus *Nono*$^{+/Y}$ testis sections (top) and SOX9-positive Sertoli cells in adult *Nono*$^{\Delta ER/Y}$ versus *Nono*$^{+/Y}$ testis sections (bottom). Scale bars, 20 μm. $^{**}P < 0.01$, $^{***}P < 0.001$, by *t* test. (**F**) The RNA binding capacity of NONO protein in *Nono*$^{+/Y}$ and *Nono*$^{\Delta ER/Y}$ mouse testis was measured by q-2C assays. Each band was measured by ImageJ software. All the intensities of 2C bands are relative to their input. $^{*}P < 0.05$, by *t* test. (**G**) Venn diagrams show the overlap of NONO binding peaks (left) and its target genes (right) between *Nono*$^{+/Y}$ and *Non*$^{\Delta ER/Y}$ testes. (**H**) Scatter plots compare the relative intensity of each target gene in NONO and NONO$^{\Delta ER}$ eCLIP-seq in the testis. Each point shows log FC between IP and input in NONO$^{\Delta ER}$ (*y* axis) and NONO (*x* axis) of one gene. (**I**) Uniform manifold approximation and projection (UMAP) shows cell-type clustering of 46,495 single cells from the adult *Nono*$^{+/Y}$ and *Nono*$^{\Delta ER/Y}$ mouse testes, with three replicates for each genotype shown in fig. S18D. The 17 cell types (indicated at the bottom) were assigned based on their marker gene expression in fig. S18E. SG: spermatogonia; SC: spermatocyte; ST: spermatid. (**J**) (Top) Heatmap shows the relative expression of NONO CLIP target genes ($n = 939$) in *Nono*$^{+/Y}$ testicular cells. Cells were clustered based on results shown in (I). Sertoli, Sertoli cells; mitosis, mitotic cells including type A spermatogonia, type B spermatogonia, and preleptotene spermatocytes; meiosis, meiotic cells including leptotene, zygotene, pachytene, and diplotene spermatocytes; rST, round spermatids; eST, elongated spermatids. (Bottom) Density plots show distribution of the expression changes of different gene sets in *Nono*$^{+/Y}$ and *Nono*$^{\Delta ER/Y}$ testicular cells. The *x* axis shows the log FC of *Nono*$^{\Delta ER/Y}$ versus *Nono*$^{+/Y}$. The green line represents NONO CLIP target genes ($n = 939$); the orange line represents total genes ($n = 25{,}771$). (**K**) Scatter plots show the average expression of mitosis-enriched genes (red) and meiosis-enriched genes (blue) in mitotic cells and meiotic cells in *Nono*$^{+/Y}$ and *Nono*$^{\Delta ER/Y}$ mice. LogCPM, logarithm of counts per million reads.

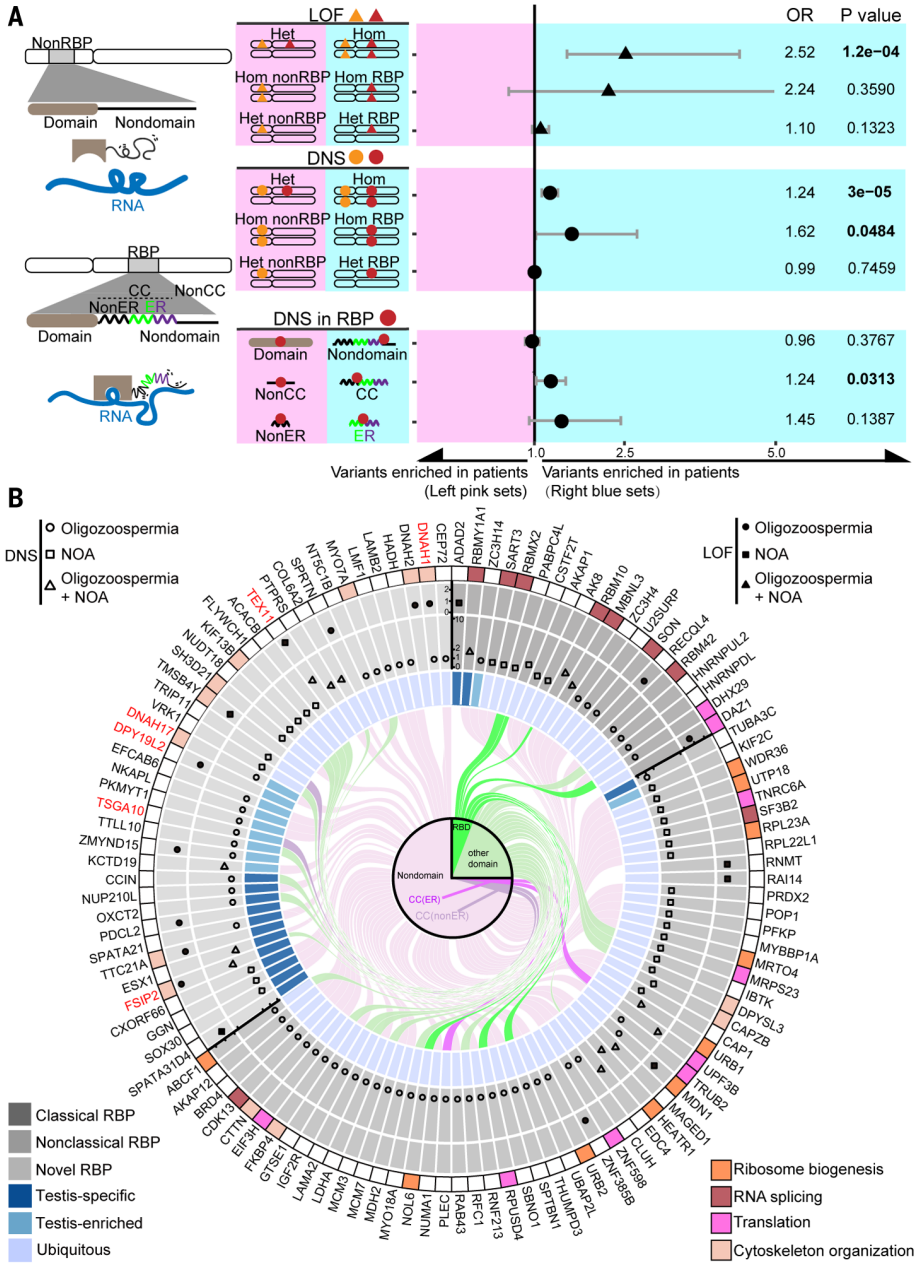


Fig. 6. The genomic architecture of human infertility with a focus on RBPs. (**A**) Forest plot shows odd ratios for different types of variants in patients with male infertility. (Left) Diagrams demonstrate the various elements of RBP and nonRBP. (Middle) Diagrams show the opposing sides, denoted by left pink sets and right blue sets, respectively, of each test. (Right) Diagrams display the estimates of odds ratios (ORs), a range of two-sided 95% confidence intervals (CIs) of ORs, and *P* values of Fisher's exact test. Note that the range of the CI of the OR for Hom LOF variants in RBP versus nonRBP is too wide (0.57 to 16.11) to be fully presented. LOF variants are marked by triangles; DNS variants are marked by circles. Variants in RBP genes are signed with red; variants in nonRBP genes are signed with yellow. Hom, homozygote; Het, heterozygote. *P* values in bold are significant at $P < 0.05$. (**B**) Circos plot lists 137 candidate Hom DNS variants in 105 RBP genes and 22 candidate Hom LOF variants in 18 RBP genes. The plot shows the following (from the inside out): (i) The pie chart clusters variants according to their locations in RBD (dark green lines), other domain (light green lines), 3E3R patch of CC (dark purple lines), non-3E3R of CC (medium purple lines), and nonCC nondomain region (light purple lines). (ii) The Sankey diagram designates the corresponding RBP of each variant. (iii) The blue-shaded circle shows the tissue specificity for each RBP: testis-specific (dark blue); testis-enriched (medium blue), and ubiquitous (light blue). (iv) The gray-shaded circle classifies RBPs into classical (dark gray), nonclassical (medium gray), and novel (light gray). Different shades in this circle indicate the sources of variants for given RBPs found in oligozoospermia-infertile men only (dot), nonobstructive azoospermia (NOA)–infertile men only (square), and both sets of infertile men (triangle). The radial height denotes the time of recurrence of variants in patients. DNS (hollow) and LOF (solid) variants are represented separately. (v) The outermost circle shows the four selected enriched GO terms among RBPs with candidate variants. RBPs having been reported for human infertility [according to the IMIGC (International Male Infertility Genomics Consortium) database] are highlighted in red.

explained by the canonical alpha helices' formation with polyampholytic sequences (*40*). The ER patches that we identified are relatively small in length (18 amino acids on average); they contain the negatively and positively charged residues that are alternating and not in long blocks of the same charge, and they prefer to reside in alpha helices. These features differ from what others have found about mixed-charge stretches, which are relatively long (50 residues or more) and often contain blocks of negatively and positively charged residues within IDRs, resulting in high local charge density for phase separation (*65*, *66*). The well-mixed linear sequence of oppositely charged residues of the ER patch may adopt random-coil conformations, which promotes the flexibility and the high local concentration of connected functional elements (*40*, *67*). Although a few polyampholytes are recently emerging to play a role other than regulating RNA, such as condensates assembly (*66*), little is known about their role as an RNA binding element. The high positional concordance of ER patches and CCs suggests that they are functionally coupled to regulate RNA.

The ER patches of three (out of four) tested RBPs (NONO, PSPC1, and UPF3B; ER mutation of UPF3B was found in infertile men) were verified biochemically to enhance RNA binding. By creating knock-in animal models, the ER patches of both tested RBPs (NONO and SAFB) were verified as a requirement for securing normal spermatogenesis. With a focus on NONO, we further found that its ER patch regulates the mitosis-to-meiosis transition of male germ cells, which can be explained by an impairment of RNA binding activity of NONO and a concomitant up-regulation of NONO target mRNA transcripts. Previous studies show that NONO predominantly binds to the 5′ ends of pre-mRNAs and that NONO knockdown induces a processing defect at the 5′ end of NONO-bound pre-mRNAs (*68*). The depth of our scRNA-seq is not sufficient for analysis of subtranscripts. Thus, we cannot exclude the possibility that some NONO targets are actually decreased in *Nono*$^{\Delta ER/Y}$ testes because of the potentially altered pre-mRNA 5′ end

processing. Generally, the NONO ER patch enhances NONO protein-RNA interactions, maintains target mRNAs expression, and ensures mitosis transition into meiosis. Loss of the NONO ER patch results in decreased NONO binding to its target mRNAs, up-regulated levels of mitotic transcripts, abnormal accumulations of Sertoli cells and mitotic spermatogonia, and impaired progression of mitotic germ cells into meiotic spermatocytes.

NONO seems to play a critical role under stress (*69*). External stimuli, such as UV irradiation, up-regulate NONO expression in the testis (*48*). Feeding increases the number of NONO-containing speckles in the liver (*70*). NONO knockout mice often exhibit a more severe phenotype under specific stimuli (*48*, *70*). In our study, ER deletion caused severe functional consequence, with a droplet-like state being transformed into a fiber-like state, in high-level NONO expression in cell lines. We therefore speculate that the NONO ER patch may augment its role in response to stress-induced higher expression of NONO.

There are some noteworthy points about the RBPs harboring CC-ER. First, CCs are especially enriched in RBPs, and ER patches are especially enriched in RBP-CCs, implying that the ER patch may play a role in CC-mediated RBP-RNA interactions rather than mediating CC formation or oligomerization, which is a general function of CCs in other kinds of nonRBPs (*71*). Second, RBPs containing CC-ER always form homo- or heterodimers, which are critical for the assembly of RNP granules. For example, DBHS family proteins SFPQ, NONO, and PSPC1 share structural similarities, having a common ER patch–embedded CC, and their dimers could form through CC interactions (*41*, *72*, *73*). However, the ER patch seems dispensable for dimer formation because it can be absent from the available structures of these dimers (*41*). This notion is supported by the stable intracellular interaction of $NONO^{\Delta ER}$ with SFPQ and PSPC1. Third, mutations in a CC of SFPQ affect SFPQ localization and polymerization and reduce SFPQ–nucleic acid interactions, and specifically, the C-terminal part of the CC that contains the ER patch mediates higher-order homotypic associations (*73*). On the basis of the above and our results, it is probable that such higher-order associations are linked to enhancing RNA binding.

Linkages between a wide spectrum of RBPs and specific diseases have not yet been established. Genetic etiology for human idiopathic infertility remains far from clear. Homozygous mutations are generally believed to be more relevant to infertility and thus are the intense focus of current research. Our whole-exome sequencing analysis strengthened this argument by uncovering enrichment of homozygous LOF and DNS variants in infertile patients and prioritized these mutations in RBPs for future studies. Most homozygous LOF variants were found in novel RBPs, implying that classical and nonclassical RBPs with crucial roles in maintaining human-development homeostasis are generally intolerant to mutations of large effects (*74*). However, the low frequencies of homozygous variants are not sufficient to explain a considerable proportion of idiopathic infertility patients. As a general scenario, idiopathic infertility could be of polygenic origin and caused by the cooperative effects of deleterious heterozygous variants (*24*). Our genome-wide heterozygous analysis supports a pathological role of heterozygous variants of both RBPs and nonRBPs in human infertility, with RBPs being more likely to be mutation-intolerant than nonRBPs at both the LOF and DNS levels. It suggests that heterozygous variants constitute an unignorable genetic component to the pathogenesis of infertility, which is in concert with a recent finding demonstrating that single heterozygous DNS variant in *HIWI* is sufficient to cause human infertility (*55*). Moreover, the pathogenic role of compound heterozygous DNS or LOF variants of RBPs was documented (*25*, *75*, *76*). We also note that the odds of finding a DNS variant in CCs and ER patches (3E3R) of RBPs are higher in patients than in controls, and are at least statistically significant for the CCs, whereas the ER patches did not show a significant difference and should be revisited with a larger cohort in the future. These findings are beginning to link nondomain elements with human infertility.

Materials and methods

mMGC purification

We isolated each type of spermatogenic cells from C57BL/6J mouse testes by STA-PUT velocity sedimentation (*77*, *78*). pSC and rST were isolated simultaneously from 16 adult testes per round of STA-PUT; SG were purified from 80 pairs of P6-8 testes per round. Testes were decapsulated, washed with DMEM twice, and incubated with 1 mg/ml collagenase IV in DMEM in 37°C water bath for 15 to 20 min. Seminiferous tubules were then gently pipetted to disperse into the solution, washed twice with DMEM, and collected after sedimentation by gravity. The sedimented tubules were further digested with 0.25% trypsin containing 1 mg/ml DNase I and gently pipetted up and down for 5 to 15 min before adding 10% FBS. Cells were pelleted by centrifugation at 500 × g for 5 min and then resuspended with wash buffer (10% FBS, 0.5% BSA, 200 μg/μl DNase I), and filtered with 40-μm cell strainer to obtain a monodisperse suspension. The single-cell suspension was loaded into a cell separation apparatus (ProScience, Canada), layered over a linear gradient (2 to 4%) of bovine serum albumin (BSA) in DMEM, and then allowed to sediment by gravity for 2.5 to 3 hours. Finally, the cells were fractionated, type-determined, and pooled into quasi-homogeneous populations. Cell purity was checked under DIC light microscope based on distinctive diameters and morphological features, and selectively validated under immunofluorescence microscope based on staining patterns of DAPI and/or protein marker antibodies. Note that the spermatogonial preparations contain no spermatocytes, which are not supposed to appear in P6 to 8 testes. pSC may be contaminated by other similar-sized meiotic cells such as zygotene and diplotene spermatocytes, and rST preparations may be contaminated by elongated spermatid. To achieve sufficient cell numbers qualified for RIC and proteomics, the STA-PUT procedure was repeated. Specifically, for each replicate of RIC, 20 rounds of STA-PUT were performed to collect ~2.5×10^8 SG, 14 rounds to collect ~2.0×10^9 rST, and 28 rounds (14 rounds synchronous with rST, plus 14 extra rounds) to collect ~5.0×10^8 pSC, respectively. A total of 96 rounds of STA-PUT provided two replicates of RIC for each cell type.

mMGC RIC

RIC was carried out according to a previously described protocol (*79*), with modifications for freshly purified mMGCs. Single types of cells were pooled and washed with ice-cold PBS three times to thoroughly remove DMEM. Before UV cross-link, we sought to remove excess liquid buffer that could not be easily pipetted out in order to minimize the loss of UV energy, considering that we used suspended cells rather than regularly cultured cells adherent tightly to the plate bottom. To this end, we spread 1 ml buffer with suspended cells through four or five 10-cm dishes sequentially. The cells were allowed to settle for awhile in each dish and then transferred with most of the buffer to the next dish, with a considerable portion of cells left in a tiny amount of the remaining buffer. Microscopy inspection showed that most settled cells dispersed well and adhered to the dish bottom. Each dish without its lid was placed on ice at a distance of 10 to 15 cm from the UV lamp and irradiated with 0.15 J/cm^2 at 254-nm UV light in a UV cross-linker device (CL-1000). Every time, half the number of cells were irradiated with UV, and in parallel, the other half were not UV cross-linked. The cells were resuspended in ice-cold PBS, pelleted in a conical tube by centrifugation at 500 × g for 5 min, snap-frozen in liquid nitrogen, and stored at −80°C until use. After collection for RIC, the frozen pellets were thawed and combined using ice-cold lysis buffer (20 mM Tris pH 7.5, 500 mM LiCl, 0.5% lithium dodecyl sulfate, 1 mM EDTA, 5 mM DTT). Oligo dT_{25} beads (NEB, S1419S) were used to capture polyA RNA. After eight stringent washes, covalently cross-linked RNA-protein complexes were released from beads by 55°C incubation. Protein eluates obtained from two RIC replicates for each cell type were preserved for MS.

Testis RBDmap

RBDmap was performed following a previously described protocol (*19*), with adaptations for fresh testicular cells. For each RBDmap replicate, 6 adult testes and 24 P28 testes were decapsulated and separately subjected to collagenase IV and trypsin enzyme digestion to obtain a mixed population of single cells. As detailed for RIC, cells were resuspended with minimal ice-cold PBS and settled to the dish bottom to facilitate UV cross-linking. RNA-protein complexes were first captured on oligo(dT) beads following standard RIC procedure. The RNA-protein complexes were then digested with Lys-C (WAKO, 125-05061) in RBP elution buffer: 20 mM Tris-HCl (pH 7.5), at 200 rpm in thermo mixer at 37°C for 8 hours, generating primary peptides which harbor all the RNA-bound and release peptides. New Oligo dT_{25} beads were incubated with the bound groups at 4°C for 1 hour. The beads were then washed with 1 × hybridization buffer (200 mM Tris-HCl pH 7.5, 5 M LiCl, 2 mM EDTA, 50 mM DTT), and the supernatant was saved as the release-peptide sample. The beads were sequentially washed by lysis buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.5% LiDS, 1 mM EDTA, 5 mM DTT), Wash-1 buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 0.1% LiDS, 1 mM EDTA, 5 mM DTT), Wash-2 buffer (20 mM Tris-HCl pH 7.5, 500 mM LiCl, 1 mM EDTA, 5 mM DTT), and Low salt buffer (20 mM Tris-HCl pH 7.5, 200 mM LiCl, 1 mM EDTA, 5 mM DTT). The beads were resuspended in an elution buffer to elute the RNA-peptide complex with 500 rpm in a thermomixer at 55°C for 10 min, and the supernatant was saved as the bound-peptide sample. New Oligo dT_{25} beads were added in primary peptide samples for another two rounds of bound and release peptide capture. All the generated bound and release peptide samples were pooled separately for mass spectrometry.

LC-MS/MS analysis

The RIC and RBDmap proteins were concentrated using Amicon Ultra Centrifugal Filters (10 kDa cutoff, Millipore) and subsequently reconstituted in lysis buffer [8 M urea, 75 mM NaCl, 50 mM Tris pH 8.2, 1% (v/v) EDTA-free protease inhibitor cocktail, 1 mM NaF, 1 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF]. The RIC and RBDmap protein lysates were then subjected to protein digestion, followed by peptide desaltation. Briefly, cysteine residues were reduced by 5 mM DTT for 25 min at 56°C followed by alkylated using 14 mM iodoacetamide for 30 min at room temperature in the dark, and quenched with DTT for an additional 15 min. Samples were then diluted to 1.6 M urea with 25 mM Tris-HCl, pH 8.2, and digested overnight at 37°C with the addition of 2 μg trypsin. Following digestion, peptide mixtures were acidified by trifluoroacetic acid (TFA) (0.4% vol/vol) (Sigma-Aldrich), desalted using a SepPak 1 cc tC18 cartridge (Waters), and then lyophilized. For RIC analysis, purified peptide samples were resuspended in Buffer A (20 mM ammonium formate, pH 10), fractionated using an XBridge BEH C18 column (Waters) to generate a total of eight fractions, and dried in vacuum.

For analyses of mMGC RIC peptides, the peptide samples were resuspended in 0.1% formic acid (FA) and analyzed using a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) coupled online to a Proxeon Easy-nLC 1000. Peptides were loaded onto a trap column (75 μm × 2 cm, Acclaim PepMap100 C18 column, 3 μm, 100 Å; DIONEX, Sunnyvale, CA) and separated by a reverse-phase microcapillary column (75 μm × 25 cm, Acclaim PepMap RSLC C18 column, 2 μm, 100 Å; DIONEX, Sunnyvale, CA) at a flow rate of 300 nl/min. The HPLC solvent A was 0.1% FA, and the solvent B was 100% acetonitrile (ACN), 0.1% FA. A 95-min linear gradient (3% to 8% buffer B for 3 min, 8% to 29% buffer B for 71 min, 29% to 41% buffer B for 12 min, 41% to 90% buffer B for 1 min, 90% buffer B for 8 min) was used for peptide analysis in the data-dependent acquisition (DDA) mode. An MS survey scan was obtained for the m/z range 350 to 1800 at a resolution of 60,000, and MS/MS spectra were acquired from the survey scan for the 20 most intense ions.

For analyses of testis RBDmap peptides, the peptide samples were resuspended in 0.1% FA and separated on an analytical column (75 μm × 160 mm, 1.9 μm, Dr. Maisch) at a flow rate of 300 nl/min in a 95-min linear gradient (3% to 5% buffer B for 5 s, 5% to 15% buffer B for 40 min, 15% to 28% buffer B for 34 min 50 s, 28% to 38% buffer B for 12 min, 30% buffer to 100% buffer B for 5 s, and 100% buffer B for 8 min) using buffers of 0.1% FA (buffer A) and 80% ACN, 0.1% FA (buffer B). The effluents from LC system were then analyzed using Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) operated in DDA mode. A full survey scan was obtained for the m/z range of 350 to 1500 with 60,000 resolution, and MS/MS was performed by high energy dissociation (HCD) in the Orbitrap with a resolution of 15,000.

Antibodies, Western blots, and immunofluorescence

Commercial antibodies: NONO (Abcam, Ab70335; ABclonal, A3800; Sangong, D199144); SAFB (ABclonal, A7927); MSH6 (ABclonal, A8795); HA (MBL, M180-3; CST, 3724S); CEP72 (Proteintech, 19928-1-AP); CENPV (Sangon, D162083); SYCP3 (Abcam, ab97672); PNA (Vector Labs, RL-1072); HNRNPQ (Abcam, Ab189405); MOV10 (Proteintech, 10370-1-AP); MILI(Abcam, Ab36764); MIWI (Abcam, Ab181056); GAPDH (ABclonal, AC002); ACTIN (Sigma, SAB5701411); TUBULIN (ABclonal, AC008); FYCO1 (Sangong, D126224); TEKT5 (Proteintech, 26108-1-AP); LIN28A (Abcam, ab279647); SOX9 (Millipore, AB5535); DNMT3A (Huabio, ET1609-31); ZBTB16 (R&D, AF2944). BOULE and PUM1 antibodies were gifts from Eugene Yujun Xu Lab (Nanjing Medical University). TEX30 and PDHA2 antibodies were generated by ABclonal.

For Western blot analysis, testes or 293T cells were homogenized in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.5% sodium deoxycholate, 1% Triton X-100, 2% SDS, 10% glycerol) added with protease inhibitor. The lysates were centrifuged at 12,800 rpm at 4°C for 30 min, and the supernatant was collected. The proteins were denatured by adding SDS loading buffer, heated for 10 min at 95°C, and loaded on SDS-PAGE. Immunofluorescence analysis of testis sections was performed as described previously (*80*). TUNEL analysis of testis sections was performed with the In situ Cell Death Detection Kit (Roche, 11684817910) according to the manufacturer's protocol.

Radioactive CLIP

Conventional CLIP of RBPs (PDHA2, TEX30, NONO, and PSPC1) was performed as previously described (*80*). For CLIP using testicular tissues, 100 mg juvenile testes from P21 (for PDHA2) or P28 (for TEX30) mice were detunicated and mildly pipetted to form discrete single seminiferous tubules in ice-cold PBS, and immediately UV cross-linked with 0.4 J/cm^2 at 254 nm for three times. For CLIP using cell lines, HA-tagged NONO or PSPC1 was transfected into 293T cells in a 10-cm dish. The dish was washed by cold PBS and UV cross-linked with 0.12 J/cm^2 at 254 nm for one time. Cross-linked samples were washed with cold PBS and pelleted in a conical tube by centrifugation at 500 × g for 5 min. Pellets were snap-frozen in liquid nitrogen and stored at −80°C until use. To begin CLIP, briefly, pellets were dissolved with 1 × PMPG (1× PBS, 2% Empigen) containing protease inhibitors (Roche, 11873580001) and RNA inhibitors (Promega, N2518), and then treated with DNase I. Then protein-RNA complexes were immunoprecipitated by 10 μg antibody using protein A Dynabeads (Invitrogen, 10002D) for rabbit antibody or protein G Dynabeads (Invitrogen, 10003D) for mouse antibody. Then RNA was labeled with [γ-32P] ATP and run on NuPAGE 4 to 12% Bis-Tris (Life Technology, NP0323BOX). The isotopic cross-linked RNA-protein complexes were transferred onto nitrocellulose membrane (Invitrogen, LC2001) and detected by a phosphor screen system (Typhoon).

eCLIP-seq

eCLIP-seq was performed as previously described (*81*, *82*). Cross-linked testicular samples for eCLIP of RBPs (PDHA2, TEX30, NONO, and $NONO^{\Delta ER}$) were prepared as for conventional CLIP. HA-tagged NONO and $NONO^{\Delta ER}$ were transfected

into 3T3-L1 cells in a 10-cm dish and cross-linked as conventional CLIP. To begin eCLIP, briefly, samples were dissociated in eCLIP lysis buffer containing protease inhibitor (Roche, 04693159001), RNase inhibitor, RNase I (Life Technology, AM2295), and Turbo DNase (Invitrogen, AM2239). Cross-linked target RNPs were captured on Dynabeads using 10 μg antibodies. Protein-bound RNAs were then dephosphorylated and ligated with 3′ RNA adapter. eCLIP and input samples (collected before immunoprecipitation) were then washed stringently, boiled from beads, run on NuPAGE 4 to 12% Bis-Tris (Life Technology, NP0323BOX), and transferred onto nitrocellulose membrane (Amersham, 10600008). Size-matched subregion of the membrane was collected and proteins were degraded by Urea and proteinase K (NEB, P8107S). Harvested RNAs were subsequently reverse-transcribed into cDNA with Affinity Script (Agilent,600107), and purified by Exo-Sap-IT (Affymetrix, 78201). The 3′ end of cDNA was ligated with a DNA adapter by T4 RNA Ligase (NEB, M0437M). Libraries were then PCR-amplified using Q5 Polymerase and run on 3% low–melting temperature agarose gel. Sizeable libraries were purified with a MinElute gel extraction kit (Qiagen, 28604) and run on the Illumina platform.

scRNA-seq libraries

Mouse testes were washed in ice-cold RPMI1640 and dissociated using Collagenase II (Sigma, V900892-100MG) and DNase I (Sigma 9003-98-9). Cell count and viability was estimated using fluorescence Cell Analyzer (Countstar Rigel S2) with AO/PI reagent after removing erythrocytes (Solarbio R1010), debris, and dead cells (Miltenyi 130-109-398/130-090-101). Finally, fresh cells were washed twice in RPMI1640 and then resuspended at 1×10^6 cells/ml in 1× PBS with 0.04% bovine serum albumin. scRNA-seq libraries were prepared using SeekOne Single Cell Full-length RNA sequence Transcriptome-seq Kit according to the manufacturer's instructions (SeekGene, Catalog No.K00801). Briefly, an appropriate number of cells were mixed with reverse transcription reagents and added to the sample wells of the SeekOne DD Chip S3. Then, Barcoded Hydrogel Beads (BHBs) and partitioning oil were dispensed into corresponding wells separately in Chip S3. Subsequently, cell-containing reverse transcription reagents and BHBs were encapsulated into emulsion droplets using SeekOne Digital Droplet System. Immediately after transferring emulsion droplets into PCR tubes, 15 cycles of annealing (ramping from 8° to 42°C) followed by a 5-min heat inactivation at 85°C were performed to obtain barcoded cDNA. Next, the barcoded cDNA was purified from broken droplet and PCR reactions were performed twice to remove the majority of ribosomal and mitochondrial cDNA. AMPure beads were used to purify cDNA from the post-PCR reaction mixture. Finally, one-fourth volume of cDNA was fragmented, end-repaired, A-tailed and ligated into sequencing adaptor. DNA amplified by index PCR contained any part of polyA or non-polyA RNA and Cell Barcode and Unique Molecular Index. The indexed sequencing libraries were purified using AMPure beads and quantified by quantitative PCR (KAPA Biosystems KK4824). The libraries were then sequenced on Illumina NovaSeq 6000 with PE150 read length.

2C assay

2C assay was performed as previously described (*31*, *32*), with modifications. Cross-linked testicular single-cell mixtures were prepared as described for testis RBDmap, and cross-linked 293T cells were prepared as described for conventional CLIP. The primer sequences for cloning are listed in table S16. Samples were then processed using an Allprep DNA/RNA mini Kit (QIAGEN, 80204). Briefly, cross-linked samples were lysed in RLTplus buffer (QIAGEN, 1053393) with 1% beta-mercaptoethanol using an ultrasonic homogenizer for 10 s on/10 s off for about 40 cycles until lysate became transparent. To eliminate DNA, the lysate was loaded on the Allprep DNA mini spin column and centrifuged at 3000 × g for 5 min. An equal volume of 70% ethanol was added to the flowthrough, mixed well, loaded onto RNeasy mini spin column, and centrifuged at 3000 × g for 5 min. The RNeasy column was then washed in sequential with RW1 and RPE wash buffers by centrifugation at 3000 × g for 3 min. 200 μl nuclease-free water was added to RNeasy column and incubated for 3 min. The free RNA and RNP complexes were eluted in nuclease-free water and collected by centrifugation at 3000 × g for 3 min. RNA concentration was measured by Nanodrop 2000. For quantitative 2C (q-2C) assays, the RNA concentration of the sample was required to vary less than 10% of that of the control. Samples were then dried into powder by vacuum centrifugal concentrator (Centtivap) and resuspended in 30 μl nuclease-free water. RNAs were digested by adding 1 μl 10 mg/ml RNase A (Thermo, EN0531) or 1 μl benzonase (Merck, E8263) at 800 rpm in thermomixer for 12 hours. Samples were then subjected to Western blot analysis.

For in vitro q-2C assay, 500 ng purified NONO or $NONO^{\Delta ER}$ proteins and 1 μg total RNAs isolated from 293T cells were incubated in 2× RNA binding buffer (10 mM HEPES pH 7.4, 50 mM KCl, 4 mM $MgCl_2$, 4 mM DTT, 0.2 mM EDTA, 7.6% glycerol, 3 mM ATP) for 30 min, transferred into a well of a 96-well plate, and UV cross-linked with 5J at 254 nm for 10 min. Eight wells of protein-RNA mixtures were then pooled and subjected to the same q-2C assay.

mRNP capture assay

mRNP capture assay was performed as a small-scale RIC (*10*). Briefly, transfected 293T cells in 10-cm dishes were UV cross-linked and lysed with 50 μl mRNA interactome lysis buffer as in RIC. 50 μl Oligo dT_{25} beads were used to capture polyA RNAs. After eight rounds of stringent washes, proteins were eluted from Oligo dT_{25} beads and then subjected to Western blot analysis.

Protein expression and purification

All coding sequences (CDS) were PCR-amplified from mouse testis cDNA. Full-length CDS sequences of PDHA2 and TEX30 were separately cloned into the pET28a expression vector. CDS of mouse NONO (1-473) and $NONO^{\Delta ER}$ (1-346, 355-473) were cloned into the pET22b expression vector. The primer sequences for cloning are listed in table S16. All constructed plasmids were transformed into *E. coli* BL21. Different proteins (PHDA2, TEX30, NONO, and $NONO^{\Delta ER}$) were expressed and purified with differential modifications of our previous protocol (*83*) (supplementary materials, materials and methods). Protein concentration in the supernatant was measured by Nanodrop 2000 (Thermo) on a 280-nm absorption peak. Proteins were stored at −80°C until use.

Nono$^{\Delta ER}$ and *Safb*$^{\Delta ER}$ knock-in mice

Knock-in mice with precise ER truncation were generated by GemPharmatech using the CRISPR-Cas9 system. sgRNA sequences were designed to direct Cas9 endonuclease to the front of Exon 9 of *Nono* to generate *Nono*$^{\Delta ER}$ allele (Fig. 5A) and to the Exon 15 of *Safb* to generate *Safb*$^{\Delta ER}$ allele (fig. S15A), respectively, by homologous recombination. Mouse zygotes were co-injected with donor, Cas9 mRNA, and sgRNA, and transferred into surrogate mothers' endometrium. All animals were handled as stipulated by the Guide for the Care and Use of Laboratory Animals at Nanjing Medical University (ethics approval ID: 2101008 and ID: 2102008).

Human male infertility diagnosis and cohort information

A total of 1046 unrelated infertile men (ranging from 19 to 51 years old) who displayed unexplained (idiopathic) NOA (n = 516) or oligozoospermia (n = 530) were recruited between June 2017 and March 2022 at the Reproductive and Genetic Hospital of CITIC-Xiangya (Changsha, China) (table S13). A group of 300 people (Han Chinese) (ranging from 20 to 73 years old) with normal childbearing history was referred to as the control group (table S13). This study was approved by the ethics committee of CITIC Xiangya Reproductive Genetics Hospital (LL-SC-2017-025 and LL-SC-2019-034).

All patients received an inquiry for detailed history to rule out those male individuals with clear causes, including iatrogenic injury, reproductive tract infection, testicular inflammation, drug exposure and other factors (*84*). We also

excluded other testicular injuries, such as testicular trauma and testicular torsion. Physical examination, including height, weight, hair distribution, mental status, and external genitalia, was conducted and all men in our cohort were normal. Genetic screening for karyotyping (G-bands at the 320- to 400-band level) and Y chromosome AZF region deletions (sY84 and sY86 for AZFa, sY127 and sY134 for AZFb, sY254 and sY255 for AZFc) was performed according to the Reproductive Medicine Professional Committee of Chinese Medical Doctor Association. In our cohort, no patients had Y chromosome AZF microdeletions or chromosomal abnormalities.

Semen analysis was conducted according to the guidelines of the World Health Organization (WHO) Laboratory Manual for the Examination and Processing of Human Semen (*85*). The individuals were diagnosed with idiopathic male infertility based on the reference (*84*).

Azoospermia was defined as a condition in which no spermatozoa were detected on three consecutive occasions when ejaculated semen was examined microscopically after centrifugal sedimentation (*86*). A non-obstructive etiology was established on the basis of the following parameters: hormone values, pH and volume of semen, testicular volume, and testicular biopsy according to guidelines (*87*, *88*) (n = 509). For some patients (n = 7), although other parametric information was insufficient to differentiate between obstructive and non-obstructive etiologies, they were also recruited as the latter based on their normal vas deferens. Any azoospermia patients with previous vasectomy, congenital absence of a vas deferens, vasal agenesis or obstructive vas deferens were excluded from our cohort (table S13).

Oligozoospermia was defined as a condition in which the concentration of spermatozoa in the ejaculated semen was below the lower limit of the reference value for semen analysis in normally fertile men (*85*). All the oligozoospermia patients in our cohort were further classified into two subtypes based on sperm concentrations: severe oligozoospermia: 0.1 to 5×10^6/ml (n = 503); extreme oligozoospermia: $< 0.1 \times 10^6$/ml (n = 27) (table S13).

Whole-exome sequencing

Genomic DNA (gDNA) from peripheral blood samples was extracted using a QIAamp DNA Blood Midi kit (Qiagen, 51106) according to the manufacturer's protocol. The qualified gDNA samples of the 1046 infertile men and 300 fertile controls were subjected to library construction, followed by WES on Illumina HiSeq 2000 or HiSeq X-TEN platforms (*89*).

Bioinformatics

RIC proteomic analysis

The mass spectrometric raw data were searched with MaxQuant (version 1.5.2.8), identified tryptic peptides assigned to contaminants were removed, and only peptides uniquely mapped to a gene model were considered. High-confidence RIC RBPs were defined as those with iBAQ in cross-linked (CL) pulldowns at least 10-fold higher than non-cross-linked (noCL) pulldowns in both RIC replicates.

GO annotation was obtained from Ensembl 75. RNA-associated GO annotation used R package mRNA interactome Hela. More specifically, proteins annotated by the GO-term "RNA-binding" and by other GO-terms related to RNA biology were classified as RNA binding and RNA-related proteins, respectively (*10*). The rest of the RBPs were classified as RNA-unrelated proteins.

A manually curated list of canonical and non-canonical RBDs was used to classify RIC RBPs as harboring a canonical or noncanonical RBD (table S1). RBPs were annotated as classical if they contained at least one canonical RBD, or classified as nonclassical if they did not contain canonical RBDs but harbored noncanonical ones or were listed in a manually curated non-classical RBP table from Liao *et al.* (*30*), or classified as novel otherwise. With reference to a manually curated list of published mSC RBPs (*30*), we classified mMGC RBPs (this study) into three families: mSC-specific, mMGC-specific, and common.

Tissue specificities of mMGC RBPs were determined by the expression specificity score (SPM) estimated from RNA-seq profiling data across 17 mouse tissues (*90*), as described in TiSGeD (*91*). Genes with SPM score ≥0.9 were classified as testis-specific, genes with SPM score ≥0.5 and <0.9 were classified as testis-enriched, and other genes were classified as widely expressed.

Human and mouse orthologous gene mapping

The mapping of orthologous genes between human and mouse was conducted by orthologsBioMART R package (*92*).

GO enrichment/network analysis

The GO enrichment was analyzed by either R package clusterProfiler (with p.adjust < 0.05) (*93*), or R package GSEA (with p.adjust < 0.05) (*94*), or StringDB web-based platform (*95*). GO enrichment network was analyzed by Metascape web-based platform (with P value < 0.05) (*96*).

TMT-based quantitative proteomics

TMT-MS data, including four spermatogenic stages of mMGCs (SG, pSC, rST, and eST) with four biological replicates for each, were analyzed using MaxQuant (version 1.2.2.5) software with UniProt proteomes (mouse) protein database, as described previously (*33*). The normalized protein abundance for each protein was used for further bioinformatics analyses. Hereon, unless specially indicated elsewhere, proteins listed as mMGC TMT-MS proteins and not as mMGC RBPs were defined as nonRBPs.

To determine the temporal pattern of protein abundance across four consecutive stages, we softly clustered proteins into 10 groups based on the average normalized protein abundance using the Mfuzz package in R (*97*).

Relative RNA binding activity analysis

The estimation of relative RNA binding activity refers to Kilchert *et al.* (*37*). RBPs not detected by TMT-MS in any single cell type were filtered out. Expression of RBPs with multiple protein isoforms were aggregated by taking the average. In total, there were 1218 distinctive RBP genes left for subsequent relative RNA binding activity analyses. Raw WCE and RIC intensity were log2-transformed and scaled within each sample. Scaled log2 (RIC/WCE) was processed with the eBayes function in the limma R package to quantify relative RNA binding activity of each RBP, and moderated t test was carried out to examine significant alteration in relative RNA binding activity between conditions.

Scoring intrinsic disorder of amino acids

For all amino acids in a protein, a score for intrinsic disorder was computed using IUPred (*98*). Amino acid residues with a score larger than 0.4 were called disordered. For each protein, the proportion of disordered residues was estimated.

RBDmap proteomic analysis

The mass spectrometric raw data were processed using MaxQuant (version 1.6.5.0) (*99*), and mass spectrometric spectra were searched using the Andromeda search engine against mouse proteins in UniProt. MS-identified tryptic peptides assigned to contaminants were removed, and only peptides uniquely mapped to a gene model were considered. A total of 5754 peptides were retained after quality control (table S7). MS-identified tryptic peptides enriched in the RNA-bound or released fractions, respectively, were mapped back to the whole proteomes of the mouse (UniProt) and extended to the two adjacent Lys-C cleavage sites to recall the original proteolytic fragment. A peptide with the average LFQ.intensity of three replicates in log1p ratio (logarithm ratio of the LFQ.intensity of Bound to Release after adding a pseudocount of 1) greater than 10 was considered RBDpeps. Finally, overlapping RBDpeps of each protein were merged. RBDpeps between protein isoforms were combined, and the UniProt ID of canonical isoform was retained.

We used three criteria to filter RBDpeps of different levels of confidence. Under the strictest criterion, we applied a moderated t test implemented in the R package limma to select those with log ratio of the average LFQ.intensity of three replicates in Bound to Release > 0 at an FDR < 1%, which generated 361 non-redundant

RBDpeps (from total 663 RBDpeps) as Stringent RBDpeps. Under the less strict criterion we selected those with the criterion log1p ratio of the average LFQ.intensity of three replicates in Bound to Release > 10, which generated 721 nonredundant RBDpeps (from total 1533 RBDpeps) as Moderate RBDpeps. Under the least strict criterion, we selected those with log1p ratio of the average LFQ.intensity of three replicates in Bound to Release > 0, which generated 793 nonredundant RBDpeps (from total 1859 RBDpeps) as Loose RBDpeps.

To annotate RBDpeps, each identified RBDpep was examined whether it overlaps with a Pfam domain (http://ftp.ebi.ac.uk/pub/databases/Pfam/) (*100*), disorder regions, or CC motifs. Note that CCs adopt an alpha-helix form and are predicted based on Pfam annotation. RBDpeps that overlap with the Pfam domain were considered domain RBDpep, and RBDpeps without overlap with Pfam domain were considered nondomain RBDpeps. Further, nondomain RBDpeps overlapped with Pfam disorder regions, and CC regions were classified as disorder region RBDpeps and CC RBDpeps, respectively. A manually curated list of published mouse somatic cell RBDpeps (*42*) was used to compare with testis RBDpeps from this study. For sequence motif search in nondomain RBDpeps, multiple sequence alignment (MSA) was built by Clustal Omega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo) with default parameters. Motifs were extracted from align part of MSA and visualized by WebLogo (https://weblogo.threeplusone.com/) with amino acids colored according to their side-chain chemistry.

For sequence motif enrichment in CC regions of mMGC RBPs and nonRBPs, motif was searched using MEME software. For a motif to be defined, we required the occurrence of the same pattern to be more than 20 times and the *P* value to be less than 0.00001.

To test enrichment of bound nondomain RBDpeps within CC, disorder, and low complexity regions, we performed Fisher's exact test using Locus Overlap Enrichment Analysis (LOLA) R package with default setting (Version 1.26.0) (*101*). All nondomain RBDpeps were the userset of LOLA analysis, all bound RBDpeps served as the userUniverse of LOLA analysis, and bound RBDpeps within the disorder regions/CC regions/low complexity regions served as the input file of LOLA analysis.

Scoring evolutionary conservation of protein residues

We obtained the normalized evolutionary conservation score for residues on each RBD-mapped protein using the web-based application ConSurf 2016 (UNIREF-90 protein database with default parameters) (*102*). Local conservation scores were calculated by sliding a 10–amino acid window along each protein and averaging the conservation score.

Screening for ER patches in proteins

We defined two classes of ER patches: at least four Glu and four Arg (4E4R) or at least three Glu and three Arg (3E3R) within a 10–amino acid window, and thus the former was the subset of the latter. To screen for these two classes of ER patch throughout mMGC RBPome and CC sequences of nonRBPome, we extracted subsequences using seqkit (*103*) sliding with step size 1 amino acid and window size 10 amino acids. Then we extracted the subsequences and protein coordinates with 3E3R and 4E4R patches, and merged the overlapping subsequences to obtain the final 3E3R and 4E4R patch within mMGC RBPome and CC sequences of nonRBPome. The procedure for screening ER patches in human orthologous proteins of mMGC RBPome and nonRBPome is the same as that in mMGC RBPome and CC sequences of nonRBPome.

Mapping protein-relative coordinates to genomic regions

We used geno2proteo (*104*) R package to map protein-relative coordinates to genomic regions of human and protein-relative coordinates to genomic regions of mouse. We used UCSC Genome browser software liftover to convert genome coordinates between different species.

Coevolution analysis

For each protein containing defined ER patches, evolutionary couplings between residues were computed by EVcouplings on local and web servers (https://evcouplings.org) (*46*). Residue pairs with stronger co-constraints or interdependence are assigned higher coevolution scores. As suggested (*105*), within each protein we defined strongly evolutionarily coupled pairs (ECs) as the top L ones, where L is the protein sequence length, and then counted the number of ECs within each ER patch. Note that nonRBPs were defined as those in TMT RBPome but out of mMGC RBPome and mSC proteome.

eCLIP-seq data analysis

eCLIP-seq data were analyzed following the ENCODE standard eCLIP pipeline (ENCODE Project Consortium, 2012; https://www.encodeproject.org/) (*106*). On the basis of trimming adapters and 3′ adapter dimers using cutadapt (V1.14), we first dropped out reads aligned to the repeat elements from UCSC with STAR (V2.7.6a) and then mapped the rest to mm10 reference genome (GENCODE vM23). PCR duplicates were removed. CLIPPER was used to call peak regions with fold enrichment ≥ 8 and Padj ≤ 0.001. In the end, we identified significantly enriched peaks as those with irreproducible discovery rate cutoff of 0.01. Binding motif among enriched peaks was predicted by MEME with motif length from five to eight. All annotations referred to GTF file from GENCODE version vM23.

To compare the NONO binding intensity in cellulo and in vivo, we used the htseq-count software (*107*) to quantify gene level count in the eCLIP IP (two replicates in cellulo and three replicates in vivo) and input samples (no replicates). Unlike expression data, eCLIP reads were enriched within a few hundred to 2000 genes, so we normalized the raw count of genes of strong signals (log of count per million ≥10) using the edgeR TMM algorithm to avoid normalization bias introduced by thousands of genes of insufficient coverages. The binding intensity is the log fold change value between the IP and input normalized count.

scRNA-seq data analysis

Data processing was primarily conducted with Seurat R package (version 4.3.0). Poor quality cells were excluded, such as cells with <200 or >8,000 expressed genes (as they are potentially cell doublets). Only genes expressed in three or more cells were used for further analysis. Cells were also discarded if their mitochondrial gene percentages were over 5%.

Raw expression data were normalized and scaled with NormalizeData function, and ScaleData function. The top 1000 genes with the highest standardized variance were identified using FindVariableFeatures function (selection.method = "vst"). Principal components analysis (PCA) was computed using RunPCA function with default parameters.

Shared nearest neighbor (SNN) graph was computed using the FindNeighbors function, taking as input the first 20 PCA dimensions. Cell clusters were defined using Louvain algorithm with the FindCluster function. For visualization uniform manifold approximation and projection (UMAP) was used. Cell type–specific markers were identified by the Seurat FindAllMarkers command, with log fold change threshold of 0.25 and minimum fractional expression threshold of 0.1. For accurate differential expression analysis between conditions, we used the pseudobulk method to aggregate expression of each sample (*108*), because treating each cell as an independent sample can lead to underestimated variance, overly small *P* values, and thus misleading results (*109*). Aggregated expressions were normalized by calcNormFactors function in the edgeR packages, and differentially expressed analyses were performed using the limma package. Developmental tracjectories were predicted by slingshot (*110*). Mitosis-enriched genes were defined as those specifically expressed in SG and preleptotene spermatocytes, while meiosis-enriched genes were those specifically expressed in meiotic spermatocytes.

Whole-exome sequencing data processing

The raw sequencing data were processed using the following steps: (i) removing reads containing sequencing adapter; (ii) removing reads whose low-quality base ratio (base quality ≤5) is

more than 50%; and (iii) removing reads whose unknown base ("N" base) ratio is more than 10%. Statistical analysis of data and downstream bioinformatics analysis were performed on this filtered high-quality data, referred to as the clean data. Clean data were aligned to the human reference genome (GRCh37) using Burrows-Wheeler Aligner (BWA) (version: 0.7.17-r1188) (*111*). Picard was used to remove duplicated sequence reads. Realignment was performed with the Genome Analysis Toolkit (GATK) (*112*). Single-nucleotide polymorphisms (SNPs) and insertions-deletions (InDels) were called using HaplotypeCaller of GATK. SNPs and InDels within low complexity regions (hg19_blackregion.bed) were removed. ANNOVAR was used to annotate the variants for functional and population frequency information (*113*). Deleterious missense variants were predicted with combined annotation-dependent depletion (CADD) (v1.3) score (*114*). CADD_phred > 15 was considered deleterious.

The RefSeq gene model was used throughout this study. Genes listed in TMT-MS data were applied in subsequent analyses. For clinical analysis, we classified genes listed in mMGC RBPome or testis RBDmap as RBP genes; genes listed in TMT-MS data but not listed in mMGC RBPome and RBDmap were classified as nonRBP genes. To assess tissue specificity of human genes, RNA-seq profiling data across 54 human tissues were extracted [https://www.proteinatlas.org. (rna_tissue_consensus.tsv)] to calculate the expression specificity score (SPM) of each RBP gene, as described in TiSGeD (*91*). Genes with SPM score ≥0.9 were classified as testis-specific, genes with SPM score ≥0.5 and <0.9 were classified as testis-enriched, and other genes were classified as ubiquitously expressed.

To examine the population structure and check sample anomaly of the patient and control groups, we performed PCA based on the genotypes of the called variants in all samples in this study and in hapmap3 (https://www.broadinstitute.org/medical-and-population-genetics/hapmap-3) using plink software (*115*) and visualized the first two components in a two-dimensional space. We performed relatedness examination by KING(v.2.2.4) (*116*) and showed all samples have no relatedness to each other in the dataset.

Candidate risk variants and gene filtering

Candidate causative rare loss of function (LOF) variants were defined using the following criteria: (i) filter is "PASS", read depth ≥ 10, GQ ≥ 20; alternative read depth/ read depth ≥ 0.2; (ii) allele frequencies in 1000 g2015aug_eas, ExAC_EAS, gnomAD_genome_EAS, and gnomAD_exome_EAS public databases were lower than 0.01 or not documented; (iii) frameshifts, premature stop gains or the disruption of core splice sites. Rare LOF pathogenic mutation genes with no LOF and rare DNS variant genes hit in control individual were defined as rare LOF pathogenic mutation genes.

Candidate causative rare damaging nonsynonymous (DNS) variants were defined using the following criteria: (i) filter is "PASS", read depth ≥ 10, GQ ≥ 20; alternative read depth/ read depth ≥ 0.2; (ii) allele frequencies in 1000 g2015aug_eas, ExAC_EAS, gnomAD_genome_EAS, and gnomAD_exome_EAS public databases were lower than 0.01 or not documented; (iii) "nonsynonymous SNV". Rare DNS pathogenic mutation genes were defined as Rare DNS variants with a Phred-scaled CADD pathogenicity prediction score > 15 and no rare DNS mutation genes hit in control individual.

Prioritization of the genetic components more relevant to the pathogenesis

We classified candidate variants into the following: homozygote/hemizygote (Hom) LOF variants of RBP, heterozygote (Het) LOF variants of RBP, Hom DNS variants of RBP, Het DNS variants of RBP, domain region DNS of RBP, nondomain region DNS of RBP, CC region DNS of RBP, nonCC region DNS of RBP, ER patch DNS of RBP, CC-nonER region DNS of RBP, and among others belonging to nonRBP.

We performed Fisher's exact test to test if the odds of finding the following type of candidate variants in the patients is significantly higher than in the controls: (i) Hom LOF variants (vs. Het LOF variants); (ii) Hom LOF variants in RBPs (vs. Hom LOF variants in nonRBPs); (iii) Het LOF variants in RBPs (vs. Het LOF variants in nonRBPs); (iv) Hom DNS variants (vs. Het DNS variants); (v) Hom DNS variants in RBPs (vs. Hom DNS variants in nonRBPs); (vi) Het DNS variants in RBPs (vs. Het DNS variants in nonRBPs); (vii) all DNS variants in domains of RBPs (vs. in nondomain regions of RBPs); (viii) all DNS variants in CC regions of RBPs (vs. in nonCC regions of RBPs); (ix) all DNS variants in CC-ER regions of RBPs (vs. in CC-nonER regions of RBPs). At the gene level [(i) to (vi)], we selected DNS variants predicted "deleterious" by any of the five algorithms aforementioned. At the subelement level [(vii) to (ix)], since Hom and Het variants were combined, we applied a more stringent criterion to select DNS variants, which should be predicted "deleterious" by at least three of the five algorithms.

REFERENCES AND NOTES

1. P. Sassone-Corsi, Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176–2178 (2002). doi: 10.1126/science.1070963; pmid: 12077401
2. S. S. Hammoud *et al*., Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* **15**, 239–253 (2014). doi: 10.1016/j.stem.2014.04.006; pmid: 24835570
3. N. Kotaja, P. Sassone-Corsi, The chromatoid body: A germ-cell-specific RNA-processing centre. *Nat. Rev. Mol. Cell Biol.* **8**, 85–90 (2007). doi: 10.1038/nrm2081; pmid: 17183363
4. D. Brawand *et al*., The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011). doi: 10.1038/nature10532; pmid: 22012392
5. M. Soumillon *et al*., Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013). doi: 10.1016/j.celrep.2013.05.031; pmid: 23791531
6. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014). doi: 10.1038/nrg3813; pmid: 25365966
7. R. Xiao *et al*., Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107–121.e18 (2019). doi: 10.1016/j.cell.2019.06.001; pmid: 31251911
8. R. K. Idler, W. Yan, Control of messenger RNA fate by RNA-binding proteins: An emphasis on mammalian spermatogenesis. *J. Androl.* **33**, 309–337 (2012). doi: 10.2164/jandrol.111.014167; pmid: 21757510
9. P. Dai *et al*., A Translation-Activating Function of MIWI/piRNA during Mouse Spermiogenesis. *Cell* **179**, 1566–1581.e16 (2019). doi: 10.1016/j.cell.2019.11.022; pmid: 31835033
10. A. Castello *et al*., Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012). doi: 10.1016/j.cell.2012.04.031; pmid: 22658674
11. A. G. Baltz *et al*., The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012). doi: 10.1016/j.molcel.2012.05.021; pmid: 22681889
12. M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018). doi: 10.1038/nrm.2017.130; pmid: 29339797
13. S. C. Kwon *et al*., The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130 (2013). doi: 10.1038/nsmb.2638; pmid: 23912277
14. B. M. Beckmann *et al*., The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.* **6**, 10127 (2015). doi: 10.1038/ncomms10127; pmid: 26632259
15. V. O. Sysoev *et al*., Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nat. Commun.* **7**, 12128 (2016). doi: 10.1038/ncomms12128; pmid: 27378189
16. H. H. Wessels *et al*., The mRNA-bound proteome of the early fly embryo. *Genome Res.* **26**, 1000–1009 (2016). doi: 10.1101/gr.200386.115; pmid: 27197210
17. J. I. Perez-Perri *et al*., The RNA-binding protein landscapes differ between mammalian organs and cultured cells. *Nat. Commun.* **14**, 2074 (2023). doi: 10.1038/s41467-023-37494-w; pmid: 37045843
18. A. Castello *et al*., Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696–710 (2016). doi: 10.1016/j.molcel.2016.06.029; pmid: 27453046
19. A. Castello *et al*., Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. *Nat. Protoc.* **12**, 2447–2464 (2017). doi: 10.1038/nprot.2017.106; pmid: 29095441
20. M. Corley, M. C. Burns, G. W. Yeo, How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol. Cell* **78**, 9–29 (2020). doi: 10.1016/j.molcel.2020.03.011; pmid: 32243832
21. J. Y. Kang *et al*., LLPS of FXR1 drives spermiogenesis by activating translation of stored mRNAs. *Science* **377**, eabj6647 (2022). doi: 10.1126/science.abj6647 pmid: 35951695
22. J. W. Schneider *et al*., Dysregulated ribonucleoprotein granules promote cardiomyopathy in RBM20 gene-edited pigs. *Nat. Med.* **26**, 1788–1800 (2020). doi: 10.1038/s41591-020-1087-x; pmid: 33188278
23. F. Gebauer, T. Schwarzl, J. Valcárcel, M. W. Hentze, RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22**, 185–198 (2021). doi: 10.1038/s41576-020-00302-y; pmid: 33235359
24. C. Krausz, A. Riera-Escamilla, Genetics of male infertility. *Nat. Rev. Urol.* **15**, 369–384 (2018). doi: 10.1038/s41585-018-0003-3; pmid: 29622783
25. L. Nagirnaja *et al*., Variant *PNLDC1*, Defective piRNA Processing, and Azoospermia. *N. Engl. J. Med.* **385**, 707–719 (2021). doi: 10.1056/NEJMoa2028973; pmid: 34347949
26. M. S. Oud *et al*., A de novo paradigm for male infertility. *Nat. Commun.* **13**, 154 (2022). doi: 10.1038/s41467-021-27132-8; pmid: 35013161
27. B. J. Houston *et al*., A systematic review of the validated monogenic causes of human male infertility: 2020 update and a discussion of emerging gene-disease relationships. *Hum. Reprod. Update* **28**, 15–29 (2021). doi: 10.1093/humupd/dmab030; pmid: 34498060

28. C. Coutton, J. Escoffier, G. Martinez, C. Arnoult, P. F. Ray, Teratozoospermia: Spotlight on the main genetic actors in the human. *Hum. Reprod. Update* **21**, 455–485 (2015). doi: 10.1093/humupd/dmv020; pmid: 25888788
29. Q. Sang, P. F. Ray, L. Wang, Understanding the genetics of human infertility. *Science* **380**, 158–163 (2023). doi: 10.1126/science.adf7760; pmid: 37053320
30. J. Y. Liao *et al*., EuRBPDB: A comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). *Nucleic Acids Res.* **48**, D307–D313 (2020). doi: 10.1093/nar/gkz823; pmid: 31598693
31. A. Panhale *et al*., CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nat. Commun.* **10**, 2682 (2019). doi: 10.1038/s41467-019-10585-3; pmid: 31213602
32. C. Asencio, A. Chatterjee, M. W. Hentze, Silica-based solid-phase extraction of cross-linked nucleic acid-bound proteins. *Life Sci. Alliance* **1**, e201800088 (2018). doi: 10.26508/lsa.201800088; pmid: 30035255
33. H. Li *et al*., A male germ-cell-specific ribosome controls male fertility. *Nature* **612**, 725–731 (2022). doi: 10.1038/s41586-022-05508-0; pmid: 36517592
34. K. Li *et al*., Panoramic transcriptome analysis and functional screening of long noncoding RNAs in mouse spermatogenesis. *Genome Res.* **31**, 13–26 (2021). doi: 10.1101/gr.264333.120; pmid: 33328167
35. N. Iguchi, J. W. Tobias, N. B. Hecht, Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7712–7717 (2006). doi: 10.1073/pnas.0510999103; pmid: 16682651
36. O. Meikar *et al*., An atlas of chromatoid body components. *RNA* **20**, 483–495 (2014). doi: 10.1261/rna.043729.113; pmid: 24554440
37. C. Kilchert *et al*., System-wide analyses of the fission yeast poly(A)$^+$ RNA interactome reveal insights into organization and function of RNA-protein complexes. *Genome Res.* **30**, 1012–1026 (2020). doi: 10.1101/gr.257006.119; pmid: 32554781
38. M. Garcia-Moreno *et al*., System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Mol. Cell* **74**, 196–211.e11 (2019). doi: 10.1016/j.molcel.2019.01.017; pmid: 30799147
39. A. I. Järvelin, M. Noerenberg, I. Davis, A. Castello, The new (dis)order in RNA regulation. *Cell Commun. Signal.* **14**, 9 (2016). doi: 10.1186/s12964-016-0132-3; pmid: 27048167
40. R. van der Lee *et al*., Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631 (2014). doi: 10.1021/cr400525m; pmid: 24773235
41. D. M. Passon *et al*., Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4846–4850 (2012). doi: 10.1073/pnas.1120792109; pmid: 22416126
42. Y. Liao *et al*., The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Rep.* **16**, 1456–1469 (2016). doi: 10.1016/j.celrep.2016.06.084; pmid: 27452465
43. I. Anishchenko, S. Ovchinnikov, H. Kamisetty, D. Baker, Origins of coevolution between residues distant in protein 3D structures. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9122–9127 (2017). doi: 10.1073/pnas.1702664114; pmid: 28784799
44. S. S. Choi, W. Li, B. T. Lahn, Robust signals of coevolution of interacting residues in mammalian proteomes identified by phylogeny-aided structural analysis. *Nat. Genet.* **37**, 1367–1371 (2005). doi: 10.1038/ng1685; pmid: 16282975
45. F. Pazos, A. Valencia, Protein co-evolution, co-adaptation and interactions. *EMBO J.* **27**, 2648–2655 (2008). doi: 10.1038/emboj.2008.189; pmid: 18818697
46. T. A. Hopf *et al*., The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**, 1582–1584 (2019). doi: 10.1093/bioinformatics/bty862; pmid: 30304492
47. T. Yamazaki *et al*., Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol. Cell* **70**, 1038–1053.e7 (2018). doi: 10.1016/j.molcel.2018.05.019; pmid: 29932899
48. S. Li *et al*., Cell-type specific role of the RNA-binding protein, NONO, in the DNA double-strand break response in the mouse testes. *DNA Repair* **51**, 70–78 (2017). doi: 10.1016/j.dnarep.2017.02.002; pmid: 28209515
49. M. Ivanova *et al*., Scaffold attachment factor B1 functions in development, growth, and reproduction. *Mol. Cell. Biol.* **25**, 2995–3006 (2005). doi: 10.1128/MCB.25.8.2995-3006.2005; pmid: 15798188
50. P. Chakraborty *et al*., LIN28A marks the spermatogonial progenitor population and regulates its cyclic expansion. *Stem Cells* **32**, 860–873 (2014). doi: 10.1002/stem.1584; pmid: 24715688
51. B. P. Hermann *et al*., The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Rep.* **25**, 1650–1667.e8 (2018). doi: 10.1016/j.celrep.2018.10.026; pmid: 30404016
52. C. D. Green *et al*., A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Dev. Cell* **46**, 651–667.e10 (2018). doi: 10.1016/j.devcel.2018.07.025; pmid: 30146481
53. L. Li *et al*., The XRN1-regulated RNA helicase activity of YTHDC2 ensures mouse fertility independently of m^{6}A recognition. *Mol. Cell* **82**, 1678–1690.e12 (2022). doi: 10.1016/j.molcel.2022.02.034; pmid: 35305312
54. H. Tan *et al*., Single-cell RNA-seq uncovers dynamic processes orchestrated by RNA-binding protein DDX43 in chromatin remodeling during spermiogenesis. *Nat. Commun.* **14**, 2499 (2023). doi: 10.1038/s41467-023-38199-w; pmid: 37120627
55. L. T. Gou *et al*., Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. *Cell* **169**, 1090–1104.e13 (2017). doi: 10.1016/j.cell.2017.04.034; pmid: 28552346
56. K. J. Karczewski *et al*., The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020). doi: 10.1038/s41586-020-2308-7; pmid: 32461654
57. J. Lykke-Andersen, M. D. Shu, J. A. Steitz, Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**, 1121–1131 (2000). doi: 10.1016/S0092-8674(00)00214-2; pmid: 11163187
58. C. C. MacDonald, P. N. Grozdanov, Nonsense in the testis: Multiple roles for nonsense-mediated decay revealed in male reproduction. *Biol. Reprod.* **96**, 939–947 (2017). doi: 10.1093/biolre/iox033; pmid: 28444146
59. J. C. Bufton *et al*., Structures of nonsense-mediated mRNA decay factors UPF3B and UPF3A in complex with UPF2 reveal molecular basis for competitive binding and for neurodevelopmental disorder-causing mutation. *Nucleic Acids Res.* **50**, 5934–5947 (2022). doi: 10.1093/nar/gkac421; pmid: 35640974
60. W. H. Hudson, E. A. Ortlund, The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014). doi: 10.1038/nrm3884; pmid: 25269475
61. T. Conrad *et al*., Serial interactome capture of the human cell nucleus. *Nat. Commun.* **7**, 11212 (2016). doi: 10.1038/ncomms11212; pmid: 27040163
62. T. A. Soboleva *et al*., A new link between transcriptional initiation and pre-mRNA splicing: The RNA binding histone variant H2A.B. *PLOS Genet.* **13**, e1006633 (2017). doi: 10.1371/journal.pgen.1006633; pmid: 28234895
63. J. Trendel *et al*., The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391–403.e19 (2019). doi: 10.1016/j.cell.2018.11.004; pmid: 30528433
64. L. K. Ford, L. Fioriti, Coiled-Coil Motifs of RNA-Binding Proteins: Dynamicity in RNA Regulation. *Front. Cell Dev. Biol.* **8**, 607947 (2020). doi: 10.3389/fcell.2020.607947; pmid: 33330512
65. H. Lyons *et al*., Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell* **186**, 327–345.e28 (2023). doi: 10.1016/j.cell.2022.12.013; pmid: 36603581
66. J. A. Greig *et al*., Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol. Cell* **77**, 1237–1250.e4 (2020). doi: 10.1016/j.molcel.2020.01.025; pmid: 32048997
67. R. K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13392–13397 (2013). doi: 10.1073/pnas.1304749110; pmid: 23901099
68. S. Zhang *et al*., NONO enhances mRNA processing of super-enhancer-associated GATA2 and HAND2 genes in neuroblastoma. *EMBO Rep.* **24**, e54977 (2023). doi: 10.15252/embr.202254977; pmid: 36416237
69. K. V. Prasanth *et al*., Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249–263 (2005). doi: 10.1016/j.cell.2005.08.033; pmid: 16239143
70. G. Benegiamo *et al*., The RNA-Binding Protein NONO Coordinates Hepatic Adaptation to Feeding. *Cell Metab.* **27**, 404–418.e7 (2018). doi: 10.1016/j.cmet.2017.12.010; pmid: 29358041
71. D. A. Parry, R. D. Fraser, J. M. Squire, Fifty years of coiled-coils and alpha-helical bundles: A close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258–269 (2008). doi: 10.1016/j.jsb.2008.01.016; pmid: 18342539
72. G. J. Knott, C. S. Bond, A. H. Fox, The DBHS proteins SFPQ, NONO and PSPC1: A multipurpose molecular scaffold. *Nucleic Acids Res.* **44**, 3989–4004 (2016). doi: 10.1093/nar/gkw271; pmid: 27084935
73. M. Lee *et al*., The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res.* **43**, 3826–3840 (2015). doi: 10.1093/nar/gkv156; pmid: 25765647
74. A. Castello, B. Fischer, M. W. Hentze, T. Preiss, RNA-binding proteins in Mendelian disease. *Trends Genet.* **29**, 318–327 (2013). doi: 10.1016/j.tig.2013.01.004; pmid: 23415593
75. B. Shi *et al*., Biallelic mutations in RNA-binding protein ADAD2 cause spermiogenic failure and non-obstructive azoospermia in humans. *Hum. Reprod. Open* **2023**, hoad022 (2023). doi: 10.1093/hropen/hoad022; pmid: 37325547
76. M. K. O'Bryan *et al*., RBM5 is a male germ cell splicing factor and is required for spermatid differentiation and male fertility. *PLOS Genet.* **9**, e1003628 (2013). doi: 10.1371/journal.pgen.1003628; pmid: 23935508
77. A. R. Bellvé, Purification, culture, and fractionation of spermatogenic cells. *Methods Enzymol.* **225**, 84–113 (1993). doi: 10.1016/0076-6879(93)25009-Q; pmid: 8231890
78. H. Gan *et al*., Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun.* **4**, 1995 (2013). doi: 10.1038/ncomms2995; pmid: 23759713
79. A. Castello *et al*., System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491–500 (2013). doi: 10.1038/nprot.2013.020; pmid: 23411631
80. K. Fu *et al*., Biological and RNA regulatory function of MOV10 in mammalian germ cells. *BMC Biol.* **17**, 39 (2019). doi: 10.1186/s12915-019-0659-z; pmid: 31088452
81. E. L. Van Nostrand *et al*., Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016). doi: 10.1038/nmeth.3810; pmid: 27018577
82. Q. Xu *et al*., Enhanced Crosslinking Immunoprecipitation (eCLIP) Method for Efficient Identification of Protein-bound RNA in Mouse Testis. *J. Vis. Exp.* **147**, e59681 (2019). doi: 10.3791/59681-v; pmid: 31132071
83. L. Yu *et al*., In Vitro Biochemical Assays using Biotin Labels to Study Protein-Nucleic Acid Interactions. *J. Vis. Exp.* **149**, e59830 (2019). doi: 10.3791/59830; pmid: 31380834
84. A. Agarwal *et al*., Male infertility. *Lancet* **397**, 319–333 (2021). doi: 10.1016/S0140-6736(20)32667-2; pmid: 33308486
85. World Health Organization, *WHO Laboratory Manual for the Examination and Processing of Human Semen* (World Health Organization, ed. 5, 2010), p. xiv.
86. Practice Committee of the American Society for Reproductive Medicine, Management of nonobstructive azoospermia: A committee opinion. *Fertil. Steril.* **110**, 1239–1245 (2018). doi: 10.1016/j.fertnstert.2018.09.012
87. H. Tournaye, C. Krausz, R. D. Oates, Concepts in diagnosis and therapy for male reproductive impairment. *Lancet Diabetes Endocrinol.* **5**, 554–564 (2017). doi: 10.1016/S2213-8587(16)30043-2; pmid: 27395770
88. P. N. Schlegel *et al*., Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. *Fertil. Steril.* **115**, 54–61 (2021). doi: 10.1016/j.fertnstert.2020.11.015; pmid: 33309062
89. C. Tan *et al*., Bi-allelic variants in DNHD1 cause flagellar axoneme defects and asthenoteratozoospermia in humans and mice. *Am. J. Hum. Genet.* **109**, 157–171 (2022). doi: 10.1016/j.ajhg.2021.11.022; pmid: 34932939
90. B. Li *et al*., A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. *Sci. Rep.* **7**, 4200 (2017). doi: 10.1038/s41598-017-04520-z; pmid: 28646208
91. S. J. Xiao, C. Zhang, Q. Zou, Z. L. Ji, TiSGeD: A database for tissue-specific genes. *Bioinformatics* **26**, 1273–1275 (2010). doi: 10.1093/bioinformatics/btq109; pmid: 20223836
92. S. Durinck *et al*., BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005). doi: 10.1093/bioinformatics/bti525; pmid: 16082012
93. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: An R package for comparing biological themes among gene

clusters. *OMICS* **16**, 284–287 (2012). doi: 10.1089/omi.2011.0118; pmid: 22455463
94. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
95. D. Szklarczyk *et al*., The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49** (D1), D605–D612 (2021). doi: 10.1093/nar/gkaa1074; pmid: 33237311
96. Y. Zhou *et al*., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019). doi: 10.1038/s41467-019-09234-6; pmid: 30944313
97. L. Kumar, M. E Futschik, Mfuzz: A software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007). doi: 10.6026/97320630002005; pmid: 18084642
98. Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347**, 827–839 (2005). doi: 10.1016/j.jmb.2005.01.071; pmid: 15769473
99. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008). doi: 10.1038/nbt.1511; pmid: 19029910
100. J. Mistry *et al*., Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021). doi: 10.1093/nar/gkaa913; pmid: 33125078
101. N. C. Sheffield, C. Bock, LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016). doi: 10.1093/bioinformatics/btv612; pmid: 26508757
102. H. Ashkenazy *et al*., ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016). doi: 10.1093/nar/gkw408; pmid: 27166375
103. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**, e0163962 (2016). doi: 10.1371/journal.pone.0163962; pmid: 27706213
104. Y. Li, E. Aguilar-Martinez, A. D. Sharrocks, Geno2proteo, a Tool for Batch Retrieval of DNA and Protein Sequences from Any Genomic or Protein Regions. *J. Integr. Bioinform.* **16**, 20180090 (2019). doi: 10.1515/jib-2018-0090; pmid: 31301672
105. A. G. Green *et al*., Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat. Commun.* **12**, 1396 (2021). doi: 10.1038/s41467-021-21636-z; pmid: 33654096
106. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
107. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi: 10.1093/bioinformatics/btu638; pmid: 25260700
108. K. D. Zimmerman, M. A. Espeland, C. D. Langefeld, A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 (2021). doi: 10.1038/s41467-021-21038-1; pmid: 33531494
109. C. Soneson, M. D. Robinson, Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018). doi: 10.1038/nmeth.4612; pmid: 29481549
110. K. Street *et al*., Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). doi: 10.1186/s12864-018-4772-0; pmid: 29914354
111. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010). doi: 10.1093/bioinformatics/btp698; pmid: 20080505
112. A. McKenna *et al*., The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010). doi: 10.1101/gr.107524.110; pmid: 20644199
113. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010). doi: 10.1093/nar/gkq603; pmid: 20601685
114. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47** (D1), D886–D894 (2019). doi: 10.1093/nar/gky1016; pmid: 30371827
115. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
116. A. Manichaikul *et al*., Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010). doi: 10.1093/bioinformatics/btq559; pmid: 20926424

ACKNOWLEDGMENTS

We thank M. Hentze (European Molecular Biology Laboratory, Germany) for kind guidance with the RBDmap experiment; A. Akhtar and A. Panhale (Max Planck Institute of Immunobiology and Epigenetics, Germany) for kind guidance with the 2C assay; X. Wu and M. Wang (Nanjing Medical University) for sharing their initial experience in RIC experiments; D. Zheng (Albert Einstein College of Medicine, USA) for bioinformatics guidance; B. Zhou (University of Chicago, USA) for advice on revision; and P. J. Wang (University of Pennsylvania, USA), R. S. Pillai (University of Geneva, Switzerland), M.-F. Liu (Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences), and E. Duan (Institute of Zoology, Chinese Academy of Sciences) for helpful comments. We thank several students from L. Ye's and K. Zheng's labs (Nanjing Medical University, China) for their co-participation in experiments, especially the collection of mMGCs, and the students from Y.-Q. Tan's lab (Central South University, China) for their genomic data analysis of the male infertility cohort. We greatly appreciate the support of high-performance computing from X. Kong (The First Affiliated Hospital of Nanjing Medical University, China) and the long-term support to this project from J. Sha and Z. Hu (Nanjing Medical University, China). **Funding:** This work was supported by the National Key R&D Program of China [2016YFA0500902 (K.Z.), 2022YFC2703500 (L.Y.), 2021YFC2700200 (X.-J.G.), 2022YFC2702604 (Y.-Q.T), and 2018YFC1003500 (K.Z.)], the National Natural Science Foundation of China [82221005 (X.-J.G.), 32170858 (K.Z.), 31771653 (K.Z), U23A20402 (Y.-Q.T), 32071133 (Y.-S.G.) and 82001613 (Y.L.)], the Excellent Youth Foundation of the Jiangsu Scientific Committee [BK20211532 (L.Y.) and BK20150047 (K.Z.)], the Postdoctoral Research Foundation of China [2020M671540 (Y.L.)], and by grants (K.Z.) from the State Key Laboratory of Reproductive Medicine and from the School of Public Health of Nanjing Medical University. **Author contributions:** K.Z., X.-J.G., and M.-Y.L. conceptualized and designed this work. Y.L., Y.-Y.W., Y.-Q.T., Q.-L.Y., L.Y., X.-J.G., M.-Y.L., and K.Z. designed experiments and/or analyses. Y.L., Q.-L.Y., H.-C.Z., L.-X.T., Z.-H.Y., W.L., C.-C.W., S.-L.H., S.R., P.-L.Y., W.-Y.J., Y.-T.Y., H.-H.T., Y.-F.W., C.-Y.L., X.L., J.X., G.-G.Y., and K.Z. carried out experiments. Y.-S.G., H.Z., and X.-J.G. conducted mass spectrometry and proteomics profiling. Y.-Q.T., L.-L.M., C.T., T.-Y.H., and G.L. collected human clinical specimen and provided whole-exome sequencing profiles. Y.-Q.C. and B.S. assisted in the generation of CRISPR/Cas9 mouse models. Y.-Q.C. prepared scRNA-seq libraries. Y.-Y.W., R.-Y.Y., L.G., and M.-Y.L. performed bioinformatics analyses. M.-Y.L., Y.-Y.W., Y.L., L.-Y.L., and C.W. analyzed whole-exome sequencing data. Y.-T.Y., W.-X.W., Y.L., Y.-Y.W., and M.-Y.L. analyzed scRNA-seq data. A.V., Y.L., L.Y., Y.-Q.T., X.-J.G., M.-Y.L., and K.Z. discussed and interpreted results. Experimental and bioinformatics parts are equally important to this work, and M.-Y.L. supervised most bioinformatics studies. Y.L. and K.Z. analyzed most experiments and compiled figures. K.Z., Y.L., and L.-X.T. designed the print summary figure. Z.S., X.-J.G., Y.-Q.T., W.L., Q.-H.S., Y.-C.G., and T.N. contributed additional specialized guidance, advice, and edits. K.Z., Y.L., M.-Y.L., Y.-Y.W, L.Y., and A.V. wrote the manuscript incorporating inputs from all authors. K.Z. organized collaboration for the overall project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The RIC-MS data and RBDmap-MS data were deposited at the ProteomeXchange Consortium through the Proteomics Identifications Database (PRIDE) under project accession nos. PXD043074 and PXD043242. The eCLIP-seq data of TEX30, PDHA2, HA-NONO, and HA-NONO$^{\Delta ER}$ were deposited in the Gene Expression Omnibus (GEO) database under project accession no. GSE232595. The eCLIP-seq data of NONO in *Nono*$^{\Delta ER/Y}$ and *Nono*$^{+/Y}$ testes were deposited in the GEO database under project accession no. GSE263329. The scRNA-seq data of *Nono*$^{\Delta ER/Y}$ and *Nono*$^{+/Y}$ testes were deposited in the GEO database under project accession no. GSE263407.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj8172
Materials and Methods
Supplementary Text
Figs. S1 to S23
Tables S1 to S16
References (*117*–*160*)

Submitted 19 July 2023; resubmitted 8 April 2024
Accepted 20 August 2024
Published online 29 August 2024
10.1126/science.adj8172