Taylor & Francis
Taylor & Francis Group

Cell Cycle

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/kccy20

Landscape of RNA editing reveals new insights into the dynamic gene regulation of spermatogenesis

Xiaodan Wang, Xiaolong Wu, Zhenshuo Zhu, Hao Li, Tongtong Li, Qun Li, Peng Zhang, Leijie Li, Dongxue Che, Xia Xiao, Tong Liu, Jinlian Hua & Mingzhi Liao

To cite this article: Xiaodan Wang, Xiaolong Wu, Zhenshuo Zhu, Hao Li, Tongtong Li, Qun Li, Peng Zhang, Leijie Li, Dongxue Che, Xia Xiao, Tong Liu, Jinlian Hua & Mingzhi Liao (2019) Landscape of RNA editing reveals new insights into the dynamic gene regulation of spermatogenesis, Cell Cycle, 18:23, 3351-3364, DOI: 10.1080/15384101.2019.1676584

To link to this article: https://doi.org/10.1080/15384101.2019.1676584

CELL CYCLE
2019, VOL. 18, NO. 23, 3351–3364
https://doi.org/10.1080/15384101.2019.1676584

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Landscape of RNA editing reveals new insights into the dynamic gene regulation of spermatogenesis

Xiaodan Wang[a*], Xiaolong Wu[b*], Zhenshuo Zhu[b], Hao Li[a], Tongtong Li[a], Qun Li [a], Peng Zhang[a], Leijie Li[a], Dongxue Che[a], Xia Xiao[a], Tong Liu[c,d], Jinlian Hua[b], and Mingzhi Liao[a]

[a]College of Life Sciences, Northwest A&F University, Yangling, China; [b]College of Veterinary Medicine, Shaanxi Centre of Stem Cells Engineering & Technology, Northwest A&F University, Yangling, China; [c]Chinese Academy of Sciences Center for Excellence in Molecular Plant Sciences, Shanghai Center for Plant Stress Biology, Chinese Academy of Sciences, Shanghai, China; [d]University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Spermatogenesis is an important physiological process associated with male infertility. As a kind of post-transcriptional regulation, RNA editings (REs) change the genetic information at the mRNA level. But whether there are REs and what's the role of REs during the process are still unclear. In this study, we integrated published RNA-Seq datasets and established a landscape of RNA REs during the development of mouse spermatogenesis. Totally, 7530 editing sites occurred in 2012 genes among all types of male germ cells were found, these sites enrich on some regions of chromosomes, including chromosome 17 and both ends of chromosome Y. We also found about half of the REs in CDSs can cause amino acids changes. Some non-synonymous REs which exist in specific genes may play important roles in spermatogenesis. Finally, we verified a non-synonymous A-to-I RNA editing site in *Cog3* and a stoploss editing in *Tssk6* during spermatogenesis. In short, we systematically analyzed the dynamic landscape of RNA editing at different stages of spermatogenesis.

ARTICLE HISTORY
Received 8 June 2019
Revised 16 September 2019
Accepted 21 September 2019

KEYWORDS
Spermatogenesis; RNA editing; Integrated RNA-Seq; conserved editing site; *Tssk6*

Introduction

Infertility is gaining increasing attention which affects about 15% couples all over the world, among which male factor is responsible, alone or in combination with female factors in about half of the cases [1–3]. Dysfunction of sperms may arise from different factors, including lifestyle, heredity, obesity, drugs, and so on. Among them, genetic abnormalities account for 15–30% of male infertility [4,5]. Genetics lead to male infertility by influencing many physiological processes, such as spermatogenesis, sperm quality and hormonal homeostasis. Therefore, revealing the molecular mechanisms of spermatogenesis is essential for the understanding of male infertility. Spermatogenesis is a complex and dynamic process leading to the continuous production of sperm [6]. This process requires the successive and coordinated expression of thousands of genes, mixed with multi-level regulations from transcriptional, post-transcriptional and translational gene regulation [7].

As a kind of post-transcriptional regulation, RNA editings (REs) change the genetic information at the mRNA level, providing flexible regulation model from mRNA to proteins and making organisms to respond environment changes quickly and easily. Transcripts of some genes should be edited to initiate an effective translation according to the need of specific tissues and developmental times. Most of the previous studies on REs focused on A-to-I editing, which is most common in humans and other primates and is often associated with cancer [8–10]. The essence of the A-to-I RNA editing is that the adenines on the double-stranded RNAs are converted to hypoxanthines by oxidative deamination reaction under the catalysis of adenosine deaminase ADARs [11]. Generally, REs occur more frequently in non-coding than in coding regions. In human, RNA editing mainly occurs in introns and

CONTACT Mingzhi Liao liaomingzhi83@163.com
*Xiaodan Wang and Xiaolong Wu contributed equally to this work
Supplemental data for this article can be accessed here.

5'- or 3'-untranslated regions (UTRs) [12–14]. Considering the complex and dynamics of spermatogenesis and the flexible regulation manner of REs, we supposed there may be close relationships between them. However, whether there are RNA-editing events in the spermatogenesis. If they exist, how REs distribute in male germ cells and what roles they play in spermatogenesis? These problems are still unclear and need to be analyzed and resolved. It has been proved that REs can make individuals better adapt to environmental changes, but abnormalities in editing level often lead to disease occurrence [15,16]. Our aim is to analyze the profile of REs during normal spermatogenesis, focusing on editing sites that occur on key reproductive genes, as abnormalities in RNA-editing levels on these genes may disrupt the performance of their normal functions, leading to Infertility, especially those that occur on the gene stop codons.

To address the above problems, we integrated all published RNA-Seq datasets and calling REs related with different types of cells during spermatogenesis. Interesting, we found there are huge amount of REs in spermatogenesis and the dominant types of REs are not only A-to-I, but also G-to-A, C-to-U and U-to-C editing in male germ cell process of mouse. Different with human, we found REs mainly occur in introns, CDSs and intergenic regions in mouse. Even more, about half of the editing events in CDSs result in amino acid changes. We also found several novel genes with non-synonymous REs are related with different male germ cells, including *Rnf17, Boll, Adad1* and *Rbmy*. As a typical example, we found non-synonymous REs in *Rbmy*, which is specific expressed in the adult testis. In addition to these genes that play an important role in spermatogenesis, we also found and validated a conserved non-synonymous A-to-I editing site between species in the gene Cog3 in germ cells. It has been reported that in human, rhesus and mouse brains, the same non-synonymous A-to-I RNA editing occurs in the gene Cog3 [17,18]. Our results supplement expand that this non-synonymous RNA editing in Cog3 is not only conserved between species but also between tissues. In summary, based on our integrated analysis and landscape of REs, we found that REs may play important roles during spermatogenesis.

Results

Data preprocessing

The framework of our works was shown as Figure 1. We collected published 12 studies of RNA-Seq datasets about spermatogenesis from Gene Expression Omnibus (GEO), Sequence Read Archive (SRA). These datasets covered 100 samples and were all sequenced with illumina machines platform. In addition to the previous 100 samples, we also collected the latest published datasets under accession number GSE75826 as validation data [19]. Our integration strategy has increased the depth of sequencing to some extent, while at the same time obtaining consistent editing sites from different samples. Compared to detecting RNA-editing sites in each sample separately, this method may improve accuracy. After data preprocessing, we identified the nucleotide changes and removed sites with overlap regions of positive and negative chains. Finally, we eliminated single-nucleotide polymorphisms (SNPs) and determined the final list of REs sites. Comparing the results of the integrated and non-integrated samples under the same identify process, we found that fewer editing sites identified from integrated samples than the unintegrated samples, but the editing sites identified by the integration method were consistent across samples. What's more, some editing sites cannot be detected in single samples due to the low sample sequencing depth, but can be identified in the integrated samples. Calculate the Simpson's diversity index of integration and non-integration results, with a minimum of 0.65 and a maximum of 0.98 (Supplemental Fig S1).

Identifying editing sites in specific cells during spermatogenesis

Based on the developmental stages, cells in testis were divided into nine different kinds (eight kinds of male germ cells and sertoli cells as control) (Supplemental Table S1), including spermatogonia (SG), leptotene spermatocytes (lepSC), pachytene spermatocytes (pacSC), primary spermatocytes before meiosis I (PSC), secondary spermatocytes (SSC), round spermatids (RS), elongative spermatids (ES) and sperm (SP). Results showed that the number of REs in sertoli cells (SC) is significantly larger than in germ cells (Figure 2(a); Table 1). It may be due to the activation of RNA synthesis in sertoli cells. On the other hand,

Figure 1. Experiment process. The whole experiment can be divided into four parts: data preprocessing, the identification of editing sites, genetic variants annotation and getting cell-specific editing sites.

the meiosis process germ cells will present chromatin condensation, diploid into haploid, nuclear degeneration, which will affect the gene expression, reduce the amount of RNA and even terminate RNA synthesis. When the sperm cells begin to elongate, RNA synthesis gradually ceases, and there is no RNA synthesis in mature spermatozoa. A considerable proportion of the RNA synthesized during the pachytene stage is preserved through spermatid development until late spermiogenesis [20–22]. Therefore, we did not analyze the RNA editing sites identified in mature spermatozoa in the subsequent analysis. Finally, 7530 editing sites among other seven types of male germ cells were found (Figure 2(a); Table 1). According to the editing level distribution of the editing sites in each cell, male germ cells have higher RNA editing level than sertoli cells (Figure 2(b); Figure 2(c)). By the way, though the samples' numbers of different kinds of cells were different, there was not significant association with the number of editing sites (Figure 2(d)), which indicates the robust results among different scales of sample size. We divided all editing sites into seven major modules based on the type of cells they exist (Figure 2(e)). The results showed that many editing sites were cell-specific, such as the performance of editing events on *Nudt6, Chd5, Wdfy1, LOC101056073, Foxn3, Cnot10* and *Zfp534* genes (Figure 2(e)). It is indicated that many RNA editing events occur at specific stages of spermatogenesis and are dynamic, which is consistent with the dynamic expression of genes during spermatogenesis.

Four most common types of nucleotide substitutions

Unlike humans, there are four dominant types of REs in mouse, including adenosine-to-inosine (A-to-I), guanosine-to-adenosine (G-to-A), cytidine-to-

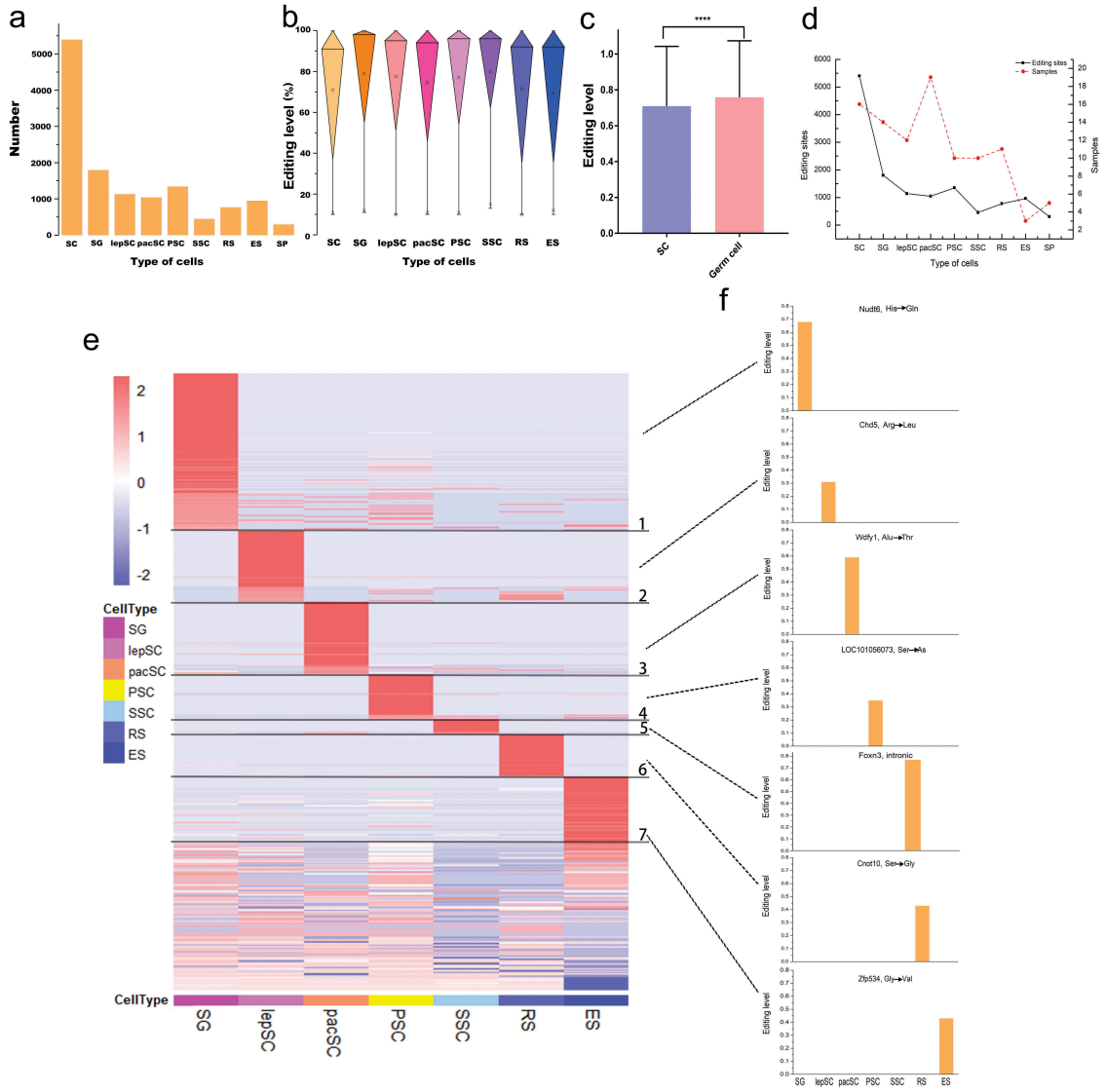

Figure 2. Editing sites identified in different types of cells during spermatogenesis. (a) The number of editing sites for specific types of cells. (b) Distribution of RNA editing levels in different germ cells. (c) Comparison of RNA editing levels between sertoli cells and male germ cells. Statistical significance was determined using *t*-test, ****$p < 0.0001$ (d) The relationship between sample size and REs number. (e) Modular RNA editing sites based on cell type. (f) Editing levels of seven genes that only undergo RNA editing in specific germ cells.

uridine (C-to-U) and uridine-to-cytidine (U-to-C) (Figure 3(a,b); Supplemental Fig S2) [23,24]. However, comparing the editing levels of these four major RNA editing events, we found that in addition to SSCs, the editing levels of A-to-I and T-to-C were higher than the other two in the other germ cells (Figure 3(c,d)). A-to-I substitutions are significantly less abundant in mouse than in human, mainly owing to the absence of *Alu* repeats in its genomes [14,23]. Two prerequisites for A-to-I RNA editing are needed, the existence of double-stranded RNAs (dsRNAs) and Adenosine Deaminase Acting on RNA (ADAR) family of enzymes. A-to-I editing is catalyzed by members of the ADAR family of enzymes which play a role in dsRNAs [25] and *Alu* sequence contains long dsRNA stem loops [26]. *Alu* sequence is the most common SINEs in humans, which is a kind of primate-specific repetitive sequences, accounts for greater than 10% of the human genome.

There is no *Alu* repeats in mouse genome, but four distinct SINEs are active in mouse, including B1, B2, ID and B4. These four families in mouse occupy a smaller portion of the genome than *Alu* in human (7.6% and 10.7%, respectively). We counted the editing sites in the repetitive regions of these eight types of cells and found that REs in

Table 1. The details of the editing sites identified in specific cells.

	SC		SG		lepSC		pacSC		PSC		SSC		RS		ES		SP	
	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes	Editing sites	#genes
All regions	5402	2008	1803	897	1140	621	1047	650	1349	700	455	364	775	521	961	478	304	190
Exonic	796	316	391	169	197	104	265	152	289	130	53	42	117	79	341	168	20	17
Synonymous	346	172	177	86	82	43	129	80	132	60	19	18	32	22	155	89	6	6
Nonsynonymous	272	152	148	83	71	52	82	64	92	59	10	9	48	36	119	75	13	13
Stopgain	11	10	4	4	2	2	3	3	8	8	0	0	3	3	3	3	0	0
Stoploss	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1

repetitive regions accounted for a lower proportion than in non-repetitive regions. What's more, less editing events occur in repetitive regions in male germ cells than in sertoli cells (Table 2).

Editing sites on different chromosomes

We analyzed the distribution of the editing sites on all the 21 chromosomes in different types of cells (Figure 4(a)). Interesting, we found that editing sites showed high density on chromosome 17, the number of editing sites on per unit length (one MB) is significantly higher than other chromosomes (Figure 4(b)), and the same chromosomal preference can also be seen in the validation data (Supplemental Figure S4). Functional enrichment analysis reveals that edited genes on chromosome 17 are almost immune-related (Figure 4(c)). As well known, take the blood testis barrier as an example, immunity play specific key roles in reproduction. The seminiferous tubules are divided into two parts by the blood testis barrier: basal and apical compartments. Meiosis, spermiogenesis and spermiation take place in the apical compartment; whereas, SG division and differentiate to pacSC occur in the basal compartment [27]. The blood-testis barrier protects spermatocytes, sperm cells, sperm against immune system attacks and pathogenic microbes, and thus acts as an immune barrier to provide a fairly stable microenvironment [28,29].

Compared with other chromosomes, there were a small number of editing sites on the Y chromosome and they only distributed at both ends of the Y chromosome. The validation samples also showed the same results. (Figure 4(a); Supplemental Fig S3; Figure 4(b)). Among Y chromosome genes, we found that non-synonymous editing occurs in gene *Rbmy* and *Erdr1*, and*Rbmy* has been found to be specifically expressed in adult testis (Figure 4(d)). It's reported that male mice deficient in *Rbmy* are sterile and knockout mutants of *Rbmy* are associated with azoospermia or severe oligospermia [30–32].

Half of the edited genes contain at least two editing sites

In total, REs were detected at 2012 protein coding genes in these germ cells and 2008 in sertoli cells. In male germ cells, 1042 of 2012 (69.68%) genes harbor

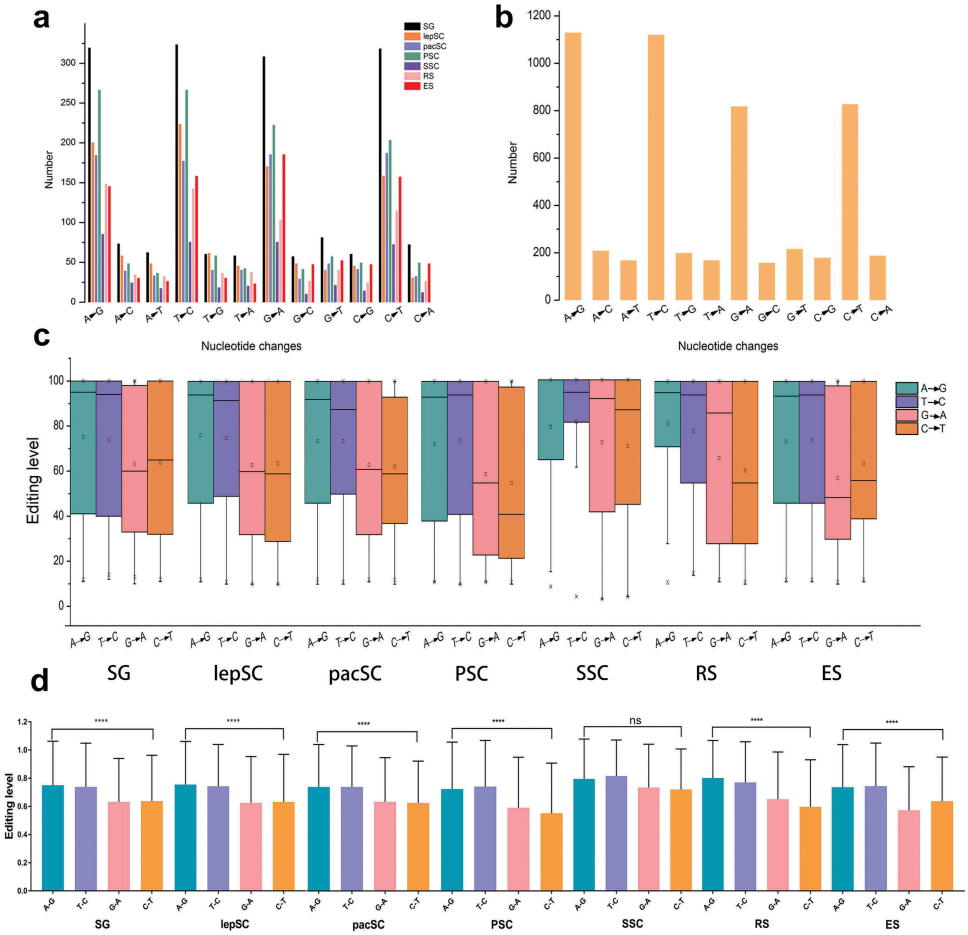

Figure 3. The number of 12 kinds of nucleotide changes. (a) Nucleotide changes in germ cells. (b) Nucleotide changes in sertoli cells. (c) Distribution of editing levels of four main types RNA editing in different germ cells. (d) Comparison of editing levels of these four main RNA editing. Statistical significance was determined using ANOVA, ****p < 0.0001, ns: non-significant.

Table 2. Editing sites in genomic repetitive regions.

	SC	SG	lepSC	pacSC	PSC	SSC	RS	ES
All sites	5402	1803	1140	1047	1349	455	775	961
Sites at repeat regions	1122	196	179	139	107	49	72	67
Percent	20.7%	10.87%	15.7%	13.28%	7.93%	10.77%	9.29%	6.97%

two or more editing sites (Figure 5(a,d)), while 78 genes showed more than 20 edited sites (Figure 5 (d)). However, in sertoli cells, 998 of 2008 (49.70%) genes harbor at least two editing sites (Figure 5(b)). Further analysis indicates that there are more than once non-synonymous editing events at most edited genes in male germ cells (Figure 5(c)). There are 19 genes that contain more than 10 non-synonymous editing sites. (Figure 5(e)).Among them, *Gm5458* and *Gm21119* show most abundant editing events, with 77 and 70 editing sites, respectively. Interesting, we found these two genes don't express in other tissues, but only show high expression level in adult mice testes (Figure 5(f,g)), which provides strong hints that they may play key roles during spermatogenesis.

Half of the editing events in CDSs lead to amino acid changes

In humans, REs primarily occur in introns and UTRs [33–35]. When analyzed the distribution of REs in different cells during spermatogenesis, we found that the editing events mainly occurred in introns, CDSs and intergenic in mouse testicular cells (Figure 6(a)).

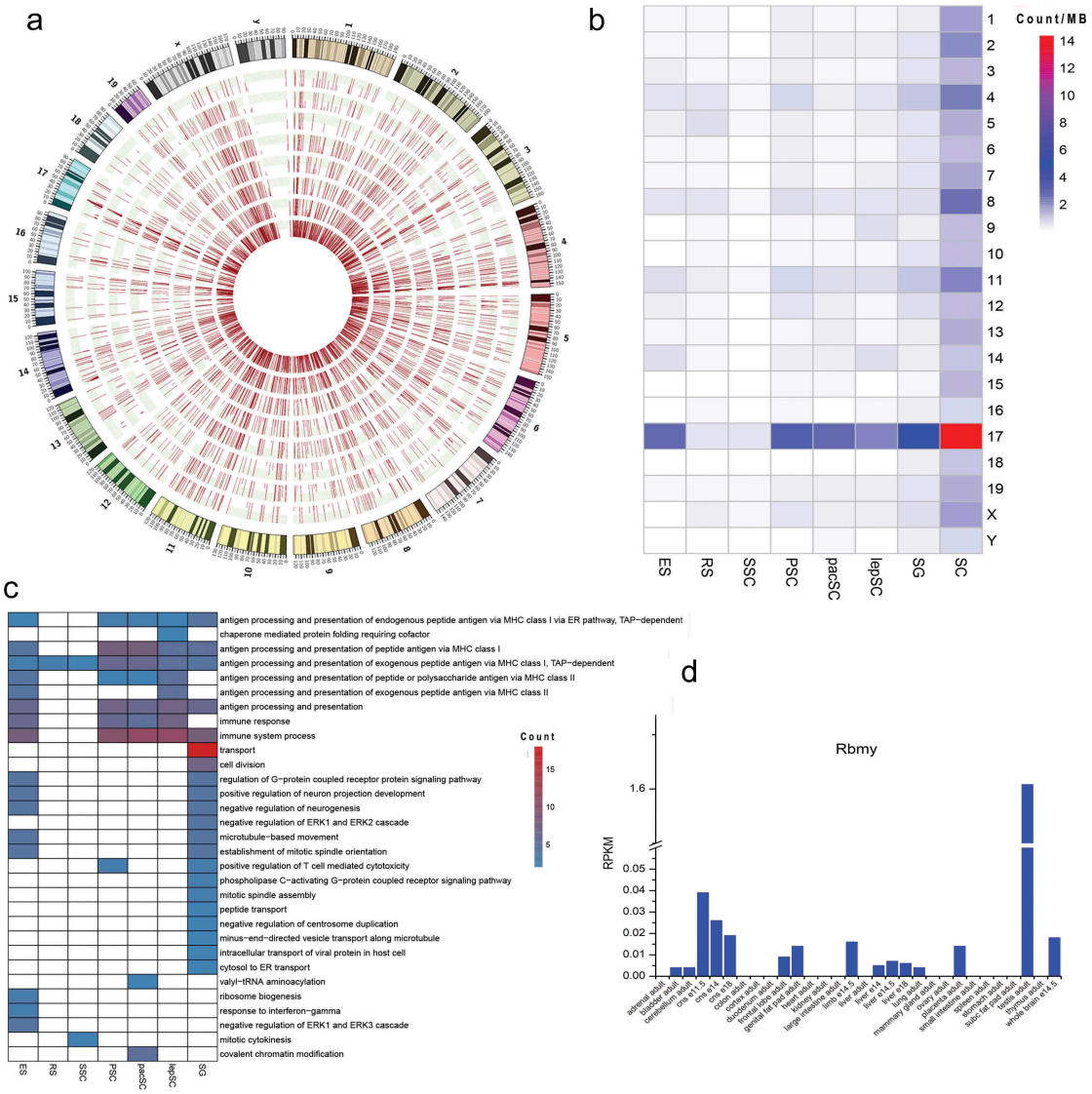

Figure 4. Editing sites on chromosome 17. (a) Distribution of editing sites on each chromosome in all types of cells and RNA editing levels in red bars. Cells are shown in concentric circles and ordered as follow from the inside: SC, SG, lepSC, pacSC, PSC, SSC, RS and ES. (b) The number of editing sites on the unit chromosome length of 21 chromosomes in each type of cell. (c) Function of edited genes on chromosome 17. (d) The expression of *Rbmy* in different tissues.

Further analyzed the editing sites in CDSs, to our surprise, about half of the editing events in CDSs resulted in amino acid changes (Figure 6(b)). On average, editing events occur in the CDSs accounted for 14%~39%. But the situation on chromosome 17 is different, on which the density of the editing sites is the highest, corresponding to 43%~62%in any germ cells, while just 33% in sertoli cells (Figure 6(a,c)). Unlike sertoli cells, in male germ cells, editing events occur more frequently in CDSs than in other region-sof chromosome 17 (Figure 6(c)). Chromosome 17 may be an active chromosome that performs some unclear important functions in spermatogenesis.

Function of RNA editing on spermatogenesis

We got different REs between male germ cells and sertoli cells (Figure 7(a)) and analyzed their related gene functions. Functional enrichment analysis results showed that many edited genes are associated with spermatogenesis, spermatid development, mitosis, cell differentiation and immunity (Figure 7(b)). Meanwhile, many genes have undergone REs at multiple stages of spermatogenesis, so we tried to link the relationships of these frequently edited genes with their function in spermatogenesis. Finally, many significant associations of gene-function were detected among all the seven developmental stages,

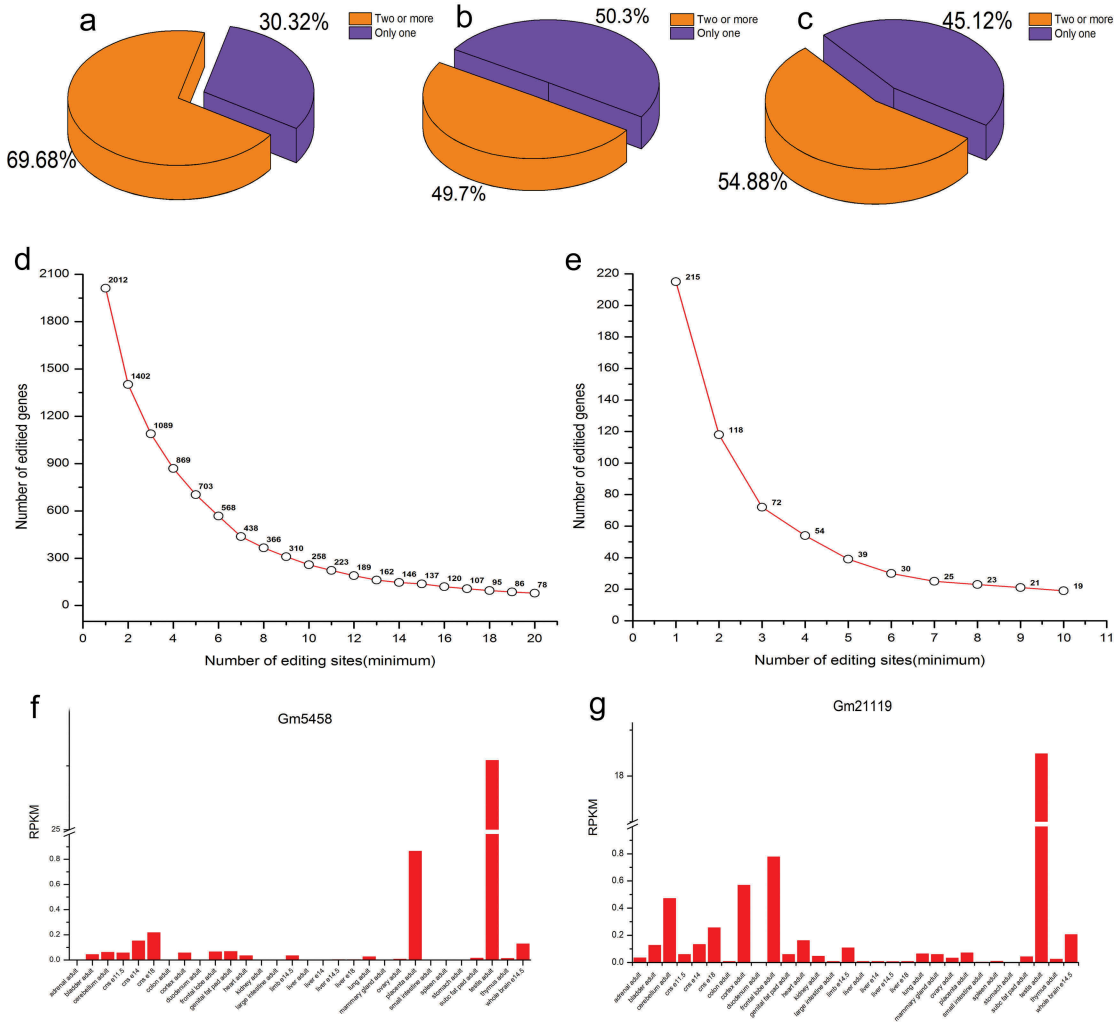

Figure 5. Genes with different numbers of editing events. (a) Proportion of genes with one editing event and genes with two or more editing events in germ cells. (b) Proportion of genes with one editing event and genes with two or more editing events in sertoli cells. (c) Proportion of genes with one nonsynonymous editing event and genes with two or more nonsynonymous editing events in germ cells. (d) The number of genes with different numbers of editing events in germ cells. (e) The number of genes with different numbers of nonsynonymous editing events in germ cells. (f) The expression of *Gm5458*across different tissues. (g) The expression of *Gm21119*across different tissues.

and we showed some typical gene-function pairs here (Figure 7(c)). Both *Ddx3x* and *Hjurp* are found to be related to chromosome segregation. *Ddx3x* has undergone REs in the seven stages of spermatogenesis, whereas *Hjurp* has been edited in four stages. In addition, *Adad1* and *Rnf17* are involved in spermatogenesis and spermatid development, while *Cxcr4* and *Cxcl12* play roles in germ cell migration. These four genes are also edited in different stages of spermatogenesis. In addition, we calculated differentially expressed genes (DEGs) between germ cells and sertoli cells, and found that all DEGs that are non-synonymously edited in germ cells are up-regulated compared to sertoli cells, and some of them are testis-specific genes or play a role in spermatogenesis, such as *Adam1b, Ddias, Speer4f1* and *Ssxb2* [36,37] (Supplemental Table S2).

RNA editing affects the translation process of genes into proteins by changing codons, and non-synonymous editing affects genes to perform their functions normally. Studies have shown that RNA editing is associated with a variety of diseases, and recent studies have shown that RNA editing may be associated with cancer, aging, cardiovascular, neurological and autoimmune diseases [38,39]. It is possible that REs on the above reproductive-related genes may alter the amino acids sequence by changing the codons of the genes, and non-synonymous REs

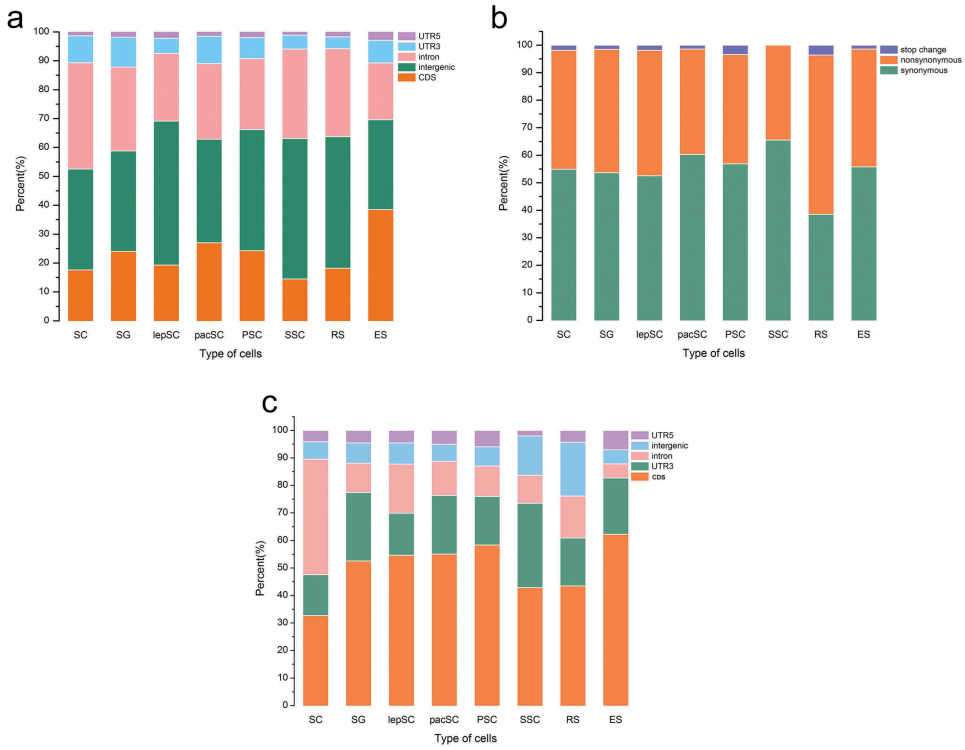

Figure 6. About half of the editing events in CDSs result in amino acid changes. (a) Distribution of the editing sites. (b) Functional consequences of the editing sites. (c) Distribution of the editing sites on chromosome 17.

occurring on the stop codon may even prolong or prematurely terminate the protein translation, thereby altering even losing these protein's original function. As a result, the spermatogenesis process is likely to be disordered. In short, these edited genes show close relationships with spermatogenesis through REs.

Validation of specific editing site and conserved RNA editing site between species and tissues

To demonstrate the reliability and effectiveness of the integration strategy, we sequenced the whole genome and transcriptome of germ cells mix of the same male adult mouse, and then detected RNA editing sites by comparing DNA-seq and RNA-seq. Previously, through the integration method, the sertoli cells were used as control, we detected editing sites specific to germ cells. These sites were compared with the results detected by DNA-seq and RNA-seq, and one-third of the sites were verified by sequencing (Supplemental Figure S5). This is mainly due to the fact that most of the germ cells in male adult mouse testis are after meiosis, so this ratio is also in line with our expectation.

Interesting, we found that there is a non-synonymous A-to-I RNA editing site in *Cog3* during mouse spermatogenesis, which have been reported previously in brains of human, rhesus and mouse [17,40]. It has been confirmed that this editing site is not only conserved between species, but also conserved across tissues [41]. Except for this conservative editing site, we also found a stoploss editing in the transcript of the testis-specific gene *Tssk6*, whose translation product is a serine/threonine kinase, it's required for chromatin remodeling after division. Studies have shown that *Tssk6* silenced mouse is infertile, and the number of sperm is reduced, sperm motility is decreased, the percentage of abnormally shaped sperm is significantly increased [42]. We find that this stoploss editing will increase the protein from the original 273 amino acids to 369 amino acids, and the protein molecular weight increased by 11kd (Figure 8(a)).

To validate the authenticity of our results, we sequenced the genome and cDNA of *Cog3* and *Tssk6* in male germ cells. By comparing the

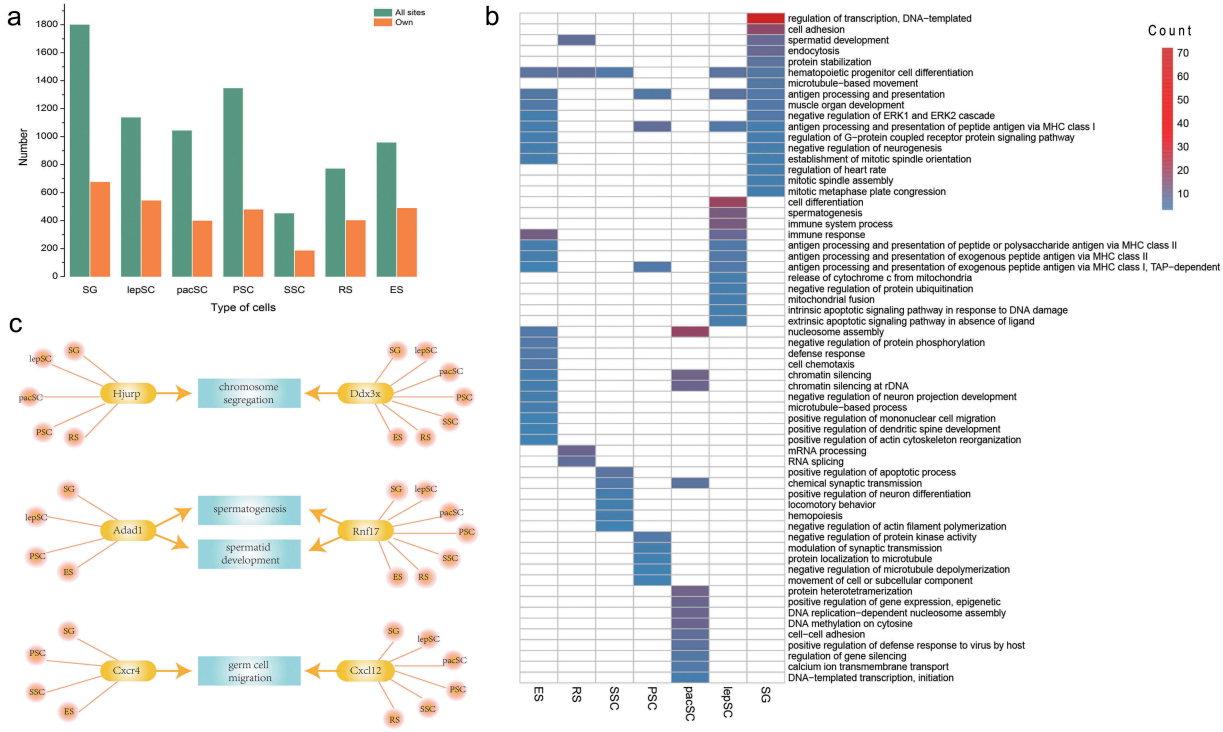

Figure 7. Significant editing events for spermatogenesis. (a) All sites in male germ cells and editing sites only in germ cells. (b) Function of genes that have editing sites only in germ cells. (c) The function of genes which are edited in multiple stages.

nucleotides of the genome and cDNA, we found that, as we detected earlier based on integrated RNA-Seq, the A-to-I RNA editing of *Cog3*, still occurs during mouse spermatogenesis (Figure 8(b)). The stop codon of the gene *Tssk6* is unable to terminate translation properly due to A-to-I editing (Figure 8(c)), and the amino acid chain grows, which may affect its protein function. Through the verification of these two sites, we confirmed the existence of RNA editing events during spermatogenesis, and REs occurred on some key reproductive genes. Once the editing levels of these genes are abnormal, the spermatogenesis process may be affected.

Discussion

In this study, we integrated all the published RNA-Seq data of all kinds of male germline cells with sertoli cells as control during mouse spermatogenesis. Based on the power of integrated data, we established the robust landscape of REs during the development of spermatogenesis. Finally, 7530 editing sites were identified in total, 2012 genes in male germ cells have presented editing events, and more than half of them harbor at least two editing sites. The two most abundant editing genes are *Gm5458* and *Gm21119*, with 77 and 70 editing sites, respectively. Both of them were found specifically expressed in testis.

Unlike human, A-to-I substitutions does not show an absolute predominant in mouse male germ cells [43], and there are four most common types of nucleotide substitutions, they are A-to-I, G-to-A, C-to-T and T-to-C. Editing sites are enriched on some regions of chromosome, with chromosome 17 as the highest density chromosome. On the Y chromosome, editing events occur specifically at both ends of the chromosome, and non-synonymous editing events only occur on the genes *Rbmy* and *Erdr1*, where *Rbmy* plays an important role in sperm formation. We hypothesize that *Rbmy* plays a role in spermatogenesis through REs as a post-transcriptional regulator. On the whole, unlike human, REs mainly occurs in introns, intergenic and CDSs in mouse, and about half of the editing events in CDSs can lead to amino acid changes. As to chromosome 17, enrichment of editing events in CDSs were most obvious. Interesting, the edited genes on chromosome 17 are almost immune-related, while as well

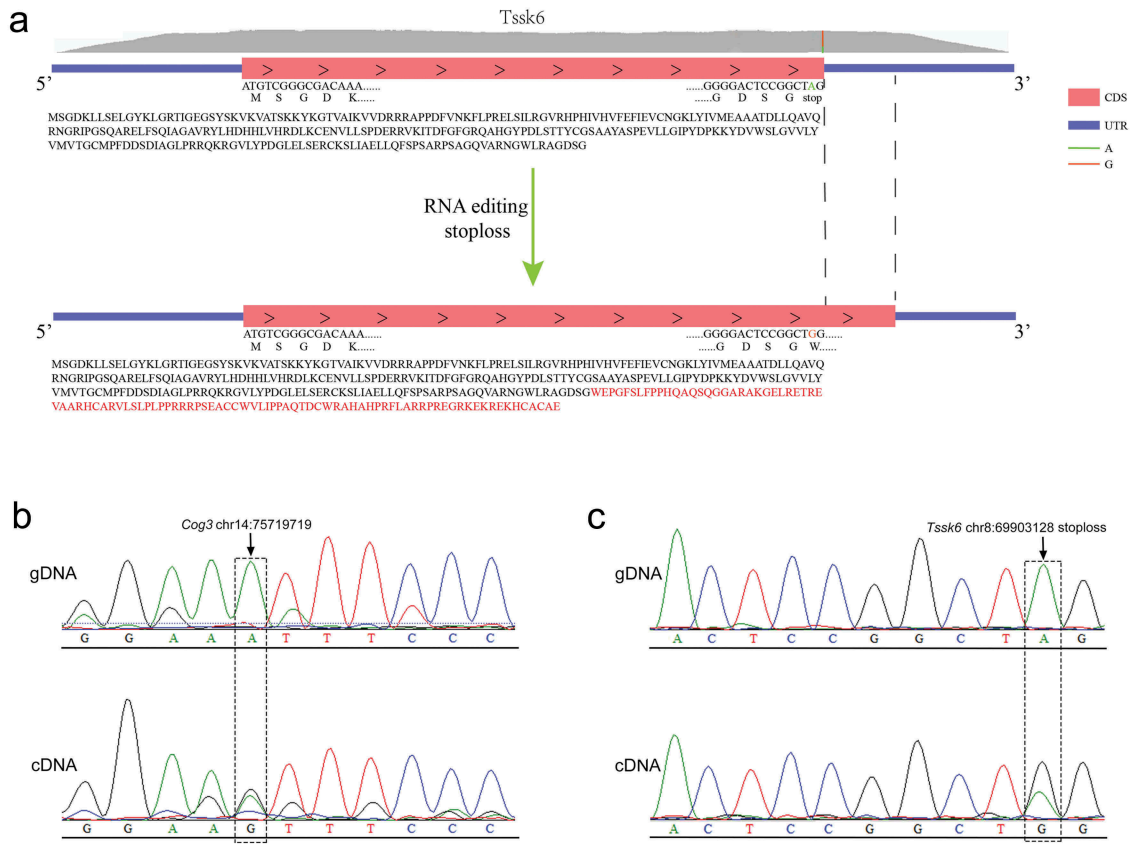

Figure 8. Validation of RNA editing site with Sanger sequencing. (a)The stoploss editing in *Tssk6* will change the original stop codon to a codon encoding an amino acid and increase the protein from the original 273 amino acids to 369 amino acids. (b) Sanger sequencing of cDNA and gDNA, at the site chr14:75,719,719 of gene *Cog3* the genomic A is highly edited to I. (c) An A-to-I editing occurred at the stop codon of the gene *Tssk6*.

known, immunity and reproductive are very closely linked. Therefore, we have reason to speculate that chromosome 17 may be a key chromosome during spermatogenesis through immune systems. What's more, edited genes are found to be related to spermatogenesis, spermatid development, mitosis, cell differentiation and immunity. Among them, some are frequently edited at multiple stages, such as *Ddx3x, Hjurp, Adad1, Rnf17, Cxcr4* and *Cxcl12*, and some are up-regulated DEGs compared to sertoli cells, such as *Adam1b, Ddias, Speer4f1*, and *Ssxb2*. Our results also add that the previously identified conserved non-synonymous editing site in *Cog3* is also conserved across tissues. A new post-transcriptional regulation of the gene *Tssk6* was discovered and confirmed by us, the stoploss editing in its transcript will translate to new protein products. In order to increase the difficulty of verification, we chose Kunming mice, which belong to different strains from the data source mice we used. Observing whether our results are applicable to mice of different strains. As we expected, our predicted RNA editing sites were validated in Kunming mice, suggesting that these editing sites may be conserved among mice of different strains. Summary, we integrated all published RNA-Seq datasets, confirmed the existence of RNA editing events and established the landscape of REs during spermatogenesis. Based on the landscape, we deconstructed molecular mechanisms for the development of male germ cells through REs from the perspective of post-transcriptional regulation.

In order to take advantages of the power of merging different studies as one union, we integrated all the published RNA-Seq datasets at the same conditions. As a typical dynamic complex process, spermatogenesis was chose to apply our strategies. Our results showed that there are much novel insights can be detected by integrating different datasets. Our strategies should be general

for other researches to make use of the huge amount of published biological datasets. Beyond RNA-Seq data, our strategies can be also used for other types of sequencing datasets.

In this study, we focused on RNA editing events at different stages of mouse spermatogenesis, but whether there are quantitative differences in RNA editing sites between different germ cells, whether these different editing sites have corresponding functional roles, and what cause these differences? These issues require further in-depth research. We analyzed the RNA editing of normal spermatogenesis. Compared with the normal spermatogenesis process, whether there is a huge difference in the number of RNA editing sites, the level of RNA editing, and the distribution of RNA editing types in the male sterile population, which requires us to analyze the data of infertile individuals in the next step. This work focused on the REs regulation for mRNA of coding proteins. But as well known, the majority of RNAs generated in the cell are non-coding RNAs (ncRNAs). In human, it is estimated that about two-third of genome is transcribed and only 2% of these transcribed RNAs are translated into proteins. The functions of ncRNAs vary from regulating transcription, translation, DNA repair and cell fate decision [44]. It's recently reported that non-coding RNA may be associated with cytoplasmic male sterility in plant [45]. What roles do non-coding RNAs play in spermatogenesis and whether they affect the formation of spermatozoa by REs are still unclear. These problems should be considered in the future researches and it will be promising for the molecular mechanism exploring of reproduction.

Materials and methods

Data collection and classification integration

RNA-seq reads of mouse spermatogenesis used to identify editing sites were downloaded from GEO and SRA, and then categorized them into nine groups according to their development stages, including spermatogonia, leptotene spermatocytes, pachytene spermatocytes, primary spermatocytes before meiosis I, secondary spermatocytes, round spermatids, elongative spermatids, sperm and sertoli cells as control. Validation datasets of spermatogonia and round spermatids were under accession number GSE75826.

Read mapping and identification of RNA editing sites

RNA-seq reads were aligned to the complete genome of mouse available in Ensembl with program HISAT2 [46]. RNA editing sites were identified by REDItools [47] and a series of stringent filter were implemented to eliminate false positive, as was described in Supplemental Methods. Known variations were downloaded from UCSC and were moved by bedtools, with details in Supplemental Methods.

Annotation of RNA editing

Annovar [48] was used to annotate RNA editing based on their related gene, while the related annotation files refGene.txt and chromFa.tar.gz were downloaded from UCSC with version mm10. Based on these data, gene-based transcriptional expression and the distribution of editing sites on chromosomes, genes and genomic elements were caculated.

Calculation of gene expression and degs

After mapping with HISAT2, the alignments files were passed to StringTie for transcript assembly and calculating the expression level of transcript. Finally, Ballgown took all the transcripts and abundances from StringTie and counted expression based on genes and looked for DEGs [49], some details in Supplemental Methods.

Mice

Five adult male Kunming mice were used in our experiments. All mice were housed in a barrier facility under natural light at 20-25℃, free access to food and water. All experimental procedures involving animals were approved by the Institutional Animal Care and Use Committee of Northwest A&F University.

Isolation of germ cells

The convoluted tubule was isolated from mouse testicles in a sterile environment and digested with trypsin to obtain single cell viability. Cells were cultured

using DMEM/high glucose (Gibco, Grand Island, NY, USA) and supplemented with 10% FBS (Gibco),100 U/ml penicillin, 100 mg/ml streptomycin, one MEM non-essential amino acids. After 4 h, when sertoli cells adhered to the bottom, germ cells were collected by centrifugation (500 g/5 min).

PCR and sanger sequencing analysis

The genome was obtained through TIANamp Genomic DNA Kit and the total RNA was extracted by RNAiso Plus using a one-step method of isothiocyanate guanium-phenol-chloroform. PCR amplification was performed using the cDNA and genome. The primer sequences were as follows: DNA primer of *Cog3* F: ACTG CAGAGAACTATCACTTTGAGG, R: AAGATG ATGTCACTCTGTTCCCTTT, DNA primer of *Tssk6* F: CCTACGACCCTAAGAAATACGACG, R: GAGGAATCAGAACCCAACAGCA. cDNA primer of *Cog3* F: TAGATGCATAGATAGGGC GGTGT, R: TGCTGTGAGAGGGTGTACTTGG. cDNA primer of *Tssk6* F: CCTACGACCCTAAGA AATACGACG, R: GAGGAATCAGAACCCAA CAGCA. The PCR products were sequenced by Sanger sequencing analysis in Sangon biotech.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was financially supported by the National Natural Science Foundation of China [grant numbers 61772431 and 31572399]; Natural Science Fundamental Research Plan of Shaanxi Province [grant number 2018JM6039]; Program of Shaanxi Province Science and Technology Innovation Team (2019TD-036).

ORCID

Qun Li http://orcid.org/0000-0002-7194-7504

References

[1] Sharlip ID, Jarow JP, Belker AM, et al. Best practice policies for male infertility. Fertil Sterility. 2002;77 (5):873–882.

[2] Pizzol D, Bertoldo A, Foresta C. Male infertility: biomolecular aspects. Biomol Concepts. 2014;5(6):449–456.

[3] Agarwal A, Mulgund A, Hamada A, et al. A unique view on male infertility around the globe. ReprodBiol Endocrinol. 2015 Apr 26;13:37. PubMed PMID: 25928197; PubMed Central PMCID: PMC4424520.

[4] O'Brien KLOF, Varghese AC, Agarwal A. The genetic causes of male factor infertility: A review. Fertil Sterility. 2010;93(1):1–12.

[5] Ferlin A. New genetic markers for male fertility. Asian J Androl. 2012 Nov;14(6):807–808. PubMed PMID: 22902905; PubMed Central PMCID: PMC3720108.

[6] Che D, Wang Y, Bai W, et al. Dynamic and modular gene regulatory networks drive the development of gametogenesis. Brief Bioinform. 2017;18(4):712–721.

[7] Bai W, Yang W, Wang W, et al. GED: a manually curated comprehensive resource for epigenetic modification of gametogenesis. Brief Bioinform. 2017 Jan;18 (1):98–104. PubMed PMID: 26851225.

[8] Paz N, Levanon EY, Amariglio N, et al. Altered adenosine-to-inosine RNA editing in human cancer. Genome Res. 2007;17(11):1586–1595.

[9] Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. Genome Med. 2013;5(11):1–13.

[10] Wang Y, Xu X, Yu S, et al. Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. Genome Res. 2017;27(7):1112.

[11] Sansam CL, Wells KS, Emeson RB. Modulation of RNA editing by functional nucleolar sequestration of ADAR2. Proc Natl Acad Sci U S A. 2003;100 (24):14018–14023.

[12] Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2004 Dec;2(12):e391. PubMed PMID: 15534692; PubMed Central PMCID: PMC526178.

[13] Blow M, Futreal PA, Wooster R, et al. A survey of RNA editing in human brain. Genome Res. 2004;14 (12):2379.

[14] Kim DD, Kim TT, Walsh T, et al. Widespread RNA Editing of Embedded Alu Elements in the Human Transcriptome. Genome Res. 2004;14(9):1719–1725.

[15] Buchumenski I, Bartok O, Ashwal-Fluss R, et al. Dynamic hyper-editing underlies temperature adaptation in Drosophila. PLoS Genet. 2017;13(7):e1006931.

[16] Yablonovitch AL, Deng P, Jacobson D, et al. The evolution and adaptation of A-to-I RNA editing. PLoS Genet. 2017;13(11):e1007064.

[17] Li JB, Levanon EY, Yoon J-K, et al. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science. 2009;324 (5931):1210–1213.

[18] Danecek P. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. Genome Biol. 2012;13(4):r26.

[19] Lin X, Han M, Cheng L, et al. Expression dynamics, relationships, and transcriptional regulations of diverse

transcripts in mouse spermatogenic cells. RNA Biol. 2016;13(10):1011–1024.
[20] Monesi V. Synthetic activities during spermatogenesis in the mouse: RNA and protein ☆. Exp Cell Res. 1965;39(1):197–224.
[21] Moore GP. DNA-dependent RNA synthesis in fixed cells during spermatogenesis in mouse. Exp Cell Res. 1971;68(2):462–465.
[22] Geremia R, Boitani C, Conti M, et al. RNA synthesis in spermatocytes and spermatids and preservation of meiotic RNA during spermiogenesis in the mouse. Cell Differ. 1977;5(5):343–355.
[23] Eisenberg E, Nemzer S, Kinar Y, et al. Is abundant A-to-I RNA editing primate-specific? Trends Genet. 2005;21(2):77–81.
[24] Neeman Y, Levanon EY, Jantsch MF, et al. RNA editing level in the mouse is determined by the genomic repeat repertoire. Rna. 2006 Oct;12(10):1802–1809. PubMed PMID: 16940548; PubMed Central PMCID: PMC1581974.
[25] Bass BL. RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem. 2002;71(1):817–846.
[26] Sinnett D, Richer C, Deragon JM, et al. Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. J Biol Chem. 1991;266(14):8675.
[27] Russell L. Movement of spermatocytes from the basal to the adluminal compartment of the rat testis. Dev Dyn. 1977;148(3):313–328.
[28] Cheng CY, Mruk DD. The blood-testis barrier and its implications for male contraception. Pharmacol Rev. 2012;64(1):16–64.
[29] Jiang XH, Bukhari I, Zheng W, et al. Blood-testis barrier and spermatogenesis: lessons from genetically-modified mice. Asian J Androl. 2014 Jul-Aug;16(4):572–580. PubMed PMID: 24713828; PubMed Central PMCID: PMC4104086.
[30] Mahadevaiah SK, Odorisio T, Elliott DJ, et al. Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. Hum Mol Genet. 1998;7(4):715–727.
[31] Tyler-Smith C. An evolutionary perspective on Y-chromosomal variation and male infertility. Int JAndrology. 2008;31(4):376–382.
[32] Yan Y, Yang X, Liu Y, et al. Copy number variation of functional RBMY1 is associated with sperm motility: an azoospermia factor-linked candidate for asthenozoospermia. Hum Reprod. 2017;32(7):1.
[33] Bahn JH, Lee JH, Li G, et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. 2012 Jan;22(1):142–150. PubMed PMID: 21960545; PubMed Central PMCID: PMC3246201.
[34] Park E, Williams B, Wold BJ, et al. RNA editing in the human ENCODE RNA-seq data. Genome Res. 2012;22(9):1626–1633.
[35] Sakurai M, Ueda H, Yano T, et al. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. Genome Res. 2014 Mar;24(3):522–534. PubMed PMID: 24407955; PubMed Central PMCID: PMC3941116.
[36] Chen YT, Alpen B, Ono T, et al. Identification and characterization of mouse genes: a multigene family on the X chromosome with restricted cancer/testis expression☆☆. Genomics. 2003;82(6):628–636.
[37] Spiess AN, Walther N, Ivell R, et al. SPEER–a new family of testis. Biol Reprod. 2003;68(6):2044–2054.
[38] Gatsiou A, Vlachogiannis N, Lunella FF, et al. Adenosine-to-Inosine RNA Editing in Health and Disease. Antioxid Redox Signal. 2018;29(9):846–863. PubMed PMID: 28762759.
[39] Gallo A, Vukic D, Michalík D, et al. ADAR RNA editing in human disease; more to it than meets the I [journal article]. Human Genet. 2017 Sept 01;136(9):1265–1278.
[40] Ramaswami G, Jin BL. RADAR: a rigorously annotated database of A-to-I RNA editing. Nucleic Acids Res. 2014;42(Databaseissue):D109.
[41] Snyder EM, Licht K, Braun RE. Testicular adenosine to inosine RNA editing in the mouse is mediated by ADARB1. Biol Reprod. 2017;96(1):244.
[42] Spiridonov NA, Lily W, Zerfas PM, et al. Identification and characterization of SSTK, a serine/threonine protein kinase essential for male fertility. Mol Cell Biol. 2005;25(10):4250–4261.
[43] Levanon K, Eisenberg E, Rechavi G, et al. Letter from the editor: adenosine-to-inosine RNA editing in Alu repeats in the human genome. EMBO Rep. 2005;6(9):831–835.
[44] Bunch H. Gene regulation of mammalian long non-coding RNA. Mol Genet Genomics. 2017;293(1):1–15.
[45] Stone JD, Kolouškova P, Sloan DB, et al. Non-coding RNA may be associated with cytoplasmic male sterility inSilene vulgaris. J Exp Bot. 2017;68(7):1599–1612.
[46] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357–360. PubMed PMID: 25751142; PubMed Central PMCID: PMC4655817.
[47] Picardi E, Pesole G. REDItools: high-throughput RNA editing detection made easy. Bioinformatics. 2013;29(14):1813–1814.
[48] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
[49] Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016 Sep;11(9):1650–1667. PubMed PMID: 27560171; PubMed Central PMCID: PMC5032908.