

# **Landscape of RNA editing reveals new insights into the dynamic gene regulation of spermatogenesis based on integrated RNA-Seq**

Xiaodan Wang<sup>1,\*</sup>, Zhenshuo Zhu<sup>2,\*</sup>, Xiaolong Wu<sup>2</sup>, Hao Li<sup>1</sup>, Tongtong Li<sup>1</sup>, Qun Li<sup>1</sup>, Peng Zhang<sup>1</sup>, Leijie Li<sup>1</sup>, Dongxue Che<sup>1</sup>, Xia Xiao<sup>1</sup>, Jinlian Hua<sup>2,#</sup>, Mingzhi Liao<sup>1,#</sup>

<sup>1</sup>College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, China

<sup>2</sup>College of Veterinary Medicine, Shaanxi Centre of Stem Cells Engineering & Technology, Northwest A&F University, Yangling, Shaanxi, China

\* The authors contribute equally to this work

<sup>#</sup>Correspondence author: Dr. Mingzhi Liao; E-mail: [liaomingzhi83@163.com](mailto:liaomingzhi83@163.com); Tel: +86 029 87092262.

<sup>#</sup>Correspondence author: Dr. Jinlian Hua; E-mail: [jinlianhua@nwsuaf.edu.cn](mailto:jinlianhua@nwsuaf.edu.cn); Tel: +86 029 87080068.

## ABSTRACT

Spermatogenesis is an important physiological process associated with male infertility. But whether there are RNA editings (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 REs during the development of mouse spermatogenesis. 7530 editing sites among all types of male germ cells were found, which enrich on some regions of chromosome, including chromosome 17 and both ends of chromosome Y. Totally, REs occur in 2012 genes during spermatogenesis, more than half of which harbor at two different sites of the same gene at least. We also found REs mainly occur in introns, coding regions (CDSs) and intergenic regions. Moreover, about half of the REs in CDSs can cause amino acids changes. Finally, based on our adult male Kunming mice, we verified that there is a non-synonymous A-to-I RNA editing site in *Cog3* during spermatogenesis, which is conserved not only between species but also across tissues. In short, based on the power of integrating RNA-Seq datasets, we provided the landscape of REs and found their dynamic changes during mouse spermatogenesis. This research strategy is general for other types of sequencing datasets and biological problems.

**KEYWORDS:** Spermatogenesis, RNA editing, Integrative analysis, RNA-Seq

## 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 (Agarwal et al., 2015; JP et al., 2002; Pizzol et al., 2014). 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 (Ferlin, 2012; O'Brien et al., 2010). Genetics leads 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 (Che et al., 2017). 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 (Bai et al., 2017).

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 (Paz et al., 2007; Slotkin and Nishikura, 2013; Wang et al., 2017). Generally, REs occur more frequently in non-coding than in coding regions. In human, RNA editing mainly occurs in introns and 5'- or 3'-untranslated regions (UTRs) (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004). Considering the complex and dynamics of spermatogenesis and the flexible regulation manner of REs, we supposed there may be close relationships between them. Though some hints have showed REs play a certain roles during spermatogenesis, no evidence supports the existence of REs, let alone the global perspective of REs function in germ cells.

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 specific cells 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 verified non-synonymous REs in *Cog3*, which is conserved not only between species but also across tissues. In summary, based on our integrated analysis and landscape of REs, we found that REs play key roles during spermatogenesis.

## RESULTS

### Data preprocessing

The framework of our works was shown as Fig. 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 illumine machines platform. In addition to the previous 100 samples, we also collected the latest published datasets under accession number GSE75826 as validation data. 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(8 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 □ (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 (Fig.2A; Table 1). It may be due to the activation of RNA synthesis in sertoli cells. On the other hand, 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(Geremia et al., 1977; Monesi, 1965; Moore, 1971). Therefore, we did not analyze the RNA editing sites identified in mature spermatozoa in the subsequent analysis. Finally, 7530 editing sites among other 7 types of male germ cells were found (Fig.2A; 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(Fig.2B;Fig.2C). 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 (Fig. 2D), 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 (Fig. 2E). 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 (Fig. 2E).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-uridine (C-to-U) and uridine-to-cytidine (U-to-C)(Fig. 3A; Fig.3B; Supplemental Fig. S2). 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 (Fig. 3C;Fig. 3D).A-to-I substitutions are significantly less abundant in mouse than in humans, mainly owing to the low representation of *Alu* repeats in its genomes(Eisenberg et al., 2005; Kim et al., 2004). 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 (Bass, 2002)and *Alu* sequence contains long dsRNA stem loops(Sinnett et al., 1991). *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 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 (Fig.4A). Interesting, we found that editing sites showed high density on chromosome 17, the number of editing sites on per unit length (1MB) is significantly higher than other chromosomes(Fig.4B), and the same chromosomal preference can also be seen in the validation data (Supplemental Fig.S4). Functional enrichment analysis reveals that edited genes on chromosome 17 are almost immune-related (Fig.4C). 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(Russell, 1977). 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(Cheng and Mruk, 2012; Jiang et al., 2014).

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. (Fig. 4A; Supplemental Fig. S3; Fig. 4B). 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 (Fig. 4D). It's reported that male mice deficient in *Rbmy* are sterile and knockout mutants of *Rbmy* are associated with azoospermia or severe oligospermia(Mahadevaiah et al., 1998; Tyler-Smith, 2008; Yan et al., 2017).

#### **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 two or more editing

sites (Fig 5A; Fig 5D), while 78 genes showed more than 20 edited sites (Fig 5D). However, in sertoli cells, 998 of 2008 (49.70%) genes harbor at least two editing sites (Fig 5B). Further analysis indicates that there are more than once non-synonymous editing events at most edited genes in male germ cells (Fig 5C). There are 19 genes that contain more than 10 non-synonymous editing sites. (Fig 5E). Among them, *Gm5458* and *Gm21119* show most abundant editing events, with 77 and 70 editing sites, respectively. Interestingly, we found these two genes don't express in other tissues, but only show high expression level in adult mice testes (Fig 5F; Fig 5G), 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 occurs in introns and UTRs (Bahn et al., 2012; Park et al., 2012; Sakurai et al., 2014). 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 (Fig 6A). Further analyzed the editing sites in CDSs, to our surprise, about half of the editing events in CDSs resulted in amino acid changes (Fig 6B). 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 (Fig 6A; Fig 6C). Unlike sertoli cells, in male germ cells, editing events occur more frequently in CDSs than in other regions of chromosome 17 (Fig 6C). 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 (Fig 7A) 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 (Fig 7B). 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, and we showed some typical gene-function pairs here (Fig 7C). 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 short, these edited genes show close relationships with spermatogenesis through REs.

### **Validation of conserved RNA editing site between species and tissues**

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 (Li et al., 2009; Ramaswami and Jin, 2014). This result demonstrated that the editing site is not only conserved between species, but also conserved across tissues. To validate the authenticity of our results, we sequenced the genome and cDNA of *Cog3* in male germ cells. By comparing the nucleotides of the genome and cDNA at this site (Fig 8), 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 which indicates an important role for the development

of male germ cells.

## 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 predominance in mouse male germ cells (Levanon et al., 2005), 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. Interestingly, the edited genes on chromosome 17 are almost immune-related, while as well 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*. Summary, we integrated all published RNA-Seq datasets 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 chosen 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.

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 (Bunch, 2017). It's recently reported that non-coding RNA may be associated with cytoplasmic male sterility in plant (Stone et al., 2017). 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  $\square$ , secondary spermatocytes, round spermatids, elongative spermatids, sperm and sertoli cells as control. Validation datasets of spermatogonia and round spermatids was 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(Kim et al., 2015). RNA editing sites were identified by REDIttools(Picardi and Pesole, 2013)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(Wang et al., 2010)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 calculated.

### **Calculation of gene expression**

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(Pertea et al., 2016), 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°C, 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, 1 MEM non-essential amino acids. After 4 hours, when sertoli cells adhered to the bottom, germ cells were collected by centrifugation(500g/5min).

### **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 guaniam-phenol-chloroform. PCR amplification was performed using the cDNA and genome.



The primer sequences were as follows: DNA primer F: ACTGCAGAGAACTATCACTTTGAGG, R: AAGATGATGTCCTCTGTTCCCTTT, cDNA primer F: TAGATGCATAGATAGGGCGGTGT, R: TGCTGTGAGAGGGTGTACTTGG. The PCR products were sequenced by Sanger sequencing analysis in Sangon biotech.

## COMPETING INTERESTS

The authors declare no competing or financial interests.

## FUNDING

This work was financially supported by the National Natural Science Foundation of China (61772431 to M. L. and 31572399 to J. H.); Natural Science Fundamental Research Plan of Shaanxi Province (2018JM6039 to M. L.).

## REFERENCES

- Agarwal, A., Mulgund, A., Hamada, A. and Chyatte, M. R.** (2015). A unique view on male infertility around the globe. *Reprod Biol Endocrinol***13**, 37.
- Athanasiadis, A., Rich, A. and Maas, S.** (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol***2**, e391.
- Bahn, J. H., Lee, J. H., Li, G., Greer, C., Peng, G. and Xiao, X.** (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res***22**, 142-150.
- Bai, W., Yang, W., Wang, W., Wang, Y., Liu, C., Jiang, Q., Hua, J. and Liao, M.** (2017). GED: a manually curated comprehensive resource for epigenetic modification of gametogenesis. *Brief Bioinform***18**, 98-104.
- Bass, B. L.** (2002). RNA editing by adenosine deaminases that act on RNA. *Annual review of biochemistry***71**, 817-846.
- Blow, M., Futreal, P. A., Wooster, R. and Stratton, M. R.** (2004). A survey of RNA editing in human brain. *Genome Research***14**, 2379.
- Bunch, H.** (2017). Gene regulation of mammalian long non-coding RNA. *Molecular Genetics and Genomics*, 1-15.
- Che, D., Wang, Y., Bai, W., Li, L., Liu, G., Zhang, L., Zuo, Y., Tao, S., Hua, J. and Liao, M.** (2017). Dynamic and modular gene regulatory networks drive the development of gametogenesis. *Briefings in bioinformatics***18**, 712-721.
- Cheng, C. Y. and Mruk, D. D.** (2012). The blood-testis barrier and its implications for male contraception. *Pharmacological reviews***64**, 16-64.
- Eisenberg, E., Nemzer, S., Kinar, Y., Sorek, R., Rechavi, G. and Levanon, E. Y.** (2005). Is abundant A-to-I RNA editing primate-specific? *Trends in Genetics***21**, 77-81.
- Ferlin, A.** (2012). New genetic markers for male fertility. *Asian J Androl***14**, 807-808.
- Geremia, R., Boitani, C., Conti, M. and Monesi, V.** (1977). RNA synthesis in spermatocytes and spermatids and preservation of meiotic RNA during spermiogenesis in the mouse. *Cell Differentiation***5**, 343-355.
- Jiang, X. H., Bukhari, I., Zheng, W., Yin, S., Wang, Z., Cooke, H. J. and Shi, Q. H.** (2014). Blood-testis barrier and spermatogenesis: lessons from genetically-modified mice. *Asian J Androl***16**, 572-580.
- JP, J., ID, S., AM, B., LI, L., M, S., AJ, T., PN, S., SS, H., A, N., MD, D., et al.** (2002). Best practice policies

for male infertility. *Fertility & Sterility***77**, 873-882.

**Kim, D., Langmead, B. and Salzberg, S. L.** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat Methods***12**, 357-360.

**Kim, D. D., Kim, T. T., Walsh, T., Kobayashi, Y., Matise, T. C., Buyske, S. and Gabriel, A.** (2004). Widespread RNA Editing of Embedded Alu Elements in the Human Transcriptome. *Genome Research***14**, 1719-1725.

**Levanon, K., Eisenberg, E., Rechavi, G. and Levanon, E. Y.** (2005). Letter from the editor: Adenosine-to-inosine RNA editing in Alu repeats in the human genome. *Embo Reports***6**, 831-835.

**Li, J. B., Levanon, E. Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., Zhang, K., Gao, Y. and Church, G. M.** (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science***324**, 1210-1213.

**Mahadevaiah, S. K., Odorisio, T., Elliott, D. J., Rattigan, A., Szot, M., Laval, S. H., Washburn, L. L., Mccarrey, J. R., Cattanch, B. M. and Lovellbadge, R.** (1998). 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. *Human Molecular Genetics***7**, 715-727.

**Monesi, V.** (1965). Synthetic activities during spermatogenesis in the mouse : RNA and protein ☆. *Experimental Cell Research***39**, 197-224.

**Moore, G. P.** (1971). DNA-dependent RNA synthesis in fixed cells during spermatogenesis in mouse. *Experimental Cell Research***68**, 462-465.

**O'Brien, K. L. O. F., Varghese, A. C. and Agarwal, A.** (2010). The genetic causes of male factor infertility: A review. *Fertility & Sterility***93**, 1-12.

**Park, E., Williams, B., Wold, B. J. and Mortazavi, A.** (2012). RNA editing in the human ENCODE RNA-seq data. *Genome Research***22**, 1626-1633.

**Paz, N., Levanon, E. Y., Amariglio, N., Heimberger, A. B., Ram, Z., Constantini, S., Barbash, Z. S., Adamsky, K., Safran, M. and Hirschberg, A.** (2007). Altered adenosine-to-inosine RNA editing in human cancer. *Genome Research***17**, 1586-1595.

**Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. and Salzberg, S. L.** (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc***11**, 1650-1667.

**Picardi, E. and Pesole, G.** (2013). RED|tools: high-throughput RNA editing detection made easy. *Bioinformatics***29**, 1813-1814.

**Pizzol, D., Bertoldo, A. and Foresta, C.** (2014). Male infertility: biomolecular aspects. *Biomolecular Concepts***5**, 449-456.

**Ramaswami, G. and Jin, B. L.** (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Research***42**, D109.

**Russell, L.** (1977). Movement of spermatocytes from the basal to the adluminal compartment of the rat testis. *Developmental Dynamics***148**, 313-328.

**Sakurai, M., Ueda, H., Yano, T., Okada, S., Terajima, H., Mitsuyama, T., Toyoda, A., Fujiyama, A., Kawabata, H. and Suzuki, T.** (2014). A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res***24**, 522-534.

**Sinnett, D., Richer, C., Deragon, J. M. and Labuda, D.** (1991). Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *Journal of Biological Chemistry***266**, 8675.

**Slotkin, W. and Nishikura, K.** (2013). Adenosine-to-inosine RNA editing and human disease. *Genome*

*Medicine*5, 1-13.

**Stone, J. D., Kolou?Kov, P., Sloan, D. B. and ?Torchov, H.** (2017). Non-coding RNA may be associated with cytoplasmic male sterility in *Silene vulgaris*. *Journal of Experimental Botany*68, 1599-1612.

**Tyler-Smith, C.** (2008). An evolutionary perspective on Y-chromosomal variation and male infertility. *International Journal of Andrology*31, 376-382.

**Wang, K., Li, M. and Hakonarson, H.** (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*38, e164.

**Wang, Y., Xu, X., Yu, S., Kang, K. J., Zhou, Z., Han, L., Tsang, Y. H., Li, J., Chen, H. and Mangala, L. S.** (2017). Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Research*27, 1112.

**Yan, Y., Yang, X., Liu, Y., Shen, Y., Tu, W., Dong, Q., Yang, D., Ma, Y. and Yang, Y.** (2017). Copy number variation of functional RBMY1 is associated with sperm motility: an azoospermia factor-linked candidate for asthenozoospermia. *Human Reproduction*32, 1.



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

	SC		SG		lepSC		pacSC		PSC		SSC		RS		ES		SP	
	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s	Editin g sites	#gene s
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

**Table 2.** Editing sites in 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%

## Figure legend

**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.

**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 7 genes that only undergo RNA editing in specific germ cells.

**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.

**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.

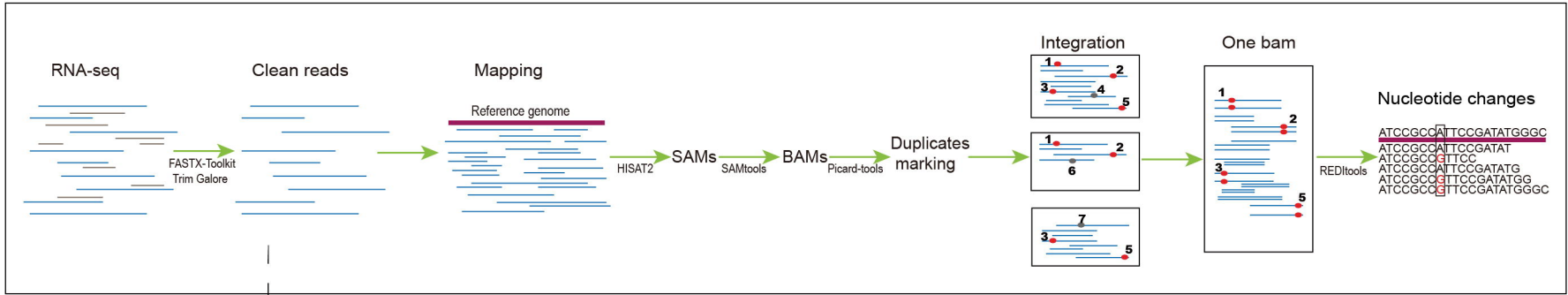
**Figure 5.** Genes with different numbers of editing events. (A) Proportion of genes with one editing event and genes with two or more editing events in germ cells. (B) Proportion of genes with one editing event and genes with two or more editing events in sertoli cells. (C) Proportion of genes with 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.

**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.

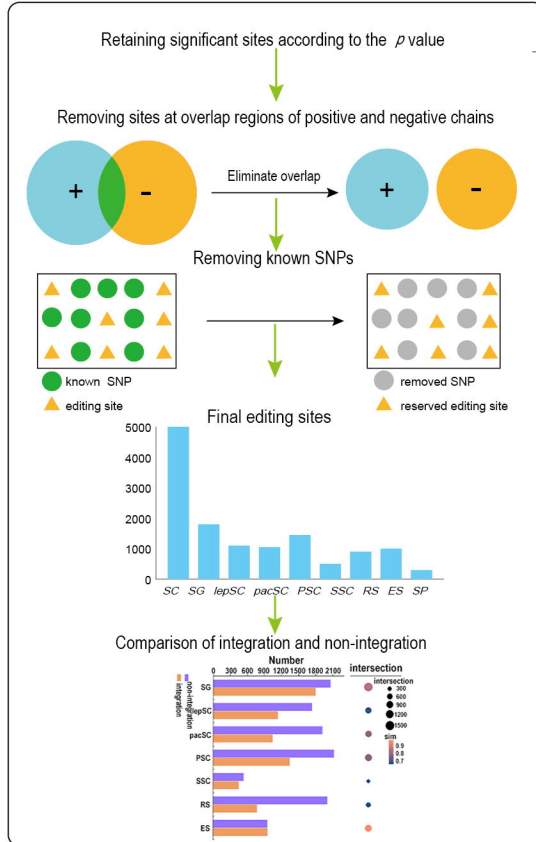
**Figure 7.** Significant editing events for spermatogenesis. (A) All sites in male germ cells and editing sites only in germ cells. (B) Function of genes that have editing sites only in germ cells. (C) The function of genes which are edited in multiple stages.

**Figure 8.** Validation of RNA editing site with Sanger sequencing. Sanger sequencing of cDNA and gDNA, at the site chr14:75719719 of gene *Cog3* the genomic A is highly edited to I.

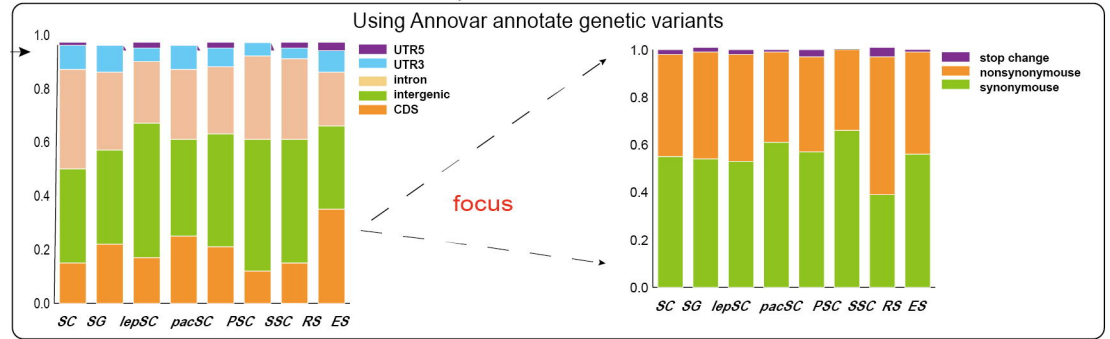
## Step 1: Data preprocessing



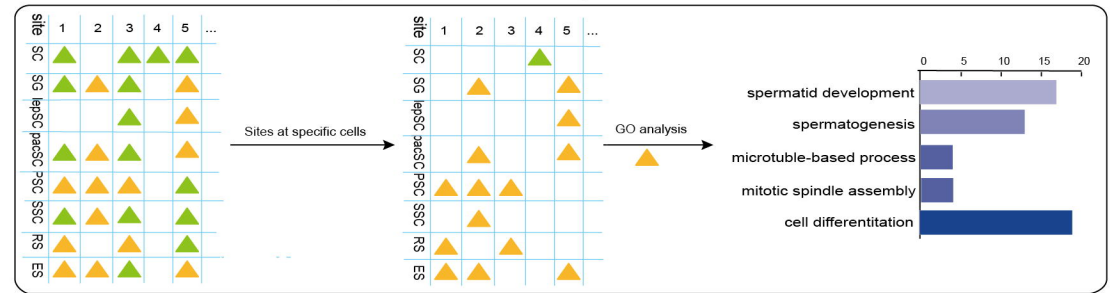
## Step 2: The identification of editing sites



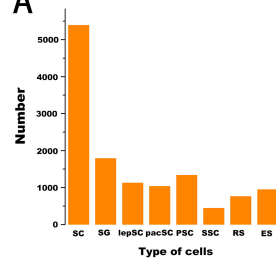
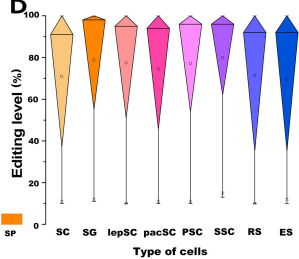
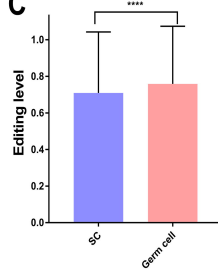
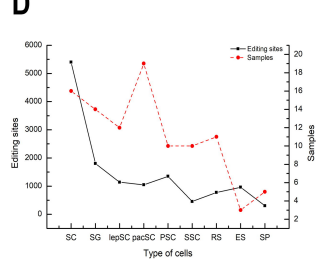
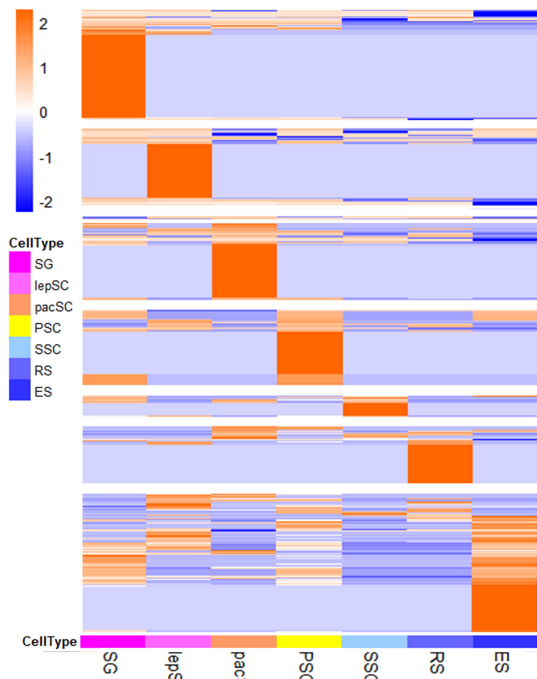
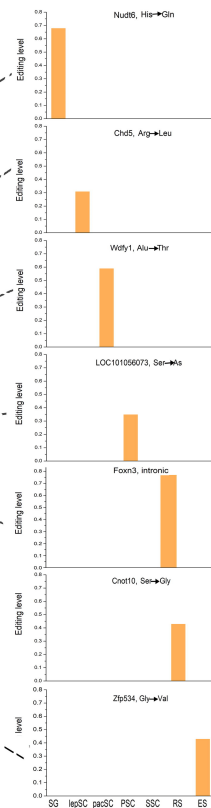
## Step 3: Annotation

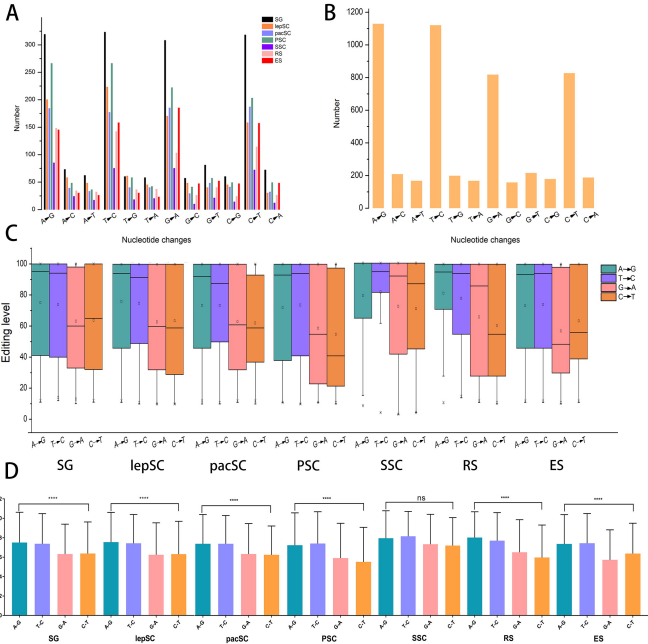


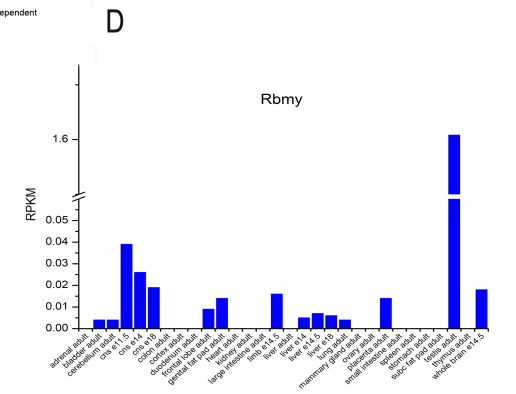
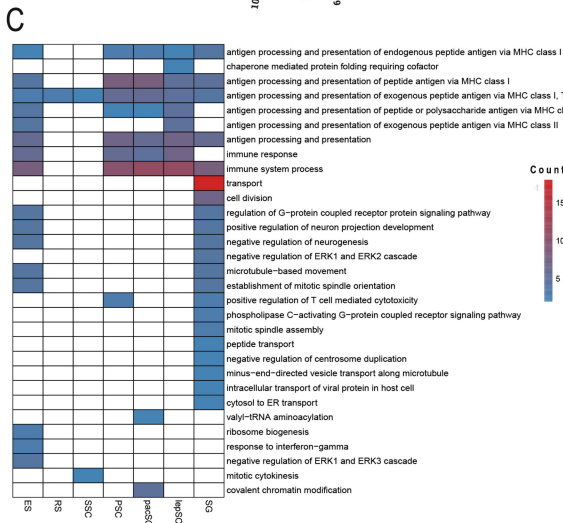
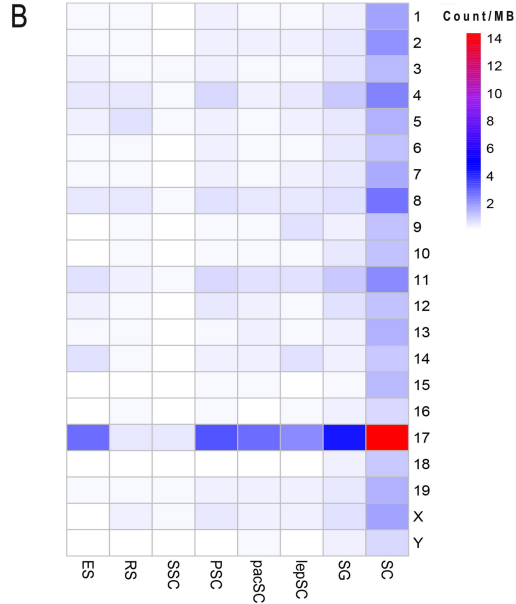
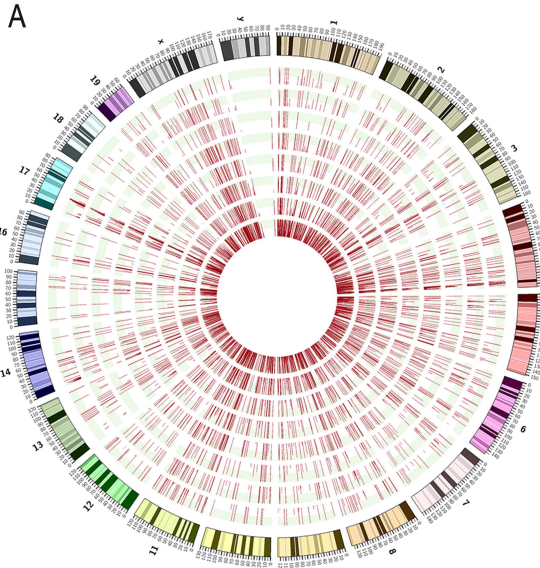
## Step 4: Different sites between germ cells and sertoli cells



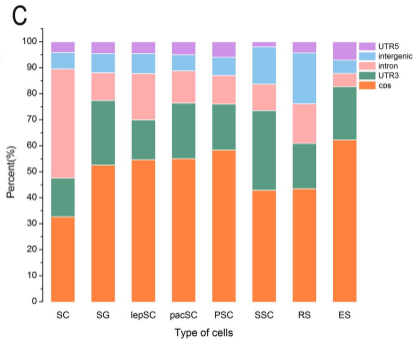
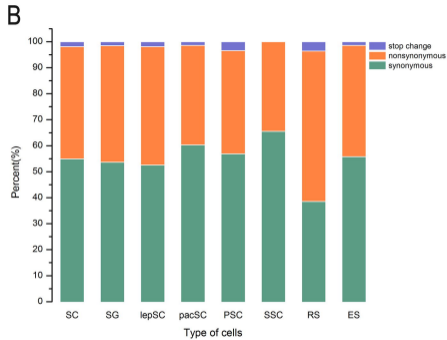
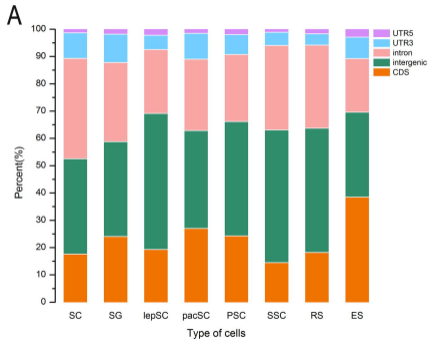


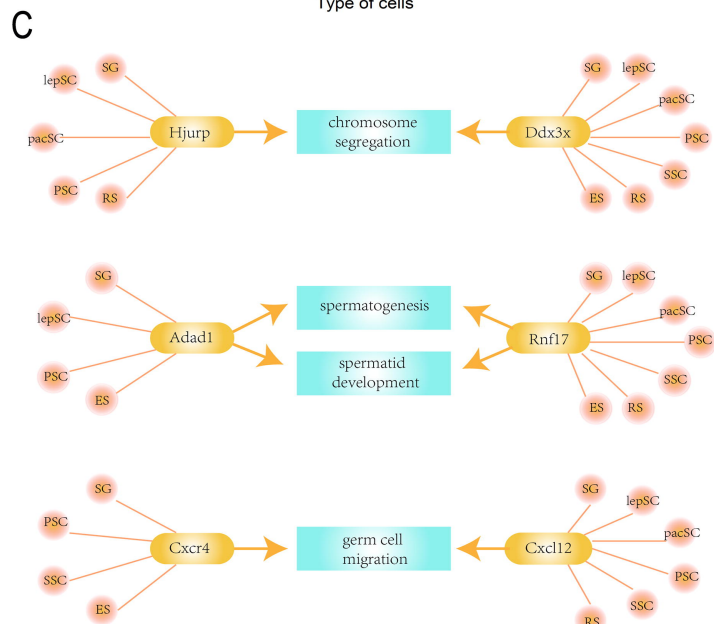
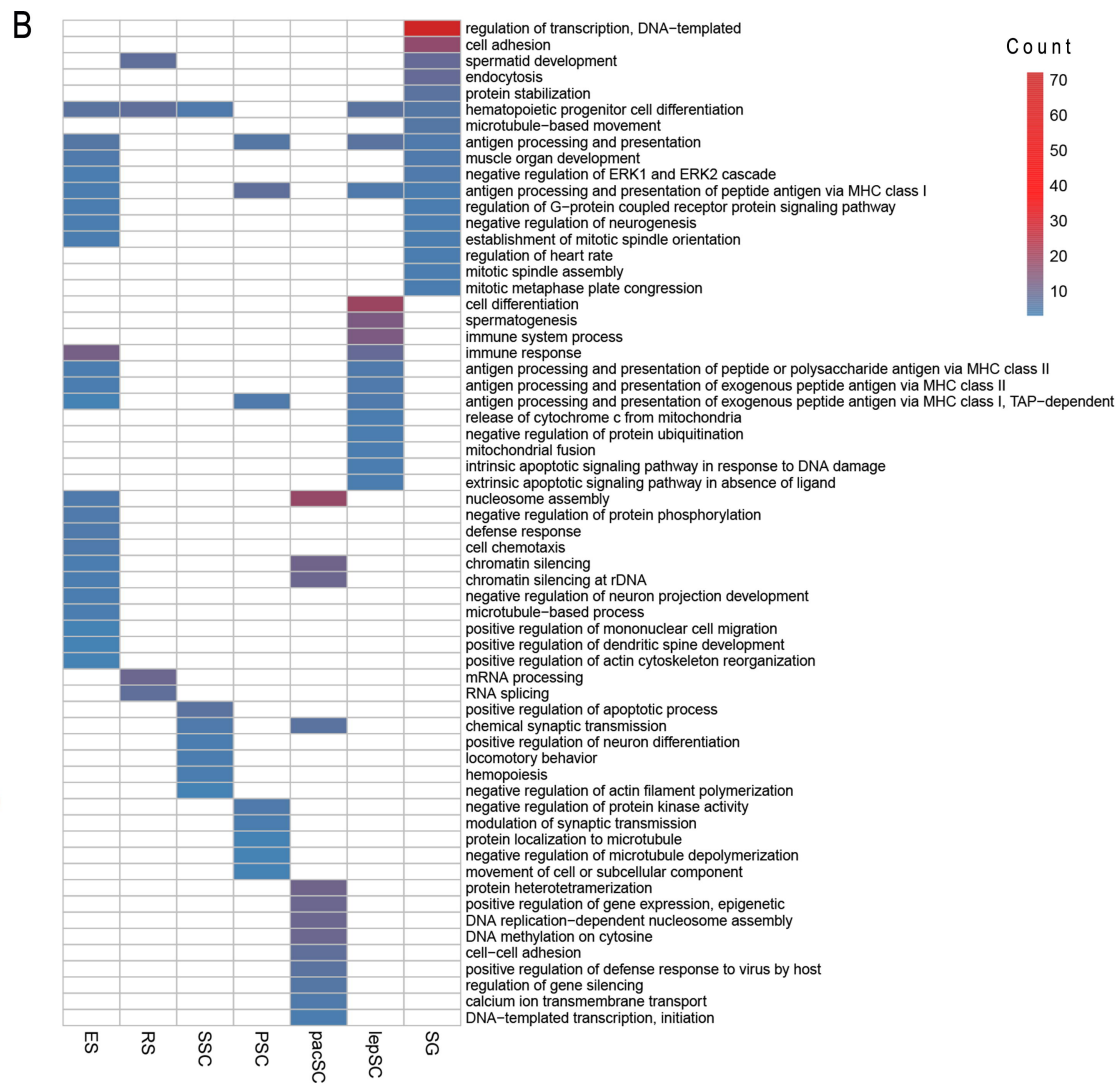
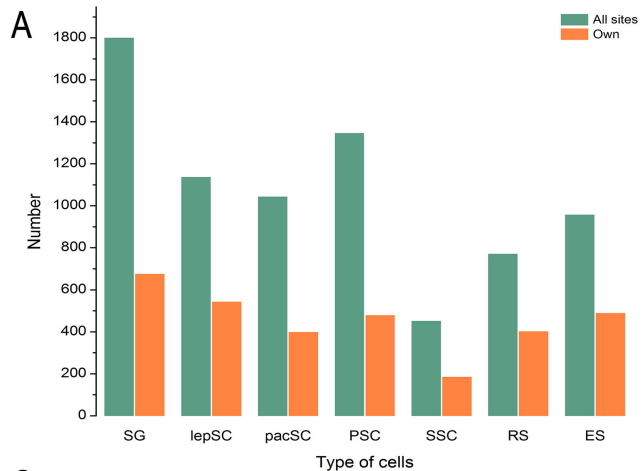
**A****B****C****D****E****F**





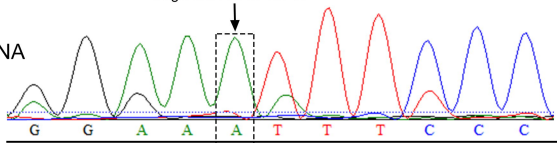






*Cog3* chr14:75719719

gDNA



cDNA

