Molecular Omics

ROYAL SOCIETY OF CHEMISTRY

RESEARCH ARTICLE

View Article Online
View Journal

Cite this: DOI: 10.1039/c9mo00163h

Specific expression and alternative splicing of mouse genes during spermatogenesis†

Qun Li, Tongtong Li, Xia Xiao, Dawood Warraich Ahmad, Ning Zhang, Hao Li, Ziyu Chen, Junyao Hou and Mingzhi Liao *

Received 6th November 2019,
Accepted 9th March 2020

DOI: 10.1039/c9mo00163h

rsc.li/molomics

Considering the high abundance of spliced RNAs in testis compared to other tissues, it is needed to construct the landscape of alternative splicing during spermatogenesis. However, there is still a lack of the systematic analysis of alternative RNA splicing in spermatogenesis. Here, we constructed a landscape of alternative RNA splicing during mouse spermatogenesis based on integrated RNA-seq data sets. Our results presented several novel alternatively spliced genes (*Eif2s3y*, *Erdr1 Uty* and *Zfy1*) in the Y chromosome with a specific expression pattern. Remarkably, the alternative splicing genes were grouped into co-expression networks involved in the microtubule cytoskeleton organization and post-transcriptional regulation of the gene expression, indicating the potential pathway to germ cell generation. Furthermore, based on the co-expression networks, we identified *Atxn2l* as a potential key gene in spermatogenesis, which presented dynamic expression patterns in different alternative splicing types. Ultimately, we proposed splicing regulatory networks for understanding novel and innovative alternative splicing regulation mechanisms during spermatogenesis. In summary, our research provides a systematic analysis of alternative RNA splicing and some novel spliced genes related to spermatogenesis.

Introduction

Spermatogenesis refers to the process of sperm development from primordial germ cells. During their development, the germ cells must maintain information and structural connection with the Sertoli cells.[1–3] Alternative splicing (AS) is an important complex biological regulatory mechanism that can increase the transcriptome complexity and its disruption can lead to human diseases.[4,5] Noticeably, alternative splicing is closely related to sperm maturation, and the dysregulation of the spliceosome can lead to obstructive azoospermia. In fact, AS occurs more frequently in highly complex organisms and tissues with a cell-type or tissue-specific manner. However, the regulation mechanism of AS remains ambiguous.

Several studies in recent years have focused on AS in germ cells. *Ptbp2*, *RNABP9 MRG5* and *BCAS2* are involved in the regulation of spermatogenesis.[6–9] AS plays an important role in gonadal sex determination in mammals.[10] The Y chromosome harbors a number of genes essential for testis development. However, it is not clear how the genes on the Y chromosome are spliced and how they work together in a regulated network.

We built an AS landscape for four spermatogenic cell types based on RNA-seq data sets and these AS events were tightly regulated in a stage-specific manner by specific co-expression networks. Here, we demonstrated that the genes in the Y chromosome presented variable alternative splicing types with specific expressions. Most importantly, *Eif2s3y*, a key gene on the Y chromosome, has different transcripts corresponding to different AS events. In addition, we identified several key AS genes during spermatogenesis by splicing regulated networks. Interestingly, *Atxn2l*, identified by the co-expression network, had variable expressions corresponding to different AS types. Collectively, these findings can provide a new understanding of AS during spermatogenesis by the co-expression regulated network.

Methods

RNA-seq

We collected and analyzed raw RNA sequence reads of four cell types (spermatogonia, spermatocytes, round spermatid and spermatid) from Gene Expression Omnibus. Raw RNA-seq data were trimmed by Trim Galore and mapped with HISAT2.[11] Then, we used SAM tools[12] to convert SAM (Sequence Alignment Map) to BAM (Binary Alignment Map) followed by sorting the output BAM files. We identified transcript and quantification from the aligned clean RNA-seq datasets with StringTie[13] and R package ballgown. We picked raw data when they met the following requirements: (1) detected in a normal physiological state without any disorder impacts; (2) subjected to next-generation

College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: liaomingzhi83@163.com

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9mo00163h

RNA-seq (100 bp, paired-end, strand-specific) with Illumina Hiseq platform; (3) retained for alternative splicing analysis when the mapping rate was higher than 90%; (4) retained for alternative splicing analysis when data had more than two biological replicates. Available RNA-seq datasets were downloaded (GSE100964, GSE75826, GSE95138 and GSE65498)[14–17] and processed as above.

Detection of AS

Detection of AS was carried out with rMATS.[18] We used sorted BAM files obtained from above as input to rMATS by comparing them to Sertoli Cell samples. Also, we provided the GTF (gene transfer format) file downloaded from Ensembles input to rMATS. We performed rMATS on all the samples at four developmental stages and generated tables of different AS events detected below the 0.05 FDR threshold.

Gene ontology

We used Cytoscape application BiNGO[19] to generate an input file for application Enrichment Map[20] to the visualized network of GO. We used R packages cluster Profiler[21] and org.Mm.eg.db to enrich genes and visualized GO terms with REVIGO.[22]

Detection of splice motif

We used Picard (http://broadinstitute.github.io/picard/) to integrate the indexed bam files and extracted ±5 bp splice site sequence. We visualized splice site with R packages ggseqlogo (method = "prob").[23]

WGCNA

Network construction was performed using the blockwise Modules function in the WGCNA package.[24] We picked the power of 12 to construct the co-expression matrix. The adjacency between the genes was calculated, and the similarity between the genes was calculated according to the adjacency. We set the minimum module size to 30 genes. We visualized modules by the plot Dendro And Colors function and exported interaction of the genes in the main module by export Network to Cytoscape functions.

Results

Overview of AS during spermatogenesis

To investigate and assess the difference in AS during spermatogenesis, different AS types (alternative 3′ and 5′ splice sites

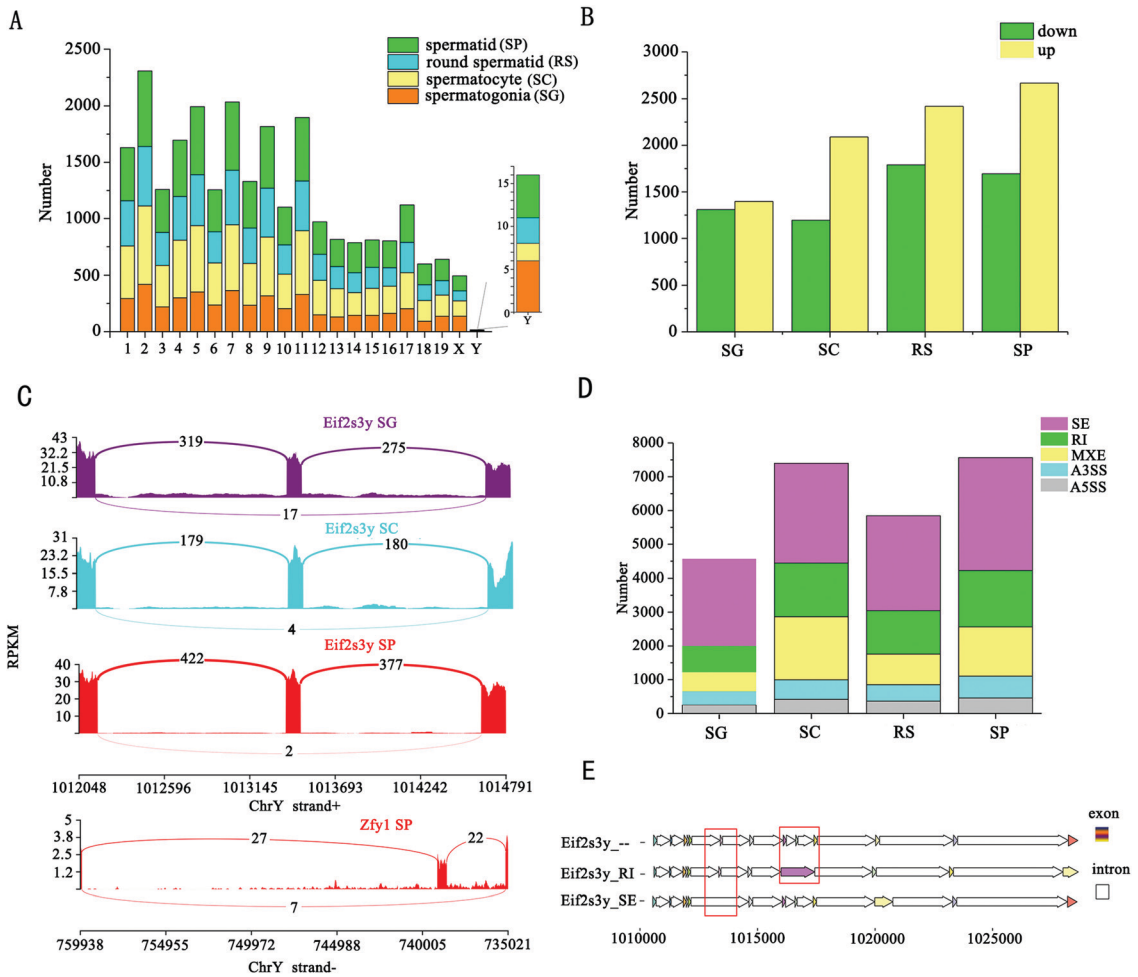

Fig. 1 Alternative splicing in spermatogenesis. (A) Distribution of developmental stage alternative splicing events across chromosomes. (B) The number of up and down expressed alternative splicing genes compared to Sertoli cells. (C) Sashimi plots of *Eif2s3y* and *zfy1*. (D) Distribution of different alternative splicing types across the developmental stage. (E) Different transcripts of *Eif2s3y* corresponding to different alternative splicing types.

(A3SS, A5SS), mutually exclusive exons (MXEs), retained introns (RIs) and skipped exons (SEs)) from clean reads were quantified using rMATS. In this study, significant AS events (FDR $\leq$ 0.05) were identified during spermatogenesis compared to that for the Sertoli cells.

In total, 5400 genes were predicted to be subjected to AS during spermatogenesis and 1505 AS genes were present in all the stages (ESI† S1). However, 25379 significant AS events were detected. This means that one gene can be alternatively spliced more than once. We counted the significant AS events on different chromosomes (Fig. 1A). Although the Y chromosome lost most of its genes,[25] we found that there was a significant difference in the numbers of AS between the Y chromosome and other chromosomes. Excitingly, four genes (*Eif2s3y*, *Erdr1*, *Uty* and *Zfy1*) in the Y chromosome were found to present AS events. Strikingly, the spermatogonia proliferation factor *Eif2s3y* plays an important role in initiating spermatogenesis, and *Uty* has a huge splicing frequency.[26,27] A lack of both *Zfy1* and *Zfy2* genes in mice leads to male infertility.[28] The Sashimi plots show the skipped exons of *Eif2s3y* and *zfy1* (Fig. 1C and E). Then, differentially expressed AS genes were detected compared to that for the Sertoli cells, and the results showed

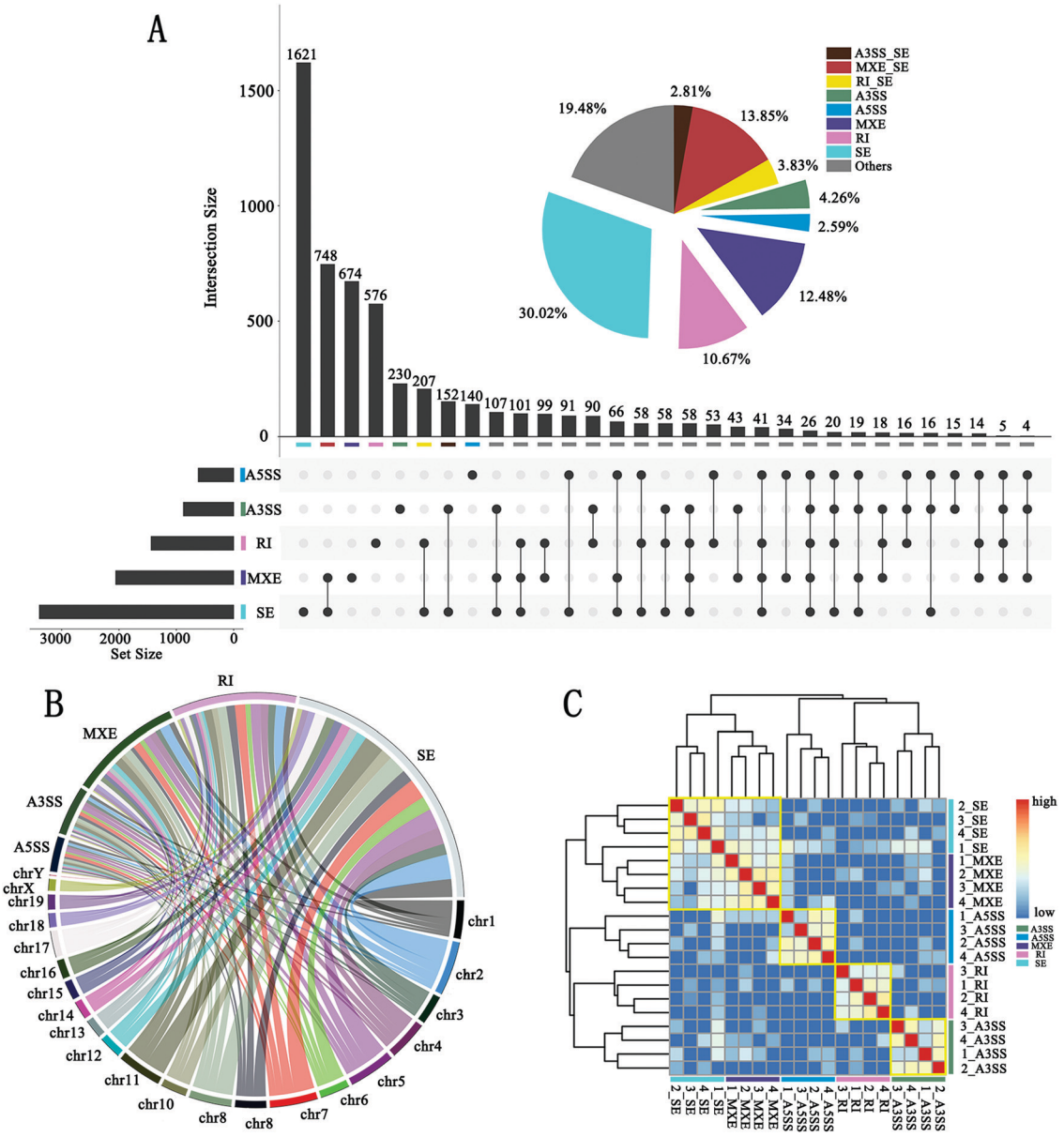

Fig. 2 Alternative splicing types in spermatogenesis. (A) UpSet plots show the alternative splicing types of a gene in spermatogenesis. The pie chart shows the distribution of different alternative splicing types; color block of histogram corresponds to the part of the pie chart. (B) Circus plot shows the distribution of alternative splicing types in different chromosomes. (C) Clustering of alternative splicing types based on the similarity of differential alternative splicing patterns in spermatogenesis. The digital number 1, 2, 3 and 4 represent spermatogonia, spermatocyte, round spermatid and spermatid, respectively. Red and blue indicate high and low similarity, respectively.

that the number of up-expressed AS genes in the four cell types was higher than that of the down-expressed AS genes (Fig. 1B). We then analyzed the distribution of AS events and found that SE is the most variable AS type followed by RI (Fig. 1D), while RI is a common AS type in plants.[29] However, the aboundance of SE is considered to be the least one among different AS form in animals.[30] In contrast, widespread intron retention in mammals regulates transcriptomes by reducing the expression of relatively low abundance transcripts.[31]

Specific AS pattern is required for gene expression in spermatogenesis

The overall function of AS involves generating multiple transcripts and increasing the complexity of gene expressions.[32] In recent years, some studies on the relationship between AS and diseases have been reported.[33,34] However, the AS mechanism of pre-mRNA is complex and remains to be explained.

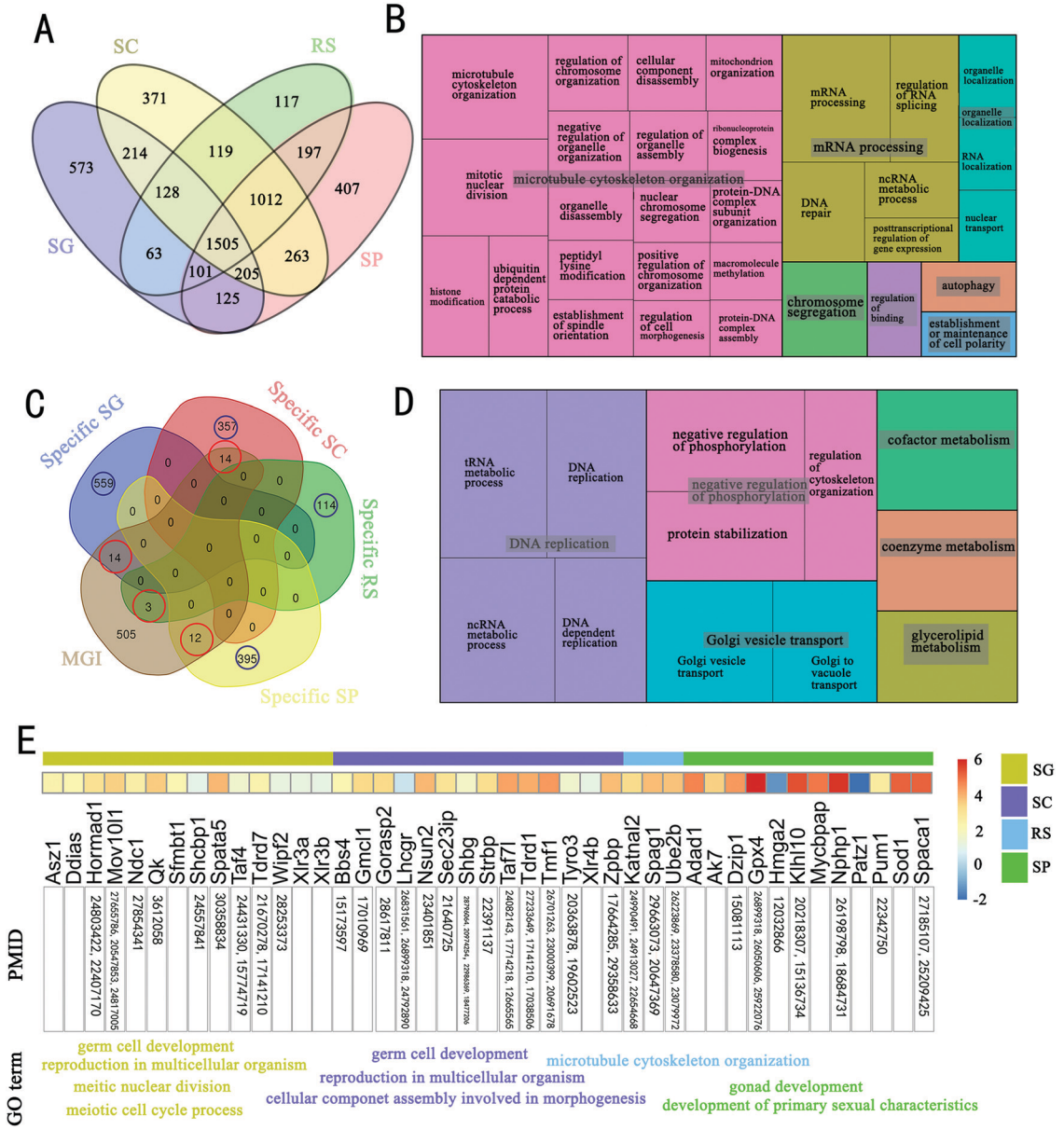

Fig. 3 Functions of alternative splicing genes in spermatogenesis. (A) Venn diagram indicating the overlapping of alternative splicing genes of 4 cell types in spermatogenesis; cyan circle indicates common alternative splicing genes in 4 cell types. (B and D) GO enrichment summarized and visualized as a TreeMap of common alternative splicing genes (cyan) and stage-specific alternative splicing genes (blue) using REVIGO, respectively. TreeMap's panel size is inversely proportional to the enrichment *p*-value. (C) Venn diagram indicating the overlapping of alternative splicing genes of 4 specific-stage cell types and Mouse Genome Informatics datasets spermatogenesis genes. Blue circles indicate specific-stage alternative splicing genes. Red circles indicate specific-stage spermatogenesis alternative splicing genes. SG: spermatogonia; SC: spermatocyte; RS: round spermatid; SP: spermatid. MGI: Mouse Genome Informatics. (E) Annotation of specific-stage spermatogenesis alternative splicing genes (red). Top: Heatmap representing the expression levels of specific-stage spermatogenesis alternative splicing genes (red). The middle box represents the study of PMID associated with the spermatogenesis of the corresponding gene. Below: color of GO term corresponds to cell types.

We found that a gene may have several AS types during spermatogenesis (Fig. 2A).[35] We also found that about 60% genes (30.02% SE, 10.67% RI, 12.48% MXE, 2.59% A5SS and 4.26% A3SS) have only one AS type, while 26 genes (less than 1%) have 5 AS types. Then, the distribution of different AS types on the chromosomes was shown with a Circos plot (Fig. 2B).

However, whether the AS types with similar cell types exhibit similarity is unknown. To address this question, similarity score based on the AS types in each spermatogenesis stage was computed. This similarity is measured as the overlapped gene's number divided by the minimum number of differential AS events between two cell types. The results show that the same AS with the same cell type presents a similar AS pattern according to our hierarchical clustering analysis (Fig. 2C). Furthermore, four blocks among them were found with yellow boxes in the clustering results. What made us interested is that SE had a similar AS pattern to that of MXE, and they were clustered into a block.

The characteristics of AS genes

In order to observe the exact AS gene sets from different stages during spermatogenesis, we detected their intersection between each other. Interestingly, we found that more than one quarter of AS genes are shared among different stages (Fig. 3A). Thus, we then performed GO enrichment analysis of the shared AS genes to detect their functional characteristics. The results show that the clusters of GO terms have a strong relationship with the microtubule cytoskeleton organization and RNA processing, including the regulation of RNA splicing and the posttranscriptional regulation of gene expressions (Fig. 3B).

Finally, we focused on the stage-specific AS genes (SSAS). The results show that they overlap with some known spermatogenesis genes from MGI (Mouse Genome Informatics, brown area) (Fig. 3C), and 43 genes are known spermatogenesis SSAS (red circles). The GO enrichment analysis of the genes in blue circles shows that these genes are associated with the DNA or RNA process without splicing terms, phosphorylation and Golgi transport (Fig. 3D). In order to explore the characteristics for the known spermatogenesis SSAS, we presented their average expression level and found that they normally showed a relatively high expression level (Fig. 3E). For these genes, we performed manual literature survey to certify their function. Our PMID panel consistently showed some famous genes related to spermatogenesis (Fig. 3E). For example, *Gpx4*

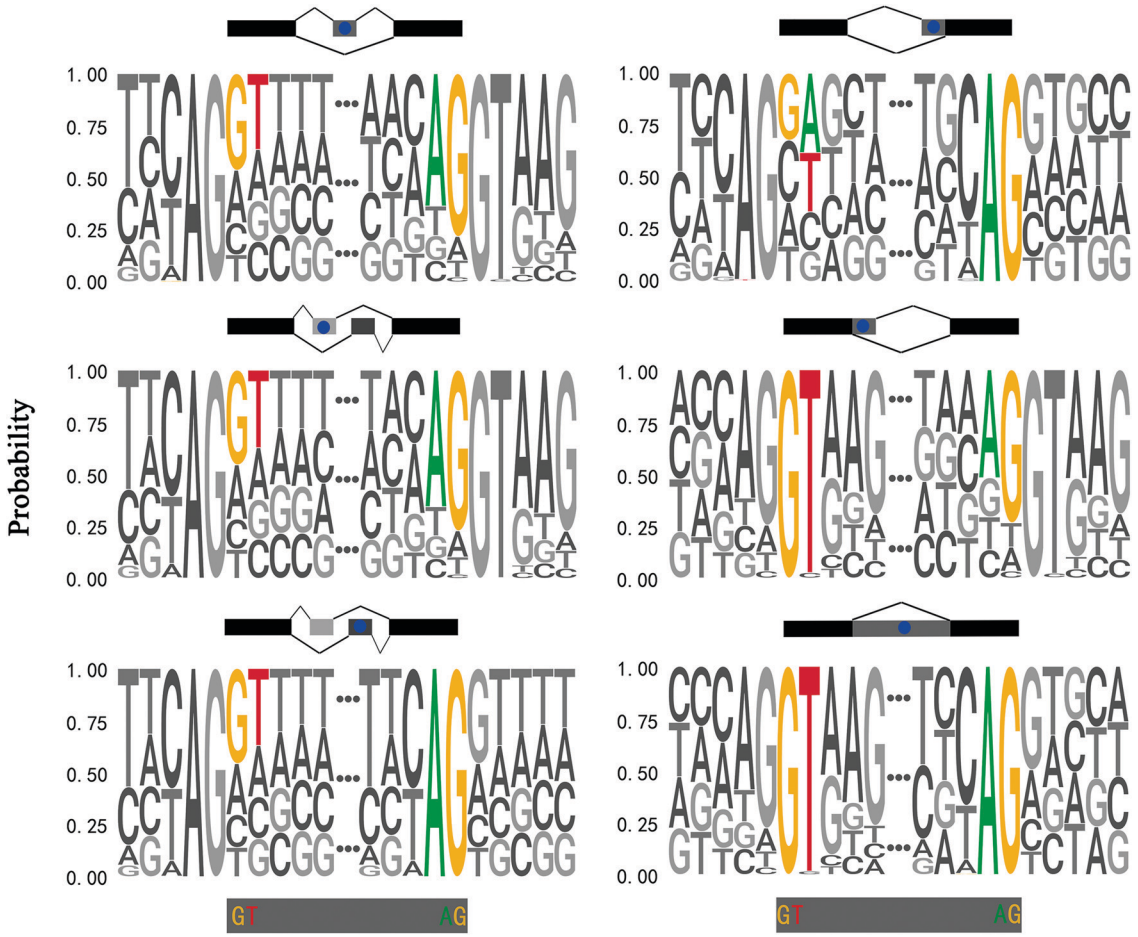

Fig. 4 Sequence logos at splice sites from alternative splicing sites. The size of the letter corresponds to the frequency at which the base appears at each position in the splicing site, and the color of the letter corresponds to splicing site conserved core. Blocks and lines represent different alternative splicing events. Blue dots are marked on the spliced intron. The grey blocks below indicate the region of alternative splicing.

(glutathione peroxidase 4), considered as an anti-apoptotic factor, plays a vital role in male fertility,[36] and the male mice targeted with mutating *Nphp1* are infertile.[37] These genes were highly correlated with germ cell development.

Diverse GT–AG alternative splice motif

The splicing process is designed to preserve the presence of exons on pre-mRNA to form mature mRNA. The process of AS is mainly divided into the assembly of the spliceosome and transesterification.[32] The splicing factor recognizes the binding domain by its N-terminal RNA and identifies the boundaries of the exons and introns. Thus, the splice site is an important feature of AS. We mapped the AS site sequences to the genome and extracted them. These sequence data are represented in the logo plots (Fig. 4). There is a relatively shorter conservative GT–AG sequence near the splicing site. However, there is still another splicing site motif, which is consistent with previous studies.[38,39] There is a difference in the proportions of GT–AG in different splicing types in spermatogenesis. The proportion of GT–AG in the retained intron (RI) is significantly higher than that for other splicing types. More interestingly, the proportion of the alternative 3′ splice sites of GA-AG is higher than the other AS types. These findings reflect that there are multiple splicing patterns, which require other splicing factors to identify and target splice sites during AS in spermatogenesis.

Weighted gene co-expression network analysis (WGCNA) of AS

Here, 1505 AS genes were grouped into nine modules (the grey module represents the genes that are not grouped into a module) by weighted gene co-expression network analysis (Fig. 5A). Then, gene function enrichment analysis of these modules was performed, and the results showed that they were highly correlated with RNA splicing (Fig. 5D). Module1 (M1), module8 (M8) and module6 (M6) have the same GO term related to the cytoskeleton. Fifty AS genes were clustered into the GO term related to the cytoskeleton. Then, we took them as an example to explore their relationship with spermatogenesis and made intersection with genes from Mouse Genome Informatics.

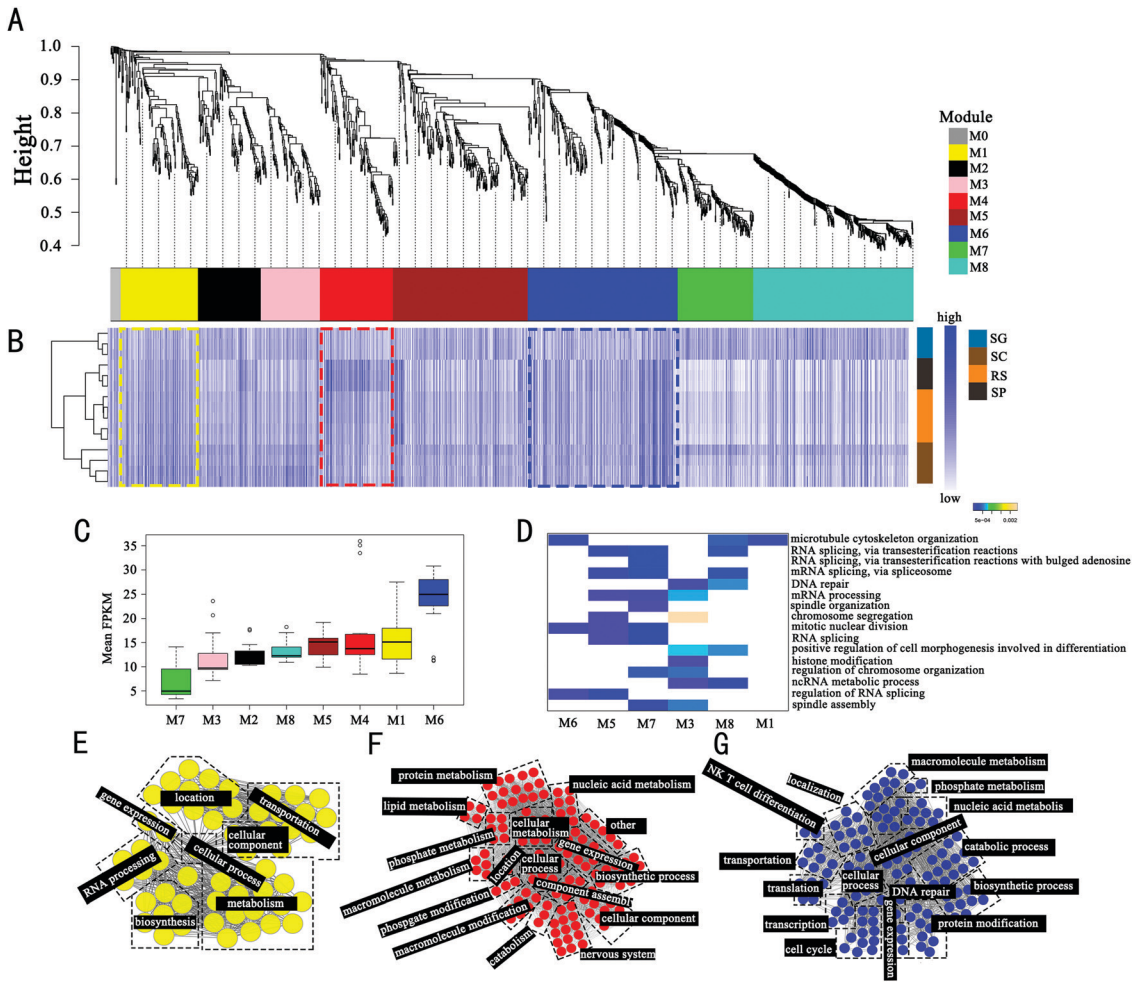

Fig. 5 Weighted gene co-expression network analysis (WGCNA) of mRNA alternative splicing. (A) Dendrogram produced by grouping genes into distinct modules using all alternative splicing genes in spermatogenesis, with the *y*-axis corresponding to the co-expression distance between genes and the *x*-axis to genes. (B) Heatmap of alternative splicing genes, with *y*-axis corresponding to cell types and the *x*-axis to genes. (C) Boxplot of modules' mean FPKM. (D) The overlapping of Go terms (p-value ≤ 0.01) among modules. (E–G) Network visualization of Gene Ontology Enrichment Analysis (GOEA) of module1, module4 and module6 using BiNGO and EnrichmentMap.

Excitingly, there are some interesting discoveries in overlapping genes by referring to the literature. Male mice with *Ttll5* (tubulin tyrosine ligase-like family member 5) deficiency show a significant decline in the reproductive capacity associated with sperm motility deficits;[40] *Poc1a* (Protein of Centriole 1 A) is essential for the normal function of both the Sertoli cells and germ cells, and exon 8 skipping of *Poc1a* leads to germ cell maturation arrest.[41]

These AS genes in different modules were clustered into four parts corresponding to four cell types in spermatogenesis according to their expression levels (Fig. 5B). We then focused on the three modules with top expression levels (color boxes in

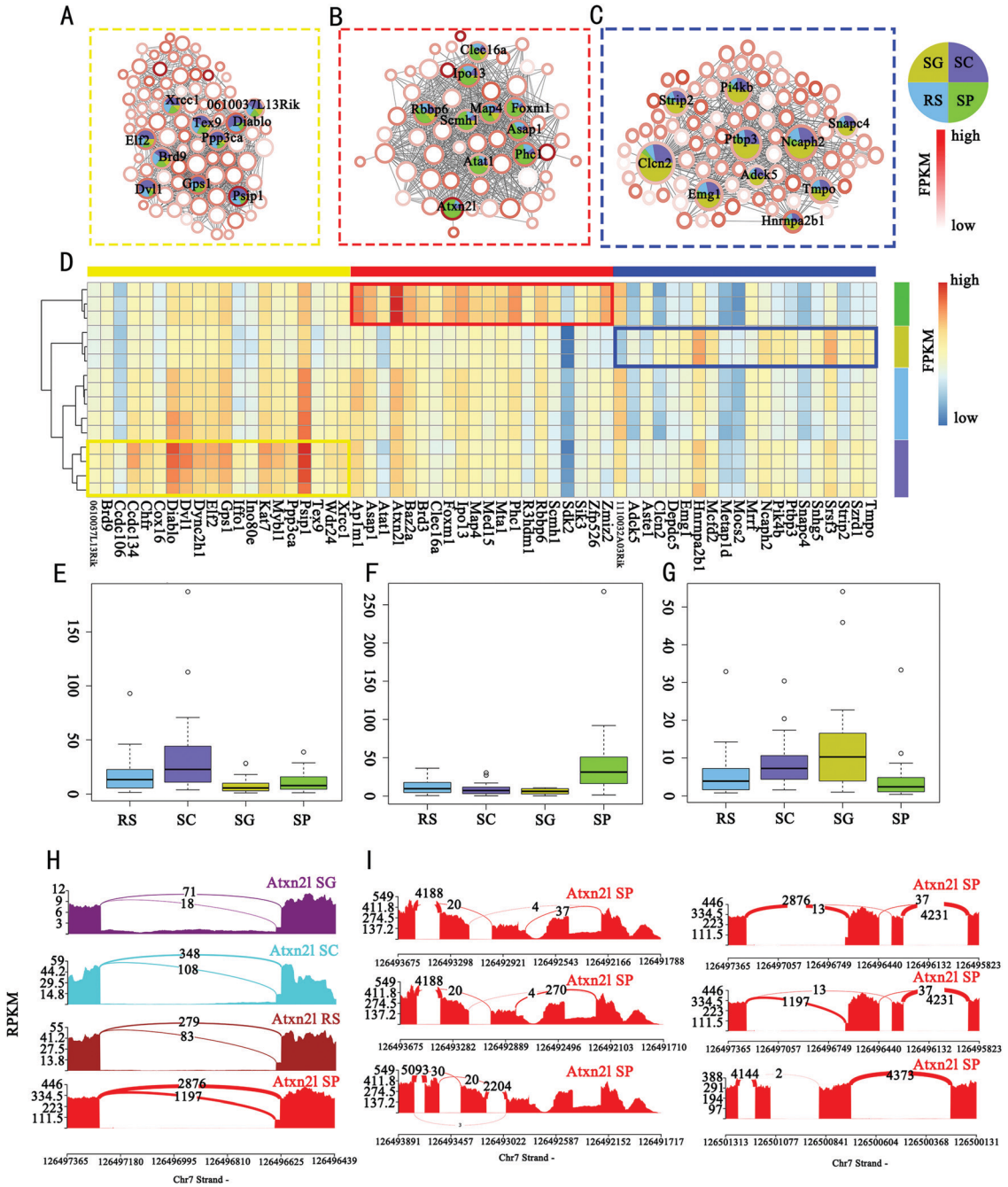

Fig. 6 Hub alternative splicing genes revealed by Weighted Correlation Network Analysis. (A–C) Network visualization of module1, module4 and module6. The top ten genes (highest degree of network) are presented as pie charts. The border-color of the circles indicates genes' expression levels; the circle size indicates the degree of network connection. SG: spermatogenesis; SC: spermatocyte; RS: round spermatid; SP: spermatid. (D) Heatmap of top 20 alternative splicing genes (highest degree of network) corresponding to modules. Higher expression level stage is highlighted with a color box. (E–G) Boxplot of module's mean FPKM. Box colors correspond to cell types. (H) Sashimi A3SS plot of *Atxn2l* of 4 cell types in spermatogenesis. (I) Sashimi MXE plot of *Atxn2l* of spermatid in spermatogenesis.

Fig. 5C) and performed gene ontology enrichment analysis. The results show that they are related to cellular metabolism, cellular location and nervous system (Fig. 5E–G).

Through the co-expression network, we identified 30 AS hub genes (selected by the top-ranked degree of network connection) (Fig. 6A–C). By manually searching the literature, we hoped to find the genes related to AS. Interestingly, several genes among them were found to process their functions through the regulation of AS. *Psip1* and *Ptbp3* were found to play important roles in the post-transcriptional regulation through AS.[42,43] Furthermore, the network presents the interaction between *Phc1* and *Scmh1* (Fig. 6B). Previous research showed that *Phc1* and *Scmh1* are the constituents of the mammalian Polycomb repressive complex 1 and participate in the regulation of spermatogenesis.[44] We speculated that our co-expression network was capable of presenting alternative splicing regulatory relationships during spermatogenesis.

In order to show more connections between the genes, we increased the depth of the network. We then focused on the top 20 genes ranked by degrees from the three modules with the highest expression, which are highlighted with the corresponding color box (Fig. 6D). Interestingly, the three modules show dynamic expression patterns and present different stages with the highest expression (Fig. 6E–G). What makes us most exciting is the results of *Atxn2l* from the red module, which shows the highest expression level in the spermatid stage. *Atxn2l* has a higher expression in testis than in the other tissues.[45] In addition, the knockdown of *Atxn2l* resulted in a 22.9% increase in blastocyst development.[46] These results certified our hypothesis preliminarily and indicated that *Atxn2l* played its role by AS in spermatogenesis. Though *Atxn2l* has common alternative 3′ splicing events during spermatogenesis in the same position (Fig. 6H), it shows a wide AS frequency with six types of MXE in spermatids with different positions (Fig. 6I).

Discussion

As an important type of the post-transcriptional regulation of the gene expression, AS can regulate binding among proteins and nucleic acids, the location of proteins, cell proliferation, cell survival and properties.[47] However, the roles played by AS during spermatogenesis are still unclear. We identified 25379 AS events and found that the AS-related genes present specific patterns during spermatogenesis. Most interestingly, AS events in *Efi2s3y* and *zfy1* were detected, which are located on the Y chromosome and are required for the development of spermatogenesis. As is well-known, the Y chromosome has a lot of decline in the process of evolution, and most of the length of the Y chromosome is not restricted by chromosome pairing, allowing for the accumulation of repeated sequences.[48,49] Our results implicate that the genes in the Y chromosome may play their roles in sex determination and spermatogenesis *via* AS.

Our results showed that the identified AS genes are related to spermatogenesis and most of these genes are consistent with previous studies. Excitingly, we detected 43 AS genes, which are known stage-specific AS genes by intersecting MGI genes with detected genes. These genes were highly related with germ cell development. On the other hand, we also found some novel AS genes, which play a potential role during spermatogenesis. Interesting, we also found that the AS genes identified in spermatogenesis have a strong relationship with the microtubule cytoskeleton organization. Actin filaments and microtubules are critical for the structural support of cells and provide a pathway for transport across epithelial cells.[50] The relationship between spermatogenesis and microtubule cytoskeleton organization needs more research in future.

We extracted a sequence near to the splice site and mapped to the genome. There is a relatively conservative GT–AG sequence near the splicing site. Most animal splicing processes rely on specific spliceosomes, which result in 5′ of introns being GT and 3′ being AG. However, there were some unusual splice site motifs in spermatogenesis. We inferred that there were other splicing factors to identify and target splice sites. Non-GT/AG splicing signals were found for circRNAs in rice, including GC/GG, CA/GC, GG/AG, GC/CG and CT/CC.[51]

We then explored the relationship between AS and expression and performed WGCNA analysis. First, we found that these genes were clustered into different gene modules with specific AS patterns in networks. Then, we focused on three modules with the top highest expression, and we found that these genes were highly correlated to RNA splicing and germ development. Interestingly, we found that the three modules showed a specific-stage expression pattern and we then highlighted them and visualized their related AS events. We identified key genes by selecting the top-ranked degree genes in the co-expression network. Finally, 30 AS hub genes in spermatogenesis with a high degree in network were detected. For these genes, we conducted a manual review of the literature to verify the effectiveness. Among these genes, *Pisp1*, *Scmh1* and *Ptbp3* presented an effective regulation of AS in spermatogenesis. *Atxn2l* has a higher expression in spermatids compared to other development stages and a higher degree of network connections compared to other genes. We found that *Atxn2l* had six kinds of MXE with a high AS frequency in spermatids. The mechanism of *Atxn2l* affecting spermatogenesis by AS needs to be further revealed.

In summary, we demonstrated that many genes, including the genes on the Y chromosome, underwent AS during spermatogenesis, and we partially verified the results through the literature. Many factors participate in the control of AS and AS events are shared among different stages during spermatogenesis and affect gene expression by type-specific and module-specific splicing patterns. We used co-expression networks to study the regulatory relationship between the alternative splicing genes and found the potentially important splicing regulatory gene *Atxn2l*. This research will provide a new insight in understanding the function and features of AS during spermatogenesis.

Conflicts of interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or

associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgements

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).

References

1 Y. Chen, L. Ma, C. Hogarth, G. Wei, M. D. Griswold and M.-H. Tong, *Development*, 2016, **143**, 1502–1511.
2 M. D. Griswold, *Semin. Cell Dev. Biol.*, 1998, **9**, 411–416.
3 Y. Hai, J. Hou, Y. Liu, Y. Liu, H. Yang, Z. Li and Z. He, *Semin. Cell Dev. Biol.*, 2014, **29**, 66–75.
4 M. M. Scotti and M. S. Swanson, *Nat. Rev. Genet.*, 2015, **17**, 19.
5 T. W. Nilsen and B. R. Graveley, *Nature*, 2010, **463**, 457.
6 M. M. Hannigan, L. L. Zagore and D. D. Licatalosi, *Cell Rep.*, 2017, **19**, 2598–2612.
7 J. Bao, C. Tang, J. Li, Y. Zhang, B. P. Bhetwal, H. Zheng and W. Yan, *PLoS Genet.*, 2014, **10**, e1004825.
8 N. Iwamori, K. Tominaga, T. Sato, K. Riehle, T. Iwamori, Y. Ohkawa, C. Coarfa, E. Ono and M. M. Matzuk, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E5408–E5415.
9 W. Liu, F. Wang, Q. Xu, J. Shi, X. Zhang, X. Lu, Z. A. Zhao, Z. Gao, H. Ma, E. Duan, F. Gao, S. Gao, Z. Yi and L. Li, *Nat. Commun.*, 2017, **8**, 14182.
10 I. U. Haussmann, Z. Bodi, E. Sanchez-Moran, N. P. Mongan, N. Archer, R. G. Fray and M. Soller, *Nature*, 2016, **540**, 301–304.
11 D. Kim, B. Langmead and S. L. Salzberg, *Nat. Methods*, 2015, **12**, 357.
12 H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin, S. Genome Project Data Processing, *Bioinformatics*, 2009, **25**, 2078–2079.
13 M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell and S. L. Salzberg, *Nat. Biotechnol.*, 2015, **33**, 290.
14 L. Wichman, S. Somasundaram, C. Breindel, D. M. Valerio, J. R. McCarrey, C. A. Hodges and A. M. Khalil, *Biol. Reprod.*, 2017, **97**, 313–323.
15 X. Lin, M. Han, L. Cheng, J. Chen, Z. Zhang, T. Shen, M. Wang, B. Wen, T. Ni and C. Han, *RNA Biol.*, 2016, **13**, 1011–1024.
16 C. Naro, A. Jolly, S. Di Persio, P. Bielli, N. Setterblad, A. J. Alberdi, E. Vicini, R. Geremia, P. De la Grange and C. Sette, *Dev. Cell*, 2017, **41**(82–93), e84.
17 K. S. McClelland, K. Bell, C. Larney, V. R. Harley, A. H. Sinclair, A. Oshlack, P. Koopman and J. Bowles, *Biol. Reprod.*, 2015, **92**, 145.
18 S. Shen, J. W. Park, Z. X. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou and Y. Xing, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E5593–E5601.
19 S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448–3449.
20 R. Isserlin, D. Merico, V. Voisin and G. D. Bader, *F1000Research*, 2014, **3**, 141.
21 G. Yu, L.-G. Wang, Y. Han and Q.-Y. He, *OMICS*, 2012, **16**, 284–287.
22 F. Supek, M. Bošnjak, N. Škunca and T. Šmuc, *PLoS One*, 2011, **6**, e21800.
23 O. Wagih, *Bioinformatics*, 2017, **33**, 3645–3647.
24 P. Langfelder and S. Horvath, *BMC Bioinf.*, 2008, **9**, 559.
25 Y. Q. S. Soh, J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves, P. J. Minx, R. S. Fulton, C. Kremitzki, N. Koutseva, J. L. Mueller, S. Rozen, J. F. Hughes, E. Owens, J. E. Womack, W. J. Murphy, Q. Cao, P. de Jong, W. C. Warren, R. K. Wilson, H. Skaletsky and D. C. Page, *Cell*, 2014, **159**, 800–813.
26 I. Laaser, F. J. Theis, M. H. D. Angelis, H.-J. Kolb and J. Adamski, *OMICS*, 2011, **15**, 141–154.
27 Y. Yamauchi, J. M. Riel, V. Ruthig and M. A. Ward, *PLoS Genet.*, 2016, **11**, e1005476.
28 N. Vernet, S. K. Mahadevaiah, Y. Yamauchi, F. Decarpentrie, M. J. Mitchell, M. A. Ward and P. S. Burgoyne, *PLoS Genet.*, 2014, **10**, e1004444.
29 Y. Marquez, J. W. Brown, C. Simpson, A. Barta and M. Kalyna, *Genome Res.*, 2012, **22**, 1184–1195.
30 E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge, *Nature*, 2008, **456**, 470.
31 U. Braunschweig, N. L. Barbosa-Morais, Q. Pan, E. N. Nachman, B. Alipanahi, T. Gonatopoulos-Pournatzis, B. Frey, M. Irimia and B. J. Blencowe, *Genome Res.*, 2014, **24**, 1774–1786.
32 Y. Wang, J. Liu, B. O. Huang, Y.-M. Xu, J. Li, L.-F. Huang, J. Lin, J. Zhang, Q.-H. Min, W.-M. Yang and X.-Z. Wang, *Biomed. Rep.*, 2015, **3**, 152–158.
33 E. El Marabti and I. Younis, *Front. Mol. Biosci.*, 2018, **5**, 80.
34 A. Srebrow and A. R. Kornblihtt, *J. Cell Sci.*, 2006, **119**, 2635.
35 J. R. Conway, A. Lex and N. Gehlenborg, *Bioinformatics*, 2017, **33**, 2938–2940.
36 I. Ingold, M. Aichler, E. Yefremova, A. Roveri, K. Buday, S. Doll, A. Tasdemir, N. Hoffard, W. Wurst, A. Walch, F. Ursini, J. P. Friedmann Angeli and M. Conrad, *J. Biol. Chem.*, 2015, **290**, 14668–14678.
37 S.-T. Jiang, Y.-Y. Chiou, E. Wang, H.-K. Lin, S.-P. Lee, H.-Y. Lu, C.-K. L. Wang, M.-J. Tang and H. Li, *Hum. Mol. Genet.*, 2008, **17**, 3368–3379.
38 Y. Shen, Z. Zhou, Z. Wang, W. Li, C. Fang, M. Wu, Y. Ma, T. Liu, L.-A. Kong, D.-L. Peng and Z. Tian, *Plant Cell*, 2014, **26**, 996–1008.
39 T. A. Thanaraj and F. Clark, *Nucleic Acids Res.*, 2001, **29**, 2581–2593.
40 N. Bedoni, L. Haer-Wigman, V. Vaclavik, V. H. Tran, P. Farinelli, S. Balzano, B. Royer-Bertrand, M. E. El-Asrag, O. Bonny, C. Ikonomidis, Y. Litzistorf, K. Nikopoulos, G. G. Yioti, M. I. Stefaniotou, M. McKibbin, A. P. Booth, J. M. Ellingford, G. C. Black, C. Toomes, C. F. Inglehearn, C. B. Hoyng, N. Bax, C. C. W. Klaver, A. A. Thiadens, F. Murisier, D. F. Schorderet, M. Ali, F. P. M. Cremers,

S. Andréasson, F. L. Munier and C. Rivolta, *Hum. Mol. Genet.*, 2016, **25**, 4546–4555.
41 K. A. Geister, M. L. Brinkmeier, L. Y. Cheung, J. Wendt, M. J. Oatley, D. L. Burgess, K. M. Kozloff, J. D. Cavalcoli, J. M. Oatley and S. A. Camper, *PLoS Genet.*, 2015, **11**, e1005569.
42 M. M. Pradeepa, H. G. Sutherland, J. Ule, G. R. Grimes and W. A. Bickmore, *PLoS Genet.*, 2012, **8**, e1002717.
43 L.-Y. Tan, P. Whitfield, M. Llorian, E. Monzon-Casanova, M. D. Diaz-Munoz, M. Turner and C. W. J. Smith, *Nucleic Acids Res.*, 2015, **43**, 5586–5600.
44 Y. Takada, K.-i. Isono, J. Shinga, J. M. A. Turner, H. Kitamura, O. Ohara, G. Watanabe, P. B. Singh, T. Kamijo, T. Jenuwein, P. S. Burgoyne and H. Koseki, *Development*, 2007, **134**, 579–590.
45 F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L.-H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y.-C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K.-H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X.-Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren and E. C. Mouse, *Nature*, 2014, **515**, 355–364.
46 N. Gross, M. G. Strillacci, F. Peñagaricano and H. Khatib, *Sci. Rep.*, 2019, **9**, 20347.
47 O. Kelemen, P. Convertini, Z. Zhang, Y. Wen, M. Shen, M. Falaleeva and S. Stamm, *Gene*, 2013, **514**, 1–30.
48 M. A. Jobling and C. Tyler-Smith, *Nat. Rev. Genet.*, 2017, **18**, 485.
49 D. Bachtrog, *Nat. Rev. Genet.*, 2013, **14**, 113–124.
50 H. Wang, R. Zhao, C. Guo, S. Jiang, J. Yang, Y. Xu, Y. Liu, L. Fan, W. Xiong, J. Ma, S. Peng, Z. Zeng, Y. Zhou, X. Li, Z. Li, X. Li, D. C. Schmitt, M. Tan, G. Li and M. Zhou, *Sci. Rep.*, 2016, **6**, 21776.
51 C.-Y. Ye, X. Zhang, Q. Chu, C. Liu, Y. Yu, W. Jiang, Q.-H. Zhu, L. Fan and L. Guo, *RNA Biol.*, 2016, **14**, 1055–1063.

