Asian Journal of Andrology (2022) 24, 248–254
www.asiaandro.com; www.ajandrology.com

Open Access

INVITED ORIGINAL ARTICLE

Male Infertility

Investigation of the genetic etiology in male infertility with apparently balanced chromosomal structural rearrangements by genome sequencing

Matthew Hoi Kin Chau[1,2,3,*], Ying Li[1,2,*], Peng Dai[4], Mengmeng Shi[1,2], Xiaofan Zhu[4], Jacqueline Pui Wah Chung[1], Yvonne K Kwok[1,2], Kwong Wai Choy[1,2,3,5], Xiangdong Kong[4], Zirui Dong[1,2,3]

Apparently balanced chromosomal structural rearrangements are known to cause male infertility and account for approximately 1% of azoospermia or severe oligospermia. However, the underlying mechanisms of pathogenesis and etiologies are still largely unknown. Herein, we investigated apparently balanced interchromosomal structural rearrangements in six cases with azoospermia/severe oligospermia to comprehensively identify and delineate cryptic structural rearrangements and the related copy number variants. In addition, high read-depth genome sequencing (GS) (30-fold) was performed to investigate point mutations causative of male infertility. Mate-pair GS (4-fold) revealed additional structural rearrangements and/or copy number changes in 5 of 6 cases and detected a total of 48 rearrangements. Overall, the breakpoints caused truncations of 30 RefSeq genes, five of which were associated with spermatogenesis. Furthermore, the breakpoints disrupted 43 topological-associated domains. Direct disruptions or potential dysregulations of genes, which play potential roles in male germ cell development, apoptosis, and spermatogenesis, were found in all cases (n = 6). In addition, high read-depth GS detected dual molecular findings in case MI6, involving a complex rearrangement and two point mutations in the gene *DNAH1*. Overall, our study provided the molecular characteristics of apparently balanced interchromosomal structural rearrangements in patients with male infertility. We demonstrated the complexity of chromosomal structural rearrangements, potential gene disruptions/dysregulation and single-gene mutations could be the contributing mechanisms underlie male infertility.

Asian Journal of Andrology (2022) **24**, 248–254; doi: 10.4103/aja2021106; published online: 07 January 2022

Keywords: azoospermia; balanced structural rearrangements; genome sequencing; male infertility; severe oligospermia

INTRODUCTION

Infertility is defined as the failure to achieve pregnancy after 12 months or more of regular unprotected intercourse.[1] Approximately 15% of couples are affected by infertility, and among them, male factor contributes to 50%.[2] Male factor infertility is often defined by abnormal semen parameters, which could be associated with other medical conditions, developmental factors, lifestyle factors, and genetics. However, the exact nature of these associations is still largely unclear.[3] The most common etiology of male infertility is primary testicular failure, resulting in quantitative impairment of spermatogenesis, accounting for 75% of male factor infertility.[4] However, the etiology of primary testicular failure is unknown in about 40% of cases, in which genetic factors are thought to contribute to a significant proportion of these cases.[5]

Genetic factors account for at least 15% of male infertility, and they have been known to contribute to all four major etiological categories, including (1) spermatogenic quantitative defects, (2) ductal obstructions or dysfunction, (3) hypothalamic–pituitary axis disturbances, and (4) spermatogenic qualitative defects.[5] The genetic landscape of male infertility is highly complex and heterogeneous, with over 2000 genes involved in spermatogenesis and testicular function.[6] Diagnosing the genetic causes of male infertility has clinical implications for prognosis of testicular sperm retrieval and personalizing therapy,[5] improving the reproductive health and general health of the couple.

Men with azoospermia have the highest risk of being affected by genetic factors (25%).[5] There are two well-known genetic causes of nonobstructive azoospermia. The most common causative genetic defects include chromosome Yq11 microdeletions of the azoospermia factor region (AZF), which can be detected in 13% of men with azoospermia.[7] In patients with complete AZFb or AZFb + AZFc deletions, testicular biopsy usually reveals Sertoli cells only or diffuse early maturation arrest; hence, these patients must use donor sperm or adoption. In contrast, men with partial AZFa deletions

[1]Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China; [2]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen 518057, China; [3]Hong Kong Hub of Paediatric Excellence, The Chinese University of Hong Kong, Hong Kong, China; [4]Genetics and Prenatal Diagnosis Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China; [5]The Chinese University of Hong Kong-Baylor College of Medicine Joint Center for Medical Genetics, Hong Kong, China.
*These authors contributed equally to this work.
Correspondence: Dr. Z Dong (elvisdong@cuhk.edu.hk)
Received: 28 July 2021; Accepted: 27 October 2021

often have hypospermatogenesis, with much better surgical sperm retrieval rates.[8]

Chromosomal abnormalities including sex chromosomal abnormalities (SCA) also cause male infertility. Among them, balanced chromosomal abnormalities (BCA; such as translocation and insertion), defined as exchange of chromosomal segments between two or more chromosomes without apparent gains or losses detectable by karyotyping, account for 0.5%–1.0% of patients with severe oligospermia or azoospermia. Male infertility caused by BCA may be explained by the failure of pairing of homologous elements on derivative chromosomes during meiosis I. There is also evidence that failure of pairing of homologous elements promotes association between the quadrivalent and the X-Y bivalent, which affects sperm count.[9] However, conception can still be achieved in a significant proportion of male BCA carriers. In addition, semen analysis showed no significant differences in BCA carriers and males with normal karyotype.[10] In contrast, GS-based investigations of BCAs have suggested that gene disruption or dysregulation caused by structural rearrangements and cryptic complexities could also contribute to male infertility.[11,12] In complex rearrangements, the number of chromosomes and breakpoints involved, the location of breakpoints, and their relative sizes are presumed to affect fertility.[13]

Although GS has been applied to investigate the molecular breakpoints of chromosomal structural rearrangements in patients with abnormal phenotypes,[14–17] standard GS (with small-insert DNA) is not well suited for identifying breakpoint junctions, particularly those that are mediated by repetitive elements.[12] To overcome this limitation, we developed an in-house mate-pair genome sequencing method which utilizes large-insert size DNA (3–8 kb) libraries, sequenced at a low read-depth.[18] Our previous studies demonstrated that this method increased the sensitivity in the detection of BCAs and demonstrated the detection of additional cryptic and complex rearrangements in karyotypically simple chromosomal abnormalities.[17] In this study, we aim to investigate BCA breakpoint junctions and the related copy number deletions/duplications by mate-pair GS and identify other genomic variants that potentially contribute to male infertility.

PATIENTS AND METHODS

Ethics approval and case recruitment

This study was approved by The Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (The Joint CUHK-NTEC CREC; approval No. 2020.046). All participants in this study have provided written informed consent. We studied six idiopathic nonobstructive azoospermic or severe oligospermic males with BCAs recruited from the Prince of Wales Hospital, 30–32 Ngan Shing Street, Shatin, New Territories, Hong Kong, China. BCAs were previously ascertained in these subjects by karyotyping (**Table 1**), and chromosome Y microdeletions have been excluded. One patient carried a simple two-way reciprocal translocation, one carried double two-way translocations, two had three-way translocations, one case had a two-way chromosomal insertion (described in our previous study[18]), and one case had both a complex insertion and a translocation.

Mate-pair genome sequencing

Genomic DNA (gDNA) was extracted with the DNeasy Blood & Tissue Kit (category No. 69506; Qiagen, Hilden, Germany) from each sample. gDNA was then quantified with a Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). After quality control (QC), a 1-μg aliquot of gDNA was sheared to 3–8 kb for mate-pair library construction.[18] The libraries were pooled and sequenced to paired-end 100 bp on an MGISEQ-2000 platform (MGI Technology Co., Ltd., Shenzhen, China). A minimal of 60 million read-pairs were generated for each case, equating to an average 4 × read-depth and 60 × to 100 × physical coverage. In addition, for each sample except MI5, 400 ng of gDNA was also subjected for standard high read-depth GS. Each library was sequenced in each lane (paired-end 150 bp) on an MGISEQ-2000 platform (MGI Technology Co., Ltd.) for a minimal of 30-fold read-depth.

Variant identification and interpretation

Genomic variants were detected using our in-house bioinformatics pipelines. (1) Structural rearrangements were detected at a resolution of 10 kb. In brief, chimeric read-pairs that aligned to different chromosomes or mapped to the same chromosome with a genomic distance of >10 kb[19,20] were clustered and filtered against a control dataset of >2500 GS data (to remove systematic errors due to short-read alignment or variants with high frequencies as polymorphisms). (2) Copy number variants (CNVs) were detected using our reported increment-ratio-of-coverage method at a resolution of 50 kb.[21–23] Rare CNVs with allele frequencies of <1% in our in-house Chinese subjects (n > 2000) were selected for review. (3) Absence of heterozygosity (AOH) was detected at a resolution of 5 Mb as previously described[24] to investigate any uniparental disomy and parental consanguinity.

The high read-depth GS data analysis involved the following steps: fastq data QC and alignment (BWA-MEM), detection of single-nucleotide variants (SNVs), and small insertions/deletions (InDels) by HaplotypeCaller version 3.4 from the Genome Analysis Toolkit (GATK; Broad Institute, MA, USA).[25] The variants were subsequently annotated using ANNOVAR[26] and InterVAR[27] with in-house and public databases.

The following criteria were used for prioritization of SNVs/InDels: (1) variants reported by ClinVAR or human gene mutation database (HGMD); (2) variants with a minor allele frequency ≤5% in the ExAC (http://exac.broadinstitute.org) and gnomAD (https://gnomad.broadinstitute.org) databases; (3) variants that are located in coding regions or exon–intron junctions; (4) variants with damaging/intolerant or splicing-change effects predicted by multiple computational algorithms (Revel, SIFT, Polyphen-2, MutationTaster, Human Splicing Finder, and MaxEntScan); and (5) variants that contain or are located in OMIM disease-causing genes. For known mutations, clinical correlation will be performed (for male infertility). Further classification of novel mutations will be performed for variants (1) located in autosomal dominant or X-linked dominant genes or (2) being homozygous or compound heterozygous in autosomal recessive or X-linked recessive genes. Potential disease-causing mutations will be selected for validation and parental confirmation when available.

Genomic variants (CNVs, AOHs, SNVs, and InDels) involving genes associated with male infertility, oligospermia or azoospermia, spermatogenesis deficiency, and ciliopathy are interpreted based on the guidelines of the American College of Medical Genetics and Genomics (ACMG) using the 5-tier classification.[28]

Variant verification and junction annotation

Rearrangement breakpoints were resolved to the single-nucleotide level by junction-specific PCR followed by Sanger sequencing, which enabled accurate junction annotation for interpretation.[29] Sanger sequencing results (fa format) were mapped to the reference genome (GRCh37) using UCSC BLAT. To analyze and assess the functional and phenotypic

Table 1: Case summary of the azoospermic/oligospermic cohort with apparently balanced structural rearrangements identified

Case ID	*Clinical indication*	*FSH (mIU ml^{-1})*	*LH (mIU ml^{-1})*	*T (ng ml^{-1})*	*Age (year)*	*Karyotype*	*Rearrangements identified by mate-pair GS (n)*	*Number of cryptic copy number variants involved in the rearrangements (n)*
MI1	Severe oligospermia	4.9	5.8	2.2	39	46,XY,t(3;19)(p21.3;q13.3)	2	0
MI2	Primary infertility, and azoospermia	NA	NA	NA	25	46,XY,t(5;9)(p13.3;p22),t(7;21)(p13;q22.1)	4	1
MI3	Severe oligospermia	5.7	5.5	5.9	24	46,XY,t(4;11;6)(q22;q21;q16)	7	0
MI4	Azoospermia	5.0	7.0	4.4	26	46,XY,t(8;12;10)(q24.1;p13;q22)	7	0
MI5	Severe oligospermia, asthenozoospermia, and teratospermia	4.4	5.5	5.9	28	46,XY,ins(6;2)(q23;p13p22)	9	0
MI6	Azoospermia	NA	NA	NA	NA	46,XY,t(4;20)(q28;q12),der(20)ins(20;3)(q12;q28q13.3)inv(3)?(q13.3q25.3)	19	4

FSH: follicle-stimulating hormone; LH: luteinizing hormone; T: testosterone; NA: not available; GS: genome sequencing

association of structural rearrangements, the breakpoint junctions were annotated for (1) direct disruption of genes, (2) disruption of regulatory elements, (3) disruption of topologically associating domain (TAD) defined for hESC cell line[30] (3D Genome browser: http://3dgenome.fsm.northwestern.edu/index.html), and (4) genes or regulatory elements within the same TAD as the breakpoint. Furthermore, the breakpoint junction sequence features were annotated to investigate the mechanisms involved in the rearrangements.

Orthogonal validation of CNVs and AOHs was performed using the 8 × 60K Fetal DNA Chip version 2.0 (Agilent Technologies, Santa Clara, CA, USA) microarray.[31,32] Sanger sequencing was performed to validate SNVs and InDels.

RESULTS

We performed mate-pair GS on six male cases with azoospermia/severe oligospermia (**Table 1**) and apparently balanced chromosomal structural rearrangements previously ascertained by G-banded chromosome analysis, who were negative for chromosome Y microdeletions. Karyotyping revealed reciprocal-balanced translocation in one case (sample ID: MI1), two independent balanced translocations in one case (MI2), three-way balanced translocations in two cases (MI3 and MI4), a two-way chromosomal insertion in one case (MI5), and a complex insertion and translocation in one case (MI6; **Table 1**). Overall, mate-pair GS detected 48 structural rearrangements. Cryptic deletions were identified in two cases (**Supplementary Table 1**). Thirty-six rearrangements were precisely mapped to single-nucleotide resolution (**Supplementary Table 2**) by Sanger sequencing, while the remaining 12 rearrangements were verified by gap-PCR but could not be resolved to nucleotide level due to the presence of repetitive elements.

In the case with the simplest rearrangement, two breakpoints were detected in a reciprocal translocation in case MI1. In the most complex case (MI6), 19 rearrangements were detected with deletions of four segments. Of the 48 rearrangements, 39 disrupted genes. Overall, a total of 30 unique RefSeq genes (**Supplementary Table 3**) and 43 unique TADs were disrupted (**Supplementary Table 4**). Potential gene disruption, dysregulation, and/or point mutations were investigated in each case to reveal the underlying mechanisms of azoospermia/severe oligospermia.

Simple-BCA: case MI1

The 39-year-old patient suffered from severe oligospermia with a karyotype of 46,XY,t(3;19)(p21.3;q13.3). Mate-pair GS revealed a simple reciprocal balanced translocation without copy number changes at the breakpoint junctions. Mate-pair GS revised the affected bands on both chromosomes to seq[GRCh37] t(3;19)(p14.3q13.2). Validated by PCR and Sanger sequencing, the breakpoint junctions were found to directly truncate the gene *ERC2* at intron 12. The expression of *ERC2* has been reported to be heat-sensitive in rat spermatocytes and round spermatid.[33] Expression of genes that are heat-sensitive may be crucial because spermatogenesis occurs at approximately 3°C lower than body temperature.[33] Mate-pair GS revealed no additional clinically significant CNVs and AOH. In addition, high read-depth GS not only confirmed the findings from mate-pair GS but also indicated that no SNVs/InDels associated with azoospermia/oligospermia were detected. As *ERC2* is likely haploinsufficient (gnomAD: pLoF=0.99), its disruption could be associated with the phenotype in the patient.

Two independent sets of BCA: case MI2

This patient has primary infertility and azoospermia. The patient has two independent balanced translocations 46,XY,t(5;9)(p13.3;p22),t(7;21)(p13;q22.1) reported by G-banded chromosome analysis. Mate-pair GS revealed two balanced translocations t(7;21)(p13;q22.11) and t(5;9)(p14.1;p23) with a 15.4-kb deletion seq[GRCh37] del(9)(p23) chr9:g.10081969_10097410del at one of the breakpoint junctions (intron 3 of gene *PTPRD*). Sanger sequencing enabled fine mapping of three rearrangements (**Figure 1**). Two breakpoints had 12-bp and 15-bp sequence insertions, and one had a 3-bp microhomology (**Supplementary Table 5**). These features suggested that the structural rearrangements were formed by DNA replication-based mechanisms.

The genes *COA1* and *PTPRD* were disrupted. Among them, *PTPRD* has been reported as a candidate gene for nonobstructive azoospermia.[34,35] *PTPRD* is also predicted to be haploinsufficient (pLoF=1; gnomAD). Although four TADs were disrupted involving a total of 41 genes, none of the genes located within the disrupted TADs was associated with azoospermia/severe oligospermia. Mate-pair GS and high read-depth GS revealed no additional SNVs/InDels, CNVs, and AOH associated with male infertility in this case. Thus, the disruption of *PTPRD* may potentially explain the infertility.

Three-way BCA: case MI3

The patient has severe oligospermia and karyotype results indicated a three-way balanced translocation 46,XY,t(4;11;6)(q22;q21;q16). However, mate-pair GS detected additional cryptic complexities underlying the translocations which resulted in an overall of seven rearrangements without CNVs detected at the breakpoint junctions (**Figure 2**). Among six rearrangements fine-mapped by Sanger sequencing, the features of the breakpoint junctions included blunt ends in four rearrangements and 1 bp microhomology in the other

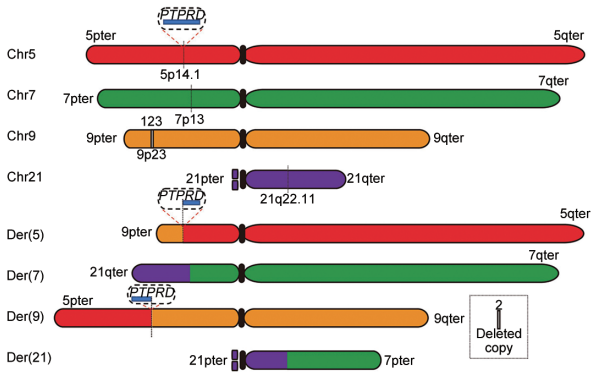

Figure 1: Two reciprocal translocations of case MI2 with a karyotype of 46,XY,t(5;9)(p13.3;p22),t(7;21)(p13;q22.1). The upper panel shows the normal chromosomes 5, 7, 9, and 21, and the lower panel shows the respective derivative chromosomes. Bars in red, green, orange, and purple indicate chromosomal segments on chromosomes 5, 7, 9, and 21, respectively. Each dotted line indicates a breakpoint junction on each chromosome. A small white bar indicates a copy number loss of 15.4 kb involving the intron 3 of *PTPRD* in chromosome 9. Disruption of candidate genes associated with male infertility is shown above the chromosomes. *PTPRD*: protein tyrosine phosphatase receptor type D; chr: chromosome; der: derivative chromosome.

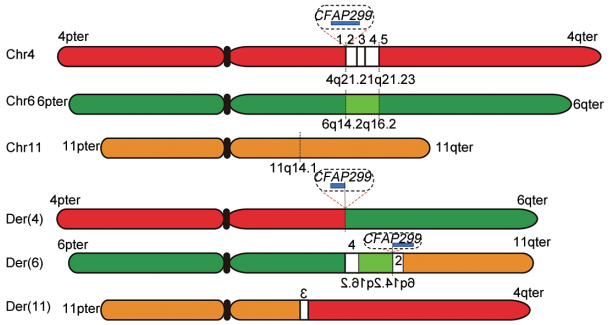

Figure 2: Three-way translocation of case MI3 with a karyotype of 46,XY,t(4;11;6)(q22;q21;q16). The upper panel shows the normal chromosomes, and the lower panel shows the derivative chromosomes. Seven rearrangements were detected by mate-pair genome sequencing, revealing more complex rearrangements than a three-way translocation. Bars in red, green, and orange indicate chromosomal segments on chromosomes 4, 6, and 11, respectively. Segments originating from chromosome 4 were found on both der(6) and der(11). Each dotted line indicates the breakpoint on each chromosome. Bars in white represent cryptic chromosomal segments on 4q21.21q21.23, while boxes in light green indicate a chromosomal segment of 6q14.2q16.2. Segments that have an inverse genomic orientation are shown by reversed numbers. Disruption of candidate genes associated with male infertility is shown above the chromosomes. *CFAP299*: cilia and flagella associated protein 299; chr: chromosome; der: derivative chromosome.

two rearrangements, suggesting end-joining as the rearrangement formation mechanism.

Genes *CEP162* and *CFAP299* were disrupted by the breakpoint junctions. Among them, the role of *CFAP299* in spermatogenesis has been implicated in animal studies. Messenger RNA of *CFAP299* is predominantly expressed in mouse testis at increased levels from 2–8-week-old testes in the cytoplasm of spermatogonia and primary spermatocytes.[36] Silencing of *CFAP299* resulted in an increase in the number of apoptotic cells and arrested cells at the G2/M phase.[36] As such, *CFAP299* plays a potential role in spermatogenesis in regulating cell apoptosis.[36] In addition, five TADs were disrupted, involving 16 genes. However, none of the genes were known to be associated with spermatogenesis. Mate-pair GS and high read-depth GS confirmed that there were no additional clinically significant SNVs/InDels, CNVs, and AOH associated with azoospermia/oligospermia detected in this case. Therefore, disruption of *CFAP299* could be a candidate cause of the infertility.

Three-way BCA: case MI4

The azoospermic patient carried a three-way reciprocal translocation: 46,XY,t(8;12;10)(q24.1;p13;q22), as shown in **Figure 3**. Mate-pair GS revealed additional cryptic complexities, with a total of seven rearrangements. There were no CNVs involved at the breakpoint junctions. All breakpoints were precisely mapped with Sanger sequencing. The sequence features showed blunt-ends in four rearrangements and 1 bp microhomology in the other three rearrangements. The formation of these rearrangements was likely mediated by end-joining mechanism.

Three genes including *EPS8*, *SOX5*, and *CTNNA3* were disrupted. *EPS8* is known to be sensitive to haploinsufficiency (gnomAD: pLoF=0.97) and has been implicated to play a role in the induction of spermatid release from seminiferous epithelium during spermiation.[37] Loss of *EPS8* at the apical Sertoli cell–spermatid interface induced premature spermatid release and infertility.[37] In addition, seven TADs were disrupted by the breakpoint junctions, which contained a total of 49 genes. Among them, *PDE3A* is a candidate gene associated with hypogonadism and Kallmann syndrome.[38] Translocation breakpoints near the *PDE3A* gene have been reported in translocation carriers with Kallmann syndrome, potentially due to positional effects.[39] In addition, female mice with *PDE3A* homozygous deletions were infertile, and ovulated oocytes were arrested in the germinal stage and could not be fertilized.[40] Mate-pair GS and high read-depth GS found no additional SNVs/InDels, CNVs, and AOH associated with male infertility in this case. The disruption of *EPS8* and the potential dysregulation of *PDE3A* are the possible explanations for infertility.

Two-way insertion: case MI5

This patient suffered from severe oligoasthenospermia and has been described in our previous study.[18] An apparently balanced chromosomal insertion was identified by karyotyping analysis: 46,XY,ins(6;2)(q23;p13p22). Mate-pair GS detected nine rearrangements and indicated that the acceptor chromosome was likely involved in a chromothripsis event. Among them, six rearrangements were fine mapped by Sanger sequencing, including 1–2 bp microhomology (n = 3), small insertion (1 bp and 9 bp; n = 2), and blunt end (n = 1).

Seven genes were directly truncated by the breakpoints; however, none of which have been implicated in azoospermia/severe oligospermia. Eight TADs involving a total of 65 genes were disrupted. Although no direct evidence has shown associations of *SLC17A5* (directly disrupted by the breakpoint junction) with male infertility, the gene *EEF1A1* whose super enhancer was highly correlated to *SLC17A5* was haploinsufficient (gnomAD: pLoF=0.98) and essential for spermatogenesis.[41] Therefore, male infertility might be explained by the dysregulation of *EEF1A1* expression through disruption of *SLC17A5*, leading to spermatogenesis failure.[18] In addition, mutations in *MCM9* are known to result in ovarian dysgenesis. *MCM9*-knockout male mice were sterile, resulted by arrest of spermatocytes in meiotic prophase I, while *MCM9*-knockout female mice were also sterile as their ovaries only contained arrested primary follicles and frequently developed tumors.[42] There was insufficient DNA to pursue high read-depth GS in this case. Dysregulation of gene *EEF1A1* and/or *MCM9* by the structural rearrangement might explain the infertility in this case.

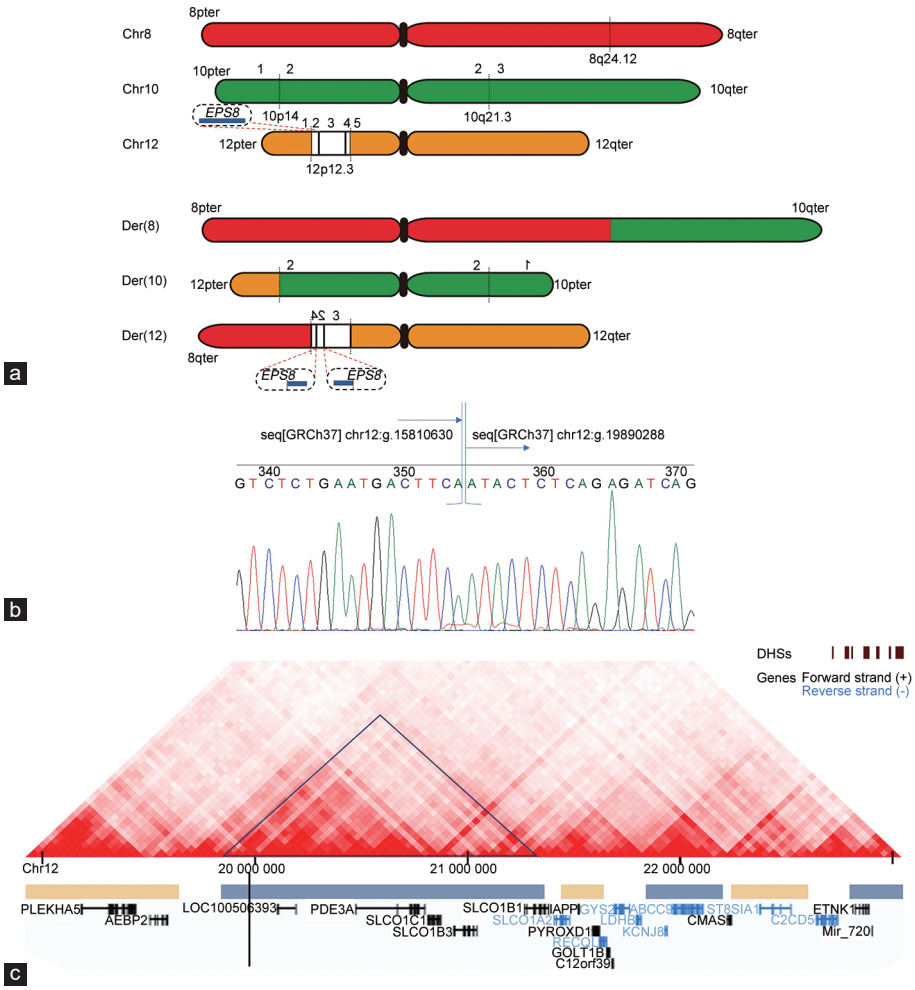

Figure 3: Three-way translocation of case MI4 with a karyotype of 46,XY,t(8;12;10)(q24.1;p13;q22). (**a**) The upper panel shows the normal chromosomes, and the lower panel shows the derivative chromosomes. Seven breakpoint junctions were detected by mate-pair genome sequencing. Bars in red, green and orange indicate chromosomal segments in chromosomes 8, 10 and 12, respectively. Each dotted line indicates the breakpoint locus on each chromosome. Bars in white represent chromosomal segments on 12p12.3. Characters in a reverse direction indicate a reverse orientation of the chromosomal segments in derivative chromosomes. Disruption of candidate genes associated with male infertility is shown above the chromosomes. (**b**) Sequence resolved breakpoint junction of a cryptic rearrangement der(12) reveals blunt end as the sequence characteristics. (**c**) The 3D genome map (http://3dgenome.fsm.northwestern.edu/) visualizing the 3D interaction and topologically associating domains. The black triangle shows a topologically associating domain disrupted by a breakpoint (shown by the vertical black line). Genes encompassed under each topologically associating domain are listed in the lower panel. *EPS8*: epidermal growth factor receptor pathway substrate 8; DHSs: DNase I hypersensitive sites; chr: chromosome; der: derivative chromosome.

Complex insertion and translocation: case MI6

This patient had a complex karyotype: 46,XY,t(4;20)(q28;q12),der(20)ins(20;3)(q12;q28q13.3)inv(3)?(q13.3q25.3) involving three chromosomes. Mate-pair GS detected multiple complex rearrangements with 19 rearrangements (**Figure 4**). Four copy number losses were found on chromosome 3, and the largest was a 1.15-Mb deletion involving the gene *CFAP91*. Mutations in *CFAP91* are known to cause spermatogenic failure 51 (OMIM#619177) in an autosomal recessive manner (OMIM*609910). However, no additional variants were detected in *CFAP91*. Among the 19 rearrangements, 12 were successfully fine mapped by Sanger sequencing. The results revealed blunt ends at five junctions, 1–2 bp microhomology at six junctions, and 1 bp insertion at one junction. The breakpoint junction features along with frequent rearrangements suggested that this complex rearrangement was likely formed by DNA replicative mechanisms (**Supplementary Table 5**).

Overall, 14 genes were directly truncated by the rearrangements including *TP63*. *TP63* is sensitive to haploinsufficiency (gnomAD: pLoF=1) and plays a role in mediating male germ cell development, apoptosis, and regulation of spermatogenesis.[43,44] In addition, 15 TADs, involving a total of 128 genes, were disrupted. Among them, three genes were implicated in male infertility, including *GMNC*, *SOX2*, and *CCDC39*. *GMNC* (also known as *Gemc1* in mice) plays a role in later stages of spermatogenesis.[45] *Gemc1*$^{-/-}$ mice had a significantly reduced number of round spermatids and elongating spermatids.[45] *SOX2* is also a candidate gene for hypogonadotropic hypogonadism and resultant testicular failure.[46] At last, mutations in *CCDC39* are known to cause primary ciliary dyskinesia and Kartagener syndrome in an autosomal recessive manner, of which the common features include situs inversus, infertility, and oligoasthenospermia.[47]

There were no additional SNVs/InDels in *CFAP91*, which was involved in a heterozygous deletion as described above. However, high read-depth GS detected two heterozygous SNVs in the gene *DNAH1*. Mutations in *DNAH1* cause spermatogenic failure through multiple morphological abnormalities of the sperm flagella and primary

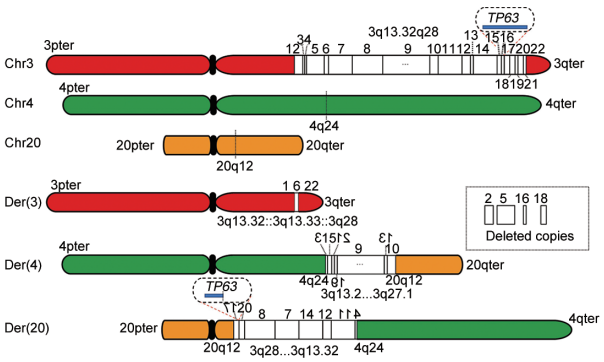

Figure 4: Complex insertion and translocation of case MI6 detected by mate-pair genome sequencing. Bars in red, green, and orange indicate chromosomal segments in chromosomes 3, 4, and 20, respectively. Each dotted line indicates the breakpoint locus on each chromosome. Bars in white present cryptic chromosomal segment in 3q13.32q28 indicating there are 22 segments after chromosome scatting. Characters in a reverse direction indicate a reverse orientation of the chromosomal segments in derivative chromosomes. Disruption of candidate genes associated with male infertility is shown above the chromosomes. Copy number losses of four chromosome segments originating on chromosome 3 are indicated in a dotted frame. *TP63*: tumor protein p63; chr: chromosome; der: derivative chromosome.

ciliary dyskinesia in an autosomal recessive manner.[48,49] However, the inheritance pattern of these two SNVs (in *cis*- or *trans*-) could not be determined as parental samples were not available. Nonetheless, male infertility in this case may be contributed by dual diagnoses of a complex rearrangement and point mutations.

DISCUSSION

In this study, we investigated the genetic etiology of six cases with male infertility and apparently balanced interchromosomal structural rearrangements by mate-pair GS in combination with high read-depth GS. Overall, our study revealed that disruption or potential dysregulation of candidate genes implicated in male germ cell development, apoptosis, and spermatogenesis could be the underlying mechanisms of male infertility in these cases.

Balanced chromosomal structural rearrangements are thought to cause male infertility through failure of pairing of homologous elements.[9] Recently, studies have shown male infertility in patients with BCAs may be caused by gene disruption or dysregulation by breakpoints or other genomic variants, for example, SNVs and InDels.[11,12] While standard high read-depth GS cannot detect chromosomal rearrangements in low complexity or highly repetitive regions of the genome,[50,51] our large-insert size DNA mate-pair GS approach had increased sensitivity in detecting structural rearrangements.[18] We identified the composition of each derivative chromosome in each case and revealed additional rearrangements cryptic to conventional G-banded chromosome analysis. We showed direct disruption or potential dysregulation of candidate genes associated with male infertility. In addition, high read-depth GS not only reported a dual molecular findings in case MI6, including a complex rearrangement and two point mutations in an autosomal recessive gene *DNAH1*, but also excluded known causative variants and noncoding regions (such as deep intronic region).

Our study on interchromosomal structural rearrangements revealed the complexities of BCAs that are largely underappreciated by G-banded chromosome analysis. Except MI1, additional cryptic rearrangements and/or copy number changes were detected in the remaining cases. The simplest was a cryptic deletion (15.4 kb) in case MI2 with two independent reciprocal translocations. In comparison, the rearrangements were most complex in MI6, which had 19 rearrangements and four deletions. The proportion of cases (5/6, 83.3%) with additional cryptic complex rearrangement identified by mate-pair GS over karyotyping was significantly higher than that reported in our previous study of males carrying BCAs in recurrent miscarriage couples (5/33, 15.2%, Fisher's exact test: $P = 0.0023$), who were able to conceive.[16] It indicated that the complexities of structural rearrangements may be correlated with the phenotypic presentation (azoospermia or severe oligospermia).

Interestingly, among these cases, direct gene disruptions and/or potential dysregulation due to TAD disruptions are the potential underlying disease-causing mechanisms. In case MI2, the cryptic deletion (15.4 kb) was located in the intron of gene *PTPRD*. In case MI6, there were four copy number losses including gene *CFAP91*; a gene known to cause spermatogenic failure 51 (OMIM#619177) in an autosomal recessive manner (OMIM*609910). We excluded the possibility of compound heterozygous disease-causing alleles in autosomal recessive genes by identifying any causative hemizygous point mutations using high read-depth GS.

The limitations of our study include (1) limited sample size as there were six cases enrolled in this study, partly due to the rarity of apparently balanced chromosomal structural rearrangements and negative AZF region deletions in patients with severe oligospermia or azoospermia (0.5%–1.0%); (2) no parental samples available to study the inheritance mode of the detected rearrangements variants; for the two point mutations identified in *DNAH1*, long-read sequencing may resolve the haplotype spanning the two variants that are approximately 33 kb apart; however, no additional sample was available for this experiment; and (3) no additional samples were available for transcriptome analysis to confirm the potential consequences of the genomic variants. Nonetheless, our study provided a comprehensive composition of genomic structural rearrangements in each case and showed that the disruption or potential dysregulation of candidate genes could be the underlying causes of male infertility.

Overall, our study provided molecular characteristics of apparently balanced interchromosomal structural rearrangements in patients with male infertility. We showed the complexity of chromosomal structural rearrangements involved in male infertility and revealed that disruption or potential dysregulation of genes potentially implicated in male gametes development, apoptosis, and spermatogenesis could be a cause for infertility.

AUTHOR CONTRIBUTIONS

MHKC, YL, KWC, XK, and ZD designed the study. JPWC collected the samples and followed up the clinical outcomes. MHKC, YL, MS, YKK, and ZD performed the analysis and data interpretation. MHKC, YL, PD and XZ conducted the validation. MHKC, YL, KWC, and ZD wrote the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

All authors declare no competing interests.

ACKNOWLEDGMENTS

This project is supported by the National Natural Science Foundation of China (No. 31801042), the Health and Medical Research Fund (No. 04152666 and No. 07180576), General Research Fund (No. 14115418), and Direct Grant (No. 2020.052).

Supplementary Information is linked to the online version of the paper on the *Asian Journal of Andrology* website.

REFERENCES

1 Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, *et al*. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril* 2009; 92: 1520–4.
2 Pfeifer S, Butts S, Dumesic D, Fossum G, Gracia C, *et al*. Diagnostic evaluation of the infertile male: a committee opinion. *Fertil Steril* 2015; 103: e18–25.
3 Choy JT, Eisenberg ML. Male infertility as a window to health. *Fertil Steri* 2018; 110: 810–4.
4 Krausz C, Consultant S. Endocrinology & metabolism male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 2011; 25: 271–85.
5 Krausz C, Riera-Escamilla A. Genetics of male infertility. *Nat Rev Urol* 2018; 15: 369–84.
6 Kasak L, Laan M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. *Hum Genet* 2021; 140: 135–54.
7 Hamada A, Esteves S, Agarwal A. A comprehensive review of genetics and genetic testing in azoospermia. *Clinics* 2013; 68: 39–60.
8 Neto FT, Bach PV, Najari BB, Li PS, Goldstein M. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol* 2016; 59: 10–26.
9 Paoloni-Giacobino A, Kern I, Rumpler Y, Djlelati R, Morris M, *et al*. Familial t(6;21)(p21.1;p13) translocation associated with male-only sterility. *Clin Genet* 2000; 58: 324–8.
10 Pastuszek E, Kiewisz J, Kulwikowska PM, Lukaszuk M, Lukaszuk K. Sperm parameters and DNA fragmentation of balanced chromosomal rearrangements carriers. *Folia Histochem Cytobiol* 2016; 53: 314–21.
11 Schilit SL, Menon S, Friedrich C, Kammin T, Wilch E, *et al*. SYCP2 translocation-mediated dysregulation and frameshift variants cause human male infertility. *Am J Hum Genet* 2020; 106: 41–57.
12 Yammine T, Reynaud N, Lejeune H, Diguet F, Rollat-Farnier P, *et al*. Deciphering balanced translocations in infertile males by next generation sequencing to identify candidate genes for spermatogenesis disorders. *Mol Hum Reprod* 2021; 27: gaab034.
13 Nguyen MH, Morel F, Pennamen P, Parent P, Douet-Guilbert N, *et al*. Balanced complex chromosome rearrangement in male infertility: case report and literature review. *Andrologia* 2015; 47: 178–85.
14 Redin C, Brand H, Collins RL, Kammin T, Mitchell E, *et al*. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* 2017; 49: 36–45.
15 David D, Freixo JP, Fino J, Carvalho I, Marques M, *et al*. Comprehensive clinically oriented workflow for nucleotide level resolution and interpretation in prenatal diagnosis of *de novo* apparently balanced chromosomal translocations in their genomic landscape. *Hum Genet* 2020; 139: 531–43.
16 Dong Z, Yan J, Xu F, Yuan J, Jiang H, *et al*. Genome sequencing explores complexity of chromosomal abnormalities in recurrent miscarriage. *Am J Hum Genet* 2019; 105: 1102–11.
17 Dong Z, Chau MH, Zhang Y, Dai P, Zhu X, *et al*. Deciphering the complexity of simple chromosomal insertions by genome sequencing. *Hum Genet* 2021; 140: 361–80.
18 Dong Z, Zhao X, Li Q, Yang Z, Xi Y, *et al*. Development of coupling controlled polymerizations by adapter-ligation in mate-pair sequencing for detection of various genomic variants in one single assay. *DNA Res* 2019; 26: 313–25.
19 Dong Z, Wang H, Chen H, Jiang H, Yuan J, *et al*. Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. *Genet Med* 2018; 20: 697–707.
20 Dong Z, Jiang L, Yang C, Hu H, Wang X, *et al*. A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing. *Hum Mutat* 2014; 35: 625–36.
21 Dong Z, Zhang J, Hu P, Chen H, Xu J, *et al*. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. *Genet Med* 2016; 18: 940–8.
22 Wang H, Dong Z, Zhang R, Chau MH, Yang Z, *et al*. Low-pass genome sequencing versus chromosomal microarray analysis: implementation in prenatal diagnosis. *Genet Med* 2020; 22: 500–10.
23 Chau MH, Wang H, Lai Y, Zhang Y, Xu F, *et al*. Low-pass genome sequencing: a validated method in clinical cytogenetics. *Hum Genet* 2020; 139: 1403–15.
24 Dong Z, Chau MH, Zhang Y, Yang Z, Shi M, *et al*. Low-pass genome sequencing-based detection of absence of heterozygosity: validation in clinical cytogenetics. *Genet Med* 2021; 23: 1225–33.
25 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297–303.
26 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38: e164.
27 Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* 2017; 100: 267–80.
28 Richards S, Aziz N, Bale S, Bick D, Das S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17: 405–24.
29 Nalbandian K, Piña-Aguilar RE, Morton CC. Resolving breakpoints of chromosomal rearrangements at the nucleotide level using Sanger sequencing. *Curr Protoc Hum Genet* 2020; 108: e107.
30 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, *et al*. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; 485: 376–80.
31 Leung TY, Vogel I, Lau TK, Chong W, Hyett JA, *et al*. Identification of submicroscopic chromosomal aberrations in fetuses with increased nuchal translucency and apparently normal karyotype. *Ultrasound Obstet Gynecol* 2011; 38: 314–9.
32 Huang J, Poon LC, Akolekar R, Choy KW, Leung TY, *et al*. Is high fetal nuchal translucency associated with submicroscopic chromosomal abnormalities on array CGH? *Ultrasound Obstet Gynecol* 2014; 43: 620–4.
33 Yadav SK, Pandey A, Kumar L, Devi A, Kushwaha B, *et al*. The thermo-sensitive gene expression signatures of spermatogenesis. *Reprod Biol Endocrinol* 2018; 16: 56.
34 Yu J, Wu H, Wen Y, Liu Y, Zhou T, *et al*. Identification of seven genes essential for male fertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in *Drosophila*. *Hum Mol Genet* 2015; 24: 1493–503.
35 Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, *et al*. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet* 2016; 48: 1462–72.
36 Li H, Dai Y, Luo Z, Nie D. Cloning of a new testis-enriched gene C4orf22 and its role in cell cycle and apoptosis in mouse spermatogenic cells. *Mol Biol Rep* 2019; 46: 2029–38.
37 Cheng CY, Mruk DD. Regulation of spermiogenesis, spermiation and blood-testis barrier dynamics: novel insights from studies on Eps8 and Arp3. *Biochem J* 2011; 435: 553–62.
38 Quaynor SD, Bosley ME, Duckworth CG, Porter KR, Kim S, *et al*. Molecular and cellular endocrinology targeted next generation sequencing approach identifies eighteen new candidate genes in normosmic hypogonadotropic hypogonadism and Kallmann syndrome. *Mol Cell Endocrinol* 2016; 437: 86–96.
39 Bhagavath B, Podolsky RH, Ozata M, Bolu E, Bick DP, *et al*. Clinical and molecular characterization of a large sample of patients with hypogonadotropic hypogonadism. *Fertil Steril* 2006; 85: 706–13.
40 Masciarelli S, Horner K, Liu C, Park SH, Hinckley M, *et al*. Cyclic nucleotide phosphodiesterase 3A-deficient mice as a model of female infertility. *J Clin Invest* 2004; 114: 196–205.
41 Chen J, Jiang D, Tan D, Fan Z, Wei Y, *et al*. Heterozygous mutation of *eEF1A1b* resulted in spermatogenesis arrest and infertility in male tilapia, *Oreochromis niloticus*. *Sci Rep* 2017; 7: 43733.
42 Lutzmann M, Grey C, Traver S, Ganier O, Maya-Mendoza A, *et al*. MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Mol Cell* 2012; 47: 523–34.
43 Wang H, Yuan Q, Niu M, Zhang W, Wen L, *et al*. Transcriptional regulation of P63 on the apoptosis of male germ cells and three stages of spermatogenesis in mice. *Cell Death Dis* 2018; 9: 76.
44 Petre-Lazar B, Livera G, Moreno SG, Trautmann E, Duquenne C, *et al*. The role of p63 in germ cell apoptosis in the developing testis. *J Cell Physiol* 2007; 210: 87–98.
45 Terré B, Lewis M, Gil-Gómez G, Han Z, Hao L, *et al*. Defects in efferent duct multiciliogenesis underlie male infertility in GEMC1, MCIDAS or CCNO deficient mice. *Development* 2019; 146: dev162628.
46 Song SH, Chiba K, Ramasamy R, Lamb D. Recent advances in the genetics of testicular failure. *Asian J Androl* 2016; 18: 350–5.
47 Merveille AC, Davis EE, Becker-Heck A, Legendre M, Amirav I, *et al*. CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet* 2011; 43: 72–8.
48 Amiri-Yekta A, Coutton C, Kherraf ZE, Karaouzène T, Le Tanno P, *et al*. Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new *DNAH1* mutations. *Hum Reprod* 2016; 31: 2872–80.
49 Ben Khelifa M, Coutton C, Zouari R, Karaouzène T, Rendu J, *et al*. Mutations in *DNAH1*, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am J Hum Genet* 2014; 94: 95–104.
50 Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, *et al*. Structural variant calling: the long and the short of it. *Genome Biol* 2019; 20: 246.
51 Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, *et al*. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020; 21: 30.

Supplementary Table 1: RefSeq genes encompassed in the copy number variation regions

Case ID	*Chromosome*	*Start*	*End*	*Length*	*Band*	*Genes*
MI2	9	10,081,969	10,097,410	15,441	9p23	*PTPRD*
MI6	3	117,792,055	118,443,877	651,823	3q13.32	*EU250752*
MI6	3	118,759,395	119911093	1,151,699	3q13.32q13.33	*IGSF11, TEX55, UPK1B, B4GALT4, ALT4-AS1, ARHGAP31, AP31-AS1, TMEM39A, POGLUT1, TIMMDC1, CD80, ADPRH, PLA1A, OPDC2, POPDC2, COX17, FAP91, CFAP91, NR1I2, GSK3B, MIR6529, GPR156, PR156*
MI6	3	189,298,977	189,332,375	33,399	3q28	*TP63*
MI6	3	189,675,304	190,261,962	586,659	3q28	*P3H2, P3H2-AS1, LDN16, CLDN1, CLDN16, TMEM207, L1RAP, IL1RAP*

Supplementary Table 2: Precise breakpoints junction features delineated by Sanger sequencing or breakpoints detected by mate-pair genome sequencing

Case ID	*Breakpoint*	*Junction*	*Confirmation method*	*Feature*
MI1	1	Chr19:(+)(42340154):ttctctt: chr3:(–)(55995592)	Sanger	Insertion, 7bp
MI1	2	Chr19:(–)(42340158)::chr3:(+)(55995594)	Sanger	Blunt end
MI2	1	Chr21:(+)(34271591): attaataaatgtccg: chr7:(+)(43760345)	Sanger	Insertion, 15bp
MI2	2	Chr7:(+)(43760342):acatttattaat: chr21:(+)(34271591)	Sanger	Insertion, 12bp
MI2	3	Chr9:(+)(100820{48~50})::chr5:(+)(2767695{5~7})	Sanger	MH, 3bp
MI2	4	Chr5:(+)(27676953)::chr9(+)(10,097,410)	gap-pcr	NA
MI3	1	Chr4:(+)(8224610{9})::chr11:(–)(8240970{1})	Sanger	MH, 1bp
MI3	2	Chr4:(+)(82204576)::chr11:(+)(82409700)	Sanger	Blunt end
MI3	3	Chr4:(–)(8594093{2})::chr4:(+)(8220457{4})	Sanger	MH, 1bp
MI3	4	Chr6:(+)(93377684)::chr4:(–)(85940930)	Sanger	Blunt end
MI3	5	Chr4(+)(81610923)::chr6:(–)(93377690)	gap-pcr	NA
MI3	6	Chr6:(–)(84894850)::chr4:(+)(81611154)	Sanger	Blunt end
MI3	7	Chr6:(+)(84894849)::chr4:(+)(82247031)	Sanger	Blunt end
MI4	1	Chr10:(+)(1141977{7}):chr10:(–)(6804828{8})	Sanger	MH, 1bp
MI4	2	Chr12:(+)(1516307{9}):chr10:(+)(1141977{7})	Sanger	MH, 1bp
MI4	3	Chr12:(–)(15166059):chr12:(+)(15810632)	Sanger	Blunt end
MI4	4	Chr12:(+)(15810630)::chr12:(+)(19890288)	Sanger	Blunt end
MI4	5	Chr12:(+)(19890255)::chr12:(+)(19978665)	Sanger	Blunt end
MI4	6	Chr12:(+)(2371854{8})::chr8:(+)(12056580{0})	Sanger	MH, 1bp
MI4	7	Chr8:(+)(120565798)::chr10:(+)(68048323)	Sanger	Blunt end
MI5	1	Chr6(+)(74138566)::chr6(+)(74346180)	gap-pcr	NA
MI5	2	Chr6:(+): 114469971::chr2:(–):58708053	Sanger	Blunt end
MI5	3	Chr2:(–): 5816279{1}::chr6:(+): 11446997{0}	Sanger	MH, 1bp
MI5	4	Chr6:(+):119963985:: aatctttta::chr2:(+):58708055	Sanger	Insertion, 9bp
MI5	5	Chr2:(+):8401992{2-3}::chr2:(+):3816018{1-2}	Sanger	MH, 2bp
MI5	6	Chr2(+)(58162618)::chr6(+)(119964044)	Gap-pcr	NA
MI5	7	Chr6(+)(162909640)::a::chr6(–)(74343934)	Sanger	Insertion ,1bp
MI5	8	Chr6(–)(7413867{0-1})::chr6(+)(16290964{2-3})	Sanger	MH, 2bp
MI5	9	Chr2:(+)(38160180*)::chr2(+)(84019924)	gap-pcr	NA
MI6	1	Chr20:(+)(40948748)::chr3:(–)(189676111)	gap-pcr	NA
MI6	2	Chr3:(+)(183349269)::chr20:(+)(40948749)	gap-pcr	NA
MI6	3	Chr3:(–)(18334986{8-9})::chr3:(–)(11874283{5-6})	Sanger	MH, 2bp
MI6	4	Chr3:(+)(117789537)::chr3:(+)(119912673)	Sanger	Blunt end
MI6	5	Chr3:(–)(125756425)::chr3:(+)(190261730)	Sanger	Blunt end
MI6	6	Chr13:(+)(19091001{5-6})::chr3:(–)(12575642{3-4})	Sanger	MH, 2bp
MI6	7	Chr3:(+)(19044791{0})::chr3:(+)(19091001{5})	Sanger	MH, 1bp
MI6	8	Chr3:(+)(18933161{8})::chr3:(+)(19044790{9})	Sanger	MH, 1bp
MI6	9	Chr3:(+)(189107560)::chr3:(+)(189105082)	gap-pcr	NA
MI6	10	Chr3:(+)(186592260)::chr3:(–)(186024159)	gap-pcr	NA
MI6	11	Chr3:(–)(182591739)::chr3:(+)(182591232)	gap-pcr	NA
MI6	12	Chr3:(+)(186676625):t: chr3:(–)(182591177)	Sanger	Insertion ,1bp
MI6	13	Chr3:(–) 122637474::chr3:(+)(120152667)	Sanger	Blunt end
MI6	14	Chr3:(+)(120152668)::chr3:(+)(190980450)	Sanger	Blunt end
MI6	15	Chr3:(+)(189295694)::chr3:(–)(190980449)	Sanger	Blunt end
MI6	16	Chr4:(+)(10139960{2~3})::chr3:(–)(11873198{6~7})	Sanger	MH, 2bp

Contd...

Supplementary Table 2: Precise breakpoints junction features delineated by Sanger sequencing or breakpoints detected by mate-pair genome sequencing

Case ID	*Breakpoint*	*Junction*	*Confirmation method*	*Feature*
MI6	17	Chr3:(–)(11873198{6-7})::chr4:(+)(10139960{2-3})	Sanger	MH, 2bp
MI6	18	Chr3:(–)(118443877)::chr3:(+)(189107419)	gap-pcr	NA
MI6	19	Chr3:(–)(190910012)::chr3:(–)(190447866)	gap-pcr	NA

MH: microhomology; NA: not available

Supplementary Table 3: RefSeq genes directly truncated by one or more breakpoints. Genes associated with male infertility are bolded

Genes		
ATP11B	*IGSF11*	***EPS8***
CEP162	*IL1RAP*	***ERC2***
CFAP299	*LOC105377975*	*FSTL1*
CGAS	*OSTN*	*GPR156*
COA1	*P3H2*	*HDAC2-AS2*
CTNNA3	*PRKN*	*TBC1D2*
DGKG	***PTPRD***	***TP63***
EMCN	*PTPRT*	*VRK2*
RMDN2	*SLC17A5*	*SOX5*
SEMA5B	*SLC41A3*	*ST6GAL1*

Supplementary Table 4: RefSeq genes located within topologically associated domains disrupted by one or more breakpoints

Genes							
ADCY5	*COA1*	*FUNDC2P2*	*LINC00945*	*LOC101929130*	*MIR7151*	*PTPRT*	*SOX2*
ADIPOQ	*CORO2A*	*FXR1*	*LINC01122*	*LOC101929353*	*MIR8066*	*PURPL*	*SOX2-OT*
ADIPOQ-AS1	*CRYGS*	*GABBR2*	*LINC01206*	*LOC102723582*	*MIR944*	*PYDC2*	*SOX5*
ALDH1L1	*CTNNA2*	*GALM*	*LINC01216*	*LOC102724152*	*MRAP2*	*QKI*	*SRSF7*
ALDH1L1-AS1	*CTNNA3*	*GART*	*LINC01480*	*LOC102724604*	*MRPL32*	*RABL3*	*SST*
ALG1L	*CYB5R4*	*GMNC*	*LINC01489*	*LOC105374060*	*MRPS24*	*RERG*	*ST6GAL1*
ANKS6	*CYP1B1*	*GPR156*	*LINC01515*	*LOC105374312*	*MTO1*	*RERG-AS1*	*STK17A*
ANXA2P3	*CYP1B1-AS1*	*GSK3B*	*LINC01548*	*LOC105377879*	*MYLK-AS1*	*RFC4*	*STRAP*
ARHGDIB	*DDIT4L*	*GTF2E1*	*LINC01690*	*LOC105377975*	*NANS*	*RIPPLY2*	*STXBP5L*
ART4	*DDX43*	*H2AJ*	*LINC01793*	*LOC105378098*	*NDUFB4*	*RMDN2*	*SUCLG1*
ATL2	*DERA*	*H2AZ1-DT*	*LINC01795*	*LOC107984208*	*NR1I2*	*RMDN2-AS1*	*SYNJ1*
ATP11B	*DGKG*	*H4-16*	*LINC01809*	*LOC285762*	*NT5DC1*	*RNU6-1*	*TAF2*
B3GNT8	*DHFRP3*	*HACD2*	*LINC01815*	*LRRC58*	*OLIG1*	*RNU6-2*	*TBC1D2*
BANK1	*DKFZp451B082*	*HDAC2*	*LINC01883*	*LRRTM1*	*OLIG2*	*RNU6-7*	*TBCCD1*
BCKDHA	*DMAC2*	*HDAC2-AS2*	*LINC01994*	*LRRTM3*	*OOEP*	*RNU6-8*	*TMEM207*
BCL11A	*DMRTC2*	*HECW1*	*LINC01995*	*LUARIS*	*OSTN*	*RNU6-9*	*TMEM50B*
BCL6	*DNAJC12*	*HECW1-IT1*	*LINC02020*	*LYPD4*	*OSTN-AS1*	*ROPN1B*	*TMEM91*
BLVRA	*DNAJC19*	*HGD*	*LINC02024*	*MAATS1*	*P3H2*	*RPL39L*	*TP63*
BMP3	*DNAJC28*	*HNRNPLL*	*LINC02035*	*MAN1A1*	*PACRG*	*RTP1*	*TPI1P3*
C12orf60	*DONSON*	*HS3ST5*	*LINC02041*	*MAP3K7*	*PACRG-AS1*	*RTP2*	*TPRG1*
C21orf62	*DPPA5*	*HSPBAP1*	*LINC02043*	*MAP6D1*	*PACRG-AS2*	*RTP4*	*TRIM14*
C21orf62-AS1	*DTX3L*	*IFNAR1*	*LINC02049*	*MASP1*	*PACRG-AS3*	*SEC22A*	*UBE2D4*
CAHM	*ECHDC3*	*IFNAR2*	*LINC02052*	*MCM9*	*PARL*	*SEMA5B*	*URGCP*
CASC6	*EEF1A1*	*IFNGR2*	*LINC02103*	*MGP*	*PARP14*	*SFTA1P*	*URGCP-MRPS24*
CCDC39	*EMCN*	*IGSF11*	*LINC02398*	*MIR1255A*	*PARP15*	*SLC17A5*	*USP6NL*
CCDC50	*ENPP2*	*IGSF11-AS1*	*LINC02468*	*MIR198*	*PARP9*	*SLC39A8*	*UTS2B*
CCDC66	*EPS8*	*IL10RB*	*LINC02534*	*MIR3938*	*PAXBP1*	*SLC41A3*	*VRK2*
CDH9	*ERC2*	*IL10RB-DT*	*LINC02613*	*MIR3943*	*PAXBP1-AS1*	*SLC49A4*	*WBP11*
CEACAM21	*ERC2-IT1*	*IL1RAP*	*LINC02614*	*MIR4282*	*PCAT19*	*SLCO1B1*	*WDFY3-AS2*
CEACAM3	*ERICH4*	*KCNQ5*	*LINC02670*	*MIR4300*	*PDE3A*	*SLCO1B3*	*WDR5B*
CEACAM4	*ERP27*	*KCNQ5-AS1*	*LINC02671*	*MIR4300HG*	*PDE6H*	*SLCO1B3-SLCO1B7*	*WNT5A*
CEACAM5	*ETNK1*	*KCNQ5-IT1*	*LINCR-0002*	*MIR4432*	*PDIA5*	*SLCO1B7*	*YEATS2*
CEACAM6	*EXOSC5*	*KHDC1*	*LOC100131635*	*MIR4432HG*	*POLQ*	*SLCO1C1*	*YEATS2-AS1*
CEACAM7	*FAM162A*	*KHDC1L*	*LOC101926968*	*MIR4447*	*POLR2J4*	*SMCO3*	
CELF2	*FAM184A*	*KHDC3L*	*LOC101927138*	*MIR4643*	*PPP3CA*	*SNAP91*	
CELF2-AS1	*FAM86JP*	*KLHL24*	*LOC101927159*	*MIR548B*	*PRKG2*	*SNAR-I*	
CELF2-AS2	*FANCL*	*KLHL6*	*LOC101928107*	*MIR548I1*	*PRKN*	*SNORA101A*	
CELF2-DT	*FGF12*	*KLHL6-AS1*	*LOC101928441*	*MIR5682*	*PSMA2*	*SNORA63D*	
CEP162	*FGF12-AS1*	*KPNA1*	*LOC101928882*	*MIR6501*	*PTPRD*	*SNORA63E*	
CFAP299	*FLJ20021*	*LINC00710*	*LOC101928942*	*MIR6529*	*PTPRD-AS2*	*SNORD141A*	
CGAS	*FRK*	*LINC00888*	*LOC101928961*	*MIR6854*	*PTPRO*	*SNORD141B*	
CGAS	*FSTL1*	*LINC00901*	*LOC101929106*	*MIR7110*	*PTPRO*	*SON*	

Supplementary Table 5: Breakpoint junction sequence feature summary

Breakpoint junction feature	*Number*	*Proportion (%)*
Microhomology[a]	15	41.7
Small insertion[b]	6	16.7
Blunt end	15	41.7
Total	36	

[a]Microhomology ranged from 1–3 bp; [b]Small insertion ranged from 1–15 bp