Zh. Kozhabek[1,2*], M. Pang[2], Q. Zh. Zhao[2], J.Y. Yi[2], W.D. Huang[2]

[1]Saken Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan;
[2]Reproductive medicine center, XinJiang JiaYin Hospital, Urumchi, China
[*]Corresponding author: jaynahojabek@yahoo.com

Copy number variation in female infertility and candidate gene screening for common infertility-related diseases

To investigate the correlation between the gene copy number variation and female infertility we collected 3962 female infertility samples and analyzed copy number variation (CNV) using high-throughput sequencing technologies. In this study 269 CNVs were found in 246 samples, 17 of which were new CNVs. The occurrence of CNVs was mostly found in X chromosome, and some candidate genes related to female infertility were screened. We also found some high frequency CNVs, which contain important functional genes. This study filled the blank of CNV research on female infertility and discovered the characteristics of CNV (CNV preference, recurrent CNV), which provided genetic reference for female infertility.

Keywords: copy number variation, high-throughput sequencing technology, female infertility, candidate gene, sex chromosomes, gene screening.

Introduction

CNVs have been found to be responsible for a wide range of human diseases [1; 21], CNVs on sex chromosomes are more likely to play key roles in germ cell development [2]. In the past two decades male infertility caused by CNV has been continuously reported [3], especially the sex chromosome CNV (Y chromosomal micro-deletion) has been widely studied [4]. Compared with the male infertility-related CNV, the correlation between female infertility and CNV is rarely reported.

In recent years the next generation sequencing (NGS) has been widely applied in chromosome aneuploidy testing and CNV detection. NGS has unique advantages of low cost, short period, high resolution, high accuracy, etc. Wang et al. used high-throughput sequencing to analyze the CNVs in spontaneous abortion. The results showed that the sensitivity of high-throughput sequencing was consistent with the result of CGH [5].

This study mainly analyzed CNVs of 3962 female infertility patients using high-throughput sequencing technology, discovered the CNV characteristics of Chinese female infertility population, and provided genetic reference for the possible causes of female infertility.

Materials and Methods

Sample collection and DNA extraction

This study was approved by the Clinical Research Ethics Committee of the Xin Jiang Jia Yin hospital (JY2017012). All patients signed written information consent before participation. 3962 blood samples of female infertility patients were collected from January 2017 to December 2018. DNA was extracted using TIANamp Genomic DNA Kit (Tiangen biochemical reagent co.).

Library construction and sequencing

After DNA fragmentation, 150–300 bp DNA was obtained, and then the library was constructed using Ion Xpress Plus Fragment Library Kit (Life Technologies). The library samples were quantitatively mixed into a chip for sequencing. The total data of each sample was about 2.5 M, and the average sequencing depth was 0.1X.

Data analysis

All sequencing reads were aligned to human genome reference sequences (version: NCBI Build 37/hg19) by TMAP software. Meanwhile, duplicate sequences were removed by Picard software. Then each chromosome was divided into 40 kb non-overlapping bins, and the number of reads mapping to each bin was calculated. We normalized the GC percentages in each bin by LOWESS regression.

Circular binary segmentation (CBS), a reliable algorithm that is widely used in the analysis of comparative genomic hybridization arrays, allowed us to precisely define the change points by partitioning chromo-

somes into regions of equal copy numbers. The CNVs obtained by analysis were compared to the DGV and other related databases, and the normal polymorphism CNVs were filtered out.

Results and Discussion

Classification statistics and clinical phenotypes of infertility

Among the 3962 infertile women, primary infertility identified in 1549 patients, the average age of infertility was 32.15±4.54, and the mean years of infertility were 4.35±3.21. 2413 patients with secondary infertility, the average age of infertility was 34.84±5.70, and the mean years of infertility were 3.46±3.15 (Table 1). The clinical phenotypes of patients were as follows: Oviduct obstruction, adhesion, or hydrops; Endometritis; Uterine polyps; Pelvic inflammatory disease; Ovarian cyst; Fibroma; Premature ovarian failure; Polycystic ovary syndrome; Recurrent abortion; Reduced multivariable reserve; Hypogonadism; Hypothyroidism; Endometriosis; Uterine leiomyoma, etc. (Table 1).

T a b l e 1

The characteristics of samples

Parameters	Primary infertility	Secondary infertility	P Value	Total
Number, pieces	1549	2413		3962
Mean age, years	32.15±4.54	34.84±5.70	3.7346E-07	33.76±5.43
The mean years of infertility, pieces	4.35±3.21	3.46±3.15	4.9382E-03	3.9±3.21
The number of CVN samples, pieces	121	148		269

The detection of CNV

After the DGV filtration, we detected 269 CNVs in 246 samples among 3962 female infertility patients. The detection rate of CNV in female infertility was 6.21%, NGS cannot detect chromosomal structural abnormalities, the detection rate of CNV may be lower than actual value. We report for the first time that CNVs in female infertility patients mostly occur on X chromosome (54 CNVs in 49 samples), followed by chromosome 16 (31 CNVs in 30 samples) (Figure 1).
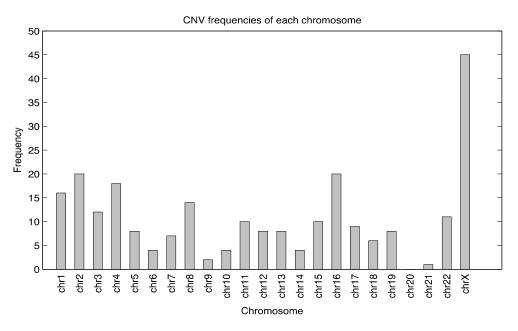

Figure 1. Frequencies of CNVs on different chromosomes

The frequency of micro-duplication was higher than micro-deletion, the average size of deleted fragments is larger than duplicated fragments (Reject outliers, 1.83M>1.51M). In this study 17 new CNVs were found. The new CNV means which is not recorded or overlapping region is less than 40% in available database such as DGV, ISCA, and DECIPHER.

Screening of candidate genes

Of 3962 patients, 133 patients affected by fibroids, 41 affected by premature ovarian failure, 92 affected by polycystic ovarian syndrome, and one affected by Ovarian dysgenesis (Table 2). Some of these patients can identify chromosomal rearrangement. According to previous reports [6, 7], we searched relevant databases and screened some candidate genes that might lead to these diseases (Table 2).

T a b l e 2

Phenotypic distribution of infertility patients

Clinical phenotype	Number	CNV
Endometritis	1506	86
Metropolypus	734	41
No significant symptom, except infertility	461	24
Fibroids	375	21
Uterine hydrops and adhesions	344	15
Oviduct blocking	231	8
Endometriosis	119	8
Recurrent abortion	117	28
Polycystic ovarian syndrome	92	17
Diminished ovarian reserve	43	5
Premature ovarian failure	32	15
Hypothyroidism	32	4
Chocolate cyst of ovary	28	1
Oviduct absence	27	2
Hydatidiform mole	19	1
Unicornuate uterus	16	2
Mediastinal uterus	16	1
Incomplete mediastinal uterus	7	0
Hyperprolactinemia	3	0
Fertilization failure	3	0
Ovarian dysgenesis	1	1
Uterus duplex	1	0
White lesion of vulva	1	1
Without menstruation, with hormone maintenance	1	1

The preference of CNV in female infertility

Our result showed that CNVs are mostly detected in chromosome X (20.07%, 54/269) and chromosome 16 (11.52%, 31/269) in female infertility. X chromosome was the key of female sex organ development, and the long arm of X chromosome was important area of gonad development. When X chromosome was mutated, it is easy to cause gonad dysplasia and fertility decline [8]. The high frequency CNV on chromosome 16 is even more surprising. Studies have shown that the most chromosomal abnormalities in abortion tissue samples detected on chromosome 16 [9]. We believe that our results partly explain the reason why chromosome 16 abnormalities are frequently detected in abortion tissue samples (Table 3). The CNV of the female genome causes abnormal gamete formation or abnormal meiosis of the oocyte, thus leading to abortion.

T a b l e   3 **Candidate genes for infertility-related diseases**

CNV type	Case ID	Region	Start-end position	Size	Disease-related gene in OMIM	Patient's phenotype
Del	Case 1	1q43	240400001–242960000	1.56Mb	*FMN2*	POF
Del	Case 2	1q32.2	212699171–212781031	0.12Mb	*ATF3*	Fibroids
Del	Case 3	2p11.2	84704631–85193625	0.26Mb	*DNAH6*	POF
Dup	Case 4, Case 5	3q22-q23	138659146..138666000	1.20Mb	*FOXL2*	Identified in 2 PCOS patients
Dup	Case 6	6q23.1	132268931–132269962	1.20 Mb	*CTGF*	Fibroids
Dup	Case 7	7p31.1	110411109–111216302	0.61Mb	*IMMP2L*	OD
Dup	Case 8	7p22.1	102436591–103012420	0.22Mb	*FBXL13, CRYZP1, ARMC10, RPL19P12, NAPEPLD*	POF
Dup	Case 9	9q33	120456128..120480000	1.02Mb	*TLR4*	endometriosis
Del	Case 10	12q24.32	125182036–125400631	0.76Mb	*SCARB1, UBC, MIR5188*	POF
Del	Case 11	15q25.1-q25.2	80760001–82520000	1.76Mb	*EFTUD1P1, C15orf26*	POF
Dup	Case 12	17q12	35960001–36120101	0.16Mb	*SYNRG, DDX52, HNF1B, DUSP14*	POF
Dup	Case 13	22q13.2	43072361–43490182	0.38 Mb	*MCAT, ARFGAP3, A4GALT, BIK, PACSIN2, TTLL1*	POF
Del, Dup	Case 14, Case 15	Xp22.31	6609312–8102139	0.17 Mb	*PNPLA4, STS, HDHD1, VCX*	POF, PCOS
Dup	Case 16	Xq21.12	76617292–76987631	1.87Mb	*PGAM4, GLUD2a, ATP7A, COX7B, ATRX, FGF16*	POF
Dup	Case 17	Xq12	66710369–67539032	0.58Mb	*AR, OPHN1*	POF

New CNVs

17 new CNVs were detected in this study. Although these CNV-related diseases are included in OMIM, whether they are associated with female infertility has not yet been confirmed, but they can provide some genomics and genetics reference for female infertility.

Recurrently found CNVs

We identified several CNVs with high frequency. Xp22.31 was detected in 24 patients. The clinical phenotypes were salpingitis, uterine fibroids, endometritis, polycystic ovary syndrome, endometriosis, ovarian chocolate cyst, fibroids, recurrent abortion, abnormal tubal development, familial X-linked ichthyosis, white lesion of vulva, and decreased ovarian reserve function. The region including STS, VCX2, VCX, VCX3A, PNPLA4, HDHD1, etc. It was reported that the deletion, reduplication, and rearrangement of Xp22.31 can lead to X-linked ichthyosis. Li et al. believed that the Xp22.31 duplicated patients have large clinical differences and highly variable phenotypes [10]. Xq13–26 was a key area of ovarian development [8], so mutation of Xq13–26 will lead to abnormal expression of sex gene and sex hormone level, which may affect fertility. In 2010 Krause et al. found the Xp22.31 in two male infertility patients [11]. Xp22.31 may play an important role in female reproduction.

In 4 patients we detected chromosomal rearrangements affecting the TBX1, CDC45, COMT genes in 22q11.21. The clinical phenotypes of the patients were primary infertility (one of them fibroids, one was endometriosis and metropolypus, two were metropolypus). TBX1 gene encodes a T-box transcription factor, which regulates the transcription and expression of a series of genes, thus affecting the fertility. CDC45 is associated with DNA replication and the COMT gene encodes catechol-o-methyltransferase. 22q11.2 was found in two MRKH syndrome (Mayer-Rokitansky-kuster-hauser syndrome, congenital Vaginal Deletion) [12]. 22q11.2 micro-duplication syndrome has been extensively studied, especially in prenatal diagnosis. Low copy duplication in this region occurs in homologous recombination and mediates non-allele recombination, resulting in rearrangement of 22q11.2.

We detected rearrangements in 15q11.2 affecting GOLGA6L1, OR4M2, OR4N4, POTE, and POTE2 genes in 4 patients. The clinical phenotypes of this patients were endometritis (two cases) and fibroid. Abnormality of 15q11.2 was also found in normal people. Some research suggest that duplication of 15q11.2 is associated with male infertility [13]. Some studies report that POTE plays an important role in human sperm maturation [14]. Others believe that GOLGA8C replication affects male infertility.

Rearrangements in 16p13.11 were found in 14 patients, including primary infertility in 5 patients, secondary infertility in 9 patients. They affected by premature ovarian failure (2 cases), tubal obstruction, hypogonadism, and uterine polyp. The variant region is susceptible to neurocognitive impairment and contains 20 protein-coding genes, including KIAA0430, NDE1, and RRN3. KIAA0430 plays a key role in regulating oogenesis, inhibiting transposon translocation during female meiosis, which is related to reproductive system integrity. NDE1 protein is located at the poles of centrosome and mitotic spindle, which mainly interacts with LIS-1 protein and fibrin. NDE1-LIS1-fibrin complex participates in neurogenesis through mediating a series of intracytoplasmic activities. The proliferation of cells affects the development of cerebral cortex and nerve. RRN3 is associated with the initiation of effective transcription by RNA polymerase I. Different individuals have different clinical manifestations. Ullmann et al. reported for the first time in 2007 a significant association between 16p13.11 duplication and autism [15]. In Hannel's study of mental retardation and multiple congenital abnormalities (MR/MCA) 16p13.11 regions deletion was found to be a serious pathogen, and duplication leads to benign outcomes [16]. Ramalingam and Tropeano reported that 16p13.11 associated with copies of duplicated pathogens. The main clinical symptoms were associated with neurodevelopment [17]. Paciorkowski et al. showed that 16p13.11 micro-deletion was associated with fetal brain development [18].

The 17p12 region contains COX10 and PMP22 genes. COX10 is expressed in the heart, skeletal muscle, and testis, which has certain effect on male azoospermia [19]. PMP22 gene encodes a plasma membrane integration glycoprotein, which is related to the peripheral nervous system. In the study about azoospermic men, this site was detected in three patients.

Polycystic ovary syndrome, endometriosis, Premature Ovarian Failure, and fibroids are not simply caused by single gene mutation or chromosomal abnormality, but affected by the cumulative effect of multiple genes. They generally interact with environmental factors, and belong to polygenic hereditary diseases. Because there are hereditary factors in it, the disease tends to be familial; but it does not conform to Mendel's hereditary law. Based on previous studies [6, 7], we are screened several related genes, and the results are also a validation for these studies.

Conclusion

Female infertility is one of the main factors affecting reproductive health. The most common influencing factors are ovulation dysfunction, endometrial abnormalities, rearrangement in sex chromosomes, single gene mutations, and hypogonadism [20, 22]. Because of following factors: environmental factors, individual differences, lack of male phenotype, diversity of characteristic phenotype, bias of investigation, difference of research scale, and diversity of races and regions, it is difficult to obtain consistent clinical research results. But with the development of human genomic research the genetic reason of female infertility will be clearer. We try to exam the correlation between genetic disorder and female infertility. In conclusion, we found infertility related CNVs mostly occurred in X chromosome and high frequency CNVs which contained important functional genes. Also, some candidate genes that related to female infertility were screened. This study filled the blank of CNV research on female infertility and discovered the characteristics of CNV (CNV preference, recurrent CNV), which provided genetic reference for female infertility.

References

1   Stankiewicz, P. & Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. *Annual Review of Medicine, 61,* 437–455.

2   Junhao, Y. et al. (2011). DYZ1 copy number variation, Y chromosome polymorphism and early recurrent spontaneous abortion/early embryo growth arrest. *European Journal of Obstetrics Gynecology & Reproductive Biology, 159,* 371–374.

3   Eggers, S. et al. (2015). Copy number variation associated with meiotic arrest in idiopathic male infertility. *Fertility & Sterility, 103,* 214–219.

4   Yang, Y. et al. (2003). Studies on molecular epidemiology of Y chromosome azoospermia factor micro-deletions in Chinese patients with idiopathic azoospermia or severe oligozoospermia. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 20,* 385–389.

5 Wang, M. Z. et al. (2017). Semiconductor Sequencing Analysis of Chromosomal Copy Number Variations in Spontaneous Miscarriage. *Medical Science Monitor International Medical Journal of Experimental & Clinical Research, 23,* 5550–5557.

6 Ledig, S., Röpke, A. & Wieacker, P. (2010). Copy Number Variants in Premature Ovarian Failure and Ovarian Dysgenesis. *Sexual Development, 4,* 225–232.

7 Tšuiko, O. et al. (2016). Copy number variation analysis detects novel candidate genes involved in follicular growth and oocyte maturation in a cohort of premature ovarian failure cases. *Human Reproduction, 31,* 1913–1925.

8 Gardner, R. J. M., Sutherland, G. R. & Shaffer, G. L. (1997). Chromosome Abnormalities and Genetic Counseling. *American Journal of Human Genetics, 60,* 1567.

9 Nicopoullos, J. D. M. et al. (2008). The role of sperm aneuploidy as a predictor of the success of intracytoplasmic sperm injection? *Human Reproduction, 23,* 240.

10 Li, F. et al. (2010). Interstitial micro-duplication of Xp22.31: Causative of intellectual disability or benign copy number variant? *European Journal of Medical Genetics, 53,* 93–99. Retrieved from https://doi.org/10.1016/j.ejmg.2010.01.004

11 Csilla, K. et al. (2012). High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. Plos One, 7, e44887.

12 Cheroki, C. et al. (2010). Report of a del22q11 in a patient with Mayer-Rokitansky-Küster-Hauser (MRKH) anomaly and exclusion of WNT-4, RAR-gamma, and RXR-alpha as major genes determining MRKH anomaly in a study of 25 affected women. *American Journal of Medical Genetics, Part A, 140A,* 1339–1342.

13 Céline, D. et al. (2015). Prenatal diagnosis of 24 cases of micro-duplication 22q11.2: an investigation of phenotype-genotype correlations. *Prenatal Diagnosis, 35,* 35–43.

14 Tomoko, I. et al. (2008). Expression of POTE protein in human testis detected by novel monoclonal antibodies. *Biochem Biophys Res Commun., 365,* 603–608.

15 Reinhard, U. et al. (2010). Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Human Mutation, 28,* 674–682.

16 Hannes, F.D. et al. (2009). Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *Journal of Medical Genetics, 46,* 223.

17 Ramalingam, A. et al. (2011). 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. *Journal of Human Genetics, 56,* 541, Retrieved from https://doi.org/10.1038/jhg.2011.42

18 Paciorkowski, A. R. et al. (2013). Deletion 16p13.11 uncovers NDE1 mutations on the non-deleted homolog and extends the spectrum of severe microcephaly to include fetal brain disruption. *American Journal of Medical Genetics, Part A, 161,* 1523–1530.

19 Bo, Y. et al. (2009). Expression of COX10 in Human Non-Obstructive Azoospermia Testes. *National Journal of Andrology, 15,* 599.

20 Signore, F. et al. (2020). The Role of Number of Copies, Structure, Behavior and Copy Number Variations (CNV) of the Y Chromosome in Male Infertility. *Genes, 11(1),* 40.

21 Patel, B. et al. (2018). Comprehensive genetic testing for female and male infertility using next-generation sequencing. *J Assist Reprod Genet., 35 (8);* 1489–1496.

22 Zhang, R. et al. (2019). Prevalence of chromosomal abnormalities identified by copy number variation sequencing in high-risk pregnancies, spontaneous abortions, and suspected genetic disorders. *J Int Med Res., 47 (3);*1169–1178.

Ж. Қожабек, М. Пан, Ч.Ж. Жау, Ж.Й. И., У.Д. Хуан

Әйелдер бедеулігіндегі геном көшірме нөмірінің өзгеруі және жалпы бедеулікке байланысты аурулар үшін кандидат гендер скринингі

Әйелдердің бедеулігі мен геном көшірме нөмірінің өзгеруі (copy number variation, CNV) арасындағы корреляцияны зерттеу үшін 3962 беду әйелдердің қан үлгілері жиналды және жаңа буынды жоғары өнімді секвенирлеу технологиясын (high-throughput sequencing technology) пайдаланып, геном көшірме нөмірлерінің өзгеруіне анализ жасалды. Осы зерттеуде 246 дана үлгі материалдан 269 дана CNV табылды. Олардың ішінде 17 дана жаңа CNV байқалды және CNV-дің пайда болуы көбінесе Х хромосомасында кездеседі. Кейбір кандидат гендердің әйелдердің бедеулігіне қатысы бар екендігі анықталды. Сонымен қатар, кейбір жоғары жиілікті CNV-дің маңызды функционалды гендерді қамтитыны байқалды. Бұл зерттеу әйелдер бедеулігінің CNV зерттеуіндегі олқылықтың орнын толтырады, себебі әйелдерді бедеулікке кіріптар ететін генетикалық факторлар (болжамды, CNV, қайталанатын CNV) талданды.

Кілт сөздер: көшірме нөмірінің өзгеруі, жаңа буынды жоғары өнімді секвенирлеу технологиясы, әйелдер бедеулігі, кандидат гендер, жыныс хромосомасы, ген скринингі.

Ж. Кожабек, М. Пан, Ч.Ж. Жау, Ж. Й. И, У.Д. Хуан

Вариация числа копий при женском бесплодии и скрининг генов-кандидатов общих заболеваний, связанных с бесплодием

Чтобы исследовать корреляцию между вариацией числа копий генома и женским бесплодием, авторами собраны 3962 образца крови женщин, страдающих бесплодием, и проанализирована вариация чисел копий (CNV) с помощью технологий высокопроизводительного секвенирования. В этом исследовании 269 CNV были обнаружены в 246 образцах, 17 из которых были новыми CNV. Наличие CNV в основном обнаруживается в Х-хромосоме, были проверены некоторые гены-кандидаты, связанные с женским бесплодием. Кроме того, обнаружено несколько высокочастотных CNV, содержащих важные функциональные гены. Это исследование заполнило пробел в исследовании CNV по женскому бесплодию и выявило характеристики CNV (предпочтительная, CNV, повторяющаяся CNV), которые обеспечивают генетический ориентир женского бесплодия.

Ключевые слова: вариация числа копий, технология высокопроизводительного секвенирования, женское бесплодие, ген-кандидат, половые хромосомы, генетический скрининг.