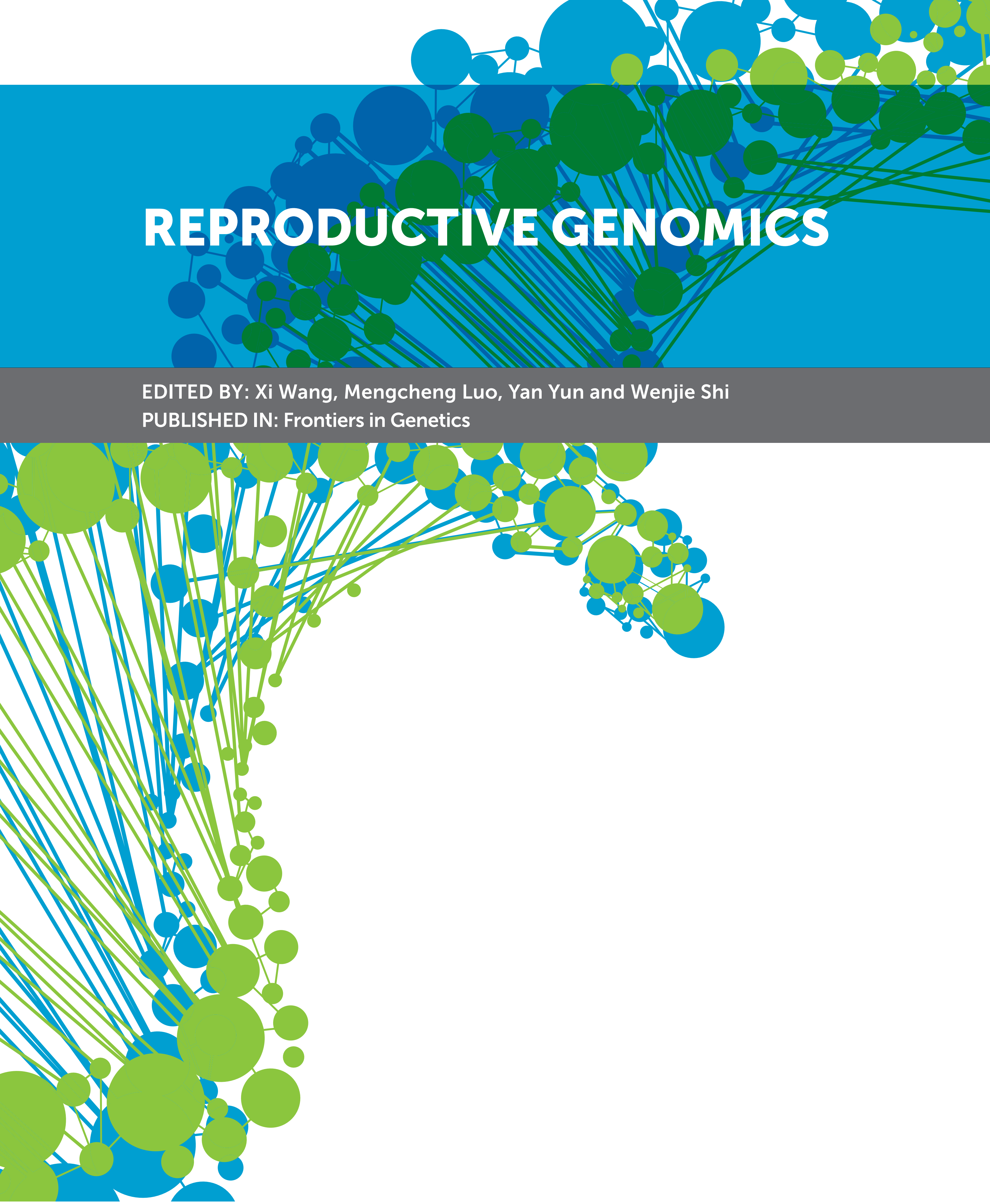


REPRODUCTIVE GENOMICS

EDITED BY: Xi Wang, Mengcheng Luo, Yan Yun and Wenjie Shi
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-869-1
DOI 10.3389/978-2-88976-869-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

REPRODUCTIVE GENOMICS

Topic Editors:
Xi Wang, Nanjing Medical University, China
Mengcheng Luo, Wuhan University, China
Yan Yun, University of California, Davis, United States
Wenjie Shi, Academy of Military Medical Sciences (AMMS), China

Citation: Wang, X., Luo, M., Yun, Y., Shi, W., eds. (2022). Reproductive Genomics. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-869-1

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 23 August 2022
DOI 10.3389/fgene.2022.1002458

OPEN ACCESS

EDITED AND REVIEWED BY
Jared C. Roach,
Institute for Systems Biology (ISB),
United States

*CORRESPONDENCE
Rong Liu,
liurong19840901@whu.edu.cn
Xi Wang,
xiwang@njmu.edu.cn
Mengcheng Luo,
luomengcheng@whu.edu.cn

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 25 July 2022
ACCEPTED 28 July 2022
PUBLISHED 23 August 2022

CITATION
Liu R, Yun Y, Shu W, Wang X and Luo M (2022), Editorial:
Reproductive genomics.
Front. Genet. 13:1002458.
doi: 10.3389/fgene.2022.1002458

Editorial: Reproductive genomics

Rong Liu[1]*, Yan Yun[2], Wenjie Shu[3], Xi Wang[4]* and Mengcheng Luo[1]*

[1]Hubei Provincial Key Laboratory of Developmentally Originated Disease, Department of Human Histology and Embryology, Taikang Medical School (School of Basic Medical Sciences), Wuhan University, Wuhan, China, [2]Department of Microbiology and Molecular Genetics, University of California, Davis, CA, United States, [3]Beijing Institute of Microbiology and Epidemiology, Beijing, China, [4]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China

KEYWORDS

reproductive biology, reproductive medicine, reproductive genomics, reproductive genetics, single-cell sequencing, omics, preimplantation genetic testing, risk factors

Editorial on the Research Topic
Reproductive genomics

In all sexually reproducing animals, a new life starts with a zygote, which derives from the successful fusion of a mature oocyte with a mature sperm. Despite the complicated journey taken for a zygote to become a well-developed individual, the production of the mature gametes (i.e., sperms and oocytes) also relies on the normally programmed gametogenesis of both the male and female. This early development is comprised of many complicated and delicate processes, accompanying with numerous molecular regulation and metabolic changes. For example, PLZF (OMIM: 176797) and c-KIT (OMIM: 164920) are associated with self-renewal and differentiation of spermatogonial stem cells (Buaas et al., 2004; Costoya et al., 2004; Zhang et al., 2013), DNA-binding protein inhibitor ID2 (OMIM: 600386) is the key transcription factor in mouse primordial follicle formation (He et al., 2021a). Reprogramming of epigenome is essential for the early embryo development (Xia and Xie, 2020; Xu R. et al., 2021).

There is an urgent need to obtain a better understanding of the fundamental reproductive biology. This need is driven by two factors. First, there has been a decline in birth rate during the past half century (Skakkebaek et al., 2022). Second, infertility in human caused by genetic defects is a barrier for many couples (Zorrilla and Yatsenko, 2013; Krausz and Riera-Escamilla, 2018). Recently, considerable efforts have been devoted to achieving better understanding of the molecular networks, regulatory programs, and germline-soma or soma-soma communications during gametogenesis and embryogenesis. In particular, the rapid advancement of high-throughput sequencing technologies and ultra-low-input (or even single-cell) omics approaches have accelerated novel discoveries in reproductive development e.g., (Wang et al., 2018; Qiao et al., 2020; Xu K. et al., 2021; He et al., 2021b; Tyser et al., 2021; Yan et al., 2021; Wu et al., 2022; Xiong et al., 2022). Towards this direction, Qian et al. profiled the transcriptomes of 14,315 single testicular cells from adult zebrafish testes by single-cell RNA sequencing (scRNA-seq) using the 10x

Genomics Chromium platform, identified ten distinguishable cell types with novelly revealed marker genes, and characterized interactions between somatic cells and germ cells through ligand-receptor analysis. This study provides an important resource for the studies of spermatogenesis in zebrafish, as well as mechanisms associated with human male infertility by using zebrafish as a model organism (Hoo et al., 2016). In addition to the application of single-cell omics to reproductive biology, Dodlapati et al. developed artificial intelligence (AI)-based novel computational approaches for the imputation of single-cell DNA methylome profiles with ultra-low coverage. The high effectiveness of the new algorithm was demonstrated by its application to bovine oocytes and early embryos, highlighting the potential for reconstructing epigenome-mediated transcriptional regulatory networks at the single-cell level in early animal development. Focusing on gene expression at the RNA level, RNA-sequencing was utilized to study oocyte development between different sex chromosome complement (by Yamazaki et al.) and between species (by Zhang et al.). More specifically, Yamazaki et al. compared transcriptomes among oocytes derived from XX, XO (i.e., monosomy X) and sex-reversed XY female mice (Vaz et al., 2020; Yamazaki et al., 2022) along follicular growth up to maximum size allowed without significant morphological differences. They found losing a copy of X chromosome is the dominant effect on gene expression changes in only XO oocytes, whereas the transcriptome landscape in XY oocytes is associated with the expression of Y-linked genes. Meanwhile, Zhang et al. performed comparative transcriptomic analysis and weighted gene co-expression network analysis (WGCNA) to compare the transcriptomes of donkey oocytes to that of cattle, sheep, pigs, and mice. They uncovered unique aspects of gene expression of donkey oocyte development from germinal vesicle (GV) to metaphase II (MII), thus providing new insights into the key regulators in donkey oocyte development.

Besides the efforts made in the fundamental research of reproductive biology, the accessibility of high-throughput sequencing and accumulated knowledge on inherited diseases have led to an extensive development and application of new technologies in screening, diagnosis and prevention of reproduction and pregnancy-related diseases (Hu et al., 2011; Hou et al., 2013). In order to develop an efficient and cost-effective method for screening pathogenic genes in infertility patients, Yuan et al. designed a target-sequencing panel containing 22 female infertility-related genes and applied such genetic screening to 68 patients with primary infertility or recurrent pregnancy loss. The authors demonstrated that the target-sequencing approach can be applied not only for genetic screening in IVF clinics, but also for uncovering novel pathogenic variants in the infertility-related genes. To reduce the risk of pregnancy loss, Pei et al. developed a clinically applicable method for detecting chromosomal reciprocal translocations in human embryos by long-read Nanopore sequencing and breakpoints region analysis. With this approach, the accurate and precise identification of balanced translocations provides a way of selecting embryos with normal karyotype for transfer into the uterus. Due to the association of reciprocal translocations and reproductive problems (Morin et al., 2017), aiming at providing more appropriate genetic counseling for couples with autosomal reciprocal translocations on their chances of producing normal blastocysts, Xie et al. evaluated several factors that may affect the unbalanced rearrangement of reciprocal translocations. The authors analyzed the meiotic segregation patterns in 10,846 blastocysts from 2,871 preimplantation genetic testing cycles of reciprocal translocation carriers, and found decreased proportions of alternate segregation in blastocysts when an Acr-ch, female sex, and lower TAR1 were involved. To avoid aneuploid embryo transfers in patients with implantation failure and pregnancy loss, Chen et al. analyzed the correction of time lapse-based blastocyst morphological scores (TLBMSs) with mosaic levels. In the study, high-resolution next-generation sequencing (NGS) was applied to evaluate the mosaic level of a blastocyst, and time-lapse embryo assessments were refined at a uniform time-point. With 918 biopsied blastocysts, Chen et al. showed that the redefined blastocyst morphology components and the converted TLBMSs are significantly correlated with all of the threshold levels of mosaicism. Expanded carrier screening (ECS) is applied to identify at-risk couples who carry heterozygous disease-causing variants and to avoid birth defects (Martin et al., 2015). Since the current expanded carrier sequencing panel cannot fully cover the variant spectrum in the East Asian population (Guo and Gregg, 2019), Tong et al. aimed to reveal the carrier spectrum in the Chinese population and to delineate an expanded carrier gene panel suitable for Chinese people. They screened 2,234 couples and found 94.5% of them were carriers of at least one disease-causing variant, and at-risk couple rate was 9.8%, highlighting the necessity of establishing a Chinese population-tailored ECS gene panel and conducting ECS for couples before receiving assisted reproductive technology. Serving as a successful example of preimplantation genetic testing (PGT), Ren et al. reported a case of preventing the transmission of a disease-causing *IL2RG* (OMIM: 308380) variant in a family. This variant will cause X-linked severe combined immune deficiency (SCID-X1), which is a recessive monogenic hereditary disease. In this case, Sanger sequencing for validating the mutated allele and linkage analysis based on single nucleotide polymorphism (SNP) haplotype via NGS were performed simultaneously. The authors then transferred an embryo without copy number variation and inherited variants at the *IL2RG* gene, and saw a healthy girl was born finally. This case demonstrates the usefulness of PGT in preventing mutated allele transmission.

Related to the assisted reproductive technology, Wang et al. compared different methods for cryopreserving human ejaculated and testicular spermatozoa by analyzing molecular and metabolic features related to sperm quality, and demonstrated that direct −80 °C freezing could be a viable alternative to liquid nitrogen vapor freezing for short-term human sperm storage. In recent years, totipotent or pluripotent stem cells have been gradually used to model embryonic development for mechanism studies. Towards a better understand of the maintenance and developmental potential of human extended pluripotent stem (hEPS) cells, An et al. constructed inducible WDR36 (OMIM: 609669) knockdown and WDR36-overexpressing hEPS cell lines, and their data demonstrated that WDR36 safeguards the self-renewal and pluripotency of hEPS cells.

So far, accumulating studies, including those in this Research Topic, have gained many novel insights into spatiotemporal regulatory networks during gametogenesis and embryonic development, leading to applications such as molecular diagnosis of various reproductive diseases. To better diagnose and treat infertile couples and to partially address the declining birth rate, both fundamental and translational research is still desired. We anticipate that continued development and refinement of technologies will continue to drive research in this field.

Author contributions

RL, XW, and ML wrote the editorial with input from YY and WS, and all the authors have approved the work for publication.

Acknowledgments

We wish to thank all the authors and reviewers for their valuable contributions to ensure high quality articles for this Research Topic and we hope this Research Topic will be of interest to the research community.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Buaas, F. W., Kirsh, A. L., Sharma, M., McLean, D. J., Morris, J. L., Griswold, M. D., et al. (2004). Plzf is required in adult male germ cells for stem cell self-renewal. *Nat. Genet.* 36 (6), 647–652. doi:10.1038/ng1366

Costoya, J. A., Hobbs, R. M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., et al. (2004). Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat. Genet.* 36 (6), 653–659. doi:10.1038/ng1367

Guo, M. H., and Gregg, A. R. (2019). Estimating yields of prenatal carrier screening and implications for design of expanded carrier screening panels. *Genet. Med.* 21 (9), 1940–1947. doi:10.1038/s41436-019-0472-7

He, Y., Chen, Q., Dai, J., Cui, Y., Zhang, C., Wen, X., et al. (2021a). Single-cell RNA-Seq reveals a highly coordinated transcriptional program in mouse germ cells during primordial follicle formation. *Aging Cell* 20 (7), e13424. doi:10.1111/acel.13424

He, Y., Chen, Q., Zhang, J., Yu, J., Xia, M., and Wang, X. (2021b). Pervasive 3'-UTR isoform switches during mouse oocyte maturation. *Front. Mol. Biosci.* 8, 727614. doi:10.3389/fmolb.2021.727614

Hoo, J. Y., Kumari, Y., Shaikh, M. F., Hue, S. M., and Goh, B. H. (2016). Zebrafish: A versatile animal model for fertility research. *Biomed. Res. Int.* 2016, 9732780. doi:10.1155/2016/9732780

Hou, Y., Fan, W., Yan, L., Li, R., Lian, Y., Huang, J., et al. (2013). Genome analyses of single human oocytes. *Cell* 155 (7), 1492–1506. doi:10.1016/j.cell.2013.11.040

Hu, Z., Xia, Y., Guo, X., Dai, J., Li, H., Hu, H., et al. (2011). A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat. Genet.* 44 (2), 183–186. doi:10.1038/ng.1040

Krausz, C., and Riera-Escamilla, A. (2018). Genetics of male infertility. *Nat. Rev. Urol.* 15 (6), 369–384. doi:10.1038/s41585-018-0003-3

Martin, J., Asan, Yi, Y., Alberola, T., Rodriguez-Iglesias, B., Jimenez-Almazan, J., et al. (2015). Comprehensive carrier genetic test using next-generation deoxyribonucleic acid sequencing in infertile couples wishing to conceive through assisted reproductive technology. *Fertil. Steril.* 104 (5), 1286–1293. doi:10.1016/j.fertnstert.2015.07.1166

Morin, S. J., Eccles, J., Iturriaga, A., and Zimmerman, R. S. (2017). Translocations, inversions and other chromosome rearrangements. *Fertil. Steril.* 107 (1), 19–26. doi:10.1016/j.fertnstert.2016.10.013

Qiao, Y., Ren, C., Huang, S., Yuan, J., Liu, X., Fan, J., et al. (2020). High-resolution annotation of the mouse preimplantation embryo transcriptome using long-read sequencing. *Nat. Commun.* 11 (1), 2653. doi:10.1038/s41467-020-16444-w

Skakkebaek, N. E., Lindahl-Jacobsen, R., Levine, H., Andersson, A. M., Jorgensen, N., Main, K. M., et al. (2022). Environmental factors in declining human fertility. *Nat. Rev. Endocrinol.* 18 (3), 139–157. doi:10.1038/s41574-021-00598-8

Tyser, R. C. V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* 600 (7888), 285–289. doi:10.1038/s41586-021-04158-y

Vaz, B., El Mansouri, F., Liu, X., and Taketo, T. (2020). Premature ovarian insufficiency in the XO female mouse on the C57BL/6J genetic background. *Mol. Hum. Reprod.* 26 (9), 678–688. doi:10.1093/molehr/gaaa049

Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., et al. (2018). Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell* 23 (4), 599–614. e594. doi:10.1016/j.stem.2018.08.007

Wu, X., Lu, M., Yun, D., Gao, S., Chen, S., Hu, L., et al. (2022). Single-cell ATAC-Seq reveals cell type-specific transcriptional regulation and unique chromatin accessibility in human spermatogenesis. *Hum. Mol. Genet.* 31 (3), 321–333. doi:10.1093/hmg/ddab006

Xia, W., and Xie, W. (2020). Rebooting the epigenomes during mammalian early embryogenesis. *Stem Cell Rep.* 15 (6), 1158–1175. doi:10.1016/j.stemcr.2020.09.005

Xiong, Z., Xu, K., Lin, Z., Kong, F., Wang, Q., Quan, Y., et al. (2022). Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat. Cell Biol.* 24 (6), 968–980. doi:10.1038/s41556-022-00928-6

Xu, K., Wang, R., Xie, H., Hu, L., Wang, C., Xu, J., et al. (2021a). Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis* 10 (10), 66. doi:10.1038/s41389-021-00355-6

Xu, R., Li, C., Liu, X., and Gao, S. (2021b). Insights into epigenetic patterns in mammalian early embryos. *Protein Cell* 12 (1), 7–28. doi:10.1007/s13238-020-00757-z

Yamazaki, W., Tan, S. L., and Taketo, T. (2022). Role of the X and Y chromosomes in the female germ cell line development in the mouse (*Mus musculus*). *Sex. Dev.*, 1–10. doi:10.1159/000521151

Yan, R., Gu, C., You, D., Huang, Z., Qian, J., Yang, Q., et al. (2021). Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing. *Cell Stem Cell* 28 (9), 1641–1656.e7. e1647. doi:10.1016/j.stem.2021.04.012

Zhang, L., Tang, J., Haines, C. J., Feng, H., Lai, L., Teng, X., et al. (2013). c-kit expression profile and regulatory factors during spermatogonial stem cell differentiation. *BMC Dev. Biol.* 13, 38. doi:10.1186/1471-213X-13-38

Zorrilla, M., and Yatsenko, A. N. (2013). The genetics of infertility: Current status of the field. *Curr. Genet. Med. Rep.* 1 (4), 247–260. doi:10.1007/s40142-013-0027-1

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/fgene.2021.792604

Effects of the Sex Chromosome Complement, XX, XO, or XY, on the Transcriptome and Development of Mouse Oocytes During Follicular Growth

*Wataru Yamazaki[1,2], Dunarel Badescu[3,4], Seang Lin Tan[2,5,6], Jiannis Ragoussis[3,4] and Teruko Taketo[1,2,5,7]**

[1]Department of Surgery, McGill University, Montreal, QC, Canada, [2]Research Institute of McGill University Health Centre, Montreal, QC, Canada, [3]Department of Human Genetics, McGill University, Montreal, QC, Canada, [4]McGill University Genome Centre, Montreal, QC, Canada, [5]Department of Obstetrics and Gynecology, McGill University, Montreal, QC, Canada, [6]OriginElle Fertility Clinic and Women's Health Centre, Montreal, QC, Canada, [7]Department of Biology, McGill University, Montreal, QC, Canada

Edited by:
Yan Yun,
University of California, Davis,
United States

Reviewed by:
Akihiko Sakashita,
Keio University, Japan
Malgorzata Kloc,
Houston Methodist Research Institute,
United States

***Correspondence:**
Teruko Taketo
teruko.taketo@mcgill.ca

Specialty section:
This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: *10 October 2021*
Accepted: *26 November 2021*
Published: *20 December 2021*

Citation:
Yamazaki W, Badescu D, Tan SL,
Ragoussis J and Taketo T (2021)
Effects of the Sex Chromosome
Complement, XX, XO, or XY, on the
Transcriptome and Development of
Mouse Oocytes During
Follicular Growth.
Front. Genet. 12:792604.
doi: 10.3389/fgene.2021.792604

The sex chromosome complement, XX or XY, determines sexual differentiation of the gonadal primordium into a testis or an ovary, which in turn directs differentiation of the germ cells into sperm and oocytes, respectively, in eutherian mammals. When the X monosomy or XY sex reversal occurs, XO and XY females exhibit subfertility and infertility in the mouse on the C57BL/6J genetic background, suggesting that functional germ cell differentiation requires the proper sex chromosome complement. Using these mouse models, we asked how the sex chromosome complement affects gene transcription in the oocytes during follicular growth. An oocyte accumulates cytoplasmic components such as mRNAs and proteins during follicular growth to support subsequent meiotic progression, fertilization, and early embryonic development without *de novo* transcription. However, how gene transcription is regulated during oocyte growth is not well understood. Our results revealed that XY oocytes became abnormal in chromatin configuration, mitochondria distribution, and *de novo* transcription compared to XX or XO oocytes near the end of growth phase. Therefore, we compared transcriptomes by RNA-sequencing among the XX, XO, and XY oocytes of 50–60 µm in diameter, which were still morphologically comparable. The results showed that the X chromosome dosage limited the X-linked and autosomal gene transcript levels in XO oocytes whereas many genes were transcribed from the Y chromosome and made the transcriptome in XY oocytes closer to that in XX oocytes. We then compared the transcript levels of 3 X-linked, 3 Y-linked and 2 autosomal genes in the XX, XO, and XY oocytes during the entire growth phase as well as at the end of growth phase using quantitative RT-PCR. The results indicated that the transcript levels of most genes increased with oocyte growth while largely maintaining the X chromosome dosage dependence. Near the end of growth phase, however, transcript levels of some X-linked genes did not increase in XY oocytes as much as XX or XO oocytes, rendering their levels much lower than those in XX oocytes. Thus, XY oocytes established a distinct transcriptome at the end of growth phase, which may be associated with abnormal chromatin configuration and mitochondria distribution.

Keywords: ovary, oocyte, XO female, XY sex reversal, transcriptome, mitochondria, BMP15, KDM5B

INTRODUCTION

The sex chromosome complement, XX or XY, determines sexual differentiation of the gonadal primordium into a testis or an ovary, which in turn directs differentiation of the germ cells into spermatogenesis or oogenesis, respectively, in eutherian mammals (McLaren, 1988; Berta et al., 1990; Koopman et al., 1990). When gonadal sex is reversed, however, the germ cell sex becomes discordant with the chromosomal sex. Both sex-reversed XX males and XY females encounter infertility, indicating that functional germ cell differentiation requires the presence of proper sex chromosome complement. While essential roles of the Y chromosome in spermatogenesis have been well defined, the cause of infertility in the XY oocyte has remained an enigma (Taketo, 2015; Yamauchi et al., 2016).

In fetal ovaries, the germ cells enter meiosis to become oocytes and go through the Meiotic Prophase I (MPI), in which homologous chromosomes pair and recombine. In the XX germ cells, one of the two X-chromosomes is initially inactivated like in somatic cells, but it becomes reactivated prior to the entry into meiosis, and the two transcriptionally active X-chromosomes efficiently pair and recombine during the MPI progression (de Napoles et al., 2007; Sugimoto and Abe, 2007; Chuva De Sousa Lopes et al., 2008). In perinatal ovaries, the oocytes reach the end of MPI and form primordial follicles with the neighboring granulosa cells. The oocytes remain arrested at this stage in the ovarian reserve in the entire reproductive life. Upon puberty, a cohort of primordial follicles is recruited into the growth phase, during which granulosa cells vigorously proliferate to form multiple layers while the oocyte increases in volume and undergoes active transcription and translation. At the end of growth phase, the oocytes, now named fully-grown (FG) or germinal vesicle (GV) oocytes, shut down transcription and become competent for going through the first meiotic division and reaching the second meiotic metaphase (MII). Upon fertilization, the MII-oocytes undergo the second meiotic division to initiate embryonic development. The oocytes of XY female mice can go through oogenesis to reach the MII-stage, but fail in embryonic development, resulting in poor fertility or infertility (reviewed by (Taketo, 2015)). We predict that one dosage of X chromosome and the presence of Y chromosome together alter the mRNAs and proteins accumulated during follicular growth and make the XY oocyte incompetent for embryonic development. The oocytes of XO female mice are competent for embryonic development. However, the normal XX oocyte carries two transcriptionally active X chromosomes, and if and how the absence of one X chromosome affects the transcriptome in the XO oocyte remains to be clarified.

In the present study, we asked how the sex chromosome complement affects the oocyte during follicular growth. We used the B6.Y^{TIR} congenic mouse strain to produce XY females. The B6.Y^{TIR} mouse was established by placing the Y chromosome from a variant of *Mus musculus domesticus* caught in Tirano, Italy (TIR) onto the C57BL/6J (B6) genetic background by repeating backcross (Eicher et al., 1982; Nagamine et al., 1987). The gonadal sex is reversed because the SRY protein encoded on the Y^{TIR} chromosome has polymorphic differences from that encoded on the Y^{B6} chromosome, and fails to sufficiently upregulate its target *Sox9* gene on B6, which is essential for testicular differentiation (Coward et al., 1994; Taketo et al., 2005; Park et al., 2011). All stages of follicles can be seen in prepubertal B6.Y^{TIR} (XY herein) ovaries, but very few MII-oocytes complete the second meiotic division or initiate embryonic development after fertilization (Taketo-Hosotani et al., 1989; Amleh et al., 1996; Villemure et al., 2007; Zhu et al., 2017). The failure in embryonic development can be attributed to cytoplasmic defects; when the XY oocyte nucleus has been transferred into an enucleated XX oocyte, the reconstructed oocyte develops into a healthy pup after *in vitro* fertilization and embryo transfer (Obata et al., 2008). This rescue is more efficient when the oocyte nucleus is replaced at the GV stage than at the MII-stage, suggesting that the XY ooplasm becomes defective largely by the end of growth phase. By 2 months of age, the XY female mouse retains very few or no follicles and fails to initiate estrous cycle (Taketo-Hosotani et al., 1989; Amleh and Taketo, 1998).

In humans, monosomy 45.X (XO) embryos rarely survive *in utero*, and those who have reached the term suffer from congenital abnormalities and infertility, termed Turner's syndrome (Turner, 1938; Singh and Carr, 1966; Ogata and Matsuo, 1995; Hook and Warburton, 2014). In mice, by contrast, most XO embryos survive to term and show no gross anomalies except for lower body weights than their XX littermates (Cattanach, 1962; Burgoyne et al., 1983). These striking somatic differences between the two species can be attributed to the greater number of genes that escape from X chromosome inactivation in humans compared to mice (Berletch et al., 2010; Yang et al., 2010; Tukiainen et al., 2017). Moreover, XO female mice are fertile, suggesting that one X chromosome is sufficient for rendering the oocyte to become competent for embryonic development in the mouse. Nonetheless, XX and XO oocytes are not equal when exogenous genes are expressed (Vernet et al., 2014b; Hamada et al., 2020). In the current study, we produced XO females by using the male mouse carrying *Patchy Fur* (*Paf*) mutation on the X chromosome, which was provided on the C3H/HeSnJ background from the Jackson Laboratory. The *Paf* mutation in males causes a high incidence of X-Y non-disjunction at the first meiotic division and sires XO daughters (Lane and Davisson, 1990; Korobova et al., 1998; Burgoyne and Evans, 2000). In order to compare XO oocytes with XY oocytes on the same genetic background, we backcrossed the *Paf* mutation onto B6 (Vaz et al., 2020). The XO female in this breeding scheme inherits the maternal wild-type X chromosome, like the B6.Y^{TIR} female, and exhibits an early oocyte loss, infertility or subfertility with high embryonic lethality, the features of which are shared with Turner's syndrome in humans (Vaz et al., 2020).

In the current study, we first compared transcriptomes in XX, XO, and XY oocytes of the largest size in which no difference was yet apparent at morphological or global transcriptional activity. We then monitored the changes in transcript levels of selected genes during the entire growth phase. The results show dominant effects of the X chromosome dosage in XO oocytes and a unique transcriptome landscape associated with Y-linked gene transcription in XY oocytes.

MATERIALS AND METHODS

Mice

All animal experiments were performed in accordance with the Canadian Council on Animal Care and approved by the McGill University Animal Care Committee. B6.Y^{TIR} males (N65-76 backcross generations) were crossed with B6 females (Jackson Laboratory, Bar Harbor, ME) to produce XY females and their XX littermates. *Paf* breeding pair on the C3H/HeSnJ background was purchased from the Jackson Laboratory and backcrossed to B6 in our mouse colony. *Paf* carrier males (N6-8 backcross generations) were crossed with B6 females to produce XO females. The day of delivery was defined as 0 days postpartum (dpp). For the use of pups at 18 dpp or younger, their ear punches and tail biopsies were taken for identifying XY and XO females, respectively, 1 day prior to the experiment. For the use of pups at 25 dpp or older, ear punches and tail biopsies were taken upon weaning at 20–25 dpp. Ear punches were digested in a fresh lysis buffer containing 25 mM NaOH and 0.2 mM EDTA disodium (pH 8.0) at 95°C for 40 min, and subject to PCR amplification of the Y-linked *Zfy* gene using the conditions described previously. From tail biopsies, total RNAs were extracted using TRIzol (Invitrogen, Burlington, ON) according to the manufacture's protocol, and dissolved in RNase-free water. The RNA samples were subject to cDNA synthesis using M-MLV Reverse Transcriptase (Invitrogen, Thermo-Fisher Scientific, St Laurent, QC) according to the manufacture's protocol and subsequently to PCR amplification of the *Xist* transcript (Kay et al., 1994), which was present in XX females and absent in XO females.

Collection of Oocytes During the Growth Phase

Ovaries were collected from XX, XO, and XY females at 8, 10, 12, and 18 dpp, and individually dissociated into single cells by the method previously described (Taketo, 2012) with minor modifications. In brief, ovaries were treated with 0.1% collagenase (Sigma-Aldrich, St. Louis, MO) in Eagle's minimum essential medium containing Hank's salts supplemented with 0.25 mM Hepes buffer (both from GIBCO, Long Island, NY) (named MEM-H) for 30 min at 37°C, followed by 0.25% trypsin (Worthington Biochemical, Lakewood, NJ) in Rinaldini solution for 20 min at 37°C. After washings, ovaries were dissociated in phosphate buffer saline (PBS) in microfuge tubes by repeated pipetting, and then centrifuged. The pellet was resuspended in 0.5–2.0 ml (depending on genotypes and age of females) M2 medium, transferred into a Petri dish, and the diameters of oocytes were measured under an inverted microscope (Leica DM IRB, Germany).

Collection of Fully-Grown oocytes

XX, XO, and XY females at 25–29 dpp were injected intraperitoneally with 10 IU equine chronic gonadotropin (Sigma-Aldrich) and euthanized 46 h later. FG (GV-stage) oocytes surrounded with cumulus cells (COC) and spontaneously denuded oocytes (DNO) were collected after puncturing large antral follicles with a pair of 26-gauge needles in the M2 medium supplemented 5.0 μM milrinone (Sigma-Aldrich). The oocytes in COCs were stripped off the cumulus cells by repeating pipetting through a fine glass needle.

Nuclear Transfer in Fully-Grown Oocytes

The oocytes denuded from COCs were incubated in MEM-α supplemented with 5% FBS, 75 μg/ml penicillin G potassium salt and 50 μg/ml streptomycin (all from GIBCO) and 5.0 μM milrinone for 1 h at 37°C with 5% CO_2 in a humidified atmosphere to create the perivitelline space between the zona pellucida and oocyte. The zona pellucida over the perivitelline space was slit with a fine glass knife in M2 containing milrinone under an inverted microscope. The nucleus was removed with a small amount of ooplasm and transferred into another enucleated oocyte using a fine glass pipette using CellTram® vario (Eppendorf, Hamburg, Germany). The recipient oocytes were placed between two gold electrodes, 0.5 mm apart, in a fusion chamber filled with M2 containing milrinone, and electro-pulsed at 55 V for 50 μs in ECFG21 Super Electro Cell Fusion Generator (NEPAGENE, Chiba, Japan). Oocytes were then incubated for fusion in MEM-α supplemented with FBS, antibiotics and milrinone for 1 h. The reconstructed oocytes were cultured in MEM-α supplemented with 250 μM sodium pyruvate (GIBCO), 5% FBS, penicillin and streptomycin, 300 ng/ml FSH (Sigma-Aldrich) and 5.0 μM milrinone for 13–14 h.

Chromatin Configuration

Oocytes collected from COCs, spontaneously denuded oocytes, and oocytes in the growth phase (50–60 μm in diameter) were fixed with 2% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) in microtubule stabilizing buffer (Messinger and Albertini, 1991) at room temperature for 1 h. The oocytes were then washed three times in PBS containing 3% BSA and blocked in PBS containing 5% FBS and 0.1% Triton X-100 overnight at 4°C. After three washes, oocytes were mounted in Prolong Antifade Mounting Medium containing DAPI (Molecular Probe, Eugene, OR) on Plus-coated histology slides (Thermo-Fisher Scientific). Images were captured and examined under a confocal microscope (Zeiss LSM 780, Germany).

Mitochondria Distribution

FG-oocytes and oocytes in the growth phase (40–50 µm in diameter) were stained with 200 and 400 nM, respectively, Mito-Tracker Orange CMTMRos (Molecular Probes, Themo-Fisher Sientific) at 37°C for 30 min with 5% CO_2 in a humidified atmosphere, followed by fixation, washing and blocking as described above. After three washes, the oocytes were mounted in Prolong Antifade Mounting Medium containing DAPI on Plus-coated histology slides. Images were captured and examined under a confocal microscope. Fluorescence intensity of Mito-Tracker Orange CMTMRos was measured by ZEN software (Carl ZEISS MicroImaging).

De Novo Transcription

The *de novo* transcriptional activity in FG-oocytes and oocytes in the growth phase (50–70 µm in diameter) was detected using Click-iT™ RNA Alexa Fluor™ 488 Imaging kit (Invitrogen, Thermo-Fisher Scientific). In brief, oocytes were incubated in 1 mM 5-ethynyl uridine (EU) for 1 h, followed by fixation, washing and blocking as described above. After three washes, the oocytes were fluorescence stained according to the manufacture's protocols and mounted in Prolong Antifade Mounting Medium containing DAPI on Plus-coated histology slides. Images were captured and analysed under a confocal microscope. EU fluorescence intensity in the nucleus was measured in individual oocytes by ZEN software. For FG-oocytes, the mean intensity of the non-surrounded-nucleolus (NSN)-type XX oocytes was set as 1.0 and the relative intensity in each oocyte was calculated in every experiment. For the oocytes in the growth phase, the mean intensity of the XX oocytes of 60–65 µm in diameter was set as 1.0 for calculating the relative intensity in all oocytes in every experiment.

Immunofluorescence Staining of H3K4me3 in Oocytes

FG-oocytes and oocytes in the growth phase (50–70 µm in diameter) were fixed as described above and incubated with rabbit monoclonal anti tri-methyl-histone H3 (Lys4) antibody (#9751, Cell Signaling Technology, New England Biolabs, Whitby, ON) (1:200) at 4°C overnight. After three washes, oocytes were incubated with goat-anti-rabbit IgG-FITC (Jackson ImmunoResearch, West Grove, PA) (1:500) at room temperature for 1 h. After three washes, the oocytes were mounted in Prolong Antifade Mounting Medium containing DAPI on Plus-coated histology slides. Images were captured and examined under a confocal microscope. Fluorescence intensity was measured in individual oocytes by ZEN software. For the oocytes in the growth phase, the mean intensity of the XX oocytes of 60–70 µm in diameter was set as 1.0 for calculating the relative intensity in all oocytes in every experiment. For FG-oocytes, the mean fluorescence intensity in the surrounded-nucleolus (SN)-type XX oocytes was set as 1.0.

RNA Preparation and Sequencing

In order to perform transcriptomic analysis of small numbers of oocytes, we applied the Smart-seq3 method (Hagemann-Jensen et al., 2020) with some modifications as follows. A total of 30 oocytes (10 oocytes of 50–55 µm and 20 oocytes of 55–60 µm in diameter) in 5 µL volume per sample were pooled into 10 µL of lysis buffer [4.54 µM polyT primer, 0.128% Triton-X 100, 3 mM each dNTP, 1 U/µl RNase inhibitor, 4.7 attomoles (diluted by 220,000 fold) of "ERCC spike-in Mix 1" (ThermoFisher #4456740)] in biological triplicate for each genotype. The diluted stock of "ERCC spike-in Mix 1" should correspond to 3–6% of total reads sequenced per sample of 30 oocytes based on theoretical calculations. Lysis was carried out in a thermocycler at 72°C for 3 min, 4°C for 10 min, and 25°C for 1 min, and the lysed samples were stored at −80°C. cDNA synthesis and library preparation was performed with an in-house Smart-seq3 protocol using modifications for first and second strand cDNA synthesis as follows: 1) oligo-dTVN: /5Me-isodC/AGATGTGTATAAGAGACAGN(12)ACT(30)VN: 2) Template Switching Oligo (TSO): /5Me-isodC//iisodG//iMe-isodC/AGATGTGTATAAGAGACAGN(12)ACGCrGrGrG and Invitrogen SuperScript IV Reverse Transcriptase (Thermo-Fisher Scientific): 3) The Nextera PCR primer TCGTCGGCAGCGTCA GATGTGTATAAGAGACAG was used for single primer cDNA amplification, using Advantage 2 Polymerase Mix (Takara). NGS libraries were generated as described (Hagemann-Jensen et al., 2020) and quality control was performed using electropherogram profiling on a Caliper HS DNA. Chip Sequencing was performed on an Illumina NovaSeq 6000 SP lane in Paired-end 150 bp mode.

Data Processing and Bioinformatic Analysis

The primary reads were aligned to the GRCm38 mouse genome and transcriptome from Gencode, using Hisat2, in stranded and paired-end mode (Leek and Storey, 2007; Zhang et al., 2019). Reads from the 5′ exons were demultiplexed and UMIs counted, using regular expressions according to the 5′ end sequencing patterns on an Apache Spark cluster (Zaharia et al., 2016). The read counts, corresponding to 5′ UMIs, were normalized using ERCC spike-ins (Ritchie et al., 2015). Quality control was performed using principal component analysis and heatmap analysis of the top 500 highly and differentially expressed genes, using a distance matrix of the Spearman correlation coefficient. Batch effects were assessed using Surrogate Variable Analysis (*sva*, R bioconductor package) (Leek and Storey, 2007), and included as covariates into the linear model using the R Bioconductor *limma* package for differential expression (Ritchie et al., 2015). We used the *lv2Transformer*, an improved version of the default log-voom transformation, for *limma* provided by the *countTransformers* R package (Zhang et al., 2019). For all other purposes including clustering, batch effects were regressed-out using *cleaningY* function of the *jaffelab* R package (Collado-Torres et al. https://github.com/LieberInstitute/jaffelab, R package version 0.99.31).

All expressed genes present in the normalized expression matrix were used for scatterplots in three-way comparisons among XX, XO, and XY oocytes. Differentially expressed genes (DEGs) were selected in three-way comparisons at $p < 0.05$ and their overlapping was analysed in Venn diagrams as follows. Directional over or under expressed DEGs were obtained from the three-way comparisons, totaling 6 DEG lists. According to the focus on gain or loss by the X chromosome or the Y chromosome, we selected 4 comparison groups out of the 6 DEG lists. A schematic diagram is presented with arrows pointing towards the directional over expression in each figure. Venn diagrams were constructed by intersection of 4 DEG lists.

For Y-linked gene analyses, p-values were adjusted to Rate False Discover Rate (FDR) using Benjamini-Hochberg (https://www.statisticshowto.com/benjamini-hochberg-procedure/).

For the X-linked DEGs identified as X chromosome dosage dependent, the ratio of transcript levels in XO and XY oocytes against that in XX oocytes was calculated using a lognormal distribution, fitted using R-package *EnvStats*, giving the mean, standard deviation and confidence internals. The ratio of all X-linked vs. autosomal gene transcript levels was calculated using bootstrap analysis framework from *boot* R-package as follows. Gene expression was separated into 6 equally sized bins, and genes were assigned to each category. Sampling with replacement, with a size of 100, repeated for 5,000 times, was performed from the X-linked genes and autosomal genes separately. The median for each extracted sample was calculated, and then a ratio of the X:A transcript levels was calculated. Its distribution was represented as boxplot, for each expression level bin and genotype. A similar procedure was applied for calculating the ratio of mean transcript levels in XO and XY oocytes against that in XX oocytes, using only the highly expressed genes in bin 6. Comparison of X, Y and autosomal homologous gene transcript levels was also performed using the R statistical framework, based on the geometric mean of the replicate data (Olivier et al., 2008), having lognormal standard deviation intervals represented.

qRT-PCR

Total RNA was extracted from 10–30 pooled oocytes of each genotype and size range with RNeasy RNA isolation kit (Qiagen, Montreal, QC), and subject to cDNA synthesis using oligo(dT) (Life Technologies, Thermo-Fisher Scientific) in total 20 μL. One μl of each cDNA solution was subject to qPCR of a gene in duplicate using a FastStart Essential DNA Green Master with LightCycler® 96 Instrument (Roche, Mannheim, Germany). Primers used for qPCR are listed in **Supplementary Table S1**. Endogenous *Ppia* levels were used for normalization in each cDNA sample and qRT-PCR procedure.

Statistical Analyses

All experiments except for RNA-Seq were repeated at least three times. When proportions of oocytes in different categories were compared between two genotypes, χ^2-test was used. Where values were normally distributed, data of three genotypes were presented as means ± SEM and statistically analyzed by two-sided t-test or one-way ANOVA followed by Tukey's honestly significant difference (HSD) test. Where values did not fit into normal distribution, data were presented in a 25–75% percent box plot with the median in line, and statistical difference was evaluated by the Dunn's multiple comparison test.

RESULTS

Chromatin Configuration

We first compared the chromatin configuration in XX, XO, and XY oocytes with DAPI staining. As previously reported (Debey et al., 1993; Zuccotti et al., 1995), we observed decondensed chromatin configuration in the nuclei of most XX oocytes in the mid growth phase (50–60 μm), named non-surrounded nucleolus (NSN), and progressive condensation of chromatin around the nucleolus, named partially surrounded nucleolus (PSN) to surrounded nucleolus (SN), by the end of growth phase (**Figure 1A**). When the XX FG-oocytes collected from COCs were examined, 95.7% showed SN-type chromatin configuration while the rest showed PSN- and NSN-type chromatin configuration (**Figure 1B**). However, when the spontaneously denuded FG-oocytes (DNO) were examined, only 44.3% were of SN-type while 24.3 and 31.4% were of PSN- and NSN-type, respectively. The frequencies of oocytes with the three types of chromatin configuration were significantly different between the oocytes of the two origins at $p < 0.001$ by χ^2-test. Similarly, 91.8% of the XO FG-oocytes collected from COCs were of SN-type while only 44.4% of the spontaneously denuded XO FG-oocytes were SN-type with a significant difference at $p < 0.001$. By contrast, although most XY oocytes in the growth phase were seen with NSN-type chromatin configuration, comparable with XX and XO oocytes, none of the XY FG-oocytes showed typical PSN- or SN-type chromatin configuration. Instead, 83.6 and 4.9% of the XY oocytes collected from COCs showed chromatin condensation along the nuclear envelop (NE) in addition to SN- and PSN-type condensation, respectively. NE-type chromatin condensation was never seen in XX or XO oocytes. The remaining 11.5% of the XY oocytes from COCs showed NSN-type chromatin configuration. Spontaneously denuded XY FG-oocytes also showed NE-SN- and NE-PSN-type chromatin configuration in addition NSN-type chromatin configuration, but neither typical SN- nor PSN-type. These results suggest that the chromatin condensation was not affected by the lack of one X chromosome in XO oocytes while it became abnormal in XY oocytes at the end of growth phase.

Chromatin configuration can be determined within the nucleus or affected by cytoplasmic components. To distinguish these two possibilities, we performed nuclear transfer between XX and XY FG-oocytes collected from COCs. In controls, the nucleus of an XX oocyte was transferred into an enucleated XX oocyte, referred as XX(N)-XX(C), and the nucleus of an XY oocyte was transferred into an enucleated XY oocyte, as XY(N)-XY(C). All control reconstructed oocytes showed SN- and NE-SN-type chromatin configuration as expected for XX and XY oocytes without manipulation (**Figures 1C,D**). When the nucleus of an XY oocyte was transferred into an enucleated XX oocyte, presented as XY(N)-XX(C), 92.1% of the reconstructed oocytes showed NE-SN-type nuclei and were not statistically different from control XY(N)-XY(C) oocytes.

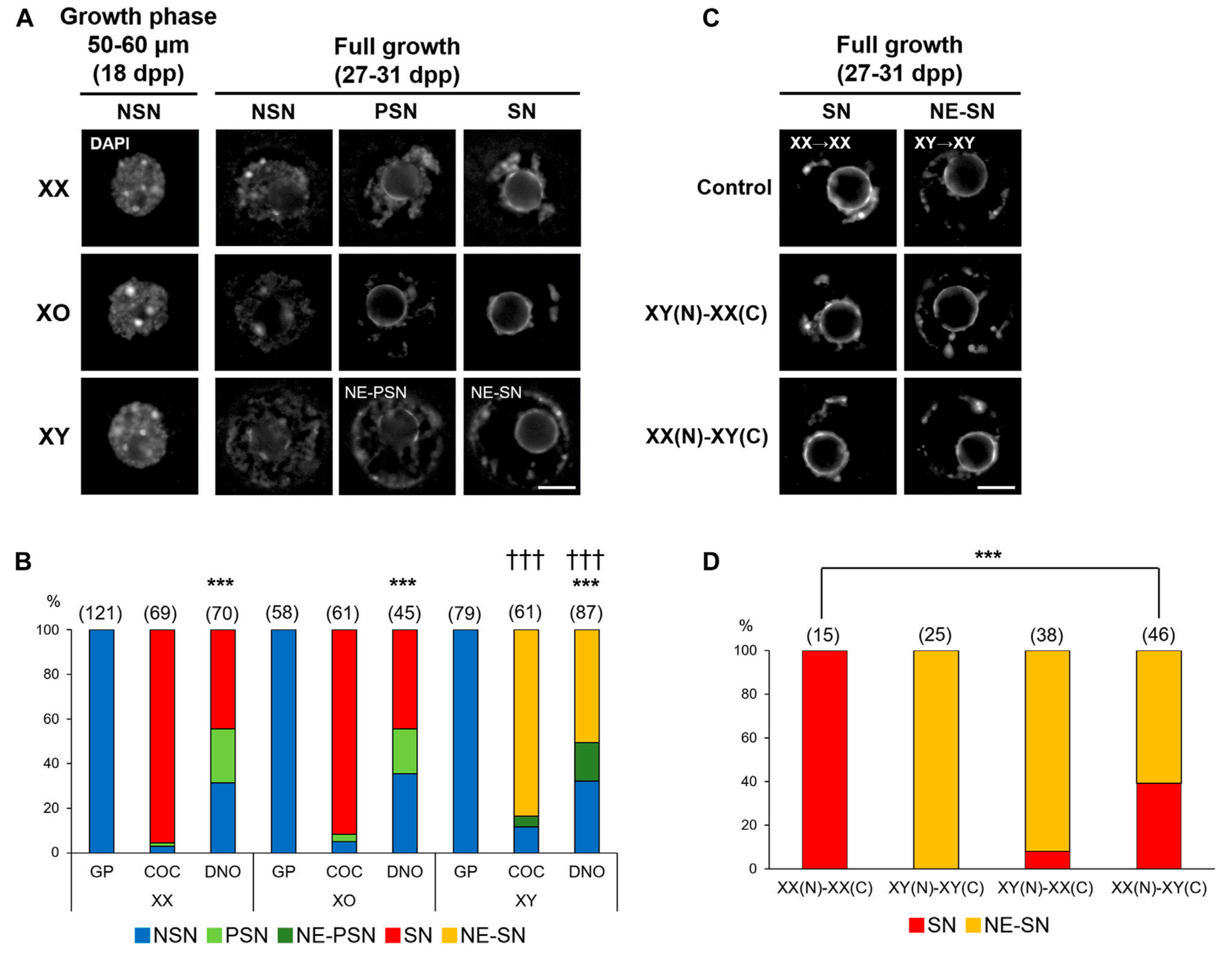


FIGURE 1 | Chromatin configuration in XX, XO, and XY oocytes in the growth phase. **(A)** Representative images of chromatin configuration with DAPI staining. XX and XO oocytes showed three types; non-surrounded nucleolus (NSN), partially surrounded nucleolus (PSN), and surrounded nucleolus (SN), whereas XY oocytes showed chromatin condensation along the nuclear envelop (NE) in addition to PSN or SN-type condensation. Scale bar, 10 µm. **(B)** Percentages of oocytes with different types of chromatin configuration. Oocytes examined were either in the growth phase (GP), at full growth collected from cumulus-oocyte complexes (COC), or spontaneously denuded at the collection of COCs (DNO). The total number of oocytes examined is given above each column. *** and ††† indicate statistical differences compared with the oocytes from COCs of the same genotype and with XX and XO oocytes in the same category, respectively, at $p < 0.001$ by X^2-test. **(C)** Representative images of chromatin configuration in the nucleus (N) after transfer into an enucleated oocyte cytoplasm (C). Scale bar, 10 µm. **(D)** Percentages of oocytes with different types of chromatin configuration in the reconstructed oocytes. The total number of oocytes examined is given above each column. *** indicates statistical difference from the control oocytes, XX(N)-XX(C), at $p < 0.001$ by X^2-test.

However, when the nucleus of an XX oocyte was transferred into an enucleated XY oocyte, referred as XX(N)-XY(C), 60.9% of the reconstructed oocytes showed NE-SN-type chromatin configuration with a significant difference from control XX(N)-XX(C) oocytes at $p < 0.001$. These results suggest that the cytoplasm of XY oocytes was mainly responsible for the abnormal chromatin condensation along the nuclear envelop. However, the chromatin configuration in the XY nucleus appears to have been irreversibly altered by the end of growth phase and could not be corrected by brief contact with the XX cytoplasm. Therefore, we cannot exclude the possibility that the XY nucleus per se also contributed to the abnormal chromatin configuration.

Distribution of Mitochondria With Active Membrane Potential

Chromatin configuration is known to be affected by cytoplasmic components (Inoue et al., 2008) such as mitochondria, which can be easily visualized. Accordingly, we examined the distribution of mitochondria by staining with MitoTracker Orange, which indicates the high mitochondrial membrane potential. As shown in **Figures 2A,C**, metabolically active mitochondria were concentrated in an area near the nucleus of 60–80% oocytes of 40–50 µm, regardless of the genotype. When the nucleus was positioned off the center, mitochondria were concentrated in the wider side of cytoplasm. With further

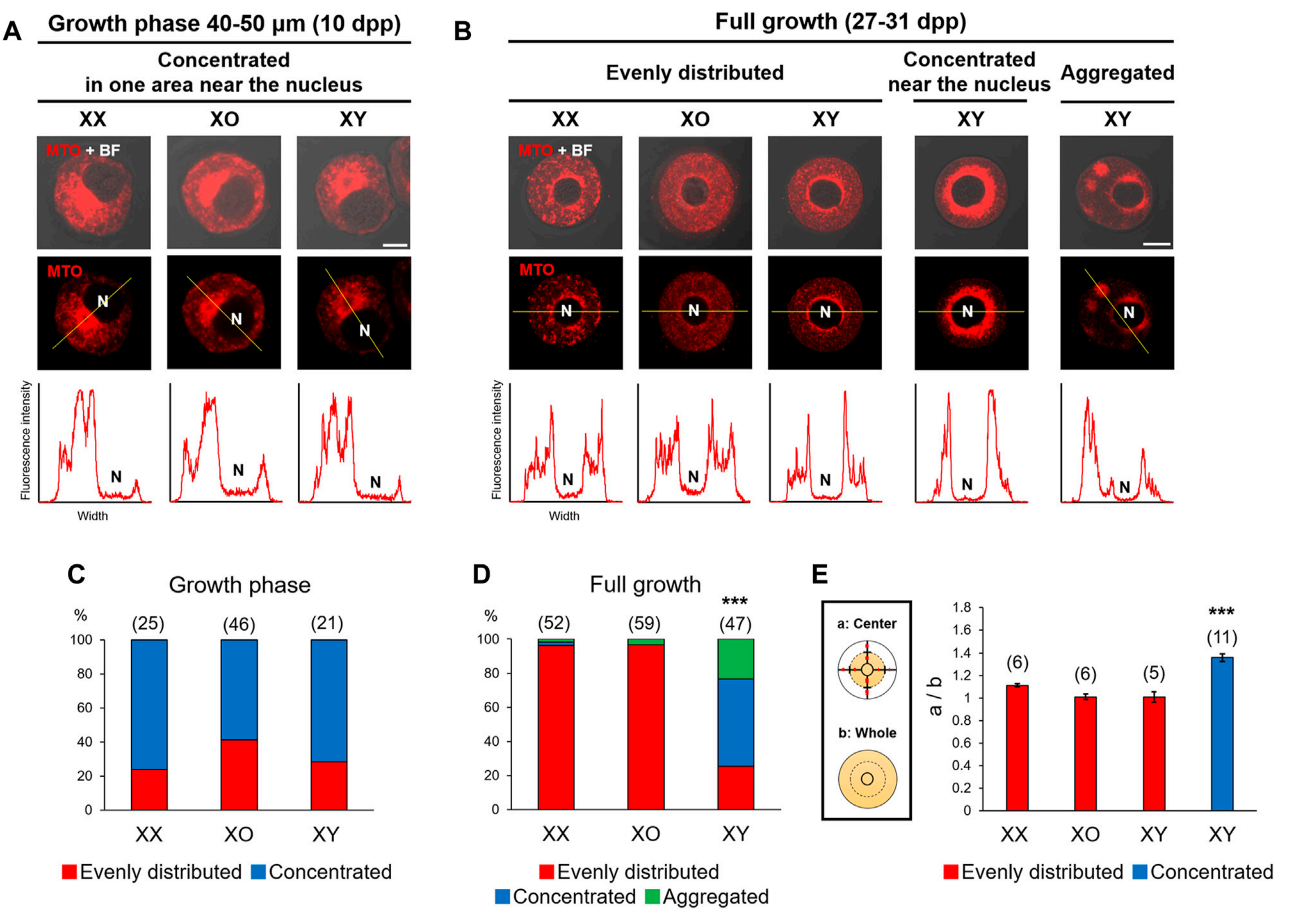


FIGURE 2 | Distribution of mitochondria with high membrane potential in XX, XO, and XY oocytes in the growth phase. **(A,B)** Representative images of mitochondria staining with MitoTracker Orange (MTO). The image taken with phase contrast in the bright field (BF) is overlayed on the top panel. The intensity of fluorescence signals along the yellow line drawn in the picture is shown beneath each picture. N, nucleus. Scale bar, 10 μm. **(A)** XX and XY oocytes of 40–50 μm. **(B)** XX, XO, and XY FG-oocytes. In the majority of XX and XO oocytes, MTO staining was evenly distributed in the cytoplasm except for the narrow area around the nucleus. In XY oocytes, MTO staining was more concentrated near the nucleus or aggregated in the cytoplasm. Scale bar, 20 μm. **(C)** Percentages of XX and XY oocytes in the growth phase with different types of mitochondria distribution. The total number of oocytes examined is given above each column. **(D)** Percentages of XX, XO, and XY FG-oocytes with different types of mitochondria distribution. *** indicates statistical difference from XX or XO oocytes at $p < 0.001$ by X^2-test. **(E)** Ratios of the total intensity of fluorescence signals in the central cytoplasmic area to that in the whole area. Each oocyte was divided by the half distance between the nuclear envelop and the ooplasmic membrane. Each column indicates the mean ± SEM. The number of oocytes examined is given above each column. *** indicates statistical difference in the XY oocyte with concentrated mitochondria, compared to XX, XO, and XY oocytes with evenly distributed mitochondria at $p < 0.001$ by one-way ANOVA followed by Tukey's honestly significant difference (HSD) test.

oocyte growth, mitochondria became evenly distributed in the cytoplasm of most XX and XO oocytes while the nucleus relocated to the centre (**Figures 2B,D**). By contrast, only 25.5% of XY FG-oocytes showed evenly distributed mitochondria while 51.1% showed highly concentrated mitochondria around the nucleus. The relative intensity of MTO fluorescence in the perinuclear area of these XY oocytes was significantly higher than in the oocytes which were categorized as "evenly distributed" (**Figure 2E**). The remaining 23.4% of XY oocytes showed mitochondria in a few large aggregates in the cytoplasm (**Figures 2B,D**). We also stained some oocytes with MitoTracker Green, which was independent of membrane potential, and found similar mitochondria distribution (not shown). Thus, mitochondria distribution was comparable in XX, XO, and XY oocytes during the growth phase and became abnormal in the XY oocytes near the end of growth phase.

De Novo Transcription

The chromatin configuration, NSN or SN, in XX FG-oocytes is associated with transcriptional silencing and competence for embryonic development (Bouniol-Baly et al., 1999; de la Fuente and Eppig, 2001). In our results, XY FG-oocytes showed abnormal chromatin condensation along the nuclear envelope (NE), distinct from typical NSN- or SN-type. Therefore, we examined *de novo* transcription in these oocytes by EU incorporation (**Figure 3A**). De novo transcription was quiescent in XX and XO FG-oocytes with SN-type chromatin configuration as well as in XY FG-oocytes with NE-SN-type chromatin configuration. By contrast, EU

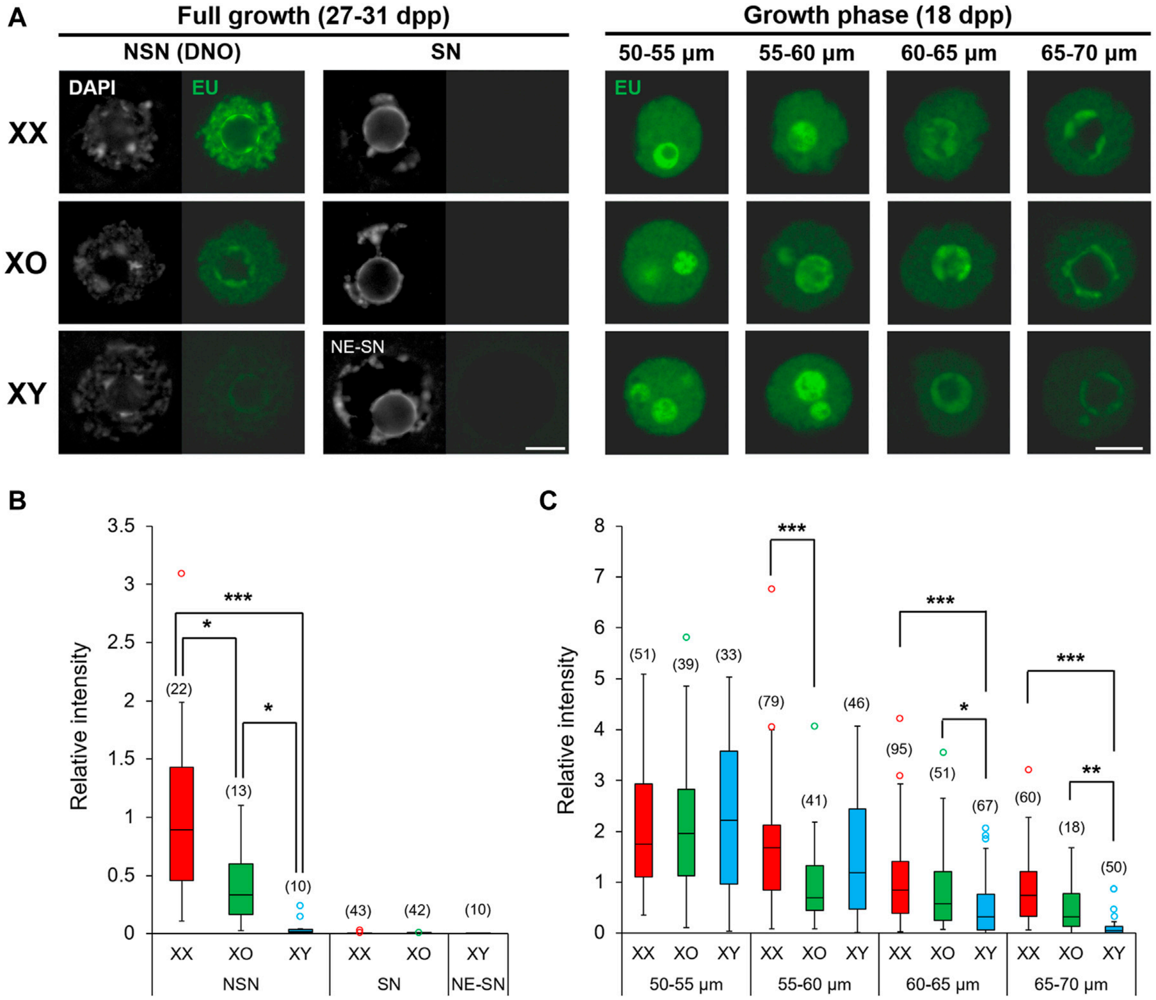


FIGURE 3 | *De novo* transcription in XX, XO, and XY oocytes in the growth phase. **(A)** Representative images of chromatin configuration (DAPI) with NSN, SN or NE-SN-type and EU incorporation (EU) in XX, XO, and XY FG-oocytes. Scale bar, 10 µm. **(B,C)** The relative fluorescence intensity of EU incorporation in the fully grown oocytes with NSN, SN or NE-SN-type chromatin configuration **(B)** or during growth **(C)**. Each box plot indicates the median with 1st and 3rd quartiles. The thin vertical line indicates the range from minimum to maximum values. ° indicates outlier. The total number of oocytes examined is given above each column. *, **, and *** indicate statistical differences at $p < 0.05$, 0.01 and 0.001, respectively, by Dunn's multiple comparison test.

staining intensity was diverse among the oocytes of three genotypes with NSN-type chromatin configuration; very high in XX oocytes, very low in XY oocytes, and intermediate in XO oocytes. We found large variability of fluorescence signal intensity among the oocytes of the same genotype, which often did not fit into the normal distribution. Therefore, we presented the results by the median and interquartile range (IQR) with minimum and maximum values, and statistically analysed the data by the Dunn's multiple comparison test (**Figure 3B**). The results indicated significant differences in all three-way comparisons.

We then examined the oocytes during the growth phase and measured the relative intensity of EU staining (See Methods). The results showed that EU staining intensity was comparable among XX, XO and XY oocytes of 50–55 µm, the smallest size examined (**Figure 3C**). The EU staining intensity then gradually declined in XX oocytes with further growth to low but still detectable levels at 65–70 µm in diameter. For comparison, EU staining intensity significantly declined in XO oocytes of 55–60 µm compared to XX or XY oocytes of the same size and remained lower than XX oocytes although without significant difference afterwards. The EU staining intensity in XY oocytes was initially comparable with that in XX oocytes up to 55–60 µm, but it declined rapidly to become significantly lower than XX or XO oocytes at 60–65 µm. Thus, *de novo* transcription declined early in XO oocytes, while it declined slightly later but further in XY oocytes approaching the end of growth phase, compared to XX oocytes. Thus, both the absence of the second X chromosome and the presence of the Y chromosome affected the global transcription during oocyte growth.

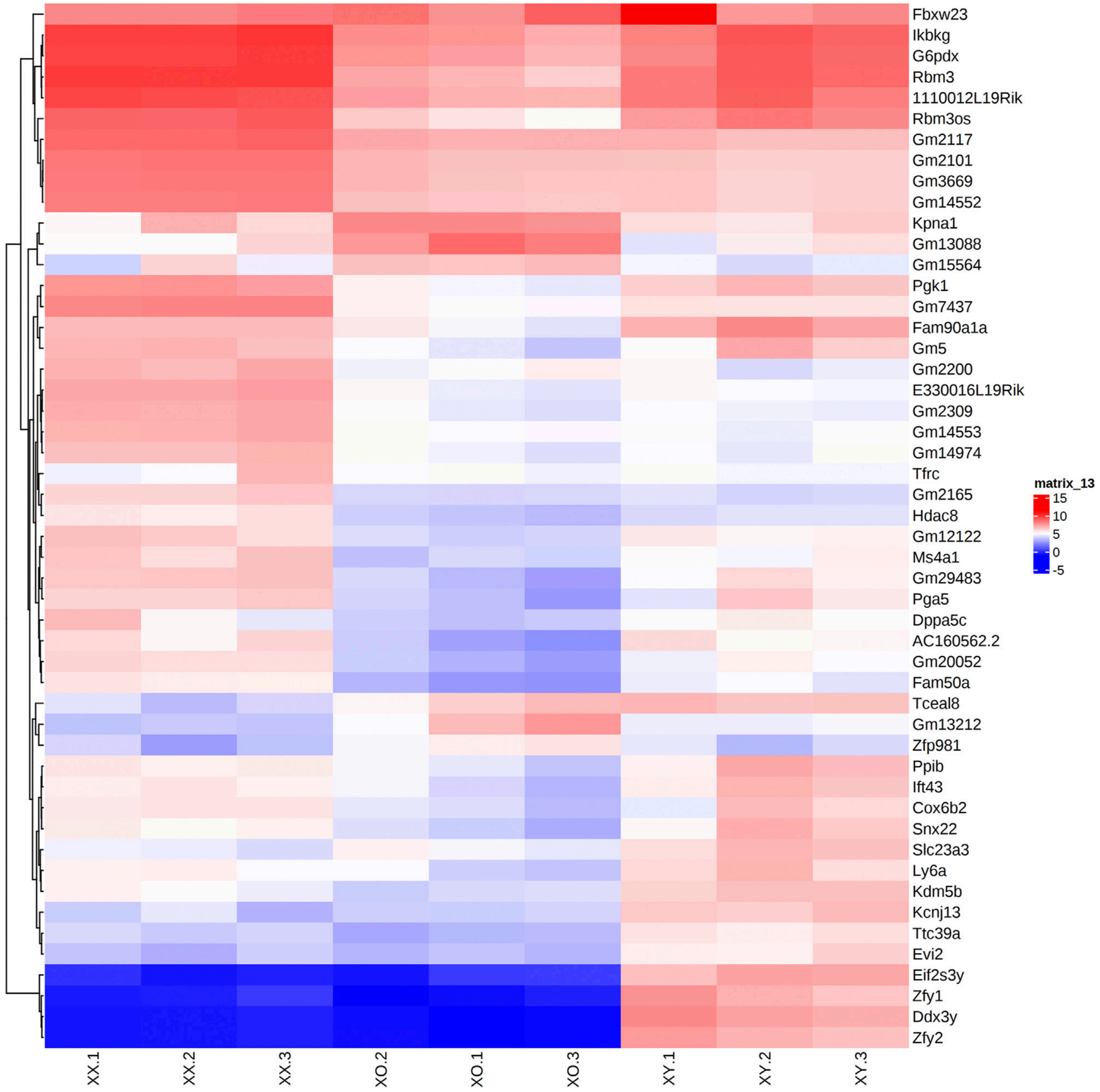
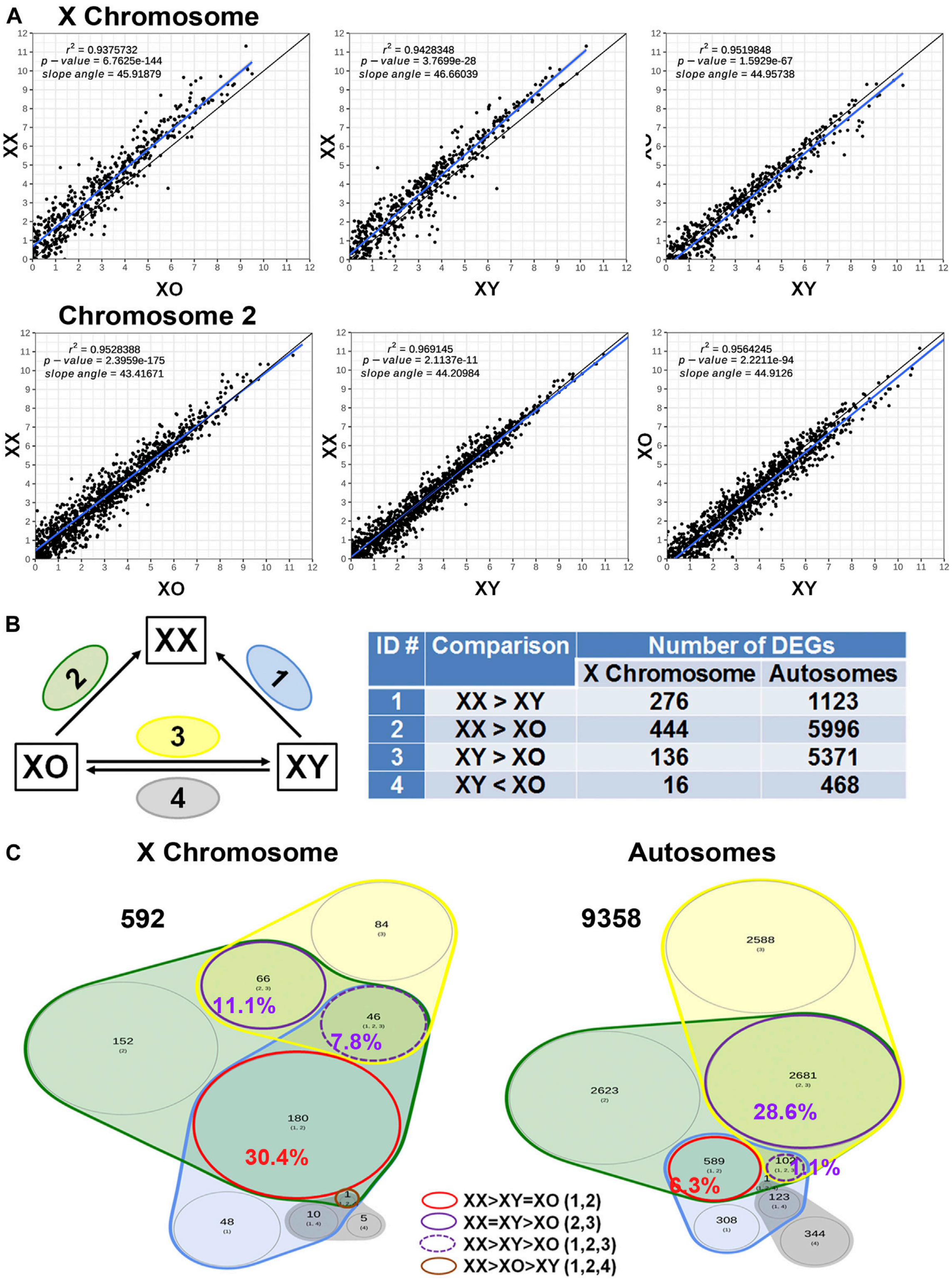


FIGURE 4 | Heatmap (Log2CPM) of the top 50 highly and differentially expressed genes in biological triplicates of XX, XO, and XY oocytes, based on the distance matrix of the Spearman correlation coefficient.

Differentially Expressed Genes in the Oocytes by RNA-Sequencing

To elucidate the role of sex chromosome complement in establishing the transcriptome in the oocytes during the growth phase, we analyzed differentially expressed genes (DEGs) by RNA-Sequencing (RNA-Seq) among XX, XO, and XY oocytes of 50–60 μm, all of which were still transcriptionally active (See above). To reduce the bias, 10 oocytes of 50–55 μm and 20 oocytes of 55–60 μm were pooled in each sample. The sequencing data, as well as full length coverage tracks, raw 5′ UMI, and normalized expression matrices reported in this study have been deposited in the Gene Expression Omnibus website with accession code GSE184153.

Across all three genotypes, a total of 28,051 genes were detected, while sample quality was interrogated using principal component analysis (not shown) and heatmap analysis of the 500 highly expressed and most variable genes, using a distance matrix of the Spearman correlation coefficient (**Figure 4**). The results



ID #	Comparison	Number of DEGs	
		X Chromosome	Autosomes
1	XX > XY	276	1123
2	XX > XO	444	5996
3	XY > XO	136	5371
4	XY < XO	16	468

FIGURE 5 | Sex chromosome dosage dependent differentially expressed genes (DEGs). **(A)** Scatterplots of all expressed genes on the X chromosome and Chromosome 2 in XX vs. XO, XX vs. XY, and XO vs. XY oocytes. **(B)** The total number of DEGs ($p < 0.05$) in four comparison groups with focus on gain by the second X chromosome. **(C)** Venn diagrams to indicate the overlapping of DEGs among four comparison groups. The area is proportional to the number of DEGs. Transcript levels of the genes in a red circle were higher in XX oocytes than those in XO or XY oocytes without difference between XO and XY oocytes. The genes in a solid purple circle were equally higher in XX and XY oocytes than XO oocytes. The genes in a broken purple circle were higher in the order of XX, XY, and XO oocytes. The single gene (*Bmp15*) in a brown circle was lower in XY oocytes than in XX or XO oocytes.

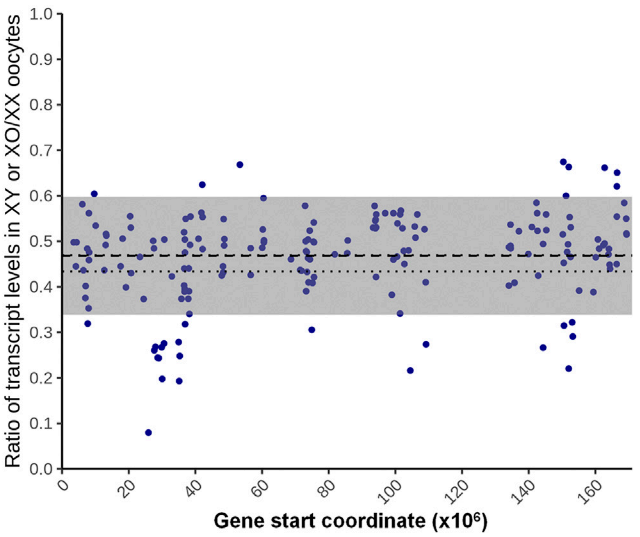


FIGURE 6 | X chromosome dosage dependent differentially expressed genes (DEGs). The ratio of the average transcript levels of individual DEGs in XO and XY oocytes against those in XX oocytes. The X-axis indicates the distance from the proximal end of the X chromosome. The thick broken line indicates the geometrical mean, thin broken lines indicate the confidence of mean, and gray area indicates the mean ± SD.

showed the highest consistency among the biological replicates of the same genotype although XX1, XO2, and XY1 showed less similarity to their other two replicates. Transcript levels (5′UMI reads) of all X-linked genes were scatter-plotted in three-way comparisons, XX vs. XO, XX vs. XY, and XO vs. XY oocytes (**Figure 5A**), all of which showed strong coefficiency ($r^2 > 0.937$). We separated X-linked and autosomal genes because of the X chromosome dosage difference among the three genotypes. The vast majority of X-linked genes were in fact shifted upward from the 1:1 line in XX vs. XO and XX vs. XY oocytes whereas most genes were scattered around the 1:1 line in XO vs. XY oocytes. For comparison, most expressed genes from Chromosome 2 were scattered around the 1:1 line in all three-way comparisons. To reveal the X chromosome dosage effects, we selected DEGs at $p < 0.05$ in four-way comparison groups, 1) XX > XY, 2) XX > XO, 3) XY > XO, and 4) XY < XO, as given in **Figure 5B**. The total number of X-linked DEGs was 276, 444, 136, and 16, respectively, whereas the total number of autosomal DEGs was 1,123, 5,996, 5,371, and 468, respectively. Overlapping of these DEGs among four comparison groups was shown in Venn diagrams (**Figure 5C**), where the area is proportional to the number of DEGs.

X Chromosome Dosage Dependent DEGs

We first asked how the X chromosome dosage reflected into X-linked gene transcript levels. Of the total 592 DEGs found in all four comparison groups, 180 (30.4%) DEGs had higher transcript levels in XX oocytes than in XO or XY oocytes without significant difference between XO and XY oocytes (red circle). These genes represent the DEGs which were predominant affected by the X chromosome dosage. Other genes may also be X chromosome dosage-dependent, but they were affected by other components such as the Y chromosome. For comparison, of the total 9,358 autosomal DEGs, only 589 (6.3%) showed the same relationship, confirming that transcript levels of the 180 X-linked genes represented the X chromosome dosage dependent DEGs. Examples are *Atrx*, which encodes a chromatin remodeling factor (de la Fuente et al., 2004b; Balboula et al., 2015), *Eif2s3x*, a translation initiation factor, and *Pdha1*, pyruvate dehydrogenase E1.

For the 180 genes identified as X chromosome dosage dependent X-linked DEGs, we used a fitted lognormal distribution of the data, and calculated the ratio of geometric mean transcript levels in XO and XY oocytes against those in XX oocytes. We then plotted the ratios at their loci on the X chromosome, which clustered within 0.47 ± 0.13 (mean ± SD) (**Figure 6**). We also calculated the ratio of transcript levels in XO or XY oocytes separately against those in XX oocytes, which were comparable at 0.44 ± 0.14 and 0.48 ± 0.14, respectively. These results indicate that the X-linked DEGs in this category were transcribed from each X chromosome, independent of the presence of another X chromosome or the Y chromosome.

The X chromosome dosage may negatively affect the transcript levels of other genes. A reverse relationship between the X chromosome dosage and transcript levels was found when DEGs were analysed by overlapping among XX < XO, XX < XY, XO > XY, or XO < XY comparison groups with the focus on the loss by the second X chromosome (**Supplementary Figure S1**). 10 (4.3%) out of total 232 X-linked DEGs and 283 (4.2%) out of 6,801 autosomal DEGs showed lower transcript levels in XX oocytes than XO or XY oocytes without difference between XO and XY oocytes. Examples are *Prps2*, phosphoribosyl pyrophosphate synthetase, and *Tceal8*, transcription elongation factor A. Since no bias towards X-linked DEGs is seen in this category, these results may suggest that the transcript levels of X chromosome dosage dependent DEGs negatively affected other genes including some X-linked genes.

X-Linked Gene Dosage Compensation

X-linked gene expression in somatic cells is known to be evolutionally adjusted by two mechanisms; 1) the transcript levels are upregulated to match those by autosomal genes in XY cells (gene dosage compensation) and 2) one of the two X chromosomes is inactivated in XX cells to match XY cells (X inactivation) (Nguyen and Disteche, 2006; Deng et al., 2014). Previous studies have reported that the primordial germ cells that have arrived at gonads follow these principles, but the oocytes violate them starting with the X chromosome reactivation prior to the onset of meiosis (Fukuda et al., 2015; Sangrithi et al., 2017). Since the oocytes during the growth phase have not yet been tested for the X-linked gene dosage compensation, we analysed our normalized RNA-Seq data of all genes with detectable transcript levels excluding Y-linked genes. Due to a large variation of reads, we divided all genes evenly into 6 bins according to their transcript levels and calculated the X:A ratio in each bin (**Supplementary Figure S2A**). The results showed that the X:A ratio was 1.0 in all XX, XO, and XY oocytes when transcript levels were low (Bins 1–3). However, when transcript

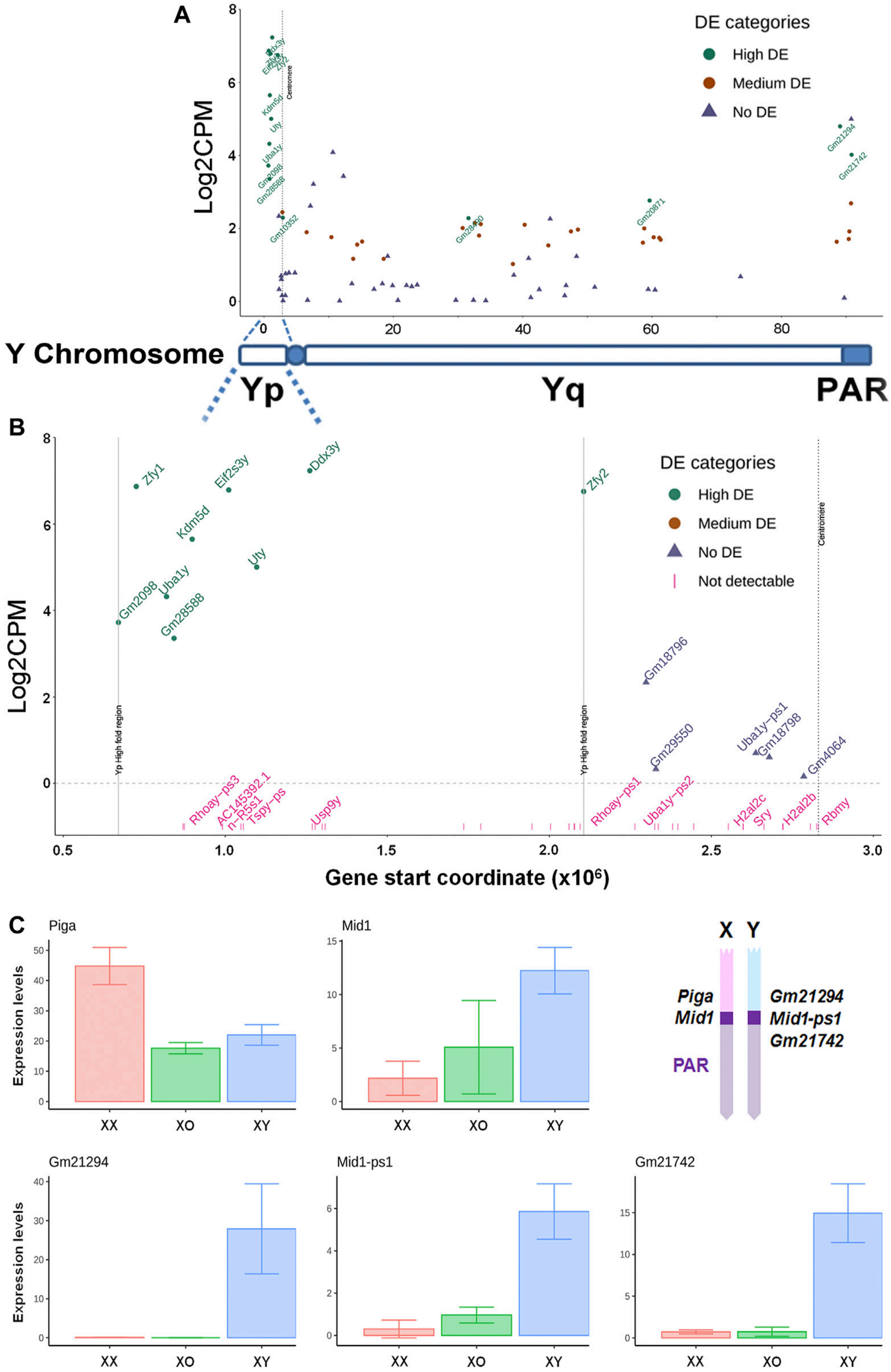


FIGURE 7 | Transcript levels of Y-linked genes in XY oocytes, compared to XX oocytes. **(A)** The entire Y chromosome. High DE, Log2FC ≥ 3. Medium DE, 1 < Log2FC < 3, No DE, Log2FC ≤ 1. **(B)** The Y short arm alone. The X axis indicates the distance from the distal end of Yp. The genes indicated in pink are undetectable. **(C)** Transcript levels of X and Y homologous DEGs near the boundary of pseudoautosomal region (PAR). Geometrical mean ± SD. The diagram **(top, right)** indicates the approximate positions of examined genes on the X and Y chromosomes.

levels were the highest (Bin 6), the X:A ratio increased to 1.5 in XX oocytes while it remained near 1.1 in both XO and XY oocytes. Similar trend was found to a lesser extent for the genes with the intermediate transcript levels (Bins 4 and 5). Thus, X-linked gene transcript levels were adjusted during oocyte growth in two distinct manners. For a half of genes, which were transcribed at low levels, their transcript levels were upregulated to match those of autosomal genes in XO and XY oocytes (typical X-linked gene dosage compensation) while they were suppressed to match those of autosomal genes in XX oocytes by a mechanism other than X inactivation. By contrast, for another half of X-linked genes, which were transcribed at higher levels, their transcript levels corresponded to the X chromosome dosage.

Similarly to **Figure 6**, we calculated the ratio of the transcript levels of individual X-linked genes in Bin 6 by dividing the mean levels in XO and XY oocytes by those in XX oocytes (Supplemental Figure S2B). The ratios distributed within 0.58 ± 0.27 (mean ± SD) along the entire X chromosome. The mean increased by 0.1 and SD was doubled compared to the ratio distribution in **Figure 6** where only DEGs were analysed. These results indicate that highly transcribed X-linked genes generally follow the X chromosome dosage dependence.

Y-Linked-Gene Transcription

Our XY female mouse model provides a unique opportunity for examining whether and how Y-linked genes are transcribed outside the male germline. From 1,570 reads of Y-linked genes in our data, 39 genes were detected in XY oocytes (**Figures 7A,B**) while very few reads were detectable near threshold levels in XO or XX oocytes. On the short arm, the nine genes with high transcript levels (Log2FC ≥ 3.0) were clustered within a narrow region except for *Zfy2*. On the long arm, 21 repetitive-sequence-genes, which are not yet well defined, were transcribed at medium levels (1 < Log2FC ≤ 3.0). By contrast, well known genes such as *Sry*, *Usp9y*, and *Rbmy* on the short arm and *Sly* and *Ssty1/2* on the long arm were undetectable in XY oocytes, suggesting that these genes may have been actively repressed. Of the 13 highly transcribed genes, *Gm21294* is proximal to the boundary of pseudoautosomal regions (PAR) on the Y long arm and its transcript levels were detectable only in XY oocytes as expected (**Figure 7C**). The PAR boundary is located in the intron 3–4 of *Mid1* (*Trim18*) on the X chromosome while its truncated form *Mid1-ps1* as well as *Gm21742* are located within the PAR on the Y chromosome (Palmer et al., 1997; Perry et al., 2001; Lu et al., 2013). In our results, transcript levels of both *Mid1-ps1* and *Gm21742* were much higher in XY oocytes than in XX or XO oocytes (**Figure 7C**). Transcript levels of X-linked *Mid1* were also higher in XY oocytes than in XX or XO oocytes although the differences were more moderate. These results were unexpected as genes in the PAR should behave like autosomal genes. By contrast, *Piga*, which is proximal to the boundary of the PAR on the X chromosome, was transcribed at higher levels in XX oocytes than XO or XY oocytes, representing the X chromosome dosage dependent X-linked DEGs. Thus, Y-linked genes at the boundary of and within the PAR were more actively transcribed in XY oocytes than in XX or XO oocytes during the growth phase.

Sex Chromosome Dosage Dependent Differentially Expressed Genes

We next asked whether the XO oocyte has the disadvantage of missing one sex chromosome compared to the XX or XY oocyte which carries two sex chromosomes. Indeed, in the Venn diagram with focus on the gain by the second X chromosome (**Figure 5C**), 2,681 (28.6%) out of all 9,358 autosomal DEGs showed comparable transcript levels between XX and XY oocytes, but significantly lower in XO oocytes (solid purple circle). In addition, 102 (1.1%) autosomal genes were lower in XY oocytes than in XX oocytes, but still higher than in XO oocytes (broken purple circle). These results may suggest that homologous genes on X and Y chromosomes share similar effects to maintain the transcript levels of many autosomal genes.

Of the highly expressed Y-linked genes, *Zfy1/2* is known to have distinct biological activities from their X homolog *Zfx* (Luoh et al., 1997; Shpargel et al., 2012). However, *Ddx3y* shares RNA helicase activity and may be exchangeable with its X homolog *Ddx3x* (Sekiguchi et al., 2004; Matsumura et al., 2019). *Eif2s3y* and *Eif2s3x* also share redundant functions in the proliferation of spermatogonia although they exhibit distinct activities when overexpressed in the ES-derived female germline (Yamauchi et al., 2016; Hamada et al., 2020). *Uty* encodes a protein, which can partially compensate for the embryonic lethality by null mutation of its X-homolog *Utx* (*Kdm6a*) although UTY has a lower histone demethylase activity than UTX (Shpargel et al., 2012; Welstead et al., 2012). These homologs may have contributed to the regulation of sex chromosome dosage dependent autosomal DEGs. To explore this possibility, we compared the transcript levels of X- and Y-homologs, and autosomal homologs if available, in XX, XO and XY oocytes (**Figure 8**). The results show a consistent trend that transcript levels of X-homologs were comparable between XO and XY oocytes and twice higher in XX oocytes, fitting into the X chromosome dosage dependent X-linked DEGs category. Furthermore, the transcript levels of Y-homologs were comparable with those of X homologs in XY oocytes, making the sum comparable between XX and XY oocytes (**Figure 8A**). Thus, simple comparison of transcript levels supports our hypothesis that the Y chromosome may compensate for the deficiency of one X chromosome in XY oocytes. One exceptional case is *Kdm5* with lysine-specific demethylase activity, which has an X-homolog (*Kdm5c*), a Y-homolog (*Kdm5d*) and two autosomal homologs (*Kdm5a* and *Kdm5b*). While transcript levels of *Kdm5a* were consistent in the oocytes of all three genotypes, those of *Kdm5b* were much higher in XY oocytes, bringing the sum of all homologs to the highest, compared to XX or XO oocytes. Another exception is the *Zfx/y* pair, where the transcript levels of Y-homologs *Zfy1/2* surpassed those of X-homolog *Zfx*.

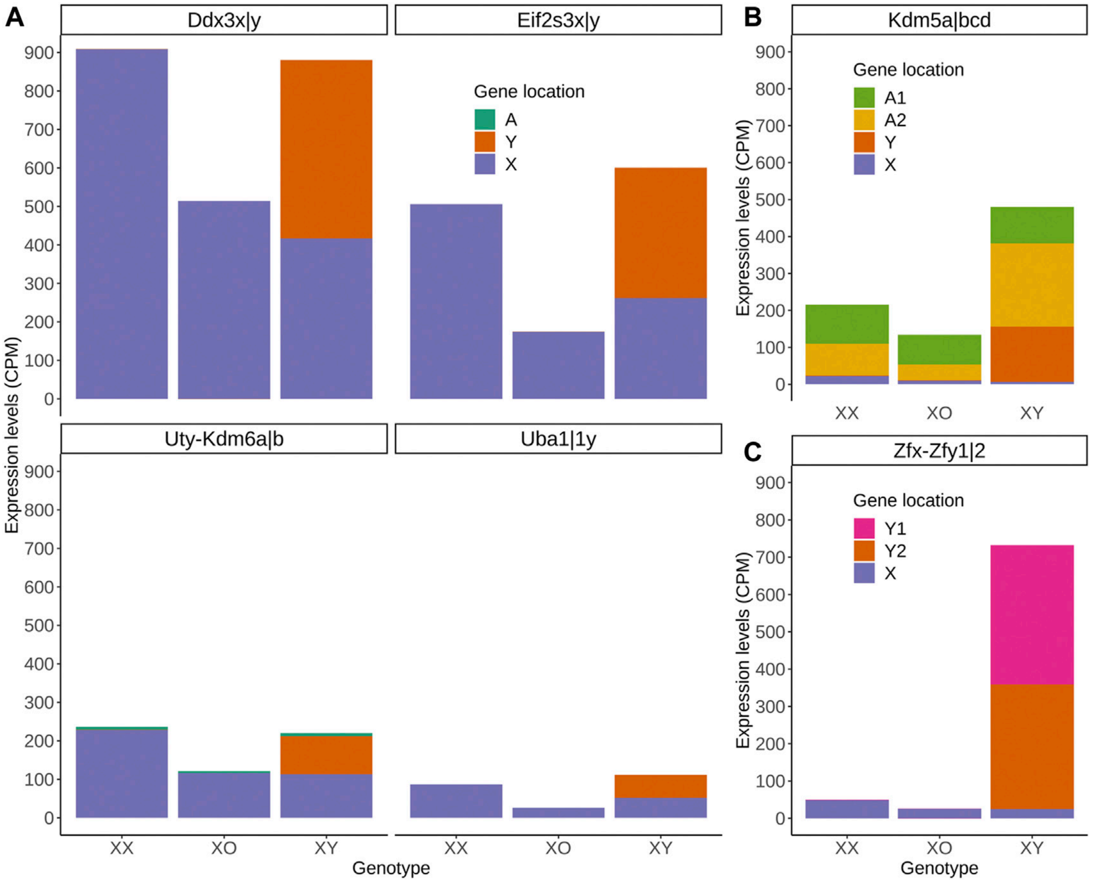


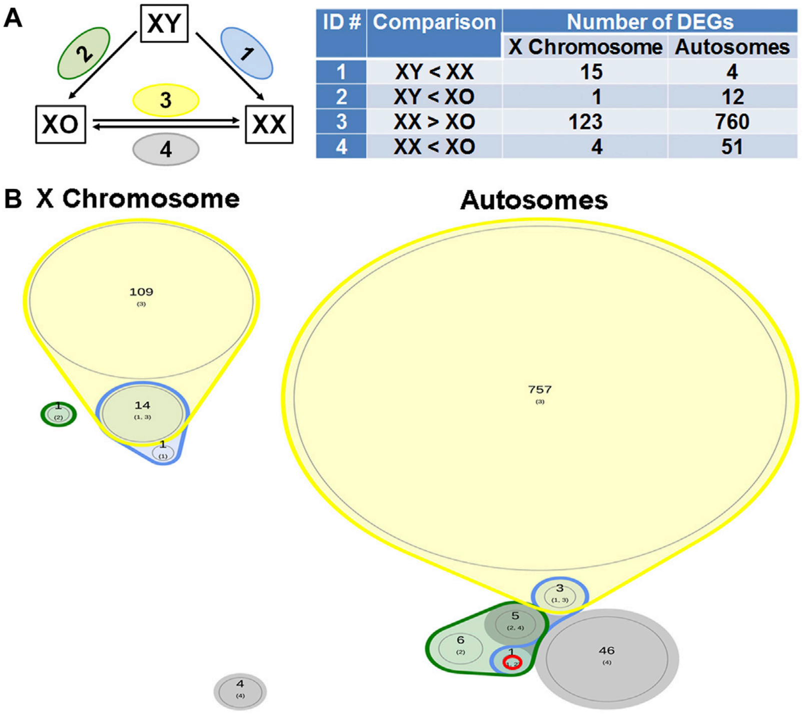
FIGURE 8 | Transcript levels of X-, Y-, and autosomal homologs in XX, XO, and XY oocytes in the growth phase. **(A)** X- and Y-homologs with (*Kdm6*) or without (others) an autosomal homolog. **(B)** X- and Y-homologs (*Kdm5c* and *Kdm5d*, respectively) and two autosomal homologs (A1 = *Kdm5a*, A2 = *Kdm5b*). **(C)** X- (*Zfx*) and two Y-homologs (Y1 = *Zfy1*, Y2 = *Zfy2*).

Differentially Expressed Genes in XY Oocytes Compared to XX and XO Oocytes

If any gene is responsible for the cytoplasmic defects in XY oocytes, such gene must be differentially expressed compared to both XX and XO oocytes. In the Venn diagram with focus on the loss by the second Y chromosome (**Figure 9**), none out of 129 X-linked DEGs and only one gene, *Larp6*, out of 818 autosomal DEGs showed lower transcript levels in XY oocytes than in both XX and XO oocytes. *Larp6* encodes a RNA-binding protein with diverse functions including tRNA processing, non-coding RNA metabolism, and ribosomal biogenesis (Maraia et al., 2017) By interrogating the Venn diagram in **Figure 5C**, the transcript levels of X-linked *Bmp15* also turned out to be the lowest in XY oocytes among the three genotypes. *Bmp15* encodes an oocyte secretary factor that promotes follicular growth and ovulation (Yan et al., 2001; Su et al., 2004; Sugiura et al., 2007). In the reverse relationship, 2 genes, *Tnmd* and *Gm15726*, out of 131 X-linked DEGs and 9 out of 1,028 autosomal DEGs were found at higher levels in XY oocytes than both XX and XO oocytes (**Figure 10**). X-linked *Tnmd* encodes a type II transmembrane protein, which has potent anti-angiogenic activity (Shukunami et al., 2005). An autosomal gene *Kcnj13* encodes an inwardly rectifying potassium channel (KIR7.1), conserved in vertebrates, and KCNJ13-regulated membrane potential modulates actin organization in tracheal smooth muscle cells (Yin et al., 2018).

Changes in Gene Transcript Levels in XX, XO, and XY Oocytes During Follicular Growth

To determine when DEGs became differential among XX, XO, and XY oocytes during follicular growth, we compared the transcript levels of selected genes in the oocytes during the entire growth phase as well as at the end of growth phase (FG-oocytes) using quantitative RT-PCR (qRT-PCR). Since oocytes of different sizes were collected from ovaries at different ages for technical necessity, we compared the transcript levels in the oocytes of 40–50 μm at 8 and 12 dpp and those of 50–60 μm at 12 and 18 dpp, but the results were consistent (not shown) and combined.



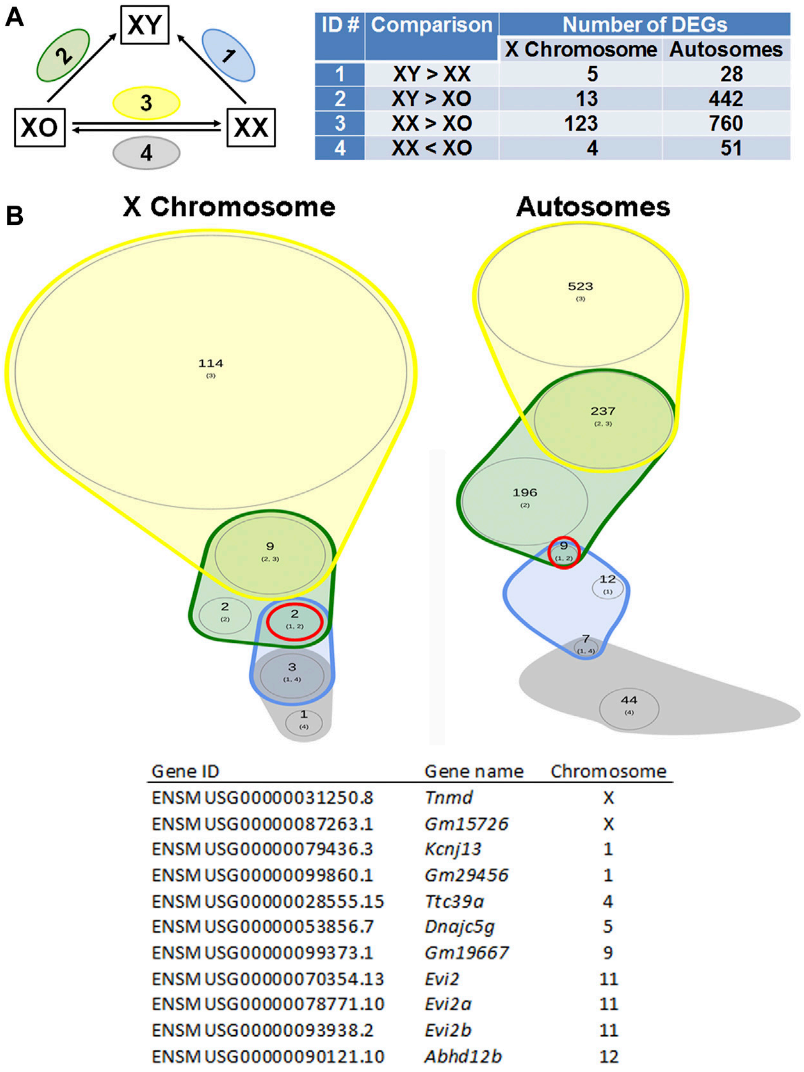
ID #	Comparison	Number of DEGs	
		X Chromosome	Autosomes
1	XY < XX	15	4
2	XY < XO	1	12
3	XX > XO	123	760
4	XX < XO	4	51

FIGURE 9 | DEGs lower in XY oocytes than in XX or XO oocytes. **(A)** The total number of DEGs ($p < 0.05$, Log2FC > 1) in four comparison groups with focus on loss by the Y chromosome. **(B)** Venn plots to indicate the overlapping of DEGs among four comparison groups. Only one autosomal gene, *Larp6* on Chromosome 9 (ENSMUSG00000034839.5), was found in this category.

We first determined Cq values of *Ppia* and *Tubb5*, selected from most consistently highly expressed genes among XX, XO, and XY oocytes of 50–60 µm in the RNA-Seq data, using cDNA aliquots from pooled XY oocytes in biological triplicates (**Supplementary Figure S3**). Based on relative consistency in the Cq values throughout the oocyte growth, we chose *Ppia* for normalization of Cq values of other genes in each cDNA sample and qPCR amplification.

We then determined the transcript levels of three Y-linked genes, *Ddx3y*, *Uba1y*, and *Zfy1/2*, in XY oocytes (**Supplementary Figure S4**). *Ddx3y* encodes a protein with characteristics of Dead-box RNA helicase, which is dispensable for male fertility (Matsumura et al., 2019). *Uba1y* (also named *Ube1y*) encodes a protein with homology to a ubiquitin activating enzyme although its function remains unknown (Mazeyrat et al., 2001). *Zfy1* and *Zfy2* encode transcription factors possessing zinc finger domains and an acidic domain (Koopman et al., 1991; Decarpentrie et al., 2012). Because of the highly homologous sequences, we could not design primers to differentially amplify *Zfy1* and *Zfy2*, and the results of qRT-PCR indicate the sum of *Zfy1* and *Zfy2* transcript levels. Our qRT-PCR results showed that the transcript levels of *Ddx3y* and *Uba1y* were low in the oocytes of 20–30 µm, gradually increased with oocyte growth, and peaked in FG-oocytes. The transcript levels of *Zfy1/2* increased slightly in the oocytes from 40 to 60 µm but remained at low levels throughout the growth phase. This is the only case that did not agree with the results of RNA-Seq, in which the transcript levels of both *Zfy1* and *Zfy2* were as high as those of *Ddx3y* and *Uba1y*. Quantitation of transcript levels by RNA-Seq should be more accurate since qRT-PCR is limited by the design of primers.

We selected *Eif2s3x* and *Atrx* to represent X chromosome dosage dependent X-linked DEGs and *Bmp15*, whose transcript levels were further decreased in XY oocytes than XO oocytes, in our RNA-Seq data. *Eif2s3x* encodes a eukaryotic translation initiation factor (Ehrmann et al., 1998) while *Atrx* encodes a helicase of the SWI/SNF2 family of chromatin remodeling proteins and its transcripts are enriched in pericentromeric heterochromatins in pachytene oocytes as well as metaphase spermatogonia (Baumann et al., 2008; Levy et al., 2015; Lovejoy et al., 2020). Our qRT-PCR results showed that the transcript levels of *Eif2s3x* were stable in the oocytes from 20 to 60 µm and then drastically increased with further oocyte growth. The transcript levels of *Bmp15* and *Atrx* were very low in the oocytes of 30–40 µm and gradually increased with further oocyte growth and peaked in FG-oocytes. Transcript levels of all three genes were almost always significantly lower in XY oocytes than in XX oocytes all through the growth phase. Transcript levels in XO oocytes were also lower than those in XX oocytes although significant difference was not always found. At the end of growth phase, transcript levels of both *Bmp15* and *Atrx* in XY oocytes were lower than those in XO oocytes and only 13 and 21% of their respective levels in XX oocytes.



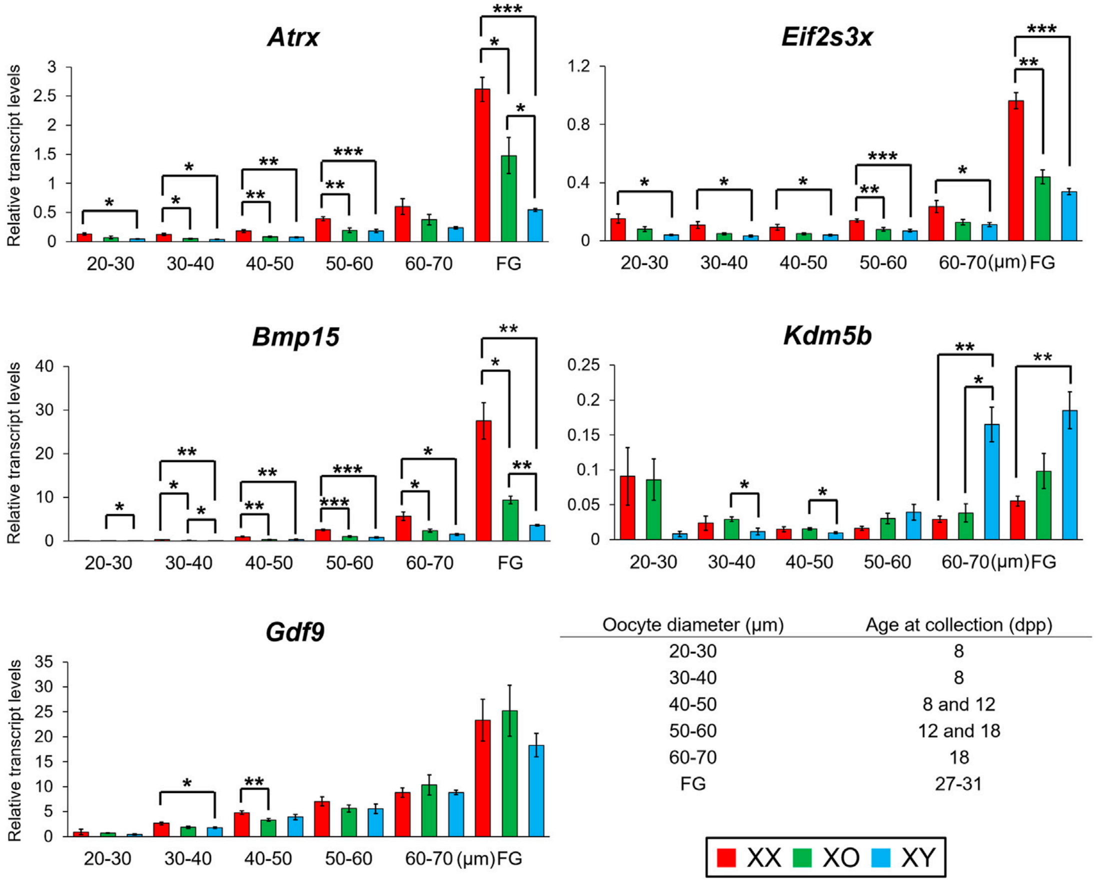
ID #	Comparison	Number of DEGs	
		X Chromosome	Autosomes
1	XY > XX	5	28
2	XY > XO	13	442
3	XX > XO	123	760
4	XX < XO	4	51

Gene ID	Gene name	Chromosome
ENSMUSG00000031250.8	*Tnmd*	X
ENSMUSG00000087263.1	*Gm15726*	X
ENSMUSG00000079436.3	*Kcnj13*	1
ENSMUSG00000099860.1	*Gm29456*	1
ENSMUSG00000028555.15	*Ttc39a*	4
ENSMUSG00000053856.7	*Dnajc5g*	5
ENSMUSG00000099373.1	*Gm19667*	9
ENSMUSG00000070354.13	*Evi2*	11
ENSMUSG00000078771.10	*Evi2a*	11
ENSMUSG00000093938.2	*Evi2b*	11
ENSMUSG00000090121.10	*Abhd12b*	12

FIGURE 10 | DEGs higher in XY oocytes than in XX or XO oocytes. **(A)** The total number of DEGs ($p < 0.05$, Log2FC > 1) in four comparison groups with focus on gain by the Y chromosome. **(B)** Venn plots to indicate the overlapping of DEGs among four comparison groups. Two X-linked genes and nine autosomal genes were found in this category as listed in the table.

We chose *Kdm5b* as an autosomal gene whose transcript levels were higher in XY oocytes than XX or XO oocytes in the RNA-Seq data. *Kdm5b* (*Jarid1b*) encodes a histone demethylase of H3K4 involved in gene transcriptional activation (Zhang et al., 2016). *Kdm5b* is particularly interesting as it was the dominant transcripts among *Kdm5* homologs in the oocytes of growth phase, and specifically affected by the XY chromosomal complement (**Figure 8**). Five alternative splicing isoforms including up to 27 exons have been reported for *Kdm5b* (ENSMUSG00000042207.17) (**Supplementary Figure S5A**). We designed three sets of primers to cover the junctions of Exons 1–2, 13–14, and 26–27. Only the amplicon of Exons 26–27 was detectable in FG-oocytes while all sets of primers yielded amplicons of expected sizes in the total RNA from ovary and testis (**Supplementary Figures S5B,C**). In fact, the entire RNA-Seq data of *Kdm5b* showed consistent transcript levels corresponding only to Exons 23–27 in the oocytes of 50–60 μm (**Supplementary Figures S5B**). Our qRT-PCR results revealed that *Kdm5b* transcript levels were detectable at moderate levels in XX and XO oocytes while barely detectable in XY oocytes of 20–30 μm (**Figure 11**). *Kdm5b* transcript levels then decreased in both XX and XO oocytes and remained low in the oocytes of all genotypes until the levels in XY oocytes drastically increased to be significantly higher than those in XX or XO oocytes of 60–70 μm. No more difference was found between XO and XY FG-oocytes as the transcript levels in XY oocytes did not increase as much as in XO oocytes at the end of growth phase.

We selected *Gdf9* as a gene whose transcript levels were not different among XX, XO, and XY oocytes of 50–60 μm in the RNA-Seq data. *Gdf9* encodes a TGFβ superfamily member, known to be secreted by the oocyte to facilitate follicular



Oocyte diameter (μm)	Age at collection (dpp)
20-30	8
30-40	8
40-50	8 and 12
50-60	12 and 18
60-70	18
FG	27-31

🟥 XX 🟩 XO 🟦 XY

FIGURE 11 | Changes in transcript levels of individual genes in XX, XO and XY oocytes during follicular growth. Transcript levels of X-linked and autosomal genes normalized to *Ppia* transcript levels. *, **, and *** indicate significance at $p < 0.05$, 0.01, and 0.001, respectively, among XX, XO, and XY oocytes of the same diameter by *t*-test.

growth beyond the primary follicle stage (Dong et al., 1996). Our qRT-PCR results (**Figure 11**) show that *Gdf9* transcript levels increased with oocyte growth to peak in FG-oocytes. Minor differences were recognized in the oocytes of small sizes, but no more difference was found among XX, XO, and XY oocytes of 50 μm or larger, in consistent with the RNA-Seq results.

Enrichment of H3K4me3 in the Nuclei of XX and XY Oocytes

Oocytes undergo dynamic epigenetic modifications of histones as well as *de novo* DNA methylation during the growth phase (Lawrence et al., 2016; Gahurova et al., 2017; Stäubli and Peters, 2021). Methylation of specific lysine on histone H3 within a gene domain is associated with gene transcription, chromatin configuration and ultimately the competence of oocytes for embryonic development. For example, methyltransferase SETD2 is crucial for H3K36 trimethylation (H3K36me3) and consequent *de novo* DNA methylation while suppressing ectopic H3K4me3 and H3K27me3 during the oocyte growth (Xu et al., 2019). H3K4me3 in the gene promoter is generally associated with active transcription while H3K27me3 is associated with transcription repression and H3K9me3 causes constitutive heterochromatinization and gene silencing (Puschendorf et al., 2008). Furthermore, EHMT2 (G9A/GLP) is responsible for H3K9me2 affecting chromatin organization in oocytes (Yeung et al., 2019). The methylation status of each site is set on by the balancing act between lysine-specific methyltransferases and demethylases. We have already shown that *Kdm5b* encoding H3K4me3 demethylase was highly expressed in XY oocytes compared to XO or XX oocytes (**Figures 8**, **11**). We interrogated other genes known to be involved in histone modifications, but none was found among the DEGs in our RNA-Seq data. To examine whether the higher transcript levels of *Kdm5b* in XY oocytes reflected into histone modifications, we immunofluorescence-stained for H3K4me3 in XX and XY oocytes of 50–70 μm and at the FG-stage (**Figure 12**). The results showed that H3K4me3 was enriched in patches over the nuclei of XX and XY oocytes in the growth phase (**Figure 12A**). H3K4me3 staining intensity on average in XY oocytes was significantly lower than that in XX oocytes of 50–60

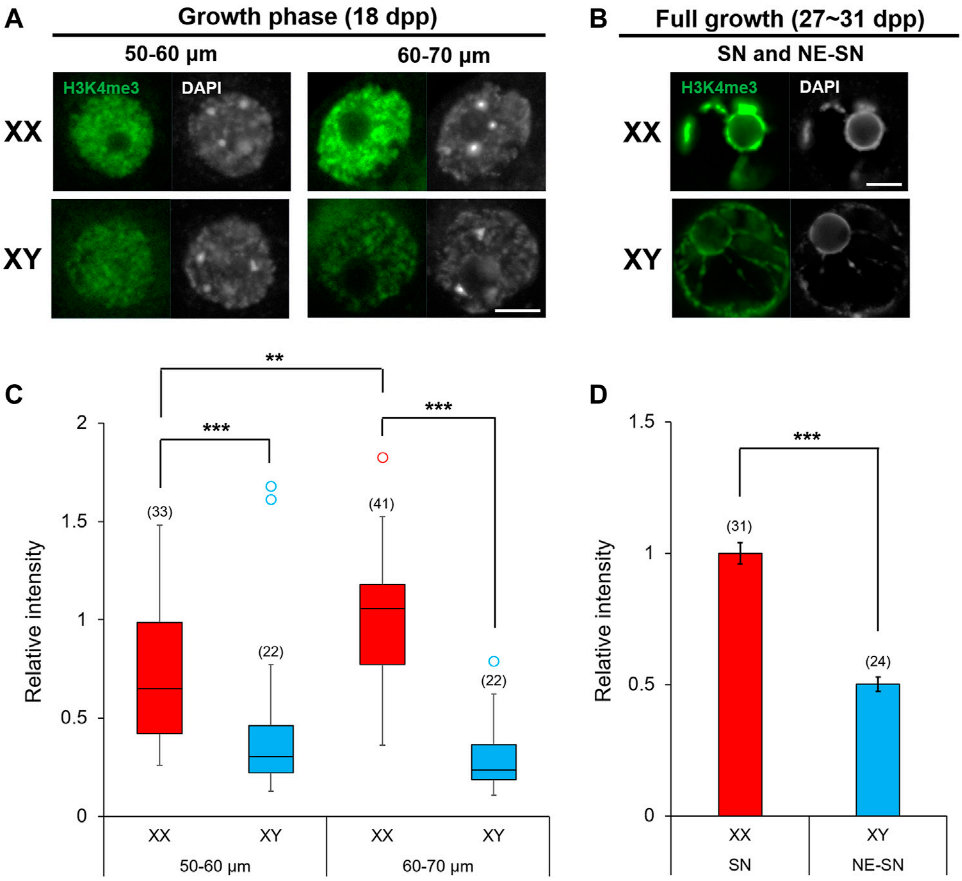


FIGURE 12 | Enrichment of H3K4me3 in XX and XY oocytes in the growth phase. **(A,B)** H3K4me3 immunoflourescence staining in the nuclei of oocytes in the late growth phase and at full growth with DAPI counterstaining. Scale bar, 10 µm. **(C)** Relative intensity of H3K4me3 staining in the oocytes of growth phase. Each box plot indicates the median with 1st and 3rd quartiles. The thin vertical line indicates the range from minimum to maximum values. ° indicates outlier. The total number of oocytes examined is given above each column. ** and *** indicate statistical differences at $p < 0.01$ and 0.001, respectively, by Mann-Whitney U test. **(D)** Relative intensity of H3K4me3 staining in FG-oocytes. Each column indicates the mean ± SEM. The number of oocytes examined is given above each column. *** indicates statistical difference at $p > 0.001$, by Welch's *t*-test.

or 60–70 µm. Furthermore, H3K4me3 staining intensity increased in XX oocytes from 50–60 to 60–70 µm while it failed to do so in XY oocytes (**Figure 12C**). In both XX and XY FG-oocytes, strong H3K4me3 signals were observed along the condensed chromatin around the nucleolus (**Figure 12B**). However, the total H3K4me3 signals over the nucleus was significantly lower in XY FG-oocytes than in XX FG-oocyte (**Figure 12D**). These results indicate that the higher transcript levels of *Kdm5b* were associated with lower H3K4me3 accumulation in the XY oocyte nucleus in the late growth phase.

DISCUSSION

An oocyte accumulates cytoplasmic components such as mRNAs and proteins during follicular growth to support the subsequent meiotic progression, fertilization, and early embryonic development in the absence of *de novo* transcription. However, how gene transcription is regulated during oocyte growth is not well understood. Our current study focused on the transcript levels of X- and Y-linked genes and their association with the genome-wide transcriptome in oocytes. The XY oocyte became abnormal in chromatin configuration, mitochondria distribution and the global *de novo* transcription near the end of growth phase. Therefore, we compared transcriptome among XX, XO and XY oocytes of 50–60 µm, where no difference was yet apparent in their morphology or *de novo* transcription activity. Our results revealed that the X chromosome dosage exerted dominant effects on the transcriptome in XO oocytes, whereas the presence of the Y chromosome or transcripts of Y-linked genes made the transcriptome in XY oocytes closer to that in XX oocytes. Nonetheless, XY oocytes established a distinct transcriptome at the end of growth phase.

RNA-Sequencing Methods Used in This Study

To accommodate the limited quantity of RNAs from 30 oocytes per sample, we used Smart-seq3 protocol (Hagemann-Jensen et al., 2020). Version 2 of the protocol was already known to work well for single cells for a standard amount of 2 pg total RNA (Picelli et al., 2014). The new version of the protocol features a higher specificity and sensitivity by using a 5′UMI. The standard oligo-dTVN and Template Switching Oligo (TSO) primers were modified, enlarging from 8 to 12N UMIs, which accommodated a relatively larger than single-cell library complexity of our small bulk samples. Besides the existing 5′UMI which was used for our present data analysis, we introduced a 3′UMI in order to be able to keep compatibility with other single-cell quantification methods like 10xGenomics or our previous Nanopore data (Oikonomopoulos et al., 2016). The 3′ end can also be sequenced by using same Nextera Tn5 binding site. Sequencing of both ends allowed for accurate PCR primer design of maximally spanning isoforms, being able to match both transcription start and end sites (TSS/TES).

Transcription of X-Linked Genes in XX, XO, and XY Oocytes

X and Y chromosomes have evolved from a pair of ancestral autosomes associated with the acquisition of the *SRY* male determining gene on the Y chromosome in mammals (Nagai, 2001; Wallis et al., 2008). While the X chromosome has retained most of its ancestral genes, the Y chromosome has lost most (Soh et al., 2014). This loss of ancestral genes from the Y chromosome led to a disparity in the dosage of X-linked versus autosomal genes between the two sexes and rendered compensation in two ways during evolution. While X-linked genes get upregulated to balance the output of X-linked and autosomal genes in XY somatic cells (X-linked gene dosage compensation), one of the two X chromosomes is inactivated (X inactivation) to balance the output of X-linked genes in XX somatic cells (Deng et al., 2013). However, neither mechanism operates in the germline (Fukuda et al., 2015; Sangrithi et al., 2017) (this study). In the mouse, when the germ cells arrive at the gonadal primordia and initiate sexual differentiation, they display both X-linked gene dosage compensation and X chromosome inactivation like somatic cell lines. Therefore, the X:A ratio is 1.0 in all XX, XO, and XY oogonia. However, when the inactive X chromosome is reactivated in XX oocytes at the onset of meiosis, the X:A ratio transiently increases to 1.7 and then declines to 1.0 in neonatal ovaries, while that in XO oocytes declines to 0.5. Since the two X chromosomes remain transcriptionally active in XX oocytes, transcript levels of X-linked genes appear to be adjusted to those of autosomal genes by suppressing the transcription of both X chromosomes. Such manner of sex chromosome dosage compensation is known in human preimplantation embryos as well as non-mammalian species such as *Drosophila* and *C. elegans* (Parkhurst and Meneely, 1994; Charlesworth, 1996; Soh et al., 2014; Petropoulos et al., 2016).

Our current results revealed two types of X-linked gene dosage compensation in the oocytes of the growth phase. For the genes with low transcript levels, the X:A ratio remained 1.0 in XX oocytes while it was increased to 1.0 in XO and XY oocytes. It appears that the X-linked gene dosage compensation is transiently lost in fetal and neonatal oocytes, but it is resumed in XO and XY oocytes during the resting stage or upon the entry into the growth phase. For the genes with high transcript levels, however, the X:A ratio increased to match the double X chromosome dosage in XX oocytes. Our RNA-Seq results identified 180 X-linked genes at lower transcript levels in XO or XY oocytes than XX oocytes with their ratios clustered around 0.47 along the entire X chromosome. Our qRT-PCR results revealed that *Eif2s3x*, *Bmp15*, and *Atrx* genes maintained the X chromosome dosage dependence during the entire oocyte growth. Thus, some X-linked genes appeared to be relieved from repression to enhance their transcription in XX oocytes during follicular growth. While XO oocytes maintained a half dosage of X-linked genes compared to XX oocytes until the end of growth phase, transcript levels of *Bmp15* and *Atrx* did not increase in XY oocytes as much as XO oocytes. The premature termination of *de novo* transcription in XY oocytes may have contributed to this difference. Alternatively, since FG-oocytes have a unique mechanism to store mRNAs, XY oocytes may have a limited capacity in such a storage mechanism.

Transcription of Y-Linked Genes in XY oocytes

Mammalian Y chromosomes used to be considered heterochromatic and gene poor. However, the male-specific region of the Y chromosome (MSY) of the B6 mouse strain has now been fully sequenced and turned out to be 99.9% euchromatic and encodes about 700 proteins (Soh et al., 2014). The mouse MSY retains only 2% of ancestral genes while all but 45 of the MSY's genes were newly acquired and massively amplified during evolution, and consist of three major Y gene families, *Sly* and *Srsy* and *Ssty* (Bishop and Hatat, 1987; Prado et al., 1992). These Y gene families have X homologs that were also newly acquired and amplified on the X chromosome (Soh et al., 2014). Of the ancestral genes, a 1.6 Mb region on the short arm contains seven single-copy genes, *Ddx3y*, *Eif2s3y*, *Kdm5d*, *Uba1y*, *Uty*, *Usp9y*, *Sry* and a pair of duplicated genes, *Zfy1* and *Zfy2* (Bellott et al., 2014). The remaining 0.4 Mb is ampliconic and contains one gene family *Rbmy*. Many of these ancestral genes are ubiquitously expressed, consistent with the speculation that surviving ancestral genes are widely expressed and regulate dosage-sensitive gene expression (Lahn and Page, 1997; Bellott et al., 2014). In our study, all these genes except for *Usp9y*, *Sry*, and *Rbmy* were transcribed at high levels in the XY oocytes of 50–60 µm. By contrast, the newly acquired ampliconic gene families on the long arm, *Sly*, *Ssty*, and *Srsy*, which are reported to be expressed predominantly in the male germline (Toure et al., 2004; Reynard and Turner, 2009), were undetectable in XY oocytes. However, other newly identified but poorly characterized genes on the Y long arm were transcribed at moderate levels in XY oocytes. These genes may be allowed for expression due to a lack of repressive regulators in oocytes. It remains to be determined whether Y-linked genes on the long arm exert biological activities in the oocyte.

Zfy2 was a candidate Y-linked gene for making the XY oocyte incompetent for embryonic development (Vernet et al., 2016). Both *Zfy1* and *Zfy2* are known to play critical roles in spermatogenesis and spermiogenesis (Vernet et al., 2014a; Vernet et al., 2016; Nakasuji et al., 2017). Furthermore, expression of exogenous *Zfy2*, but not *Zfy1*, in a XO female mouse on a mixed genetic background makes their oocytes incompetent for embryonic development (Vernet et al., 2014b). However, these studies used mice carrying the *Mus musculus molossinu*-type *Zfy2*, which harbors an 18 bp deletion compared to *Zfy1* (Nagamine et al., 1990). By contrast, *Mus musculus domesticus*-type *Zfy2* in the XY female used in this study harbors no such deletion, and none of the minor polymorphic differences (one or two consecutive nucleotide changes) between *Zfy1* and *Zfy2* would contribute to a functional difference. In our RNA-Seq data, both *Zfy1* and *Zfy2* were highly expressed in XY oocytes of 50–60 μm. However, qRT-PCR results showed rather low *Zfy1*/*Zfy2* transcript levels throughout the oocyte growth. Our current results do not exclude but do cast doubt on the contribution of *Zfy2* to the cytoplasmic defects of XY oocytes carrying the $\mathrm{Y}^{\mathrm{TIR}}$ chromosome.

Sex Chromosome Dosage Dependent DEGs

We presumed that XY oocytes are inferior to XO oocytes because of the presence of the ectopic Y chromosome and their failure to support embryonic development. However, our current results do not concur with this simplistic prediction. The transcript levels of the one thirds of autosomal DEGs were comparable between XX and XY oocytes but significantly lower in XO oocytes of 50–60 μm. XO oocytes appear to accumulate sufficient mRNAs by the end of growth phase to become competent for embryonic development. Nonetheless, the haplo-insufficiency of many gene products during the growth phase may make XO oocytes more vulnerable to genetic manipulations than XX oocytes. As discussed above, expression of a *Zfy2* transgene causes infertility in XO but not in XX female mice (Vernet et al., 2014b). Our current results show that the transcriptome landscape in XO oocytes is far off that in XX oocytes, more so than in XY oocytes of 50–60 μm. We predicted that the haplo-deficiency in essential X-linked genes was compensated by their Y-linked homologs to make XY oocytes closer to XX oocytes. In general, transcript levels of X homologs corresponded to the X chromosome dosage while those of Y homologs raised the sum in XY oocytes to match that in XX oocytes. Whether X and Y homologs share biological activities with comparable potencies in oocytes remain to be tested.

An interesting case is the genes in the PAR of sex chromosomes. In theory, these genes must behave like autosomal genes. However, two Y-linked genes within the PAR (*Gm21742* and *Mid1-ps1*) and one X-linked gene at the PAR boundary (*Mid1*) were transcribed at much higher levels in XY oocytes than in XX oocytes of 50–60 μm. Their transcript levels in XO oocytes were slightly higher than in XX oocytes, but still much lower than in XY oocytes. Thus, some genes within the PAR did not act like autosomal genes in oocytes. This phenomenon appears to be unique to the PAR since X-linked *Piga* and Y-linked *Gm21294* proximal to the PAR were transcribed as expected; X chromosome dosage dependent and XY specific, respectively. We speculate that the higher order structure or position of the Y chromosome in the XY oocyte nucleus may have favored high transcript levels of genes in the PAR.

Oocyte-Specific Alternative Splicing of *Kdm5b*

Most of multiexon-containing genes are subject to alternatively splicing. RNA-Seq data provide a great opportunity to identify oocyte-specific splicing isoforms. In our current study, we found such an example for *Kdm5b*. KDM5B, which catalyzes H3K4 demethylation, is presumed to be a transcriptional repressor since H3K4me3 is enriched at transcriptional start sites of active genes (Santos-Rosa et al., 2002; Schneider et al., 2004). However, the role of KDM5B in gene transcription is more complex and depends on cell type and stage. For example, depletion of KDM5B leads to lower H3K4me3 levels in promoter regions and higher H3K4me3 levels in gene body regions in mouse embryonic stem cells (He and Kidder, 2017). Consequently, KDM5B plays an integral role in regulating RNA Polymesase II occupancy, transcriptional initiation and elongation, and alternative splicing of target genes. In our current study, we found a novel oocyte-specific *Kdm5b* isoform, which was upregulated in XY oocytes near the end of growth phase. This isoform includes the catalytic domain and has the potential to exert the demethylase activity. The results with immunofluorescence staining indicated that the higher levels of *Kdm5b* transcript levels reflected into lower enrichment of H3K4me3, the target of KDM5B, in the nucleus of XY oocytes compared to XX oocytes during the growth phase.

Genes That Are Not Differentially Expressed

It is important to note that some genes which are known to play critical roles in the development of oocyte competence were not differentially expressed among XX, XO, and XY oocytes in our current study. Examples are *Gdf9* and *Tgf-β*, two oocyte secretory factors that are essential for proliferation, differentiation, and metabolism of surrounding follicular cells; *Gja1*, *Gja4*, *Cdh1*, *Cdh2*, and *Ptk2*, which are involved in the oocyte-granulosa cell communication; and *Zfp36l2*, which is important for global transcriptional silencing. TBP2, a TATA binding protein, is a central basal transcriptional regulator that drives cell type-specific features in the oocyte and is known to upregulate both *Gdf9* and *Bmp15* (Gazdag et al., 2009). However, we did not find a difference in the *Tbp2* transcript levels among XX, XO and XY oocytes, suggesting that the transcript levels of *Brmp15* were lower in XY oocytes due to the X chromosome dosage rather than availability of transcription factors. Finally, other than *Kdm5b*, none of the genes involved in post-translational histone modifications or DNA methylation were among the DEGs. Such results are consistent with our previous finding that the nucleus of XY FG-oocytes can generate healthy pups after its transfer into the XX oocyte cytoplasm (Obata et al., 2008).

Abnormal Morphological and Transcriptional Features in XY Oocytes at the End Of Growth Phase

Near the end of growth phase, XY oocytes exhibited abnormal chromatin configuration and mitochondria distribution, associated with premature silencing of *de novo* transcription. We speculate that all the morphological abnormalities observed in XY FG-oocytes are due to the altered gene transcription during the growth phase. We cannot pinpoint a gene or genes in this effect. However, X-linked *Bmp15* particularly attracted our attention. Its transcript levels initially corresponded to the X chromosome dosage, however, while they rapidly increased in XX oocytes with further growth, they increased much slowly in XY oocytes and their difference reached 9:1 at the end of growth phase. *Bmp15* encodes a TGFβ family member, known to be secreted by the oocyte to promote proliferation and glucose metabolism in the neighboring granulosa cells (Yan et al., 2001; Su et al., 2004; Sugiura et al., 2007). Since oocytes do not express the glucose transporter, they rely on the supply from the granulosa cells of glycolysis intermediates such as lactate and pyruvate for ATP production by mitochondria (Su et al., 2004; Su et al., 2008; Sutton-McDowall et al., 2010). Glucose metabolism through the pentose phosphate pathway is also crucial for the antioxidant defense and nucleotides supply in the oocyte (Richani et al., 2021). Our previous study has indeed shown that XY FG-oocytes produce lesser amounts of paracrine factors that regulate glycolytic gene expression in its companion granulosa cells, resulting in lower ATP contents in both granulosa cells and the enclosed oocyte (Xu et al., 2014). It is possible that low BMP15 levels indirectly affected global metabolism and gene transcription in XY oocytes during the growth phase.

Larp6 was the only gene to be found with lower transcript levels in XY oocytes compared to XX and XO oocytes of 50–60 μm in our RNA-Seq data. LARPs are a family of evolutionarily conserved RNA-binding proteins with diverse functions including tRNA processing, non-coding RNA metabolism, ribosomal biogenesis and mRNA translation (Maraia et al., 2017). Therefore, the low LARP6 levels may have contributed to the premature silencing of global gene transcription in XY oocytes. LAPR6 is also known to play a key role in mRNA translocation and translation at the protrusion of mesenchymal cells in culture (Dermit et al., 2020). Although the oocyte does not protrude, its surrounding follicular cells send numerous transzonal projections to the oocyte for exchanging molecules and metabolites (El-Hayek et al., 2018). It would be an attractive possibility that low *Larp* transcript levels are associated with poor communication with follicular cells and consequent lower metabolic supply.

By contrast, X-linked *Tnmd* was found to be transcribed at higher levels in XY oocytes than XO and XX oocytes in our RNA-Seq data. *Tnmd* encodes a type II transmembrane protein, which has potent anti-angiogenic activity necessary for cartilage and retina differentiation (Shukunami et al., 2005). *TNMD* is also highly expressed in human adipose tissue and involved in insulin resistance and glucose metabolism (Senol-Cosar et al., 2016). Altered expression of this gene may also have indicated abnormal metabolism in XY oocytes.

The unusual chromatin configuration in XY FG-oocytes puzzled us. It is known that the transition from NSN- to SN-type chromatin configuration is accompanied by a dynamic relocation of centromeres towards close apposition with the perinucleolar heterochromatin rim (de la Fuente et al., 2004a; Bonnet-Garnier et al., 2012). However, in addition to SN- and PSN-type condensation, XY oocytes exhibited chromatin condensation along the nuclear envelop, which was never seen in XX or XO oocytes. It has been reported that the perinucleolar centromere accumulation involves chromatin remodeling proteins such as ATRX (de la Fuente et al., 2004b). ATRX may have contributed to the abnormal chromatin configuration in XY oocytes since *Atrx* transcript levels in XY oocytes were lower than in XX oocytes throughout the growth phase and even lower (21%) at the end of growth phase. Lower ATRX accumulation may have allowed for the transcription of repetitive sequences such as transposable elements, which are normally repressed in oocytes. However, no ectopic expression or upregulation of genes near centromeres was found in XY oocytes in our RNA-Seq data.

The abnormal chromatin condensation along the nuclear envelop was frequently, but not always, accompanied by accumulation of mitochondria with high membrane potential surrounding the nucleus in XY FG-oocytes. For comparison, mitochondria were homogeneously distributed across the cytoplasm in XX and XO oocytes in agreement with previous reports (Dumollard et al., 2004; Al-Zubaidi et al., 2019). In our RNA-Seq data, none of mitochondrial genes were included in the DEGs in XY oocytes vs. XX and XO oocytes. However, mitochondrial function depends on the coordination of mitochondrial and nuclear genomes, the latter of which encode ~1,000 proteins and have not yet been thoroughly interrogated in our study. Mitochondrial activity, mediated by ATP production, affects histone modifications and transcription, and ultimately the developmental competence of oocytes (Van Blerkom, 2011; Matilainen et al., 2017). While all XY oocytes exhibited abnormal chromatin configuration, fewer oocytes showed abnormal mitochondrial distribution. Therefore, the altered mitochondrial distribution is more likely a consequence of other metabolic defects in XY oocytes. Nonetheless, further studies are needed to establish the hierarchical relationship between differential gene transcription, epigenetic modification, chromatin configuration and mitochondrial distribution, all of which appeared to be interrupted in the XY oocytes near the end of growth phase.

DATA AVAILABILITY STATEMENT

The sequencing data, as well as full length coverage tracks, raw 5'UMI, and normalized expression matrices reported in this study have been deposited in the Gene Expression Omnibus website with accession code GSE184153.

ETHICS STATEMENT

All animal experiments were performed in accordance with the Canadian Council on Animal Care and approved by the McGill University Animal Care Committee.

AUTHOR CONTRIBUTIONS

WY, DB, ST, JR, and TT contributed to the conception and design of the work. WY and DB acquired and analyzed the data. WY, DB, JR, and TT interpreted the results. WY, DB, ST, JR, and TT drafted the work and revised the intellectual content. WY, DB, ST, JR, and TT approved the version to be published. WY, DB, ST, JR, and TT agreed to be accountable for all aspects of the work.

FUNDING

This work was supported by the grant from the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-2018-04464) to TT and Canada Foundation for Innovation (CFI, 33408), CFI-JELF award (32557), Compute Canada RAC wst-164 and a Genome Canada GTP award to JR.

ACKNOWLEDGMENTS

We thank Min Fu and Shibo Feng (Molecular Imaging Platform, McGill University Health Center) for assisting our confocal microscope use and Yu Chang Wang (McGill Genome Center) for performing Smart-seq3 RNA-seq.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.792604/full#supplementary-material

REFERENCES

Al-Zubaidi, U., Liu, J., Cinar, O., Robker, R. L., Adhikari, D., and Carroll, J. (2019). The Spatio-Temporal Dynamics of Mitochondrial Membrane Potential during Oocyte Maturation. *Mol. Hum. Reprod.* 25, 695–705. doi:10.1093/molehr/gaz055

Amleh, A., Ledee, N., Saeed, J., and Taketo, T. (1996). Competence of Oocytes from the B6.Y^{DOM} Sex-Reversed Female Mouse for Maturation, Fertilization, and Embryonic Development *In Vitro*. *Dev. Biol.* 178, 263–275. doi:10.1006/dbio.1996.0217

Amleh, A., and Taketo, T. (1998). Live-borns from XX but Not XY Oocytes in the Chimeric Mouse Ovary Composed of B6.Y^{TIR} and XX Cells. *Biol. Reprod.* 58, 574–582. doi:10.1095/biolreprod58.2.574

Balboula, A. Z., Stein, P., Schultz, R. M., and Schindler, K. (2015). RBBP4 Regulates Histone Deacetylation and Bipolar Spindle Assembly during Oocyte Maturation in the Mouse. *Biol. Reprod.* 115, 128298. doi:10.1095/biolreprod.115.128298

Baumann, C., Schmidtmann, A., Muegge, K., and De La Fuente, R. (2008). Association of ATRX with Pericentric Heterochromatin and the Y Chromosome of Neonatal Mouse Spermatogonia. *BMC Mol. Biol.* 9, 29. doi:10.1186/1471-2199-9-29

Bellott, D. W., Hughes, J. F., Skaletsky, H., Brown, L. G., Pyntikova, T., Cho, T.-J., et al. (2014). Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators. *Nature* 508, 494–499. doi:10.1038/nature13206

Berletch, J. B., Yang, F., and Disteche, C. M. (2010). Escape from X Inactivation in Mice and Humans. *Genome Biol.* 11, 213. doi:10.1186/gb-2010-11-6-213

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., et al. (1990). Genetic Evidence Equating SRY and the Testis-Determining Factor. *Nature* 348, 448–450. doi:10.1038/348448a0

Bishop, C. E., and Hatat, D. (1987). Molecular Cloning and Sequence Analysis of a Mouse Y Chromosome RNA Transcript Expressed in the Testis. *Nucl. Acids Res.* 15, 2959–2969. doi:10.1093/nar/15.7.2959

Bonnet-Garnier, A., Feuerstein, P., Chebrout, M., Fleurot, R., Jan, H.-U., Debey, P., et al. (2012). Genome Organization and Epigenetic marks in Mouse Germinal Vesicle Oocytes. *Int. J. Dev. Biol.* 56, 877–887. doi:10.1387/ijdb.120149ab

Bouniol-Baly, C., Hamraoui, L., Guibert, J., Beaujean, N., Szollosi, M. S., and Debey, P. (1999). Differential Transcriptional Activity Associated with Chromatin Configuration in Fully Grown Mouse Germinal Vesicle Oocytes. *Biol. Reprod.* 60, 580–587. doi:10.1095/biolreprod60.3.580

Burgoyne, P. S., and Evans, E. (2000). A High Frequency of XO Offspring from $X^{Paf}Y^{*}$ Male Mice: Evidence that the *Paf* Mutation Involves an Inversion Spanning the X PAR Boundary. *Cytogenet. Cel Genet* 91, 57–61. doi:10.1159/000056819

Burgoyne, P. S., Tam, P. P. L., and Evans, E. P. (1983). Retarded Development of XO Conceptuses during Early Pregnancy in the Mouse. *J. Reprod. Fert* 68, 387–393. doi:10.1530/jrf.0.0680387

Cattanach, B. M. (1962). XO Mice. *Genet. Res. Camb* 3, 487–490. doi:10.1017/s0016672300003335

Charlesworth, B. (1996). The Evolution of Chromosomal Sex Determination and Dosage Compensation. *Curr. Biol.* 6, 149–162. doi:10.1016/s0960-9822(02)00448-7

Chuva De Sousa Lopes, S. M., Hayashi, K., Shovlin, T. C., Mifsud, W., Surani, M. A., and McLaren, A. (2008). X Chromosome Activity in Mouse XX Primordial Germ Cells. *Plos Genet.* 3, e30. doi:10.1371/journal.pgen.0040030

Coward, P., Nagai, K., Chen, D., Thomas, H. D., Nagamine, C. M., and Lau, Y. C. (1994). Polymorphism of a CAG Trinucleotide Repeat within Sry Correlates with B6.Y^{DOM} Sex Reversal. *Nat. Genet* 6, 245–250. doi:10.1038/ng0394-245

de la Fuente, R., and Eppig, J. J. (2001). Transcriptional Activity of the Mouse Oocyte Genome: Companion Granulosa Cells Modulate Transcription and Chromatin Remodeling. *Dev. Biol.* 229, 224–236. doi:10.1006/dbio.2000.9947

de la Fuente, R., Viveiros, M. M., Burns, K. H., Adashi, E. Y., Matzuk, M. M., and Eppig, J. J. (2004a). Major Chromatin Remodeling in the Germinal Vesicle (GV) of Mammalian Oocytes Is Dispensable for Global Transcriptional Silencing but Required for Centromeric Heterochromatin Function. *Dev. Biol.* 275, 447–458. doi:10.1016/j.ydbio.2004.08.028

de la Fuente, R., Viveiros, M. M., Wigglesworth, K., and Eppig, J. J. (2004b). ATRX, a Member of the SNF2 Family of Helicase-ATPase, Is Required for Chromosome Alignment and Meiotic Spindle Organization in Metaphase II Stage Mouse Oocyte. *Dev. Biol.* 272, 1–14. doi:10.1016/j.ydbio.2003.12.012

de Napoles, M., Nesterova, T., and Brockdorff, N. (2007). Early Loss of Xist RNA Expression and Inactive X Chromosome Associated Chromatin Modification in Developing Primordial Germ Cells. *PLos ONE* 2, e860. doi:10.1371/journal.pone.0000860

Debey, P., Szöllösi, M. S., Szöllösi, D., Vautier, D., Girousse, A., and Besombes, D. (1993). Competent Mouse Oocytes Isolated from Antral Follicles Exhibit

Different Chromatin Organization and Follow Different Maturation Dynamics. *Mol. Reprod. Dev.* 36, 59–74. doi:10.1002/mrd.1080360110

Decarpentrie, F., Vernet, N., Mahadevaiah, S. K., Longepied, G., Streichemberger, E., Aknin-Seifer, I., et al. (2012). Human and Mouse ZFY Genes Produce a Conserved Testis-specific Transcript Encoding a Zinc finger Protein with a Short Acidic Domain and Modified Transactivation Potential. *Hum. Mol. Genet.* 21, 2631–2645. doi:10.1093/hmg/dds088

Deng, X., Berletch, J. B., Ma, W., Hiatt, J. B., Noble, W. S., Shendure, J., et al. (2013). Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Dev. Cel.* 25, 55–68. doi:10.1016/j.devcel.2013.01.028

Deng, X., Berletch, J. B., Nguyen, D. K., and Disteche, C. M. (2014). X Chromosome Regulation: Diverse Patterns in Development, Tissues and Disease. *Nat. Rev. Genet.* 15, 367–378. doi:10.1038/nrg3687

Dermit, M., Dodel, M., Lee, F. C., Azman, M. S., Schwenzer, H., Jones, J. L., et al. (2020). Subcellular mRNA Localization Regulates Ribosome Biogenesis in Migrating Cells. *Dev. Cel.* 55, 298–313. doi:10.1016/j.devcel.2020.10.006

Dong, J., Albertini, D. F., Nishimori, K., Kumar, T. R., Lu, N., and Matzuk, M. M. (1996). Growth Differentiation Factor-9 Is Required during Early Ovarian Folliculogenesis. *Nature* 383, 531–535. doi:10.1038/383531a0

Dumollard, R., Marangos, P., Fitzharris, G., Swan, K., Duchen, M., and Carroll, J. (2004). Sperm-triggered [Ca $^{2+}$] Oscillations and Ca $^{2+}$ Homeostasis in the Mouse Egg Have an Absolute Requirement for Mitochondrial ATP Production. *Development* 131, 3057–3067. doi:10.1242/dev.01181

Ehrmann, I. E., Ellis, P. S., Mazeyrat, S., Duthie, S., Brockdorff, N., Mattei, M. G., et al. (1998). Characterization of Genes Encoding Translation Initiation Factor eIF-2g in Mouse and Human: Sex Chromosome Localization, Escape from X-Inactivation and Evolution. *Hum. Mol. Genet.* 7, 1725–1737. doi:10.1093/hmg/7.11.1725

Eicher, E. M., Washburn, L. L., Whitney, J. B. I., and Morrow, K. E. (1982). *Mus Poshiavinus* Y Chromosome in the C57BL/6J Murine Genome Causes Sex Reversal. *Science* 217, 535–537. doi:10.1126/science.7089579

El-Hayek, S., Yang, Q., Abbassi, L., FitzHarris, G., and Clarke, H. J. (2018). Mammalian Oocytes Locally Remodel Follicular Architecture to Provide the Foundation for Germline-Soma Communication. *Curr. Biol.* 28, 1124–1131. doi:10.1016/j.cub.2018.02.039

Fukuda, A., Tanino, M., Matoba, R., Umezawa, A., and Akutsu, H. (2015). Imbalance between the Expression Dosages of X-Chromosome and Autosomal Genes in Mammalian Oocytes. *Scientific Rep.* 5, e14101. doi:10.1038/srep14101

Gahurova, L., Tomizawa, S.-i., Smallwood, S. A., Stewart-Morgan, K. R., Saadeh, H., Kim, J., et al. (2017). Transcription and Chromatin Determinants of De Novo DNA Methylation Timing in Oocytes. *Epigenetics & chromatin* 10, 25. doi:10.1186/s13072-017-0133-5

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., et al. (2009). TBP2 Is Essential for Germ Cell Development by Regulating Transcription and Chromatin Condensation in the Oocyte. *Genes Dev.* 23, 2210–2223. doi:10.1101/gad.535209

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A. J., et al. (2020). Single-cell RNA Counting at Allele and Isoform Resolution Using Smart-Seq3. *Nat. Biotechnol.* 38, 708–714. doi:10.1038/s41587-020-0497-0

Hamada, N., Hamazaki, N., Shimamoto, S., Hikabe, O., Nagamatsu, G., Takada, Y., et al. (2020). Germ Cell-Intrinsic Effects of Sex Chromosomes on Early Oocyte Differentiation in Mice. *PLoS Genet.* 16, e1008676. doi:10.1371/journal.pgen.1008676

He, R., and Kidder, B. L. (2017). H3K4 Demethylase KDM5B Regulates Global Dynamics of Transcription Elongation and Alternative Splicing in Embryonic Stem Cells. *Nucleic Acids Res.* 45, 6427–6441. doi:10.1093/nar/gkx251

Hook, E. B., and Warburton, D. (2014). Turner Syndrome Revisited: Review of New Data Supports the Hypothesis that All Viable 45, X Cases Are Cryptic Mosaics with a rescue Cell Line, Implying an Origin by Mitotic Loss. *Hum. Genet.* 133, 417–424. doi:10.1007/s00439-014-1420-x

Inoue, A., Nakajima, R., Nagata, M., and Aoki, F. (2008). Contribution of the Oocyte Nucleus and Cytoplasm to the Determination of Meiotic and Developmental Cometence in Mice. *Hum. Reprod.* 23, 1377–1384. doi:10.1093/humrep/den096

Kay, G. F., Penny, G. D., Patel, D., Ashwarth, A., Brockdorff, N., and Rastan, S. (1994). Expression of Xist during Mouse Development Suggests a Role in the Initiation of X Chromosome Inactivation. *Cell* 72, 171–182. doi:10.1016/0092-8674(93)90658-d

Koopman, P., Gubbary, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male Development of Chromosomally Female Mice Transgenic for Sry. *Nature* 351, 117–121. doi:10.1038/351117a0

Koopman, P., Münsterberg, A., Capel, B., Vivian, N., and Lovell-Badge, R. (1990). Expression of a Candidate Sex-Determining Gene during Mouse Testis Differentiation. *Nature* 348, 450–452. doi:10.1038/348450a0

Korobova, O., Lane, P. W., Perry, J., Palmer, S., Ashworth, A., Davisson, M. T., et al. (1998). Patchy Fur, a Mouse Coat Mutation Associated with X–Y Nondisjunction, Maps to the Pseudoautosomal Boundary Region. *Genomics* 54, 556–559. doi:10.1006/geno.1998.5528

Lahn, B. T., and Page, D. C. (1997). Functional Coherence of the Human Y Chromosome. *Science* 278, 675–680. doi:10.1126/science.278.5338.675

Lane, P. W., and Davisson, M. T. (1990). Patchy Fur (Paf), a Semidominant X-Linked Gene Associated with a High Level of X-Y Nondisjunction in Male Mice. *J. Hered.* 81, 43–50. doi:10.1093/oxfordjournals.jhered.a110923

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* 32, 42–56. doi:10.1016/j.tig.2015.10.007

Leek, J. T., and Storey, J. D. (2007). Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genet.* 3, e161. doi:10.1371/journal.pgen.0030161

Levy, M. A., Kernohan, K. D., Jiang, Y., and Bérubé, N. G. (2015). ATRX Promotes Gene Expression by Facilitating Transcriptional Elongation through Guanine-Rich Coding Regions. *Hum. Mol. Genet.* 24, 1824–1835. doi:10.1093/hmg/ddu596

Lovejoy, C. A., Takai, K., Huh, M. S., Picketts, D. J., and de Lange, T. (2020). ATRX Affects the Repair of Telomeric DSBs by Promoting Cohesion and a DAXX-dependent Activity. *PLoS Biol.* 18, e3000594. doi:10.1371/journal.pbio.3000594

Lu, T., Chen, R., Cox, T. C., Moldrich, R. X., Kurniawan, N., Tan, G., et al. (2013). X-linked Microtubule-Associated Protein, Mid1, Regulates Axon Development. *Proc. Natl. Acad. Sci.* 110, 19131–19136. doi:10.1073/pnas.1303687110

Luoh, S.-W., Bain, P. A., Polakiewicz, R. D., Goodheart, M. L., Gardner, H., Jaenisch, R., et al. (1997). *Zfx* Mutation Results in Small Animal Size and Reduced Germ Cell Number in Male and Female Mice. *Development* 124, 2275–2284. doi:10.1242/dev.124.11.2275

Maraia, R. J., Mattijssen, S., Cruz-Gallardo, I., and Conte, M. R. (2017). The La and Related RNA-binding Proteins (LARPs): Structures, Functions, and Evolving Perspectives. *Wiley Interdiscip. Rev. RNA* 8, e1430. doi:10.1002/wrna.1430

Matilainen, O., Quirós, P. M., and Auwerx, J. (2017). Mitochondria and Epigenetics–Crosstalk in Homeostasis and Stress. *Trends Cell Biology* 27, 453–463. doi:10.1016/j.tcb.2017.02.004

Matsumura, T., Endo, T., Isotani, A., Ogawa, M., and Ikawa, M. (2019). An Azoospermic Factor Gene, Ddx3y and its Paralog, Ddx3x Are Dispensable in Germ Cells for Male Fertility. *J. Reprod. Dev.* 65, 121–128. doi:10.1262/jrd.2018-145

Mazeyrat, S., Saut, N., Grigoriev, V., Mahadevaiah, S. K., Ojarikre, O. A., Rattigan, A., et al. (2001). A Y-Encoded Subunit of the Translation Initiation Factor Eif2 Is Essential for Mouse Spermatogenesis. *Nat. Genet* 29, 49–53. doi:10.1038/ng717

McLaren, A. (1988). Somatic and Germ-Cell Sex in Mammals. *Phil Trans. R. Soc. Lond. B* 322, 3–9. doi:10.1098/rstb.1988.0109

Messinger, S. M., and Albertini, D. F. (1991). Centrosome and Microtubule Dynamics during Meiotic Progression in the Mouse Oocyte. *J. Cel Sci* 100, 289–298. doi:10.1242/jcs.100.2.289

Nagai, K. (2001). Molecular Evolution of Sry and Sox Gene. *Gene* 270, 161–169. doi:10.1016/s0378-1119(01)00479-6

Nagamine, C. M., Chan, K., Hake, L. E., and Lau, Y.-F. C. (1990). The Two Candidate Testis-Determining Y Genes (Zfy-1 and Zfy-2) Are Differentially Expressed in Fetal and Adult Mouse Tissues. *Genes Dev.* 4, 63–74. doi:10.1101/gad.4.1.63

Nagamine, C. M., Taketo, T., and Koo, G. C. (1987). Studies on the Genetics of *Tda-1* XY Sex Reversal in the Mouse. *Differentiation* 33, 223–231. doi:10.1111/j.1432-0436.1987.tb01561.x

Nakasuji, T., Ogonuki, N., Chiba, T., Kato, T., Shiozawa, K., Yamatoya, K., et al. (2017). Complementary Critical Functions of *Zfy1* and *Zfy2* in Mouse Spermatogenesis and Reproduction. *PLoS Genet.* 13, e1006578. doi:10.1371/journal.pgen.1006578

Nguyen, D. K., and Disteche, C. M. (2006). Dosage Compensation of the Active X Chromosome in Mammals. *Nat. Genet* 38, 47–53. doi:10.1038/ng1705

Obata, Y., Villemure, M., Kono, T., and Taketo, T. (2008). Transmission of Y Chromosomes from XY Female Mice Was Made Possible by the Replacement of Cytoplasm during Oocyte Maturation. *Proc. Natl. Acad. Sci. USA* 105, 13918–13923. doi:10.1073/pnas.0802680105

Ogata, T., and Matsuo, N. (1995). Turner Syndrome and Female Sex Chromosome Aberrations: Deduction of the Principal Factors Involved in the Development of Clinical Features. *Hum. Genet.* 95, 607–629. doi:10.1007/BF00209476

Oikonomopoulos, S., Wang, Y. C., Djambazian, H., Badescu, D., and Ragoussis, J. (2016). Benchmarking of the Oxford Nanopore MinION Sequencing for Quantitative and Qualitative Assessment of cDNA Populations. *Sci. Rep.* 6, e31602. doi:10.1038/srep31602

Olivier, J., Johnson, W. D., and Marshall, G. D. (2008). The Logarithmic Transformation and the Geometric Mean in Reporting Experimental IgE Results: what Are They and when and Why to Use Them? *Ann. Allergy Asthma Immunol.* 100, 333–337. doi:10.1016/s1081-1206(10)60595-9

Palmer, S., Perry, J., Kipling, D., and Ashworth, A. (1997). A Gene Spans the Pseudoautosomal Boundary in Mice. *Proc. Natl. Acad. Sci. USA* 94, 12030–12035. doi:10.1073/pnas.94.22.12030

Park, S., Zeidan, K. T., Shin, J. S., and Taketo, T. (2011). SRY Upregulation of SOX9 Is Inefficient and Delayed, Allowing Ovarian Differentiationthe B6. Y^{TIR} *Gonad. Differ.* 82, 18–27. doi:10.1016/j.diff.2011.04.007

Parkhurst, S. M., and Meneely, P. M. (1994). Sex Determination and Dosage Compensation: Lessons from Flies and Worms. *Science* 264, 924–932. doi:10.1126/science.8178152

Perry, J., Palmer, S., Gabriel, A., and Ashworth, A. (2001). A Short Pseudoautosomal Region in Laboratory Mice. *Genome Res.* 11, 1826–1832. doi:10.1101/gr.203001

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., et al. (2016). Single-cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 165, 1012–1026. doi:10.1016/j.cell.2016.03.023

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-Seq from Single Cells Using Smart-Seq2. *Nat. Protoc.* 9, 171–181. doi:10.1038/nprot.2014.006

Prado, V. F., Lee, C.-H., Zahed, L., Vekemans, M., and Nishioka, Y. (1992). Molecular Characterization of a Mouse Y Chromosomal Repetitive Sequence that Detects Transcripts in the Testis. *Cytogenet. Cel Genet* 61, 87–90. doi:10.1159/000133375

Puschendorf, M., Terranova, R., Boutsma, E., Mao, X., Isono, K.-I., Brykczynska, U., et al. (2008). PRC1 and Suv39h Specify Parental Asymmetry at Constitutive Heterochromatin in Early Mouse Embryos. *Natue Genet.* 40, 411–420. doi:10.1038/ng.99

Reynard, L. N., and Turner, J. M. A. (2009). Increased Sex Chromosome Expression and Epigenetic Abnormalities in Spermatids from Male Mice with Y Chromosome Deletion. *J. Cel Sci* 122, 4239–4248. doi:10.1242/jcs.049916

Richani, D., Dunning, K. R., Thompson, J. G., and Gilchrist, R. B. (2021). Metabolic Co-dependence of the Oocyte and Cumulus Cells: Essential Role in Determining Oocyte Developmental Competence. *Hum. Reprod. Update* 27, 27–47. doi:10.1093/humupd/dmaa043

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Sangrithi, M. N., Royo, H., Mahadevaiah, S. K., Ojarikre, O., Bhaw, L., Sesay, A., et al. (2017). Non-Canonical and Sexually Dimorphic X Dosage Compensation States in the Mouse and Human Germline. *Dev. Cel.* 40, 289–301. doi:10.1016/j.devcel.2016.12.023

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. T., et al. (2002). Active Genes Are Tri-methylated at K4 of Histone H3. *Nature* 419, 407–411. doi:10.1038/nature01080

Schneider, R., Bannister, A. J., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 Lysine 4 Methylation Patterns in Higher Eukaryotic Genes. *Nat. Cel. Biol.* 6, 73–77. doi:10.1038/ncb1076

Sekiguchi, T., Iida, H., Fukumura, J., and Nishimoto, T. (2004). Human DDX3Y, the Y-Encoded Isoform of RNA Helicase DDX3, Rescues a Hamster Temperature-Sensitive ET24 Mutant Cell Line with a DDX3X Mutation. *Exp. Cel. Res.* 300, 213–222. doi:10.1016/j.yexcr.2004.07.005

Senol-Cosar, O., Flach, R. J. R., DiStefano, M., Chawla, A., Nicoloro, S., Straubhaar, J., et al. (2016). Tenomodulin Promotes Human Adipocyte Differentiation and Beneficial Visceral Adipose Tissue Expansion. *Nat. Commun.* 7, e10686. doi:10.1038/ncomms10686

Shpargel, K. B., Sengoku, T., Yokoyama, S., and Magnuson, T. (2012). UTX and UTY Demonstrate Histone Demethylase-independent Function in Mouse Embryonic Development. *PLoS Genet.* 8, e1002964. doi:10.1371/journal.pgen.1002964

Shukunami, C., Oshima, Y., and Hiraki, Y. (2005). Chondromodulin-I and Tenomodulin: a New Class of Tissue-specific Angiogenesis Inhibitors Found in Hypovascular Connective Tissues. *Biochem. biophysical Res. Commun.* 333, 299–307. doi:10.1016/j.bbrc.2005.05.133

Singh, R. P., and Carr, D. H. (1966). The Anatomy and Histology of XO Human Embryos and Fetuses. *Anat. Rec.* 155, 369–384. doi:10.1002/ar.1091550309

Soh, Y. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., et al. (2014). Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell* 159, 800–813. doi:10.1016/j.cell.2014.09.052

Stäubli, A., and Peters, A. H. (2021). Mechanisms of Maternal Intergenerational Epigenetic Inheritance. *Curr. Opin. Genet. Dev.* 67, 151–162. doi:10.1016/j.gde.2021.01.008

Su, Y.-Q., Sugiura, K., Wigglesworth, K., O'Brien, M. J., Affourtit, J. P., Pangas, S. A., et al. (2008). Oocyte Regulation of Metabolic Cooperativity between Mouse Cumulus Cells and Oocytes: BMP15 and GDF9 Control Cholesterol Biosynthesis in Cumulus Cells. *Development* 135, 111–121. doi:10.1242/dev.009068

Su, Y.-Q., Wu, X., O'Brien, M. J., Pendola, F. L., Denegre, J. N., Matzuk, M. M., et al. (2004). Synergistic Roles of BMP15 and GDF9 in the Development and Function of the Oocyte-Cumulus Cell Complex in Mice: Genetic Evidence for an Oocyte-Granulosa Cell Regulatory Loop. *Dev. Biol.* 276, 64–73. doi:10.1016/j.ydbio.2004.08.020

Sugimoto, M., and Abe, K. (2007). X Chromosome Reactivation Initiates in Nascent Primordial Germ Cells in Mice. *Plos Genet.* 3, e116. doi:10.1371/journal.pgen.0030116

Sugiura, K., Su, Y.-Q., Diaz, F. J., Pangas, S. A., Sharma, S., Wigglesworth, K., et al. (2007). Oocyte-derived BMP15 and FGFs Cooperate to Promote Glycolysis in Cumulus Cells. *Development* 134, 2593–2603. doi:10.1242/dev.006882

Sutton-McDowall, M. L., Gilchrist, R. B., and Thompson, J. G. (2010). The Pivotal Role of Glucose Metabolism in Determining Oocyte Developmental Competence. *Reproduction* 139, 685–695. doi:10.1530/rep-09-0345

Taketo, T., Lee, C.-H., Zhang, J., Li, Y., Lee, C.-Y. G., and Lau, Y.-F. C. (2005). Expression of SRY Proteins in Both normal and Sex-Reversed XY Fetal Mouse Gonads. *Dev. Dyn.* 233, 612–622. doi:10.1002/dvdy.20352

Taketo, T. (2012). Microspread Ovarian Cell Preparations for the Analysis of Meiotic Prophase Progression with Improved Recovery by Cytospin Centrifugation. *Meth Mol. Biol.* 825, 173–181. doi:10.1007/978-1-61779-436-0_13

Taketo, T. (2015). The Role of Sex Chromosomes in Mammalian Germ Cell Differentiation: Can the Germ Cells Carrying X and Y Chromosomes Differentiate into fertile Oocytes? *Asian J. Androl.* 17, 360–366. doi:10.4103/1008-682X.143306

Taketo-Hosotani, T., Nishioka, Y., Nagamine, C., Villalpando, I., and Merchant-Larios, H. (1989). Development and Fertility of Ovaries in the B6.Y^{DOM} Sex-Reversed Female Mouse. *Development* 107, 95–105. doi:10.1242/dev.107.1.95

Toure, A., Szot, M., Mahadevaiah, S. K., Rattigan, A., Ojarikre, O. A., and Burgoyne, P. S. (2004). A New Deletion of the Mouse Y Chromosome Long Arm Associated with the Loss of Ssty Expression, Abnormal Sperm Development and Sterility. *Genetics* 166, 901–912. doi:10.1093/genetics/166.2.901

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., et al. (2017). Landscape of X Chromosome Inactivation across Human Tissues. *Nature* 550, 244–248. doi:10.1038/nature24265

Turner, H. H. (1938). A Syndrome of Infantilism, Congenital Webbed Neck, and Cubitus Valgus. *Endocrinology* 23, 566–574. doi:10.1210/endo-23-5-566

Van Blerkom, J. (2011). Mitochondrial Function in the Human Oocyte and Embryo and Their Role in Developmental Competence. *Mitochondrion* 11, 797–813. doi:10.1016/j.mito.2010.09.012

Vaz, B., El Mansouri, F., Liu, X., and Taketo, T. (2020). Premature Ovarian Insufficiency in the XO Female Mouse on the C57BL/6J Genetic Background. *Mol. Hum. Reprod.* 26, 678–688. doi:10.1093/molehr/gaaa049

Vernet, N., Mahadevaiah, S. K., de Rooij, D. G., Burgoyne, P. S., and Ellis, P. J. (2016). Zfy Genes Are Required for Efficient Meiotic Sex Chromosome Inactivation (MSCI) in Spermatocytes. *Hum. Mol. Genet.* 25, 5300–5310. doi:10.1093/hmg/ddw344

Vernet, N., Mahadevaiah, S. K., Yamauchi, Y., Decarpentrie, F., Mitchell, M. J., Ward, M. A., et al. (2014a). Mouse Y-Linked *Zfy1* and *Zfy2* Are Expressed during the Male-specific Interphase between Meiosis I and Meiosis II and Promote the 2nd Meiotic Division. *PLoS Genet.* 10, e1004444. doi:10.1371/journal.pgen.1004444

Vernet, N., Szot, M., Mahadevaiah, S. K., Ellis, P. J. I., Decarpentrie, F., Ojarikre, O. A., et al. (2014b). The Expression of Y-Linked *Zfy2* in XY Mouse Oocytes Leads to Frequent Meiosis 2 Defects, a High Incidence of Subsequent Early Cleavage Stage Arrest and Infertility. *Development* 141, 855–866. doi:10.1242/dev.091165

Villemure, M., Chen, H.-Y., Kurokawa, M., Fissore, R. A., and Taketo, T. (2007). The Presence of X- and Y-Chromosomes in Oocytes Leads to Impairment in the Progression of the Second Meiotic Division. *Dev. Biol.* 301, 1–13. doi:10.1016/j.ydbio.2006.10.034

Wallis, M. C., Waters, P. D., and Graves, J. A. M. (2008). Sex Determination in Mammals - before and after the Evolution of *SRY. Cell Mol Life Sci* 65, 3182–3195. doi:10.1007/s00018-008-8109-z

Welstead, G. G., Creyghton, M. P., Bilodeau, S., Cheng, A. W., Markoulaki, S., Young, R. A., et al. (2012). X-linked H3K27me3 Demethylase Utx Is Required for Embryonic Development in a Sex-specific Manner. *Proc. Natl. Acad. Sci.* 109, 13004–13009. doi:10.1073/pnas.1210787109

Xu, B., Noohi, S., Shin, J. S., Tan, S. L., and Taketo, T. (2014). Bi-directional Communication with the Cumulus Cells Is Involved in the Deficiency of XY Oocytes in the Components Essential for Proper Second Meiotic Spindle Assembly. *Dev. Biol.* 385, 242–252. doi:10.1016/j.ydbio.2013.11.004

Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A. B., et al. (2019). SETD2 Regulates the Maternal Epigenome, Genomic Imprinting and Embryonic Development. *Nat. Genet.* 51, 844–856. doi:10.1038/s41588-019-0398-7

Yamauchi, Y., Riel, J. M., Ruthig, V. A., Ortega, E. A., Mitchell, M. J., and Ward, M. A. (2016). Two Genes Substitute for the Mouse Y Chromosome for Spermatogenesis and Reproduction. *Science* 351, 514–516. doi:10.1126/science.aad1795

Yan, C., Wang, P., DeMayo, J., DeMayo, F. J., Elvin, J. A., Carino, C., et al. (2001). Synergistic Roles of Bone Morphogenetic Protein 15 and Growth Differentiation Factor 9 in Ovarian Function. *Mol. Endocrinol.* 15, 854–866. doi:10.1210/mend.15.6.0662

Yang, F., Babak, T., Shendure, J., and Disteche, C. M. (2010). Global Survey of Escape from X Inactivation by RNA-Sequencing in Mouse. *Genome Res.* 20, 614–622. doi:10.1101/gr.103200.109

Yin, W., Kim, H.-T., Wang, S., Gunawan, F., Wang, L., Kishimoto, K., et al. (2018). The Potassium Channel KCNJ13 Is Essential for Smooth Muscle Cytoskeletal Organization during Mouse Tracheal Tubulogenesis. *Nat. Commun.* 9, e2815. doi:10.1038/s41467-018-05043-5

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., et al. (2016). Apache Spark: a Unified Engine for Big Data Processing. *Commun. ACM* 59, 56–65. doi:10.1145/2934664

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., et al. (2016). Allelic Reprogramming of the Histone Modification H3K4me3 in Early Mammalian Development. *Nature* 537, 553–557. doi:10.1038/nature19361

Zhang, Z., Yu, D., Seo, M., Hersh, C. P., Weiss, S. T., and Qiu, W. (2019). Novel Data Transformations for RNA-Seq Differential Expression Analysis. *Scientific Rep.* 9, e4820. doi:10.1038/s41598-019-41315-w

Zhu, J.-Q., Tan, S.L, and Taketo, T. (2017). A lack of coordination between sister-chromatids segregation and cytokinesis in the oocytes of B6. Y^{TIR} (XY) sex-reversed female mice. *Sci Rep* 7, e960. doi:10.1038/s41598-017-00922-1

Zuccotti, M., Piccinelli, A., Rossi, P. G., and Garagna, S. R. (1995). Chromatin Organization during Mouse Oocyte Growth. *Mol. Reprod. Dev.* 41, 479–485. doi:10.1002/mrd.1080410410

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 December 2021
doi: 10.3389/fgene.2021.783826

Blastocyst Morphology Based on Uniform Time-Point Assessments is Correlated With Mosaic Levels in Embryos

Chien-Hong Chen[1†], *Chun-I Lee*[1,2,3†], *Chun-Chia Huang*[1], *Hsiu-Hui Chen*[1], *Shu-Ting Ho*[1], *En-Hui Cheng*[1], *Pin-Yao Lin*[1], *Chung-I Chen*[1], *Tsung-Hsien Lee*[1,2,3*] *and Maw-Sheng Lee*[1,2,3*]

[1]*Division of Infertility, Lee Women's Hospital, Taichung, Taiwan,* [2]*Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan,* [3]*Department of Obstetrics and Gynecology, Chung Shan Medical University Hospital, Taichung, Taiwan*

OPEN ACCESS

Edited by:
Yan Yun,
University of California, Davis,
United States

Reviewed by:
Silvestre Sampino,
Polish Academy of Sciences, Poland
Alison Campbell,
CARE Fertility Nottingham,
United Kingdom

***Correspondence:**
Tsung-Hsien Lee
jackth.lee@gmail.com
Maw-Sheng Lee
msleephd@gmail.com

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *27 September 2021*
Accepted: *06 December 2021*
Published: *22 December 2021*

Citation:
Chen C-H, Lee C-I, Huang C-C, Chen H-H, Ho S-T, Cheng E-H, Lin P, Chen C-I, Lee T-H and Lee M-S (2021) Blastocyst Morphology Based on Uniform Time-Point Assessments is Correlated With Mosaic Levels in Embryos. Front. Genet. 12:783826. doi: 10.3389/fgene.2021.783826

Avoiding aneuploid embryo transfers has been shown to improve pregnancy outcomes in patients with implantation failure and pregnancy loss. This retrospective cohort study aims to analyze the correlation of time-lapse (TL)-based variables and numeric blastocyst morphological scores (TLBMSs) with different mosaic levels. In total, 918 biopsied blastocysts with time-lapse assessments at a uniform time-point were subjected to next-generation sequencing–based preimplantation genetic testing for aneuploidy. In consideration of patient- and cycle-related confounding factors, all redefined blastocyst morphology components of low-grade blastocysts, that is, expansion levels (odds ratio [OR] = 0.388, 95% confidence interval [CI] = 0.217–0.695; OR = 0.328, 95% CI = 0.181–0.596; OR = 0.343, 95% CI = 0.179–0.657), inner cell mass grades (OR = 0.563, 95% CI = 0.333–0.962; OR = 0.35, 95% CI = 0.211–0.58; OR = 0.497, 95% CI = 0.274–0.9), and trophectoderm grades (OR = 0.29, 95% CI = 0.178–0.473; OR = 0.242, 95% CI = 0.143–0.411; OR = 0.3, 95% CI = 0.162–0.554), were less correlated with mosaic levels ≤20%, <50%, and ≤80% as compared with those of top-grade blastocysts ($p < 0.05$). After converting blastocyst morphology grades into scores, high TLBMSs were associated with greater probabilities of mosaic levels ≤20% (OR = 1.326, 95% CI = 1.187–1.481), <50% (OR = 1.425, 95% CI = 1.262–1.608), and ≤80% (OR = 1.351, 95% CI = 1.186–1.539) ($p < 0.001$). The prediction abilities of TLBMSs were similar for mosaic levels ≤20% (AUC = 0.604, 95% CI = 0.565–0.642), <50% (AUC = 0.634, 95% CI = 0.598–0.671), and ≤80% (AUC = 0.617, 95% CI = 0.576–0.658). In conclusion, detailed evaluation with TL monitoring at the specific time window reveals that redefined blastocyst morphology components and converted numeric TLBMSs are significantly correlated with all of the threshold levels of mosaicism. However, the performance of TLBMSs to differentiate blastocysts with aberrant ploidy risk remains perfectible.

Keywords: blastocyst morphology, mosaic levels, preimplantation genetic testing for aneuploidy, time-lapse monitoring, high-resolution next-generation sequencing

INTRODUCTION

Several studies have suggested that time-lapse (TL) monitoring can enhance aneuploid embryo identification, particularly based on blastulation kinetics (Campbell et al., 2013a; Campbell et al., 2013b; Basile et al., 2014; Minasi et al., 2016; Del Carmen Nogales et al., 2017; Mumusoglu et al., 2017; Desai et al., 2018). Campbell et al. revealed that the period from insemination to the earliest signs of compaction, to the earliest signs of cavitation (tSB), and to the blastocyst with a full-filled blastocoel (tB) are prolonged in aneuploid embryos compared with those in euploid embryos. A risk classification model of aneuploidy based on two blastulation kinetics parameters (tSB and tB) has been established, which has the potential to select embryos with increased rates of fetal heart beat and live birth during fresh embryo transfer cycles without PGT-A (Campbell et al., 2013a; Campbell et al., 2013b). Similar results were obtained from a dataset of 928-blastocysts and confirmed that euploid embryos have fast kinetics related to blastocyst development (Minasi et al., 2016). Moreover, Mumusoglu et al. (2017) further considered the confounding factors related to patient and ovarian stimulation and demonstrated that tB and the time from insemination to the expanded blastocyst stage can moderately predict aneuploidy (Mumusoglu et al., 2017). Using TL technology, dynamic dysmorphisms of cleavage-stage embryos, such as multinucleation (MN), reverse cleavage (RC), irregular chaotic division (ICD), or direct cleavage (DC), can be evaluated in more detail, and two or more dysmorphisms are correlated with a reduced euploidy probability (Desai et al., 2018). However, these studies have determined the ploidy status, at least in part if not all, of the biopsied embryos through array comparative genomic hybridization (aCGH), which has insufficient sensitivity for the detection of diploid–aneuploid mosaicism (van Echten-Arends et al., 2011; Munne and Wells 2017; Sachdev et al., 2017).

Studies have demonstrated the potential of the postimplantation development of transferred mosaic embryos (Greco et al., 2015; Munne et al., 2017; Viotti et al., 2021). Importantly, the application of high-resolution next-generation sequencing (hr-NGS) may enhance sensitivity to the level of 10 MB, through which the detailed ploidy characteristics of embryos, for example, chromosomal gains and losses, different mosaic levels, and whole-or segmental-chromosome aneuploidy, are efficiently identified (Kane et al., 2016; Munne and Wells 2017). Clinical results have demonstrated that compared with array CGH counterparts, hr-NGS significantly improves the pregnancy outcomes of single-euploid embryo transfer cycles (Friedenthal et al., 2018), and evidence supports that healthy babies can result from embryos with low or high mosaic levels (Lee et al., 2020b; Lin et al., 2020). However, the clinical outcomes of mosaic embryos appear to be significantly affected by mosaic levels (<50% vs. ≥50%), mosaic types (chromosome numbers), and blastocyst morphology (Viotti et al., 2021). It would be interesting to evaluate the correlation of the mosaic status with embryonic characteristics. To date, whether any embryonic feature has the ability to distinguish embryos with different mosaicism levels remains to be investigated.

Numerous experimental studies have delineated a link between blastocyst morphology and chromosomal abnormalities (Alfarawati et al., 2011; Capalbo et al., 2014; Minasi et al., 2016; Barash et al., 2017; Wang et al., 2018; Zhan et al., 2020). Individual morphological components of blastocyst evaluation, that is, expansion degrees, inner cell mass (ICM) grades, and trophectoderm (TE) grades, may be correlated with the ploidy status independently (Minasi et al., 2016; Zhan et al., 2020). The combination of expansion, ICM, and TE grades could therefore result in complication for ranking the embryos. Thus, several studies have combined morphological components to categorize blastocysts based on quality (Capalbo et al., 2014; Wirleitner et al., 2016) and have even considered the blastocyst biopsy day and morphological features for generating a blastocyst score (Zhan et al., 2020). Although these modified approaches demonstrate that blastocyst morphology is highly associated with its ploidy (Capalbo et al., 2014; Zhan et al., 2020), embryo grading variability may occur due to multiple reasons, such as subjective judgments of individual embryologists, laboratory settings, and observation time windows. One of the advantages of TL monitoring is objectively and accurately defined observation time after insemination or ICSI. The hypothesis was therefore raised that blastocyst morphology observed at a definite time period by TL monitoring might be precisely correlated with the mosaic level in an embryo. If the hypothesis is correct, such blastocyst morphology scores may be used to improve embryo selection without the help of AI or labor-intensive interpretation of the whole TL videos.

In the current study, hr-NGS was applied to classify the mosaic level of a blastocyst. We refined embryo assessments at a uniform time-point by using the definition based on TL monitoring and aimed to continuously improve our understanding regarding their relationship with different mosaic levels. In addition, the effectiveness of comprehensive assessment, which converted blastocyst morphology into a numeric score, in predicting ploidy was evaluated by considering the effects of potential confounding variables.

MATERIALS AND METHODS

Study Design and Patient Selection

The data were collected at Lee Women's Hospital from January 2017 to August 2018. In total, 210 infertile couples undergoing PGT-A and TL cultivation were enrolled into this retrospective cohort study, which was approved by the Institutional Review Board of Chung Sun Medical University, Taichung, Taiwan (approval number CS19039). Patients with surgical sperm retrieval ($n = 4$) and without qualified blastocysts for PGT-A ($n = 28$) were excluded. A total of 918 biopsied blastocysts from 178 couples were analyzed in this study.

Controlled Ovarian Stimulation

All the procedures were conducted following the standard protocols stated in our previous report (Lee et al., 2019). Briefly, in this study, controlled ovarian stimulation was applied using the gonadotrophin-releasing hormone (GnRH) agonist long protocol or the GnRH antagonist protocol. Patients who were assigned the GnRH agonist long protocol received leuprolide acetate (Lupron; Takeda Chemical Industries, Osaka, Japan) subcutaneously (0.05–0.1 mg/day) from day 21 of the first cycle to the day of human chorionic gonadotropin (hCG; Ovidrel; Serono, Modugno, Italy) injection (250 μg). Exogenous gonadotropin (Gonal-F; Serono or Menopur; Ferring, São Paulo, Brazil) was then administered from day 3 of the second cycle until the leading follicle was ≥18 mm. Patients who were assigned the GnRH antagonist protocol received the gonadotropin regimen, same as that in the agonist long protocol, and additional cetrorelix acetate (Cetrotide; Merck Serono, Geneva, Switzerland) injections (0.25 mg/day) were given subcutaneously before hCG injection. Oocyte maturation was triggered through hCG administration, and ultrasound-guided ovum retrieval was performed approximately 36 h after hCG injection (Lee et al., 2020a).

Cultivation, Insemination, and Micromanipulation

The TL culture system, EmbryoScope+ (Vitrolife, Kungsbacka, Sweden), was used in this study. The cultivation condition consisting of an atmosphere of 5% O_2, 5% CO_2, and 90% N_2 at 37°C was applied for oocytes and embryos. The media covered with paraffin oil (Vitrolife) were equilibrated in an incubator for at least 2 h before use, including the fertilization medium (SAGE Biopharma, Bedminster, NJ, United States) with 15% serum protein substitute (SPS; SAGE Biopharma) for conventional insemination or intracytoplasmic sperm injection (ICSI), cleavage medium (SAGE Biopharma) with 15% SPS, and blastocyst medium (SAGE Biopharma) with 15% SPS for embryo culture. On the basis of the World Health Organization Laboratory Manual for the Examination and Processing of Human Semen, abnormal sperm quality was defined if at least one of the following properties was found: oligospermia (sperm concentration $\leq 15 \times 10^6$/ml), asthenospermia (total motility ≤40%), and teratospermia (normal sperm form ≤4%). ICSI was used for all of the mature oocytes from couples with oligoasthenoteratozoospermia (OAT) or for approximately half of the mature oocytes from couples with non-OAT.

Time-Lapse Assessment and Morphological Grading

Embryo morphology and developmental dynamics were noninvasively observed through the capture of images with 11 focal planes at 10-min intervals using a TL device. According to guidelines by Vitrolife and published nomenclature (Ciray et al., 2014), the exact timings of developmental events and morphological features after complete pronuclei fading (tPNf) were annotated on a daily basis. **Supplementary Table S1** provides the definitions of blastulation-related kinetics, including the time to accomplish compaction (tM), tSB, tB, the time interval between tM and tB (tB-tM), and the time interval between tSB and tB (tB-tSB), as well as cleavage aberrations at early cleavage stages (≤8-cell stage), including delayed division (DD), DC, RC, ICD, MN at the two-cell stage (MN2), and MN at the four-cell stage (MN4). In order to synchronized the observation time window of individual blastocysts, the morphological evaluation was performed at 118 h post insemination (hpi) by all of recorded images of embryonic development using redefined descriptions of embryo expansion, TE quality, and ICM quality through TL monitoring (**Supplementary Figure S1** and **Supplementary Table S2**). A score from 1 to 7 was then assigned to the expansion level (level of <1 = 1, level 1 = 3, level 2 = 5, and level 3 = 7), and a score from 0 to 2 was assigned to ICM and TE groups (grade A = 2, grade B = 1, and grade C or less = 0). The TL-based blastocyst morphological score (TLBMS) was then developed to represent the morphological grading of a blastocyst using the following formula: expansion score + (ICM score × TE score).

Determination of Mosaic Levels

After *in vitro* cultivation, embryos were assigned into different groups according to morphological assessments at 118 hpi, that is TLBMS, embryo expansion, TE quality, and ICM quality. The expanded blastocysts with at least grade B of ICM or at least grade B of TE on day 5 (around 120 hpi) or day 6 (around 140 hpi) were selected for embryo biopsy. Five to eight TE cells were then carefully separated from a blastocyst through micromanipulation by using inverted microscopy with a laser system (Chen et al., 2017). Biopsied cells were rinsed with phosphate-buffered saline first and then placed into an RNAse–DNAse-free polymerase chain reaction tube. The remaining blastocyst was incubated in a tri-gas incubator for ≥3 h and then subjected to cryopreservation by using the Cryotech vitrification method (Cryotech, Tokyo, Japan). The mosaic levels of biopsied samples were determined according to the manufacturer's instructions of hr-NGS, including the SurePlex DNA amplification system (Illumina, San Diego, CA, United States) for the extraction and preparation of genomic DNA of TE cells, the VeriSeq PGS workflow (Illumina) for the preparation of genomic DNA libraries, the VeriSeq DNA Library Prep Kit (Illumina) for the normalization of each DNA library's concentration, and the Miseq System with Miseq Reagent Kit v3 (Illumina) for DNA sequencing of individual libraries. Finally, the generated bioinformatic data were analyzed using BlueFuse Multi software (Illumina), and segmental changes in individual chromosomes were defined as altered segments of ≥10 MB in size. The mosaic level was determined by the percentage of aneuploid cells in a TE biopsy specimen. If the embryo had more than one mosaic chromosomal region, the highest value of mosaic levels was used for analysis (Viotti et al., 2021). The ploidy status of each sample was determined according to the following criteria: 1) mosaic levels ≤20% (euploidy); 2) mosaic levels between >20% and <50% (low-level mosaicism); 3) mosaic levels between ≥50% and ≤80% (high-level mosaicism); and 4) mosaic levels >80% (aneuploidy). The hr-NGS data for all embryos have been deposited to the NCBI SRA database, and the BioProject accession number is PRJNA782420.

TABLE 1 | Patient and cycle characteristics.

No. of patients	**178**
Female age (years)	35.6 ± 4.7
Male age (years)	39.4 ± 6.6
Infertility diagnosis (%)	
Tubal factor	9 (5.1)
Ovulatory	14 (7.9)
Male factor	16 (9.0)
Multiple factors	139 (78.1%)
AMH levels (ng/mL or %)	4.7 ± 3.8
<1.1	8 (4.5)
1.1–2.0	35 (19.7)
>2.0	135 (75.8)
BMI (kg/m^2)	21.8 ± 3.2
PGT-A attempts (%)	
1	171 (96.1)
2	5 (2.8)
3	2 (1.1)
Sperm quality (%)	
Normal	153 (86.0)
Abnormal	25 (14.1)
No. of cycles	187
Type of stimulation protocol (%)	
GnRH antagonist	56 (30.0)
GnRH agonist long	131 (70.1)
Oocytes retrieved	16.8 ± 9.2
Mature oocytes	14.1 ± 7.9
Normal fertilization embryos	10.7 ± 6.0
Blastocysts biopsied	4.9 ± 3.2
Blastocysts analyzed	918
Fertilization methods (%)	
Conventional oocyte insemination	412 (44.9)
Intracytoplasmic sperm injection	506 (55.1)
Oocyte sources (%)	
Autologous	814 (88.7)
Donor	104 (11.3)
Expansion levels (%)	
Level 3	157 (17.1)
Level 2	692 (75.4)
Level 1	66 (7.2)
Level <1	3 (0.3)
ICM grades (%)	
A	200 (21.8)
B	605 (65.9)
C or less	113 (12.3)
TE grades (%)	
A	83 (9.0)
B	546 (59.5)
C or less	289 (31.5)
Biopsy day (%)	
Day 5	595 (64.8)
Day 6	323 (35.2)
PGT-A diagnosis (%)	
Euploidy (%)	320 (34.9)
Low-level mosaicism (%)	242 (26.4)
High-level mosaicism (%)	111 (12.1)
Aneuploidy (%)	245 (26.7)

Data are n (%) or mean ± standard deviation. The abbreviations "AMH", "BMI", "PGT-A", and "GnRH" denoted anti-mullerian hormone, body mass index, preimplantation genetic test for aneuploidy, and gonadotropin-releasing hormone, respectively.

Statistical Analysis

SPSS (v 20.0; IBM Corporation, United States) or Prism (version 6.0 h; GraphPad Software, United States) were used for statistical analysis. $p < 0.05$ was considered significant in all analyses. Generalized estimating equation (GEE) analysis was used to assess the associations between the probability of a specific ploidy status and individual independent variables in unadjusted (univariate) or adjusted (multivariate) models. The ploidy status examined included three thresholds of mosaic levels, that is, ≤20%, <50%, and ≤80%. The receiver operating characteristic (ROC) curve was applied to estimate the prediction performance of TLBMS and determine cutoff values. Finally, the rates of euploidy, low-level mosaicism, high-level mosaicism, and aneuploidy between groups were compared using Fisher's exact test.

RESULTS

The patient- and cycle-characteristics were described in **Table 1**. All morphokinetic and morphological features of biopsied blastocysts ($n = 918$) with known mosaic levels, comprising 320 euploid blastocysts, 242 low-level mosaic blastocysts, 111 high-level mosaic blastocysts, and 245 aneuploid blastocysts, were evaluated using TL images. The potential confounding variables associated with embryo ploidy were analyzed, including factors related to patient (at the start of infertility treatment) or cycle characteristics, that is female age, anti-Müllerian hormone (AMH), body mass index (BMI), male age, oocyte numbers, mature oocyte numbers, ovarian stimulation protocols (agonist long vs. antagonist protocols), oocyte sources (autologous vs. donor oocytes), sperm quality (abnormal vs. normal sperm), and artificial insemination methods (ICSI vs. conventional insemination); factors related to the kinetics of blastocoel formation, that is, tM, tSB, tB, tB-tM, and tB-tSB; factors related to embryo dysmorphisms, that is, DD, DC, RC, ICD, MN2, and MN4; and factors related to blastocyst morphology, that is, expansion levels, ICM quality, and TE quality.

Association Between Patient- or Cycle-Related Confounding Variables and Mosaic Levels

Firstly, this study aimed to identify significant confounding factors derived from non-embryonic sources. The correlation between individual patient confounding factors and mosaic levels was tested through univariate regression analysis by using the GEE model (**Table 2**). Data revealed that female age (odds ratio [OR] = 0.956, 95% confidence interval [CI] = 0.931–0.981, $p < 0.05$) was associated with mosaic levels ≤20%; female age (OR = 0.932, 95% CI = 0.905–0.96, $p < 0.001$), mature oocyte numbers (OR = 1.017, 95% CI = 1–1.033, $p < 0.05$), and oocyte sources (autologous vs. donor oocytes, OR = 0.57, 95% CI = 0.332–0.976, $p < 0.05$) were significantly associated with mosaic levels <50%; and female age (OR = 0.92, 95% CI = 0.884–0.958, $p < 0.001$), oocyte numbers (OR = 1.021, 95% CI = 1.004–1.038, $p < 0.05$), mature oocyte numbers (OR = 1.024, 95% CI = 1.004–1.044, $p < 0.05$), and oocyte sources (autologous vs. donor oocytes, OR = 0.351, 95% CI = 0.157–0.785, $p < 0.05$) were significantly associated with mosaic levels ≤80%. According to the results,

TABLE 2 | The correlations between patient- or cycle-characteristics and embryo ploidy status in the current dataset.

Variables	Mosaic level ≤20% (Euploid)				Mosaic level <50% (Euploid and low-level mosaic)				Mosaic level ≤80% (Non-aneuploid)			
	OR	95% CI		*P*	OR	95% CI		*P*	OR	95% CI		*P*
		Lower	Upper			Lower	Upper			Lower	Upper	
Female age	0.956	0.931	0.981	<0.05	0.932	0.905	0.96	<0.001	0.92	0.884	0.958	<0.001
AMH	0.988	0.943	1.035	NS	0.99	0.962	1.019	NS	0.989	0.957	1.022	NS
BMI	0.981	0.942	1.021	NS	0.975	0.939	1.013	NS	0.98	0.936	1.026	NS
Male age	0.986	0.968	1.005	NS	0.988	0.968	1.009	NS	0.989	0.967	1.011	NS
Oocyte numbers	1.009	0.997	1.021	NS	1.013	0.999	1.027	NS	1.021	1.004	1.038	<0.05
Mature oocyte numbers	1.01	0.996	1.023	NS	1.017	1	1.033	<0.05	1.024	1.004	1.044	<0.05
Agonist long protocol vs. antagonist protocol[a]	1.024	0.735	1.427	NS	0.81	0.568	1.156	NS	0.854	0.581	1.256	NS
Autologous oocytes vs. donor oocytes[a]	0.762	0.543	1.069	NS	0.57	0.332	0.976	<0.05	0.351	0.157	0.785	<0.05
Abnormal sperm vs. normal sperm[a]	1.064	0.733	1.543	NS	0.869	0.615	1.229	NS	0.825	0.527	1.293	NS
ICSI vs. COI[a]	1.045	0.811	1.347	NS	0.94	0.721	1.225	NS	0.928	0.694	1.24	NS

The univariate generalized estimating equation (GEE) analysis in a logistic regression setting was used for statistical analysis. The abbreviations "OR", "CI", "P", "NS", "AMH", "BMI", "ICSI", and "COI" denoted odds ratio, confidence interval, p-value*, not significant, anti-mullerian hormone, body mass index, intracytoplasmic sperm injection, and conventional oocyte insemination, respectively.*

[a] *Indication of a reference group in the GEE, model.*

female age, oocyte numbers, mature oocyte numbers, and oocyte sources were considered to be significant confounding factors from non-embryonic sources.

Correlation Between Adjusted Embryo Variables and Mosaic Levels

This study attempted to evaluate the associations between embryonic variables and ploidy status. Considering the significant patient- and cycle-related confounding factors, that is, female age, mature oocyte numbers, and oocyte sources, the multivariate GEE analysis demonstrated that blastocyst expansion levels (level ≤1 vs. level 3, OR = 0.388, 95% CI = 0.217–0.695, $p < 0.005$; level 2 vs. level 3, OR = 0.488, 95% CI = 0.347–0.688, $p < 0.001$), ICM grades (grade C or less vs. grade A, OR = 0.563, 95% CI = 0.333–0.962, $p < 0.05$), and TE grades (grade C or less vs. grade A, OR = 0.29, 95% CI = 0.178–0.473, $p < 0.001$; grade B vs. grade A, OR = 0.586, 95% CI = 0.39–0.883, $p < 0.05$) were associated with mosaic levels ≤20%. Furthermore, tSB (OR = 0.978, 95% CI = 0.961–0.996, $p < 0.05$), tB (OR = 0.975, 95% CI = 0.96–0.991, $p < 0.005$), tB-tM (OR = 0.957, 95% CI = 0.934–0.98, $p < 0.001$), blastocyst expansion levels (level ≤1 vs. level 3, OR = 0.328, 95% CI = 0.181–0.596, $p < 0.001$; level 2 vs. level 3, OR = 0.48, 95% CI = 0.323–0.712, $p < 0.001$), ICM grades (grade C or less vs. grade A, OR = 0.35, 95% CI = 0.211–0.58, $p < 0.001$; grade B vs. grade A, OR = 0.604, 95% CI = 0.413–0.882, $p < 0.01$), and TE grades (grade C or less vs. grade A, OR = 0.242, 95% CI = 0.143–0.411, $p < 0.001$; grade B vs. grade A, OR = 0.584, 95% CI = 0.358–0.953, $p < 0.05$) were associated with mosaic levels <50%. Finally, tSB (OR = 0.977, 95% CI = 0.957–0.997, $p < 0.05$), tB (OR = 0.973, 95% CI = 0.955–0.991, $p < 0.005$), tB-tM (OR = 0.942, 95% CI = 0.918–0.966, $p < 0.001$), tB-tSB (OR = 0.963, 95% CI = 0.928–0.999, $p < 0.05$), MN4 (MN4 vs. non-MN4, OR = 1.655, 95% CI = 1.004–2.729, $p < 0.05$), blastocyst expansion levels (level ≤1 vs. level 3, OR = 0.343, 95% CI = 0.179–0.657, $p < 0.005$; level 2 vs. level 3, OR = 0.523, 95% CI = 0.327–0.835, $p < 0.01$), ICM grades (grade C or less vs. grade A, OR = 0.497, 95% CI = 0.274–0.9, $p < 0.05$), and TE grades (grade C or less vs. grade A, OR = 0.3, 95% CI = 0.162–0.554, $p < 0.001$) were associated with mosaic levels ≤80% (**Table 3**). The results indicated that the grading of blastocyst morphology components were positively associated with all of threshold levels of mosaicism. However, the variables of blastulation kinetics, such as tB, tSB, tB-tM, and tB-tSB, were negatively associated with the mosaic level <50% or ≤80%, and the variable of embryo dysmorphisms, i.e., MN4, had a positive association with the mosaic level ≤80%. Additionally, when incorporating the blastocyst morphology component (e.g., expansion levels, ICM grades, or TE grades) with tB and MN4 in the multivariate logistic regression models, the associations between tB and embryo ploidy were attenuated to be non-significant and MN4 was merely correlated with the mosaic level ≤80% (OR = 1.703–2.16, $p < 0.05$) (**Supplementary Tables S3–S5**). The non-aneuploid (mosaic level ≤80%) rate of the blastocysts with MN4 (81.5%) was higher than the blastocysts without MN4 (72.2%) ($p < 0.05$), which resulted from the significantly raised rate of high-level mosaicism in the MN4 group (19.4%) (**Supplementary Figure S2A**). By contrast, the blastocyst morphology components still had the statistically significant correlations with mosaic levels ≤20% (OR = 0.225–0.528), <50% (OR = 0.204–0.614) or ≤80% (OR = 0.261–0.554), indicating this redefined blastocyst morphology could be capable of categorizing the embryos with different mosaic levels (**Supplementary Figure S3** and **Supplementary Tables S3–S5**).

Effectiveness of Time Lapse-Based Blastocyst Morphological Score as a Predictor for Embryo Ploidy

In order to provide a comprehensive inspection of blastocyst morphology, this study tried to convert the grading of morphological components into TLBMSs and investigated its correlations with different mosaic levels. After adjustment of patient- and cycle-related confounding factors (i.e., female age, mature oocyte numbers, and oocyte sources) as well as the

TABLE 3 | The correlations between embryonic variables and embryo ploidy status in consideration of patient- and cycle-confounding factors.

Variables	Mosaic level ≤20% (Euploid)				Mosaic level <50% (Euploid and low-level mosaic)				Mosaic level ≤80% (Non-aneuploid)			
	[a]OR	95% CI		*P*	[a]OR	95% CI		*P*	[a]OR	95% CI		*P*
		Lower	Upper			Lower	Upper			Lower	Upper	
Kinetics of blastocyst formation												
tM	0.996	0.979	1.013	NS	1.001	0.984	1.018	NS	1.008	0.989	1.028	NS
tSB	0.99	0.97	1.009	NS	0.978	0.961	0.996	<0.05	0.977	0.957	0.997	<0.05
tB	0.983	0.965	1.001	NS	0.975	0.96	0.991	<0.005	0.973	0.955	0.991	<0.005
tB-tM	0.978	0.954	1.002	NS	0.957	0.934	0.98	<0.001	0.942	0.918	0.966	<0.001
tB-tSB	0.962	0.925	1	NS	0.969	0.934	1.006	NS	0.963	0.928	0.999	<0.05
Embryo dysmorphisms												
DD vs. non-DD[b]	0.768	0.357	1.654	NS	0.521	0.239	1.134	NS	0.702	0.341	1.445	NS
DC vs. non-DC[b]	1.25	0.693	2.254	NS	1.168	0.605	2.257	NS	1.926	0.765	4.852	NS
RC vs. non-RC[b]	0.794	0.38	1.658	NS	0.805	0.41	1.578	NS	1.298	0.591	2.854	NS
ICD vs. non-ICD[b]	0.795	0.240	2.632	NS	1.329	0.33	5.359	NS	2.896	0.628	13.359	NS
MN2 vs. non-MN2[b]	0.895	0.652	1.230	NS	0.898	0.648	1.244	NS	1.433	0.971	2.114	NS
MN4 vs. non-MN4[b]	1.066	0.682	1.668	NS	1.035	0.679	1.578	NS	1.655	1.004	2.729	<0.05
Blastocyst morphology assessments												
Expansion level ≤1 vs. 3[b]	0.388	0.217	0.695	<0.005	0.328	0.181	0.596	<0.001	0.343	0.179	0.657	<0.005
Expansion level 2 vs. 3[b]	0.488	0.347	0.688	<0.001	0.48	0.323	0.712	<0.001	0.523	0.327	0.835	<0.01
ICM ≤ C vs. ICM A[b]	0.563	0.333	0.962	<0.05	0.35	0.211	0.58	<0.001	0.497	0.274	0.9	<0.05
ICM B vs. ICM A[b]	0.795	0.566	1.116	NS	0.604	0.413	0.882	<0.01	0.689	0.434	1.095	NS
TE ≤ C vs. TE A[b]	0.29	0.178	0.473	<0.001	0.242	0.143	0.411	<0.001	0.3	0.162	0.554	<0.001
TE B vs. TE A[b]	0.586	0.39	0.883	<0.05	0.584	0.358	0.953	<0.05	0.614	0.34	1.111	NS

The multivariate generalized estimating equation (GEE) analysis in a logistic regression setting was used for statistical analysis. The abbreviations "OR", "CI", "P", and "NS" denoted odds ratio, confidence interval, p-value, *and not significant, respectively. Other morphokinetic and morphological abbreviations were described in the* ***Supplementary Table S1****.*

[a] *Indicating the adjusted OR, by female age, mature oocyte numbers, and oocyte sources.*

[b] *Indicating a reference group in the GEE model.*

TABLE 4 | Evaluation of the relationship with embryo ploidy status and prediction abilities of time-lapse blastocyst morphology scores.

Ploidy status	[a]OR	95% CI		*P*	AUC	95% CI		*P*
		Lower	Upper			Lower	Upper	
Mosaic level ≤20% (Euploid)	1.326	1.187	1.481	<0.001	0.604	0.565	0.642	<0.001
Mosaic level <50% (Euploid and low-level mosaic)	1.425	1.262	1.608	<0.001	0.634	0.598	0.671	<0.001
Mosaic level ≤80% (Non-aneuploid)	1.351	1.186	1.539	<0.001	0.617	0.576	0.658	<0.001

The multivariate generalized estimating equation (GEE) analysis in a logistic regression setting and receiver-operating characteristic analysis were used for statistical analysis. The abbreviations "OR", "CI", "AUC", and "P" denoted odds ratio, confidence interval, area under the curve, and p-value, *respectively.*

[a] *Indicating the adjusted OR, by female age, mature oocyte numbers, oocyte sources, MN4, and tB.*

embryo features except for blastocyst morphology, i.e., tB and MN4, the results revealed that TLBMS was significantly associated with mosaic levels ≤20% (OR = 1.326, 95% CI = 1.187–1.481, $p < 0.001$), <50% (OR = 1.425, 95% CI = 1.262–1.608, $p < 0.001$), and ≤80% (OR = 1.351, 95% CI = 1.186–1.539, $p < 0.001$) (**Table 4**). The ROC curve demonstrated that TLBMS was statistically associated with mosaic levels. The values of area under the curve (AUC) were 0.604, 0.634, and 0.617 for mosaic levels ≤20%, <50%, and ≤80%, respectively (**Table 4**). The cutoff values of TLBMS and female age were 6 and 36 years, respectively. Embryos were classified into four group according to cutoff values. Between the groups with TLBMSs ≥6 and <6, the rates of mosaic levels ≤20% (41.4 vs. 30.3% in the age group of <36 years, $p < 0.05$; 39 vs. 16.8% in the age group of ≥36 years, $p < 0.0001$), mosaic levels <50% (72.8 vs. 55.1% in the age group of <36 years, $p < 0.001$; 63.7 vs. 35.5% in the age group of ≥36 years, $p < 0.0001$), and mosaic levels ≤80% (83.4 vs. 72.4% in the age group of <36 years, $p < 0.01$; 72.9 vs. 51% in the age group of ≥36 years, $p < 0.0001$) appeared to be different (**Figure 1**). Moreover, this study further combined TLBMSs with the occurrence of MN4 for embryo selection. The results demonstrated that a higher rate of high-level mosaicism was observed in the good morphology blastocysts (TLBMSs ≥6) with MN4 (20%) than the good morphology blastocysts without MN4 (9.3%) ($p < 0.05$), suggesting deselection of MN4 embryos could be considered to reduce the risk of selecting good morphology blastocysts with high-level mosaicism (**Supplementary Figure S2D**).

DISCUSSION

As mentioned in a literature review, the diploid–aneuploid mosaicism in an embryo can originate from paternal, maternal, and external factors (Taylor et al., 2014). Firstly, this retrospective study was proposed to identify the possible

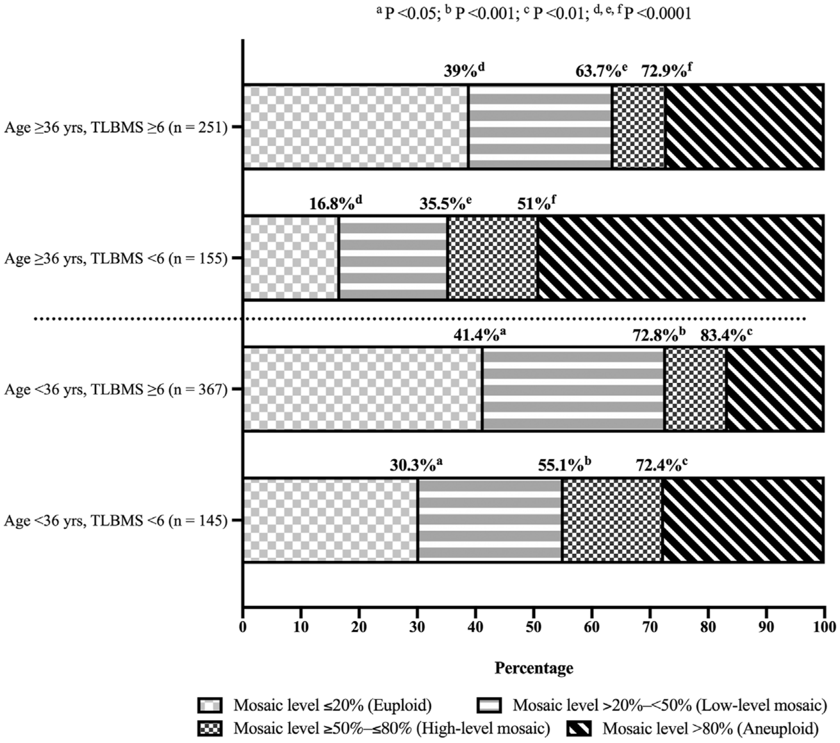


FIGURE 1 | Distribution of blastocysts with different mosaic levels according to time-lapse blastocyst morphology scores. Fisher exact test was used for statistical analysis and women were stratified into age <36 years and ≥36 years groups. Same superscript letters indicated statistically significant. Abbreviations "n", "yrs", and "TLBMS" denoted the number of embryos, years, and time-lapse blastocyst morphology score, respectively.

confounding factors from the non-embryonic source. In our dataset, the univariate GEE demonstrated that several patient or cycle characteristics could be the confounding factors of mosaic levels, that is, female age, oocyte sources, and numbers of oocytes and mature oocytes, in addition to embryo parameters (**Table 2**). Several reports have shown that maternal age is considerably associated with oocyte aneuploidy incidence (Jones 2008; Gianaroli et al., 2010; Greaney et al., 2018; Gruhn et al., 2019). Gruhn et al. (2019) demonstrated that the proportion of aneuploid oocytes follows a U-shaped curve according to female age, in which the risk of meiotic segregation errors is decreased in the 20–32 years age group (Gruhn et al., 2019). Similarly, the current study revealed that female age (23–44 years) was the major confounding factor, which was negatively associated with mosaic levels at all three thresholds (≤20%, <50%, or ≤80%). Furthermore, advanced age was one of the reasons for the association of autologous oocytes and reduced oocyte numbers with increased mosaic levels in our dataset. Taking female age into account revealed that oocyte numbers and oocyte sources were no longer significantly associated with ploidy status (**Supplementary Table S6**). Schaeffer et al. (2020) analyzed aneuploidy rates and IVF outcomes by comparing patient and donor, showing that with increasing maternal age there are more substantial differences in aneuploidy rates between the patient and the donor embryos. Moreover, Irani et al. (2020) reveals that euploidy rates are similar between groups with different numbers of retrieved oocytes after age stratification. Notably, the present study showed that male factors, that is, male age and semen quality, were not significantly related to mosaic levels. These results are supported by the findings of a previous study that used young donor oocytes for analysis, which revealed that advanced paternal age was not associated with higher aneuploidy rates (Dviri et al., 2021). Moreover, as per blastocysts obtained, sperm factors did not exert significant effects on the euploidy rate, although using poor-quality sperm for insemination appears to reduce fertilization and blastocyst formation rates (Mazzilli et al., 2017). Furthermore, univariate GEE analysis demonstrated that other patient or cycle variables evaluated in this study, such as BMI, AMH, ovarian stimulation protocols, and artificial insemination methods, were not significantly correlated with embryo ploidy; these findings were similar to previous reports (Deng et al., 2020; Irani et al., 2020; Cascales et al., 2021; Pipari et al., 2021; Stovezky et al., 2021).

This study was designed to assess the importance of embryonic characteristics in embryo ploidy. Considering confounding variables from nonembryonic factors, that is, maternal age, mature oocyte numbers, and oocyte sources, the present study revealed that delayed blastocyst formation was positively associated with increased mosaic levels (**Table 3**). Two studies, which applied TL monitoring and hr-NGS simultaneously for PGT-A cycles, observed that aneuploid blastocysts exhibited a significant delay in tSB or tB as compared with euploid blastocysts (Lee et al., 2019; Martin et al., 2021). Multivariate GEE analysis in this study further demonstrated a meaningful correlation of tSB and tB with mosaic levels when the threshold was set at 50%. Correspondingly, prolonged blastulation intervals (tB-tM or tB-tSB) accompanied with mosaic levels ≥50% and >80% were present. However, this correlation was not detected between blastocyst kinetics and euploidy (mosaic levels ≤20%). A consistent observation with our previous study (Lee et al., 2019) revealed that blastulation kinetics of low-level mosaic blastocysts are comparable with those of euploid blastocysts (**Supplementary Table S7**). When low-level mosaic blastocysts were included in the reference category (mosaic level >20%) of a binary logistic regression, the associations of blastocyst kinetics with mosaic level ≤20% appeared to be non-significant.

On the other hand, considering patient- and cycle-related confounding factors, the results of this study indicated that the correlation of individual developmental dysmorphisms with embryo ploidy was not identified, except for MN4 (**Table 3**). Desai et al. (2018) reported a similar influence of developmental dysmorphisms and concluded that individual dysmorphisms, including MN, RC, DC, and ICD, are not associated with euploidy (Desai et al., 2018). In their study, MN embryos were examined at two-to five-cell stages, with the pooling of MN2 and MN4 embryos for euploidy analysis by using aCGH or hr-NGS, which might lead to an inconsistent result for MN embryos as compared with our study. A mouse model was used to prove that MN occurrence at the early developmental stage affects blastomere ploidy and compromises blastocyst developmental potential. Notably, after MN-derived blastocysts were transferred to surrogate mothers, the pregnancy loss rate was not significantly increased by the adverse effects of early chromosome segregation errors (Mashiko et al., 2020). A drastic decrease in MN incidence from two-to four-cell stages and altered morphokinetic features, such as delayed timings of early divisions and prolonged periods of two-cell and four-cell stages, support that a repair mechanism exists for self-correction of MN embryos during early cell divisions (Balakier et al., 2016). According to our data, in biopsied blastocysts, the MN incidence was decreased from the two-cell stage (23.2%, 213/918) to the four-cell stage (11.8%, 108/918), suggesting a repair mechanism might exist during early cell divisions (**Supplementary Table S7**). However, if the early chromosome segregation errors continued at the four-cell stage, self-correction of MN embryos might not be completed in time, resulting in the high potential of high-level mosaicism (**Supplementary Figure S2** and **Supplementary Table S7**).

The TL incubation system is designed to enable four-dimension-like observation, which offers a comprehensive and consistent inspection of embryo morphology and dynamics. One of the aims of this study was to modify the morphological evaluation method of a blastocyst at a specific time window (118 hpi) based on traditional assessments (Gardner et al., 2000) and TL monitoring requirements (Kragh et al., 2019). In our dataset, the fertilization methods (ICSI vs. conventional insemination) appeared to have non-significant associations with the timing of a blastocyst with a cavity beginning to form (OR = 1.338, 95% CI = 0.362–4.946, $p > 0.05$) and the timing of a blastocyst with a cavity beginning to expand (OR = 1.46, 95% CI = 0.362–5.891, $p > 0.05$). The same observation time-point was therefore applied for the blastocysts derived from ICSI or conventional insemination. The annotations of ICM and TE were conducted more thoroughly to generalize the Gardner blastocyst grading system to time-lapse imaging (Kragh et al., 2019). The annotations of expansion levels emphasized the specific developmental features, including an embryo with the cavity beginning to form (level 1), to expand (level 2), and to herniate (level 3), which were difficult to determine with the conventional method (**Supplementary Figure S1**). This study found that TL-based morphological components are still considerably associated with mosaic levels at all thresholds (**Table 3**). Unlike in previous studies (Capalbo et al., 2014; Minasi et al., 2016; Wang et al., 2018), the expansion levels of blastocysts in this study were evaluated at a uniform time window (118 hpi), which could be considered an evaluation of the developmental speed; blastocysts with expansion levels ≤1 (OR = 0.388, 0.328, and 0.343) and 2 (OR = 0.488, 0.480, and 0.523) had significantly lower probabilities of mosaic levels ≤20%, <50%, and ≤80% compared with blastocysts with expansion level 3. The expansion results of TL-based assessments not only revealed the correlation of the blastocyst expansion status with embryo ploidy (Minasi et al., 2016; Wang et al., 2018), but also reflected the importance of the developmental speed on blastocyst formation with lower mosaic levels, which had been evidenced by morphokinetic observations in the current and previous studies (Campbell et al., 2013a; Mumusoglu et al., 2017). On the other hand, in accordance with previous studies (Fragouli et al., 2014; Minasi et al., 2016; Wang et al., 2018), this study indicated a greater likelihood of euploidy among blastocysts with good quality ICM or TE. However, in previous studies, under-sensitive PGT-A approaches and unfixed periods of ICM and TE assessments, for example, D5, D6, or even D7, may result in inconsistent findings. Therefore, in our study, multivariate GEE analysis was conducted to evaluate the associations of ICM and TE quality, which was assessed at the specific time window of cultivation, with embryo ploidy using hr-NGS. In this circumstance, as compared with blastocysts with grade A of ICM or TE, the blastocysts with grade C or less of ICM or TE had a higher risk to be non-euploid (>20%), high-level mosaic (≥50%), and aneuploid (>80%) embryos. Similar risks of non-euploidy or aneuploidy were found between blastocysts with grade A and grade B of ICM, but the blastocysts with grade B of ICM were still at an increased risk of mosaic levels ≥50%. Moreover, similar risks were only observed in aneuploidy between blastocysts with grade A and grade B of TE. These results suggest that the morphology of TE and ICM influenced, at least, the discrimination between mosaic levels <50% and

≥50% among blastocysts with moderate to high quality of ICM or TE (**Table 3**). This study thus demonstrated that all the three refined morphology criteria of a blastocyst were meaningfully correlated with embryo ploidy (**Supplementary Figure S3**).

In order to make an effort to improve intuitive decision-making in clinical practice, the current study designed a TLBMS formula as a ranking tool of blastocyst morphology. Furthermore, patient- and cycle-related factors, that is, female age, mature oocyte numbers, and oocyte sources, and significant embryo factors, that is, tB and MN4, were considered to be confounding factors and adjusted in GEE analysis, which revealed an apparent correlation of TLBMS with embryo ploidy (**Table 4**). Previously, at least two functional formulae were available, which used multiplication or addition for the calculation of parameters from expansion levels, ICM grades, and TE grades to convert blastocyst morphology grades into numbers (Rehman et al., 2007; Zhan et al., 2020). Compared with our study, Rehman et al. (2007) used the multiplication formula to generate a wide range of blastocyst quality scores (BQSs = 1–54, a total of 17 scores), leading to the nonlinear correlation with pregnancy outcomes. The blastocyst biopsy day was not considered in the formula for BQS, and the correlation of BQS with embryo ploidy was not investigated (Rehman et al., 2007). Zhan et al. (2020) further improved the formula by using addition instead of multiplication for calculation, which narrows down the score range (score 3–15, a total of 13 scores) and ranks blastocysts considering the biopsy day. Each grade of the expansion status, ICM quality, and TE quality was assigned a score based on the implantation rate, resulting in a linear relationship between blastocyst morphology scores and euploidy rates and acceptable prediction abilities (AUC = 0.683–0.698) (Zhan et al., 2020). However, this updated BMS uses a complicated grading system for expansion (nine levels), ICM (seven grades), and TE (seven grades), which may increase participant bias by embryologists. Inconsistent results of the embryo ploidy status might appear when different PGT-A platforms were used in a study (Zhan et al., 2020). Applying this scoring system to our dataset revealed slightly lower ploidy prediction abilities (AUC = 0.585 for mosaic levels ≤20%, AUC = 0.628 for mosaic levels <50%, and AUC = 0.613 for mosaic levels ≤80%; data not shown). Distinctly, our TLBMS considering two separated parts of blastocyst morphology, that is, the expansion level and ICM/TE quality, used a simplified formula with a combination of addition and multiplication; a score from expansion levels with distinct definition in TL monitoring (four scores) and a score from ICM and TE combinations (four scores) were added. The expansion levels were considered to be the major part of our scoring system. Owing to only three expansion levels in this study, the scores differed by 2 each time to enlarge the gap between two continues levels. The blastocysts with the expansion level <2 would only had the scores from expansion levels and the minimum score of 1 was thus used for the expansion level of <1 to avoid a given total score of TLBMS to be zero. Additionally, this study revealed that the blastocysts with a low-quality ICM or a low-quality TE had reduced probabilities to be euploid, euploid and low-level mosaic, or non-aneuploid blastocysts. A score of zero was therefore assigned to the grade C ICM and the grade C TE. The combination score, which was generated by multiplication of a ICM score by a TE score, was used to represent the ICM/TE quality. Consequently, the blastocysts with the ICM of grade C or the TE of grade C (i.e., AC, BC, CB, and CA) would have a score of zero, the BB quality blastocysts had the score of 1, the AB or BA quality blastocysts had the score of 2, and the AA quality blastocysts had the best score of 4 (**Supplementary Table S2**). As compared with previous studies, TLBMS (score 1–11, a total of 9 scores) provided comparable or even better differentiation capabilities for different thresholds of mosaic levels (AUC = 0.604–0.634; **Table 4**). Notably, when incorporating TLBMSs with the occurrence of MN4 for embryo selection, this study revealed the high-level mosaicism rate of good morphology blastocysts (TLBMSs ≥6) with MN4 was higher than those without MN4 (**Supplementary Figure S2**). Compared with low-level mosaic blastocysts, previous studies have demonstrated high-level mosaic blastocysts have adverse pregnancy outcomes (e.g., decreased rates of implantation, ongoing pregnancy, and birth) (Viotti et al., 2021) and an increased miscarriage rate (Lin et al., 2020). In order to reduce the possibility of selecting blastocysts with high-level mosaicism, this study thus suggested to lower the priority of good morphology blastocysts with MN4 for embryo selection.

This single-center study had several limitations. The major limitation was its retrospective nature, which lacked randomization and might have resulted in the risk of selection bias. Therefore, this study collected biopsied blastocysts from female patients with a wide age range (23–44 years) for the analysis and took the possible confounding factors into account. Although the sample size of this study (918 blastocysts from 178 patients) was relatively small as compared with a previous study (Zhan et al., 2020), the GEE model was successful applied for analysis of repeated measurements. In comparison with a previous investigation with a large dataset (Munne et al., 2017), a higher mosaicism rate was revealed in this study (38.5 versus 21.8%), which could result from different clinical and laboratory settings between infertility centers (Sachdev et al., 2017; Lee et al., 2019). Even though the mosaicism detected by NGS was applied in donor oocyte cycles, the rates between individual infertility centers still range widely from 17 to 47% (Sachdev et al., 2017). Further investigation would be conducted to evaluate the effects of clinical settings and laboratory techniques on embryonic mosaicism. Multilayered TL images were evaluated at 118 hpi to reach a consistent observation regarding blastocyst morphology, but the implementation of TL technology might increase the financial cost and need specific training. Moreover, the current method was established based on our protocols. Further considerations on the clinical settings of individual IVF laboratories were required prior to application of TLBMS. Nevertheless, the present study was designed to determine the association of blastocyst morphology with embryo ploidy. The derived TLBMS as a ranking tool was positively associated with mosaic levels ≤20%, <50%, and ≤80% ($p < 0.001$) (**Table 4**), which could potentially select a blastocyst with a better ploidy status for both young (<36 years) and older (≥36 years) women undergoing infertility treatment (**Figure 1**). Embryonic factors, such as developmental potency and embryo ploidy, have major influences on the following successful pregnancy. Our previous study demonstrated that the improved implementation rate of the euploid embryo is accompanied by an increased score of the KIDScore D5 algorithm, suggesting the potential to discriminate euploid blastocysts with different developmental competence (Lee et al., 2019). The clinical application of TLBMS could be advised as follows: 1) select the embryo (e.g., TLBMS ≥6): with a relatively low risk of aneuploidy for PGT-A; 2) reduce the biopsy number if the patient demanded reduction in the financial burden; 3) directly combine the KIDScore D5 algorithm in non-PGT cycles; and 4) enhance the performance of noninvasive ploidy evaluation, such as the artificial intelligence approach (Lee et al., 2021).

CONCLUSION

Time-lapse assessments at the uniform time-point demonstrate significant correlations of embryonic variables derived from blastulation kinetics, cleavage dysmorphisms, and blastocyst morphology with ploidy status. However, only the morphological variables of a blastocyst (i.e., expansion levels, ICM grades, and TE grades) are meaningfully associated with all of the threshold levels of mosaicism. The TLBMS that is converted from blastocyst morphological components appears to be negatively correlated with aberrant ploidy status but the predictive performance remains limited.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: http://www.ncbi.nlm.nih.gov/bioproject/ under the accession number PRJNA782420

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Chung Sun Medical University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

M-SL and T-HL formulated this study. C-HC, C-IL, C-CH, H-HC, S-TH, E-HC, C-IC and P-YL collected and processed the data. C-HC, C-IL, and H-HC carried out analyses. C-HC and C-IL wrote the manuscript. All authors reviewed the manuscript and provided editorial feedback.

FUNDING

This study was generously supported by grants (NSC 101-2314-B-040-007 and MOST-110-2314-B-040-005) from the Ministry of Science and Technology, Executive Yuan, Taiwan, Republic of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.783826/full#supplementary-material

REFERENCES

Alfarawati, S., Fragouli, E., Colls, P., Stevens, J., Gutiérrez-Mateo, C., Schoolcraft, W. B., et al. (2011). The Relationship between Blastocyst Morphology, Chromosomal Abnormality, and Embryo Gender. *Fertil. Sterility* 95, 520–524. doi:10.1016/j.fertnstert.2010.04.003

Balakier, H., Sojecki, A., Motamedi, G., and Librach, C. (2016). Impact of Multinucleated Blastomeres on Embryo Developmental Competence, Morphokinetics, and Aneuploidy. *Fertil. Sterility* 106, 608–614. doi:10.1016/j.fertnstert.2016.04.041

Barash, O. O., Ivani, K. A., Willman, S. P., Rosenbluth, E. M., Wachs, D. S., Hinckley, M. D., et al. (2017). Association between Growth Dynamics, Morphological Parameters, the Chromosomal Status of the Blastocysts, and Clinical Outcomes in IVF PGS Cycles with Single Embryo Transfer. *J. Assist. Reprod. Genet.* 34, 1007–1016. doi:10.1007/s10815-017-0944-0

Basile, N., Nogales, M. d. C., Bronet, F., Florensa, M., Riqueiros, M., Rodrigo, L., et al. (2014). Increasing the Probability of Selecting Chromosomally normal Embryos by Time-Lapse Morphokinetics Analysis. *Fertil. Sterility* 101, 699–704. doi:10.1016/j.fertnstert.2013.12.005

Campbell, A., Fishel, S., Bowman, N., Duffy, S., Sedler, M., and Hickman, C. F. L. (2013a). Modelling a Risk Classification of Aneuploidy in Human Embryos Using Non-invasive Morphokinetics. *Reprod. BioMedicine Online* 26, 477–485. doi:10.1016/j.rbmo.2013.02.006

Campbell, A., Fishel, S., Bowman, N., Duffy, S., Sedler, M., and Thornton, S. (2013b). Retrospective Analysis of Outcomes after IVF Using an Aneuploidy Risk Model Derived from Time-Lapse Imaging without PGS. *Reprod. BioMedicine Online* 27, 140–146. doi:10.1016/j.rbmo.2013.04.013

Capalbo, A., Rienzi, L., Cimadomo, D., Maggiulli, R., Elliott, T., Wright, G., et al. (2014). Correlation between Standard Blastocyst Morphology, Euploidy and Implantation: an Observational Study in Two Centers Involving 956 Screened Blastocysts. *Hum. Reprod.* 29, 1173–1181. doi:10.1093/humrep/deu033

Cascales, A., Lledó, B., Ortiz, J. A., Morales, R., Ten, J., Llácer, J., et al. (2021). Effect of Ovarian Stimulation on Embryo Aneuploidy and Mosaicism Rate. *Syst. Biol. Reprod. Med.* 67, 42–49. doi:10.1080/19396368.2020.1850908

Chen, H.-H., Huang, C.-C., Cheng, E.-H., Lee, T.-H., Chien, L.-F., and Lee, M.-S. (2017). Optimal Timing of Blastocyst Vitrification after Trophectoderm Biopsy for Preimplantation Genetic Screening. *PLoS One* 12, e0185747. doi:10.1371/journal.pone.0185747

Ciray, H. N., Campbell, A., Agerholm, I. E., Aguilar, J., Chamayou, S., Esbert, M., et al. (2014). Proposed Guidelines on the Nomenclature and Annotation of Dynamic Human Embryo Monitoring by a Time-Lapse User Group. *Hum. Reprod.* 29, 2650–2660. doi:10.1093/humrep/deu278

Del Carmen Nogales, M., Bronet, F., Basile, N., Martínez, E. M., Liñán, A., Rodrigo, L., et al. (2017). Type of Chromosome Abnormality Affects Embryo Morphology Dynamics. *Fertil. Sterility* 107, 229–235. doi:10.1016/j.fertnstert.2016.09.019

Deng, J., Kuyoro, O., Zhao, Q., Behr, B., and Lathi, R. B. (2020). Comparison of Aneuploidy Rates between Conventional *In Vitro* Fertilization and Intracytoplasmic Sperm Injection in *In Vitro* Fertilization-Intracytoplasmic Sperm Injection Split Insemination Cycles. *F&S Rep.* 1, 277–281. doi:10.1016/j.xfre.2020.07.006

Desai, N., Goldberg, J. M., Austin, C., and Falcone, T. (2018). Are Cleavage Anomalies, Multinucleation, or Specific Cell Cycle Kinetics Observed with Time-Lapse Imaging Predictive of Embryo Developmental Capacity or Ploidy? *Fertil. Sterility* 109, 665–674. doi:10.1016/j.fertnstert.2017.12.025

Dviri, M., Madjunkova, S., Koziarz, A., Madjunkov, M., Mashiach, J., Nekolaichuk, E., et al. (2021). Is There an Association between Paternal Age and Aneuploidy? Evidence from Young Donor Oocyte-Derived Embryos: a Systematic Review and Individual Patient Data Meta-Analysis. *Hum. Reprod. Update* 27, 486–500. doi:10.1093/humupd/dmaa052

Fragouli, E., Alfarawati, S., Spath, K., and Wells, D. (2014). Morphological and Cytogenetic Assessment of Cleavage and Blastocyst Stage Embryos. *Mol. Hum. Reprod.* 20, 117–126. doi:10.1093/molehr/gat073

Friedenthal, J., Maxwell, S. M., Munné, S., Kramer, Y., Mcculloh, D. H., Mccaffrey, C., et al. (2018). Next Generation Sequencing for Preimplantation Genetic Screening Improves Pregnancy Outcomes Compared with Array Comparative Genomic Hybridization in Single Thawed Euploid Embryo Transfer Cycles. *Fertil. Sterility* 109, 627–632. doi:10.1016/j.fertnstert.2017.12.017

Gardner, D. K., Lane, M., Stevens, J., Schlenker, T., and Schoolcraft, W. B. (2000). Blastocyst Score Affects Implantation and Pregnancy Outcome: towards a Single Blastocyst Transfer. *Fertil. Sterility* 73, 1155–1158. doi:10.1016/s0015-0282(00)00518-5

Gianaroli, L., Magli, M. C., Cavallini, G., Crippa, A., Capoti, A., Resta, S., et al. (2010). Predicting Aneuploidy in Human Oocytes: Key Factors Which Affect the Meiotic Process. *Hum. Reprod.* 25, 2374–2386. doi:10.1093/humrep/deq123

Greaney, J., Wei, Z., and Homer, H. (2018). Regulation of Chromosome Segregation in Oocytes and the Cellular Basis for Female Meiotic Errors. *Hum. Reprod. Update* 24, 135–161. doi:10.1093/humupd/dmx035

Greco, E., Minasi, M. G., and Fiorentino, F. (2015). Healthy Babies after Intrauterine Transfer of Mosaic Aneuploid Blastocysts. *N. Engl. J. Med.* 373, 2089–2090. doi:10.1056/NEJMc1500421

Gruhn, J. R., Zielinska, A. P., Shukla, V., Blanshard, R., Capalbo, A., Cimadomo, D., et al. (2019). Chromosome Errors in Human Eggs Shape Natural Fertility over Reproductive Life Span. *Science* 365, 1466–1469. doi:10.1126/science.aav7321

Irani, M., Canon, C., Robles, A., Maddy, B., Gunnala, V., Qin, X., et al. (2020). No Effect of Ovarian Stimulation and Oocyte Yield on Euploidy and Live Birth Rates: an Analysis of 12 298 Trophectoderm Biopsies. *Hum. Reprod.* 35, 1082–1089. doi:10.1093/humrep/deaa028

Jones, K. T. (2008). Meiosis in Oocytes: Predisposition to Aneuploidy and its Increased Incidence with Age. *Hum. Reprod. Update* 14, 143–158. doi:10.1093/humupd/dmm043

Kane, S. C., Willats, E., Bezerra Maia e Holanda Moura, S., Hyett, J., and Da Silva Costa, F. (2016). Pre-Implantation Genetic Screening Techniques: Implications for Clinical Prenatal Diagnosis. *Fetal Diagn. Ther.* 40, 241–254. doi:10.1159/000449381

Kragh, M. F., Rimestad, J., Berntsen, J., and Karstoft, H. (2019). Automatic Grading of Human Blastocysts from Time-Lapse Imaging. *Comput. Biol. Med.* 115, 103494. doi:10.1016/j.compbiomed.2019.103494

Lee, C.-I., Chen, C.-H., Huang, C.-C., Cheng, E.-H., Chen, H.-H., Ho, S.-T., et al. (2019). Embryo Morphokinetics Is Potentially Associated with Clinical Outcomes of Single-Embryo Transfers in Preimplantation Genetic Testing for Aneuploidy Cycles. *Reprod. BioMedicine Online* 39, 569–579. doi:10.1016/j.rbmo.2019.05.020

Lee, C.-I., Chen, H.-H., Huang, C.-C., Chen, C.-H., Cheng, E.-H., Huang, J. Y., et al. (2020a). Effect of Interval between Human Chorionic Gonadotropin Priming and Ovum Pick-Up on the Euploid Probabilities of Blastocyst. *Jcm* 9, 1685. doi:10.3390/jcm9061685

Lee, C.-I., Cheng, E.-H., Lee, M.-S., Lin, P.-Y., Chen, Y.-C., Chen, C.-H., et al. (2020b). Healthy Live Births from Transfer of Low-Mosaicism Embryos after Preimplantation Genetic Testing for Aneuploidy. *J. Assist. Reprod. Genet.* 37, 2305–2313. doi:10.1007/s10815-020-01876-6

Lee, C.-I., Su, Y.-R., Chen, C.-H., Chang, T. A., Kuo, E. E.-S., Zheng, W.-L., et al. (2021). End-to-end Deep Learning for Recognition of Ploidy Status Using Time-Lapse Videos. *J. Assist. Reprod. Genet.* 38, 1655–1663. doi:10.1007/s10815-021-02228-8

Lin, P.-Y., Lee, C.-I., Cheng, E.-H., Huang, C.-C., Lee, T.-H., Shih, H.-H., et al. (2020). Clinical Outcomes of Single Mosaic Embryo Transfer: High-Level or Low-Level Mosaic Embryo, Does it Matter? *Jcm* 9, 1695. doi:10.3390/jcm9061695

Martín, Á., Rodrigo, L., Beltrán, D., Meseguer, M., Rubio, C., Mercader, A., et al. (2021). The Morphokinetic Signature of Mosaic Embryos: Evidence in Support of Their Own Genetic Identity. *Fertil. Sterility* 116, 165–173. doi:10.1016/j.fertnstert.2020.12.031

Mashiko, D., Ikeda, Z., Yao, T., Tokoro, M., Fukunaga, N., Asada, Y., et al. (2020). Chromosome Segregation Error during Early Cleavage in Mouse Pre-implantation Embryo Does Not Necessarily Cause Developmental Failure after Blastocyst Stage. *Sci. Rep.* 10, 854. doi:10.1038/s41598-020-57817-x

Mazzilli, R., Cimadomo, D., Vaiarelli, A., Capalbo, A., Dovere, L., Alviggi, E., et al. (2017). Effect of the Male Factor on the Clinical Outcome of Intracytoplasmic Sperm Injection Combined with Preimplantation Aneuploidy Testing: Observational Longitudinal Cohort Study of 1,219 Consecutive Cycles. *Fertil. Sterility* 108, 961–972. doi:10.1016/j.fertnstert.2017.08.033

Minasi, M. G., Colasante, A., Riccio, T., Ruberti, A., Casciani, V., Scarselli, F., et al. (2016). Correlation between Aneuploidy, Standard Morphology Evaluation and Morphokinetic Development in 1730 Biopsied Blastocysts: a Consecutive Case Series Study. *Hum. Reprod.* 31, 2245–2254. doi:10.1093/humrep/dew183

Mumusoglu, S., Yarali, I., Bozdag, G., Ozdemir, P., Polat, M., Sokmensuer, L. K., et al. (2017). Time-lapse Morphokinetic Assessment Has Low to Moderate Ability to Predict Euploidy when Patient- and Ovarian Stimulation-Related Factors Are Taken into Account with the Use of Clustered Data Analysis. *Fertil. Sterility* 107, 413–421. doi:10.1016/j.fertnstert.2016.11.005

Munné, S., Blazek, J., Large, M., Martinez-Ortiz, P. A., Nisson, H., Liu, E., et al. (2017). Detailed Investigation into the Cytogenetic Constitution and Pregnancy Outcome of Replacing Mosaic Blastocysts Detected with the Use of High-Resolution Next-Generation Sequencing. *Fertil. Sterility* 108, 62–71. doi:10.1016/j.fertnstert.2017.05.002

Munné, S., and Wells, D. (2017). Detection of Mosaicism at Blastocyst Stage with the Use of High-Resolution Next-Generation Sequencing. *Fertil. Sterility* 107, 1085–1091. doi:10.1016/j.fertnstert.2017.03.024

Pipari, A., Guillen, A., Cruz, M., Pacheco, A., and Garcia-Velasco, J. A. (2021). Serum Anti-müllerian Hormone Levels Are Not Associated with Aneuploidy Rates in Human Blastocysts. *Reprod. BioMedicine Online* 42, 1211–1218. doi:10.1016/j.rbmo.2021.03.006

Rehman, K. S., Bukulmez, O., Langley, M., Carr, B. R., Nackley, A. C., Doody, K. M., et al. (2007). Late Stages of Embryo Progression Are a Much Better Predictor of Clinical Pregnancy Than Early Cleavage in Intracytoplasmic Sperm Injection and *In Vitro* Fertilization Cycles with Blastocyst-Stage Transfer. *Fertil. Sterility* 87, 1041–1052. doi:10.1016/j.fertnstert.2006.11.014

Sachdev, N. M., Maxwell, S. M., Besser, A. G., and Grifo, J. A. (2017). Diagnosis and Clinical Management of Embryonic Mosaicism. *Fertil. Sterility* 107, 6–11. doi:10.1016/j.fertnstert.2016.10.006

Schaeffer, E., Porchia, L. M., Neumann, A., Luna, A., Rojas, T., and López-Bayghen, E. (2020). Embryos Derived from Donor or Patient Oocytes Are Not Different for *In Vitro* Fertilization Outcomes when PGT Allows Euploid Embryo Selection: a Retrospective Study. *Clin. Translational Med.* 9, 14. doi:10.1186/s40169-020-0266-1

Stovezky, Y. R., Romanski, P. A., Bortoletto, P., and Spandorfer, S. D. (2021). Body Mass index Is Not Associated with Embryo Ploidy in Patients Undergoing *In Vitro* Fertilization with Preimplantation Genetic Testing. *Fertil. Sterility* 116, 388–395. doi:10.1016/j.fertnstert.2021.02.029

Taylor, T. H., Gitlin, S. A., Patrick, J. L., Crain, J. L., Wilson, J. M., and Griffin, D. K. (2014). The Origin, Mechanisms, Incidence and Clinical Consequences of Chromosomal Mosaicism in Humans. *Hum. Reprod. Update* 20, 571–581. doi:10.1093/humupd/dmu016

Van Echten-Arends, J., Mastenbroek, S., Sikkema-Raddatz, B., Korevaar, J. C., Heineman, M. J., Van Der Veen, F., et al. (2011). Chromosomal Mosaicism in

Human Preimplantation Embryos: a Systematic Review. *Hum. Reprod. Update* 17, 620–627. doi:10.1093/humupd/dmr014

Viotti, M., Victor, A. R., Barnes, F. L., Zouves, C. G., Besser, A. G., Grifo, J. A., et al. (2021). Using Outcome Data from One Thousand Mosaic Embryo Transfers to Formulate an Embryo Ranking System for Clinical Use. *Fertil. Sterility* 115, 1212–1224. doi:10.1016/j.fertnstert.2020.11.041

Wang, A., Kort, J., Behr, B., and Westphal, L. M. (2018). Euploidy in Relation to Blastocyst Sex and Morphology. *J. Assist. Reprod. Genet.* 35, 1565–1572. doi:10.1007/s10815-018-1262-x

Wirleitner, B., Schuff, M., Stecher, A., Murtinger, M., and Vanderzwalmen, P. (2016). Pregnancy and Birth Outcomes Following Fresh or Vitrified Embryo Transfer According to Blastocyst Morphology and Expansion Stage, and Culturing Strategy for Delayed

Zhan, Q., Sierra, E. T., Malmsten, J., Ye, Z., Rosenwaks, Z., and Zaninovic, N. (2020). Blastocyst Score, a Blastocyst Quality Ranking Tool, Is a Predictor of Blastocyst Ploidy and Implantation Potential. *F&S Rep.* 1, 133–141. doi:10.1016/j.xfre.2020.05.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 January 2022
doi: 10.3389/fgene.2021.810900

Identifying Balanced Chromosomal Translocations in Human Embryos by Oxford Nanopore Sequencing and Breakpoints Region Analysis

Zhenle Pei[1], Ke Deng[1], Caixai Lei[1], Danfeng Du[1], Guoliang Yu[2], Xiaoxi Sun[1], Congjian Xu[1] and Shuo Zhang[1]**

[1]Shanghai Ji Ai Genetics and IVF Institute, Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China, [2]Chigene (Beijing) Translational Medical Research Center Co. Ltd., Beijing, China

Edited by:
Xi Wang,
Nanjing Medical University, China

Reviewed by:
Chuan-Le Xiao,
Sun Yat-sen University, China
Yueqiu Tan,
Central South University, China

***Correspondence:**
Congjian Xu
pghunion@163.com
Shuo Zhang
chnszhang@163.com

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *08 November 2021*
Accepted: *13 December 2021*
Published: *18 January 2022*

Citation:
Pei Z, Deng K, Lei C, Du D, Yu G, Sun X, Xu C and Zhang S (2022) Identifying Balanced Chromosomal Translocations in Human Embryos by Oxford Nanopore Sequencing and Breakpoints Region Analysis. Front. Genet. 12:810900. doi: 10.3389/fgene.2021.810900

Background: Balanced chromosomal aberrations, especially balanced translocations, can cause infertility, recurrent miscarriage or having chromosomally defective offspring. Preimplantation genetic testing for structural rearrangement (PGT-SR) has been widely implemented to improve the clinical outcomes by selecting euploid embryos for transfer, whereas embryos with balanced translocation karyotype were difficult to be distinguished by routine genetic techniques from those with a normal karyotype.

Method: In this present study, we developed a clinically applicable method for reciprocal translocation carriers to reduce the risk of pregnancy loss. In the preclinical phase, we identified reciprocal translocation breakpoints in blood of translocation carriers by long-read Oxford Nanopore sequencing, followed by junction-spanning polymerase chain reaction (PCR) and Sanger sequencing. In the clinical phase of embryo diagnosis, aneuploidies and unbalanced translocations were screened by comprehensive chromosomal screening (CCS) with single nucleotide polymorphism (SNP) microarray, carrier embryos were diagnosed by junction-spanning PCR and family haplotype linkage analysis of the breakpoints region. Amniocentesis and cytogenetic analysis of fetuses in the second trimester were performed after embryo transfer to conform the results diagnosed by the presented method.

Results: All the accurate reciprocal translocation breakpoints were effectively identified by Nanopore sequencing and confirmed by Sanger sequencing. Twelve embryos were biopsied and detected, the results of junction-spanning PCR and haplotype linkage analysis were consistent. In total, 12 biopsied blastocysts diagnosed to be euploid, in which 6 were aneuploid or unbalanced, three blastocysts were identified to be balanced translocation carriers and three to be normal karyotypes. Two euploid embryos were subsequently transferred back to patients and late prenatal karyotype analysis of amniotic fluid cells was performed. The outcomes diagnosed by the current approach were totally consistent with the fetal karyotypes.

Conclusions: In summary, these investigations in our study illustrated that chromosomal reciprocal translocations in embryos can be accurately diagnosed. Long-read Nanopore sequencing and breakpoint analysis contributes to precisely evaluate the genetic risk of disrupted genes, and provides a way of selecting embryos with normal karyotype, especially for couples those without a reference.

Keywords: balanced translocation, preimplantation genetic testing, long-read sequencing, breakpoint PCR, haplotype linkage analysis

INTRODUCTION

As one of the most prevalent genomic structural rearrangements, balanced translocations occur during the process of genetic material between two separate chromosomes interchanging. Regardless of parental age, balanced translocations are frequently linked to infertility. The overall prevalence of balanced translocations is 0.25% in the general population (Jacobs et al., 1992), 1.1% in patients with infertility (Clementini et al., 2005), 4.5% in patients with a history of recurrent abortion (Sugiura-Ogasawara et al., 2004), and 9.2% in couples with more than three first-trimester abortions (Sampson et al., 2004). Carriers with balanced translocations are usually phenotypically healthy, but they suffer from a high rate of imbalanced gametes due to aberrant meiotic segregation (Ford and Clegg 1969; Brandriff et al., 1986). Additionally, balanced chromosomal translocations are capable of inducing duplications, microdeletions and disruption of genes related to infertility (Donker et al., 2017; Schilit et al., 2020). These carriers may sustain higher probability of infertility, repeated spontaneous miscarriages, and chromosomally defective offspring (Campbell et al., 1995).

Preimplantation genetic testing (PGT) offers an effective way for reciprocal translocation carriers to reduce the risk of pregnancy loss due to abnormal chromosomal segregation. PGT can be divided into three categories, namely PGT for aneuploidy (PGT-A), PGT for monogenic (PGT-M), and PGT for structural rearrangement (PGT-SR). Breakpoint sequences cannot be recovered by traditional analytical techniques, such as fluorescence *in situ* hybridization (FISH) (Schluth-Bolard et al., 2013). There is an urgent demand for innovative techniques to precisely distinguish breakpoints or linkage polymorphism markers for embryo diagnosis and genetic risk assessment of carriers. Multiple techniques have been developed to identify copy numbers variants (CNVs) of chromosome fragments, e.g. next-generation sequencing (NGS) (Chow et al., 2018), quantitative polymerase chain reaction (qPCR) (Cimadomo et al., 2018a), single nucleotide polymorphism microarray (SNP array) (Treff et al., 2011a), and array comparative genomic hybridization (aCGH) (Fragouli et al., 2011). However, none of these genetic methods can be employed to discriminate between the carrier and translocation-free embryos. In recent years, new PGT approaches have achieved significant advancement. Owing to preimplantation genetic haplotyping (PGH), translocation carriers can be accurately distinguished from noncarriers with genome-wide haplotype linkage analysis (Zhang et al., 2017; Zhang et al., 2021). Mapping Allele with Resolved Carrier Status (MaReCs) was successfully performed on 96 human embryos from 13 reciprocal translocation carriers and 12 Robertsonian translocation carriers (Xu et al., 2017). Hu, L *et al.* diagnosed 13 carrier blastocysts from 15 balanced blastocysts with the aid of NGS following micro-dissecting junction region preimplantation genetic diagnosis (MicroSeq-PGD). Moreover, they confirmed that this technique could be employed to discriminate reciprocal translocation carriers (Hu et al., 2016).

As a high-throughput sequencing or massively parallel/deep sequencing technology, NGS has been proven to be revolutionary in the genomic age. SNPs and indels are frequently observed with short-read sequencing (SRS), in which reads with the typical length of 100–200 bp proves to be adequate (Eggertsson et al., 2017). However, SRS complicates the genotyping and characterization of structural rearrangements, and the number of chromosomal rearrangements discovered per person in large-scale SRS studies has been restricted to 2,000–11,000 (Sudmant et al., 2015; Abel et al., 2020). Moreover, NGS and array-based methods also fail to identity noncarrier from euploid carrier embryos in PGT-SR cycles, or to find the position of breakpoint junctions (Tsiatis et al., 2010). By contrast, long-read approaches could provide the read length that is well-qualified for breakpoint identification (Goodwin et al., 2016). Long-read sequencing (LRS) provided by Oxford Nanopore Technologies (ONT) and PacBio single molecular real-time (SMRT) sequencing have ushered in a broad prospect of chromosomal aberration genotyping and phasing in recent years (Chaisson et al., 2015; Deamer et al., 2016; Jain et al., 2018). Those readings that are 10 kb or longer on average may be determined by LRS, which conduces to improving the detection of structural variations (SVs) (Lu et al., 2016). As a result, long reads allow for the traversing of complicated or repetitive sections with a single continuous read, thus removing the ambiguity regarding the position or size of genomic components. Notably, LRS is more sensitive and accurate than SRS in mapping SVs across the genome. In a study of LRS at the population scale, more than 22,636 SVs were identified per individual, and this size was three to five times more than that observed in SRS data (Beyter et al., 2021). This benefit is especially noticeable in repeat regions, such as tandem-repeat (TR) regions (Collins et al., 2020). Nanopore sequencing, originally proposed in 2012, is a single-molecule LRS technique capable of directly mapping the DNA structure of a native single-stranded DNA (ssDNA) molecule

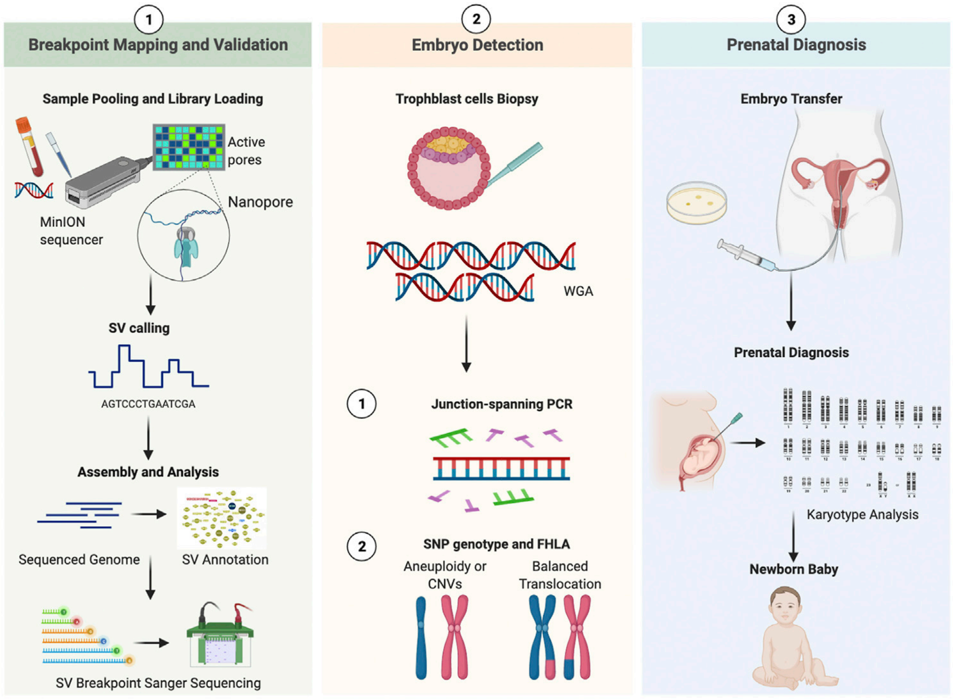


FIGURE 1 | Working pipeline of the study. Legend: The workflow contains three phases: breakpoint mapping and validation, embryo detection, prenatal diagnosis. The procedure is universal for both patients. SV, structural variation; WGA, whole genomic amplification; CNV, copy Number Variation; SNP, single nucleotide polymorphism; FHLA, family haplotype linkage analysis.

(Pennisi 2012). When whole-exome sequencing fails to differentiate the breakpoint area in Alu elements, Nanopore sequencing discovers a 7 kb segment deletion in *G6PC* gene, which is related to glycogen storage disorder type Ia (Miao et al., 2018). Furthermore, another widespread long-read technology, PacBio SMRT, can be employed to identify integrated nucleotides based on the fluorescence of the nucleotide released after phosphate chain cleavage (Rhoads and Au 2015). The SMRT sequencing has been successfully employed to screen out a 2184 bp chromosomal deletion in *PRKAR1A* gene, which is one of its first clinical applications in detecting *de novo* structural aberrations in patients *via* LRS (Merker et al., 2018). With the advancement of bioinformatics tools, single-molecule LRS has established itself as a cutting-edge method for PGT-SR cycles of balanced aberration carriers with no need for aneuploid embryos or samples from immediate family as references, which bridges gaps between existing reference assemblies (Hu et al., 2019).

In this study, whole-genome sequencing was implemented in two translocation carriers with LRS provided by the ONT. Besides, PCR and Sanger sequencing were adopted to confirm the discovery of reciprocal translocation breakpoints. In addition, the junction-spanning PCR and haplotype linkage analysis combined with comprehensive chromosomal screening (CCS) were applied to detect biopsied cells of blastocysts. This genetic method, based on Nanopore sequencing data from two carriers, established the groundwork for the larger-scale application of PGT-SR, which promotes the investigation of the SNP allele frequency spectrum and even contributes to the exploration in genomic areas that SRS technologies have not reached.

MATERIALS AND METHODS

Study Design

This study consisted of three sections (**Figure 1**). In the preclinical **section 1**, the parental blood sample was analyzed in an attempt to accurately decipher translocation breakpoints based on long-read, Nanopore sequencing data, and the results were validated by PCR and Sanger sequencing. The outline was generated with BioRender. In the clinical **section 2**, the study of whole genome amplified (WGA) DNA of embryos was performed, followed by breakpoint PCR and haplotype analysis. In the clinical **section 3**, the above-mentioned results were compared with those from prenatal diagnosis.

Patients Recruited

Two patients with cytogenetically confirmed chromosomal translocation who underwent assisted reproductive technologies (ART) were enrolled. Both patients experienced repeated miscarriages. The informed consent forms of both patients were obtained. The karyotype of patient 1 was 46, XX, t(1,2)(1q44; 2q31). While that of patient 2 was 46, XY, t(12; 14)(p13; q24). The karyotype analysis of these two carriers' parents was also performed. 10 ml of peripheral blood sample was collected from both couples and their family members during recruitment.

High-Resolution Breakpoint Mapping by Long-Read Nanopore Sequencing

According to the protocols of Oxford Nanopore, UK, after genomic DNA (5 μg) of samples passed the quality inspection, the BluePippin (Sage Science, MA, United States) automatic nucleic acid fragment recovery system was employed to cut and recycle the large fragments. After purification, both ends of the DNA fragment were repaired and the addition A reaction was performed. Subsequently, the sequencing connector was connected, and finally Qubit was adopted to perform the precise detection of the constructed DNA library. The sequencing process was executed on R9.4 flow cells with GridION X5. For the SV phasing pipeline, reads were mapped to the human reference genome (GRCh37/hg19) with NGMLR, followed by sensitive SV predictions based on Sniffles. SV was annotated when the overlap degree between sequencing data and public database was greater than or equal to 50% (the distance between INS and INS of public database shall be less than or equal to 1000 bp). Moreover, such databases as 1000 genome phase3, DGV gold standard CNV, dbVar nstd37 and Decipher were utilized during the analysis. According to the phenotype of this disease, OMIM, HPO, Clinvar and other databases were searched with a view to identifying disease-related genes.

Confirmatory Breakpoint PCR Primer Design and Sanger Sequencing Validation

The 600 kb sequences flanking the assumed breakpoints were searched from the UCSC genome website (GRCh37/hg19). Primer3 software and PrimerBank website were adopted to design primers for reference sequences and balanced reciprocal translocation breakpoint junction sequences. PCR was conducted according to the recommended protocols of manufacturers (RR001A; Takara). PCR products of breakpoint junction sequences were visualized on a 1.5% agarose gel, which was used to analyze the bands of normal and derivative chromosomes. The accurate breakpoint positions were confirmed by mapping Sanger sequencing sequences to GRCh37/hg19 human reference genome by minimap2 (version 2.10) and BLAT. If breakpoints lead to gene disruption or fusion, these genes would be retrieved on the website Online Mendelian Inheritance in Man (OMIM).

Blastocyst Biopsy and WGA

Standard procedures were conducted in the ART process. Briefly, pituitary desensitization was conducted with controlled ovarian hyper-stimulation (COH) based on individual situations. In general, metaphase II (MII) oocytes were generated by intracytoplasmic sperm injection (ICSI), and subsequently were cultured for 5–6 days into the blastocyst stage. In this study, the criteria for grading blastocysts was in line with the standards recommended by Schoolcraft *et al* (Schoolcraft et al., 1999). Approximately 3–10 trophectoderm cells were biopsied and subsequently put into PCR tubes. The multiple displacement amplification method was utilized for WGA (QIAGEN, Hilden, Germany). CCS was performed on all the biopsied blastocysts with SNP-array.

Junction-Spanning PCR and Haplotype Linkage Analysis

Junction-spanning PCR was performed according to the recommended protocols of manufacturers (RR001A; Takara). Junction-spanning PCR primers were the same as confirmatory breakpoint PCR primers in blood above, which included primers for forward reference sequence, reverse reference sequence, forward breakpoint junction sequence and reverse breakpoint junction sequence. As described previously, the SNP microarray could generate genome-wide SNP genotypes. The haplotype was created with informative SNPs, including the translocation breakpoint regions, the whole translocation chromosomes and the corresponding normal homologous chromosomes in the couple, reference and embryos. An unbalanced embryo or family member of carriers was used as reference to establish haplotypes. Informative SNPs were selected based on the criteria of homozygous in the spouse and heterozygous in the patient. Moreover, SNPs of the patient's parents or other family members shall be homozygous if these SNPs are regarded as references to establish haplotypes. The haplotypes of both the translocated chromosome and the homologous chromosome could delineate the recombination of the breakpoint region.

Embryo Transfer, Prenatal Diagnosis and Postnatal Follow-Up

Selected euploid blastocysts were transferred into the uterus of patients either 5 days after ovulation during a normal menstrual cycle or 5 days after ovulation initiated by progesterone treatment. The majority of two blastocysts were transferred, and the single blastocyst transfer was recommended for the patient with well-cryopreserved embryos. The clinical pregnancy was diagnosed if an intrauterine gestational sac with a heartbeat was detected by ultrasound imaging 30–40 days after embryo transfer. Amniocentesis was administered in the second trimester for pregnant patients. To validate the PGT results, amniocentesis fluid samples of fetuses were utilized for karyotyping analysis.

RESULTS

In Vitro Fertilization (IVF) and CCS Results

In this study, both patients with chromosomal translocations underwent two IVF cycles. Both patients had one cycle (**Supplementary Table S1**). The information and COH results of these patients are listed in **Supplementary Table S1**. CCS with SNP-array was performed on all the 12 biopsied blastocysts, and six blastocysts were diagnosed as unbalanced abnormalities

TABLE 1 | Detailed genetic testing results of the biopsied blastocysts.

Patient	Number of biopsied blastocysts	Grade of blastocysts	Molecular karyotype	Status of balanced translocation		Karyotype of fetus amniotic fluid
				Sanger method	PGH method	
Patient 1	Embryo-1	5BB	1q44*1; 2q32.1q37.3*3	Unbalanced	Unbalanced	
	Embryo-2	5BC	1q44*3; 2q32.1q37.3*1	Unbalanced	Unbalanced	
	Embryo-3	5BB	(1–22)*1,(XN)*1	Carrier	Carrier	Not transplanted
	Embryo-4	5BB	(1–22)*1,(XN)*1	Normal	Normal	Not transplanted
	Embryo-5	5BC	(1–22)*1,(XN)*1	Normal	Normal	46, XN
	Embryo-6	5BC	(1–22)*1,(XN)*1	Normal	Normal	Not transplanted
Patient 2	Embryo-1	5BC	12p13.33p13.31*3; 14q23.3q32.33*1	Unbalanced	Unbalanced	
	Embryo-2	5BB	(1–22)*1,(XN)*1	Carrier	Carrier	46,XN,t(12; 14)(p13.31; q23.3)
	Embryo-3	5BB	12p13.33p13.31*3; 14q23.3q32.33*1	Unbalanced	Unbalanced	
	Embryo-4	5CB	12p13.31q24.33*1; 14q11.1q23.3*3	Unbalanced	Unbalanced	
	Embryo-5	5BB	(1–22)*1,(XN)*1	Carrier	Carrier	Not transplanted
	Embryo-6	5BB	12p13.33p13.31*1; 14q23.3q32.33*3	Unbalanced	Unbalanced	

related to translocation. Among six blastocysts analyzed to be chromosomal balanced, three blastocysts were diagnosed as carriers and the other three were diagnosed to be normal. The specific results are listed in **Table 1**. The bioinformatics pipeline incorporating several analytical tools is presented in **Figure 1**.

High-Resolution Mapping of Translocation Breakpoints With ONT Platform

Whole-genome and long-read sequencing analyses were performed on all subjects to obtain the accurate coordinates of breakpoints. Besides, low-coverage (~20× coverage) Nanopore long-read sequencing was performed. In addition, SVs detected by Sniffles were screened out if the number of reads supporting SVs is greater than or equal to 2. The histogram number of SV statistics is presented in **Figures 2A,B**. The length distribution diagram of SVs between 30 bp and 1 Mbp is presented in **Figure 2C**. Further, 96.0 and 67.2 G bases were generated with the average length of 16,679 and 17,677 bp, respectively, for both patients (**Table 2** and **Supplementary Tables S2, S3**). Mean mapping rates for both patients were larger than 96%. Additionally, chromosome translocations for both patients were found. In patient 1, a pair of translocation breakpoints were detected in 1q44 and 2q31. Translocation breakpoints spanning chr1:244,573,198-244,573,199 and chr2:182,319,607-182,319,608 were identified; twenty-two reads supported the translocation. The reliability was verified by Integrative Genomics View (IGV) and Ribbon (**Figure 3A**). While, in patient 2, a pair of translocation breakpoints were detected in 12p13 and 14q23. Translocation breakpoints spanning chr12:5,959,384-5,959,385 and chr14: 66,307,382-66,307,383 were identified; 17 reads supported the translocation. The reliability was verified by IGV and Ribbon.

By checking these breakpoints in the UCSC Genome Browser, two breakpoints were found inside introns of genes *ADSS* and *ANO2* in samples 1 and 2. Both breakpoints disrupted the normal gene structures, which resulted in the exchange of chromosomal segments, thereby potentially disturbing the gene function due to the movement of the gene fragment from one chromosome to another. In addition, among the four breakpoints, two breakpoints were identified inside the intergenic region. The breakpoint spanning chr2:182,319,607-182,319,608 (patient 1) was located between gene *LINC01934* and *ITGA4*. Besides, the breakpoint spanning chr14:66,307,382-66, 307, 383 (patient 2) was located between gene *FUT8* and *CCDC196*. Further, it could be found that no obvious deletion or duplication was caused by breakpoints. However, there was no obvious pathogenic phenotype of carriers from whom the above two samples were obtained, other than primary infertility. All these observations demonstrated the benefits of long reads for breakpoint characterization in genomic regions of low complexity.

Validation of Breakpoints by Sanger Sequencing

To further verify translocation calls identified by Nanopore sequencing, PCR and Sanger sequencing were performed on the precise adjacent sequences of rearrangement breakpoints. The primer information is presented in **Figure 4E**. Primers for forward reference sequence, reverse reference sequence, forward breakpoint junction sequence and reverse breakpoint junction sequence were designed for each translocation breakpoint. An ideogram of normal chromosomes and derivate chromosomes of patients 1 and 2 was established. For patients 1 and 2, both PCR products and breakpoint junction sequences were observed, which validated the two translocation variations. In addition, the target PCR bands reflecting translocated chromosomes could be found in patients 1 and 2. The length of PCR products was in line with the expected product size. The precise breakpoint positions detected by Sanger sequencing for all rearranged fragments are shown in **Table 2** and **Figure 4C**. The results of sequencing in these breakpoint regions illuminated the complexity of human genome shuffling.

Embryo Biopsies, Junction-Spanning PCR and Haplotype Linkage Analysis

To evaluate the feasibility of Nanopore sequencing in PGT-SR, junction-spanning PCR analysis was performed on 12 WGA products of 12 euploid embryos, in an attempt to identify the noncarrier embryos and carrier embryos based on translocation breakpoints. For trophectoderm biopsies, it was found that there

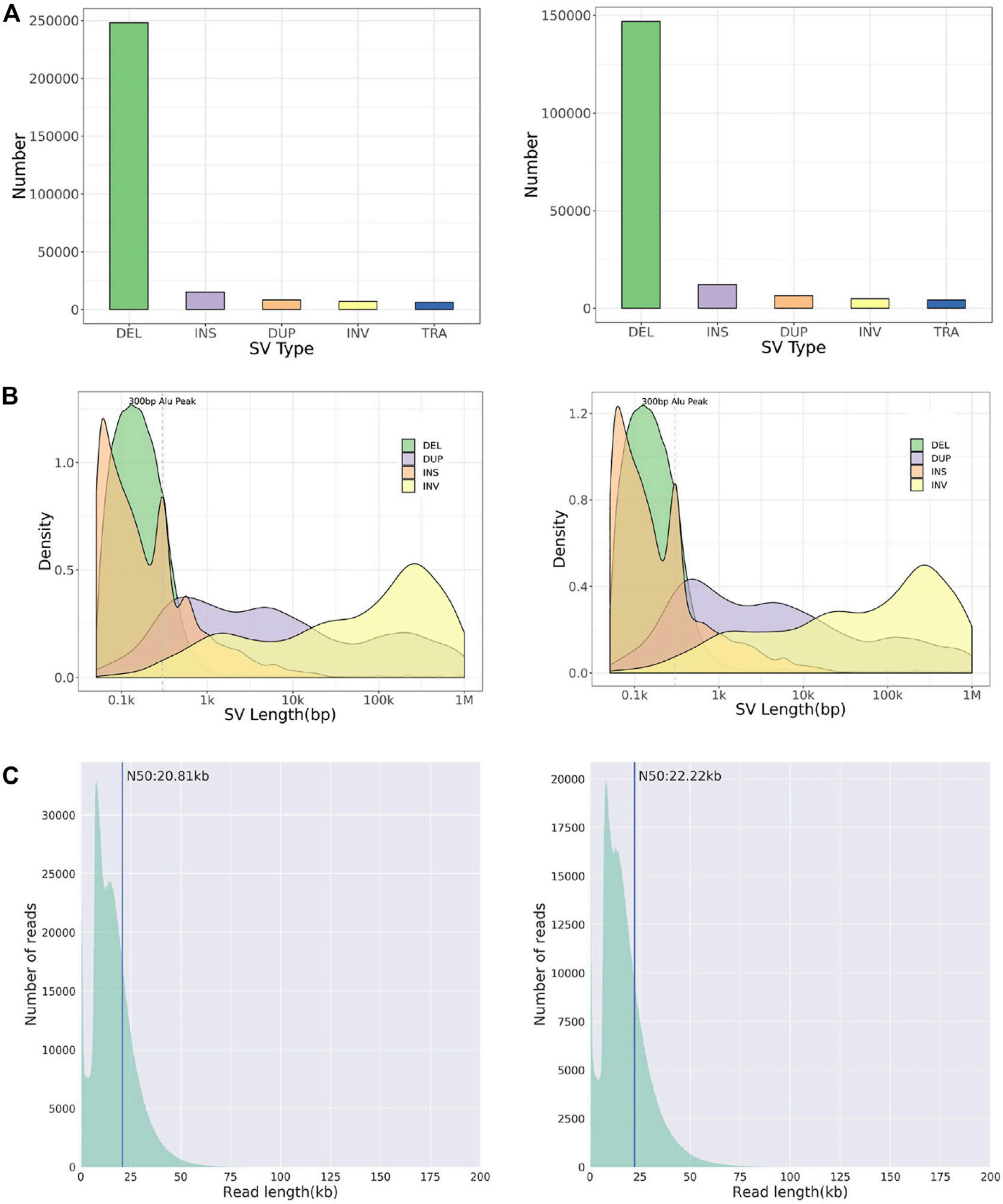


FIGURE 2 | Chromosome structural aberration results of long-read Nanopore sequencing. Legend: The results of patient 1 (left) and 2 (right) are listed respectively. **(A)** Histogram of number of different types of SVs. **(B)** Length distribution of different types of SVs. **(C)** Diagram of reads length distribution. DEL, Deletion; INS, Insertion; DUP, Duplication; INV, Inversion; TRA, Translocation.

was 100% concordance (12/12) of breakpoint PCR for patients 1 and 2. The results of junction-spanning breakpoint PCR are presented in **Table 3** and **Figure 4D**. These findings indicated that these three blastocysts were reciprocal translocation carriers, the other three were normal noncarriers, and the remaining six were unbalanced embryos. The designed primers' information on the reference sequences and balanced reciprocal translocation breakpoint junction sequences is presented in **Figure 4E**.

Meanwhile, the haplotype analysis approach of breakpoints was adopted. The detailed haplotype phasing results of the biopsied blastocysts are listed in **Table 4**. Informative SNPs were successfully generated adjacent to the breakpoint regions, which enabled the detection of the rearranged chromosomes and the corresponding normal homologous chromosomes in samples 1 and 2 (**Table 5**). Subsequently, haplotypes were employed to discriminate between embryos with balanced

TABLE 2 | Summary of breakpoint characterization results of the translocation by Nanopore and sanger sequencing.

Patient	Karyotype of peripheral blood cells	Depth (X)	Mapped sequencing reads, *n*	Mapped sequencing bases, *n*	Coverage rate (%)	Spanning break points reads, *n*	Break point position verified by sanger sequencing	SV segment length (%HLA)	Disrupted gene (break point)
Patient 1	46,XX,t(1; 2)(q44; q32.1), mat	32.01	5,570,349	93,191,761,872	96.83	22	244573198	4.68 Mb (1.88%)	ADSS (Intronic region)
						22	182319607–182319608	60.88 Mb (25.03%)	Intergenic region
Patient 2	46,XY,t(12; 14)(p13.31; q23.3), pat	22.39	3,661,947	64,963,815,620	96.39	17	5959384	5.96 Mb (4.45%)	ANO2 (Intronic region)
						17	66307374–66307382	41.04 Mb (38.23%)	Intergenic region

Reference sequence: GRCH37/hg19 reference genome.

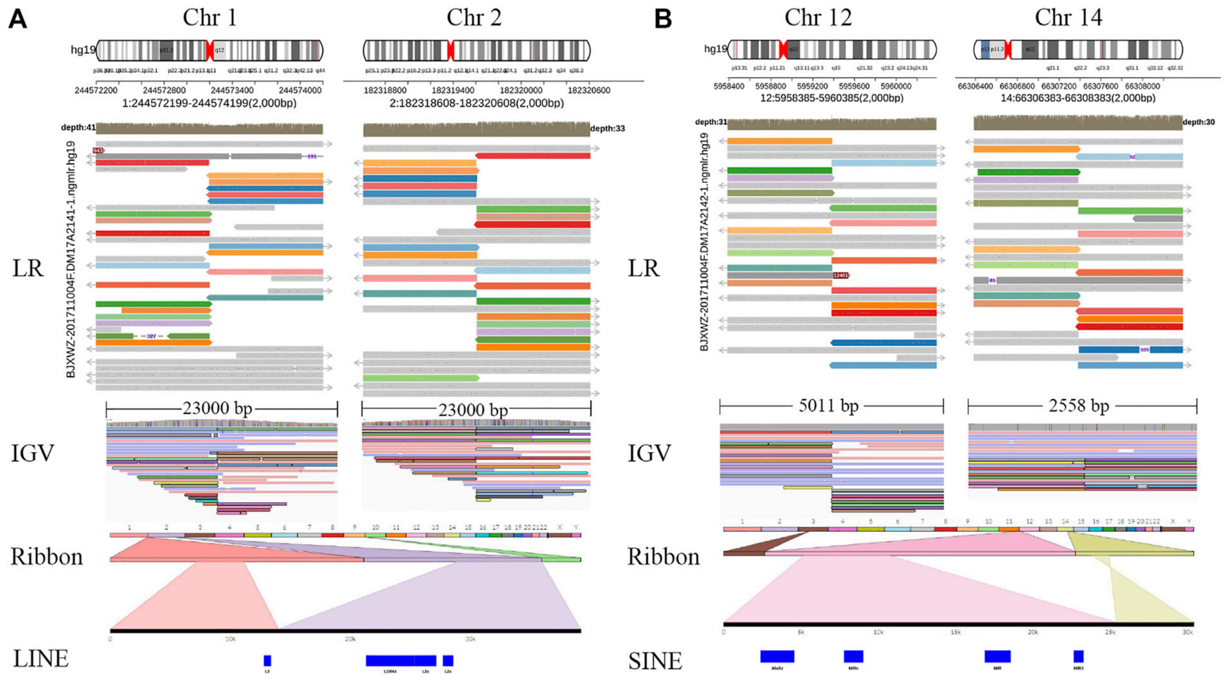
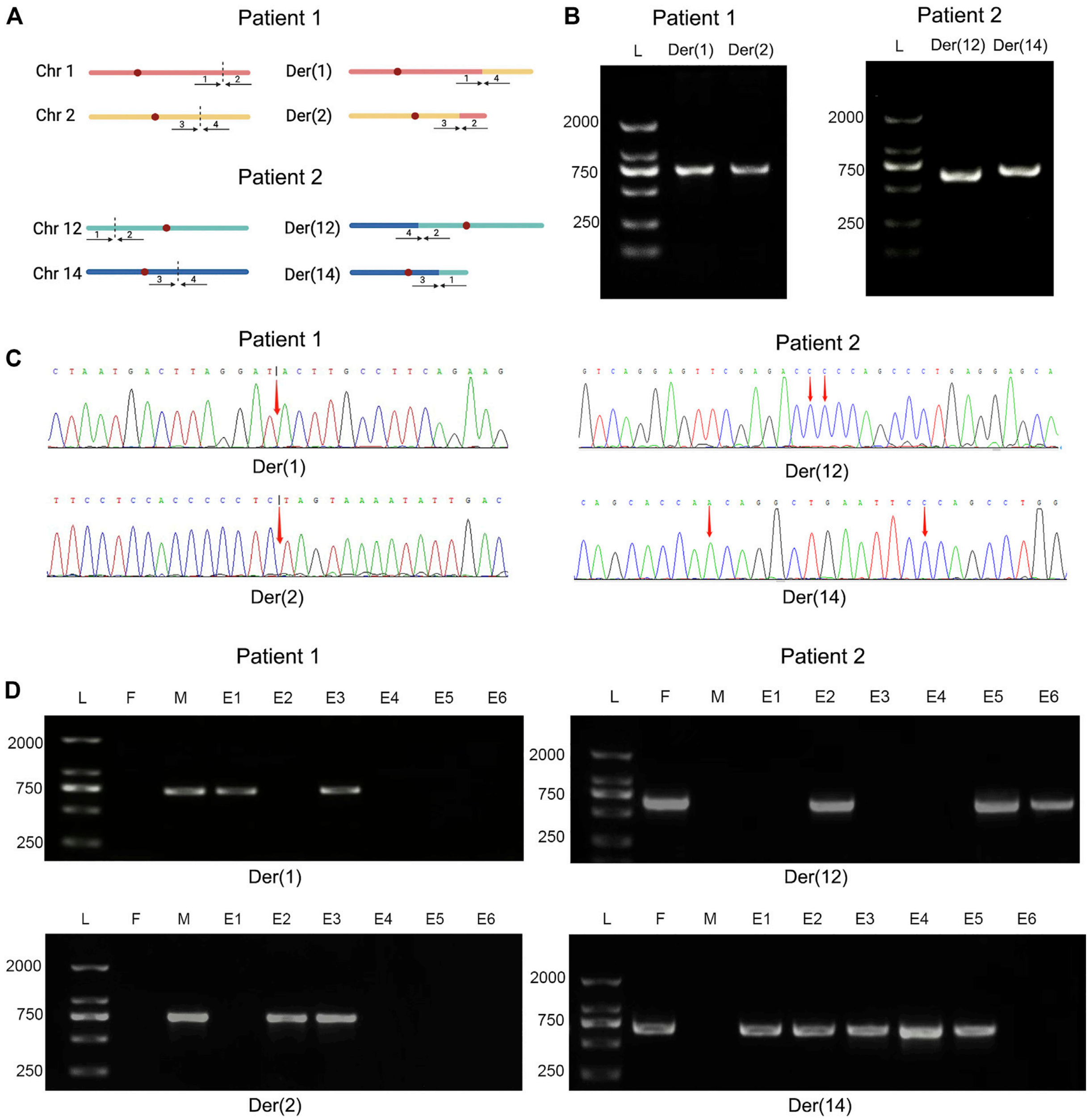


FIGURE 3 | Chromosome rearrangement annotation and visualized analysis of Nanopore sequencing data. Legend: Translocation breakpoints detected by long-read Nanopore sequencing. Two cases, patients 1–2, were represented by **(A,B)** respectively. For each case, the top panel projected the chromosome pattern which translocation occurred, and LR alignments covering the breakpoint junctions. The panel in the bottom shows SR alignments (BWA-MEM; IGV-SR) and LR alignments (NGMLR; IGV-LR) across breakpoints using IGV. In Ribbon, chromosome 1 is shown in light orange and chromosome 2 in light purple, 22 recombined reads of chromosome 1 and 2 can be seen. Chromosome 12 is shown in pink, chromosome 14 is shown in yellow, 17 recombined reads of chromosome 12 and 14 can be seen. LR, long read; SR, short read; LINE, long interspersed nuclear elements; SINE, short interspersed nuclear elements; IGV, Integrative Genomics View; BWA-MEM, Burrows-Wheeler Aligner maximal exact matches.

translocation chromosomes and those with structurally normal chromosomes through linkage analysis. A total of 12 embryos were detected. Among them, three embryos were detected with normal karyotypes and three were with balanced translocation karyotypes, the results of haplotype analysis were in line with PCR analysis as expected.

Clinical Outcomes and Validation of Efficacy

Two euploid embryos were selected and transferred into the uterus in both patients. Subsequently, both patients were pregnant, amniocentesis was performed and two healthy babies were delivered successfully. It could be verified that



E

Lab ID	Primer	Primer sequence	Size of PCR product (bp)
Patient 1	Der (1) F	AATCTCCACAGTAGGCATCCA	724
	Der (1) R	GACAAGCAGTCAGTGGAGTGG	
	Der (2) F	AATAGGGTGAGTCCACATTAGAA	740
	Der (2) R	TTTTCCTTGTATCGTTTTGTTG	

Lab ID	Primer	Primer sequence	Size of PCR product (bp)
Patient 2	Der (12) F	CAGTGGAAATGTCCTTTGAGC	626
	Der (12) R	AATGTGGTGGGTTAGGGAGTAG	
	Der (14) F	TGGACCCCACCGAGTTATG	680
	Der (14) R	GGCTGAGCTACTTGCTGGA	

FIGURE 4 | Comparative analysis of breakpoint PCR validation between Nanopore sequencing and embryo biopsies. Legend: **(A)** Ideogram of normal chromosomes and derivate chromosomes of patient 1 (left) and 2 (right). Dotted portion represents breakpoint identification. **(B)** PCR validation of translocation variants. PCR products of patient 1 (left) and 2 (right) were shown, respectively. **(C)** Schematic diagrams of precise breakpoints characteristics of patient 1 (left) and 2 (right) by Sanger sequencing. The coordinates of breakpoints were showed (red arrows). **(D)** Junction-spanning PCR and AGE validation of patient 1 (left) and 2 (right), along with biopsied embryos after PGT-SR cycles, respectively. For both individual, four fragments, two pairs of breakpoint junctions, i.e. der (1) and der (2), der (12) and der (14) were verified respectively. L, DNA ladder; M, mother; F, father; E, embryo; PCR, polymerase chain reaction; AGE, agarose gel electrophoresis. **(E)** Primers used for amplification of translocation breakpoints. F, forward primers; R, reverse primers; bp, base pairs.

TABLE 3 | Detailed junction-spanning PCR results of the biopsied blastocysts.

Patient	Number of biopsied blastocysts	Junction-spanning PCR results		Comprehensive result
		Der (1)	Der (2)	
Patient 1	Embryo-1	Positive	Negative	Unbalanced
	Embryo-2	Negative	Positive	Unbalanced
	Embryo-3	Positive	Positive	Translocation carrier
	Embryo-4	Negative	Negative	Normal
	Embryo-5	Negative	Negative	Normal
	Embryo-6	Negative	Negative	Normal
		Der (12)	Der (14)	
Patient 2	Embryo-1	Negative	Positive	Unbalanced
	Embryo-2	Positive	Positive	Translocation carrier
	Embryo-3	Negative	Positive	Unbalanced
	Embryo-4	Negative	Positive	Unbalanced
	Embryo-5	Positive	Positive	Translocation carrier
	Embryo-6	Positive	Negative	Unbalanced

TABLE 4 | Detailed haplotype phasing results of the biopsied blastocysts.

Patient	Number of biopsied blastocysts	Reference of haplotype phasing	Haplotype results of breakpoint regions (2 Mb)		Comprehensive result
Patient 1		Grandmother	1q44	2q32.1	
	Embryo-1		Derivative	Normal	Unbalanced
	Embryo-2		Normal	Derivative	Unbalanced
	Embryo-3		Derivative	Derivative	Translocation carrier
	Embryo-4		Normal	Normal	Normal
	Embryo-5		Normal	Normal	Normal
	Embryo-6		Normal	Normal	Normal
Patient 2		Grandfather	12p13.31	14q23.3	
	Embryo-1		Normal	Derivative	Unbalanced
	Embryo-2		Derivative	Derivative	Translocation carrier
	Embryo-3		Normal	Derivative	Unbalanced
	Embryo-4		Normal	Derivative	Unbalanced
	Embryo-5		Derivative	Derivative	Translocation carrier
	Embryo-6		Derivative	Negative	Unbalanced

there was 100% consistency between the bioinformatics analysis results of PGT and the cytogenetic results. Newborns born in PGT-SR cycles were all healthy. Besides, the existence of the rearranged chromosomes was also detected through linkage analysis. Therefore, the predicting accuracy of the proposed method was further demonstrated by the above-mentioned results.

DISCUSSION

Despite the fact that many carriers with balanced translocation can achieve successful live births through PGT treatment, these carriers may transmit the parental translocation aberrations to their offspring. In this situation, their children may also suffer from infertility in the future. Moreover, the genetic risk induced by *de novo* chromosomal translocation usually cannot be assessed in phenotypically normal balanced translocation carriers. There are still challenges for reliable molecular breakpoint detection of carrier embryos from those with normal karyotype due to short read lengths and application limitations.

Numerous techniques are being developed to facilitate breakpoint deciphering during PGT-SR cycles. The resolution ratio of the FISH technique is approximately 100 kilobase to 1 mega base in size (Cui et al., 2016). However, due to the fact that FISH is restricted by the requirements of specific fluorescent probes, complex procedures, and ambiguous fluorescence signals, it is necessary to develop more novel methods to accomplish a precise translocation breakpoint analysis (Wilton et al., 2009). Recently, Treff *et al.* reported that SNP genotype could be employed to distinguish two normal human embryos from 46 embryos (Treff et al., 2011b). Besides, the qPCR-based comprehensive chromosome analysis can be conducted to detect chromosomal aberrations in PGT cycles (Cimadomo et al., 2018b). With the significant advancement of sequencing technologies, NGS is becoming a potentially applicable approach for breakpoint detection and chromosomal

TABLE 5 | Summary of informative SNPs used to establish the haplotypes of the diploid blastocysts.

Patient	Embryo number	Breakpoint position	The total number of informative SNPs in breakpoint (2 Mb)	The total number of informative SNPs in chromosome	The number of recombination informative SNPs	The location of recombination in chromosome	Whether recombination happens in the breakpoint regions?
Patient-1	Embryo-3	Chr1:244573198	18	1694	127	1:1-12186274 (p36.33p36.22)	No
					423	1:48083274-107951986 (p33p13.3)	
		Chr2:182319607	4	1770	341	2:27807350-65522654 (p23.3p14)	No
					328	2:193324405-234037544 (q32.3q37.1)	
	Embryo-4	Chr1:244573198	18	1681	828	1:34419776-163565845 (p35.1q23.3)	No
					226	1: 205474838-231902291 (q32.1q42.2)	
					44	1:245135196-249250621 (q44)	
		Chr2:182319607	3	1722	196	2:102068311-131602187 (q11.2q21.1)	No
					335	2:193324405-243199373 (q32.3q37.3)	
	Embryo-5	Chr1:244573198	18	1666	503	1:38670335-82095444 (p34.3p31.1)	No
					264	1:156883028-201751539 (q23.1q32.1)	
		Chr2:182319607	4	1750	648	2:1-66217787 (p25.3p14)	No
					24	2:240229300-243199373 (q37.3)	
	Embryo-6	Chr1:244573198	15	1625	649	1:1-65798457 (p36.33p31.3)	No
		Chr2:182319607	4	1662	366	2:23021659-64943105 (p24.1p14)	No
					171	2:222699721-243199373 (q36.1q37.3)	
Patient-2	Embryo-2	Chr12:5959384	22	890	14	12:1-2019430 (p13.33)	No
					812	12:25776583-133851895 (p12.1q24.33)	
		Chr14:66307374	20	671	94	14:1-29117385 (p13q11.2)	No
	Embryo-5	Chr12:5959384	25	997	30	12:1-3538270 (p13.33p13.32)	No
					679	12:43737510-133851895 (q12q24.33)	
		Chr14:66307374	20	722	87	14:96027947-107349540 (q32.13q32.33)	No

rearrangement identification in clinical settings (Tan et al., 2014; Trautmann et al., 2019). In fact, the transition from multiplex PCR to universal processes with WGA, followed by SNP array or NGS marks a significant advancement over the last decade. The waiting period of subjects has been substantially reduced as a result of reduced lab workload and omission of customized preclinical inspections (De Rycke and Berckmoes 2020). Genotype-based NGS has been regarded as a promising platform for future PGT since NGS allows an all-in-one solution for chromosome aberration assessment (Popovic et al., 2020). However, short-read sequencing approaches are still restricted by their unfavorable sensitivity (only 10–70% of chromosome rearrangements can be detected) (Huddleston et al., 2017) and very large false positive rates (Teo et al., 2012; Sudmant et al., 2015). Moreover, on the ground that rearranged breakpoints typically occurred in complex regions, such as GC-bias regions, AT-rich short tandem-repeat regions and long low-copy repeat (LCR) regions, complex segments may be prone to be misinterpreted and not all translocation breakpoints can be mapped due to the short-read length of NGS (Wang et al., 2017).

Currently, new PGT-SR methods, e.g. PGH, are being developed, and chromosomal haplotyping has been effectively utilized to distinguish normal human embryos from translocation carriers (Xu et al., 2017; Zhang et al., 2017). The SNP haplotyping approach depends on the analysis of informative SNP markers to flank breakpoints. As for SNP-array techniques, however, it is necessary to take aneuploid embryos with the translocated chromosome or samples from the carriers' parents as references. Nowadays, Nanopore long-read sequencing is becoming a technically powerful and a clinically practical approach. Long reads-based sequencing could significantly promote the development of the phasing of chromosome aberrations at an unprecedented scale. LRS outperforms SRS in the determination of complex rearranged regions (Lu, Giordano, and Ning 2016). It has been indicated that a portable Nanopore-based sequencer can be employed to perform fast preimplantation genetic diagnosis onsite on five samples within 2 h (Wei et al., 2018). 10–30× coverage of long reads was reported to cover ~80% of full types of SV calls with ~80% precision or even higher (Kato et al., 2012). In this study, a reliable strategy was developed to distinguish noncarrier from carrier embryos in PGT-SR cycles for balanced translocation. Based on the breakpoint analysis of the blood from translocation carriers through Nanopore sequencing and Sanger sequencing, junction-spanning PCR and haplotype linkage analysis were then performed in embryos, the results of which demonstrates that this present method can be employed to distinguish translocation-free euploid embryos in PGT-SR cycles.

To the best of our knowledge, this is one of the first studies exploring the detection of normal karyotype in embryos. Besides, single-molecule long-reading sequencing is creatively introduced into PGT-SR cycles for clinical application. Compared with existing methods, there are several distinct advantages in this approach. First, in comparison to existing technologies, Nanopore long-read sequencing could be employed to decipher high-resolution breakpoint mapping directly and localize the disrupted genes precisely. Second, Nanopore long-read sequencing provides an average length of 10 kb, which significantly increases opportunities for mapping the overlapped breakpoints of chimeric reads. Third, it takes only 1.5 h for Nanopore sequencing library preparation, instead of costing days setting up the NGS library. Fourth, the cost of detecting breakpoint with Nanopore sequencing in blood samples is close to that with NGS. It is reported that 10× coverage of the whole genome can be detected with 10–15 MinION flowcells at a cost of $1000–$2000 in 7 days of sequencing duration (Middelkamp et al., 2017). Nonetheless, there are still some drawbacks in this approach. Firstly, the method is not feasible for Robertsonian translocation carriers, due to the fact that their breakpoints are in the highly repetitive centromeric regions of subtelocentric chromosomes. Secondly, the results of breakpoint PCR may be affected by the location of breakpoints.

Carriers with structural rearrangements suffer a high risk of producing unbalanced gametes due to aberrant separation during meiosis. A 1:1 rate of noncarrier to carrier embryos can be determined among all the biopsied embryos in this study. This result is similar to a previous study, in which 49% of 126 human balanced translocation embryos have been diagnosed as noncarrier embryos and the other 51% have been diagnosed as carrier embryos (Treff et al., 2016). The frequency of unbalanced chromosome recombinants seems to be affected by multiple factors, such as the coordinates of breakpoints, the size of the centric/translocated fragment, the region involved, and the chromosome affected as reported (Zhang et al., 2018; Mateu-Brull et al., 2019). Although there are different segregation patterns, the overwhelming meiotic segregants are 2:2 adjacent-1, followed by 2:2 alternate. Among all possible gametes, only two from alternate separation patterns have normal or balanced chromosomes, and the remaining gametes are genetically unbalanced. According to the theoretical basis of segregation theory, the incidence of normal or balanced gametes is relatively low (Scriven et al., 1998). It has been found that chromosome rearrangement could interfere with the correct separation of other chromosomes by disrupting the arrangement of chromosomes during meiosis I. Additionally, it has been reported that there is a trend of a higher rate of abnormal segregation patterns with a smaller size of translocated segments as well as smaller chromosomes has been reported (Jalbert et al., 1980). Besides, quadrivalents with acrocentric chromosomes are related to a higher percentage of unbalanced adjacent-1 segregants, and asymmetric quadrivalents are related to a higher incidence of adjacent-2 segregants (Zhang et al., 2018; Mateu-Brull et al., 2019).

The detected breakpoints in chromosomes can lead to gene interruption, and gene fusion, and exert position effect at the breakpoint junctions. Although the transcriptional outcomes of most chromosome structural aberrations have not been investigated, there is a risk to break genetic rules for rearrangement-mediated breakpoints that fuse or disrupt relevant genes (Canela et al., 2017). Due to the fact that the included patients have normal phenotypes, the interruption may not produce a pathogenic phenotype or may cause an undetectable phenotype. In this study, insertion sequences of less than 40 bases, which are induced by fallible non-homologous end-joining (NHEJ) or microhomology-mediated break-induced replication (MMBIR) after DNA double strand breaks down at the conjunction with translocation (Smith et al., 2005), are detected adjacent to the breakpoint junctions in all four derivative chromosomes. The majority of the constitutional translocations are not recurrent, and microhomology is a typical characteristic at breakpoint junctions. Moreover, although the sequence at breakpoints can be modified by resection, insertion, and translocation, it has been suggested in recent studies that mutation may also occur in sequences from breakpoint junctions (Weckselblatt and Rudd 2015). Chromosome structural rearrangements may also impose position effects at the breakpoint junctions, which could diversify the expression of related intact genes. It has been reported that in the translocation between chromosomes 11 and 22, the aberrant nuclear localization of translocated segments contributes to the altered expression of several genes on chromosomes (Harewood et al., 2010). Further, it is necessary to investigate the phenotypes of cis and trans position effects of structural aberrations when WGS breakpoint mapping fails to pinpoint genes associated with diseases (Gilissen et al., 2014).

In summary, techniques for breakpoint analysis based on long-read Nanopore sequencing are currently being investigated. Based on the clinical validation, this method has been demonstrated to have favorable feasibility. The strategy proposed in this study may provide a way of selecting embryos with normal karyotypes during PGT-SR treatment cycles. Junction-spanning PCR in embryos proves to be alternative, especially for those without a reference. The sensitivity and specificity of this strategy require to be further verified due to a relatively small sample size.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Obstetrics and gynecology hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CX, XS, and SZ conceived the hypothesis, SZ and CL collected the cases, GY performed the Nanopore sequencing and data analysis, SZ and ZP preformed the embryo testing and data analysis, ZP wrote the manuscript. KD, DD, and CL wrote, revised the manuscript and provided edits. All authors contributed to the final manuscript and approved the submitted version.

FUNDING

The research was supported by shanghai science and Technology innovation action Plan Program (18411953800 and 20Y11907200).

ACKNOWLEDGMENTS

We express deepest gratitude to all the families enrolled in our research. We also thank the staff of the genetics laboratory and IVF laboratory for their invaluable efforts and contribution relating to the experiment. We thank YF Z for polishing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.810900/full#supplementary-material

REFERENCES

Abel, H. J., Larson, D. E., Larson, D. E., Regier, A. A., Chiang, C., Das, I., et al. (2020). Mapping and Characterization of Structural Variation in 17,795 Human Genomes. *Nature* 583, 83–89. doi:10.1038/s41586-020-2371-0

Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H. P., Bjornsson, E., Jonsson, H., et al. (2021). Long-read Sequencing of 3,622 Icelanders Provides Insight into the Role of Structural Variants in Human Diseases and Other Traits. *Nat. Genet.* 53, 779–786. doi:10.1038/s41588-021-00865-4

Brandriff, B, Gordon, L, Ashworth, L. K, Littman, V, Watchmaker, G, and Carrano, A. V (1986). Cytogenetics of Human Sperm: Meiotic Segregation in Two Translocation Carriers. *Am. J. Hum. Genet.* 38, 197–208.

Campbell, S, Uhlmann, W. R, Duquette, D, Johnson, M. P, and Evans, M. I (1995). Pregnancy Outcome When Both Members of a Couple Have Balanced Translocations. *Obstet. Gynecol.* 85, 844–846. doi:10.1016/0029-7844(94)00246-a

Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., et al. (2017). Genome Organization Drives Chromosome Fragility. *Cell* 170, 507–521. doi:10.1016/j.cell.2017.06.034

Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., et al. (2015). Resolving the Complexity of the Human Genome Using Single-Molecule Sequencing. *Nature* 517, 608–611. doi:10.1038/nature13907

Chow, J. F. C, Yeung, W. S. B, Lee, V. C. Y, Lau, E. Y. L, and Ng, E. H. Y (2018). Evaluation of Preimplantation Genetic Testing for Chromosomal Structural Rearrangement by a Commonly Used Next Generation Sequencing Workflow. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 224, 66–73. doi:10.1016/j.ejogrb.2018.03.013

Cimadomo, D., Capalbo, A., Levi-Setti, P. E., Soscia, D., Orlando, G., Albani, E., et al. (2018a). Associations of Blastocyst Features, Trophectoderm Biopsy and Other Laboratory Practice with Post-warming Behavior and Implantation. *Hum. Reprod.* 33, 1992–2001. doi:10.1093/humrep/dey291

Cimadomo, D., Rienzi, L., Romanelli, V., Alviggi, E., Levi-Setti, P. E., Albani, E., et al. (2018b). Inconclusive Chromosomal Assessment After Blastocyst Biopsy: Prevalence, Causative Factors and Outcomes After Re-biopsy and Re-vitrification. A Multicenter Experience. *Hum. Reprod.* 33, 1839–1846. doi:10.1093/humrep/dey282

Clementini, E, Palka, C, Iezzi, I, Stuppia, L, Guanciali-Franchi, P, and Tiboni, G. M (2005). Prevalence of Chromosomal Abnormalities in 2078 Infertile Couples

Referred for Assisted Reproductive Techniques. *Hum. Reprod.* 20, 437–442. doi:10.1093/humrep/deh626

Collins, R. L, Brand, H, Karczewski, K. J, Zhao, X, Alföldi, J, Francioli, L. C., et al. (2020). A Structural Variation Reference for Medical and Population Genetics. *Nature* 581, 444–451. doi:10.1038/s41586-020-2287-8

Cui, C, Shu, W, and Li, P (2016). Fluorescence In Situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. *Front. Cel Dev. Biol.* 4, 89. doi:10.3389/fcell.2016.00089

De Rycke, M, and Berckmoes, V (2020). Preimplantation Genetic Testing for Monogenic Disorders. *Genes* 11, 871. doi:10.3390/genes11080871

Deamer, D, Akeson, M, and Branton, D (2016). Three Decades of Nanopore Sequencing. *Nat. Biotechnol.* 34, 518–524. doi:10.1038/nbt.3423

Donker, R. B, Vloeberghs, V, Groen, H, Tournaye, H, van Ravenswaaij-Arts, C. M. A, and Land, J. A (2017). Chromosomal Abnormalities in 1663 Infertile Men with Azoospermia: The Clinical Consequences. *Hum. Reprod.* 32, 2574–2580. doi:10.1093/humrep/dex307

Eggertsson, H. P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., et al. (2017). Graphtyper Enables Population-Scale Genotyping Using Pangenome Graphs. *Nat. Genet.* 49, 1654–1660. doi:10.1038/ng.3964

Ford, C. E, and Clegg, H. M (1969). Reciprocal Translocations. *Br. Med. Bull.* 25, 110–114. doi:10.1093/oxfordjournals.bmb.a070659

Fragouli, E., Alfarawati, S., Daphnis, D. D., Goodall, N.-n., Mania, A., Griffiths, T., et al. (2011). Cytogenetic Analysis of Human Blastocysts with the Use of FISH, CGH and aCGH: Scientific Data and Technical Evaluation. *Hum. Reprod.* 26, 480–490. doi:10.1093/humrep/deq344

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., et al. (2014). Genome Sequencing Identifies Major Causes of Severe Intellectual Disability. *Nature* 511, 344–347. doi:10.1038/nature13394

Goodwin, S, McPherson, J. D, and McCombie, W. R (2016). Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nat. Rev. Genet.* 17, 333–351. doi:10.1038/nrg.2016.49

Harewood, L., Schütz, F., Boyle, S., Perry, P., Delorenzi, M., Bickmore, W. A., et al. (2010). The Effect of Translocation-Induced Nuclear Reorganization on Gene Expression. *Genome Res.* 20, 554–564. doi:10.1101/gr.103622.109

Hu, L., Cheng, D., Gong, F., Lu, C., Tan, Y., Luo, K., et al. (2016). Reciprocal Translocation Carrier Diagnosis in Preimplantation Human Embryos. *EBioMedicine* 14, 139–147. doi:10.1016/j.ebiom.2016.11.007

Hu, L., Liang, F., Cheng, D., Zhang, Z., Yu, G., Zha, J., et al. (2019). Location of Balanced Chromosome-Translocation Breakpoints by Long-Read Sequencing on the Oxford Nanopore Platform. *Front. Genet.* 10, 1313. doi:10.3389/fgene.2019.01313

Huddleston, J., Chaisson, M. J. P., Steinberg, K. M., Warren, W., Hoekzema, K., Gordon, D., et al. (2017). Discovery and Genotyping of Structural Variation from Long-Read Haploid Genome Sequence Data. *Genome Res.* 27, 677–685. doi:10.1101/gr.214007.116

Jacobs, P. A, Browne, C, Gregson, N, Joyce, C, and White, H (1992). Estimates of the Frequency of Chromosome Abnormalities Detectable in Unselected Newborns Using Moderate Levels of Banding. *J. Med. Genet.* 29, 103–108. doi:10.1136/jmg.29.2.103

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore Sequencing and Assembly of a Human Genome with Ultra-long Reads. *Nat. Biotechnol.* 36, 338–345. doi:10.1038/nbt.4060

Jalbert, P, Sele, B, and Jalbert, H (1980). Reciprocal Translocations: A Way to Predict the Mode of Imbalanced Segregation by Pachytene-Diagram Drawing. *Hum. Genet.* 55, 209–222. doi:10.1007/BF00291769

Kato, T, Kurahashi, H, and Emanuel, B. S (2012). Chromosomal Translocations and Palindromic AT-Rich Repeats. *Curr. Opin. Genet. Dev.* 22, 221–228. doi:10.1016/j.gde.2012.02.004

Lu, H, Giordano, F, and Ning, Z (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics* 14, 265–279. doi:10.1016/j.gpb.2016.05.004

Mateu-Brull, E., Rodrigo, L., Peinado, V., Mercader, A., Campos-Galindo, I., Bronet, F., et al. (2019). Interchromosomal Effect in Carriers of Translocations and Inversions Assessed by Preimplantation Genetic Testing for Structural Rearrangements (PGT-SR). *J. Assist. Reprod. Genet.* 36, 2547–2555. doi:10.1007/s10815-019-01593-9

Merker, J. D., Wenger, A. M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., et al. (2018). Long-read Genome Sequencing Identifies Causal Structural Variation in a Mendelian Disease. *Genet. Med.* 20, 159–163. doi:10.1038/gim.2017.86

Miao, H., Zhou, J., Yang, Q., Liang, F., Wang, D., Ma, N., et al. (2018). Long-read Sequencing Identified a Causal Structural Variant in an Exome-Negative Case and Enabled Preimplantation Genetic Diagnosis. *Hereditas* 155, 32. doi:10.1186/s41065-018-0069-1

Middelkamp, S., van Heesch, S., Braat, A. K., de Ligt, J., van Iterson, M., Simonis, M., et al. (2017). Molecular Dissection of Germline Chromothripsis in a Developmental Context Using Patient-Derived iPS Cells. *Genome Med.* 9, 9. doi:10.1186/s13073-017-0399-z

Pennisi, E (2012). Search for Pore-Fection. *Science* 336, 534–537. doi:10.1126/science.336.6081.534

Popovic, M, Dhaenens, L, Boel, A, Menten, B, and Heindryckx, B (2020). Chromosomal Mosaicism in Human Blastocysts: The Ultimate Diagnostic Dilemma. *Hum. Reprod. Update* 26, 313–334. doi:10.1093/humupd/dmz050

Rhoads, A, and Au, K. F (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* 13, 278–289. doi:10.1016/j.gpb.2015.08.002

Sampson, J. E., Ouhibi, N., Lawce, H., Patton, P. E., Battaglia, D. E., Burry, K. A., et al. (2004). The Role for Preimplantation Genetic Diagnosis in Balanced Translocation Carriers. *Am. J. Obstet. Gynecol.* 190, 1707–1711. doi:10.1016/j.ajog.2004.02.063

Schilit, S. L. P., Menon, S., Friedrich, C., Kammin, T., Wilch, E., Hanscom, C., et al. (2020). SYCP2 Translocation-Mediated Dysregulation and Frameshift Variants Cause Human Male Infertility. *Am. J. Hum. Genet.* 106, 41–57. doi:10.1016/j.ajhg.2019.11.013

Schluth-Bolard, C., Labalme, A., Cordier, M.-P., Till, M., Nadeau, G., Tevissen, H., et al. (2013). Breakpoint Mapping by Next Generation Sequencing Reveals Causative Gene Disruption in Patients Carrying Apparently Balanced Chromosome Rearrangements with Intellectual Deficiency And/or Congenital Malformations. *J. Med. Genet.* 50, 144–150. doi:10.1136/jmedgenet-2012-101351

Schoolcraft, W. B, Gardner, D. K, Lane, M, Schlenker, T, Hamilton, F, and Meldrum, D. R (1999). Blastocyst Culture and Transfer: Analysis of Results and Parameters Affecting Outcome in Two In Vitro Fertilization Programs. *Fertil. Sterility* 72, 604–609. doi:10.1016/s0015-0282(99)00311-8

Scriven, P. N, Handyside, A. H, and Ogilvie, C. M (1998). Chromosome Translocations: Segregation Modes and Strategies for Preimplantation Genetic Diagnosis. *Prenat. Diagn.* 18, 1437–1449. doi:10.1002/(sici)1097-0223(199812)18:13<1437::aid-pd497>3.0.co;2-p

Smith, J. A, Waldman, B. C, and Waldman, A. S (2005). A Role for DNA Mismatch Repair Protein Msh2 in Error-Prone Double-Strand-Break Repair in Mammalian Chromosomes. *Genetics* 170, 355–363. doi:10.1534/genetics.104.039362

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., et al. (2015). An Integrated Map of Structural Variation in 2,504 Human Genomes. *Nature* 526, 75–81. doi:10.1038/nature15394

Sugiura-Ogasawara, M, Ozaki, Y, Sato, T, Suzumori, N, and Suzumori, K (2004). Poor Prognosis of Recurrent Aborters with Either Maternal or Paternal Reciprocal Translocations. *Fertil. Sterility* 81, 367–373. doi:10.1016/j.fertnstert.2003.07.014

Tan, Y., Yin, X., Zhang, S., Jiang, H., Tan, K., Li, J., et al. (2014). Clinical Outcome of Preimplantation Genetic Diagnosis and Screening Using Next Generation Sequencing. *GigaSci* 3, 30. doi:10.1186/2047-217X-3-30

Teo, S. M, Pawitan, Y, Ku, C. S, Chia, K. S, and Salim, A (2012). Statistical Challenges Associated with Detecting Copy Number Variations with Next-Generation Sequencing. *Bioinformatics* 28, 2711–2718. doi:10.1093/bioinformatics/bts535

Trautmann, M., Cyra, M., Isfort, I., Jeiler, B., Krüger, A., Grünewald, I., et al. (2019). Phosphatidylinositol-3-kinase (PI3K)/Akt Signaling Is Functionally Essential in Myxoid Liposarcoma. *Mol. Cancer Ther.* 18, 834–844. doi:10.1158/1535-7163.MCT-18-0763

Treff, N. R, Tao, X, Schillings, W. J, Bergh, P. A, Scott, R. T, Jr, and Levy, B (2011b). Use of Single Nucleotide Polymorphism Microarrays to Distinguish Between Balanced and Normal Chromosomes in Embryos from a Translocation Carrier. *Fertil. Sterility* 96, e58–e65. doi:10.1016/j.fertnstert.2011.04.038

Treff, N. R., Northrop, L. E., Kasabwala, K., Su, J., Levy, B., and Scott, R. T., Jr (2011a). Single Nucleotide Polymorphism Microarray-Based Concurrent Screening of 24-chromosome Aneuploidy and Unbalanced Translocations in Preimplantation Human Embryos. *Fertil. Sterility* 95, 1606–1612. doi:10.1016/j.fertnstert.2010.11.004

Treff, N. R., Thompson, K., Rafizadeh, M., Chow, M., Morrison, L., Tao, X., et al. (2016). SNP Array-Based Analyses of Unbalanced Embryos as a Reference to Distinguish between Balanced Translocation Carrier and Normal Blastocysts. *J. Assist. Reprod. Genet.* 33, 1115–1119. doi:10.1007/s10815-016-0734-0

Tsiatis, A. C., Norris-Kirby, A., Rich, R. G., Hafez, M. J., Gocke, C. D., Eshleman, J. R., et al. (2010). Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve Analysis for the Detection of KRAS Mutations. *J. Mol. Diagn.* 12, 425–432. doi:10.2353/jmoldx.2010.090188

Wang, L., Shen, J., Cram, D. S., Ma, M., Wang, H., Zhang, W., et al. (2017). Preferential Selection and Transfer of Euploid Noncarrier Embryos in Preimplantation Genetic Diagnosis Cycles for Reciprocal Translocations. *Fertil. Sterility* 108, 620–627. doi:10.1016/j.fertnstert.2017.07.010

Wapner, R. J., Martin, C. L., Levy, B., Ballif, B. C., Eng, C. M., Zachary, J. M., et al. (2012). Chromosomal Microarray Versus Karyotyping for Prenatal Diagnosis. *N. Engl. J. Med.* 367, 2175–2184. doi:10.1056/NEJMoa1203382

Weckselblatt, B, and Rudd, M. K (2015). Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet.* 31, 587–599. doi:10.1016/j.tig.2015.05.010

Wei, S, Weiss, Z. R, Gaur, P, Forman, E, and Williams, Z (2018). Rapid Preimplantation Genetic Screening Using a Handheld, Nanopore-Based DNA Sequencer. *Fertil. Sterility* 110, 910–916. doi:10.1016/j.fertnstert.2018.06.014

Wilton, L, Thornhill, A, Traeger-Synodinos, J, Sermon, K. D, and Harper, J. C (2009). The Causes of Misdiagnosis and Adverse Outcomes in PGD. *Hum. Reprod.* 24, 1221–1228. doi:10.1093/humrep/den488

Xu, J., Zhang, Z., Niu, W., Yang, Q., Yao, G., Shi, S., et al. (2017). Mapping Allele with Resolved Carrier Status of Robertsonian and Reciprocal Translocation in Human Preimplantation Embryos. *Proc. Natl. Acad. Sci. USA* 114, E8695–E8702. doi:10.1073/pnas.1715053114

Zhang, S., Lei, C., Wu, J., Sun, H., Zhou, J., Zhu, S., et al. (2018). Analysis of Segregation Patterns of Quadrivalent Structures and the Effect on Genome Stability During Meiosis in Reciprocal Translocation Carriers. *Hum. Reprod.* 33, 757–767. doi:10.1093/humrep/dey036

Zhang, S., Lei, C., Wu, J., Xiao, M., Zhou, J., Zhu, S., et al. (2021). A Comprehensive and Universal Approach for Embryo Testing in Patients with Different Genetic Disorders. *Clin. Translational Med.* 11, e490. doi:10.1002/ctm2.490

Zhang, S., Lei, C., Wu, J., Zhou, J., Sun, H., Fu, J., et al. (2017). The Establishment and Application of Preimplantation Genetic Haplotyping in Embryo Diagnosis for Reciprocal and Robertsonian Translocation Carriers. *BMC Med. Genomics* 10, 60. doi:10.1186/s12920-017-0294-x

Conflict of Interest: Author GY was employed by the company Chigene (Beijing) Translational Medical Research Center Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fgene.2022.839207

Comparative Transcriptomics Uncover the Uniqueness of Oocyte Development in the Donkey

Fa-Li Zhang[1,2], Shu-Er Zhang[3], Yu-Jiang Sun[1,4], Jun-Jie Wang[1] and Wei Shen[1]**

[1]College of Life Sciences, Key Laboratory of Animal Reproduction and Biotechnology in Universities of Shandong, Qingdao Agricultural University, Qingdao, China, [2]College of Animal Science and Veterinary Medicine, Shandong Agricultural University, Tai'an, China, [3]Animal Husbandry General Station of Shandong Province, Jinan, China, [4]Dongying Vocational Institute, Dongying, China

The donkey is an important domestic animal, however the number of donkeys world-wide is currently declining. It is therefore important to protect their genetic resources and to elaborate the regulatory mechanisms of donkey reproduction, particularly, oocyte development. Here, we adopted comparative transcriptomic analysis and weighted gene co-expression network analysis (WGCNA) to uncover the uniqueness of donkey oocyte development compared to cattle, sheep, pigs, and mice, during the period from germinal vesicle (GV) to metaphase II (MII). Significantly, we selected 36 hub genes related to donkey oocyte development, including wee1-like protein kinase 2 (WEE2). Gene Ontology (GO) analysis suggested that these genes are involved in the negative regulation of cell development. Interestingly, we found that donkey specific differentially expressed genes (DEGs) were involved in RNA metabolism and apoptosis. Moreover, the results of WGCNA showed species-specific gene expression patterns. We conclude that, compared to other species, donkey oocytes express a large number of genes related to RNA metabolism to maintain normal oocyte development during the period from GV to MII.

Keywords: donkey, oocyte development, WGCNA, comparative transcriptomic, WEE2

Edited by:
Mengcheng Luo,
Wuhan University, China

Reviewed by:
Shao-Chen Sun,
Nanjing Agricultural University, China
Teng Zhang,
Inner Mongolia Normal University, China
Jinlian Hua,
Northwest A&F University, China

***Correspondence:**
Wei Shen
wshen@qau.edu.cn
shenwei427@163.com
Jun-Jie Wang
junjieseven0717@163.com

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *19 December 2021*
Accepted: *11 January 2022*
Published: *28 January 2022*

Citation:
Zhang F-L, Zhang S-E, Sun Y-J, Wang J-J and Shen W (2022) Comparative Transcriptomics Uncover the Uniqueness of Oocyte Development in the Donkey. Front. Genet. 13:839207. doi: 10.3389/fgene.2022.839207

INTRODUCTION

The donkey (*Equus asinus*) is a descendent of the African wild ass and is a common domestic beast of burden (Beja-Pereira et al., 2004). It can also provide meat and milk, and in particular donkey-hide gelatin, mostly composed of collagen, which is a traditional Chinese medicinal material (Kim et al., 2018). There is a broad consensus that, owing to the high content of linoleic acid, donkey meat is very palatable (Polidori et al., 2009). However, it is worrying that the number of donkeys is currently declining sharply owing to agricultural mechanization and the development of transportation vehicles.

Animal reproductive biotechnology, including *in vitro* fertilization and embryo transfer (IVF-ET) and its derivative technologies, is a perfect technology for saving endangered species (Bai et al., 2020). However, a limitation of this technology with donkeys is that the rate of *in vitro* oocyte maturation is relatively low (Goudet et al., 2016). Therefore, a deeper understanding of the characteristics of donkey oocyte development is needed, focusing on the meiotic maturation of oocytes, during the period from germinal vesicle (GV) to metaphase II (MII). The period from GV–MII marks the maturation of an oocyte, both nuclear and cytoplasmic (Eppig, 1996; Zhao et al., 2020). Studying the core regulatory factors of oocyte development during this period, and the precise and meticulous regulatory signaling pathways are of great significance for improving the *in vitro* maturation rate of donkey oocytes.

Previous studies mainly focused on economic traits, comparing donkeys with other livestock, but little research has addressed gene expression characteristics (Aganga et al., 2003; Polidori et al., 2008; Tian et al., 2020). Currently, research regarding single-cell RNA-seq (scRNA-seq) data of donkey oocytes suggests that the differentially expressed genes (DEGs) during GV–MII are related to the meiotic cell cycle, mitochondrial activity, and so on (Li Z. et al., 2021). The research suggested that wee1-like protein kinase 2 (WEE2) was involved in donkey oocyte development. Moreover, a study in mice indicated that Wee2-deficient caused fertilization failure and female infertility (Sang et al., 2018). WEE2 is indispensable for ensuring exit from meiosis in oocytes and promote pronuclear formation (Oh et al., 2011). However, the function of WEE2 is still unclear so far. In 2013, the donkey genome was been reported for the first time by Orlando et al., and in 2018, they released an updated version (Orlando et al., 2013; Renaud et al., 2018). In 2020, the Dezhou donkey genome was reported by Wang et al. (Wang C. et al., 2020). The reports of these reference genomes provide us with convenient conditions for studying the gene expression characteristics of donkey oocyte development.

RNA sequencing (RNA-seq) is able to determine the sum of all transcribed RNAs of tissues or cells under specific conditions, and has been widely used in animal and plant research (Stark et al., 2019). However, the limitation of this technology is that it requires large number of cells, which is not ideal for precious cell types such as oocytes. Excitingly, the emergence of scRNA-seq solves the problem of how to obtain the transcriptome status of a very small number of cells (Picelli et al., 2013). Here, we collected GV phase and MII phase scRNA-seq data from donkeys, cattle, sheep, pigs, and mice, and applied comparative transcriptomics analysis and weighted gene co-expression network analysis (WGCNA). Surprisingly, including WEE2, we identified 36 hub genes related to donkey oocyte development during GV–MII stages. This study provides novel information about key regulators of donkey oocyte development, which can provide a theoretical basis for the protection of donkey germplasm resources.

MATERIALS AND METHODS

Data Collection

We collected transcriptome data of oocyte development during the period from germinal vesicle (GV) to metaphase II (MII) in different species, including donkeys, cattle, sheep, pigs, and mice. All RNA sequencing (RNA-seq) data was derived from public databases, including sequence read archive (SRA), gene expression omnibus (GEO), and genome sequence archive (GSA) with the following accession numbers: donkey (PRJNA763991) (Li Z. et al., 2021), cattle (CRA005589) (Li M.-H. et al., 2021), pig (GSE160334) (Du et al., 2021), sheep (GSE148022) (Wang J.-J. et al., 2020), and mouse (GSE119906) (Qian et al., 2019).

Workflow of RNA-Seq Data Processing

In order to ensure the accuracy of data analysis, we dealt with raw sequencing data, not processed files. Firstly, the FastQC (v0.11.8) was used to check the raw RNA-seq data and based on the quality control report, unqualified data was eliminated (Andrews, 2010). Next, Fastp (v0.23.1) was used for further quality control, and in this step, low-quality, unqualified reads were removed (Chen et al., 2018). As important sequence alignment software, STAR (v2.7.0f) was selected for sequence alignment to a reference genome, and we directly used the "*--outSAMtype BAM SortedByCoordinate*" parameter to generate a BAM format file (Dobin et al., 2013). The reference genome of each species was as follows: *Equus asinus* (assembly ASM1607732v2), *Bos taurus* (assembly ARS-UCD1.2), *Ovis aries* (assembly ARS-UI_Ramb_v2.0), *Sus scrofa* (assembly Sscrofa11.1), and *Mus musculus* (assembly GRCm38.p5). Finally, FeatureCounts (v1.6.3) was processed to generate gene counts (Liao et al., 2014).

Differentially Expressed Genes Analysis

The R package DESeq2 (v1.32.0) was adopted to identify the DEGs of oocytes during GV–MII (Love et al., 2014). The judgment threshold of significantly different DEGs was "|log2fold change| > 2 and p-value < 0.05". In order to quantify the amount of gene expression, we used a custom R script to calculate fragments per kilobase of exon model per million mapped fragments (FPKM) (Zhao et al., 2021).

Comparative Analysis of Oocyte Development Among Species

To obtain a comprehensive understanding of the uniqueness of donkey oocyte development during GV–MII, compared to other species, we used comparative transcriptomic analysis. Firstly, we used uniform conditions to obtain DEGs during GV–MII. Next, taking gene function annotations into consideration, the gene symbols in donkeys, cattle, sheep, and pigs, were converted into murine homologous gene symbol IDs via the R package gprofiler2 (v0.2.1) (Kolberg et al., 2020). Finally, we compared the similarities and differences between donkey and cattle, sheep, pig, and mouse oocyte development during GV–MII.

Principal Component Analysis (PCA) and Hierarchical Clustering Analysis

PCA of gene expression counts in this study were performed using the R package DESeq2 (v1.32.0), and the top 2 PCs were displayed by scatter plot (Love et al., 2014). The FPKM matrix of gene expression in the examined species, underwent hierarchical clustering analysis and the R package ggtree (v3.0.4) was used for visualization (Yu et al., 2017).

Weighted Gene Co-expression Network Analysis (WGCNA)

The R package WGCNA (v1.70-3) was used to uncover the correlation between genes (Langfelder and Horvath, 2008). First, FPKM was used to normalize gene expression levels among all species in the study. Next, in order to construct a mixed matrix of the gene expression of different species, the R package gprofiler2 (v0.2.1) was selected to unify the gene symbol IDs (Kolberg et al., 2020). Further, the function

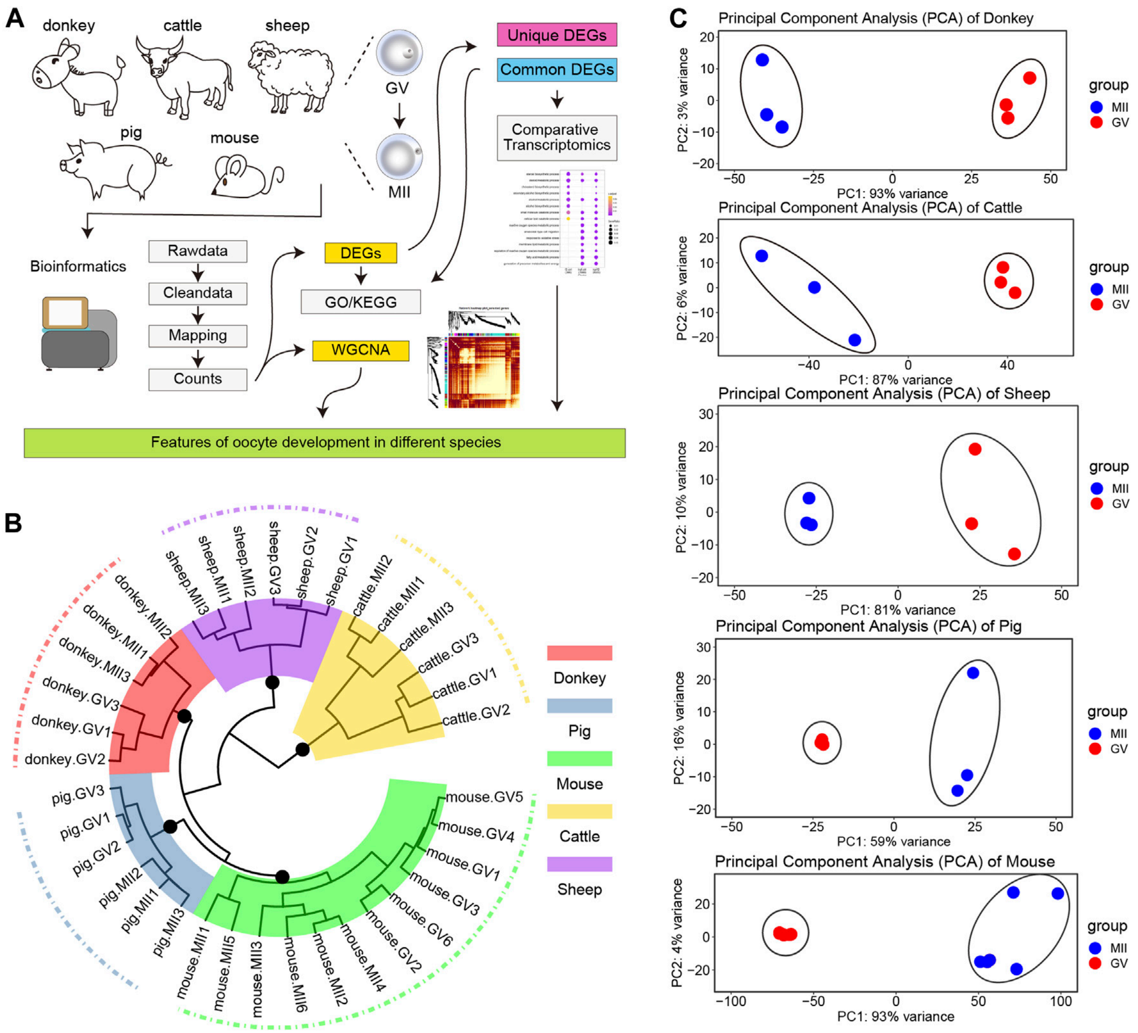


FIGURE 1 | Transcriptome landscape of oocyte development in different species. **(A)** The experimental design flow chart of this study. **(B)** Hierarchical clustering of the transcriptome landscape of oocyte development in different species; different colors represent different species: red, donkey; steel blue, pig; green, mouse; gold, cattle; and purple, sheep **(C)** Principal component analysis (PCA) of different species; from top to bottom: donkey, cattle, sheep, pig, and mouse.

pickSoftThreshold() of R package WGCNA to ensure a scale-free network, the soft threshold of β was set to 16. A hierarchical clustering dendrogram, a heat map, and a topology overlap matrix (TOM) were used to show the relationship between the functional modules and genes. Cytoscape software (v3.8.2) was used to exhibit the top 10 topological overlap relationships in functional modules related to donkey oocyte development during GV–MII.

Identification of Hub Genes and Protein-Protein Interaction Network Analysis

The hub genes, referring to highly interconnected nodes in functional modules, are considered as functionally important genes in WGCNA (Langfelder and Horvath, 2008). The hub genes were selected by the module membership and gene significance (MM & GS) method, and the conditional threshold was set as MM > 0.98 and SG > 0.8. Next, the hub genes were placed into the PPI network analysis through STRING (v11.5) (https://string-db.org/) to observe the interaction between genes.

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis

DEGs were processed for GO and KEGG analysis using clusterProfiler (v4.0.5) (Wu et al., 2021). Considering that different function annotation libraries may lead to differences in results, Metascape, a web-based and timeously updated biological annotation database was also used for the same analysis of hub genes (Zhou et al., 2019).

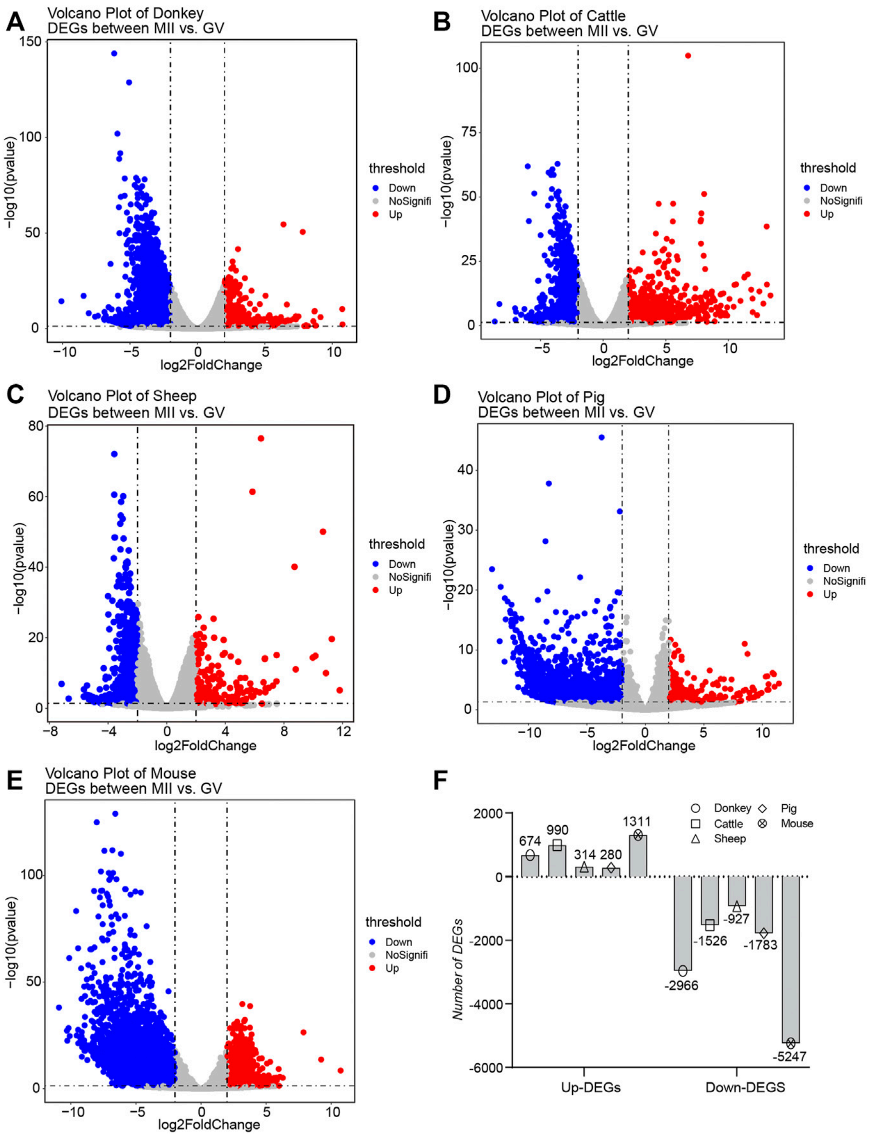


FIGURE 2 | Overview of the differentially expressed genes (DEGs) during oocyte development in different species **(A–E)** The volcano plot exhibits the differential expression levels of oocytes during the period from germinal vesicle (GV) to metaphase II (MII) in different species, including donkey, cattle, sheep, pig, and mouse. **(F)** The bar plot shows the number of DEGs during the period from GV to MII in different species.

RESULTS

Overview of the Transcriptome Landscape of Oocytes in Different Species

Committed to obtaining key regulatory factors and signal pathways for donkey oocyte development, we designed the experimental program shown in **Figure 1A**. First, we collected the single cell RNA sequencing data of GV and MII phase oocytes in the studied species (the detail information in **Supplementary Table S1**). To achieve improved analysis results, we took a strict bioinformatics analysis process (as described in Materials and Methods). Next, we focused on the DEGs of the different species, and adopted a plan of pairwise comparison and overall comprehensive comparison. Moreover, we carried out WGCNA of all uniformed detected genes of oocyte development during GV–MII in different species (**Figure 1A**). Unsurprisingly, it uncovered that the transcriptome profile of oocyte development during GV–MII between different species was species-specific, and in the same species, the hierarchical clustering indicated that there was a clear difference between GV and MII stages (**Figure 1B**). Furthermore, in all species, the GV and MII stages showed a clear dividing line, which indicated that the transcription characteristics of these two periods were clearly separated (**Figure 1C**).

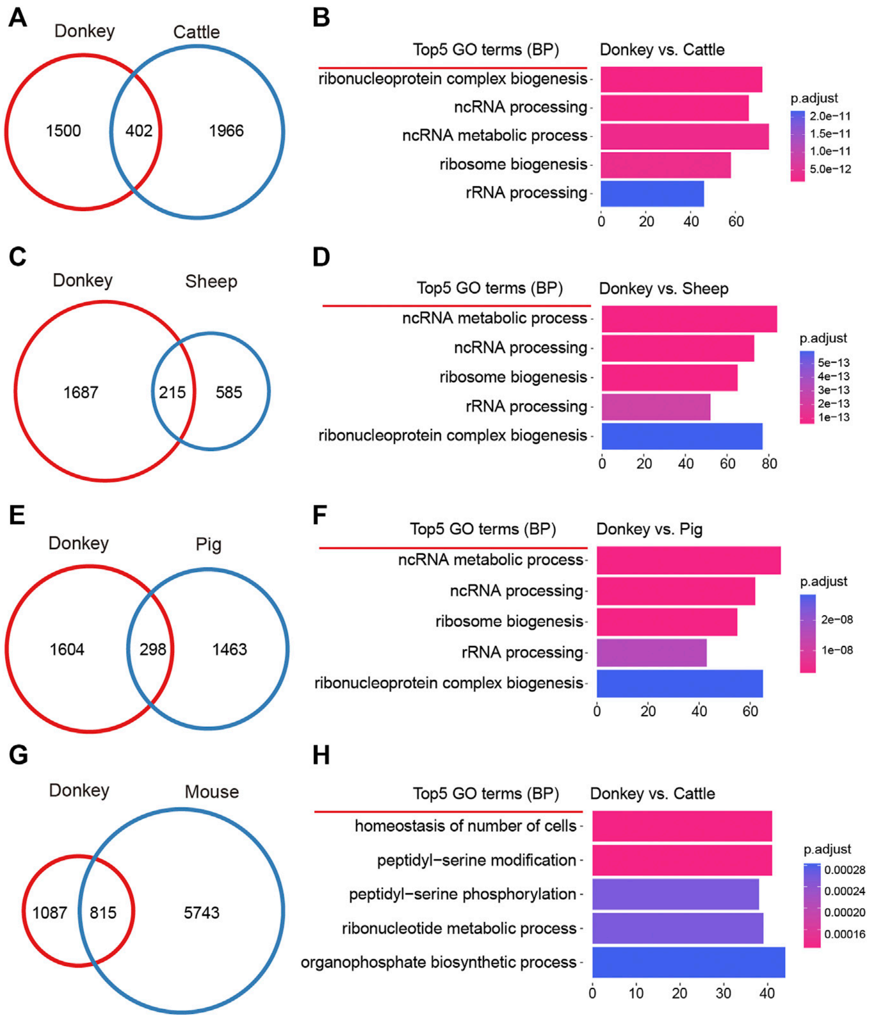


FIGURE 3 | Comparative analysis of donkey oocyte development during the period from germinal vesicle (GV) to metaphase II (MII). **(A)** The differences in oocyte development between donkeys and cattle during GV–MII. **(B)** The top5 GO (biological process) terms of donkey unique differentially expressed genes (DEGs) compared to cattle. **(C)** The differences in oocyte development between donkeys and sheep during GV–MII. **(D)** The top5 GO (biological process) terms of donkey unique DEGs compared to sheep. **(E)** The differences in oocyte development between donkeys and pigs during GV–MII. **(F)** The top5 GO (biological process) terms of donkey unique DEGs compared to pigs. **(G)** The differences in oocyte development between donkeys and mice during GV–MII. **(H)** The top5 GO (biological process) terms of donkey unique DEGs compared to mice.

Identification of the DEGs in Oocytes From GV to MII in Different Species

There is a broad consensus that DEGs tend to serve as key factors with the potential to regulate the transformation of cell fate (Stark et al., 2019; Zhang et al., 2019). Consequently, we collected detailed statistics of all the DEGs in oocyte development during GV–MII in different species. For the donkey, the number of down-regulation DEGs (down-DEGs) was 2,966, and the number of up-DEGs was 674 (**Figure 2A** and detail in **Supplementary Table S2**). For cattle, the number of down-DEGs was 1,526, and the number of up-DEGs was 990 (**Figure 2B** and detail in **Supplementary Table S3**). For sheep, the number of down-DEGs was 927, and the number of up-DEGs was 314 (**Figure 2C** and detail in **Supplementary Table S4**). For pigs, the number of down-DEGs was 1783, and the number of up-DEGs was 280 (**Figure 2D** and detail in **Supplementary Table S5**). For mice, the number of down-DEGs was 5,247, and the number of up-DEGs was 1,311 (**Figure 2E** and detail in **Supplementary Table S6**). In general, the number of down-DEGs was much greater than the number of up-DEGs; moreover, the number of DEGs detected in sheep was the least among all species (**Figure 2F**).

Difference of Oocyte Development in the Donkey Compared With Other Species

Organisms of different species have species-specific development (Yamashita et al., 2000); we may therefore ask whether oocyte development in different species also has species-specificity? To address this question, we made a detailed comparison of DEGs during oocyte development in donkeys and other species. For

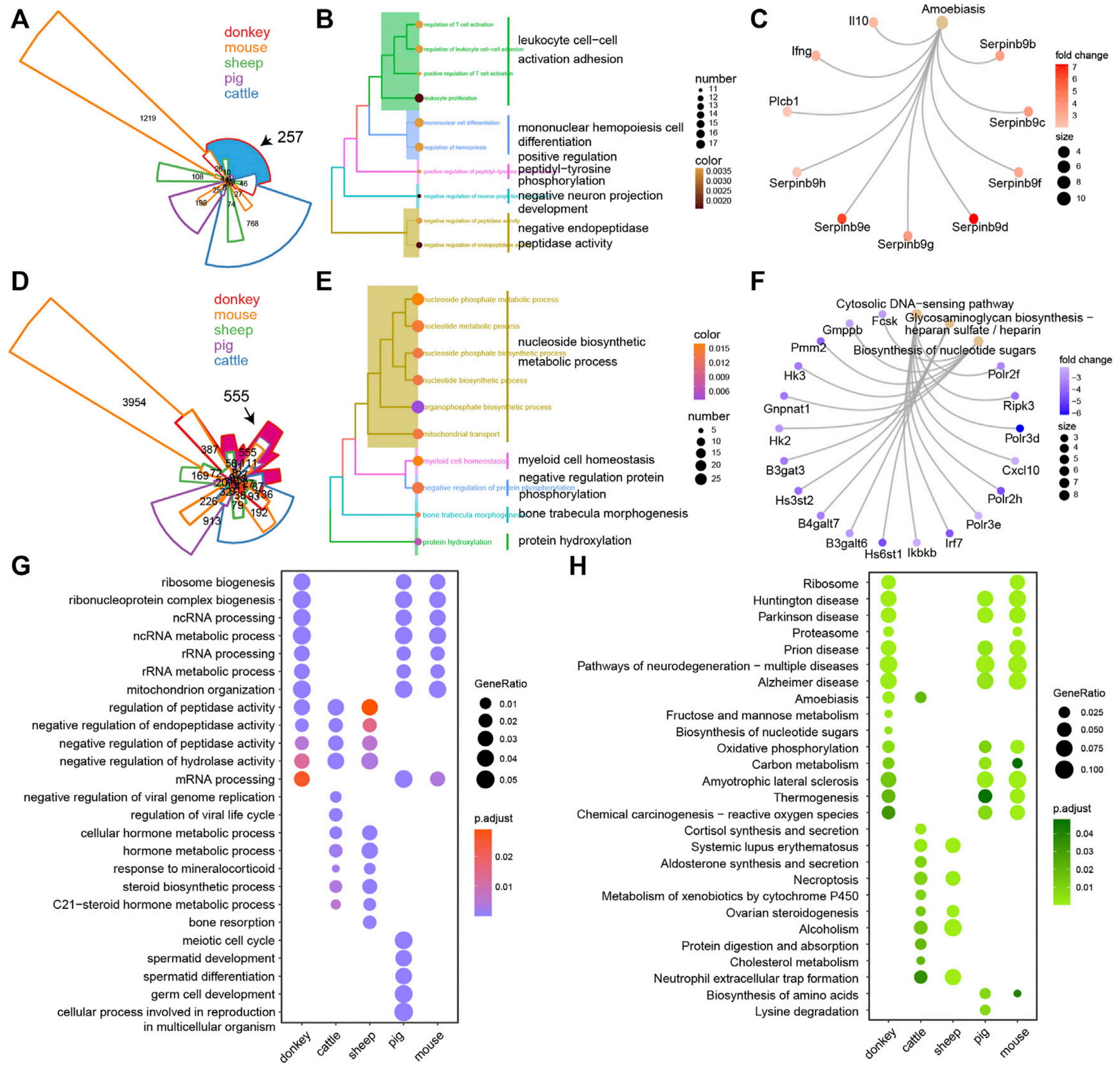


FIGURE 4 | The unique transcriptional characteristics of donkey oocyte development during the period from germinal vesicle (GV) to metaphase II (MII). **(A)** Venn plot of the up-regulated differentially expressed genes (up-DEGs) in different species during GV–MII. The black arrow refers to the donkey-specific up-DEGs compared to other species. The area size represents the number of genes, and the area marked in blue contains the donkey-specific up-DEGs. **(B and C)** The GO (biological process) terms and KEGG annotation of donkey-specific up-DEGs, respectively. **(D)** Venn plot of the down-DEGs in different species during GV–MII. The black arrow refers to the donkey-specific down-DEGs compared to other species. The area size represents the number of genes, and the area marked in rose red indicates the donkey-specific down-DEGs. **(E and F)** The GO (biological process) terms and KEGG annotation of the donkey-specific down-DEGs, respectively. **(G and H)** The dot plot shows the GO (biological process) terms and KEGG annotation of the DEGs of donkeys and other species from GV–MII, respectively.

donkeys and cattle, when comparing the DEGs of oocyte development during GV–MII, we found a total of 402 identical DEGs; and for donkeys, there were 1,500 donkey specific DEGs (**Figure 3A**). On the 1,500 donkey specific DEGs, we performed GO analysis, and found that the top5 terms focused on non-coding RNA (ncRNA) metabolic process (**Figure 3B**). For donkeys and sheep, we found a total of 215 identical DEGs, and 1,687 donkey specific DEGs (**Figure 3C**). GO analysis of the 1,687 donkey specific DEGs found that the top5 terms focused on the ncRNA metabolic process (**Figure 3D**). For donkeys and pigs, we found a total of 298 identical DEGs, and 1,604 donkey specific DEGs (**Figure 3E**) of which GO analysis indicated that the top5 terms focused on the ncRNA metabolic process (**Figure 3F**). For donkeys and mice, we found 815 identical DEGs, with 1,087 of donkey specific DEGs (**Figure 3G**) of which GO analysis, indicated that the top5 terms included the homeostasis of number of cells (**Figure 3H**).

The Uniqueness of Donkey Oocyte Development During GV–MII

In order to better understand the intricacies of donkey oocyte development, we comprehensively analyzed the similarities and differences of DEGs in oocyte development between the studied species. For up-DEGs, compared with all other tested species, there were 257 donkey-specific genes (**Figure 4A**). Furthermore, results of GO and KEGG analysis suggested that the 257 donkey specific up-DEGs were involved in cell differentiation and immune process (**Figures 4B,C**). Interleukin 10 (IL10) is a cytokine involved in inflammation and immunosuppression, and is engaged in the regulation of cell growth and differentiation. Moreover, it participates in apoptosis (Dhingra et al., 2009). Interestingly, we found that IL-10 was a donkey specific up-DEG (**Figure 4C**). For down-DEGs, there were 555 donkey specific genes (**Figure 4D**). GO and KEGG analysis indicated that the 555 donkey specific down-DEGs participated in nucleoside biosynthetic metabolic process and cytosolic DNA-sensing pathway (**Figures 4E,F**). To our surprise, we found that C-X-C motif chemokine 10 (CXCL10) was a donkey specific down-DEG (**Figure 4F**). Because CXCL10 is an inflammatory cytokine and involved in apoptosis, it stands to reason that these donkey specific down-DEGs are also involved in apoptosis. Moreover, we constructed a list of DEGs of all detected species, including up-DEGs and down-DEGs, and used clusterProfiler's *compareCluster* function to perform GO and KEGG function analysis on them. For GO analysis, compared with other species, the donkey was mainly reflected in the process of RNA metabolism regulation (**Figure 4G**). For KEGG analysis, not surprisingly, compared with other species, donkeys had significant enrichment in their ribosomal regulation process, which was well combined with the ncRNA and rRNA metabolism annotation in GO terms (**Figure 4H**).

WGCNA of Oocyte Development in Different Species

WGCNA, increasingly used in bioinformatics analysis, is able to accurately pick out the hub genes related to traits of interest in a complex gene expression profile (Pei et al., 2017). In order to make the correlation between genes conform to the scale-free network distribution, the appropriate soft threshold must be set correctly. Here, we set the soft threshold at 16 to assure the downstream analysis of WGCNA (**Supplementary Figure S1A**). Next, 20 gene functional modules were identified and exhibited as a hierarchical clustering dendrogram and a module eigengene adjacency heat map (**Supplementary Figure S1B**). The turquoise module contained the greatest number of genes (**Supplementary Figure S1B**). Moreover, the heat map of module-sample relationships elucidated that the different functional modules were species-specific, and donkey-specific functional modules were yellow, blue, and light cyan (**Figure 5A**). The heat map of module-trait relationships also suggested that the different functional modules were species-specific, and donkey-specific functional modules were yellow, blue, and light cyan (**Figure 5B**). In order to identify the core functional modules that regulate the development of donkey oocytes, we constructed a hybrid matrix pool between the traits of donkeys and functional modules, and performed cluster analysis. The results showed that the yellow modules were closely related to the development of donkey oocytes (**Figure 5C**).

Identification of Hub Genes Involved in Donkey Oocyte Development

The TOM plot indicated the interactions between gene functional modules, and the results revealed an obvious strong interactive relationship between different functional modules, which was related to the species specificity between functional modules (**Supplementary Figure S1C**). In order to further explore the regulatory relationship between genes in the yellow functional module, first the GO analysis of genes in the yellow functional module showed that these genes were involved in meiotic cell division such as DNA repair, and apoptosis process such as transcriptional regulation by the TP53 pathway (**Figure 6A**). Moreover, the top 10 genes with the strongest topology overlap relationship were exhibited as a circle graph (**Figure 6B**). Notably, the hub genes were selected by the MM & GS method (**Figure 6C**); finally, we obtained a total of 36 hub genes. After observing the expression characteristics of these hub genes in different species, we found that these genes, including wee1-like protein kinase 2 (WEE2) that is a notable meiotic gene (Hanna et al., 2010), were highly expressed in donkey oocytes, compared to other species (**Figure 6D**). In order to systematically understand the interaction of these hub genes, protein-protein interaction (PPI) network analysis showed that the proteins translated by these hub genes did not directly interact strongly, which suggested that these hub genes were involved in different biological pathways to regulate the specificity of donkey oocyte development (**Figure 6E**). Furthermore, GO analysis showed that these hub genes were related to the negative regulation of cell development (**Figure 6F**).

DISCUSSION

Donkey (*Equus asinus*), was domesticated around 3000 B.C. and became a traditional domestic animal whose ancestors are believed to include the African wild ass (Beja-Pereira et al., 2004; Rossel et al., 2008). With agricultural and industrial development and mechanization, the donkey has become somewhat redundant as a beast of burden and the global donkey population has notably reduced. China has the largest number of donkeys in Asia, estimated at more than three million in 2017 (Tian et al., 2020). However, as the demand for donkey-hide gelatin products increases, the number of donkeys in China is decreasing. There is now a need to address aspects of donkey breeding in order to preserve the species. Unfortunately, there is little research regarding the reproductive performance of donkeys (Tian et al., 2020; Li Z. et al., 2021). Hence, in this study, comparative transcriptomic and WGCNA were carried out to reveal the developmental characteristics of donkey oocytes, which would provide a theoretical basis for the protection of donkey germplasm.

The current study made an in-depth exploration of gene expression characteristics by thoroughly comparing the similarities and differences in oocyte development during GV–MII between donkeys and cattle, sheep, pigs, and mice (Qian et al., 2019; Wang et al., 2020b; Du et al., 2021; Li et al.,

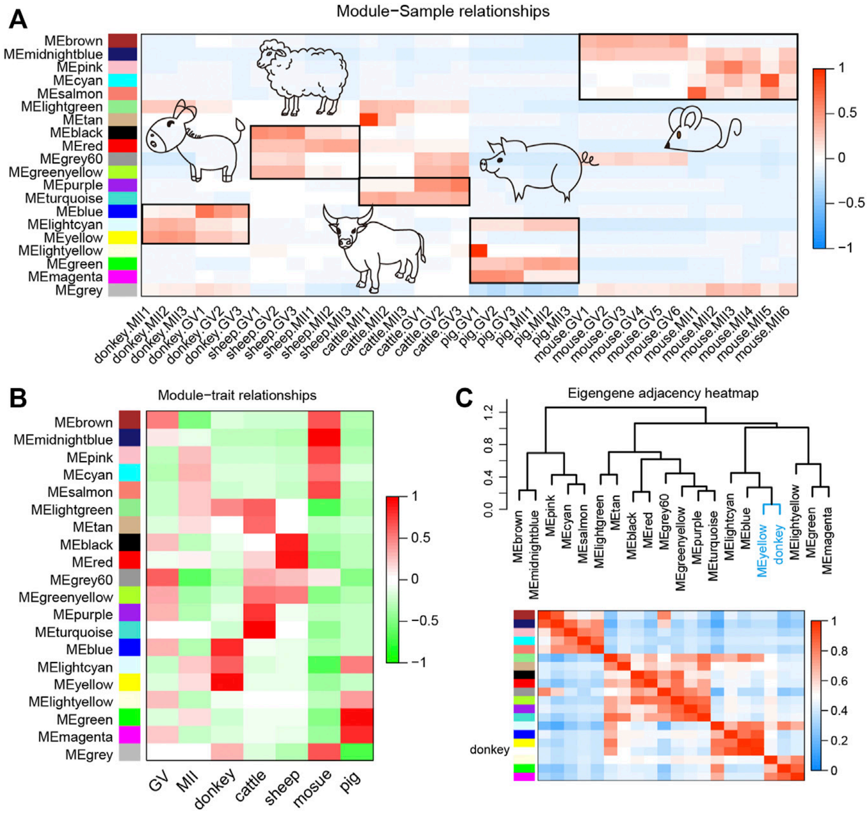


FIGURE 5 | Weighted gene co-expression network analysis (WGCNA) of oocyte development during the period from germinal vesicle (GV) to metaphase II (MII) in different species. **(A)** The heat map shows the correlation between different samples and functional modules identified by WGCNA. **(B)** The heatmap shows the correlation between different phenotypes and functional modules identified by WGCNA. **(C)** The correlation between donkey and functional modules identified by WGCNA. The yellow functional module is the key functional module involved in donkey oocyte development.

2021a; Li et al., 2021b), Furthermore, the release of the newly assembled donkey reference genome and the update of bioinformatics analysis tools (Wang C. et al., 2020), allowed us to annotate donkey genes more accurately, and to obtaining more refined results. Interestingly, in terms of transcriptome characteristics, we found that the GV and MII phases formed two groups of cells with a clear dividing line (**Figure 1**). Consistent with the results of previous studies, we found that in the studied species there were far fewer up-DEGs than down-DEGs (**Figure 2**). This phenomenon may be due to the gene transcription repression that occurs from the GV phase to the MII phase (De La Fuente et al., 2004); previous studies report that maternal mRNA is selectively degraded during development from GV to MII (Su et al., 2007; Zhao et al., 2020). Many studies suggested that non-coding RNAs (ncRNAs) were related to oocyte development (Barragán et al., 2017; Bouckenheimer et al., 2018). Interestingly, a study reported that ncRNAs were related to age and ovarian reserve in human (Barragán et al., 2017). But, the reports of ncRNAs in donkey oocyte development is few.

Unexpectedly, comparative transcriptomic analysis showed that donkey-specific DEGs are involved in the immune response, compared with other species. IL10 is an anti-inflammatory cytokine that is involved in apoptosis (Dhingra et al., 2009), and in this study, we found that it was a donkey specific up-DEG (**Figure 4C**). It has been reported in the literature that the expression of IL10 in granulosa cells is related to the success rate of IVF (Çavus and Deger, 2020). Moreover, the expression of IL10 attenuates the apoptosis rate of cells (Dhingra et al., 2009). However, few studies reported the role of IL10 in donkey oocytes. Interestingly, donkey specific down-DEGs, such as CXCL10 (**Figure 4F**), also participated in immune processes. It is reported that, CXCL10 serves as a marker of apoptosis (Sui et al., 2004), and the overexpression of CXCL10 increases the apoptosis of cells (Sui et al., 2006). The up-regulation of anti-apoptotic factors and the down-regulation of pro-apoptotic factors in donkey oocytes may be related to the maintenance of cell survival during oocyte development.

WGCNA was used to find co-expressed gene modules and to explore the relationship between gene networks and the phenotype of interest, as well as the core genes in the network; it is an effective method for analyzing the regulatory relationship of complex gene expression patterns (Pei et al., 2017). There is a consensus that the mechanism of oocyte development and maturation is species-specific (Yamashita

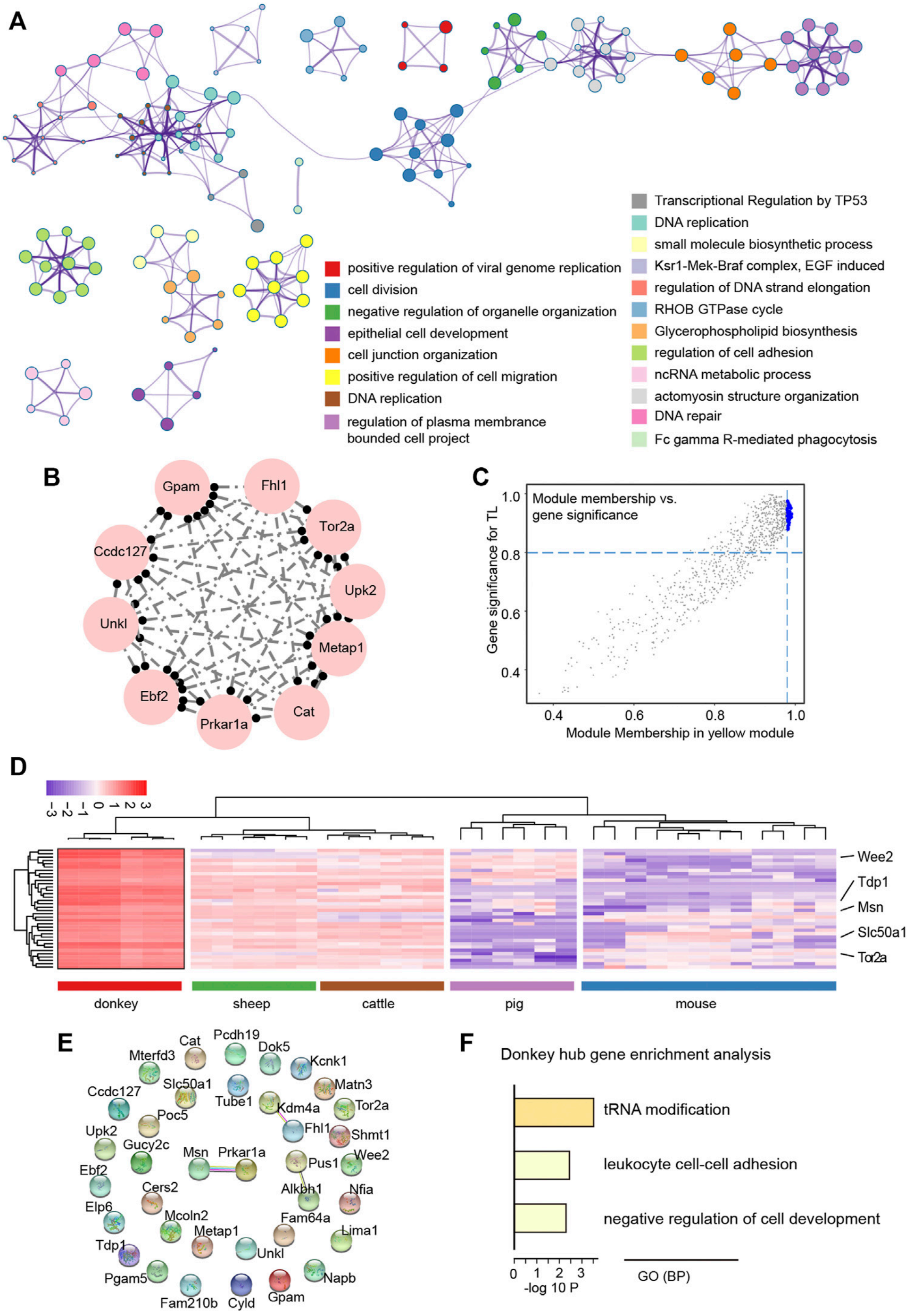


FIGURE 6 | Functional exploration of hub genes related to donkey oocyte development identified by weighted gene co-expression network analysis (WGCNA). **(A)** The GO and KEGG annotation of genes in the yellow functional module. **(B)** The top 10 topological overlap relationship in the yellow functional module. **(C)** The scatter plot shows the relationship between module membership and gene significance (MM & GS). The large blue dots represent hub genes related to donkey oocyte development. **(D)** The heat map shows the expression levels of hub genes selected by the yellow functional module. **(E)** The protein-protein interaction network of hub genes. The connection between proteins represents the degree of protein interaction. **(F)** The GO (biological process) terms of hub genes related to donkey oocyte development.

et al., 2000). However, WGCNA data in the current study showed similar results for all studied species (**Figure 5C**). Subsequently, we explored the genes in the modules most related to donkeys and the results suggested that, consistent with the results of comparative transcriptomic analysis, these genes are closely related to the meiotic cell cycle and immune processes. Moreover, we selected 36 hub genes including WEE2, which was consistent with previous reports. Previous research reported that the overexpression of WEE2 is able to postpone the reentry of oocytes into meiosis in both mice and monkeys (Hanna et al., 2010). Furthermore, Wee2-deficient mice exhibit fertilization failure and female infertility (Sang et al., 2018).

CONCLUSION

Overall, our research brought new perspectives regarding the development of donkey oocytes. Compared with other species, donkey oocytes express a large number of genes related to RNA metabolism to maintain normal oocyte development during GV–MII. Our study offers a theoretical basis for improving oocyte maturation in the donkey.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Qingdao Agricultural University.

AUTHOR CONTRIBUTIONS

F-LZ and WS designed the study and WS managed funding. F-LZ did the bioinformatics. S-EZ and Y-JS collected the data, F-LZ and J-JW drafted the manuscript. The final manuscript approved by all authors.

FUNDING

This work was supported by the Start-up Fund for High-level Talents of Qingdao Agricultural University, Major Agricultural Application Technology Innovation Projects of Shandong Province (SD2019XM008), Modern Agricultural Industrial Technology System of Shandong Province (SDAIT-27), and Taishan Scholar Construction Foundation of Shandong Province (ts20190946) of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.839207/full#supplementary-material

Supplementary Figure S1 | (A) The evaluation of WGCNA genes of oocyte development during GV–MII in different species. Left panel, analysis of the scale-free fit index for various soft thresholding powers. The blue line represents a scale-free fit index of 0.9. Right panel, analysis of the mean connectivity for various soft-thresholding powers. **(B)** The hierarchical cluster tree shows functional modules distinguished by WGCNA. Bottom part, different colors represent different modules. **(C)** The heat map plot shows the topological overlap matrix (TOM) among 500 randomly selected genes. Light purple color depict a small overlap, and the red color indicates a greater overlap.

REFERENCES

Aganga, A., Aganga, A., Thema, F., and Obocheleng, K. (2003). Carcass Analysis and Meat Composition of the Donkey. *Pakistan J. Nutr.* 2, 138–147. doi:10.3923/pjn.2003.138.147

Andrews, S. (2010). "FastQC: A Quality Control Tool for High Throughput Sequence Data," in *Babraham Bioinformatics* (Cambridge, United Kingdom: Babraham Institute).

Bai, F., Wang, D. Y., Fan, Y. J., Qiu, J., Wang, L., Dai, Y., et al. (2020). Assisted Reproductive Technology Service Availability, Efficacy and Safety in mainland China: 2016. *Hum. Reprod.* 35, 446–452. doi:10.1093/humrep/dez245

Barragán, M., Pons, J., Ferrer-Vaquer, A., Cornet-Bartolomé, D., Schweitzer, A., Hubbard, J., et al. (2017). The Transcriptome of Human Oocytes Is Related to Age and Ovarian reserve. *MHR: Basic Sci. Reprod. Med.* 23, 535–548. doi:10.1093/molehr/gax033

Beja-Pereira, A., England, P. R., Ferrand, N., Jordan, S., Bakhiet, A. O., Abdalla, M. A., et al. (2004). African Origins of the Domestic Donkey. *Science* 304, 1781. doi:10.1126/science.1096008

Bouckenheimer, J., Fauque, P., Lecellier, C. H., Bruno, C., Commes, T., Lemaître, J. M., et al. (2018). Differential Long Non-Coding RNA Expression Profiles in Human Oocytes and Cumulus Cells. *Sci. Rep.* 8, 2202. doi:10.1038/s41598-018-20727-0

Çavus, Y., and Deger, U. (2020). Investigation of ADAMTS-1, BAX and IL-10 Expressions in Granulosa Cells Determinanting Fertilization Success in IVF. *Int. J. Morphol.* 38, 427–434. doi:10.4067/S0717-95022020000200427

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560

De La Fuente, R., Viveiros, M. M., Burns, K. H., Adashi, E. Y., Matzuk, M. M., and Eppig, J. J. (2004). Major Chromatin Remodeling in the Germinal Vesicle (GV) of Mammalian Oocytes Is Dispensable for Global Transcriptional Silencing but Required for Centromeric Heterochromatin Function. *Develop. Biol.* 275, 447–458. doi:10.1016/j.ydbio.2004.08.028

Dhingra, S., Sharma, A. K., Arora, R. C., Slezak, J., and Singal, P. K. (2009). IL-10 Attenuates TNF- -Induced NF B Pathway Activation and Cardiomyocyte Apoptosis. *Cardiovasc. Res.* 82, 59–66. doi:10.1093/cvr/cvp040

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Du, Z. Q., Liang, H., Liu, X. M., Liu, Y. H., Wang, C., and Yang, C. X. (2021). Single Cell RNA-Seq Reveals Genes Vital to *In Vitro* Fertilized Embryos and Parthenotes in Pigs. *Sci. Rep.* 11, 14393. doi:10.1038/s41598-021-93904-3

Eppig, J. (1996). Coordination of Nuclear and Cytoplasmic Oocyte Maturation in Eutherian Mammals. *Reprod. Fertil. Dev.* 8, 485–489. doi:10.1071/rd9960485

Goudet, G., Douet, C., Kaabouba-Escurier, A., Couty, I., Moros-Nicolás, C., Barrière, P., et al. (2016). Establishment of Conditions for Ovum Pick up and IVM of Jennies Oocytes toward the Setting up of Efficient IVF and *In Vitro* Embryos Culture Procedures in Donkey (*Equus A*). *Theriogenology* 86, 528–535. doi:10.1016/j.theriogenology.2016.02.004

Hanna, C. B., Yao, S., Patta, M. C., Jensen, J. T., and Wu, X. (2010). WEE2 Is an Oocyte-Specific Meiosis Inhibitor in Rhesus Macaque Monkeys1. *Biol. Reprod.* 82, 1190–1197. doi:10.1095/biolreprod.109.081984

Kim, J.-S., Kim, D., Kim, H.-J., and Jang, A. (2018). Protection Effect of Donkey Hide Gelatin Hydrolysates on UVB-Induced Photoaging of Human Skin Fibroblasts. *Process Biochem.* 67, 118–126. doi:10.1016/j.procbio.2018.02.004

Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2--an R Package for Gene List Functional Enrichment Analysis and Namespace

Conversion Toolset G: Profiler. *F1000Research* 9, 709. doi:10.12688/f1000research.24956.1

Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC bioinformatics* 9, 559–613. doi:10.1186/1471-2105-9-559

Li, M.-H., Niu, M.-H., Feng, Y.-Q., Zhang, S.-E., Tang, S.-W., Wang, J.-J., et al. (2021a). Establishment of lncRNA-mRNA Network in Bovine Oocyte between Germinal Vesicle and Metaphase II Stage. *Gene* 791, 145716. doi:10.1016/j.gene.2021.145716

Li, Z., Song, X., Yin, S., Yan, J., Lv, P., Shan, H., et al. (2021b). Single-Cell RNA-Seq Revealed the Gene Expression Pattern during the *In Vitro* Maturation of Donkey Oocytes. *Genes* 12, 1640. doi:10.3390/genes12101640

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* 30, 923–930. doi:10.1093/bioinformatics/btt656

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Oh, J. S., Susor, A., and Conti, M. (2011). Protein Tyrosine Kinase Wee1B Is Essential for Metaphase II Exit in Mouse Oocytes. *Science* 332, 462–465. doi:10.1126/science.1199211

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., et al. (2013). Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse. *Nature* 499, 74–78. doi:10.1038/nature12323

Pei, G., Chen, L., and Zhang, W. (2017). WGCNA Application to Proteomic and Metabolomic Data Analysis. *Methods Enzymol.* 585, 135–158. doi:10.1016/bs.mie.2016.09.016

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells. *Nat. Methods* 10, 1096–1098. doi:10.1038/nmeth.2639

Polidori, P., Vincenzetti, S., Cavallucci, C., and Beghelli, D. (2008). Quality of Donkey Meat and Carcass Characteristics. *Meat Sci.* 80, 1222–1224. doi:10.1016/j.meatsci.2008.05.027

Polidori, P., Cavallucci, C., Beghelli, D., and Vincenzetti, S. (2009). Physical and Chemical Characteristics of Donkey Meat from Martina Franca Breed. *Meat Sci.* 82, 469–471. doi:10.1016/j.meatsci.2009.03.001

Qian, Y., Liao, J., Suen, A. H. C., Lee, A. W. T., Chung, H. S., Tang, N. L. S., et al. (2019). Comparative Analysis of Single-Cell Parallel Sequencing Approaches in Oocyte Application. *Int. J. Biochem. Cel Biol.* 107, 1–5. doi:10.1016/j.biocel.2018.12.003

Renaud, G., Petersen, B., Seguin-Orlando, A., Bertelsen, M. F., Waller, A., Newton, R., et al. (2018). Improved De Novo Genomic Assembly for the Domestic Donkey. *Sci. Adv.* 4, eaaq0392. doi:10.1126/sciadv.aaq0392

Rossel, S., Marshall, F., Peters, J., Pilgram, T., Adams, M. D., and O'connor, D. (2008). Domestication of the Donkey: Timing, Processes, and Indicators. *Proc. Natl. Acad. Sci.* 105, 3715–3720. doi:10.1073/pnas.0709692105

Sang, Q., Li, B., Kuang, Y., Wang, X., Zhang, Z., Chen, B., et al. (2018). Homozygous Mutations in WEE2 Cause Fertilization Failure and Female Infertility. *Am. J. Hum. Genet.* 102, 649–657. doi:10.1016/j.ajhg.2018.02.015

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA Sequencing: The Teenage Years. *Nat. Rev. Genet.* 20, 631–656. doi:10.1038/s41576-019-0150-2

Su, Y.-Q., Sugiura, K., Woo, Y., Wigglesworth, K., Kamdar, S., Affourtit, J., et al. (2007). Selective Degradation of Transcripts during Meiotic Maturation of Mouse Oocytes. *Develop. Biol.* 302, 104–117. doi:10.1016/j.ydbio.2006.09.008

Sui, Y., Potula, R., Dhillon, N., Pinson, D., Li, S., Nath, A., et al. (2004). Neuronal Apoptosis Is Mediated by CXCL10 Overexpression in Simian Human Immunodeficiency Virus Encephalitis. *Am. J. Pathol.* 164, 1557–1566. doi:10.1016/s0002-9440(10)63714-5

Sui, Y., Stehno-Bittel, L., Li, S., Loganathan, R., Dhillon, N. K., Pinson, D., et al. (2006). CXCL10-Induced Cell Death in Neurons: Role of Calcium Dysregulation. *Eur. J. Neurosci.* 23, 957–964. doi:10.1111/j.1460-9568.2006.04631.x

Tian, F., Wang, J., Li, Y., Yang, C., Zhang, R., Wang, X., et al. (2020). Integrated Analysis of mRNA and miRNA in Testis and Cauda Epididymidis Reveals Candidate Molecular Markers Associated with Reproduction in Dezhou Donkey. *Livestock Sci.* 234, 103885. doi:10.1016/j.livsci.2019.103885

Wang, C., Li, H., Guo, Y., Huang, J., Sun, Y., Min, J., et al. (2020a). Donkey Genomes Provide New Insights into Domestication and Selection for Coat Color. *Nat. Commun.* 11, 6014–6015. doi:10.1038/s41467-020-19813-7

Wang, J.-J., Niu, M.-H., Zhang, T., Shen, W., and Cao, H.-G. (2020b). Genome-Wide Network of lncRNA-mRNA during Ovine Oocyte Development from Germinal Vesicle to Metaphase II *In Vitro*. *Front. Physiol.* 11, 1019. doi:10.3389/fphys.2020.01019

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). ClusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *The Innovation* 2, 100141. doi:10.1016/j.xinn.2021.100141

Yamashita, M., Mita, K., Yoshida, N., and Kondo, T. (2000). Molecular Mechanisms of the Initiation of Oocyte Maturation: General and Species-Specific Aspects. *Prog. Cel. Cycle Res.* 4, 115–129. doi:10.1007/978-1-4615-4253-7_11

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. Y. (2017). Ggtree : An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol. Evol.* 8, 28–36. doi:10.1111/2041-210x.12628

Zhang, F.-L., Li, N., Wang, H., Ma, J.-M., Shen, W., and Li, L. (2019). Zearalenone Exposure Induces the Apoptosis of Porcine Granulosa Cells and Changes Long Noncoding RNA Expression to Promote Antiapoptosis by Activating the JAK2-STAT3 Pathway. *J. Agric. Food Chem.* 67, 12117–12128. doi:10.1021/acs.jafc.9b05189

Zhao, Z. H., Meng, T. G., Li, A., Schatten, H., Wang, Z. B., and Sun, Q. Y. (2020). RNA-Seq Transcriptome Reveals Different Molecular Responses during Human and Mouse Oocyte Maturation and Fertilization. *BMC genomics* 21, 475–511. doi:10.1186/s12864-020-06885-4

Zhao, Y., Li, M.-C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., et al. (2021). TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-Seq Data from the NCI Patient-Derived Models Repository. *J. Translational Med.* 19, 1–15. doi:10.1186/s12967-021-02936-w

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nat. Commun.* 10, 1523–1610. doi:10.1038/s41467-019-09234-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer TZ declared a past co-authorship with several of the authors JJW, WS to the handling editor.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fgene.2021.815270

A Simple and Efficient Method to Cryopreserve Human Ejaculated and Testicular Spermatozoa in –80°C Freezer

Xiaohan Wang[1†], *Fangting Lu*[2†], *Shun Bai*[2†], *Limin Wu*[2], *Lingli Huang*[2], *Naru Zhou*[2], *Bo Xu*[2], *Yangyang Wan*[2], *Rentao Jin*[2*], *Xiaohua Jiang*[2*] and *Xianhong Tong*[1*]

[1]Provincial Hospital Affiliated to Anhui Medical University, Hefei, China, [2]Division of Life Sciences and Medicine, Reproductive and Genetic Hospital, The First Affiliated Hospital of USTC, University of Science and Technology of China, Hefei, China

Edited by:
Mengcheng Luo,
Wuhan University, China

Reviewed by:
Jinmin Gao,
Shandong Normal University, China
Jasmer Dalal,
Lala Lajpat Rai University of Veterinary and Animal Sciences, India

***Correspondence:**
Xiaohua Jiang
biojxh@ustc.edu.cn
Xianhong Tong
tong68xianhong@163.com,
Rentao Jin
jrtao2001@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 15 November 2021
Accepted: 27 December 2021
Published: 28 January 2022

Citation:
Wang X, Lu F, Bai S, Wu L, Huang L, Zhou N, Xu B, Wan Y, Jin R, Jiang X and Tong X (2022) A Simple and Efficient Method to Cryopreserve Human Ejaculated and Testicular Spermatozoa in –80°C Freezer. Front. Genet. 12:815270. doi: 10.3389/fgene.2021.815270

Human autologous sperm freezing involves ejaculated sperm, and testicular or epididymal puncture sperm freezing, and autologous sperm freezing is widely used in assisted reproductive technology. In previous studies, researchers have tried to cryopreserve sperm from mammals (rats, dogs, etc.) using a –80°C freezer and have achieved success. It is common to use liquid nitrogen vapor rapid freezing to cryopreserve human autologous sperm. However, the operation of this cooling method is complicated, and the temperature drop is unstable. In this study, we compared the quality of human ejaculation and testicular sperm after liquid nitrogen vapor rapid freezing and –80°C freezing for the first time. By analyzing sperm quality parameters of 93 ejaculated sperm and 10 testicular sperm after liquid nitrogen vapor rapid freezing and –80°C freezing, we found reactive oxygen species (ROS) of sperm of the –80°C freezer was significantly lower than liquid nitrogen vapor rapid freezing. Regression analysis showed that progressive motility, ROS, and DNA fragmentation index (DFI) in post-thaw spermatozoa were correlated with sperm progressive motility, ROS, and DFI before freezing. For the freezing method, the –80°C freezer was positively correlated with the sperm progressive motility. Among the factors of freezing time, long-term freezing was negatively correlated with sperm progressive motility and ROS. Although freezing directly at –80°C freezer had a slower temperature drop than liquid nitrogen vapor rapid freezing over the same period, the curves of the temperature drop were similar, and slight differences in the freezing point were observed. Furthermore, there were no statistically significant differences between the two methods for freezing testicular sperm. The method of direct –80°C freezing could be considered a simplified alternative to vapor freezing for short-term human sperm storage. It could be used for cryopreservation of autologous sperm (especially testicular sperm) by *in vitro* fertilization centers.

Clinical Trial Registration: (website), identifier (ChiCTR2100050190).

Abbreviations: DFI, DNA fragmentation index; FITC-PSA, fluorescein isothiocyanate-pisum sativum agglutinin; HDS, high DNA stainability; ICSI, intracytoplasmic sperm injection; IVF, *in vitro*fertilization; MMP, mitochondrial membrane potential; RT, room temperature; ROS, reactive oxygen species; SCSA, the sperm chromatin structure assay; CASA, computer-assisted semen analysis; SPSS, statistical product and service solutions; SD, standard deviations; TESA, testicular sperm aspiration; ZP, zona pellucida, AR, acrosome reaction.

Keywords: human sperm, cryopreservation, −80°C freezer, testicular sperm, liquid nitrogen vapor rapid freezing

INTRODUCTION

Human sperm cryopreservation has been widely used for human reproduction. Autologous sperm freezing is the general method applied *in vitro* fertilization (IVF) laboratories in a variety of circumstances ranging from fertility preservation for cancer patients to the clinical management of male infertility (Trottmann et al., 2007).

The first attempt to freeze semen dates can be traced back to 1776 when Abbot et al. (Royere, et al., 1996) reported that snow could store sperm by cooling. Advances in cryobiology seen over the past decades and cryoprotective medium, in particular, have significantly aided sperm freezing-resuscitation technology (Sherman, 1973). A variety of freezing methods have been discovered with the development of cryopreservation technology: slow freezing, liquid nitrogen vapor rapid freezing, and vitrification (Tao et al., 2020). Currently, the mainstream method for semen freezing or freezing a small number of spermatozoa samples is liquid nitrogen vapor rapid freezing (vapor rapid freezing) (Huang et al., 2020). The standard cryopreservation method normally involves freezing human sperm in liquid nitrogen vapor to −80°C and then stored in liquid nitrogen. Previous reports had shown that samples are placed at 1–10 cm above the liquid nitrogen surface for 10–30 min, followed by storage in liquid nitrogen (Gwatkin, 1991; Le et al., 2019). Vapor rapid freezing cannot control the rate of temperature drop caused by the volatilization of liquid nitrogen (Di Santo et al., 2012). Another limitation of liquid nitrogen vapor rapid freezing is that it cannot cool many semen samples simultaneously.

Freezing semen samples by directly placing them in a −80°C freezer is the most practical means because a −80°C freezer is readily available. A few researchers have reported that several mammalian spermatozoa can be successfully frozen by direct placement in a −80°C freezer. For example, Marcello Raspa et al. (2017), Raspa et al. (2018a), Raspa et al. (2018b), reported that mouse spermatozoa can be frozen, transported, shared, and stored at −80°C for a long time without a significant loss of viability. Pezo et al. (2017) demonstrated that semen freezing and storage using a −80°C ultra-freezer is an effective technique for the long-term preservation of canine spermatozoa. Parkes. (1945) indicated that human sperm cryopreservation without cryoprotectants at −79°C offered an advantage over storage at −196°C. In addition, Liu et al. (2016) concluded that storage of neat semen samples at −80°C caused milder damage to sperm DNA than storage at −196°C mixed with cryoprotectants. In 2012, Sanchez et al. (2012) concluded that there were no significant differences in sperm progressive motility, the integrity of mitochondrial membrane potential (MMP) or DNA fragmentation for vitrified swim-up human sperm either at −196°C under liquid nitrogen or at -80°C. However, Vaz et al. (2018) found storage of human sperm at −80°C freezer up to 96 hours was detrimental to sperm viability. Therefore, the impact of −80°C freezer on human sperm needs to be further explored.

The cryopreservation of limited spermatozoa from men afflicted with nonobstructive azoospermia using testicular sperm aspiration (TESA) can avoid repeated surgery and promote the preservation of fertility (Miller et al., 2017). Current research on the freezing of testicular or epididymal sperm is mainly focused on single sperm freezing. In 1997, Cohen et al. (1997) first described a novel cryopreservation technique for single sperm using an empty zona pellucida (ZP). Various single sperm cryopreservation carriers have been proposed in the last 20 years, including ZPs (Hsieh et al., 2000; Just et al., 2004), cryoloops (Desai et al., 2004), culture dishes (Sereni et al., 2008), cell sleepers (Coetzee et al., 2016), cryotops (Endo et al., 2012), and novel sperm vitrification devices (Berkovitz et al., 2018). Each cryopreservation method has limitations; in particular, the usage of these carriers requires experienced technicians to select and capture single sperm, which always results in very few available sperm after this process. Therefore, frozen testicular sperm with single sperm has not become widespread, and an ideal container that can be universally used needs to be developed.

To explore whether freezing sperm at −80°C freezer is feasible for human ejaculate and testicular spermatozoa, we evaluated the effects of freezing sperm at −80°C freezer and vapor rapid freezing techniques on sperm quality, and simplified the procedure and equipment for freezing human ejaculate and testicular spermatozoa.

MATERIALS AND METHODS

The procedures in this study were approved by the Ethical Committee of The First Affiliated Hospital of the University of Science and Technology of China (2021-KY-040), and informed consent was obtained from all subjects.

Source of Sperm Samples

The study included 93 ejaculated sperm from normozoospermic men who sought fertility evaluation at the Reproductive Center of The First Affiliated Hospital of University of Science and Technology of China between March 2021 and June 2021. And ten testicular tissue from patients who were diagnosed with obstructive azoospermia. Based on medical history and seminal examination findings, patients with vasectomy, varicocele, cryptorchidism, or genital infection were excluded from the analysis. Patients receiving any medication or antioxidant supplementation in 3 months before the study were also excluded.

Experimental Design

A diagram of the experimental design is shown in **Figure 1**. The experiments were split into four sections.

Section 1

Thirty ejaculated sperm from normozoospermic men were analyzed for progressive motility and viability by vapor rapid freezing or direct −80°C freezing. The sperm samples were thawed

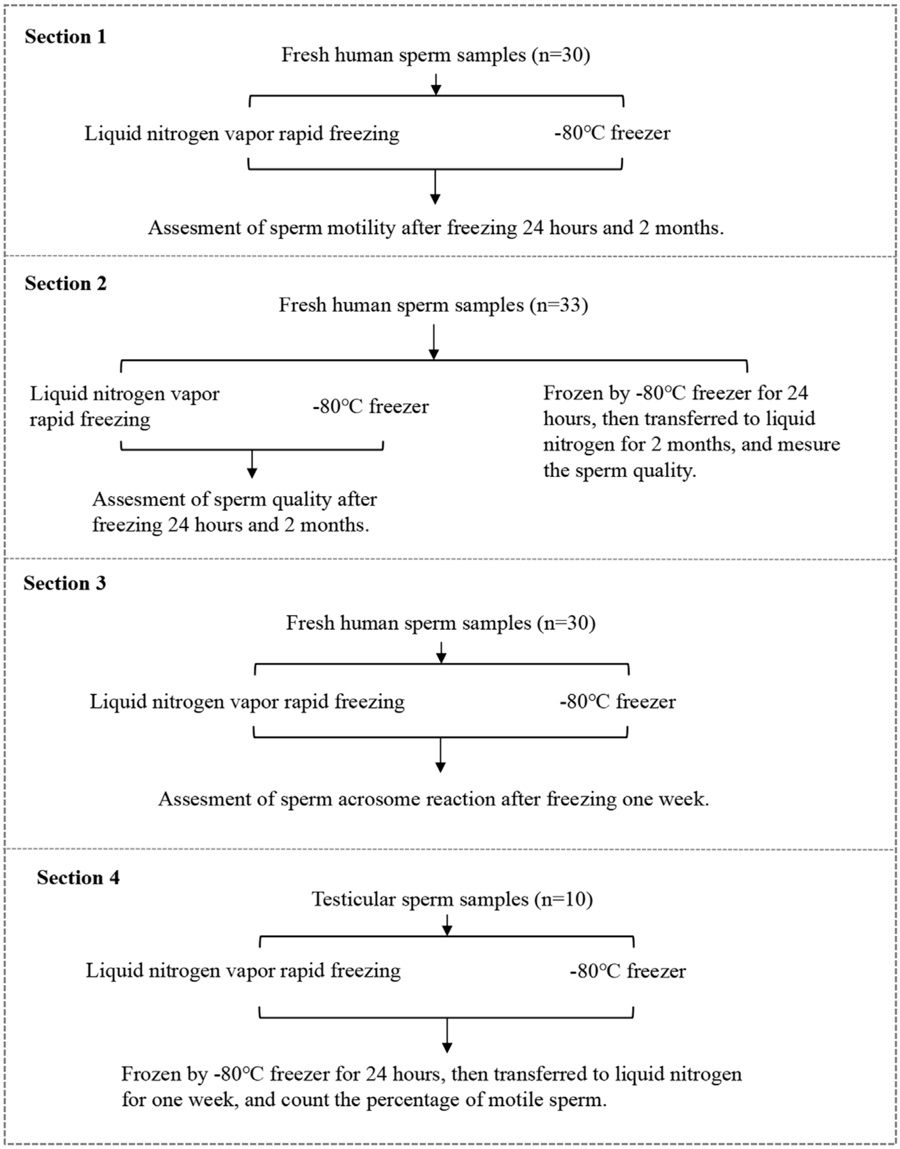


FIGURE 1 | Diagrammatic representation of the experimental design.

and sperm parameters were measured 24 h (short-term) or 2 months (long-term) later.

Section 2

Thirty-three ejaculated sperm from normozoospermic men were analyzed for sperm quality by vapor rapid freezing or direct −80°C freezing, kept for 24 h in a −80°C freezer and then moved into liquid nitrogen. The sperm samples were thawed and sperm parameters were measured 24 h (short-term) or 2 months (long-term) later.

Overall, there were five groups in section 1 and section 2 (I) frozen by vapor rapid freezing for short-term (short vapor group); (II) frozen by vapor rapid freezing for long-term (long vapor group); (III) frozen in -80°C freezers for short-term (short freezer group); (IV) frozen in −80°C freezers for long-term (long freezer group); (V) frozen in −80°C freezers for 24 h and then immersed in liquid nitrogen for 2 months (liquid nitrogen group).

Section 3

Thirty ejaculated sperm from normozoospermic men were analyzed for acrosome reaction after a week by vapor rapid freezing and direct −80°C freezing.

Section 4

Vapor rapid freezing and direct −80°C freezing were used to freeze 10 testicular sperm for a week. This section was designed to compare the percentage of motile sperm between vapor rapid freezing and freezing at −80°C.

Method of –80°C Freezing

The liquefied ejaculated semen was diluted with sperm freezing media (ORIGIO, Måløv, Denmark) (ratio 1:1), and 1 ml of the suspension was pipetted into a Nunc cryotube vial (1.8 ml; catalog number: 375418, Thermo Fisher Scientific, Jiangsu, China), which was kept at room temperature (RT) for 10 min and subsequently placed in a horizontal position in the freezer. Some samples were transferred to liquid nitrogen 24 h later. The purpose of placing the container in a horizontal position was to minimize the heat difference between the two ends during freezing (Di Santo et al., 2012).

Method of Liquid Nitrogen Vapor Rapid Freezing

After 1 ml of suspension was pipetted into a 1.8 ml aseptic cryotube, the samples were kept at RT for 10 min and subsequently placed horizontally 8–10 cm above the liquid nitrogen surface. Fifteen minutes later, it was submerged in liquid nitrogen and stored at −196°C in a tank full of liquid nitrogen for 24 h or 2 months.

Method of Ejaculated Sperm Thawing

Thawing was performed as described previously for a specific sperm freezing medium with some modifications (ORIGIO, Måløv, Denmark). The cryotube was removed from the liquid nitrogen or −80°C freezer and submerged in warm water (37°C) for 10 min. Post-thaw sperm progressive motility, viability, reactive oxygen species (ROS), MMP, DNA fragmentation index (DFI), and high DNA stainability (HDS) were analyzed.

Cryopreserved Spermatozoa Collected by TESA

After removing the seminiferous tubules from the testicles, the sample was placed in a dish with 5% 3-(N-Morpholino) propanesulphonic acid with gentamicin (G-MOPS^{TM}, Vitrolife, Sweden) medium and minced with a needle connected to a 1 ml syringe under a dissecting microscope. Then, the sample was examined under an inverted microscope (Olympus, Tokyo, Japan) at 400× magnification. Once spermatozoa (motile or not motile) were observed in the dish, the suspension was mixed by pipetting and transferred to a tube with SpermRinse (Vitrolife, Sweden), and then the tube was placed upright for 5 min. After removing the large sludge at the bottom using a straw, the supernatant was centrifuged at 250 g for 5 min, and approximately 0.5 ml of sediment was reserved. An equal volume of sperm freezing medium (ORIGIO, Måløv, Denmark) was added dropwise onto the sediment, and the solution was carefully mixed after each addition, followed by incubation at RT for 10 min. Each sample was divided into two tubes, which were frozen using vapor rapid freezing or –80°C freezers. After 24 h, the samples frozen in the -80°C freezer were transferred to liquid nitrogen.

Thawing of Spermatozoa Collected by TESA

The sperm obtained from the testicular puncture and frozen for 1 week were thawed in a 37°C-water bath for 10 min. Then, we used SpermRinse to remove the cryoprotectant in the semen sample and centrifuged it to remove the supernatant. After adding pentoxifylline to the sample for sperm activation, the sample was filled into the sperm counting pool, and the number of sperm and the percentage of motile sperm were counted under a microscope.

Temperature Curve Assessment

The temperature changes during freezing were determined using a portable, multiuse industrial data logger (OM-CP-OCTPRO, Omega Engineering, United States). The temperature sensor uses a thermocouple K matched with the industrial data logger.

Routine Semen Analyses

Routine semen analyses were performed using computer-assisted semen analysis (CASA) to determine progressive motility. We analyzed a minimum of six fields of view per chamber and, at least 200 sperm were evaluated in each chamber according to World Health Organization (WHO) guidelines (WHO, 2010).

Sperm viability was evaluated by eosin-nigrosine staining (Ankebio, China). At least 200 spermatozoa were analyzed with an optical microscope (magnification 1000×). Sperm with red heads were considered nonviable (membrane-damaged), whereas sperm showing no color were considered alive (membrane-intact) (Rarani et al., 2019).

The Sperm Chromatin Structure Assay

Sperm Chromatin Structure Assay (SCSA) was measured by flow cytometry according to the protocol based on Evenson et al. (1980). A commercial kit (Cellpro, China) was used for the evaluation of SCSA. Damaged chromatin in the sperm nucleus after acid treatment forms a single chain and emits red or orange fluorescence upon binding the dye acridine orange; normal sperm chromatin in the nucleus maintains the integrity of the double-stranded structure after acid treatment and emits green fluorescence when combined with acridine orange. At least 5,000 cells were counted per sample. The SCSA parameters include DFI defined as the percentage of denatured sperm DNA and HDS defined as the percentage of spermatozoa with abnormally high DNA stainability.

Reactive Oxygen Species Assessment

Reactive oxygen species (ROS) were determined using a Sperm Reactive Oxygen Species Detection Kit (Ankebio, China), flow cytometry was used for detection, and each sample contained at least 5,000 cells (Mahfouz et al., 2009).

Mitochondrial Membrane Potential Assessment

To measure the mitochondrial membrane potential (MMP) of the sperm, a sperm mitochondrial staining kit (JC-1 fluorescent staining method, Ankebio, China) was used. The MMP in sperm cells can be labeled with fluorescent dyes. The fluorescent probes gather in the mitochondria and emit red fluorescence when there is high MMP. At least 5,000 cells were counted per sample.

Sperm Acrosome Reaction Assessment Statistical Analysis

According to the WHO guidelines, the acrosome reaction (AR) was assessed by Fluorescein isothiocyanate-Pisum sativum

TABLE 1 | Characteristics and sperm parameters in patients from sections 1, 2 and 3.

Characteristics	Section 1 (*n* = 30)	Section 2 (*n* = 33)	Section 3 (*n* = 30)	Total (*n* = 93)
Age (year)	30.80 ± 3.99	32.82 ± 4.87	33.07 ± 6.03	32.23 ± 4.96
Sperm volume (ml)	4.08 ± 1.17	3.48 ± 1.29	4.00 ± 1.47	3.85 ± 1.31
Sperm concentration (million/mL)	70.94 ± 47.23	61.08 ± 37.85	87.62 ± 36.40	73.21 ± 40.49
Progressive motility (%)	36.08 ± 9.40	36.28 ± 12.13	44.93 ± 7.16	39.10 ± 9.56
Viability (%)	73.95 ± 13.59	71.01 ± 14.12	68.02 ± 9.51	70.99 ± 12.41
ROS (%)	—	10.25 ± 9.38	—	—
MMP (%)	—	88.81 ± 14.18	—	—
DFI (%)	—	9.90 ± 4.22	—	—
HDS (%)	—	6.75 ± 2.75	—	—

ROS, reactive oxygen species; MMP, mitochondrial membrane potential; DFI, DNA, fragmentation index; HDS, high DNA, stainable. Data were presented as mean ± SD.

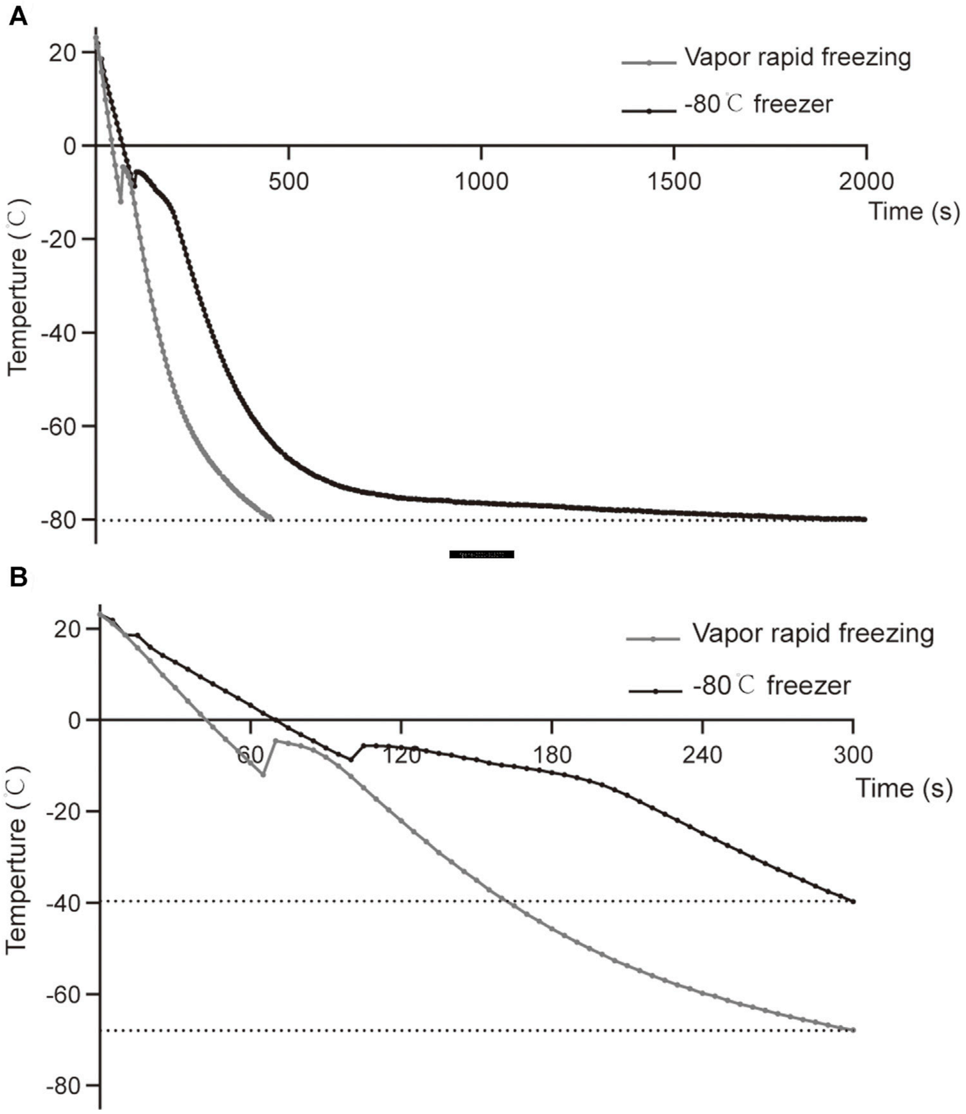


FIGURE 2 | Comparison of temperature curves of vapor rapid freezing and −80°C freezer. **(A)** The temperature drop curve of sperm samples from RT to −80°C in liquid nitrogen vapor rapid freezing and −80°C freezer. **(B)** The temperature drop curve of the sperm samples in the first 300 s of liquid nitrogen vapor rapid freezing and −80°C freezer.

TABLE 2 | Comparison of sperm motility and viability between liquid nitrogen vapor rapid freezing and –80°C freezer.

Groups	Progressive motility (%)	Viability (%)
Short vapor group (*n* = 63)	14.67 ± 4.99	44.33 ± 9.92
Short freezer group (*n* = 63)	15.53 ± 5.40	46.02 ± 10.84
p-value	0.052	0.056
Long vapor group (*n* = 63)	12.97 ± 4.94	41.02 ± 9.75
Long freezer group (*n* = 63)	12.87 ± 4.84	40.77 ± 9.52
p-value	0.793	0.741
Long vapor group (*n* = 33)	11.77 ± 5.19	38.69 ± 10.18
Liquid nitrogen group (*n* = 33)	11.10 ± 4.86	37.43 ± 8.67
p-value	0.091	0.120

Short vapor group: frozen by vapor rapid freezing for 24 h; Short freezer group: frozen by –80°C freezer for 24 h; Long vapor group: frozen by vapor rapid freezing for 2 months; Long freezer group: frozen by –80°C freezer for 2 months; Liquid nitrogen group: frozen by –80°C freezers for 24 h and then immersing it in liquid nitrogen for 2 months; p-value *was derived from paired* t-test. *Data were presented as mean ± SD.*

agglutinin (FITC-PSA, Sigma-Aldrich, St. Louis, America) staining. After rinsing sperm with phosphate buffer saline, sperm were fixed with 4% (w/v) paraformaldehyde or 10 min, mounted on slides, then air-dried and incubated 2 h in the dark at 4°C with 25 mg/L FITC-PSA. Sperm were washed with PBS and examined by fluorescence Nikon Eclipse 80i microscopy (n b 200 sperm/ sample) (Nikon Inc., Tokyo, Japan).

The results were analyzed by using the program Statistical Product and Service Solutions (SPSS) Statistics 23.0 (SPSS Inc., Chicago, IL, United States). Data are expressed as the means ± standard deviation (SD). The comparison between the two groups was performed by paired-samples *t*-test. The percentage of motile sperm was expressed as proportion, and *p*-value was derived from the chi-square test. Regression analysis was used to compare the correlation between sperm parameters. A statistical value of $p < 0.05$ was considered statistically significant.

RESULTS

Characteristics of the Study Population

A total of ninety-three specimens from patients attending the reproductive center were enrolled. The characteristics of the study population are shown in **Table 1**. Among the 93 participants, the mean age was 32.23 ± 4.96 years and the mean semen volume was 3.85 ± 1.31 ml. The mean of sperm parameters, such as sperm concentration, viability, and progressive motility, were all above the reference established by the WHO (2010).

Freezing Temperature Curve of Vapor Rapid Freezing and –80°C Freezing

The sample was placed in liquid nitrogen vapor for 5 min, and the cooling rate was –18°C/min down from RT to –65°C. The drop from RT to –80°C took approximately 410 s (**Figure 2**, **Supplementary Table 1**). In the –80°C freezer, the sample was placed in liquid nitrogen vapor for 5 min, the cooling rate was approximately -12°C/ min down from RT to –40°C, and the drop from RT to –80°C took approximately 1990 s (**Figure 2**, **Supplementary Table 1**). Although freezing at –80°C has a slower temperature drop than vapor rapid freezing during the same period, the curves of the temperature drop are similar between the two methods. Notably, slight differences in the temperature were observed at the freezing point (**Figure 2B**, **Supplementary Table 1**).

Effect of Different Cryopreservation Methods and Different Storage Times on Sperm Quality

As shown in **Table 2**, no significant difference was observed in progressive motility and viability across the different freezing methods ($p > 0.05$).

The ROS of sperm of the freezer group was significantly lower than that of the vapor group (**Table 3**). However, the HDS of sperm in the long freezer group was significantly higher than that of sperm in the long vapor group ($p < 0.05$). There was no significant difference in the other parameters between the different methods ($p > 0.05$).

We compared the state of the sperm after freezing and found that progressive motility and viability were significantly decreased (**Figures 3A,B**). Specifically, compared with short-term freezing, long-term freezing significantly decreased sperm motility. There was no significant difference in ROS before and after freezing (**Figure 3C**), while long-term freezing decreased MMP

TABLE 3 | Comparison of sperm function between liquid nitrogen vapor rapid freezing and –80°C freezer.

Groups	ROS (%)	MMP (%)	DFI (%)	HDS (%)
Short vapor group (*n* = 33)	9.53 ± 8.47	84.95 ± 17.73	9.41 ± 4.62	7.87 ± 2.95
Short freezer group (*n* = 33)	7.45 ± 7.14	87.04 ± 14.68	9.45 ± 4.47	8.20 ± 2.87
p-value[a]	0.036	0.611	0.805	0.133
Long vapor group (*n* = 33)	12.94 ± 9.46	75.91 ± 19.39	9.53 ± 4.74	8.46 ± 2.90
Long freezer group (*n* = 33)	11.64 ± 8.13	75.30 ± 21.19	9.72 ± 5.30	9.10 ± 2.88
Liquid nitrogen group (*n* = 33)	12.48 ± 9.35	77.75 ± 20.18	9.47 ± 4.44	8.65 ± 2.63
p-value[b]	0.022	0.713	0.634	0.017
p-value[c]	0.655	0.339	0.791	0.343

Short vapor group: frozen by vapor rapid freezing for 24 h; Short freezer group: frozen by –80°C freezer for 24 h; Long vapor group: frozen by vapor rapid freezing for 2 months; Long freezer group: frozen by –80°C freezer for 2 months; Liquid nitrogen group: frozen by –80°C freezers for 24 h and then immersing it in liquid nitrogen for 2 months.

[a] *p-value of short freezer group versus short vapor group (paired* t-test).

[b] *p-value of long freezer group versus long vapor group (paired* t-test).

[c] *p-value of liquid nitrogen group versus long vapor group (paired* t-test). *Data were presented as mean ± SD.*

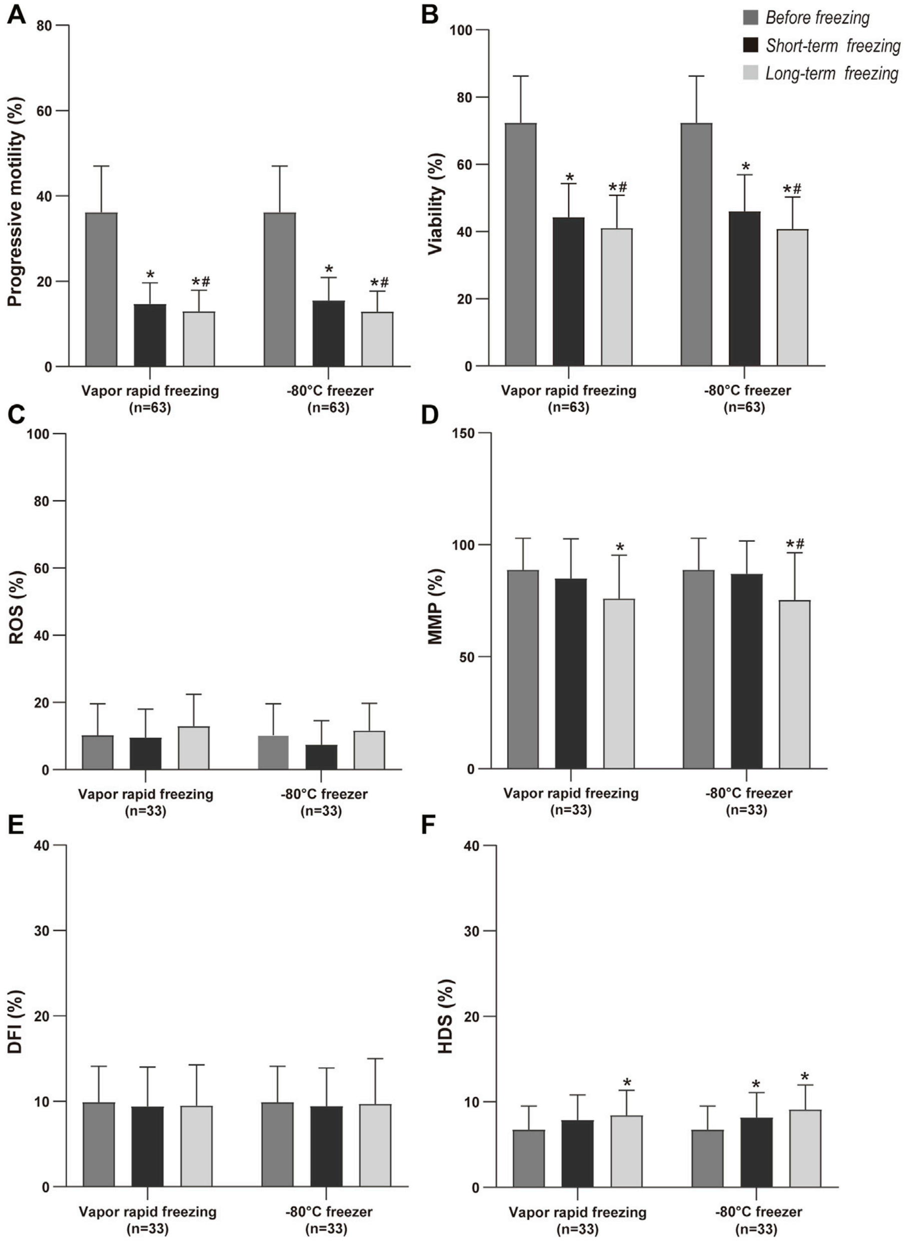


FIGURE 3 | Comparison of sperm quality by cryopreservation time before freezing, short term and long-term freezing. Values are mean and bar are SD; *p-value < 0.05 compare with before freezing (paired t-test); $^{\#}$p-value < 0.05 compare with short-term freezing (paired t-test).

(**Figure 3D**). Moreover, long-term storage of sperm had a similar DFI with short-term storage (**Figure 3E**). In addition, freezing also resulted in a higher HDS in the sperm (**Figure 3F**).

Effect of Different Cryopreservation Methods on Sperm Acrosome Reaction

A total of 30 men were included and significant differences (9.23 ± 4.61 vs. 14.62 ± 6.86 vs. 14.38 ± 6.05, $p < 0.05$) were observed for acrosome reaction (AR) within before and freeze-thawed spermatozoa (**Figure 4**). To be noted, there was no significant difference in the AR between vapor rapid freezing and −80°C freezer (14.62 ± 6.86 vs. 14.38 ± 6.05, $p > 0.05$).

Linear Regression Between Sperm Quality After Freeze-Thawing and Sperm Parameters

In regression analysis, we found that progressive motility in post-thaw spermatozoa was positively correlated with both progressive motility before freezing and the direct −80°C freezing methods but negatively correlated with ROS, DFI

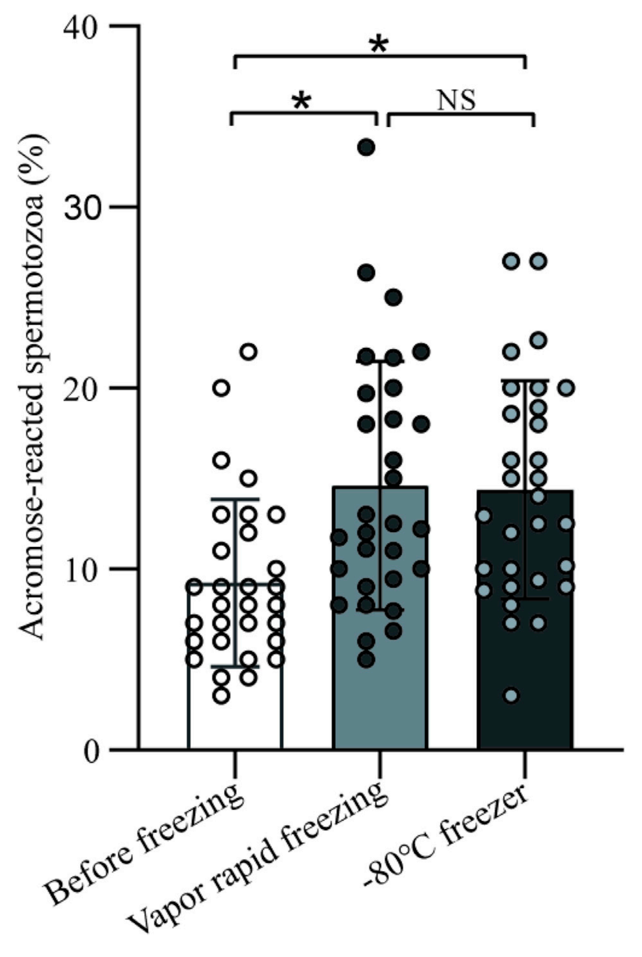


FIGURE 4 | The impact on sperm acrosome of vapor rapid freezing and −80°C freezer. Sperm treated by freezing had significant higher acrosome reaction than before freezing (*p*-value < 0.05). There was no significant difference in the acrosome reaction between vapor rapid freezing and −80°C freezer (*p*-value > 0.05).

before freezing, and storage time ($p < 0.05$, **Table 4**). ROS in post-thaw spermatozoa was positively correlated with ROS before freezing and storage time ($p < 0.05$). DFI in post-thaw spermatozoa was positively correlated with DFI before freezing but negatively correlated with ROS before freezing ($p < 0.05$).

Motile Sperm Percentage of Testicular Sperm Samples After Cryopreservation With Vapor Rapid Freezing and –80°C Freezer

To analyze the outcomes of freeze-thawed spermatozoa collected by TESA, we compared the parameters of sperm conserved by vapor rapid freezing and stored at −80°C (**Table 5**). We counted a total of 62 motile sperm, accounting for 16.89% of the total sperm, when vapor rapid freezing was used. A total of 65 motile sperm, accounting for 17.02% of the total sperm, were counted when freezing at −80°C. There was no statistically significant difference between the two methods for freezing testicular sperm ($p = 0.965$).

DISCUSSION

This study compared the effects of a −80°C freezer and vapor rapid freezing on sperm quality, and we found that there was no significant difference in progressive motility and viability across the different freezing methods. Correlation analysis showed that progressive motility, ROS, and DFI in the post-thaw spermatozoa were correlated with sperm characteristics before freezing, methods and storage time.

Some studies have reported that freezing-thawing decreases sperm quality and function (Satirapod et al., 2012; Lusignan et al., 2018; Le et al., 2019). For example, proteomic analysis of sperm showed significant changes in proteins related to motility, viability, and acrosomal integrity of sperm compared with the fresh state (Wang et al., 2014). It has also been suggested that

TABLE 4 | Results from the regression analyses after freezing sperm quality.

Parameters	Progressive motility after freezing		ROS after freezing		DFI after freezing	
	β (95%CI)	***p*-value**	**β (95%CI)**	***p*-value**	**β (95%CI)**	***p*-value**
Progressive motility before freezing	0.30 (0.25–0.35)	<0.001	0.06 (-0.06–0.18)	0.331	−0.14 (−0.20–0.08)	<0.001
ROS before freezing	−0.13 (−0.23–0.01)	0.010	0.19 (0.03–0.34)	0.021	−0.05 (−0.13–0.04)	0.314
DFI before freezing	−0.44 (−0.65–0.23)	<0.001	−0.06 (−0.47–0.30)	0.577	0.93 (0.82–1.04)	<0.001
Time						
Short-term	Ref	Ref	Ref	Ref	Ref	Ref
Long-term	−2.49 (−3.84–1.14)	<0.001	3.80 (1.02–6.59)	0.008	0.19 (−0.74–1.12)	0.688
Method						
Vapor rapid freezing	Ref	Ref	Ref	Ref	Ref	Ref
−80°C freezer	1.64 (0.29–2.99)	0.018	−1.69 (−4.48–1.09)	0.231	0.12 (−0.81–1.05)	0.804

β, regression analyses coefficients; 95% CI, 95% confidence interval.

TABLE 5 | Motile sperm percentage of testicular sperm samples after cryopreservation with vapor rapid freezing and -80°C freezer.

Sample number	Percentage of motile sperm (%)		*p*-value
	Vapor rapid freezing	-80°C freezer	
1	7/53 (13.21)	6/48 (12.50)	-
2	7/38 (18.42)	4/24 (16.67)	-
3	4/29 (13.79)	6/31 (19.35)	-
4	3/33 (9.09)	4/38 (10.52)	-
5	13/57 (22.81)	10/48 (20.83)	-
6	4/24 (16.67)	6/37 (16.21)	-
7	3/19 (15.79)	3/18 (16.67)	-
8	15/63 (23.81)	17/72 (23.61)	-
9	2/20 (10.00)	3/27 (11.11)	-
10	4/31 (12.90)	6/39 (15.38)	-
Total	62/367 (16.89)	65/382 (17.02)	0.965

p-value was derived from chi-square test.

rapid changes in intracellular ice crystal formation and osmolarity during the cryopreservation process may lead to changes in the carbohydrate composition and the membrane proteins, which can disrupt membrane structures and reduce sperm viability (Pedersen and Lebech, 1971). The production of ROS and lower antioxidant enzyme activity in sperm induce apoptotic pathways, which can lead to a reduction in sperm viability (Di Santo et al., 2012). In our study, we found the same results as previous research that showed that after freezing-thawing, sperm viability and progressive motility were both decreased significantly. ROS increased and MMP continued to decrease as the freezing time extended. However, compared with vapor rapid freezing, there was no significant difference in progressive motility, viability, MMP or DFI of sperm frozen in a −80°C freezer. Previously, Rahana et al. (2011) also reported similar results that there was no significant difference in human sperm motility and DFI between -85°C and conventional liquid nitrogen cryopreservation. Previous studies observed a decrease in AR after the freezing-thawing process (Gomez-Torres et al., 2017; Rahiminia et al., 2017). Here we also observed that the acrosome status was vulnerable to the freezing-thawing process, but the percentages of AR decrease for different cryopreservation methods was similar, which further confirmed the efficiency of −80°C freezer.

In regression analysis, compared with vapor rapid freezing, −80°C freezing was positively correlated with sperm progressive motility. Our study also indicated that freezing in a −80°C freezer resulted in lower ROS than vapor rapid freezing, which confirmed that a slow temperature drop would result in lower ROS damage. During freezing, the cooling rate of vapor rapid freezing was quicker than that of the −80°C freezer, and thus vapor rapid freezing took less time to reach the freezing point and had a lower freezing point temperature than the −80°C freezer (**Supplementary Table 1**). In the freezing process, water tends to chill beyond its freezing point without forming ice, which is known as supercooling. Studies have confirmed that super-cooling can cause damages of sperm in mice and human (Check et al., 1995; Mazur and Koshimoto, 2002). Compared with vapor rapid freezing, the samples frozen in a −80°C freezer have a higher freezing point that is closer to −6°C (**Supplementary Table 1**). The WHO (2010) first recommended that a human sperm freezing program decline from RT (22–25°C) to −6°C, and we speculated that −6°C was the freezing point of semen with cryoprotectant. Therefore, samples in a −80°C freezer were less supercooled than vapor rapid freezing, which may result in less damage. Notably, the HDS of sperm frozen in a −80°C freezer for a long time was higher than that in liquid nitrogen. Compared with vapor rapid freezing, the −80°C freezer has a more stable and slower cooling rate and can meet the demand for sperm freezing so that the freezing step is simplified. Thus, although liquid nitrogen storage is irreplaceable and is still the first choice for long-term sperm cryopreservation at a low temperature of −196°C, a −80°C freezer can be an alternative method for short-term sperm storage.

In this study, we used 1.8 ml cryotubes and a −80°C freezer to simplify the steps of freezing testicular sperm. We have proven that −80°C freezers and vapor rapid freezing have similar effects on testicular sperm. Previously, the method of freezing single sperm captured under a microscope using intracytoplasmic sperm injection (ICSI) pipettes equipped with a micromanipulator is widely used for testicular sperm (Coetzee et al., 2016). However, more sperm can be obtained after freezing all the testicular sperm in a freezing tube and a −80°C freezer compared with the freezing of single sperm. Using cryotubes will provide enough sperm after freezing-thawing, and it is better to select motile sperm for ICSI. Furthermore, the efficiency of single sperm freezing is low and cannot meet the requirements of storing numerous samples. The freezing of testicular sperm in a −80°C freezer in a freezer tube ensures that all sperm in the testicular tissue are captured, it only takes a short time and simplifies the process, and several samples can be processed simultaneously.

The present study was based on a limited sample size and normal semen meeting the WHO standard, and further studies of −80°C freezers in clinical applications with sperm samples from oligo-astheno-teratozoospermia are warranted. This was a preliminary study to improve the freezing of testicular sperm, and it was necessary to freeze tremendous testicular sperm samples to avoid instability. *In vitro* fertilization tests, follow-up embryo development, and implantation are also the focus of future research.

In conclusion, this study demonstrated that short-term storage of sperm at −80°C freezer could be a viable alternative to liquid nitrogen vapor rapid freezing at −196°C due to their comparable post-thaw results and lower ROS. During long-term freezing, the −80°C freezer is expected to be a cooling process that can provide an option except for liquid nitrogen vapor rapid freezing. An improved sperm freezing process was also preliminarily explored in this study. Additional study is necessary to confirm the clinical value of the freezing testicular sperm method.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee of The First Affiliated Hospital of the University of Science and Technology of China. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XW and FL contributed to sample preparation, sample analysis, data analysis, and wrote the manuscript. SB and XJ contributed to manuscript preparation, supervised the analysis, and critically revised the manuscript. LW, LH, NZ and RJ recruited study subjects, participated in sample collection, and provided clinical information. BX and YW recruited study subjects and controls. SB and XJ performed the statistical analysis. XT and XJ conceptualized and designed the study, supervised the analysis, critically revised the manuscript, edited the paper, and gave final approval. All authors read and approved the final manuscript.

FUNDING

This work was supported by National Key R&D Program of China (2018YFC1003900), National Natural Science Foundation of China (No.81971446), Open Project Fund from Key Laboratory of Reproduction Regulation of NHC (KF 2020-07) and Natural Science Foundation of Qinghai (2019-HZ-823).

ACKNOWLEDGMENTS

The authors thank the teacher and staff of Reproductive Center of the First Affiliated Hospital of University of Science and Technology of China. We also thank the technicians and research staff of the Institute of graduate, Anhui Medical University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2021.799504/full#supplementary-material

REFERENCES

Berkovitz, A., Miller, N., Silberman, M., Belenky, M., and Itsykson, P. (2018). A Novel Solution for Freezing Small Numbers of Spermatozoa Using a Sperm Vitrification Device. *Hum. Reprod.* 33 (11), 1975–1983. doi:10.1093/humrep/dey304

Check, D. J., Katsoff, D., and Check, J. H. (1995). Effect of an Intermediate Hold with Vapor Freezing on Subsequent Hypoosmotic Swelling in Thawed Sperm. *Arch. Androl.* 35 (1), 79–81. doi:10.3109/01485019508987857

Coetzee, K., Ozgur, K., Berkkanoglu, M., Bulut, H., and Isikli, A. (2016). Reliable Single Sperm Cryopreservation in Cell Sleepers for Azoospermia Management. *Andrologia* 48 (2), 203–210. doi:10.1111/and.12434

Cohen, J., Garrisi, G. J., Congedo-Ferrara, T. A., Kieck, K. A., Schimmel, T. W., and Scott, R. T. (1997). Cryopreservation of Single Human Spermatozoa. *Hum. Reprod.* 12 (5), 994–1001. doi:10.1093/humrep/12.5.994

Desai, N., Culler, C., and Goldfarb, J. (2004). Cryopreservation of Single Sperm from Epididymal and Testicular Samples on Cryoloops: Preliminary Case Report. *Fertil. Sterility* 82, S264–S265. doi:10.1016/j.fertnstert.2004.07.706

Di Santo, M., Tarozzi, N., Nadalini, M., and Borini, A. (2012). Human Sperm Cryopreservation: Update on Techniques, Effect on DNA Integrity, and Implications for ART. *Adv. Urol.* 2012, 1–12. doi:10.1155/2012/854837

Endo, Y., Fujii, Y., Kurotsuchi, S., Motoyama, H., and Funahashi, H. (2012). Successful Delivery Derived from Vitrified-Warmed Spermatozoa from a Patient with Nonobstructive Azoospermia. *Fertil. Sterility* 98 (6), 1423–1427. doi:10.1016/j.fertnstert.2012.07.1128

Evenson, D. P., Darzynkiewicz, Z., and Melamed, M. R. (1980). Relation of Mammalian Sperm Chromatin Heterogeneity to Fertility. *Science* 210 (4474), 1131–1133. doi:10.1126/science.7444440

Gomez-Torres, M. J., Medrano, L., Romero, A., Fernández-Colom, P. J., and Aizpurúa, J. (2017). Effectiveness of Human Spermatozoa Biomarkers as Indicators of Structural Damage during Cryopreservation. *Cryobiology* 78, 90–94. doi:10.1016/j.cryobiol.2017.06.008

Gwatkin, R. B. L. (1991). "Handbook of the Laboratory Diagnosis and Treatment of Infertility," in *Molecular Reproduction and Development*. Editors B. Keel and B. Webster. (Boca Raton, FL: CRC Press), 29, 431. doi:10.1002/mrd.1080290117

Hsieh, Y.-Y., Tsai, H.-D., Chang, C.-C., and Lo, H.-Y. (2000). Cryopreservation of Human Spermatozoa within Human or Mouse Empty Zona Pellucidae. *Fertil. Sterility* 73 (4), 694–698. doi:10.1016/s0015-0282(99)00612-3

Huang, W.-J., Zhang, D., Hong, Z.-W., Chen, Z.-B., Dong, L.-H., Zhang, Y., et al. (2020). Sequential Interval Micro-droplet Loading in Closed Hemi-Straw Carrier System: A Convenient and Efficient Method for Ultra-rapid Cryopreservation in Extreme Oligozoospermia. *Cryobiology* 93, 75–83. doi:10.1016/j.cryobiol.2020.02.005

Just, A., Gruber, I., Wöber, M., Lahodny, J., Obruca, A., and Strohmer, H. (2004). Novel Method for the Cryopreservation of Testicular Sperm and Ejaculated Spermatozoa from Patients with Severe Oligospermia: A Pilot Study. *Fertil. Sterility* 82 (2), 445–447. doi:10.1016/j.fertnstert.2003.12.050

Le, M. T., Nguyen, T. T. T., Nguyen, T. T., Nguyen, V. T., Nguyen, T. T. A., Nguyen, V. Q. H., et al. (2019). Cryopreservation of Human Spermatozoa by Vitrification versus Conventional Rapid Freezing: Effects on Motility, Viability, Morphology and Cellular Defects. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 234, 14–20. doi:10.1016/j.ejogrb.2019.01.001

Liu, T., Gao, J., Zhou, N., Mo, M., Wang, X., Zhang, X., et al. (2016). The Effect of Two Cryopreservation Methods on Human Sperm DNA Damage. *Cryobiology* 72 (3), 210–215. doi:10.1016/j.cryobiol.2016.04.004

Lusignan, M. F., Li, X., Herrero, B., Delbes, G., and Chan, P. T. K. (2018). Effects of Different Cryopreservation Methods on DNA Integrity and Sperm Chromatin Quality in Men. *Andrology* 6 (6), 829–835. doi:10.1111/andr.12529

Mahfouz, R., Sharma, R., Lackner, J., Aziz, N., and Agarwal, A. (2009). Evaluation of Chemiluminescence and Flow Cytometry as Tools in Assessing Production of Hydrogen Peroxide and Superoxide Anion in Human Spermatozoa. *Fertil. Sterility* 92 (2), 819–827. doi:10.1016/j.fertnstert.2008.05.087

Mazur, P., and Koshimoto, C. (2002). Is Intracellular Ice Formation the Cause of Death of Mouse Sperm Frozen at High Cooling Rates? *Biol. Reprod.* 66, 1485–1490. doi:10.1095/biolreprod66.5.1485

Miller, N., Biron-Shental, T., Pasternak, Y., Belenky, M., Shefi, S., Itsykson, P., et al. (2017). Fertility Outcomes after Extended Searches for Ejaculated Spermatozoa in Men with Virtual Azoospermia. *Fertil. Sterility* 107 (6), 1305–1311. doi:10.1016/j.fertnstert.2017.04.005

Parkes, A. S. (1945). Preservation of Spermatozoa at Low Temperatures. *Bmj* 2 (4415), 212–213. doi:10.1136/bmj.2.4415.212

Pedersen, H., and Lebech, P. E. (1971). Ultrastructural Changes in the Human Spermatozoon after Freezing for Artificial Insemination. *Fertil. Sterility* 22 (2), 125–133. doi:10.1016/S0015-0282(16)38048-7

Pezo, F., Cheuquemán, C., Salinas, P., and Risopatrón, J. (2017). Freezing Dog Semen Using −80 °C Ultra-freezer: Sperm Function and *In Vivo* Fertility. *Theriogenology* 99, 36–40. doi:10.1016/j.theriogenology.2017.05.007

Rahana, A. R., Ng, S. P., Leong, C. F., and Rahimah, M. D. (2011). Comparison between Mechanical Freezer and Conventional Freezing Using Liquid Nitrogen in Normozoospermia. *Singapore Med. J.* 52 (10), 734–737. PubMed (nih.gov).

Rahiminia, T., Hosseini, A., Anvari, M., Ghasemi-Esmailabad, S., and Talebi, A. R. (2017). Modern Human Sperm Freezing: Effect on DNA, Chromatin and Acrosome Integrity. *Taiwanese J. Obstet. Gynecol.* 56 (4), 472–476. doi:10.1016/j.tjog.2017.02.004

Rarani, F. Z., Golshan-Iranpour, F., and Dashti, G. R. (2019). Correlation between Sperm Motility and Sperm Chromatin/DNA Damage before and after Cryopreservation and the Effect of Folic Acid and Nicotinic Acid on post-thaw Sperm Quality in Normozoospermic Men. *Cell Tissue Bank* 20 (3), 367–378. doi:10.1007/s10561-019-09775-6

Raspa, M., Fray, M., Paoletti, R., Montoliu, L., Giuliani, A., and Scavizzi, F. (2018b). A New, Simple and Efficient Liquid Nitrogen Free Method to Cryopreserve Mouse Spermatozoa at −80 °C. *Theriogenology* 119, 52–59. doi:10.1016/j.theriogenology.2018.06.020

Raspa, M., Fray, M., Paoletti, R., Montoliu, L., Giuliani, A., and Scavizzi, F. (2018a). Long Term Maintenance of Frozen Mouse Spermatozoa at −80 °C. *Theriogenology* 107, 41–49. doi:10.1016/j.theriogenology.2017.10.036

Raspa, M., Guan, M., Paoletti, R., Montoliu, L., Ayadi, A., Marschall, S., et al. (2017). Dry Ice Is a Reliable Substrate for the Distribution of Frozen Mouse Spermatozoa: A Multi-Centric Study. *Theriogenology* 96, 49–57. doi:10.1016/j.theriogenology.2017.04.003

Royere, D., Barthelemy, C., Hamamah, S., and Lansac, J. (1996). Cryopreservation of Spermatozoa: a 1996 Review. *Hum. Reprod. Update* 2 (6), 553–559. doi:10.1093/humupd/2.6.553

Sanchez, R., Risopatrón, J., Schulz, M., Villegas, J. V., Isachenko, V., and Isachenko, E. (2012). Vitrified Sperm banks: the New Aseptic Technique for Human Spermatozoa Allows Cryopreservation at −86 °C. *Andrologia* 44 (6), 433–435. doi:10.1111/j.1439-0272.2012.01314.x

Satirapod, C., Treetampinich, C., Weerakiet, S., Wongkularb, A., Rattanasiri, S., and Choktanasiri, W. (2012). Comparison of Cryopreserved Human Sperm from Solid Surface Vitrification and Standard Vapor Freezing Method: on Motility, Morphology, Vitality and DNA Integrity. *Andrologia* 44 (s1), 786–790. doi:10.1111/j.1439-0272.2011.01267.x

Sereni, E., Bonu, M., Fava, L., Sciajno, R., Serrao, L., Preti, S., et al. (2008). Freezing Spermatozoa Obtained by Testicular fine Needle Aspiration: a New Technique. *Reprod. BioMedicine Online* 16 (1), 89–95. doi:10.1016/S1472-6483(10)60560-3

Sherman, J. K. (1973). Synopsis of the Use of Frozen Human Semen since 1964: State of the Art of Human Semen Banking. *Fertil. Sterility* 24 (5), 397–412. doi:10.1016/S0015-0282(16)39678-9

Tao, Y., Sanger, E., Saewu, A., and Leveille, M.-C. (2020). Human Sperm Vitrification: the State of the Art. *Reprod. Biol. Endocrinol.* 18 (1), 17. doi:10.1186/s12958-020-00580-5

Trottmann, M., Becker, A. J., Stadler, T., Straub, J., Soljanik, I., Schlenker, B., et al. (2007). Semen Quality in Men with Malignant Diseases before and after Therapy and the Role of Cryopreservation. *Eur. Urol.* 52 (2), 355–367. doi:10.1016/j.eururo.2007.03.085

Vaz, C. R., Lamim, T., Salvador, R. A., Batschauer, A. P. B., Amaral, V. L. L., and Til, D. (2018). Could Cryopreserved Human Semen Samples Be Stored at -80°C? *JBRA Assist. Reprod.* 22 (2), 108–112. doi:10.5935/1518-0557.20180016

Wang, S., Wang, W., Xu, Y., Tang, M., Fang, J., Sun, H., et al. (2014). Proteomic Characteristics of Human Sperm Cryopreservation. *Proteomics* 14 (2-3), 298–310. doi:10.1002/pmic.201300225

WHO (2010). *WHO Laboratory Manual for the Examination and Processing of Human Semen*. 5th ed edn. Geneva: World Health Organization.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Genetics

ORIGINAL RESEARCH
published: 11 March 2022
doi: 10.3389/fgene.2022.851719

Single Cell Transcriptome Sequencing of Zebrafish Testis Revealed Novel Spermatogenesis Marker Genes and Stronger Leydig-Germ Cell Paracrine Interactions

Peipei Qian [1†], ***Jiahui Kang*** [1†], ***Dong Liu*** [2*] **and *Gangcai Xie*** [1*]

[1]Institute of Reproductive Medicine, Medical School, Nantong University, Nantong, China, [2]School of Life Sciences, Key Laboratory of Neuroregeneration of Jiangsu and Ministry of Education, Co-innovation Center of Neuroregeneration, Nantong University, Nantong, China

Spermatogenesis in testis is an important process for sexual reproduction, and worldwide about 10–15 percent of couples suffer from infertility. It is of importance to study spermatogenesis at single cell level in both of human and model organisms. Currently, single-cell RNA sequencing technologies (scRNA-seq) had been extensively applied to the study of cellular components and its gene regulations in the testes of different species, including human, monkey, mouse, and fly, but not in zebrafish. Zebrafish was a widely used model organism in biology and had been extensively used for the study of spermatogenesis in the previous studies. Therefore, it is also important to profile the transcriptome of zebrafish testis at single cell level. In this study, the transcriptomes of 14, 315 single cells from adult male zebrafish testes were profiled by scRNA-seq, and 10 cell populations were revealed, including Leydig cell, Sertoli cell, spermatogonia cell (SPG), spermatocyte, and spermatids. Notably, thousands of cell-type specific novel marker genes were identified, including *sumo3b* for SPG, *krt18a.1* for Sertoli cells, *larp1b* and *edrf1* for spermatids, which were also validated by RNA *in situ* hybridization experiments. Interestingly, through Ligand-Receptor (LR) analyses, zebrafish Leydig cells demonstrated stronger paracrine influence on germ cells than Sertoli cells. Overall, this study could be an important resource for the study of spermatogenesis in zebrafish and might also facilitate the study of the genes associated with human infertility through using zebrafish as a model organism.

Keywords: single cell, testis, spermatogenesis, zebrafish, transcriptome, Leydig cell, paracrine

Edited by:
Xi Wang,
Nanjing Medical University, China

Reviewed by:
Xianbin Su,
Shanghai Jiao Tong University, China
Deshou Wang,
Southwest University, China

****Correspondence:***
Dong Liu
tom@ntu.edu.cn
Gangcai Xie
gangcai@ntu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *10 January 2022*
Accepted: *27 January 2022*
Published: *11 March 2022*

Citation:
Qian P, Kang J, Liu D and Xie G (2022) Single Cell Transcriptome Sequencing of Zebrafish Testis Revealed Novel Spermatogenesis Marker Genes and Stronger Leydig-Germ Cell Paracrine Interactions.
Front. Genet. 13:851719.
doi: 10.3389/fgene.2022.851719

INTRODUCTION

In animals, testis is the primary male reproductive organ that generates sperms for sexual reproduction. It had been estimated that 10–15% of couples worldwide suffer from infertility, and male factors attribute to about 20–30% of all factors leading to infertility (Babakhanzadeh et al., 2020). Spermatogenesis in testis is an important process to generate sperms through undergoing meiosis I and II, which involves a sequential of cell types, such as spermatogonia stem cells (SSC), spermatocytes derived from SSCs through mitotic cell division, haploid spermatids through meiotic cell division, and the supporting somatic cells (eg., Leydig cell and Sertoli cell) (Nishimura and L'Hernault, 2017).

The advent of single-cell RNA sequencing (scRNA-seq) has greatly enlarged our understanding of the transcriptome landscape for various organs in different species. During past years, scRNA-seq had been used to characterize cell heterogeneity and identify novel cell types in various studies, including studies for development, cancer, and kidney diseases (Potter, 2018). Notably, scRNA-seq had been widely used to study spermatogenesis in mammals, such as the studies of murine spermatogenesis (Chen et al., 2018; Green et al., 2018; Lukassen et al., 2018; Grive et al., 2019; Jung et al., 2019), human testis single-cell level studies for different age stages (Guo et al., 2018; Sohni et al., 2019; Guo et al., 2020), and single-cell level evolutionary studies in macaques (Lau et al., 2020; Shami et al., 2020).

In contrast to mammals, due to lacking sex chromosomes, the sex determination and timing of testis differentiation is different in zebrafish (Orban et al., 2009). Besides, zebrafish testis is composed by cystic structure for spermatogenesis (Schulz et al., 2010), while in mammals non-cystic structure seminiferous tubule is the basic units for spermatogenesis. Although there are several studies (Chen et al., 2018; Green et al., 2018; Guo et al., 2018; Lukassen et al., 2018; Grive et al., 2019; Jung et al., 2019; Sohni et al., 2019; Guo et al., 2020; Lau et al., 2020; Shami et al., 2020) for mammalian testis at single-cell level, it is still lacking for the study of zebrafish testis by scRNA-seq.

In this study, we applied 10X genomics scRNA-seq technology to the study of pooled adult zebrafish testes, and profiled the transcriptome of 14, 315 single testicular cells. In total, we identified 10 distinguishable cell types in zebrafish testis, including four types of Spermatids, one type of Spermatocyte, two types of Spermatogonia cells (SPG), two known somatic supporting cells (Leydig and Sertoli cells). Besides identification of basic cell types in zebrafish testis, novel marker genes were also revealed and experimentally validated by this study. Furthermore, through Ligand-Receptor (LR) analyses, more LR interactions were found between somatic cell and germ cell in Leydig than in Sertoli, which indicates stronger paracrine influence from Leydig cell than from Sertoli cell in zebrafish testis. Overall, our study might be a valuable resource for spermatogenesis studies in zebrafish testis.

MATERIALS AND METHODS

Sample Preparation and Single Cell RNA Sequencing

5 month-old adult AB line zebrafish (*Danio rerio*) were prepared and treated with tricaine methanesulfonate (MS-222, 0.25%) on ice for 15 min before experiment. The study was conducted in line with the Chinese law for the Protection of Animals, and the animals were treated properly. The testes samples collected from those zebrafish were first washed by PBS three times, and then were digested in 10 ml 0.25% trypsin at 37°C for 15 min, during which process the tissues were pipetted up and down every 3 min. The cell suspension was first filtered through 70 μm nylon mesh after stopping the digestion process by DMEM (with 10% FBS), and then were centrifuged at the speed of 1,000 rpm for 10 min. After centrifugation, the cells were suspended in 1 ml DMEM (with 10% FBS) and were further filtered through 40 μm nylon mesh. Before loading onto the 10x chromium chip, the cells were washed by BSA DPBS (0.04%, three times) and were resuspended to a concentration of 800~1,000 cells/μl (viability >85%). The single cell library preparation was carried out according to the manufacturer's protocol (Chromium Single Cell 3′ Reagent Kits V3 Chemistry). The cDNA library was further size filtered and sequenced on the Illumina NovaSeq 6,000 System following 150 bp paired-end sequencing protocol.

ScRNA-Seq Analysis

Cell ranger software (v5.0.0) from 10X genomics was used to preprocess the raw sequencing data, and primary UMI filtering was carried out by Cell ranger following criteria "No N contained in the UMI," "all base quality should be no less than 10," and "not a homopolymer sequence." For cell ranger processing, zebrafish genome and gene annotation from Ensembl were used as reference datasets (Ensembl Gene 102). Then, R package Seurat (V4.0.0) was used for further data filtering, data normalization, cell clustering, and cluster-level marker gene discovery. The UMI matrix data was filtered at two dimensions: gene filtering and cell filtering. For gene filtering, only the genes that were detected at least in 5 cells were retained. For cell filtering, the retained cells should contain the number of expressed genes within the range between 200 and 5,000, and the proportion of UMIs from mitochondria should be less than 5%. The raw UMI matrix was normalized by sctransform, which fitted the data through regularized negative binomial regression. Three types of features were used for data regression: total number of expressed genes, total number of UMIs, and the percentage of UMIs from mitochondria. In order to cluster cells, the principal components (PC) were computed by principal components analysis (PCA), and the first fifteen PCs were used for cluster finding. K nearest neighbors (KNN) method was applied to cell clustering, where the K was set to be 20. Finally, the marker genes for each cell cluster were statistically computed by Wilcoxon Rank Sum test, and only the genes detected in at least 25% of the cells (in either of the two tested cell clusters) were included for testing. The marker genes were defined as the genes with p value less than 0.01 and the log transformed (base 2) fold change larger than 0.25.

To annotate the cell clusters identified in this study, the testicular cell-type marker genes were collected from previous studies in zebrafish (Leal et al., 2009; Chen et al., 2013; Assis et al., 2016; Safian et al., 2016; Lin et al., 2017; de Castro Assis et al., 2018; Takemoto et al., 2020) or other species (Hermann et al., 2018). In the further analyses, only the collected marker genes showing cell-type specific expression patterns in our single cell studies were used for annotation. The final list of known marker genes and their sources can be found in **Supplementary Table S1**.

Ligand-Receptor Based Cell-Cell Communication Analyses

In order to study the cell-cell communication among cell populations, the Ligand-Receptor (LR) pairs were analyzed, which followed the procedure with minor modifications described in one recently published paper (Shi et al., 2021) about LR studies in mouse. In detail, CellTalkDB (Shao et al., 2020) was selected as the database for LR analyses, which contained 3,398 manually curated human LR pairs. The orthologous genes between human and zebrafish were downloaded from Ensembl biomart (Ensembl Genes 102), and then 13,004 one-one orthologous genes were retained for LR analyses. SingleCellSignalR (Cabello-Aguilar et al., 2020) was applied to the inference of intercellular LR interaction networks, where the default LR database was replaced by CellTalkDB and only the genes found in the zebrafish-human one-one orthologous gene list were included for the analysis.

H&E Staining

Adult zebrafish (5 months after fertilization) were anesthetized in 0.25% ms-222 (ethyl 3-aminobenzoic; Sigma, E10505) in filtered system water. Zebrafish testes were fixed with 4% paraformaldehyde (Sigma, P6148) overnight at 4°C. Then, the fixed samples were embedded in OCT (Sakura, 4583) after dehydration of sugar, sectioned at 10 μm thickness, and stained with hematoxylin and eosin (BBI, E607318).

RNA *in situ* Hybridization

Probe fragments were first amplified from zebrafish testes cDNA libraries (primers information in **Supplementary Table S2**) and then inserted into pGEM-T-easy vector (Promega, A1360). Digoxigenin-labeled antisense RNA probes were made by using DIG-RNA labeling Kit (Roche, 11175025910). Zebrafish testes were cut into 12 μm slices by cryotome under temperature between −25°C and −23°C. Slides were first fixed with 4% PFA, then digested with 0.1% Proteinase K, followed by washing in PBS. Lastly, the digested slides were first incubated with the digoxygenin-labelled RNA probes at 68°C overnight and then incubated with alkaline phosphatase-conjugated anti-DIG antibody, followed by treatment with AP-substrate NBT/BCIP solution (Roche, 11681451001) for visualization. Brightfield images of sections were obtained with Nikon Eclipse Ni-U microscope.

Data Visualization and GO Enrichment Analyses

To visualize the cell clusters and the gene expression patterns in two-dimensional graphs, Uniform Manifold Approximation and Projection (UMAP) was applied to dimensional reduction, where the first 15 PCs from PCA were used as the input for UMAP. ClusterProfiler (Yu et al., 2012) (version 3.18.0) was used for GO enrichment analyses, and the p values were corrected by Benjamini-Hochberg (BH) method. Only the GO terms with BH corrected p value less than 0.05 were retained as significant ones.

Data and Software Availability

The raw sequence data reported in this paper had been deposited in the Genome Sequence Archive (Wang et al., 2017) in National Genomics Data Center (Members, 2021), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number CRA003925 that are publicly accessible at https://bigd.big.ac.cn/gsa. All the codes used for this study can be found in https://github.com/gangcai/zebTestis.

RESULTS

Single Cell Transcriptome Profiling of Zebrafish Testis

Testes from five adult male zebrafish were pooled for 10X genomics single RNA sequencing, initially there were 16,032 single cells detected, and after quality filtering 14,315 single cells were remained and used for further analyses (**Figure 1**; **Supplementary Figure S1A**). 10 different cell types were identified, including spermatogonia subpopulation 1 and 2 (c3, c5), spermatocyte (c1), early round spermatid (c4), middle round spermatid (c6), late round spermatid (c0), elongated spermatid (c2), Leydig (c9), Sertoli (c7), and a small proportion of red blood cells (c8) (**Figure 2A**). In detail, the cell type with highest proportion was late round spermatid, and the second highest one was spermatocyte, which were 3,889 and 2,396 single cells respectively (**Supplementary Figure S1B**). Notably, the proportion of two somatic cells (Leydig and Sertoli) is relatively small compared to the germ cells, in total, we identified 58 Leydig cells and 133 Sertoli cells (**Supplementary Figure S1B**). The top three cell types with largest number of detected genes are SPG2, early round spermatid and SPG1, and the average number of genes is also positively corrected with the average number of UMIs (**Figure 2B**; **Supplementary Figures S1C,D**). Lower gene expression was detected in late-stage cell types of spermatogenesis (elongated spermatids, and late round spermatids) (**Supplementary Figure S1D**). All the cell types identified contain lower proportions of expressed mitochondrial genes (0.5–2% on average, **Figure 2B**), which indicates rare contamination of apoptotic, stressed, or lower quality cells.

The cell types were annotated based on known marker genes examined in zebrafish testis or the testis of other species (**Supplementary Table S1**; **Supplementary Figure S2**). For example, *star* was reported as the marker gene of Leydig cell (Lin et al., 2017), and our scRNA-seq data clearly showed that it was specifically expressed in cell cluster c9. Furthermore, based on previous studies, *sycp2* had been recognized as a marker gene for spermatocyte cells (Takemoto et al., 2020), while *dazl* was mainly expressed in SPG and weakly detected at later spermatogenesis stages (Chen et al., 2013). In this study, *sycp2* was found to be significantly enriched in c1, and *dazl* was discovered to be highly enriched in c3 and c5. **Supplementary Table S3**

Novel Zebrafish Spermatogenesis Marker Genes

Next, the cluster level marker genes were examined, in total, thousands of marker genes were identified for each cell type,

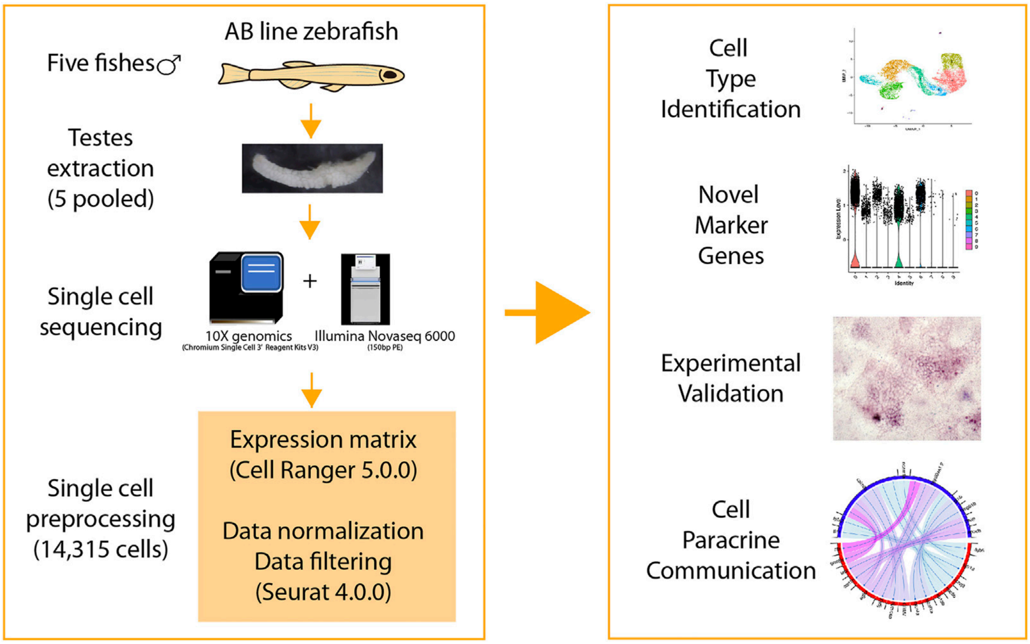


FIGURE 1 | Overview of this study.

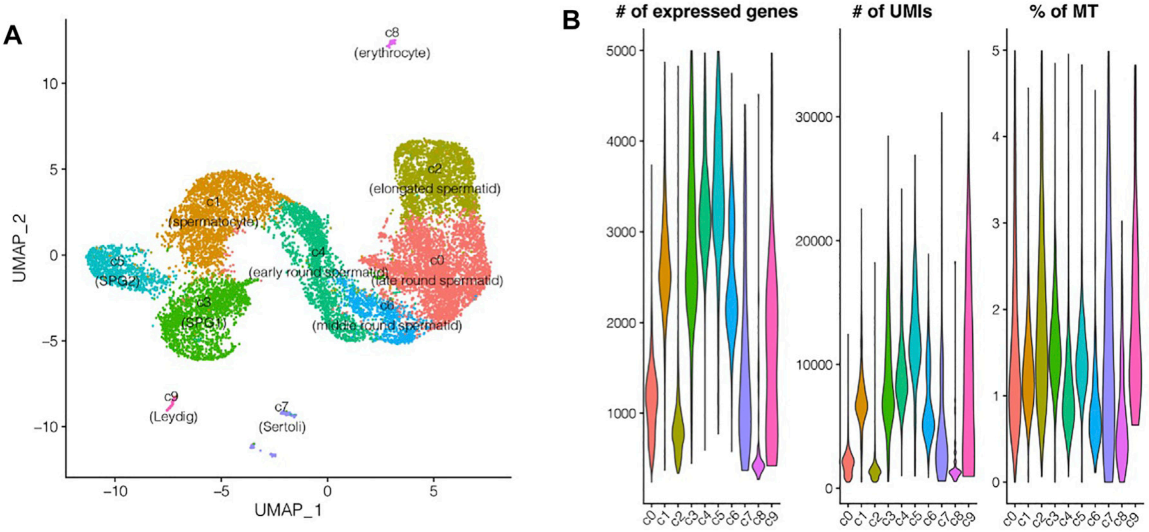


FIGURE 2 | Single cell RNA sequencing of zebrafish testes. **(A)** UMAP representation of 10 cell populations. **(B)** Violin plots for the distribution of the number of expressed genes, UMIs, and the percentage of mitochondrial UMIs.

including 1,449 marker genes for SPG1, 1,271 marker genes for SPG2, 418 genes for Sertoli cells, and 412 genes for Leydig cells (**Figure 3A**; **Supplementary Table S3**). The top 10 marker genes based on the rankings of gene expression foldchange can be found in **Figure 3B**, including *star* for Leydig (c9), *sycp2* for Spermatocyte (c1), *hbba1*/*hbba2* for red blood cell (c8).Based on the marker genes, GO enrichment analysis was performed: "cilium assembly," "cilium movement," "cilium organization" GO terms were significantly enriched in spermatids (c4, c6, c0, c2), and "DNA recombination," "chromatin organization," "mRNA metabolic process" GO terms were significantly enriched in early-stage cell types of spermatogenesis (c1, c3, c5) (**Figure 3C**). Notably, "ribosome assembly" and other ribosome related terms were enriched in non-spermatids cell types, which might indicate lack of translation events in spermatids.

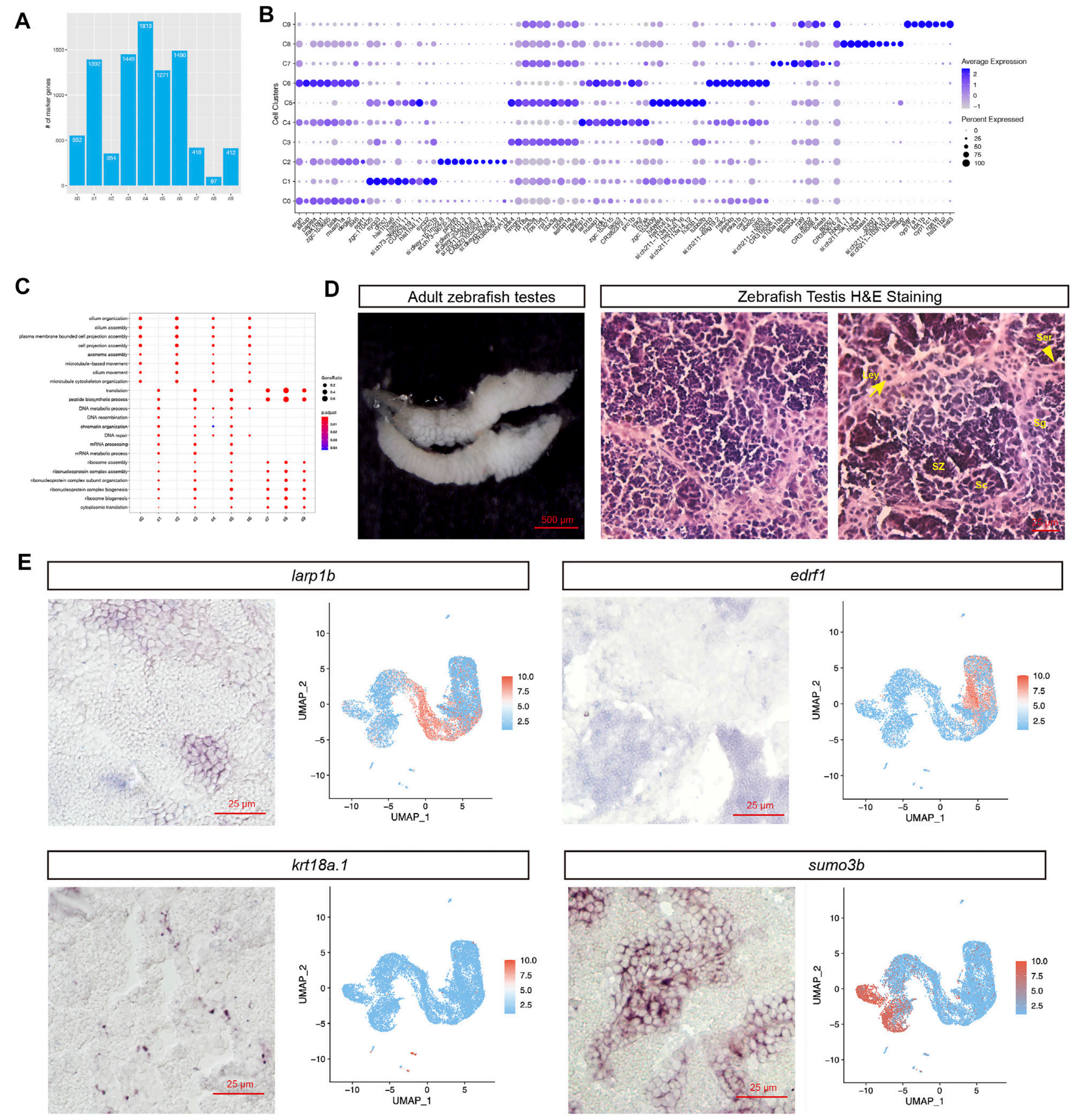


FIGURE 3 | Studies of novel marker genes for each cell population. **(A)** Distribution of the number of marker genes in each cell population. **(B)** Dotplot visualization of top 10 marker genes for each cell cluster. **(C)** GO term enrichment analysis of the marker genes. **(D)** H&E staining of different cell types (Ser, Sertoli cell; Sg, Spermatogonia; Sc, Spermatocyte; SZ: spermatozoa; Ley: Leydig cell). **(E)** RNA *in situ* hybridization of novel marker genes. (For each marker gene, both of the ISH experimental result (Left) and scRNA-seq UMAP representation of gene expression were presented (Right))

The cluster-level marker genes provide a way to find novel spermatogenesis marker genes. We found *edrf1* (erythroid differentiation regulatory factor 1) was specifically expressed at elongated spermatids, and *larp1b* (La ribonucleoprotein 1B) RNA was enriched in round spermatids (**Figure 3B**). We further carried out RNA *in situ* hybridization (ISH) to validate the novel testicular cell type specific marker genes. Firstly, for better recognition of each cell types in zebrafish testis, we annotated the cell types based on the cell shapes and relative locations in the H&E staining image (**Figure 3D**). As we expected, the ISH staining for *larp1b* and *edrf1* RNA showed positive signal in round spermatids and elongated spermatids respectively, and

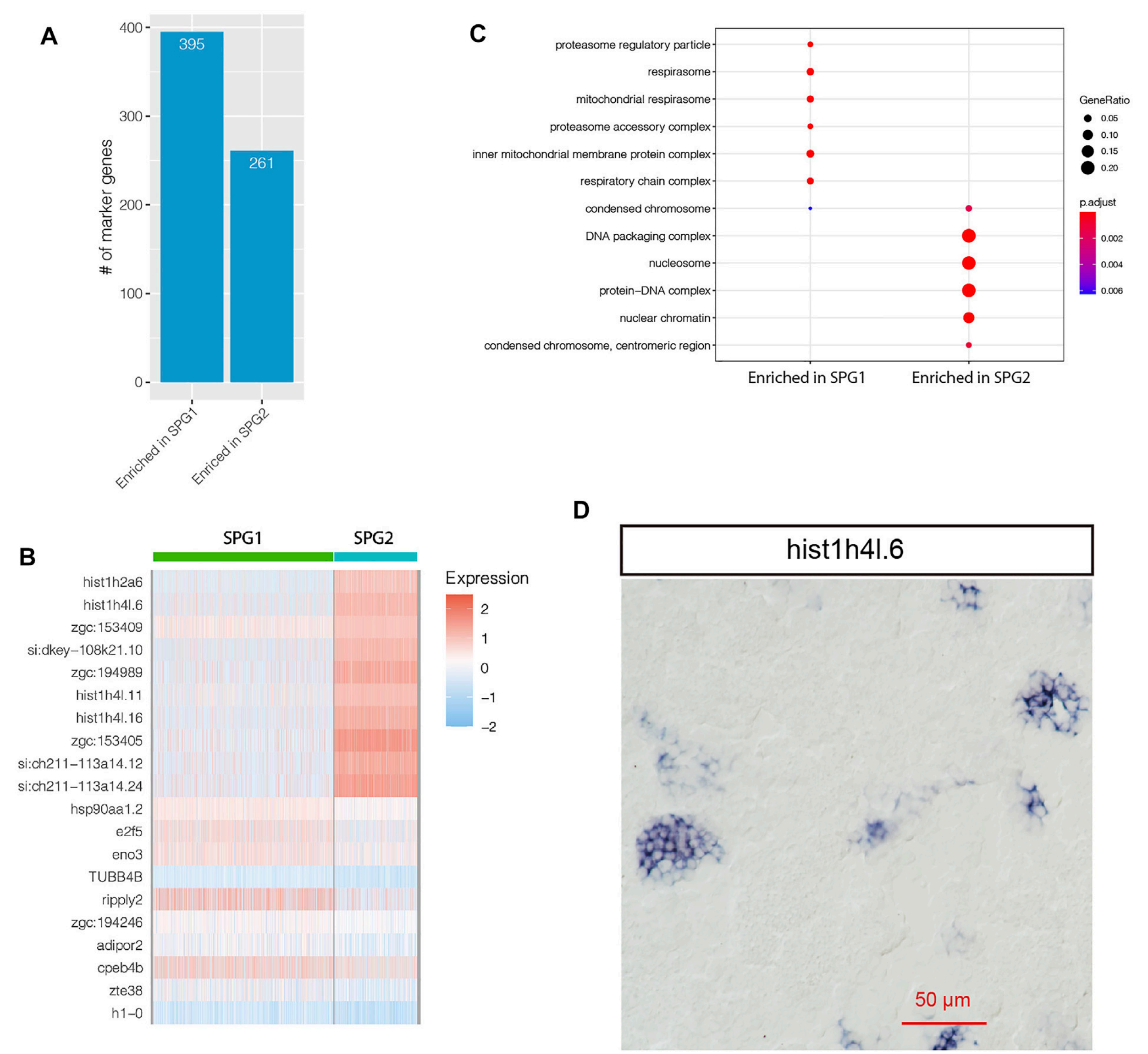


FIGURE 4 | Comparative study of SPG1 and SPG2. **(A)** Distribution of the number of genes enriched in SPG1 or SPG2. **(B)** Heatmap visualization of top 10 marker genes. **(C)** GO term enrichment analysis of the enriched genes. **(D)** RNA *in situ* hybridization result for SPG2 specifically expressed gene hist1h4l.6.

the ISH staining for *krt18a.1* and *sumo3b* showed positive signal in Sertoli and Spermatogonia (SPG) cells respectively (**Figure 3E**). Furthermore, we also provided all the novel marker genes for each zebrafish testicular cell type in **Supplementary Table S3**.

Subpopulation Study of Spermatogonia Cells Identified Marker Genes for SPG1 and SPG2

For zebrafish spermatogonia cells, two subpopulations were found in this study: SPG1 and SPG2 (**Figure 2A**). In total, 656 genes were significantly differentially expressed between SPG1 and SPG2 (**Figure 4A**, **Supplementary Table S4**). We found histone related genes involved in heterochromatin assembly were enriched in SPG2, such as *hist1h2a6*, *hist1h4l.6*, *hist1h4l.11* (**Figure 4B**). Further GO enrichment analysis based on SPG subpopulation specifically expressed genes confirmed that the GO terms such as “condensed chromosome,” “DNA packaging complex” were significantly enriched in SPG2 (**Figure 4C**). In human, there were three types of spermatogonia: Type A dark, Type A pale, and Type B, where type A pale went through division every seminiferous epithelial cycle (Kolasa et al., 2012). In zebrafish, there were also three types of spermatogonia: type A undifferentiated (A_{und}), type A differentiated (A_{diff}), and type B spermatogonia (Schulz et al., 2010). The enrichment of GO term “condensed chromosome” suggests more histone proteins involved for chromosome DNA packaging, as shown in **Figure 4B** histone protein genes such as

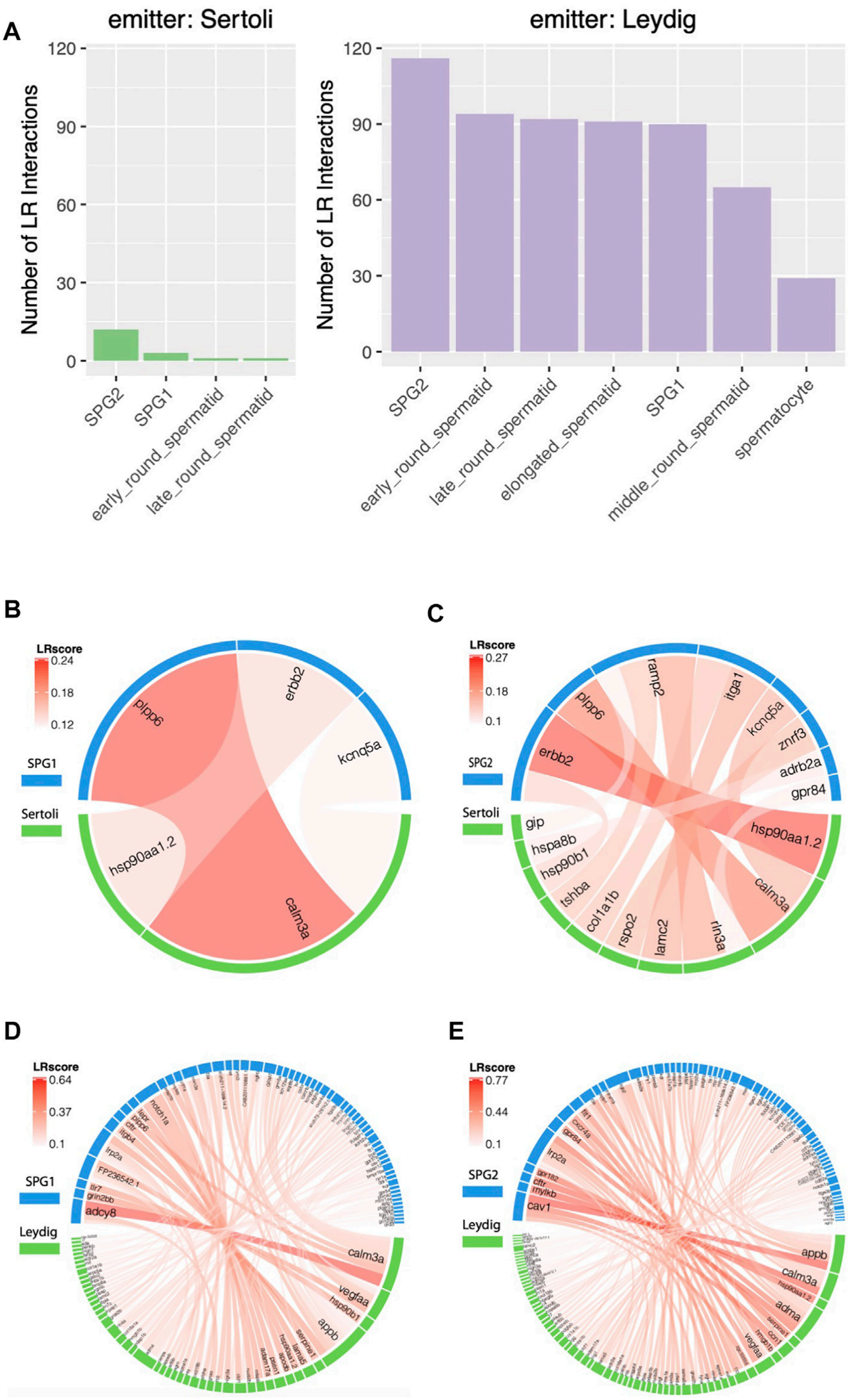


FIGURE 5 | Comparative studies of paracrine influence of Leydig and Sertoli cells. **(A)** Summary of the Ligand-Receptor (LR) interactions between somatic and germ cells. **(B)** LR interaction chord diagram for the influence of Sertoli cells on SPG1. **(C)** LR interaction chord diagram for the influence of Sertoli cells on SPG2. **(D)** LR interaction chord diagram for the influence of Leydig cells on SPG1. **(E)** LR interaction chord diagram for the influence of Leydig cells on SPG2.

hist1h2a6, *hist1h4l.16* were significantly enriched in SPG2. Further ISH experiment for *hist1h4l.6* RNA confirmed its highly specific expression pattern in spermatogonia cells (**Figure 4D**). The enriched GO terms and high expression level of histone protein genes in SPG2 suggests that SPG2 was actively undergoing mitosis and differentiation, which was in line with the definition of type A differentiated spermatogonia in zebrafish (A_{diff}) and type B spermatogonia. In contrast to SPG2, SPG1 was not enriched with histone protein genes and DNA packaging complex (**Figure 4C**), which indicates that SPG1 was not actively involved in mitosis, and it might be the type A undifferentiated spermatogonia cells (A_{und}). New marker genes for A_{und} were found based on our scRNA-seq datasets, such as *eno3*, *e2f5*, and *ripply2* (**Figure 4B**). Interestingly, *ENO3* gene might be involved in muscle regeneration, and had been recognized as one of the marker genes during muscle regenerative phase of maturation (Forcina et al., 2020). Furthermore, *E2F5* gene belongs to *E2F* family of transcription factors, and it might play a role as a tumor suppressor gene and control the cell cycle process (Chen et al., 2009).

Leydig Cells Manifest Stronger Paracrine Influence on Germ Cells

Lastly, we studied the influence of somatic cells on zebrafish germ cells and the spermatids. As shown in **Figure 5A**, more LR interactions were observed between Leydig-Germ cells than Sertoli-Germ cells, which demonstrated a stronger influence of Leydig cell on germ cells than Sertoli cells. However, based on previous studies, Sertoli cell played important roles in germ cell differentiation, survival, development, and physiological functioning (Schulz et al., 2010), which made us expect more LR interactions emitted from Sertoli than Leydig. In our study (**Figure 5A**), both of the two somatic cells (Sertoli and Leydig, as LR interaction signal emitter) had most abundant paracrine interactions with SPG2 (as LR interaction signal receiver), while the number of LR interactions emitted from Leydig (116) is much more than the number of LR interactions emitted from Sertoli (12). Among two spermatogonia cells, both of Leydig and Sertoli cells exhibit higher influence on SPG2 than SPG1. The detail LR interactions between somatic-SPG cells were shown by chord diagrams (**Figures 5B–E**). *Calm3a-plpp6* was predicted with highest LR interaction score (LRscore) between Sertoli (emitter) and SPG1 (receiver) (**Figure 5B**), and *hsp90aa1.2-erbb2* showed the strongest paracrine interaction between Sertoli and SPG2 (**Figure 5C**). The top LR interactions between Leydig and SPG1 include *calm3a-adcy8*, *vegfaa-grin2bb*, *hsp90b1-tlr7*, *appb-FP236542.1* (**Figure 5D**), and the top LR interactions between Leydig and SPG2 include *appb-cav1*, *calm3a-mylkb*, *hsp90aa1.2-cftr*, *adma-gpr182* (**Figure 5E**). We also analyzed other paracrine interactions between each testicular cell, and the full list of LR interactions and their LR interaction scores can be found in **Supplementary Table S5**. Our result based on the comparative analyses of the paracrine interactions between somatic-germ cells suggested the importance for the future study of the paracrine influence of Leydig on Germ cells.

DISCUSSION

In this study, for the first time, the whole transcriptome of the zebrafish testis was sequenced at single cell level. In total, the transcriptome information of 14, 315 single testicular cells were profiled, and 10 cell-clusters were revealed, including spermatids, spermatocytes, SPG, Leydig cells, and Sertoli cells. In this study, due to the high cost of single cell sequencing, only one sample was sequenced, however, testes from five zebrafish were pooled for single cell sequencing to reduce the influence of individual variations. Our study not only illustrated new marker genes for each cell population of zebrafish testis, but also revealed the stronger paracrine influence of Leydig cells on Germ cells.

Furthermore, thousands of novel marker genes were identified, which could be important for further functional examination of each cell types in zebrafish testis. For example, sumo3b and stmn1a were discovered to be specifically expressed in SPG, *pimr93* and *si:ch73-367j5.3* were specifically expressed in the elongated spermatids. Recently, the adult testis transcriptome of ray-finned fish named orange-spotted grouper (*Epinephelus coioides*) was profiled at single cell level, and several novel marker genes were identified for testicular cells (Wu et al., 2021). Among the cell-marker genes identified, the genes such as *pprc1*, *top2b* were identified as the marker genes for spermatogonia and spermatocyte cells in both of orange-spotted grouper fish and zebrafish respectively. However, discrepancies had also been found in the two species, such as *supt16h* had been identified as top marker genes for spermatogonia cells in orange-spotted grouper fish but showed a broader expression pattern from spermatogonia cells to spermatids in our study (**Supplementary Figure S3**), which indicates cell type differential gene expression in the two types of fishes.

Next, based on the ligand-receptor paracrine interaction analyses, our study revealed stronger cell-cell communications between Leydig and germ cells, and the key players of LR interactions involved were also revealed in this study. For Sertoli cells, we only found 15 LR interactions between itself and SPG (SPG1 and SPG2), however, this number in Leydig is 206. It had been suggested that Sertoli cell is important for SPG stem cell niche by providing paracrine and other signals (França et al., 2016), which made us expect more LR interactions from Sertoli than from Leydig to SPG. To check possible factors that might increase Leydig-SPG LR interactions, we examined the influence of sequencing depth and the number of detected genes in Leydig and Sertoli for their paracrine interactions with SPG. Indeed, we found a higher average sequencing depth for Leydig (9,140 UMIs) compared to Sertoli (5,459 UMIs) (**Supplementary Figure S1C**), but the difference is not as large as the one for LR interactions (206 vs 15). Furthermore, Leydig and Sertoli cells contained similar number of average detected genes, which suggests its influence on the LR interactions might be small (**Supplementary Figure S1D**). This information indicates that both differences for sequencing-depth and the number of detected genes in the two somatic cell types cannot fully explain the observation of much stronger Leydig-SPG paracrine interactions in zebrafish testis, and future work is needed to study the paracrine influence of Leydig cells to the germ cells.

Finally, the transcriptome of the testes from other species had been extensively profiled at single cell levels, including human (Sohni et al., 2019; Guo et al., 2020; Guo et al., 2018; Shami et al., 2020), mouse (Chen et al., 2018; Green et al., 2018; Lukassen et al., 2018), monkey (Shami et al., 2020), and fly (Witt et al., 2019). However, as far as we know, before this study, there was no single cell level transcriptome study in zebrafish testis. Our study on the single cell RNA sequencing of zebrafish testes could be an important complementary resource for future study of spermatogenesis in zebrafish. Furthermore, to illustrate how the findings in zebrafish study could also contribute to human infertility studies, we examined the male infertility genes in GWAS Catalog database (https://www.ebi.ac.uk/gwas/)(Buniello et al., 2019). In total, we found nine genes that were recorded to be associated with "male infertility" or "non-obstructive azoospermia," and three of them (*CLASP2*, *CRACR2A*, *SOX5*) were identified in this study as zebrafish testis cell-type specific marker genes (**Supplementary Table S6**). Notably, *clasp2*, *cracr2a* and *sox5* were identified as the marker genes for spermatocyte (c1), middle round spermatid (c6), and early round spermatid (c4) respectively in zebrafish testis (**Supplementary Table S3**), which indicates that zebrafish could be used as a model to study the genes that are associated with human infertility and might help to reveal the underlying mechanism.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://bigd.big.ac.cn/gsa, CRA003925.

ETHICS STATEMENT

The animal study was reviewed and approved by Ethics Committee of Nantong University.

AUTHOR CONTRIBUTIONS

GX, DL supervised and designed this project. GX, DL, PQ wrote the manuscript. GX, PQ, and JK analyzed the data. PQ and JK performed the experiments.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (31900484 received by GX; 2018YFA0801004 received by DL), Natural Science Foundation of Jiangsu Province (BK20190924 received by GX; BK20180048, and BRA2019278 received by DL).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.851719/full#supplementary-material

REFERENCES

Assis, L. H. C., Crespo, D., Morais, R. D. V. S., França, L. R., Bogerd, J., and Schulz, R. W. (2016). INSL3 Stimulates Spermatogonial Differentiation in Testis of Adult Zebrafish (*Danio rerio*). *Cell Tissue Res* 363 (2), 579–588. doi:10.1007/s00441-015-2213-9

Babakhanzadeh, E., Nazari, M., Ghasemifar, S., and Khodadadian, A. (2020). Some of the Factors Involved in Male Infertility: A Prospective Review. *Ijgm* Vol. 13, 29–41. doi:10.2147/ijgm.s241099

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of Published Genome-wide Association Studies, Targeted Arrays and Summary Statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012. doi:10.1093/nar/gky1120

Cabello-Aguilar, S., Alame, M., Kon-Sun-Tack, F., Fau, C., Lacroix, M., Colinge, J., et al. (2020). SingleCellSignalR: Inference of Intercellular Networks from Single-Cell Transcriptomics. *Nucleic Acids Res.* 48 (10), e55. doi:10.1093/nar/gkaa183

Chen, H.-Z., Tsai, S.-Y., and Leone, G. (2009). Emerging Roles of E2Fs in Cancer: an Exit from Cell Cycle Control. *Nat. Rev. Cancer* 9 (11), 785–797. doi:10.1038/nrc2696

Chen, S. X., Bogerd, J., Schoonen, N. E., Martijn, J., de Waal, P. P., and Schulz, R. W. (2013). A Progestin (17α,20β-Dihydroxy-4-Pregnen-3-One) Stimulates Early Stages of Spermatogenesis in Zebrafish. *Gen. Comp. Endocrinol.* 185, 1–9. doi:10.1016/j.ygcen.2013.01.005

Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., et al. (2018). Single-cell RNA-Seq Uncovers Dynamic Processes and Critical Regulators in Mouse Spermatogenesis. *Cell Res* 28 (9), 879–896. doi:10.1038/s41422-018-0074-y

de Castro Assis, L. H., de Nóbrega, R. H., Gómez-González, N. E., Bogerd, J., and Schulz, R. W. (2018). Estrogen-induced Inhibition of Spermatogenesis in Zebrafish Is Largely Reversed by Androgen. *J. Mol. Endocrinol.* 60 (4), 273–284. doi:10.1530/jme-17-0177

Forcina, L., Cosentino, M., and Musarò, A. (2020). Mechanisms Regulating Muscle Regeneration: Insights into the Interrelated and Time-dependent Phases of Tissue Healing. *Cells* 9 (5), 9051297. doi:10.3390/cells9051297

França, L. R., Hess, R. A., Dufour, J. M., Hofmann, M. C., and Griswold, M. D. (2016). The Sertoli Cell: One Hundred Fifty Years of beauty and Plasticity. *Andrology* 4 (2), 189–212. doi:10.1111/andr.12165

Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., et al. (2018). A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. *Develop. Cel* 46 (5), 651–667. doi:10.1016/j.devcel.2018.07.025

Grive, K. J., Hu, Y., Shu, E., Grimson, A., Elemento, O., Grenier, J. K., et al. (2019). Dynamic Transcriptome Profiles within Spermatogonial and Spermatocyte Populations during Postnatal Testis Maturation Revealed by Single-Cell Sequencing. *Plos Genet.* 15 (3), e1007810. doi:10.1371/journal.pgen.1007810

Guo, J., Grow, E. J., Mlcochova, H., Maher, G. J., Lindskog, C., Nie, X., et al. (2018). The Adult Human Testis Transcriptional Cell Atlas. *Cel Res* 28 (12), 1141–1157. doi:10.1038/s41422-018-0099-2

Guo, J., Nie, X., Giebler, M., Mlcochova, H., Wang, Y., Grow, E. J., et al. (2020). The Dynamic Transcriptional Cell Atlas of Testis Development during Human Puberty. *Cell Stem Cell* 26 (2), 262–276. doi:10.1016/j.stem.2019.12.005

Hermann, B. P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K. N., Chen, I.-C., et al. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cel Rep.* 25 (6), 1650–1667. e1658. doi:10.1016/j.celrep.2018.10.026

Jung, M., Wells, D., Rusch, J., Ahmad, S., Marchini, J., Myers, S. R., et al. (2019). Unified Single-Cell Analysis of Testis Gene Regulation and Pathology in Five Mouse Strains. *Elife* 8. doi:10.7554/eLife.43966

Kolasa, A., Misiakiewicz, K., Marchlewicz, M., and Wiszniewska, B. (2012). The Generation of Spermatogonial Stem Cells and Spermatogonia in Mammals. *Reprod. Biol.* 12 (1), 5–23. doi:10.1016/s1642-431x(12)60074-6

Lau, X., Munusamy, P., Ng, M. J., and Sangrithi, M. (2020). Single-Cell RNA Sequencing of the Cynomolgus Macaque Testis Reveals Conserved Transcriptional Profiles during Mammalian Spermatogenesis. *Develop. Cel* 54 (4), 548–566. doi:10.1016/j.devcel.2020.07.018

Leal, M. C., de Waal, P. P., García-López, Á., Chen, S. X., Bogerd, J., and Schulz, R. W. (2009). Zebrafish Primary Testis Tissue Culture: an Approach to Study Testis Function *Ex Vivo*. *Gen. Comp. Endocrinol.* 162 (2), 134–138. doi:10.1016/j.ygcen.2009.03.003

Lin, Q., Mei, J., Li, Z., Zhang, X., Zhou, L., and Gui, J.-F. (2017). Distinct and Cooperative Roles of Amh and Dmrt1 in Self-Renewal and Differentiation of Male Germ Cells in Zebrafish. *Genetics* 207 (3), 1007–1022. doi:10.1534/genetics.117.300274

Lukassen, S., Bosch, E., Ekici, A. B., and Winterpacht, A. (2018). Single-cell RNA Sequencing of Adult Mouse Testes. *Sci. Data* 5, 180192. doi:10.1038/sdata.2018.192

Members, C-N. (2021). Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2021. *Nucleic Acids Res.* 49 (D1), D18–D28. doi:10.1093/nar/gkaa1022

Nishimura, H., and L'Hernault, S. W. (2017). Spermatogenesis. *Curr. Biol.* 27 (18), R988–R994. doi:10.1016/j.cub.2017.07.067

Orban, L., Sreenivasan, R., and Olsson, P. E. (2009). Long and Winding Roads: Testis Differentiation in Zebrafish. *Mol. Cel Endocrinol* 312 (1-2), 35–41. doi:10.1016/j.mce.2009.04.014

Potter, S. S. (2018). Single-cell RNA Sequencing for the Study of Development, Physiology and Disease. *Nat. Rev. Nephrol.* 14 (8), 479–492. doi:10.1038/s41581-018-0021-7

Safian, D., Morais, R. D. V. S., Bogerd, J., and Schulz, R. W. (2016). Igf Binding Proteins Protect Undifferentiated Spermatogonia in the Zebrafish Testis against Excessive Differentiation. *Endocrinology* 157 (11), 4423–4433. doi:10.1210/en.2016-1315

Schulz, R. W., de França, L. R., Lareyre, J.-J., LeGac, F., Chiarini-Garcia, H., Nobrega, R. H., et al. (2010). Spermatogenesis in Fish. *Gen. Comp. Endocrinol.* 165 (3), 390–411. doi:10.1016/j.ygcen.2009.02.013

Shami, A. N., Zheng, X., Munyoki, S. K., Ma, Q., Manske, G. L., Green, C. D., et al. (2020). Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Develop. Cel* 54 (4), 529–547. doi:10.1016/j.devcel.2020.05.010

Shao, X., Liao, J., Li, C., Lu, X., Cheng, J., and Fan, X. (2020). CellTalkDB: a Manually Curated Database of Ligand-Receptor Interactions in Humans and Mice. *Brief Bioinform* 22 (4), bbaa269. doi:10.1093/bib/bbaa269

Shi, J., Fok, K. L., Dai, P., Qiao, F., Zhang, M., Liu, H., et al. (2021). Spatio-temporal Landscape of Mouse Epididymal Cells and Specific Mitochondria-Rich Segments Defined by Large-Scale Single-Cell RNA-Seq. *Cell Discov* 7 (1), 34. doi:10.1038/s41421-021-00260-7

Sohni, A., Tan, K., Song, H.-W., Burow, D., de Rooij, D. G., Laurent, L., et al. (2019). The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cel Rep.* 26 (6), 1501–1517. doi:10.1016/j.celrep.2019.01.045

Takemoto, K., Imai, Y., Saito, K., Kawasaki, T., Carlton, P. M., Ishiguro, K.-i., et al. (2020). Sycp2 Is Essential for Synaptonemal Complex Assembly, Early Meiotic Recombination and Homologous Pairing in Zebrafish Spermatocytes. *Plos Genet.* 16 (2), e1008640. doi:10.1371/journal.pgen.1008640

Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., et al. (2017). GSA: Genome Sequence Archive *. *Genomics, Proteomics & Bioinformatics* 15 (1), 14–18. doi:10.1016/j.gpb.2017.01.001

Witt, E., Benjamin, S., Svetec, N., and Zhao, L. (2019). Testis Single-Cell RNA-Seq Reveals the Dynamics of De Novo Gene Transcription and Germline Mutational Bias in Drosophila. *Elife* 8, 47138. doi:10.7554/eLife.47138

Wu, X., Yang, Y., Zhong, C., Wang, T., Deng, Y., Huang, H., et al. (2021). Single-Cell Atlas of Adult Testis in Protogynous Hermaphroditic Orange-Spotted Grouper, *Epinephelus coioides*. *Int. J. Mol. Sci.* 22 (22), 222212607. doi:10.3390/ijms222212607

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 10 May 2022
doi: 10.3389/fgene.2022.865103

Target-Sequencing of Female Infertility Pathogenic Gene Panel and a Novel TUBB8 Loss-of-Function Mutation

Hongxia Yuan[1]†, Jianhua Chen[1]†, Na Li[2], Hui Miao[2], Yao Chen[3], Shuyan Lyu[4], Yu Qiao[4], Guangping Yang[1], Hui Luo[1], Liangliang Chen[1], Fei Mao[3], Lingli Huang[3], Yanni He[1], Saifei Hu[1], Congxiu Miao[2]*, Yun Qian[3]* and Ruizhi Feng[1,3]*

[1]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China, [2]The Reproduction Engineer Key Laboratory of Shanxi Health Committee, Department of Reproductive Genetics, Institute of Reproduction and Genetics of Changzhi Medical College, Heping Hospital of Changzhi Medical College, Changzhi, China, [3]Reproductive Medical Center of the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China, [4]The Affiliated Huaian No. 1 People's Hospital of Nanjing Medical University, Huaian, China

Edited by:
Wenjie Shi,
Academy of Military Medical Sciences (AMMS), China

Reviewed by:
Yuan Gao,
Shandong University, China
James A Poulter,
University of Leeds, United Kingdom

***Correspondence:**
Congxiu Miao
mcxms@163.com
Yun Qian
qianyun@njmu.edu.cn
Ruizhi Feng
ruizhifeng@njmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *29 January 2022*
Accepted: *07 March 2022*
Published: *10 May 2022*

Citation:
Yuan H, Chen J, Li N, Miao H, Chen Y, Lyu S, Qiao Y, Yang G, Luo H, Chen L, Mao F, Huang L, He Y, Hu S, Miao C, Qian Y and Feng R (2022) Target-Sequencing of Female Infertility Pathogenic Gene Panel and a Novel TUBB8 Loss-of-Function Mutation. Front. Genet. 13:865103. doi: 10.3389/fgene.2022.865103

Genetic screening is an important approach for etiology determination and helps to optimize administration protocols in reproductive centers. After the first pathogenic gene of female infertility was reported in 2016, more and more new pathogenic genes were discovered, and we sought to develop an efficient and cost-effective method for genetic screening in patients. In this study, we designed a target-sequencing panel with 22 female infertility-related genes, namely, TUBB8, PATL2, WEE2, and PANX1 and sequenced 68 primary infertility (PI) and recurrent pregnancy loss (RPL) patients. We sequenced 68 samples reaching an average depth of 1559× and detected 3,134 variants. Among them, 62.2% were synonymous single-nucleotide variants (SNVs) and 36.3% were non-synonymous SNVs. The remaining 1.5% are indels (insertions and deletions) and stop-gains. DNAH11 and TUBB8 are the two genes that mutated most frequently. We also found a novel TUBB8 variant (c.898_900del; p.300_300del), proved its loss-of-function mechanism, and profiled the interactome of the wild-type (WT) and mutant TUBB8 proteins. Overall, this target-sequencing method provides an efficient and cost-effective approach for screening in IVF clinics and will support researchers for the discovery of new pathogenic variants.

Keywords: genetic screening, female infertility, target sequencing, TUBB8, mutation

INTRODUCTION

Genetic screening is a newly developed technology based on the widespread next-generation sequencing (Pös et al., 2019). Compared with the comprehensive but expensive approach of genome sequencing, exome sequencing has already been used for new pathogenic genes discovery in patient cohorts and preimplantation genetic testing in reproductive centers. For known genetic disorders, the simple and cost-effective method to detect pathogenic genes is target sequencing. Designing hybridization probes or multiple PCR primers, genes of interest can be enriched and sequenced, thus enabling researchers to detect variants in hot-spot regions. Targeted sequencing has been widely used in numerous areas such as genetic disorders

(Eggers et al., 2016), pharmacogenetics (Gordon et al., 2016), and cancer (Matsunaga, 2009). With profiles of actionable gene alterations, the ovarian cancer target sequencing makes it possible to find variants responding to molecular targeting drugs. Tumors containing variants on PIK3CA, AKT1, and PTEN are targetable by P13K/AKT/mTOR inhibitors (Takenaka et al., 2015). Another crucial application is prenatal screening. While traditional ultrasonography at the latter stage of pregnancy is usually used to diagnose skeletal dysplasia, target sequencing of recurrent pathogenic variants in FGFR3, COL1A2, etc. in the early gestational stage meets the need of precise diagnosis of skeletal dysplasia (Ching-Yuan Wang et al., 2021).

Infertility affects about 48.5 million couples in the world (Chiware et al., 2021). Among multiple factors contributing to this complex disease, genetic causes have been drawn increasing attentions (Miyamoto et al., 2017; Beke et al., 2019). Genetic screening has been applied in some reproductive centers for years, especially in male infertile patients (Liu et al., 2020). Screening in genetic abnormalities, such as Y microdeletions and chromosomal aberration, and known pathogenic genes (CFTR for obstructive azoospermia, NR5A1 in disorders of sex development, etc.), have been routinely performed for decades (An et al., 2021). However, the pathogenic genes of female infertility were hidden until the first gene TUBB8 was reported to cause oocyte maturation arrest (Feng et al., 2016). Since then, a number of genes have been discovered in primary infertility (PI) and recurrent pregnancy loss (RPL) patients, namely, PATL2 (Chen et al., 2017a), TRIP13 (Zhang et al., 2020a), WEE2 (Sang et al., 2018), etc. These newly found genes contribute to female infertile phenotypes including oocyte maturation arrest, zygotic cleavage failure, and embryo developmental arrest. Screening of these genes in patients with idiopathic assisted reproduction failures would greatly help in diagnosis and genetic counseling.

In this study, we designed a target sequencing gene panel for detecting pathogenic genes of female infertility. A total of 68 PI/RPL patients were recruited, sequenced, and analyzed. We found two TUBB8 variants, namely, a novel single-nucleotide variant (SNV), and demonstrated its loss-of-function pathogenic mechanism. Interactome of wild-type (WT) and mutant TUBB8 proteins were also profiled. This female infertility gene panel would be an efficient and cost-effective tool for genetic screening and diagnosis, as well as a method for researchers to discover more variants, and further study genetic pathogenicity of female infertility.

MATERIALS AND METHODS

Human Subjects, Standards of Inclusion and Exclusion, and Ethical Approval

Patients were recruited from the Clinical Center of Reproductive Medicine, the second affiliated hospital of Nanjing Medical University, the affiliated Huaian No. 1 People's Hospital of Nanjing Medical University, and Heping Hospital affiliated to Changzhi Medical College during 2020.1–2021.10. All the recruited patients had a history of PI or RPL. The age of the participated patients was no more than 38 years. They were not diagnosed with other complex reproductive diseases such as polycystic ovarian syndrome, premature ovarian failure, and premature ovarian insufficiency. No chromosome defects or infertile factors were found in their spouses. Peripheral blood (5 ml) was collected from each patient. All participating patients gave informed consent. This study was approved by the Ethics Committee of Nanjing Medical University.

Panel Characterization and Information of Each Gene

Our gene panel includes 22 female infertility-related genes: TUBB8, PATL2, WEE2, PADI6, TLE6, ZP1, ZP2, ZP3, NLRP2, NLRP5, PANX1, REC114, PLCZ1, CDC20, ANAPC4, TRIP13, KPNA7, BTG4, DNAH11, CCNO, LHCGR, and FOXP3 (Sang et al., 2021; Mu et al., 2020; Furuta et al., 2000; Wang et al., 2012; Zhang et al., 2020b; Weijie Wang et al., 2021). Detailed information of each gene is shown in **Table 1**. According to their phenotypes, these genes are classified into four categories (**Figure 1**). Some genes such as TUBB8 have spectrums of phenotypes, which means that they could be responsible for more than one female infertility phenotype (Chen et al., 2017b).

Genomic DNA Extraction and Target-Panel Sequencing

Genomic DNA was extracted from the peripheral blood using TIANamp Genomic DNA KIT (TIANGEN, DP304-03) following the manufacturer's instructions. DNA concentrations were detected on Nanodrop 2000. Concentrations above 25 μg/μl and $OD_{260/280}$ between 1.8 and 2.0 were used for the next steps. DNA qualities were tested with agarose gel electrophoresis. Specialized DNA probes were designed according to the coding regions of the target genes. Hybridization capture and enrichment were performed to screen candidate variants (iGeneTech Ltd. Co., China). Sequencing was performed on Illumina NovaSeq 6000 platform (iGeneTech Ltd. Co., China). Raw reads were filtered to remove low-quality reads. Q30 ≥80% was regarded as qualified. Clean data were mapped using BWA (Burrows-Wheeler Alignment tool), and statistical analyses were conducted (iGeneTech Ltd.Co., China). Candidate variants were selected for further analysis based on the following: variants located in exons or splicing regions, non-synonymous, and non-benign as predicted by InterVar. The left variants were further screened with minor allele frequency no more than 0.1% in at least one database such as NHLBI-ESP, 1000G, ExAC, gnomAD, and the in-house database of iGeneTech. Sanger sequencing was performed for validation.

Plasmid Construction, Cell Culture, and Transfection

Eukaryotic expression vectors pCMV6 with Myc-DDK (FLAG) tags containing WT and mutant TUBB8 were kindly provided by

TABLE 1 | Genes included in the target sequencing panel and their phenotypes.

Gene	Phenotype	Locus	OMIM	30× coverage (%)
TUBB8	Oocyte maturation defects	10p15.3	616768	100
PATL2	Oocyte maturation defects	15q21.1	614661	100
WEE2	Fertilization failure	7q34	614084	100
PADI6	Early embryo arrest	1p36.13	610363	100
TLE6	Early embryo arrest	19p13.3	612399	100
ZP1	Abnormal zona pellucida formation	19p13.3	195000	100
ZP2	Abnormal zona pellucida formation	16p12.3-p12.2	182888	100
ZP3	Abnormal zona pellucida formation	7q11.23	182889	100
NLRP2	Early embryo arrest	19q13.42	609364	100
NLRP5	Early embryo arrest	19q13.43	609658	100
PANX1	Oocyte death	11q21	608420	100
REC114	Oocyte death	15q24.1	618421	100
PLCZ1	Fertilization failure	12p12.3	608075	100
CDC20	Meiosis defects	1p34.2	603618	100
ANAPC4	Abnormal gametogenesis and embryogenesis	4p15.2	606947	100
TRIP13	Oocyte maturation defects	5p15.33	604507	100
KPNA7	Fertilization failure	7q22.1	614107	100
BTG4	Zygotic cleavage failure	11q23.1	605673	100
LHCGR	Abnormal zona pellucida formation	2p16.3	152790	100
FOXP3	Immunodysregulation, polyendocrinopathy	Xp11.23	300292	100
DNAH11	Primary ciliary dyskinesia	7p15.3	603339	100
CCNO	Primary ciliary dyskinesia	5q11.2	607752	100

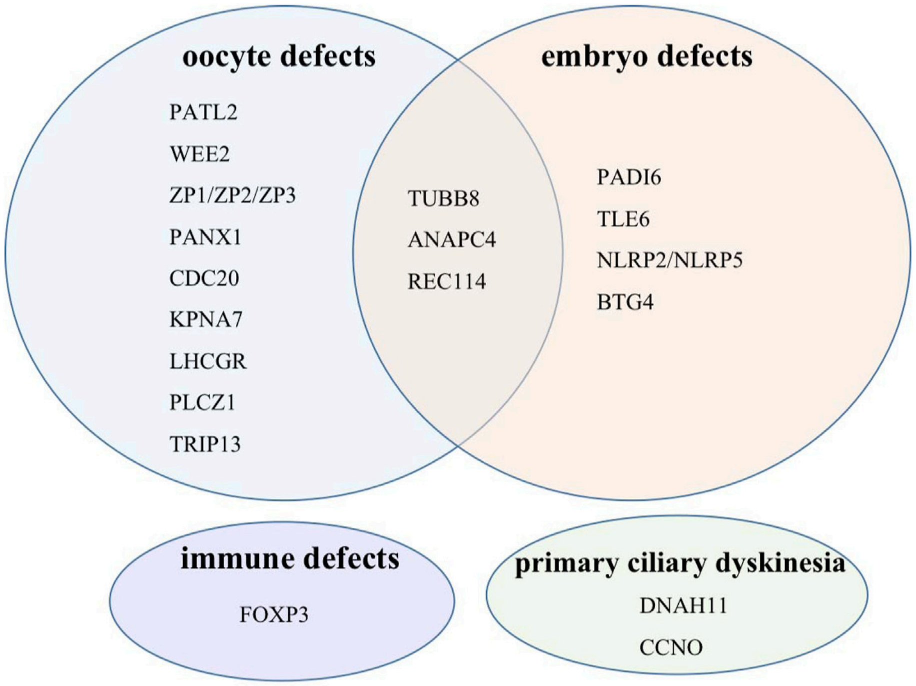


FIGURE 1 | Target genes classified into four categories according to their phenotypes.

Professors Lei Wang and Qing Sang in Fudan University. We recombined the vector to add a C-terminal AviTag for protein purification. pEF1a-BirA-V5-neo was purchased from MiaoLing Plasmid Platform. Variants were introduced using Mut Express II Fast Mutagenesis Kit 2 (Vazyme, C214). The HEK293T cell line was purchased from National Collection of Authenticated Cell

Culture and cultured with DMEM adding 10% fetal bovine serum and penicillin/streptomycin at 37°C with 5% CO_2. The cells were transiently transfected using Lipofectamine 2000 reagent following the standard protocols.

Immunofluorescence and Confocal Microscopy

Twenty-four hours after seeding, HEK 293T cells were transiently transfected with WT and mutant C-terminally Myc-DDK (FLAG)–tagged TUBB8 plasmids. The cells were gently washed three times using PBS 48 h after transfection. Then, 2% paraformaldehyde was applied to fix the cells. Permeabilization was performed with 0.5% Triton X-100 for 30 min at room temperature after being washed with PBS three times. Nonspecific binding was blocked by 5% BSA at room temperature for 1 h. The cells were then incubated in the dark at 4°C overnight with diluted antibodies: anti-FLAG M2-Cy3 (sigma, #A9594, 1:500), anti-α tubulin (cell signaling technology, #5063S, Alexa flour 488 conjugated, 1:250), and DAPI (Beyotime, #C1002, 1:1000). The cells were washed three times using 0.5% Triton X-100, adding 0.5% Tween-20 before mounting. The confocal images were taken using Leica TCS SP8 confocal laser scanning microscope.

Cells Lysis and Western Blots

Buffers used for cell harvest and purification were in reference to *Cytoskeleton Dynamics* (Maiato, 2020). Forty-eight hours post-transfection, cells were gently washed once by cold D-PBS (Beyotime, #C0221D), and then 3 ml fresh D-PBS was added to collect the cells into a new 15-ml conical centrifuge tube by pipetting up and down. Centrifugation at 200 *g* for 10 min was performed, and cold lysis buffer was added after entirely removing D-PBS. At 4°C, cells lysate was rotated for 30 min and centrifuged at 14000 *g* for 10 min. The supernatant was saved to measure concentrations with BCA protein assay kit (Beyotime, #P0012). 10% SDS-PAGE gel was applied, and proteins were then transferred to the PVDF membranes. Nonspecific binding was blocked by 5% BSA (Beyotime, #ST023) at room temperature for 1 h before incubating in diluted primary antibodies overnight at 4°C: anti-Vinculin (Abcam, #ab129002, 1:10000), anti-FLAG tag mouse monoclonal antibody (YIFEIXUE BIOTECHNOLOGY, #YFMA0036, 1:5000), anti-Myc tag antibody (Abcam, #ab32, 1:1000), TBCD polyclonal antibody (Proteintech, #14867-1-AP, 1:2000), TBCA polyclonal antibody (Proteintech, #12304-1-AP, 1:2000), and HRP-Streptavidin (Beyotime, #A0303, 1:100000). Secondary antibodies, HRP-conjugated goat anti-mouse IgG (YIFEIXUE BIOTECHNOLOGY, #YFSA01, 1:7500) or HRP-conjugated mouse anti-rabbit IgG (Sangon Biotech, #D110065, 1:7500), were used to incubate the membranes for 1 h at room temperature, and then washed three times in Tris-buffered saline containing 0.05% Tween-20. Chemistar™ High-sig ECL Western Blotting Substrate (Tanon, #180-5001) was used to detected immune complexes on Tanon 4500 SF.

Protein Purification, Silver Staining, and Mass Spectrometry

Cells in 10-cm dishes were co-transfected equimolar amounts of C-terminal AviTag WT/mutant TUBB8 pCMV6 vectors and pEF1a-BirA-V5-neo. The protocols for cell harvest and lysis were the same as WB. After determining the protein concentration, equal amounts of protein were purified by Dynabeads™ M280 Streptavidin (Invitrogen, #11205D), according to the protocols from the *Cytoskeleton Dynamics* (Maiato, 2020). Approximately, 30 μl of beads were treated with blocking buffer for 1 h at room temperature. Six milligrams of proteins was added into the beads and rotated at 4°C for 2 h. Subsequently, the beads were washed seven times with a wash buffer before being boiled with 40 μl of SDS-PAGE sample-loading buffer (Beyotime, #P0015A) at 95°C for 5 min. 10% SDS-PAGE was exploited to separate purified proteins. Silver staining was performed with a commercial kit following the standard instructions (Beyotime, #P0017S). Protein analysis was accomplished with shotgun mass spectrometry by Shanghai Bioprofile.

Statistical Analysis

GO enrichment analysis was performed and visualized by clusterProfiler R package (4.0) (Wu et al., 2021).

RESULTS

Gene Panel Performance Evaluation and Sequencing Data Analysis

To evaluate the performance of our panel capture and sequencing, the average values of relevant parameters were reported as below. The QC rate [clean bases (Mb)/raw bases (Mb)] was 91.07%, and the total reads mapping rate (mapped reads/clean reads) was 98.9%. The target reads capture rate (target reads/mapped reads) was 57.1%. The target effective rate [target effective bases (Mb)/total effective bases (Mb)] was 34.15%. The mean depth of panel sequencing is 1559×, and the 30× coverage is 100%. T 10% ×coverage, which meant about 155× coverage rate was 99.97%, and T 50% ×coverage was 92.56% (**Supplementary Table S1**). Qualified data by target sequencing laid a solid foundation for the precision of subsequent analysis.

An average of 14.2 variant-bearing genes were detected per patient. A total of 3,134 variants located in exons and splicing regions were found, in which 3,088 were SNVs and 46 were indels. In 3,088 SNVs, 1952 were synonymous and 1136 variants were non-synonymous. In 46 indels, there were 26 stop-gain, 11 non-frameshift deletions, 2 frameshift deletions, 6 non-frameshift insertions, and 1 frameshift insertion (**Figure 2A**). Based on clinical interpretation, 2,979 out of 3,134 were benign according to the ACMG (the American College of Medical Genetics and Genomics) and InterVar, while 86 variants were likely benign and 49 had uncertain significance, leaving 20 of

FIGURE 2 | Panel sequencing performance. **(A)** The proportions of different types of variants. **(B)** The proportions of variants with different annotations. **(C)** The distribution of patients with different amounts of potentially significant variants. **(D)** Sequencing depth of all detected exonic and splicing variants. **(E)** Average sequencing depth of each gene. **(F)** Sequencing depth distribution of variants on each gene. **(G)** Sequencing depth distribution of each patient. **(H)** Annotation distributions of all non-synonymous and non-benign variants on each gene. **(I)** Distributions of all non-synonymous and non-benign variants on each gene from three different reproductive centers. CZ means Heping Hospital affiliated to Changzhi Medical College. HA means the affiliated Huaian No. 1 People's Hospital of Nanjing Medical University, and EFY means Reproductive Medical Center of the Second Affiliated Hospital of Nanjing Medical University. **(J)** Normalized variants distribution on each gene.

them non-annotated (**Figure 2B**). We excluded all the synonymous and benign variants, and we focused on those non-synonymous and non-benign variants which we referred to as "potentially significant" variants. In 68 patients, 15 patients had no potentially significant variants, and 19 patients had only a single potentially significant variant. Moreover, 6 patients had multiple variants on a single gene, while 28 patients had multiple variants on multiple genes (**Figure 2C**). The sequencing depth of all splicing and exonic variants showed that 74.0% of them had reached the depth of over 1000× (**Figure 2D**). The gene with the

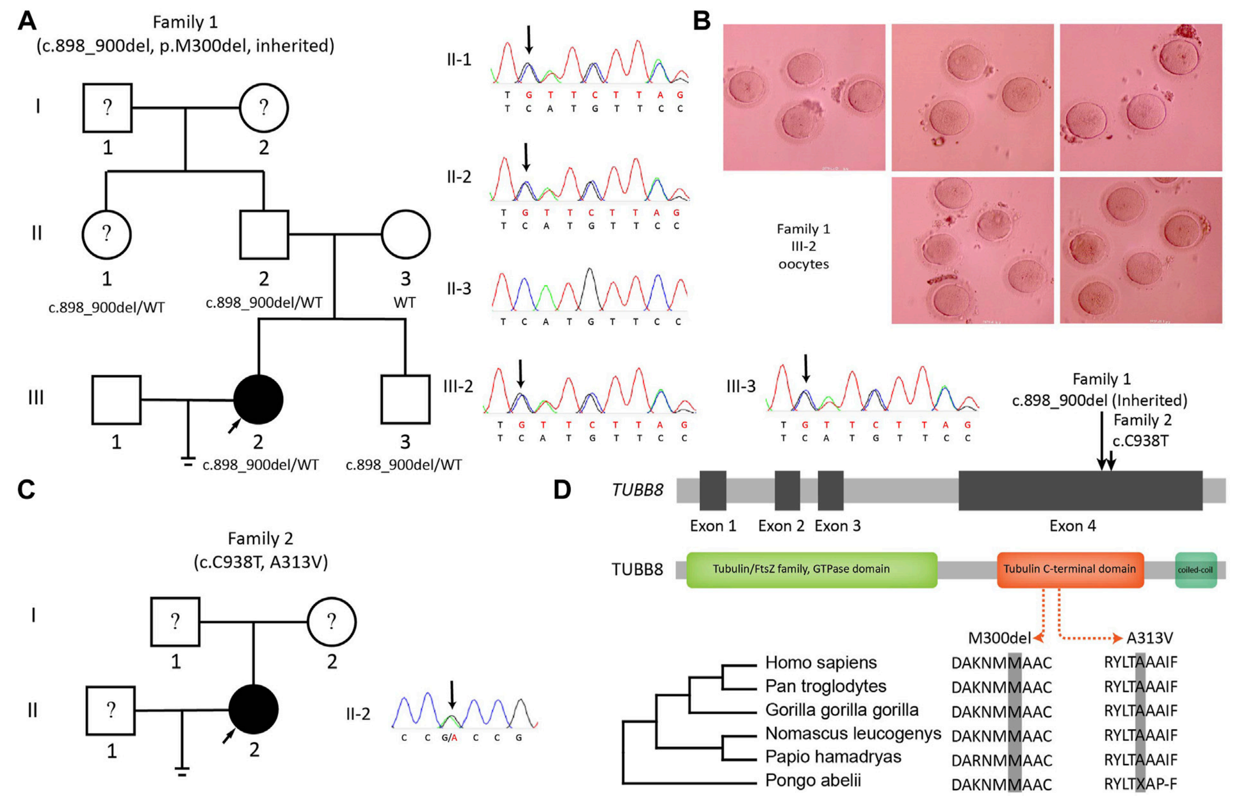


FIGURE 3 | Variants on TUBB8 and its conservation. **(A)** The pedigree and Sanger sequencing of Family 1. **(B)** Oocytes from III-2 (Family 1) arrested in MI. **(C)** The pedigree and Sanger sequencing of Family 2. **(D)** Conservations of variant sites of TUBB8.

highest average depth was FOXP3, which was above 2500×. PATL2 had the lowest average depth, which was around 500× (**Figure 2E**). REC114 and TRIP13 had no detected exonic and splicing variants. Variants sequencing depth distribution displayed that most variants were sequenced between 1000× and 2000×. PATL2 had about 60% variants sequenced no more than 500×, while some variants from TUBB8 were sequenced deeper than 5000× (**Figure 2F**). **Figure 2G** showing the depth distribution of each patient demonstrated that most sequencing depth was between 1000× and 2000×.

Based on clinical annotation, TUBB8 had the highest percentage of non-annotated and uncertain variants, and DNAH11 was the highest in the likely benign variants (**Figure 2H**). Data from three different reproductive centers showed that TUBB8 and DNAH11 were the top two most frequently variant genes (**Figure 2I**). Mutant frequencies of each gene were changed when we normalized the counts of variants in gene length. However, TUBB8 was still the most frequently variant gene (**Figure 2J**). Our target gene panel sequencing exhibited good quality in sequencing depth and coverage, though optimization is in need for better performance.

Sanger Sequencing Validation

Among the potentially significant variants we detected, when FRE (the proportion of variants bases in all detected bases) was above 0.4, the results of panel sequencing were 100% validated. We noticed that variants with FRE around 0.2 could not be validated (**Supplementary Figure S1**). This is probably due to the sensitivity limitation of Sanger sequencing. When FRE was low, PCR amplification bias made Sanger sequencing hard to detect low-frequency variants. We also found another frequent variant TUBB8 (c.167-169del:p.56-57del), which was reported in nine patients and could not be validated. It may be because this region captured by the probes is highly similar to other β-tubulin isotype genes or some duplicated fragment in the human genome such as chr18:14414. Thus, an alignment error may have occurred, and false-positive was reported.

Two Variants on TUBB8 in Two Patients

Two unrelated patients were found bearing variants on TUBB8. Patient III-2 in Family 1 was 30 years old and had been trying to conceive for 5 years. She had experienced two failed IVF cycles with 8 and 19 immature oocytes. One potentially significant variant on TUBB8 (c.898_900del:p.300_300del) was found. Sequencing of family members showed the paternal transmission of this heterozygous missense variant (**Figure 3A**). Oocytes typically arrested at MI stage are shown in **Figure 3B**. Patient II-2 in Family 2 was 29 years old and was diagnosed as RPL with two idiopathic spontaneous abortion (**Figure 3C**): a previously reported TUBB8 variant c.C938T: p.A313V with frequency of 0.007% in ExAC (Exome Aggregation Consortium, allele count: 8, allele number:

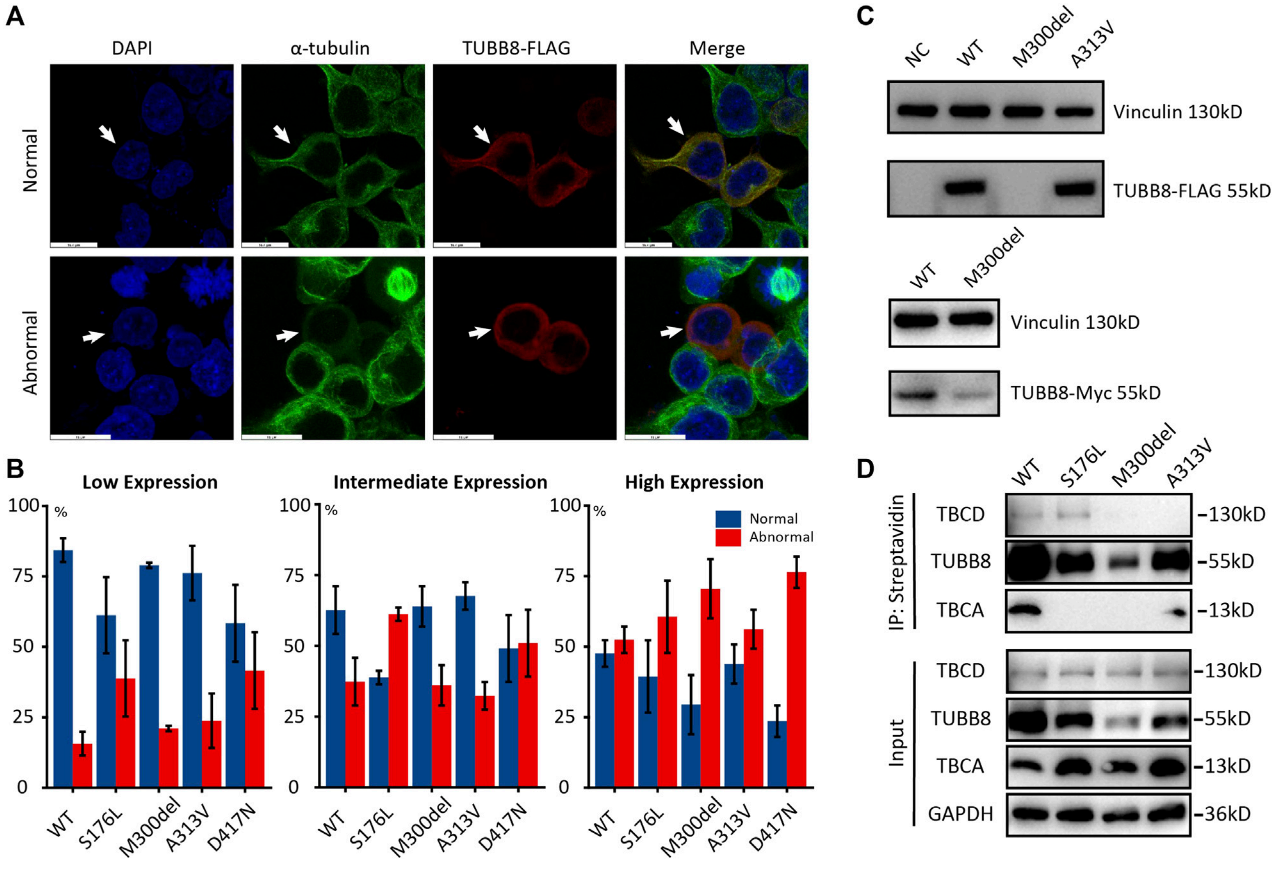


FIGURE 4 | *In vitro* disruption of microtubule network by WT and mutant TUBB8. **(A)** Microtubule phenotypes resulting from expression of wild-type and mutated TUBB8-flag transfected in HEK293T. α-tubulin (green) showed the endogenous microtubule network, and M2-FLAG (red) detected the expression of the transgene. Microtubules assembled into a regular network in the normal group, while the abnormal group showed a mottled pattern. The bar in normal pattern indicates 15.3 μm, and the bar in the abnormal pattern indicates 15 μm. **(B)** The quantitative analysis of microtubule phenotypes in panel. Approximately 200 transfected cells were examined. I bar indicates standard error. **(C)** Western blot of overexpressions of wild-type and mutant TUBB8. NC was a blank load transfection negative control. WT was overexpressions of wild-type TUBB8 in HEK293T. M300del showed no visible protein band when detected with anti-flag antibody, while WT and A313V had comparable expression levels. When repeated with anti-Myc antibody, M300del still had a blurry band. **(D)** Co-IP of TUBB8 and tubulin cofactors. WT and A313V kept interacting with both TBCD and TBCA, and S176L lost binding with TBCA but maintained interaction with TBCD. M300del had no interaction with either TBCD or TBCA.

116668) and 0.05% in gnomAD (Genome Aggregation Database, allele count: 132, allele number: 273892). The position and conservation of the two SNVs are shown in **Figure 3D**.

Disruption of Microtubule Network by Mutant TUBB8

We next transfected WT and mutant TUBB8 vectors into cultured HEK293T cells to investigate the influence of variants on microtubule network. Two different morphologies were observed. On one hand, the normal cells had well-balanced α- and β-tubulin distributions which could assemble into a regular microtubule network. The endogenous α-tubulin (green) and overexpressed TUBB8 (red) showed a similar network co-assembling into a morphologically normal cell. On the other hand, abnormal cells lost regular microtubule network showing diffused distribution of α-tubulin and TUBB8 throughout the cytoplasm (**Figure 4A**). We observed the microtubule network with low, intermediate, and high transfection levels. Apart from the SNVs detected by our target sequencing M300del and A313V, we also compared two previously reported variants S176L and D417N. Under low expression, about 84.1% cells were morphologically normal in the WT group, and mutant groups were lower than WT: S176L was 59.87%, M300del was 78.61%, A313V was 77.13%, and D417N was 55.28%. With intermediate expressions, percentages of normal cells decreased, as WT dropped to 58.83%, and so did those of the mutant groups: S176L was 38.1%, M300del was 61.60%, A313V was 68.0%, and D417N was 49.08%. Under high expression levels, normal cells only took up 46.92% in the WT group, and S176L was 29.55%, M300del was 20.44%, while A313V was 41.99%, and D417N was 20.40% (**Figure 4B** and **Supplementary Table S2**). The different disrupted tendencies in these four mutant groups might indicate distinguished mechanisms. S176L, M300del, and A313V were predicted to impact dimer stability, β-tubulin folding, or polymerization, while D417N was predicted to disrupt

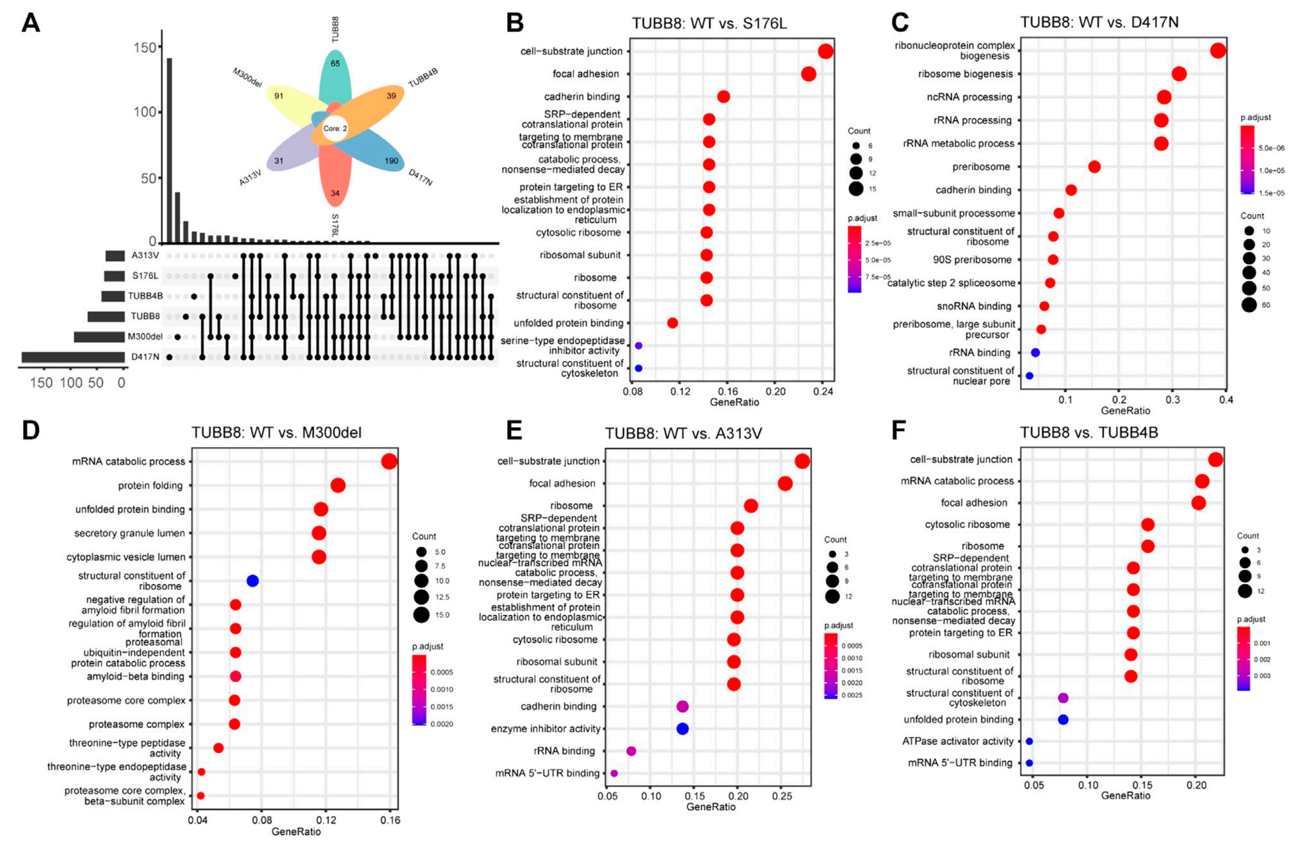


FIGURE 5 | GO enrichment analysis of WT and mutant TUBB8. **(A)** Amounts of unique and common proteins between each group. **(B)** GO enrichment analysis of differently expressed proteins between TUBB8 WT and S176L. **(C)** GO enrichment analysis of differently expressed proteins between TUBB8 WT and D417N. **(D)** GO enrichment analysis of differently expressed proteins between TUBB8 WT and M300del. **(E)** GO enrichment analysis of differently expressed proteins between TUBB8 WT and A313V. **(F)** GO enrichment analysis of differently expressed proteins between TUBB8 and TUBB4B.

microtubule network by influencing kinesin binding (Feng et al., 2016).

Western blot showed that M300del variant had an extremely decreased protein amount with no visible band, which could be a result of faster protein degradation by this variant. A313V had comparable expression levels with WT. We repeated western blot with a different tag and antibody, and M300del still showed a dramatic lowered expression, which implied the M300del variant resulted in a severely impaired protein function following a loss-of-function mechanism (**Figure 4C**).

Interactions Between TUBB8 and Tubulin Co-Factors

To identify the protein interactome of TUBB8, we transfected WT/mutant TUBB8 and TUBB4B vectors in HEK293T cells and purified proteins with M280 Dynabeads. Mass spectrometry was performed. Comparisons were made to investigate the specific interacting molecules of TUBB8 from a different β-tubulin isotype TUBB4B, so did WT vs mutant TUBB8 (S176L, M300del, A313V and D417N). Totally 1655 proteins were detected in seven groups including NC, which was a blank load transfection negative control. Differently expressed proteins were clustered using GO enrichment analysis. **Figure 5A** shows the uniquely expressed proteins of each group when compared with NC. The enrichment analysis of WT vs mutant TUBB8 is shown in **Figures 5B–E**. Comparing TUBB8 with TUBB4B, the main differently expressed proteins were enriched in the cell–substrate junction, mRNA catabolic process, and focal adhesion (**Figure 5F**). TUBB8 specifically expresses in human oocyte and early embryo while TUBB4B shows low tissue specificity. Therefore, this might lead to the different protein interactome, which is the basis of different protein functions. It is noted that M300del had been enriched to the ubiquitin-independent protein catabolic process (**Figure 5D**). Since the M300del variant showed a decreased protein expression in WB, we assumed that M300del activated a protein degradation pathway independent from ubiquitination which needs to be further investigated.

We further focused on how mutant TUBB8 disrupted microtubule network. According to previous studies, tubulin cofactors were vital in the process of tubulin heterodimer folding. During biogenesis and degradation of the α/β heterodimer, tubulin cofactors function synergistically on the folding of α- or β-tubulin monomers (Nithianantham et al., 2015). Co-immunoprecipitation showed that apart from S176L with increased expression of TBCD, other mutant TUBB8 did not change the expressions of TBCD or TBCA. M300del and A313V totally lost binding with TBCD, unlike WT or S176L. The absence of interaction between M300del/A313V and TBCD might destroy transport of the tubulin monomer and disrupt microtubule assembly. Interactions between TUBB8 and TBCA showed a different result. S176L and M300del had no interaction with TBCA, but WT/A313V still withheld the interplay (**Figure 4D**), assuming two varying mechanisms underlying the differently mutant TUBB8. TBCA mainly received β-tubulin from heterodimers and was critical in tubulin recycling (Nolasco et al., 2021). TBCD formed a trimer with regulatory GTPase ARL2 and β-tubulin (Francis et al., 2017). A previous report investigated S176L disrupting microtubule assembly by interfering GTP hydrolysis (Feng et al., 2016), so it could be speculated that the S176L variant was passed on by TBCD but failed to form heterodimers due to interfering GTP hydrolysis. M300del had no interaction with either TBCD or TBCA, implying its total failure in forming heterodimers, as M300del could not bind with tubulin cofactors at the first step. A313V could interplay with TBCA but not with TBCD. The possible reason was that M300del and A313V were buried within the protein structure, thus tubulin cofactors could not efficiently bind to M300del or A313V-bearing proteins, resulting in the destabilization of tubulin folding.

DISCUSSION

Despite the long-time use of genetic screening on other diseases, the practice of it on female infertility is still at the flagging stage. As pathogenic genes of female infertility have just been revealed in recent years, it is of great significance to develop an efficient and cost-effective screening method, especially for numerous idiopathic PI/RPL patients. To achieve this goal, we designed a target gene panel to sequence 22 known female infertility-related genes. Samples from 68 PI/RPL patients were sequenced with an average depth of 1559×. A sum of 3,134 exonic and splicing variants was detected, and potentially significant variants were further investigated. Based on this target sequencing, two TUBB8 variants were found including a novel one. We demonstrated mutant TUBB8 protein disrupting microtubule network and the loss-of-function role of M300del variant. We also revealed the interactome of WT and mutant TUBB8 and validated their different interaction patterns with TBCD and TBCA.

Panel sequencing enables specific in-depth detection of target genes. Its lower time-consumption, lower cost, and high accuracy make it a cost-effective tool for clinical application. Not only could our panel provide accurate diagnosis for patients but also could open a new window for the discovery of new pathogenic variants. For better performance, our panel needs optimization. First, the average depth of genes was unbalanced. FOXP3 had reached a mean depth of 2500×, while PATL2 did not achieve more than 1000×. The uneven depth of each gene might be a consequence of the innate characteristics of gene sequences. When genes like PATL2 have high abundance of GC or duplicated regions, capture efficiencies of specific probes might be affected. To improve the sequencing depth, extra or longer probes could be applied. Another concern is the false-positive reports in a particular position. When a target region was highly homologous with unwanted areas on the genome, it could be possible to enrich the unwanted regions or map the reads to the wrong place, thus leading to false-positive reports. One possible solution is longer probes or performing specific amplifications to increase the specificity of hybridization capture. Moreover, the current mean depth of our panel was 1559×, which might be excessive for genetic disorders (LaDuca et al., 2017). Usually 30× to 40× genome-sequencing would be adequate to give a confident sequence performance (Telenti et al., 2016), while the current clinical exome-sequencing uses 120× as a standard (Kong et al., 2018). The otoscope hearing loss panel achieved 716× per patient (Sloan-Heggen and Smith, 2016). Next, we will optimize our panel to have a balanced and reasonable sequencing depth for better performance and reduce the cost.

Some variants that we detected had no certain annotation in ACMG and InterVar. It might be that these variants are novel, or time was required for the database update. The newly published NyuWa Genome Resource providing the variation profile for Chinese population does not contain annotation of TUBB8: M300del or A313V. We also searched NHLBI Exome Sequencing Project, genomAD, human genetic variation database, ClinVar, and Online Mendelian Inheritance in Man (OMIM). We found that M300del had no annotation in these databases, and A313V was predicted to be likely pathogenic in ClinVar. The novel variant M300del expanded the spectrum of TUBB8 pathogenicity, and we confirm the pathogenic role of the two variants detected by our panel. To sum up, our target sequencing gene panel is unique in implementing to find novel and rare variants.

We also revealed that tubulin cofactors lost their binding to mutant TUBB8 in the process of heterodimer folding. TBCD passes on β-tubulin from TBCA and then transports them to the transient super-complex of TBCD/TBCE/TBCC trimer and α-tubulin (Nithianantham et al., 2015). However, M300del and A313V lost binding with TBCD, indicating the failed transport of β-tubulin monomer, which might lead to microtubule disruption. On the other hand, TBCA lost binding with S176L and M300del, indicating that mutants located in different domains of TUBB8 disturb different steps of heterodimer formation. A study on microtubule reconstruction and microtubule dynamics might help reveal further mechanisms. The interactome of M300del enriched in ubiquitin-independent protein catabolic process might imply that M300del proteins were eliminated through uncanonical pathways in an accelerated manner. For further study, more differently expressed proteins from mass spectrum would be validated, and downstream pathways and potential target molecules would be identified. It is promising to uncover deeper underlying mechanisms and bring new treatment strategies for variant-bearing patients. Finally, for the loss-of-function M300del, we are trying to conduct clinical trials of the WT TUBB8 supplement by microinjection of WT TUBB8 mRNA into the oocyte similar to previously reported (Sang et al., 2018).

The limitation of our work is the sample size. We are recruiting more patients and conducting optimization of the current sequencing panel. Moreover, we will add newly found female infertility pathogenic genes such as MEI1 (Dong et al., 2021) and FBXO43 (Weijie Wang et al., 2021). In conclusion, we designed a target-sequencing gene panel of female infertility-related genes and tested it in 68 patients. Two variants of TUBB8 were found, as well as a novel SNV M300del, and the primary loss-of-function mechanism of this novel variant was demonstrated. Further study will be aiming at both the optimization of this sequencing panel and the deeper understanding of TUBB8 pathology. We hope this target sequencing gene panel would be an efficient and cost-effective tool for genetic screen in reproductive centers to achieve a better diagnosis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ngdc.cncb.ac.cn/; [HRA002192].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HY performed the main experiments and wrote the manuscript, and JC analyzed the statistics. NL, HM, YC, SL, and YQ helped collected samples in reproductive centers. GY, HL, LC, FM, LH, YH, and SH contributed in cell culture and plasmid construction. CM, YQ, and RF designed the project and supervised the program. RF revised the manuscript.

FUNDING

This research was financially supported by grants from the National Natural Science Foundation of China (81971451, 31900605), Nature Science Foundation of Jiangsu Province (BK20190654), Innovative and Entrepreneurial Team of Jiangsu Province (JSSCTD202144), and Innovative and Entrepreneurial Talent Program of Jiangsu Province.

ACKNOWLEDGMENTS

We are sincerely grateful of Lei Wang and Qing Sang in Fudan University for kindly providing plasmids used in this study. We also appreciate the selfless and professional advice by Qiulun Lu and Da Huo from Nanjing Medical University School of Pharmacy as well as Dong Zhang and Jing Li from the State Key Laboratory of Reproductive Medicine, Nanjing Medicine University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.865103/full#supplementary-material

REFERENCES

An, M., Liu, Y., Zhang, M., Hu, K., Jin, Y., Xu, S., et al. (2021). Targeted Next-Generation Sequencing Panel Screening of 668 Chinese Patients with Non-obstructive Azoospermia. *J. Assist. Reprod. Genet.* 38 (8), 1997–2005. doi:10.1007/s10815-021-02154-9

Beke, A. (2019). “Genetic Causes of Female Infertility,”. *Genetics of Endocrine Diseases and Syndromes*. Editors P. Igaz and A. Patócs (Cham: Springer International Publishing), 111, 367–383. doi:10.1007/978-3-030-25905-1_17

Chen, B., Zhang, Z., Sun, X., Kuang, Y., Mao, X., Wang, X., et al. (2017). Biallelic Mutations in PATL2 Cause Female Infertility Characterized by Oocyte Maturation Arrest. *Am. J. Hum. Genet.* 101 (4), 609–615. doi:10.1016/j.ajhg.2017.08.018

Chen, B., Li, B., Li, D., Yan, Z., Mao, X., Xu, Y., et al. (2017). Novel Mutations and Structural Deletions in *TUBB8* : Expanding Mutational and Phenotypic Spectrum of Patients with Arrest in Oocyte Maturation, Fertilization or Early Embryonic Development. *Hum. Reprod.* 32 (2), 457–464. doi:10.1093/humrep/dew322

Ching-Yuan Wang, C.-Y., Tang, Y.-A., Lee, I.-W., Chang, F.-M., Chien, C.-W., Pan, H.-A., et al. (2021). Development and Validation of an Expanded Targeted Sequencing Panel for Non-invasive Prenatal Diagnosis of Sporadic Skeletal Dysplasia. *BMC Med. Genomics* 14 (S3), 212. doi:10.1186/s12920-021-01063-1

Chiware, T. M., Vermeulen, N., Blondeel, K., Farquharson, R., Kiarie, J., Lundin, K., et al. (2021). IVF and Other ART in Low- and Middle-Income Countries: a Systematic Landscape Analysis. *Hum. Reprod. Update* 27 (2), 213–228. doi:10.1093/humupd/dmaa047

Dong, J., Zhang, H., Mao, X., Zhu, J., Li, D., Fu, J., et al. (2021). Novel Biallelic Mutations in *MEI1:* Expanding the Phenotypic Spectrum to Human Embryonic Arrest and Recurrent Implantation Failure. *Hum. Reprod.* 36 (8), 2371–2381. doi:10.1093/humrep/deab118

Eggers, S., Sadedin, S., van den Bergen, J. A., Robevska, G., Ohnesorg, T., Hewitt, J., et al. (2016). Disorders of Sex Development: Insights from Targeted Gene Sequencing of a Large International Patient Cohort. *Genome Biol.* 17 (1), 243. doi:10.1186/s13059-016-1105-y

Feng, R., Sang, Q., Kuang, Y., Sun, X., Yan, Z., Zhang, S., et al. (2016). Mutations in *TUBB8* and Human Oocyte Meiotic Arrest. *N. Engl. J. Med.* 374 (3), 223–232. doi:10.1056/NEJMoa1510791

Francis, J. W., Newman, L. E., Cunningham, L. A., and Kahn, R. A. (2017). A Trimer Consisting of the Tubulin-specific Chaperone D (TBCD), Regulatory GTPase ARL2, and β-Tubulin Is Required for Maintaining the Microtubule Network. *J. Biol. Chem.* 292 (10), 4336–4349. doi:10.1074/jbc.M116.770909

Furuta, T., Tuck, S., Kirchner, J., Koch, B., Auty, R., Kitagawa, R., et al. (2000). EMB-30: An APC4 Homologue Required for Metaphase-To-Anaphase Transitions during Meiosis and Mitosis inCaenorhabditis Elegans. *Mol. Biol. Cel* 11 (4), 1401–1419. doi:10.1091/mbc.11.4.1401

Gordon, A. S., Fulton, R. S., Qin, X., Mardis, E. R., Nickerson, D. A., and Scherer, S. (2016). PGRNseq. *Pharmacogenet. Genomics* 26 (4), 161–168. doi:10.1097/FPC.0000000000000202

H. Maiato (Editor) (2020). *Cytoskeleton Dynamics: Methods and Potocols* (New York, NY: Humana Press).

Kong, S. W., Lee, I.-H., Liu, X., Hirschhorn, J. N., and Mandl, K. D. (2018). Measuring Coverage and Accuracy of Whole-Exome Sequencing in Clinical Context. *Genet. Med.* 20 (12), 1617–1626. doi:10.1038/gim.2018.51

LaDuca, H., Farwell, K. D., Vuong, H., Lu, H. M., Mu, W., Shahmirzadi, L., et al. (2017). Exome Sequencing Covers >98% of Mutations Identified on Targeted Next Generation Sequencing Panels. *PLOS ONE* 12 (2), e0170843. doi:10.1371/journal.pone.0170843

Liu, J. L., Peña, V., Fletcher, S. A., and Kohn, T. P. (2020). Genetic Testing in Male Infertility - Reassessing Screening Thresholds. *Curr. Opin. Urol.* 30 (3), 317–323. doi:10.1097/MOU.0000000000000764

Matsunaga, T. (2009). Value of Genetic Testing in the Otological Approach for Sensorineural Hearing Loss. *Keio J. Med.* 58 (4), 216–222. doi:10.2302/kjm.58.216

Miyamoto, T., Minase, G., Shin, T., Ueda, H., Okada, H., and Sengoku, K. (2017). Human Male Infertility and its Genetic Causes. *Reprod. Med. Biol.* 16 (2), 81–88. doi:10.1002/rmb2.12017

Mu, J., Zhang, Z., Wu, L., Fu, J., Chen, B., Yan, Z., et al. (2020). The Identification of Novel Mutations in PLCZ1 Responsible for Human Fertilization Failure and a Therapeutic Intervention by Artificial Oocyte Activation. *Mol. Hum. Reprod.* 26 (2), 80–87. doi:10.1093/molehr/gaaa003

Nithianantham, S., Le, S., Seto, E., Jia, W., Leary, J., Corbett, K. D., et al. (2015). Tubulin Cofactors and Arl2 Are Cage-like Chaperones that Regulate the Soluble αβ-tubulin Pool for Microtubule Dynamics. *eLife* 4, e08811. doi:10.7554/eLife.08811

Nolasco, S., Bellido, J., Serna, M., Carmona, B., Soares, H., and Zabala, J. C. (2021). Colchicine Blocks Tubulin Heterodimer Recycling by Tubulin Cofactors TBCA, TBCB, and TBCE. *Front. Cel Dev. Biol.* 9, 656273. doi:10.3389/fcell.2021.656273

Pös, O., Budiš, J., and Szemes, T. (2019). Recent Trends in Prenatal Genetic Screening and Testing. *F1000Research* 8, 764. doi:10.12688/f1000research.16837.1

Sang, Q., Li, B., Kuang, Y., Wang, X., Zhang, Z., Chen, B., et al. (2018). Homozygous Mutations in WEE2 Cause Fertilization Failure and Female Infertility. *Am. J. Hum. Genet.* 102 (4), 649–657. doi:10.1016/j.ajhg.2018.02.015

Sang, Q., Zhou, Z., Mu, J., and Wang, L. (2021). Genetic Factors as Potential Molecular Markers of Human Oocyte and Embryo Quality. *J. Assist. Reprod. Genet.* 38 (5), 993–1002. doi:10.1007/s10815-021-02196-z

Sloan-Heggen, C. M., and Smith, R. J. H. (2016). Navigating Genetic Diagnostics in Patients with Hearing Loss. *Curr. Opin. Pediatr.* 28 (6), 705–712. doi:10.1097/MOP.0000000000000410

Takenaka, M., Saito, M., Iwakawa, R., Yanaihara, N., Saito, M., Kato, M., et al. (2015). Profiling of Actionable Gene Alterations in Ovarian Cancer by Targeted Deep Sequencing. *Int. J. Oncol.* 46 (6), 2389–2398. doi:10.3892/ijo.2015.2951

Telenti, A., Pierce, L. C. T., Biggs, W. H., di Iulio, J., Wong, E. H. M., Fabani, M. M., et al. (2016). Deep Sequencing of 10,000 Human Genomes. *Proc. Natl. Acad. Sci. U.S.A.* 113 (42), 11901–11906. doi:10.1073/pnas.1613365113

Wang, X., Park, K.-E., Koser, S., Liu, S., Magnani, L., and Cabot, R. A. (2012). KPNA7, an Oocyte- and Embryo-specific Karyopherin?subtype, Is Required for Porcine Embryo Development. *Reprod. Fertil. Dev.* 24 (2), 382. doi:10.1071/RD11119

Weijie Wang, W., Wang, W., Xu, Y., Shi, J., Fu, J., Chen, B., et al. (2021). *FBXO43* Variants in Patients with Female Infertility Characterized by Early Embryonic Arrest. *Hum. Reprod.* 36 (8), 2392–2402. doi:10.1093/humrep/deab131

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data. *The Innovation* 2 (3), 100141. doi:10.1016/j.xinn.2021.100141

Zhang, Z., Li, B., Fu, J., Li, R., Diao, F., Li, C., et al. (2020). Bi-allelic Missense Pathogenic Variants in TRIP13 Cause Female Infertility Characterized by Oocyte Maturation Arrest. *Am. J. Hum. Genet.* 107 (1), 15–23. doi:10.1016/j.ajhg.2020.05.001

Zhang, Z., Wu, L., Diao, F., Chen, B., Fu, J., Mao, X., et al. (2020). Novel Mutations in LHCGR (Luteinizing Hormone/choriogonadotropin Receptor): Expanding the Spectrum of Mutations Responsible for Human Empty Follicle Syndrome. *J. Assist. Reprod. Genet.* 37 (11), 2861–2868. doi:10.1007/s10815-020-01931-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

CASE REPORT
published: 01 June 2022
doi: 10.3389/fgene.2022.926060

Case Report: Preimplantation Genetic Testing for X-Linked Severe Combined Immune Deficiency Caused by *IL2RG* Gene Variant

Jun Ren [1,2,3], ***Cuiting Peng*** [1,2,3], ***Fan Zhou*** [1,2,3], ***Yutong Li*** [1,2,3], ***Yuezhi Keqie*** [1,2,3], ***Han Chen*** [1,2,3], ***Hongmei Zhu*** [1,2,3], ***Xinlian Chen*** [1,2,3*†] **and** ***Shanling Liu*** [1,2,3*]

[1]Department of Medical Genetics, Center of Prenatal Diagnosis, West China Second University Hospital, Sichuan University, Chengdu, China, [2]Department of Obstetrics and Gynecology, West China Second University Hospital, Sichuan University, Chengdu, China, [3]Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China

Edited by:
Xi Wang,
Nanjing Medical University, China

Reviewed by:
Junyu Zhang,
Shanghai Jiao Tong University School of Medicine, China
Andrea Grioni,
Novartis Institutes for BioMedical Research, Switzerland

****Correspondence:***
Xinlian Chen
chenxinlian1121@163.com
Shanling Liu
sunny630@126.com

†These authors have contributed equally to this work and share last authorship

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *22 April 2022*
Accepted: *16 May 2022*
Published: *01 June 2022*

Citation:
Ren J, Peng C, Zhou F, Li Y, Keqie Y, Chen H, Zhu H, Chen X and Liu S (2022) Case Report: Preimplantation Genetic Testing for X-Linked Severe Combined Immune Deficiency Caused by IL2RG Gene Variant.
Front. Genet. 13:926060.
doi: 10.3389/fgene.2022.926060

Preimplantation genetic testing (PGT) has been increasingly used to prevent rare inherited diseases. In this study, we report a case where PGT was used to prevent the transmission of disease-caused variant in a SCID-X1 (OMIM:300400) family. SCID-X1 is an X-linked recessive inherited disease whose major clinical manifestation of immune deficiency is the significant reduction in the number of T-cells and natural killer cells. This family gave birth to a boy who was a hemizygous proband whose *IL2RG* gene was mutated (c.315T > A, p(Tyr105*), NM_000206.3, CM962677). In this case, Sanger sequencing for mutated allele and linkage analysis based on single-nucleotide polymorphism (SNP) haplotype via next-generation sequencing were performed simultaneously. After PGT for monogenic disorder, we detected the aneuploidy and copy number variation (CNV) for normal and female carrier embryos. Four embryos (E02, E09, E10, and E11) were confirmed without CNVs and inherited variants at the *IL2RG* gene. Embryo E02 (ranking 4BB) has been transferred after considering the embryo growth rate, morphology, and PGT results. Prenatal genetic diagnosis was used to detect amniotic fluid cells, showing that this fetus did not carry the variant of the *IL2RG* gene (c.315T > A). Ultimately, a healthy girl who had not carried disease-causing variants of SCID-X1 confirmed by prenatal diagnosis was born, further verifying our successful application of PGT in preventing mutated allele transmission for this SCID family.

Keywords: preimplantation genetic testing, severe combined immune deficiency, IR2RG, haplotype, rare genetic disease, next generation sequencing

INTRODUCTION

Severe combined immune deficiency (SCID, OMIM:300400) is a fatal genetic defect (Torii, 1996). Although SCID morbidity varies from country to country, there is no report on its exact incidence in China. According to research on newborn screening in the United States, SCID affects one in 58,000 infants (95% CI, 1/46000–1/80000) (Kwan et al., 2014). SCID is a prenatal disorder of T lymphocyte development (Cossu, 2010), in which the affected infant gradually develops a pediatric emergency after birth. In general, the affected infant presents severe opportunistic infections within 1 month of birth because of defects in humoral and cellular immunity (Mamcarz et al., 2019). Typical laboratory inspections show a lack of T cells, natural killer (NK) cells, and functional B cells. As a result, affected infants cannot usually live beyond their first year of life.

The most common cause of SCID is a variant of the *IL2RG* gene (OMIM: 300400), also called SCID-X1. The *IL2RG* gene, situated in Xq13.1, encodes the interleukin-2 receptor common gamma chain and is shared by several cytokine receptors necessary for the development and function of lymphocytes (Mamcarz et al., 2019). Allogeneic hematopoietic stem-cell transplantation or autologous gene therapy is considered the most effective treatment for SCID (Cavazzana-Calvo et al., 2000). Although hematopoietic stem cell transplantation (HSCT) from a matched sibling donor is effective, it can only be used in a minority of patients, as transplantation from an alternative donor is related to an increased risk of graft-versus-host disease and incomplete immune reconstitution (Mamcarz et al., 2019). Gene therapy has also shown potential in this regard, but the carcinogenicity of retroviral vectors remains to be solved. Leukemia was caused when gene therapy was first used to treat SCID-X1, owing to an insertional variant induced by enhancers of the adenovirus vectors (Fischer and Hacein-Bey-Abina, 2020). With the development of gene therapy vectors, such as lentivirus, increasingly exciting research has been carried out to prove their application value in SCID therapy (Blanco et al., 2020).

However, for these couples with an inherited disease that can be clearly diagnosed genetically, adequate genetic counseling and prenatal or preimplantation genetic diagnostic techniques are important (De Rycke and Berckmoes, 2020, Group et al., 2020). With the development of assisted reproduction and molecular genetic technology, preimplantation genetic testing (PGT) has been used in birth defect prevention and control (Group et al., 2020). PGT for monogenic disorders (PGT-M) can select unaffected embryos to prevent the transmission of disease-causing variants. PGT technology avoids the adverse effects of repeated abortion on women's physical and mental health.

Through this case, we report a PGT-M case based on MALBAC and next-generation sequencing (NGS)-based single-nucleotide polymorphism (SNP) haplotype for SCID. To further verify our PGT results, chromosome microarray (CMA) for copy number variation (CNV) analysis and Sanger sequencing for mutated alleles were conducted in amniotic cells for prenatal genetic diagnosis.

PATIENTS AND METHODS

Patients

A 31-year-old couple had given birth to a boy who was the proband. The female carrying *IL2RG* mutated allele visited the Department of Medical Genetics, West China Second University Hospital, Sichuan University. The boy was diagnosed with X-linked SCID (OMIM: 300400) in the West China Second University Hospital for repeated high fever, severe anemia, hepatosplenomegaly, immune system deficiency, coagulation dysfunction, hemangioma, and severe sepsis. The boy died at less than 1 year of age. We found that the boy had a variant in the *IL2RG* gene (c.315T > A, p(Tyr105*), NM_000206.3, CM962677). According to the standards and guidelines for sequence variant interpretation of the American College of Medical Genetics and Genomics (ACMG/AMP) (Richards et al., 2015), Clinical Interpretation of Sequence Variants (zhang et al., 2020) and ClinGen Sequence Variant Interpretation Recommendation for PM2 - Version 1.0, we evaluated this variant with PVS1 + PM2_Supporting, likely pathogenic variant. Parents received genetic counseling and signed an informed consent form. Sample collection, library preparation, NGS and data analysis were conducted at the Department of Medical Genetics, West China Second University Hospital. This study was approved by the Internal Ethical Committees of the West China Second University Hospital.

Assisted Reproductive Technology Procedure and Embryo Trophectoderm Biopsy

Controlled ovarian stimulation, intracytoplasmic sperm injection (ICSI), blastocyst culture, trophectoderm biopsy, and blastocyst transfer were conducted in the Reproductive Medicine Center of West China Second University Hospital, according to the standard protocol (McArthur et al., 2005; Schoolcraft et al., 2011). In this cycle, twelve embryos were finally developed into blastocysts, and the trophectoderm (TE) cells were biopsied on day 5 or 6 after insemination. A total of five to eight biopsied cells from the TE were transferred into 4.5 μL lysis buffer (Yikon Genomics) in 0.2 μL PCR tubes for whole-genome amplification (WGA) (MALBAC).

gDNA Extraction and WGA

Peripheral blood of the couple was collected for gDNA extraction by a DNeasy Blood and Tissue Kit (Qiagen). Although the proband (hemizygote) was deceased, the couple kept his gDNA sample. In the second trimester, amniotic cells were collected by transabdominal amniocentesis; gDNA extraction was also carried out using a DNeasy Blood and Tissue Kit (Qiagen). The buccal mucosa cells (BMCs) of the woman (variant carrier) were diluted by phosphate buffer saline; five to eight cells were selected for WGA. WGA involves multiple annealing and looping-based amplification cycles (MALBAC, Yikon genomics) applied for diluted mucosa and TE cells. Additionally, 5 μL WGA products were used for 1% agarose gel electrophoresis; a 300–2000 bp diffused band indicated successful amplification. Whole-genome products were purified using a DNA Clean-up Kit (CWBIO). All operations were performed according to manufacturers' protocols.

Variant Site Detection

PCR amplification and Sanger sequencing were conducted to validate the mutated *IL2RG* allele. Primer was designed to amplify the segment containing c.315 of *IL2RG*. Forward primer (CTCCCAG GTACCCCACTGTT) and reverse primer (TCC AATGTCCCACAGTATCCC) were designed (Primer 5.0 software) and synthesized (TsingKe Biotechnology, Beijing). PCR was performed in a 25 μL system using 2× GoldStar Best Master Mix (CWBIO) on a 96-Well Thermal Cycler Veriti Dx (Life Technologies). The amplification system contained 2 μL of primer mix, 12.5 μL of enzyme mix, gDNA, or purified WGA products as templates. The reaction condition was as follows:

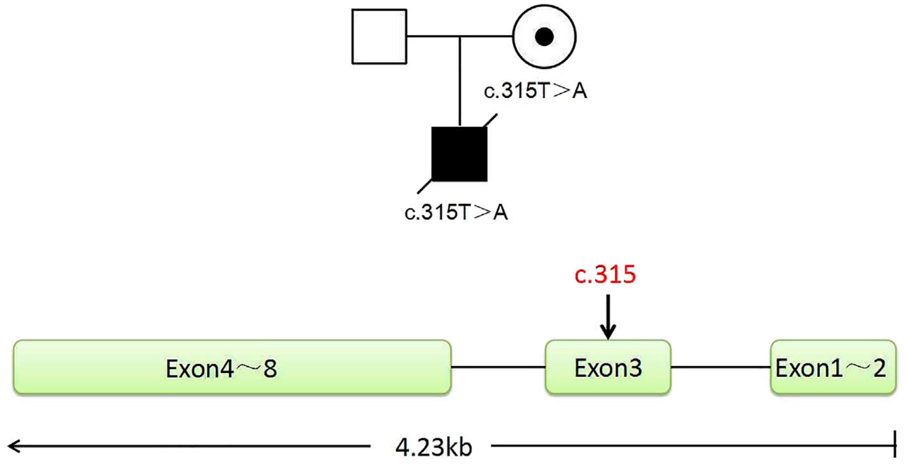


FIGURE 1 | Pedigree of the family. The proband is a deceased hemizygous patient, and the woman is a disease-causing gene carrier, and the man is normal. The *IL2RG* gene with eight exons is located in the X chromosome Xq13.1. The graph shows the position of the mutant sites of c.315, located in exon 3.

95°C for 5 min; 95°C for 30 s, 57°C for 30 s, 72°C for 40 s (35 cycles); 72°C for 5 min; held at 4°C. Sanger sequencing data were analyzed by ChromasPro software.

Before the scheme could be used for the TE cells, we used the wife's BMCs to imitate biopsied trophectoderm (TE) cells because the proband had passed away. The purpose is to test the effectiveness of the primers and the allele drop-out (ADO) rate of the WGA (WGA) method. MALBAC was applied for five to eight BMCs. WGA production of BMCs was used to amplify the target segments and Sanger sequencing.

Library Preparation and NGS

Purified WGA products were used for the PCR and SNP-based haplotype via NGS. An NGS library preparation kit (Yikon Genomics) was used to prepare the SNP library. For the CNV library, unpurified WGA products were used via an NGS library preparation kit (Yikon Genomics). All operations followed the manufacturers' protocols. Library sequencing was performed in the MiSeq Dx platform using a MiSeq Reagent Kit v3 (150-cycles) (Illumina). The raw data were automatically filtered, generating FASTQ files; the Q30 should be greater than 90%.

CNV Analysis and NGS-Based SNP Haplotype

For CNV analysis, FASTQ files were disposed of in the local analysis platform ChromGo (Yikon Genomics). More than 4 Mb deletions or duplications were reported. For CNV analysis, valid reads should be more than 1 Mb, CV (1000K_bin_size), and valid read GC contents should be in an acceptable range. To detect ADO and recombination, an NGS-based haplotype was conducted using SNP within the 2-million base pair (Mb) region, flanking the targeted gene. In this case, 60 SNPs were selected. For the SNP haplotype, bioinformatics analysis was conducted by Yikon Genomics, Ltd. The informative SNP sites are homozygous in the spouse and heterozygous in the variant carrier. Proband haplotypes were used for reference.

Embryo Transplantation and Prenatal Genetic Diagnosis

For PGT-M, embryo selection involves comprehensive considerations (Munné et al., 2019), including quality, the developmental stage of the embryo, and the results of PGT-A&M. Results were confirmed by the prenatal genetic diagnosis of amniotic fluid cells at gestational weeks 18–22^{+6}.

RESULTS

Patients and Genetic Background

In this family, this variant is the cause of X-linked SCID, whose hereditary mode is X-linked recessive inheritance (**Figure 1**). The size of the *IL2RG* gene is about 4.23 kb, containing eight exons (**Figure 1**). The c.315T > A variant occurred in exon 3. According to the ACMG/AMP guidelines, we evaluated this variant with PVS1 + PM2_Supporting as likely pathogenic. Both the husband and wife had normal karyotypes (320 bands). The peripheral blood of the couple and gDNA of the proband were detected by amplifying the target segments and Sanger sequencing (**Supplementary Figure S1**). Results showed that the female and proband were the carrier and hemizygote, respectively. The male did not carry this variant in the *IL2RG* gene (c.315T > A). These tests were carried out on the WGA production of BMCs from the carrier. The same diagnostic result was obtained as that of the peripheral blood cell test. The results indicate that the primers and MALBAC method could detect pathogenic variants in WGA products. (**Supplementary Figure S1**).

TE Cell Biopsy and Detection of Pathogenic Variant Allele

In this PGT cycle, 12 embryos finally developed into blastocysts after ICSI (**Table1**). The TE biopsy was conducted by a mechanical method on day 5 or 6 post-insemination. A total

TABLE 1 | Summary of detection results.

Biopsied Blastocysts	Gardner Grade	Copy Number Variations	SNP Haplotype	Sanger Sequencing
E1	4BC	—	male patient	c.315T > A
E2	4BB	46,XX	normal female	normal
E3	4BC	—	male patient	c.315T > A
E4	4BB	—	male patient	c.315T > A
E5	4BC	45,X,-X (×1)	abnormal detection	normal
E6	4BC	48,XX,+13 (×3),+16 (×3)	normal female	normal
E7	4B^{-}C	46,XX, -Xq (q13.3→q28,~80 Mb,×1,mos,~50%)	normal female	normal
E8	4BC	46,XX, -4q (q34.3→q35.2,~13 Mb,×1)	normal female	normal
E9	4BB	46,XX	normal female	normal
E10	4BC	46,XY	normal male	normal
E11	5BC	46,XX	normal female	normal
E12	4BC	—	male patient	c.315T > A

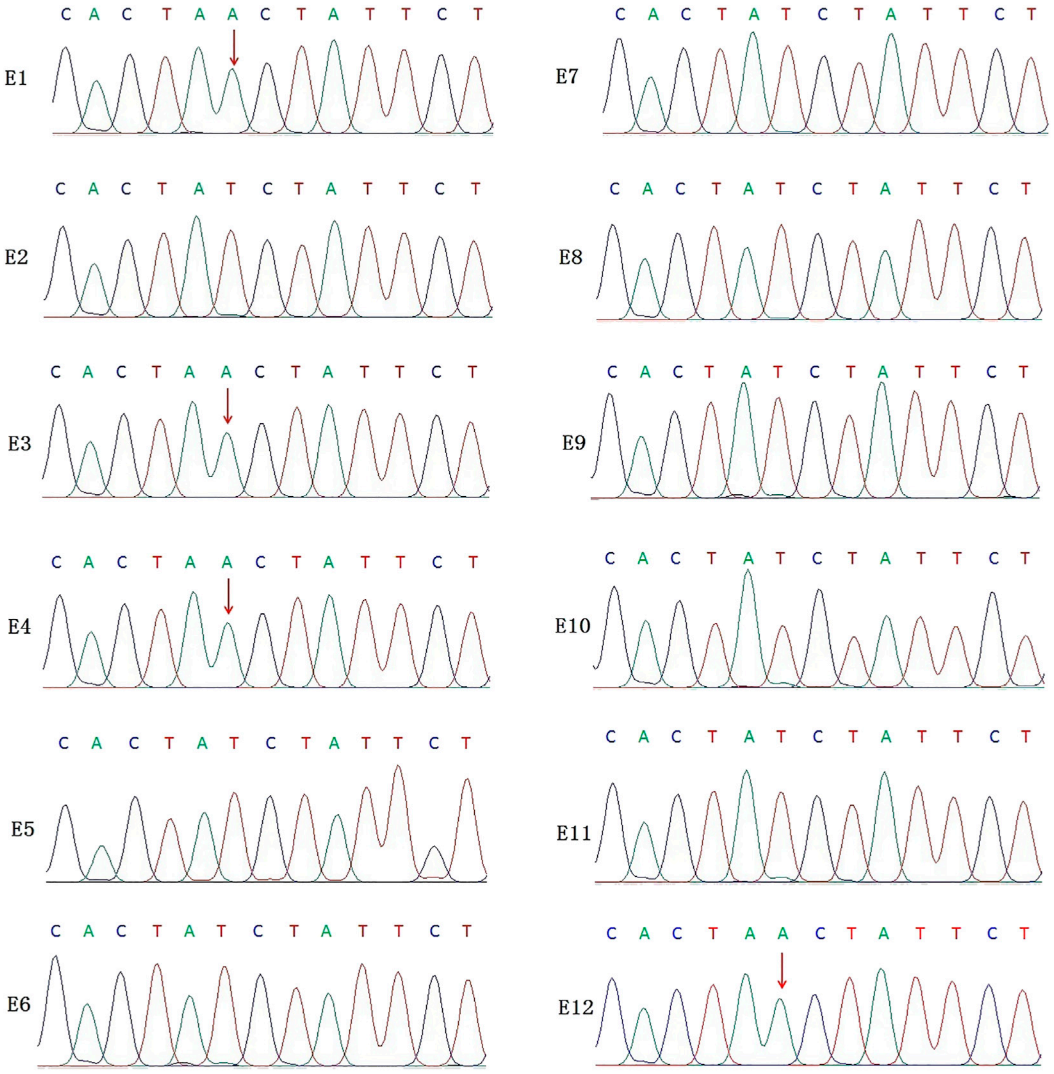


FIGURE 2 | Results of the Sanger sequencing for biopsied TE cells. The figure shows that E1, E3, E4, and E12 are carriers of the *IL2RG* c.315 (T > A) variant; no variants were detected in the remaining embryo samples. For X-linked recessive diseases, this technique alone cannot distinguish between male patients and female carriers. At the same time, misdiagnosis lead by ADO cannot be ruled out.

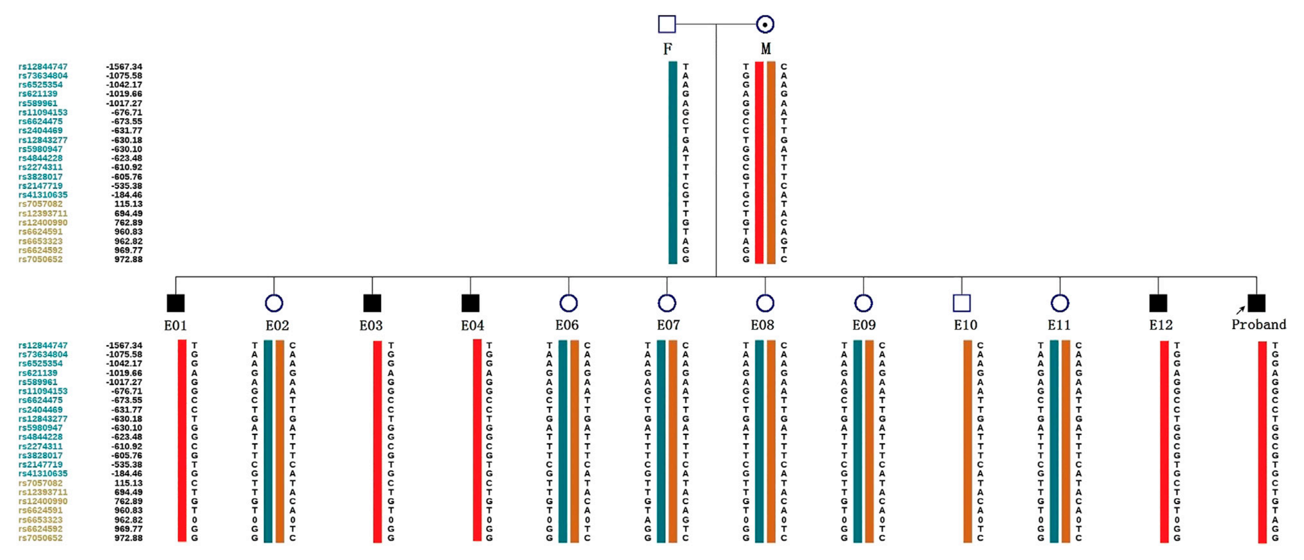


FIGURE 3 | Schematic diagram representing the SNP-based haplotype of this family members and embryos. F, M, and P represent the father, mother, and proband, respectively. The reference SNP cluster ID numbers were listed on the left side. The ID numbers highlighted in dark blue and orange refer to the upstream and downstream informative SNPs, respectively. The red bar refer to the high-risk haplotypes, the dark blue bar represent the normal or low-risk haplotype of the father, and the dark orange bar represent the normal or low-risk haplotype of the mother. Results show that the E01, E03, E04, and E12 are male patient embryos; the E02, E06, E07, E08, E09, and E11 are normal female embryos; and the E10 is a normal male embryo. E05 was not included because of anomalies found in the linkage analysis.

of five to eight cells were collected and transferred into 0.2 μL PCR tubes, to which lysis buffer was added beforehand. The MALBAC two-step method was used for WGA. To detect the variant allele, parts of WGA products were purified to obtain 100 bp–10 kb fragments and eliminate primers, enzymes, and oligonucleotides of the previous reaction. Then, the purified products were detected by amplifying the target segments and Sanger sequencing (**Figure 2**). The data show that embryos E1, E3, E4, and E12 had a variant in the *IL2RG* gene c.315 (T > A). The remaining embryos either did not indicate variants or there was morbigenous ADO. Therefore, linkage analysis is necessary to determine if there is an ADO or chromosome recombination.

SNP Haplotype

ADO may lead to misdiagnosis. To minimize its interference with the diagnostic results, SNP markers within the 2 Mb region flanking the target gene were used for linkage analysis. The peripheral blood samples of this couple, gDNA of the proband, and WGA products of embryos were applied to the SNP haplotype (**Figure 3**). For the *IL2RG* gene, 60 SNPs were chosen for linkage analysis. In this case, within 1 Mb upstream and downstream of the variant site, there were 7 and 13 SNPs, respectively. The selection of informative SNP was based on the ESHRE PGT Consortium good practice recommendations for detecting monogenic disorders (Group et al., 2020). Combined with the pedigree haplotype, the high-risk maternal chromosome was indicated by a red bar (**Figure 3**); the dark blue and dark orange represent paternal and low-risk maternal chromosomes, respectively. Results show that E02, E06, E07, E08, E09, E10, and E11 did not carry the high-risk maternal chromosome. E05 is not listed in the graph because its linkage analysis cannot be determined, as it only had the paternal X chromosome; therefore, as we think it might be X monosomy, we will confirm our conjecture through PGT-A soon. The results of pathogenic variant site detection and linkage analysis were considered, and we can conclude that E02, E06, E07, E08, E09, E10, and E11 do not carry maternal c.315 (T > A) in the *IL2RG* gene.

Preimplantation Genetic Testing for Aneuploidy (PGT-A)

The above seven embryos that did not carry the maternal variant were detected for aneuploidy. NGS-based CNV-Seq (CNV sequencing) was performed in the Illumina MiseqDx platform, indicating large than 10 Mb CNV and 30–70% mosaicism (large than 30 Mb). The CNV results are summarized in **Table 1**. No CNV larger than 10 M and aneuploidy were found in E02, E09, E10, and E11. E05, which could be determined for linkage analysis, indicated X monosomy (**Supplementary Figure S2**). E06 was aneuploidy, and E07 and E08 had fragment deletion or mosaicism (**Supplementary Figure S2**).

Embryo Transplantation and Prenatal Genetic Diagnosis

For the embryo transplant decision, parents had discussions with a reproductive doctor, embryologist and medical genetic doctor in the Reproductive Medicine-Medical Genetic MDT (Multi-disciplinary Treatment, MDT) of the West China Second University Hospital. The decision was based on the developmental stage and grade of the embryo, according to

PGT-M and PGT-A results (**Table1**). Finally, E02 was used for the transplantation, and the pregnancy was successful. At the 20th week of gestation, the pregnant woman underwent amniocentesis for prenatal genetic diagnosis. Amniotic cells were applied for CMA (cytoscan 750K, Affymetrix) (Lwvy and Wapner, 2018), amplification of the target segments, and Sanger sequencing (data not shown). The prenatal diagnosis was consistent with the preimplantation genetic testing. The woman eventually gave birth to a healthy girl who did not carry the variant.

DISCUSSION

SCID is an X-linked recessive monogenic hereditary disease characterized by defects in humoral and cellular immunity (Verbsky et al., 2012). Affected children often do not survive their first year of life due to recurrent and severe infections. PGT-M is an effective technique for blocking the transmission of this disease-causing gene. Until now, allogenic HSCT (Haddad and Hoenig, 2019) and gene therapy (Ferrua and Aiuti, 2017) are promising treatments for SCID (Suk See De Ravin, 2016). However, these two technologies still have limitations in clinical applications (Fischer and Hacein-Bey-Abina, 2020). For some inherited immunodeficiency like SCID, the PGT has become a practical technique for these at-risk couples to avoid affected pregnancies and have a healthy progeny free from genetic and chromosomal disorders (Rechitsky et al., 2018). At present, there is no mature gene therapy program for clinical application for children born with SCID, HSCT from a matched sibling donor is the most effective way. HLA matched stem cell transplantation improves significantly the outcome of transplantation treatment (Rechitsky et al., 2018). If children born with SCID are treated and protected early by pediatric ICU, they may be able to wait until PGT-M with HLA typing technique helps their family have a sibling who can provide a suitable transplantation donor.

We report a case blocking offspring with a rare inherited disease SCID by PGT-M. MALBAC is applied for WGA. An SNP-based haplotype and NGS were used for PGT-M&A by Illumina MiseqDx. Recently, MALBAC has been gradually recognized for its comparable single nucleotide variations detection efficiency, false-positive ratio, ADO ratio with MDA (Hou et al., 2015). Moreover, MALBAC is suited for the detection of CNVs because of its lower long-range variability in read mapping and weaker inherent bias in amplification than MDA (Charles F. A. de Bourcy, 2014).

Before the test protocol is applied to biopsied TE cells, we first detected gDNA of peripheral blood cells and WGA products of BMCs to evaluate the effectiveness and ADO rate of this method. Then we performed a PGT-M test on biopsied samples. For X-linked recessive inherited SCID, female carrier and normal embryo proceed to the next step of PGT-A. In linkage analysis, it was found that embryo 05 carried only the male normal X chromosome haplotype. In the Sanger sequencing of the pathogenic site, embryo 05 showed no variant. We hold the opinion that embryo 05 may be the X-monosomy inherited from the father; subsequent PGT-A results confirmed our conjecture. Therefore, to use PGT-M for X-linked genetic diseases, it is necessary to design a pair of primers to amplify the partial loci of the SRY gene or apply with PGT-A for embryo sex identification. For autosomal genetic diseases, PGT-A is also recommended after PGT-M. This helps to prevent us from misjudging the embryo in these cases. After PGT-M and PGT-A, E02, E09, E10, and E11 did not carry pathogenic variants, and no clear chromosomal abnormalities were found. E02 was eventually used for transplantation after considering the PGT results and morphological scores.

After a successful pregnancy by the PGT technique, prenatal diagnosis by amniocentesis during the second trimester is indispensable. Due to the possibility of mosaicism, biopsied trophectoderm cells represent only the genetic makeup of the placenta; there may be a normal placenta and abnormal fetal development. Therefore CMA and amplification of the target segments for Sanger sequencing were performed in amniocytes. This ensures that healthy babies are born through this double-check method.

In conclusion, PGT-M has important clinical application value in the prevention and control of birth defects. It can avoid the psychological and physical harm caused by induced labor after prenatal diagnosis. In this case, PGT (monogenic disorder and aneuploidy) combined with follow-up prenatal diagnosis helped a family with *IL2RG* gene variant have a healthy infant. We believe double detection at the blastocyst stage and second trimester, as well as the rational choice of multiple technologies can guarantee the birth of healthy children. At the same time, it also minimizes the risk of misdiagnosis caused by chromosome recombination, mosaicism, and ADO.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The raw datasets analysed during the current study are not deposited in publicly available repositories because of considerations about the security of human genetic resources and patient anonymity, but are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to Shanling Liu, sunny630@126.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of West China Second Hospital, Sichuan University. Written informed consent to participate in this study was provided by the participants'legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

Conceptualization: JR and SL; Methodology: CP, YL, YK, HC, and HZ; Formal analysis and investigation: FZ and XC;

Writing—original draft preparation: JR; Writing—review and editing: CL; Funding acquisition: SL; Resources: JR; Supervision: XC and SL.

FUNDING

This study was supported by the National Key Research and Development Program of China (2021YFC1005303) and the Technology Research and Development Program of the Science and Technology, Sichuan Province, China (2021YFS0078).

ACKNOWLEDGMENTS

We are grateful to the family for their cooperation and participation, to the embryology team in the Center of Reproductive Medicine for the help with sample preparation and to Zhen Liu from Yikon Genomics for the technical help for data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.926060/full#supplementary-material

REFERENCES

Blanco, E., Izotova, N., Booth, C., and Thrasher, A. J. (2020). Immune Reconstitution After Gene Therapy Approaches in Patients with X-Linked Severe Combined Immunodeficiency Disease. *Front. Immunol.* 11, 608653. doi:10.3389/fimmu.2020.608653

Carvalho, F., Moutou, C., Dimitriadou, E., Dreesen, J., Gimenez, C., et al.ESHRE PGT-M Working Group (2020). ESHRE PGT Consortium Good Practice Recommendations for the Detection of Monogenic Disorders. *Hum. Reprod. Open* 2020 (3), hoaa018. doi:10.1093/hropen/hoaa018

Cavazzana-Calvo, M., Hacein-Bey, S., Basile, G. d. S., Gross, F., Yvon, E., Nusbaum, P., et al. (2000). Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science* 288, 669–672. doi:10.1126/science.288.5466.669

Cossu, F. (2010). Genetics of SCID. *Ital. J. Pediatr.* 36, 76. doi:10.1186/1824-7288-36-76

de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., and Quake, S. R. (2014). A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *Plos One* 9 (8), e105585. doi:10.1371/journal.pone.0105585

De Ravin, S. S., Wu, X., Moir, S., Anaya-O'Brien, S., Kwatemaa, N., Littel, P., et al. (2016). Lentiviral Hematopoietic Stem Cell Gene Therapy for X-Linked Severe Combined Immunodeficiency. *Sci. Transl. Med.* 8 (335), 335ra57. doi:10.1126/scitranslmed.aad8856

De Rycke, M., and Berckmoes, V. (2020). Preimplantation Genetic Testing for Monogenic Disorders. *Genes* 11 (8), 871. doi:10.3390/genes11080871

Ferrua, F., and Aiuti, A. (2017). Twenty-Five Years of Gene Therapy for ADA-SCID: FromBubble Babiesto an Approved Drug. *Hum. Gene Ther.* 28 (11), 972–981. doi:10.1089/hum.2017.175

Fischer, A., and Hacein-Bey-Abina, S. (2020). Gene Therapy for Severe Combined Immunodeficiencies and beyond. *J. Exp. Med.* 217 (2), e20190607. doi:10.1084/jem.20190607

Haddad, E., and Hoenig, M. (2019). Hematopoietic Stem Cell Transplantation for Severe Combined Immunodeficiency (SCID). *Front. Pediatr.* 7, 481. eCollection 2019. doi:10.3389/fped.2019.00481

Hou, Y., Ww, K., Shi, X., Li, F., Song, L., Wu, H., et al. (2015). Comparison of Variations Detection between Whole-Genome Amplification Methods Used in Single-Cell Resequencing. *Gigascience* 4, 37. doi:10.1186/s13742-015-0068-3

Kwan, A., Abraham, R. S., Currier, R., Brower, A., Andruszewski, K., Abbott, J. K., et al. (2014). Newborn Screening for Severe Combined Immunodeficiency in 11 Screening Programs in the United States. *JAMA* 312 (7), 729–738. doi:10.1001/jama.2014.9132

Lwvy, B., and Wapner, R. (2018). Prenatal Diagnosis by Chromosomal Microarray Analysis. *Fertil. Steril.* 109, 201–212. doi:10.1016/j.fertnstert.2018.01.005

Mamcarz, E., Zhou, S., Lockey, T., Abdelsamed, H., Cross, S. J., Kang, G., et al. (2019). Lentiviral Gene Therapy Combined with Low-Dose Busulfan in Infants with SCID-X1. *N. Engl. J. Med.* 380 (16), 1525–1534. doi:10.1056/NEJMoa1815408

Mcarthur, S. J., Leigh, D., Marshall, J. T., Dd Boer, K. A., and Jansen, R. P. (2005). Pregnancies and Live Births after Trophectoderm Biopsy and Preimplantation Genetic Testing of Human Blastocysts. *Fertil. Steril.* 84 (6), 1628–1636. doi:10.1016/j.fertnstert.2005.05.063

Munn, S., Kaplan, B., Frattarelli, J. L., Child, T., Nakhuda, G., Shamma, F. N., et al. (2019). Preimplantation Genetic Testing for Aneuploidy versus Morphology as Selection Criteria for Single Frozen-Thawed Embryo Transfer in Good-Prognosis Patients: a Multicenter Randomized Clinical Trial. *Fertil. Steril.* 112 (6), 1071–1079. e7. doi:10.1016/j.fertnstert.2019.07.1346

Rechitsky, S., Pakhalchuk, T., Prokhorovich, M., Ramos, G. S., Verlinsky, O., and Kuliev, A. (2018). Preimplantation Genetic Testing for Inherited Immunodeficiency. *Hematol. Transfus. Int. J.* 6 (6), 218–220. doi:10.15406/htij.2018.06.00187

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Mutations: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Schoolcraft, W. B., Treff, N. R., Stevens, J. M., Ferry, K., Katz-Jaffe, M., and Scott, R. T. (2011). Live Birth Outcome with Trophectoderm Biopsy, Blastocyst Vitrification, and Single-Nucleotide Polymorphism Microarray–Based Comprehensive Chromosome Screening in Infertile Patients. *Fertil. Steril.* 96 (3), 638–640. doi:10.1016/j.fertnstert.2011.06.049

Tsuge, I., Matsuoka, H., Abe, T., Kamachi, Y., and Torii, S. (1996). Interleukin-2 Receptor Y-Chain Mutations in Severe Combined Immunodeficiency with B-Lymphocytes. *Eur. J. Pediatr.* 155 (12), 1018–1024. doi:10.1007/BF02532522

Verbsky, J. W., Baker, M. W., Grossman, W. J., Hintermeyer, M., Dasu, T., Bonacci, B., et al. (2012). Newborn Screening for Severe Combined Immunodeficiency; the Wisconsin Experience (2008-2011). *J. Clin. Immunol.* 32 (1), 82–88. doi:10.1007/s10875-011-9609-4

Zhang, J. Y., Yao, Y. Y., He, H. X., and Shen, J. (2020). Clinical Interpretation of Sequence Variants. *Curr. Protoc. Hum. Genet.* 106 (1), e98. doi:10.1002/cphg.98

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fgene.2022.880208

Risk Factors Affecting Alternate Segregation in Blastocysts From Preimplantation Genetic Testing Cycles of Autosomal Reciprocal Translocations

Pingyuan Xie[1,2,3†], Liang Hu[2,3,4,5†], Yangqin Peng[5†], Yue-qiu Tan[2,3,4,5], Keli Luo[2,3,4,5], Fei Gong[2,3,4,5], Guangxiu Lu[2,3,4,5] and Ge Lin[2,3,4,5]*

[1]Hunan Normal University School of Medicine, Changsha, China, [2]National Engineering and Research Center of Human Stem Cells, Changsha, China, [3]Hunan International Scientific and Technological Cooperation Base of Development and Carcinogenesis, Changsha, China, [4]NHC Key Laboratory of Human Stem Cell and Reproductive Engineering, Institute of Reproductive and Stem Cell Engineering, School of Basic Medical Science, Central South University, Changsha, China, [5]Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha, China

Reciprocal translocations are the most common structural chromosome rearrangements and may be associated with reproductive problems. Therefore, the objective of this study was to analyze factors that can influence meiotic segregation patterns in blastocysts for reciprocal translocation carriers. Segregation patterns of quadrivalents in 10,846 blastocysts from 2,871 preimplantation genetic testing cycles of reciprocal translocation carriers were analyzed. The percentage of normal/balanced blastocysts was 34.3%, and 2:2 segregation was observed in 90.0% of the blastocysts. Increased TAR1 (ratio of translocated segment 1 over the chromosome arm) emerged as an independent protective factor associated with an increase in alternate segregation ($p = 0.004$). Female sex and involvement of an acrocentric chromosome (Acr-ch) were independent risk factors that reduced alternate segregation proportions ($p < 0.001$). Notably, a higher TAR1 reduced the proportion of adjacent-1 segregation ($p < 0.001$); a longer translocated segment and female sex increased the risk of adjacent-2 segregation ($p = 0.009$ and $p < 0.001$, respectively). Female sex and involvement of an Acr-ch enhanced the ratio of 3:1 segregation ($p < 0.001$ and $p = 0.012$, respectively). In conclusion, autosomal reciprocal translocation carriers have reduced proportions of alternate segregation in blastocysts upon the involvement of an Acr-ch, female sex, and lower TAR1. These results may facilitate more appropriate genetic counseling for couples with autosomal reciprocal translocation regarding their chances of producing normal/balanced blastocysts.

Keywords: reciprocal translocation, PGT-SR, meiotic segregation patterns, next-generation sequencing, blastocysts

Edited by:
Xi Wang,
Nanjing Medical University, China

Reviewed by:
Miluse Vozdova,
Veterinary Research Institute (VRI), Czechia
Marta Olszewska,
Institute of Human Genetics (PAN), Poland

****Correspondence:***
Ge Lin
linggf@hotmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *21 February 2022*
Accepted: *11 April 2022*
Published: *02 June 2022*

Citation:
Xie P, Hu L, Peng Y, Tan Y-q, Luo K, Gong F, Lu G and Lin G (2022) Risk Factors Affecting Alternate Segregation in Blastocysts From Preimplantation Genetic Testing Cycles of Autosomal Reciprocal Translocations. Front. Genet. 13:880208. doi: 10.3389/fgene.2022.880208

INTRODUCTION

Reciprocal translocations are the most common structural chromosomal reorganizations in humans and typically involve two chromosomes, each with one breakpoint to generate two distal segments that are interchanged (Morin et al., 2017). Balanced translocation carriers are commonly phenotypically normal, and >6% of reciprocal carriers have a variety of symptoms, such as autism, intellectual disabilities, or congenital abnormalities (Warburton, 1991). Reciprocal translocations are associated with reproductive problems, such as subfertility, infertility, and recurrent miscarriage, because of genomic imbalance in the gametes (Scriven et al., 1998; Morin et al., 2017; Fatemi et al., 2021). The prevalence of reciprocal translocation is approximately 0.14% in newborns (Nielsen and Wohlert, 1991), 1.2% in infertile men with azoospermia (Stern et al., 1999; Clementini et al., 2005), and 2.4–6.9% in couples with recurrent miscarriage (Clifford et al., 1994; Stern et al., 1999).

During meiosis, the two translocated chromosomes and their two homologous normal chromosomes form a quadrivalent (Scriven et al., 1998; Tease et al., 2002). In anaphase I, the quadrivalent structure is segregated to the daughter cells through one of the following five modes: 1) alternate segregation, including segregation of a normal chromosome complement and balanced rearrangement into different poles; 2) adjacent-1 segregation, in which homologous centromeres pass to the opposite poles and lead to monosomy for one translocated segment and trisomy for the other; 3) adjacent-2 segregation, in which homologous centromeres are segregated together resulting in trisomy for one centric segment and monosomy for the other; 4) 3:1 segregation, wherein three out of four chromosomes are segregated together and may result in trisomy or monosomy; and 5) 4:0 segregation, wherein all four chromosomes are separated into one pole, leading to double-trisomy or double-monosomy. Owing to recombination, 32 possible gametes with different karyotypes may occur, and only two of these, which arise from the alternate segregation mode, can have normal/balanced genetic content (Scriven et al., 1998).

Preimplantation genetic testing (PGT) has been a common technique to exclude unbalanced embryos for transfer since 1998 (Munne et al., 1998; (Pierce et al., 1998). Originally, fluorescence *in situ* hybridization (FISH) was used to analyze the segregation mode in the spermatozoon or biopsied blastomeres (Anton et al., 2008; Lim et al., 2008). More recently, with the development of whole-genome amplification (WGA) and comprehensive chromosome screening (CCS), including single nucleotide polymorphism (SNP) arrays, comparative genomic hybridization (CGH) arrays, and next-generation sequencing (NGS) (Tan et al., 2014; Idowu et al., 2015), all components of chromosomes can be analyzed simultaneously.

Extensive studies on meiotic segregation patterns of sperm, blastomeres, and blastocysts have identified several factors that may influence the unbalanced rearrangement of reciprocal translocations, including carrier's sex and age, location of the breakpoints, chromosome type, and quadrivalent structure (Munne, 2005; Anton et al., 2008; Zhang et al., 2018; Zhang et al., 2019a; Zhang et al., 2019b). Gamete results indicate a very wide range of chromosomal abnormalities, depending on the translocation, ranging from 18 to 82% abnormal sperm (Munne, 2005). A previous study suggested that a sperm FISH analysis prior to PGD is not a reliable predictor of the PGT outcome (Haapaniemi Kouru et al., 2017). Recently, several studies have assessed meiotic segregation patterns in blastocysts using CCS, even though controversial results could be achieved (Zhang et al., 2018; Zhang et al., 2019b; Wang et al., 2019). Wang et al. (2019) demonstrated that segregation modes in blastocysts could be affected by the prsence of acrocentric chromosomes and terminal breakpoints, rather than by the carrier's sex. Zhang et al. (2019b) suggested that an acrocentric chromosome (Acr-ch) involved in translocation interacts with the carrier's sex to influence the proportion of alternate segregation for normal or balanced chromosome contents during meiotic segregation in autosomal reciprocal translocations.

In the present study, our aim was to evaluate the factors that affect alternate segregation and to provide more appropriate genetic counseling for couples with autosomal reciprocal translocation regarding their chances of producing normal/balanced blastocysts. The impact factors were analyzed according to the carrier's sex, age, translocated chromosome type, and the location of translocated chromosome breakpoints using NGS.

MATERIALS AND METHODS

Study Subjects

The study was approved by the Ethics Committee of the Reproductive and Genetic Hospital of CITIC-Xiangya.

We enrolled 2,253 couples with autosomal reciprocal translocations, whose blastocysts were tested using NGS, from October 2013 to December 2019 from the Reproductive and Genetic Hospital of CITIC-Xiangya. G-banding chromosome analysis of cultured lymphocytes from peripheral blood was performed at our hospital or other certified organizations. The karyotypes in this report are described according to the ISCN 2013 guidelines. All individuals had a history of infertility, recurrent spontaneous abortion, or pregnancy with chromosomal anomalies.

PGT-SR Procedure

Preimplantation genetic testing for chromosomal structural rearrangement (PGT-SR) was performed at the Reproductive & Genetic Hospital of CITIC-Xiangya, as previously described. In brief, ovarian stimulation was conducted based on the patient's status. Fertilization was assessed at 16–18 h post-intracytoplasmic sperm injection (ICSI) by visualization of two pronuclei and two polar bodies. All embryos were cultured in sequential media (G1 and G2; Vitrolife) to the blastocyst stage under 6% CO_2, 5% O_2, and 89% N_2 in a Cook mini-incubator. Trophectoderm (TE) biopsy was performed on day 5 or day 6, and approximately five cells were collected after zona pellucida dissection and hatching using a laser (ZILOS-tk, Hamilton Thorne). The blastocysts were then cryopreserved as previously described (Zhang et al., 2016).

Two NGS platforms were used in this study. For the initial 2,345 trophectoderm cells (from October 2013 to May 2015), whole-genome amplification (WGA) was performed using the WGA4 GenomePlex Single Cell Whole Genome Amplification kit (Sigma-Aldrich, St. Louis, MO, United States). The Illumina HiSeq2000 platform was used for NGS, and approximately 5–10 million single-end (SE) 50 bp reads were obtained for each sample. For the remaining 8,501 trophectoderm cells, WGA was performed using the REPLI-g Single Cell Kit (Qiagen, Hilden, Germany). The Illumina NextSeq platform was used for NGS, and a 10 million SE 75 bp reads were obtained for each sample. Copy number variations larger than 1–4 megabase (>1 Mb) could be detected (Tan et al., 2014; Zhou et al., 2018). The NGS-PGT platform results were confirmed using FISH or SNP arrays before clinical application, as described in our previous studies (Tan et al., 2013; Zhang et al., 2013).

Embryo Transfer

One or two cryopreserved embryos could be transferred, and only euploid blastocysts were selected for transfer. No blastocyst transfer was performed with an endometrium of < 7 mm thickness.

Statistical Analysis

The distribution of patient demographics and cytogenetic characteristics was analyzed using the Kolmogorov–Smirnov test. Continuous variables were expressed as mean ± standard deviation (SD). Categorical variables were described as the frequency and percentage. One-way analysis of variance (ANOVA) or Wilcoxon rank-sum test (alternate segregation group versus non-alternate segregation group) or Kruskal–Wallis H test (alternate segregation group versus adjacent-1 segregation group versus adjacent-2 segregation group versus 3:1 segregation group versus ND group) was used for comparison according to the distribution of continuous variables; chi-square or Fisher's exact tests were used to compare differences between categorical variables.

To evaluate associations of the patient demographics and cytogenetic characteristics with the embryo separation patterns, we utilized a multivariate generalized linear mixed model (GLMM) with random intercepts to account for multiple embryos from the same woman. A binomial distribution with a logit link function (alternate (balanced) segregation group versus unbalanced segregation group) or a multinomial distribution with a logit link function (alternate segregation group versus adjacent-1 segregation group versus adjacent-2 segregation group versus 3:1 segregation group versus not determined (ND) group) was specified for the embryo segregation mode. Covariates with statistically significant differences in univariate GLMM were considered for inclusion in the multivariate GLMM. The following covariates were considered for inclusion in the univariate GLMM: carrier sex, age, and chromosomes involved in the translocation (chromosomes #1 and #2 corresponding to the small and large chromosomes, respectively, according to the guidelines of the International System for Human Cytogenetic Nomenclature (ISCN) 2016). In addition, the inclusion of an Acr-ch, size of the involved chromosomes, size of the translocated segments (TS1 and TS2), size of the centric segments (CS1 and CS2), ratio of the longest and shortest translocated segments (TSR), ratio of the longest and shortest centric segments (CSR), and ratio of the translocated segment over the chromosome arm (TAR) were analyzed. The location of breakpoints was identified by the PGT-SR results of the karyotype; the length of both centric and translocated segments (CS and TS) was measured according to the copy number variations of the unbalanced embryos. TSR, CSR, and TAR were then calculated separately.

A receiver operating characteristic (ROC) curve was used to discriminate the predictive values of the patient demographics and clinical variables in a multivariate GLMM for the embryo separation mode (balanced group versus unbalanced group).

The Cochran–Mantel–Haenszel (CMH) test was utilized to explore the interaction effect of carrier sex and the involvement of an Acr-ch on the likelihood of every single embryo.

All statistical analyses were performed using SPSS 25.0 software (IBM, Armonk, NY, United States), and a two-sided p-value < 0.05 was considered statistically significant.

RESULTS

Patient Demographics and Clinical Characteristics

This study included 1,028 male and 1,225 female autosomal reciprocal translocation carriers. There were no significant differences in demographic characteristics between female and male carriers. In total, 31,295 oocytes originating from 2,871 oocyte retrieval cycles underwent ICSI, and 12,328 (39.4%) embryos were developed to blastocysts on days 5 or 6. Among 12,328 TE-biopsied embryos, 10,876 (88.2%) were tested and 10,846 were diagnosed successfully (**Table 1**); the percentage of normal/balanced blastocysts was 34.3% (3,716/10,846). The clinical pregnancy rate and delivery rate in the male carrier group were significantly higher than those in the female carrier group ($p < 0.05$) (**Table 1**).

Segregation Pattern Analysis

The segregation patterns were determined using PGT-SR results from translocation carrier individuals. Overall, among 10,864 diagnosed blastocysts, 2:2 segregation was observed in 9,751 (90.0%) blastocysts, of which alternate segregation was found as the most frequent segregation pattern (46.1%, 5,005/10,846), followed by adjacent-1 segregation (31.3%, 3,397/10,846) and adjacent-2 segregation (12.4%, 1,349/10,846). The details (karyotype and segregation pattern) are shown in **Supplementary Table S1**. The frequencies of 3:1 segregation and 4:0/other segregation patterns were 5.4% (585/10,846) and 4.7% (510/10,864), respectively. Among all 4:0/other segregation products, only four 4:0 products were identified, and most were classified as undetermined segregation products; therefore, we concentrated on the 2:2 and 3:1 segregation pattern.

Acr-ch Involved in the Translocation, Carrier Sex, and Lower TAR1 Could Reduce the Proportion of Balanced Embryos

As only alternate segregation could produce normal or balanced gametes, we first analyzed the factors that could affect the

TABLE 1 | Clinical characteristics and preimplantation genetic testing for chromosomal structural rearrangement outcomes of reciprocal translocation heterozygotes.

Parameter	Female carrier	Male carrier	p-value	Total
Patient number	1,028	1,225		2,253
Cycles	1,583	1,288		2,871
Female age (years)	29.97 ± 4	30.16 ± 4.04	0.488	29.87 ± 3.78
Male age (years)	31.93 ± 4.8	32.1 ± 5.04	0.151	31.81 ± 4.79
Retrieved oocytes	21,157 (13.4)	17,526 (13.6)	NS	38,684
Injected oocytes	17,175 (10.9)	14,119 (11.0)	NS	31,295
ICSI cleavage embryos	15,175 (9.6)	12,447 (9.7)	NS	27,622
Biopsied embryos	6,743 (4.3)	5,585 (4.3)	NS	12,328
Diagnosed embryos	5,995 (3.8)	4,851 (3.8)	NS	10,846
Transferable normal embryos	1918	1798		3,716
Abnormal embryos	4,077	3,053		7,130
Transferred embryos	876	852		1728
Embryo transfer cycles	830	730		1,560
Clinical pregnancies (%)	514 (61.9%)	526 (72.1%)	<0.001	1,040 (66.7%)
Spontaneous abortions (%)	76 (9.2%)	87 (11.9%)	0.082	175 (11.2%)
Deliveries (%)	437 (52.7%)	438 (60%)	0.004	875 (56.1%)
Ectopic pregnancies	1	1		

All data are n values unless stated otherwise. The results in parentheses are the mean oocyte or embryo numbers in each retrieved cycle.

TABLE 2 | Univariable and multivariable generalized linear mixed model analysis for unbalanced translocation.

	Univariate generalized linear mixed model (GLMM)		Multivariable generalized linear mixed model (GLMM)	
	OR (95% CI)	p	OR (95% CI)	p
Size 1	0.999 (0.998,1.000)	0.006		
TAR1	0.823 (0.711,0.954)	0.01	0.806 (0.696,0.933)	0.004
CS2	0.999 (0.998,1.000)	0.029		
Sex				
Female	1.295 (1.189,1.411)	<0.001	1.293 (1.187,1.408)	<0.001
Male	Reference		Reference	
Acr-ch				
With	1.22 (1.115, 1.333)	<0.001	1.208 (1.089,1.339)	<0.001
Without	Reference		Reference	

Acr-ch, acrocentric chromosome;
Size 1, size of the longer chromosome involved;
CS 2, size of the centric segments of the shorter chromosome involved;
TAR1, the ratio of the translocated segment over the chromosome arm of the longer chromosome.

proportion of alternate segregation. The blastocysts were divided into balanced (alternate segregation) and unbalanced (non-alternate segregation) groups according to their segregation patterns. The results of the univariate and multivariate GLMMs are presented in **Table 2**. All variables with $p < 0.05$ were used for multivariate analysis, including the length of the involved chromosomes, TAR1, CS2, carrier sex, and involvement of an Acr-ch. The multivariate analysis results indicated that TAR, carrier sex, and involvement of an Acr-ch could significantly influence the proportion of alternate segregation. In particular, 1) increased TAR1 reduced the non-alternate segregation ratio (odds ratio (OR), 0.806 [95% confidence interval (CI), 0.696, 0.933]; $p = 0.004$), female sex (OR, 1.293 [95% CI, 1.187, 1.408]; $p < 0.001$), and involvement of an Acr-ch (OR, 1.208 [95% CI, 1.089, 1.339]; $p < 0.001$) enhanced the proportion of non-alternate segregation. Furthermore, a ROC curve was used to evaluate the predictive values of alternate segregation. The values of area under the curve (AUC), sensitivity, and specificity were 0.708, 84.4%, and 40.3%, respectively (**Supplementary Figure S1**).

The CMH test was employed to further assess the effects of a carrier's sex and involvement of an Acr-ch on the likelihood of each single embryo in an unbalanced rearrangement. The results were summarized as follows (**Table 3** and **Figure 1A**): 1) the proportion of blastocysts carrying an unbalanced rearrangement was higher in the female group than that in the male group, both in carriers with an Acr-ch [62% versus 51.1%, $p < 0.001$, OR = 1.561, 95% CI (1.374, 1.775)] and without an Acr-ch [53.8% versus 49.8%, $p = 0.001$, OR = 1.176, 95% CI (1.07, 1.293)]. 2) The risk of an embryo with an unbalanced rearrangement was higher in carriers with an Acr-ch than in those without an Acr-ch in the female group [62e % versus 53.8%, $p < 0.001$, OR = 1.403, 95% CI (1.26, 1.563)], rather than in the male group (51.1% versus 49.8%, $p = 0.353$). The data were further stratified according to the chromosomes involved; in the Acr-ch group, the proportion of alternate segregation in female carriers, involving chr-2, -4, -6, -9,

TABLE 3 | Analysis of balanced and unbalanced segregation patterns according to carrier sex and chromosome type.

		Unbalanced	Balanced	p	OR (95%CI)	p*
With Acr-ch	Female	1,334 (62%)	816 (38%)	<0.001	1.561 (1.374,1.775)	<0.001
	Male	891 (51.1%)	851 (48.9%)			
Without Acr-ch	Female	2069 (53.8%)	1776 (46.2%)	0.001	1.176 (1.07,1.293)	
	Male	1,547 (49.8%)	1,562 (50.2%)			
Female	With Acr-ch	1,334 (62%)	816 (38%)	<0.001	1.403 (1.26,1.563)	<0.001
	Without Acr-ch	2069 (53.8%)	1776 (46.2%)			
Male	With Acr-ch	891 (51.1%)	851 (48.9%)	0.353	1.057 (0.94,1.189)	
	Without Acr-ch	1,547 (49.8%)	1,562 (50.2%)			

**p for tests of homogeneity of the odds ratio.*
Acr-ch, acrocentric chromosome.

-10, and -20, was significantly lower than that in the corresponding male carriers ($p < 0.05$). However, in the non-Acr-ch group, the proportion of alternate segregation in female carriers, involving chr-5, -8, -9, and -12 ($p < 0.05$), was significantly lower than that in the corresponding male carriers (**Figure 1B** and **Supplementary Table S2**).

Factors Influencing the Adjacent-1, Adjacent-2, and 3:1 Segregation Patterns

We identified several factors that could influence the proportions of adjacent-1, adjacent-2, and 3:1 segregation patterns (**Table 4**). In particular, the proportion of adjacent-1 segregation was affected by the TAR1, and a greater TAR1 reduced the proportion of adjacent-1 segregation [OR = 0.316, 95% CI (0.174, 0.575), $p < 0.001$]. For making simplification, we divided the TAR1 into TAR1 ≤ 0.2 group and TAR1 > 0.2 group. The results indicated that the proportion of adjacent-1 segregation was significantly higher in TAR1 ≤ 0.2 group than in TAR1 > 0.2 group (36.2% versus 30.2%, $p < 0.001$) and both in female subgroup (36.8% versus 29.2%, $p < 0.001$) and male subgroup (35.4% versus 31.5%, $p < 0.05$) (**Figure 1C** and **Supplementary Table S3**). The proportion of adjacent-2 segregation was influenced by the TS1 and carrier's sex, and longer TS1 and female sex increased the proportion of adjacent-2 segregation [OR = 1.014, 95% CI(1.003, 1.024), p = 0.009 and OR = 1.66, 95% CI(1.427, 1.931), $p < 0.001$, respectively] (**Figure 1D** and **Supplementary Table S4**); besides, the proportion of 3:1 segregation was affected by a carrier's sex and involvement of an Acr-ch, wherein female sex and involvement of an Acr-ch increased the proportion of 3:1 segregation [OR = 2.643, 95% CI (2.147, 3.253), $p < 0.001$ and OR = 1.387, 95% CI (1.075, 1.789), p = 0.012, respectively] (**Table 4**; **Figure 1E** and **Supplementary Table S5**).

DISCUSSION

After GLMM analysis of 10,846 biopsied blastocysts from 2,253 autosomal reciprocal translocation individuals, we found that carrier's sex, involvement of an Acr-ch, and lower TAR1 were independent risk factors that may be associated with the proportion of alternate segregation in autosomal reciprocal translocation. Our results also indicated that the most common pattern was alternate segregation (46.1%) in blastocysts, followed by adjacent-1 segregation. The percentage of normal/balanced embryos in diagnosed blastocysts in the present study was 34.3%, which was similar to that reported in several recently conducted studies (Idowu et al., 2015; Zhang et al., 2018; Zhang et al., 2019b; Wang et al., 2019).

Numerous studies have analyzed the meiotic segregation patterns in gametes or embryos from reciprocal translocation carriers and have demonstrated that the carrier's sex and age and the location of breakpoints might influence the meiotic segregation patterns, though their results were somewhat controversial (Benet et al., 2005; Lim et al., 2008; Lledo et al., 2010; Ye et al., 2012; Zhang et al., 2018). Regarding the carrier's sex and involvement of an Acr-ch, studies concentrating on both gametes and embryos suggested that a carrier's sex influenced the meiotic segregation patterns, and that male carriers have a significantly higher proportion of alternate segregation (Munne et al., 1998; Lin et al., 2021). However, several research studies reported that the proportions of adjacent-1, adjacent-2, or 3:1 segregation patterns were markedly different in male and female carriers, whereas this outcome was not observed in alternate segregation (Ko et al., 2010; Lledo et al., 2010; Ye et al., 2012; Zhang et al., 2018). Previous studies have indicated that participation of an Acr-ch in reorganization or a very asymmetrical configuration of the quadrivalent favored 3:1 segregation owing to the extremely small length of their arms, and thus marked asymmetry of the quadrivalent (Benet et al., 2005; Lim et al., 2008). Furthermore, the percentage of adjacent-1 products from a quadrivalent with an Acr-ch has been demonstrated to be significantly higher than that without an Acr-ch, and no significant differences in other segregation patterns have been identified (Zhang et al., 2018). The translocations with an Acr-ch exhibited a notably higher frequency of 3:1 segregation, whereas there was no significant difference in alternate segregation (Wang et al., 2019).

In contrast to a previous study, we observed no influence of a carrier's age on the meiotic segregation patterns. Overall, we found that a carrier's sex, involvement of an Acr-ch, and location of breakpoints may influence the proportion of segregation patterns. Furthermore, the proportion of alternate segregation

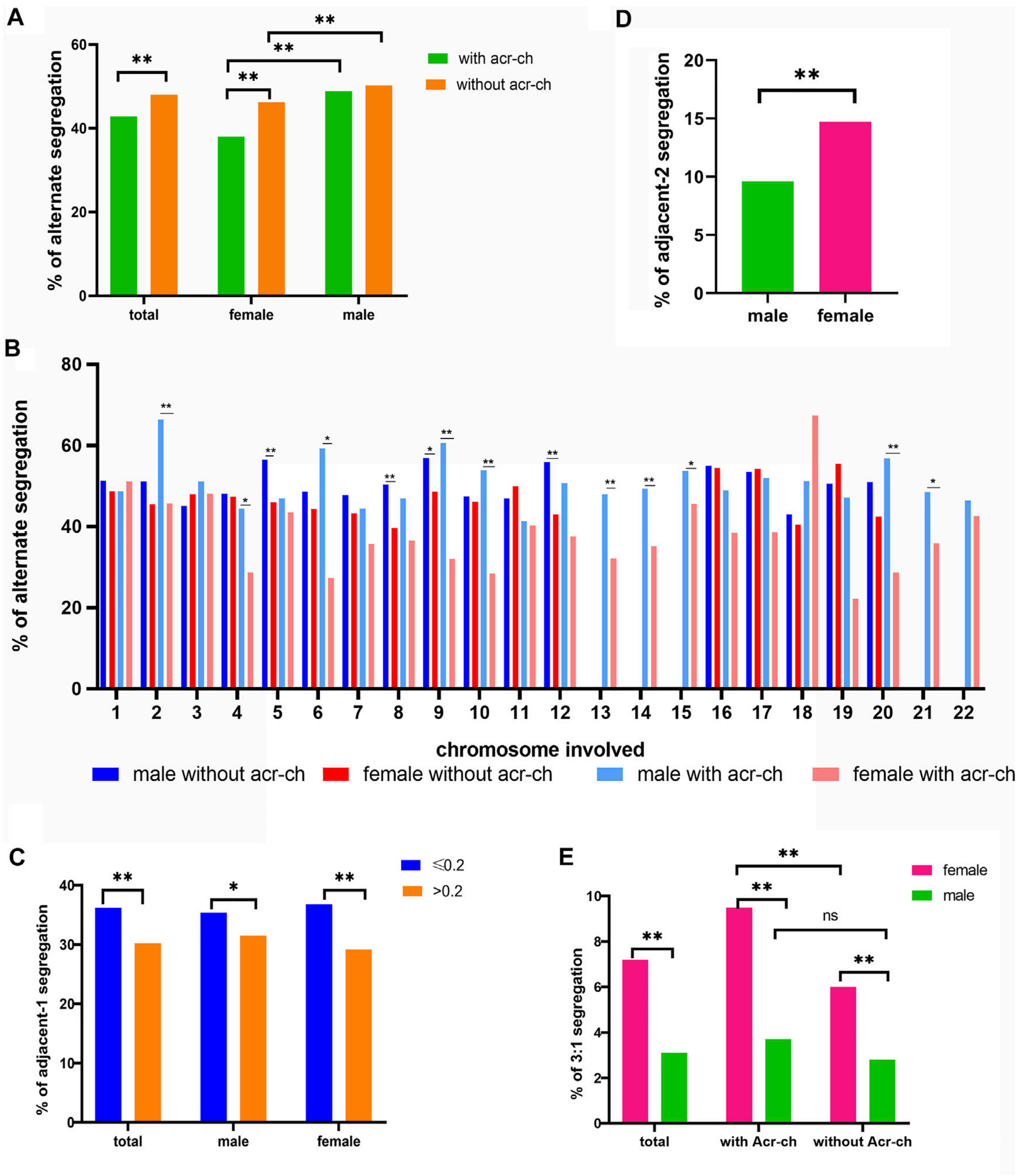


FIGURE 1 | Factors influencing meiotic segregation patterns in blastocysts for reciprocal translocation carriers. **(A)** Alternate segregation proportions in male and female carriers according to the type of chromosome involved (Acr-ch and acrocentric chromosome). **(B)** Alternate segregation proportions in male and female carriers according to the chromosome involved (Acr-ch and acrocentric chromosome). **(C)** Proportion of adjacent-1 segregation under different values of TAR1 and the carrier's sex group (TAR1, the ratio of translocated segment 1 over the chromosome arm). **(D)** Proportion of adjacent-2 segregation in female and male carriers with reciprocal translocation. **(E)** Proportion of 3:1 segregation with different sex and chromosome types involved (Acr-ch and acrocentric chromosome).

was significantly higher in male carriers than that in female carriers, particularly when an Acr-ch could interact with chr-2, -4, -6, -9, -10, and -20. Moreover, the proportion of alternate segregation was remarkably higher in female carriers with an Acr-ch than that in female carriers without an Acr-ch, rather than that in male carriers. Furthermore, the proportion of adjacent-2 and the 3:1 segregation patterns was significantly higher in female carriers than that in male carriers, although this finding was not confirmed for adjacent-1 segregation. In addition, the proportion of 3:1 segregation was notably higher in carriers with an Acr-ch than in those without an Ach-ch. Finally, our findings suggest that TAR1 could influence segregation patterns, wherein lower TAR1 could increase the proportion of adjacent-1 segregation and decrease the proportion of alternate segregation. Refer to why

TABLE 4 | Multivariable generalized linear mixed model analysis.

	Multivariable generalized linear mixed model (GLMM) OR (95% CI)	p
Adjacent-1 vs. alternate		
TAR1	0.316 (0.174,0.575)	<0.001
Adjacent-2 vs. alternate		
TS1	1.014 (1.003,1.024)	0.009
Carrier's sex		
Female	1.66 (1.427,1.931)	<0.001
Male	Reference	
3:1 vs. alternate		
Carrier's sex		
Female	2.643 (2.147,3.253)	<0.001
Male	Reference	
Acrocentric chromosome		
With	1.387 (1.075,1.789)	0.012
Without	Reference	

only TAR1 but not TAR2 could influence the segregation pattern needed further study.

Compared with previous studies, the present study possesses several strengths. First, the PGT-SR results of 10,846 blastocysts enabled statistical analysis of a relatively large number of samples in each subgroup. Second, analysis of GLMMs was advantageous because these models provided multiple observations per patient, while accounting for within-person correlations in the outcomes. These models could provide unbiased estimates in the presence of an unbalanced design (different numbers of embryos per cycle were contributed by each female patient) when imbalance in the number of embryos was not fully randomized, and the lack of balance could be accurately predicted by all measured covariates in an adjusted model. This study also has several limitations. First, our segregation pattern analysis was based on blastocysts, which may not reflect the actual segregation pattern in gametes. Furthermore, as only four embryos were diagnosed initially, approximately 1,400 blastocysts were not diagnosed in 2,871 cycles, which might have caused bias in the results.

A remarkably higher proportion of alternate segregation for normal or balanced chromosome contents was also observed in the blastocysts of male carriers than that in those of female carriers among Robertsonian translocations and pericentric inversions (Zhang et al., 2019a; Xie et al., 2019), suggesting that the checkpoints in oogenesis are not as stringent as those in spermatogenesis. The different spatial characteristics of the quadrivalent (i.e., involvement of an Acr-ch) could affect synaptic recombination in human oocytes and spermatocytes. However, understanding the detailed underlying mechanism requires further investigations.

NGS-PGT is widely used in PGT-SR to exclude unbalanced translocations and *de novo* chromosomal aberrations. However, in routine PGT, balanced and normal embryos cannot be distinguished by CCS, and recently developed techniques may allow distinction of inversion carrier embryos from those with normal chromosomes (Hu et al., 2016; Wang et al., 2017; Xu et al., 2017).

In conclusion, carrier's sex, involvement of an Acr-ch, and location of breakpoints may influence the segregation patterns in blastocysts. In addition, involvement of an Acr-ch, female sex, and lower TAR1 as independent risk factors can reduce the proportion of alternate segregation. These results may provide more appropriate information for couples with autosomal reciprocal translocation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Reproductive and Genetic Hospital of CITIC-Xiangya. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

GeL, GxL, and LH conceived the study. PX and YP performed the statistical analysis of the experiment. PX wrote the manuscript. KL and FG selected the patients. Y-qT and GL revised the manuscript. All authors critically read, commented, and approved the final manuscript.

FUNDING

This study was funded by the National Key R&D Program of China (2018YFC1004901 and 2018YFC1003100) and was supported by the Scientific Research Funds from Hunan Provincial Education Department (21C0044).

ACKNOWLEDGMENTS

We are grateful to the staff of the cytogenetic group and PGT group for providing technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.880208/full#supplementary-material

Supplementary Figure S1 | Receiver operating characteristic (ROC) curve was used to evaluate the predictive values of alternate segregation patterns.

REFERENCES

Anton, E., Vidal, F., and Blanco, J. (2008). Reciprocal Translocations: Tracing Their Meiotic Behavior. *Genet. Med.* 10, 730–738. doi:10.1097/gim.0b013e318187760f

Benet, J., Oliver-Bonet, M., Cifuentes, P., Templado, C., and Navarro, J. (2005). Segregation of Chromosomes in Sperm of Reciprocal Translocation Carriers: a Review. *Cytogenet. Genome Res.* 111, 281–290. doi:10.1159/000086901

Clementini, E., Palka, C., Iezzi, I., Stuppia, L., Guanciali-Franchi, P., and Tiboni, G. M. (2005). Prevalence of Chromosomal Abnormalities in 2078 Infertile Couples Referred for Assisted Reproductive Techniques. *Hum. Reprod.* 20, 437–442. doi:10.1093/humrep/deh626

Clifford, K., Rai, R., Watson, H., and Regan, L. (1994). Pregnancy: An Informative Protocol for the Investigation of Recurrent Miscarriage: Preliminary Experience of 500 Consecutive Cases. *Hum. Reprod.* 9, 1328–1332. doi:10.1093/oxfordjournals.humrep.a138703

Fatemi, N., Varkiani, M., Ramezanali, F., Babaabasi, B., Ghaheri, A., Biglari, A., et al. (2021). Risk Factors Associated with Recurrent Pregnancy Loss and Outcome of Pre-implantation Genetic Screening of Affected Couples. *Int. J. Fertil. Steril* 15, 269–274. doi:10.22074/IJFS.2021.137626.1027

Haapaniemi Kouru, K., Malmgren, H., White, I., Rodriguez Sanchez, A., and Syk Lundberg, E. (2017). Meiotic Segregation Analyses of Reciprocal Translocations in Spermatozoa and Embryos: No Support for Predictive Value Regarding PGD Outcome. *Reprod. BioMedicine Online* 34, 645–652. doi:10.1016/j.rbmo.2017.02.013

Hu, L., Cheng, D., Gong, F., Lu, C., Tan, Y., Luo, K., et al. (2016). Reciprocal Translocation Carrier Diagnosis in Preimplantation Human Embryos. *EBioMedicine* 14, 139–147. doi:10.1016/j.ebiom.2016.11.007

Idowu, D., Merrion, K., Wemmer, N., Mash, J. G., Pettersen, B., Kijacic, D., et al. (2015). Pregnancy Outcomes Following 24-chromosome Preimplantation Genetic Diagnosis in Couples with Balanced Reciprocal or Robertsonian Translocations. *Fertil. Sterility* 103, 1037–1042. doi:10.1016/j.fertnstert.2014.12.118

Ko, D. S., Cho, J. W., Park, S. Y., Kim, J. Y., Koong, M. K., Song, I. O., et al. (2010). Clinical Outcomes of Preimplantation Genetic Diagnosis (PGD) and Analysis of Meiotic Segregation Modes in Reciprocal Translocation Carriers. *Am. J. Med. Genet. A.* 152A, 1428–1433. doi:10.1002/ajmg.a.33368

Lim, C. K., Cho, J. W., Song, I. O., Kang, I. S., Yoon, Y.-D., and Jun, J. H. (2008). Estimation of Chromosomal Imbalances in Preimplantation Embryos from Preimplantation Genetic Diagnosis Cycles of Reciprocal Translocations with or without Acrocentric Chromosomes. *Fertil. Sterility* 90, 2144–2151. doi:10.1016/j.fertnstert.2007.10.035

Lin, L., Chen, X., Wang, J., Li, R., Ding, C., Cai, B., et al. (2021). Effect of Carriers' Sex on Meiotic Segregation Patterns and Chromosome Stability of Reciprocal Translocations. *Reprod. BioMedicine Online* 43, 1011–1018. doi:10.1016/j.rbmo.2021.08.017

Lledo, B., Ortiz, J. A., Morales, R., Ten, J., De La Fuente, P. E., Garcia-Ochoa, C., et al. (2010). The Paternal Effect of Chromosome Translocation Carriers Observed from Meiotic Segregation in Embryos. *Hum. Reprod.* 25, 1843–1848. doi:10.1093/humrep/deq111

Morin, S. J., Eccles, J., Iturriaga, A., and Zimmerman, R. S. (2017). Translocations, Inversions and Other Chromosome Rearrangements. *Fertil. Sterility* 107, 19–26. doi:10.1016/j.fertnstert.2016.10.013

Munné, S. (2005). Analysis of Chromosome Segregation during Preimplantation Genetic Diagnosis in Both Male and Female Translocation Heterozygotes. *Cytogenet. Genome Res.* 111, 305–309. doi:10.1159/000086904

Munné, S., Morrison, L., Fung, J., Márquez, C., Weier, U., Bahçe, M., et al. (1998). Spontaneous Abortions Are Reduced after Preconception Diagnosis of Translocations. *J. Assist. Reprod. Genet.* 15, 290–296. doi:10.1023/a:1022544511198

Nielsen, J., and Wohlert, M. (1991). Chromosome Abnormalities Found Among 34910 Newborn Children: Results from a 13-year Incidence Study in �rhus, Denmark. *Hum. Genet.* 87, 81–83. doi:10.1007/bf01213097

Pierce, K., Fitzgerald, L. M., Seibel, M. M., and Zilberstein, M. (1998). Preimplantation Genetic Diagnosis of Chromosome Balance in Embryos from a Patient with a Balanced Reciprocal Translocation. *Mol. Hum. Reprod.* 4, 167–172. doi:10.1093/molehr/4.2.167

Scriven, P. N., Handyside, A. H., and Ogilvie, C. M. (1998). Chromosome Translocations: Segregation Modes and Strategies for Preimplantation Genetic Diagnosis. *Prenat. Diagn.* 18, 1437–1449. doi:10.1002/(sici)1097-0223(199812)18:13<1437::aid-pd497>3.0.co;2-p

Stern, C., Pertile, M., Norris, H., Hale, L., and Baker, H. W. G. (1999). Chromosome Translocations in Couples with *In-Vitro* Fertilization Implantation Failure. *Hum. Reprod.* 14, 2097–2101. doi:10.1093/humrep/14.8.2097

Tan, Y. Q., Tan, K., Zhang, S. P., Gong, F., Cheng, D. H., Xiong, B., et al. (2013). Single-nucleotide Polymorphism Microarray-Based Preimplantation Genetic Diagnosis Is Likely to Improve the Clinical Outcome for Translocation Carriers. *Hum. Reprod.* 28, 2581–2592. doi:10.1093/humrep/det271

Tan, Y., Yin, X., Zhang, S., Jiang, H., Tan, K., Li, J., et al. (2014). Clinical Outcome of Preimplantation Genetic Diagnosis and Screening Using Next Generation Sequencing. *GigaSci* 3, 30. doi:10.1186/2047-217x-3-30

Tease, C., Hartshorne, G. M., and Hultén, M. A. (2002). Patterns of Meiotic Recombination in Human Fetal Oocytes. *Am. J. Hum. Genet.* 70, 1469–1479. doi:10.1086/340734

Wang, J., Li, D., Xu, Z., Diao, Z., Zhou, J., Lin, F., et al. (2019). Analysis of Meiotic Segregation Modes in Biopsied Blastocysts from Preimplantation Genetic Testing Cycles of Reciprocal Translocations. *Mol. Cytogenet.* 12, 11. doi:10.1186/s13039-019-0423-7

Wang, L., Shen, J., Cram, D. S., Ma, M., Wang, H., Zhang, W., et al. (2017). Preferential Selection and Transfer of Euploid Noncarrier Embryos in Preimplantation Genetic Diagnosis Cycles for Reciprocal Translocations. *Fertil. Sterility* 108, 620–627. doi:10.1016/j.fertnstert.2017.07.010

Warburton, D. (1991). De Novo balanced Chromosome Rearrangements and Extra Marker Chromosomes Identified at Prenatal Diagnosis: Clinical Significance and Distribution of Breakpoints. *Am. J. Hum. Genet.* 49, 995–1013.

Xie, P., Hu, L., Tan, Y., Gong, F., Zhang, S., Xiong, B., et al. (2019). Retrospective Analysis of Meiotic Segregation Pattern and Interchromosomal Effects in Blastocysts from Inversion Preimplantation Genetic Testing Cycles. *Fertil. Sterility* 112, 336–342. doi:10.1016/j.fertnstert.2019.03.041

Xu, J., Zhang, Z., Niu, W., Yang, Q., Yao, G., Shi, S., et al. (2017). Mapping Allele with Resolved Carrier Status of Robertsonian and Reciprocal Translocation in Human Preimplantation Embryos. *Proc. Natl. Acad. Sci. U S A.* 114, E8695–E8702. doi:10.1073/pnas.1715053114

Ye, Y., Qian, Y., Xu, C., and Jin, F. (2012). Meiotic Segregation Analysis of Embryos from Reciprocal Translocation Carriers in PGD Cycles. *Reprod. BioMedicine Online* 24, 83–90. doi:10.1016/j.rbmo.2011.08.012

Zhang, C., Zhang, C., Chen, S., Yin, X., Pan, X., Lin, G., et al. (2013). A Single Cell Level Based Method for Copy Number Variation Analysis by Low Coverage Massively Parallel Sequencing. *PLoS One* 8, e54236. doi:10.1371/journal.pone.0054236

Zhang, L., Jiang, W., Zhu, Y., Chen, H., Yan, J., and Chen, Z.-J. (2019a). Effects of a Carrier's Sex and Age on the Segregation Patterns of the Trivalent of Robertsonian Translocations. *J. Assist. Reprod. Genet.* 36, 1963–1969. doi:10.1007/s10815-019-01534-6

Zhang, L., Wei, D., Zhu, Y., Jiang, W., Xia, M., Li, J., et al. (2019b). Interaction of Acrocentric Chromosome Involved in Translocation and Sex of the Carrier Influences the Proportion of Alternate Segregation in Autosomal Reciprocal Translocations. *Hum. Reprod.* 34, 380–387. doi:10.1093/humrep/dey367

Zhang, S., Lei, C., Wu, J., Sun, H., Zhou, J., Zhu, S., et al. (2018). Analysis of Segregation Patterns of Quadrivalent Structures and the Effect on Genome Stability during Meiosis in Reciprocal Translocation Carriers. *Hum. Reprod.* 33, 757–767. doi:10.1093/humrep/dey036

Zhang, S., Luo, K., Cheng, D., Tan, Y., Lu, C., He, H., et al. (2016). Number of Biopsied Trophectoderm Cells Is Likely to Affect the Implantation Potential of Blastocysts with Poor Trophectoderm Quality. *Fertil. Sterility* 105, 1222–1227. doi:10.1016/j.fertnstert.2016.01.011

Zhou, S., Cheng, D., Ouyang, Q., Xie, P., Lu, C., Gong, F., et al. (2018). Prevalence and Authenticity of De-novo Segmental Aneuploidy (>16 Mb) in Human Blastocysts as Detected by Next-Generation Sequencing. *Reprod. BioMedicine Online* 37, 511–520. doi:10.1016/j.rbmo.2018.08.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fgene.2022.905395

WDR36 Safeguards Self-Renewal and Pluripotency of Human Extended Pluripotent Stem Cells

Shiyu An[1†], Dan Yao[2,3†], Wenyi Zhang[1], Hao Sun[1], Tianyi Yu[4], Ruizhe Jia[2*] and Yang Yang[1*]

[1]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, China, [2]Department of Obstetrics, Women's Hospital of Nanjing Medical University, Nanjing Maternity and Child Health Care Institute, Nanjing, China, [3]Fourth Clinical Medicine College, Nanjing Medical University, Nanjing, China, [4]Department of Gynecology and Obstetrics, Affiliated Zhongda Hospital, Medical School, Southeast University, Nanjing, China

OPEN ACCESS

Edited by:
Yan Yun,
University of California, Davis,
United States

Reviewed by:
Baojiang Wu,
Inner Mongolia University, China
Wei Jiang,
Wuhan University, China

****Correspondence:***
Yang Yang
yangyang11@njmu.edu.cn
Ruizhe Jia
jiaruizhe2016@163.com

†These authors have contributed equally to this work and share the first authorship

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *27 March 2022*
Accepted: *30 May 2022*
Published: *22 July 2022*

Citation:
An S, Yao D, Zhang W, Sun H, Yu T, Jia R and Yang Y (2022) WDR36 Safeguards Self-Renewal and Pluripotency of Human Extended Pluripotent Stem Cells.
Front. Genet. 13:905395.
doi: 10.3389/fgene.2022.905395

Extended pluripotent stem cells (EPS cells) have unlimited self-renewal ability and the potential to differentiate into mesodermal, ectodermal, and endodermal cells. Notably, in addition to developing the embryonic (Em) lineages, it can also make an effective contribution to extraembryonic (ExEm) lineages both *in vitro* and *in vivo*. However, multiple mysteries still remain about the underlying molecular mechanism of EPS cells' maintenance and developmental potential. WDR36 (WD Repeat Domain 36), a protein of 105 kDa with 14 WD40 repeats, which may fold into two β-propellers, participates in 18sRNA synthesis and P53 stress response. Though WDR36 safeguards mouse early embryonic development, that is, homozygous knockout of WDR36 can result in embryonic lethality, what role does WDR36 plays in self-renewal and differentiation developmental potential of human EPS cells is still a subject of concern. Here, our findings suggested that the expression of WDR36 was downregulated during human hEPS cells lost self-renewal. Through constructing inducible knockdown or overexpressing WDR36-human EPS cell lines, we found that WDR36 knockdown disrupted self-renewal but promoted the mesodermal differentiation of human EPS cells; however, overexpressing of WDR36 had little effect. Additionally, P53 inhibition could reverse the effects of WDR36 knockdown, on both self-renewal maintenance and differentiation potential of human EPS cells. These data implied that WDR36 safeguards self-renewal and pluripotency of human EPS cells, which would extend our understanding of the molecular mechanisms of human EPS cells' self-renewal and differentiation.

Keywords: self-renewal, differentiation, WDR36, p53, hEPS cells

INTRODUCTION

Pluripotent stem cells (PSCs), including embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), have an unlimited self-renewal capacity and the potential to differentiate into cell types representing the three embryonic germ layers—mesoderm, ectoderm, and endoderm (Ying et al., 2008). Due to these properties, PSCs, an excellent system for modeling embryonic development, occurrence, and development of disease, may contribute to the development of cell-replacement therapies.

Previous work has shown that the self-renewal of PSCs is coordinated by multiple transcription factors and RNA regulators (Huang et al., 2015; Abu-Dawud et al., 2018). Mouse, human, and rat

PSCs are regulated by a common subset of transcription factors for "stemness," among which OCT4, SOX2, and NANOG are constitutive core pluripotency factors (Boyer et al., 2005). OCT4-deficient embryos are viable at the morula stage but cannot form the intact inner cell mass (ICM) *in vivo* and ESC colony *in vitro*, suggesting that OCT4 plays an important role in the maintenance of ESCs (Nichols et al., 1998). SOX2 is critical for PSC self-renewal and pluripotency, and it has been shown that knockdown or conditional deletion of SOX2 leads to trophoblast differentiation (Masui et al., 2007). NANOG is a protein

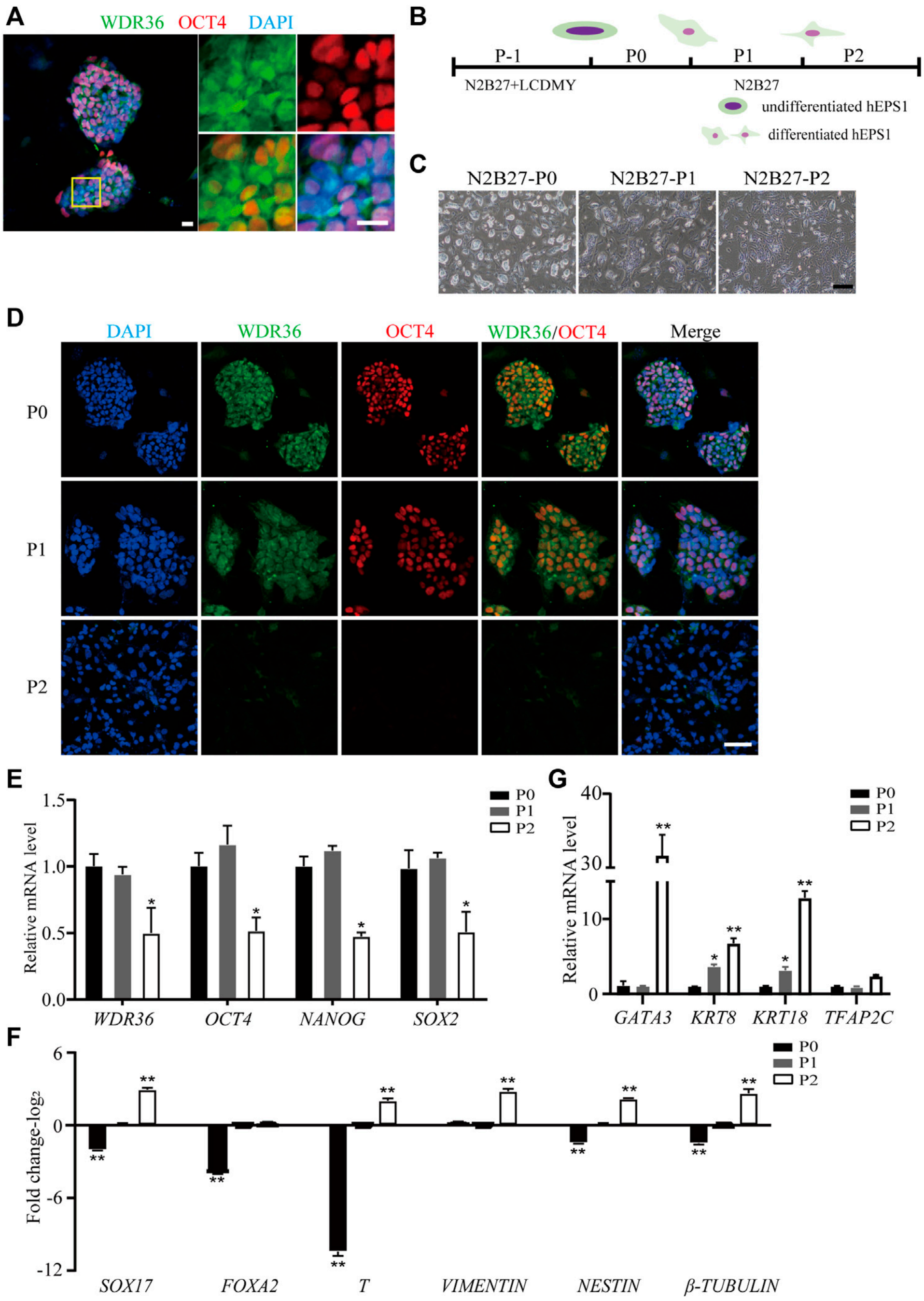


FIGURE 1 | Down-expression of WDR36 during hEPS1 cell differentiation. **(A)** Expression of WDR36 protein in hEPS1 cells. Bars = 20 μm. **(B)** Flow chart of hEPS1 cells in a spontaneous differentiation model. **(C,D)** Representative brightfield [**(C)**, Bar = 200 μm] and immunofluorescent images [**(D)**, Bar = 50 μm] of hEPS1 cells in the spontaneous differentiation assay at P0, P1, and P2, separately. **(E–G)** Expression of pluripotent genes **(E)**, embryonic germ layer-related genes **(F),** and extraembryonic differentiation-related genes **(G)** in hEPS1 cells cultured in N2B27 medium at P0, P1, and P2. *n* = 3 experiments; mean ± S. D; two-tailed Student's *t*-test. *, $0.01 < p < 0.05$; **, $p < 0.01$; no labeling indicates no statistical significance.

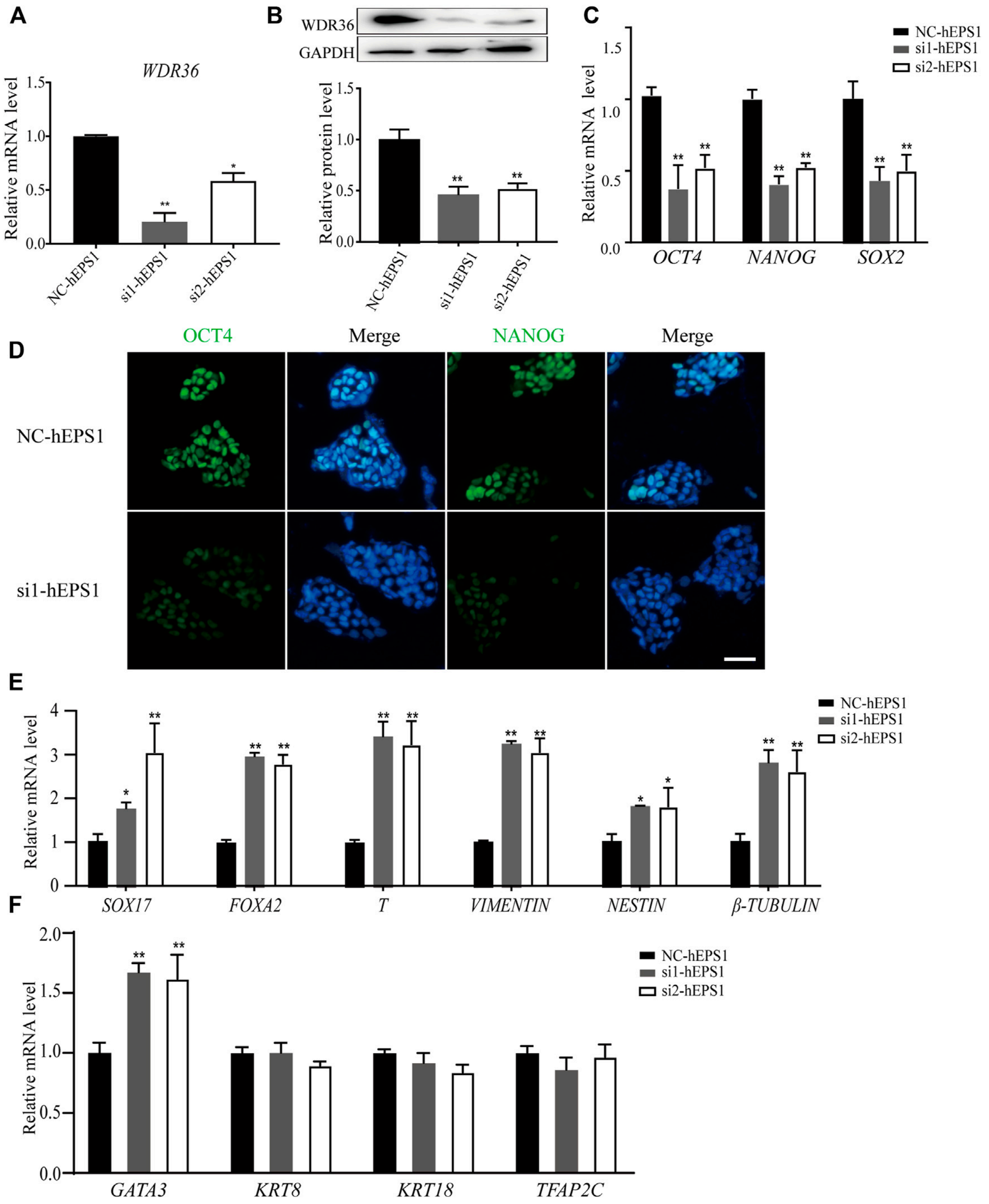


FIGURE 2 | siRNA-WDR36 would impair self-renewal of hEPS1 cells. **(A,B)** WDR36 expression level in hEPS1 cells which had been transfected with WDR36 siRNA was determined using qRT-PCR **(A)** and western blotting **(B)**. **(C)** Expression of *OCT4*, *NANOG* and *SOX2* mRNA in hEPS1 cells transfected with WDR36 siRNA. **(D)** Immunofluorescent images of *OCT4* and *NANOG* in hEPS1 cells with WDR36 siRNA. Bar = 50 μm. **(E,F)** Expression of three embryonic germ layer related genes **(E)** and extraembryonic differentiation related genes **(F)** in siRNA-WDR36 hEPS1 cells. $n = 3$ experiments; mean ± S.D; two-tailed Student's t-test. *, $0.01 < p < 0.05$; **, $p < 0.01$; no labeling indicates no statistical significance. NC-hEPS1: hEPS1 transfected with siRNA-scramble; si1-hEPS1 and si2-hEPS1: hEPS1 transfected with small interference RNA 1 or RNA 2 of WDR36 respectively.

containing a homologous structural domain that interacts with SOX2 and OCT4 to establish the stemness of PSCs. Overexpression of NANOG in mouse PSCs will promote self-renewal and stabilize the undifferentiated state by establishing endogenic self-renewal, which is independent of growth factors or small molecules (Chambers et al., 2003), whereas in human PSCs it allows feeder-free propagation for multiple passages (Darr et al., 2006). Recent studies have identified many additional transcription factors in the regulatory network of PSCs. Importantly, many of these self-renewal-related factors work together to maintain pluripotency. These factors also act as hubs between external signaling pathways and internal determinants of pluripotency (Huang et al., 2015). OCT4, SOX2, and NANOG also occupy repressed genes which encode cell lineage-specific regulators, and the strict control of these genes is essential for PSCs to maintain a stable pluripotent state or to undergo normal differentiation. For example, the precise expression level of OCT4 regulates three distinct fates of PSCs: a less than two-fold increase in OCT4 expression results in differentiation toward primitive endoderm and mesoderm. In contrast, the repression of OCT4 induces loss of pluripotency and commitment to trophectoderm (Niwa et al., 2000). Furthermore, a slowdown of the cell cycle also aids differentiation (Li and Kirschner, 2014), indicating an integral role of the cell cycle in the maintenance of pluripotency and differentiation.

More recently, Yang et al. (2017) have developed a culture condition that endowed canonical human PSCs with totipotent-like features, thereafter named extended pluripotent stem (hEPS) cells, which can efficiently contribute to both embryonic (Em) and extraembryonic (ExEm) lineages not only *in vitro* but also *in vivo*. Most notably, some laboratories have successfully generated blastocyst-like structures (EPS-blastoids), which resemble blastocysts in morphology and cell-lineage allocation and recapitulate key morphogenetic events during pre-implantation and early post-implantation development *in vitro* (Fan et al., 2021; Sozen et al., 2021), by using human EPS through lineage segregation and self-organization. However, little is known about the underlying molecular mechanism of EPS maintenance and developmental potential.

The WD40 repeat (WDR) domain is one of the most abundant protein interaction domains in the human proteome, with over 360 domains annotated to date. These structural domains are important subunits of multiprotein complexes involved in multiple signaling pathways, such as DNA damage sensing and repair, epigenetic regulation of gene expression and chromatin organization, ubiquitin signaling and protein degradation, cell cycle, and immune-related pathways (Schapira et al., 2017). WDR36, a 105-kDa protein with 14 WD40 repeats which may fold into two β-propellers (Footz et al., 2009), is expressed in multiple tissues, with significant mRNA organization in the heart, skeletal muscle, pancreas, liver, and placenta (Monemi et al., 2005). The primary structure of WDR36 protein is similar to that of Utp21, a nucleolar ribonucleoprotein (Kressler et al., 2010), which might explain why the depletion of WDR36 in human cells leads to delay of 18S rRNA processing, disruption of nucleolar morphology, and activation of the P53 stress response pathway (Gallenberger et al., 2011). The tumor suppressor gene *P53* is known to be described as the "guardian of the genome" because it can protect cells from tumor transformation. In addition to being involved in the regulation of DNA repair, apoptosis, and senescence, it also supervises processes such as self-renewal and differentiation of stem cells (Lin et al., 2005; Qin et al., 2007; Chen, 2016; Koifman et al., 2019), as well as iPSC reprogramming (Marion et al., 2009; Utikal et al., 2009). WDR5, another member having the WD40 repeat domain, was confirmed to regulate P53 stability and directly interact with P53 during ESC specification. It could also interact with OCT4, CTCF, or lncRNA to facilitate iPSC reprogramming and maintain ESC identity (Ang et al., 2011; Li et al., 2020).

Accordingly, we wonder whether WDR36 plays a pivotal role in the self-renewal and differentiation potential of hEPS cells. Through constructing inducible knockdown or overexpressing WDR36-hEPS cell lines, we found that WDR36 regulated the self-renewal of hEPS cells, providing a new regulatory target. Moreover, we determined that WDR36-knockdown promoted the differentiation of human EPS cells, while overexpressing of WDR36 had little effect. Additionally, P53 inhibition could reverse the effects of WDR36 knockdown on hEPS cells.

MATERIALS AND METHODS

Cell Culture

Human EPS cells (hEPS1 and hEPS2, two cell lines of hEPS) were cultured in the serum-free N2B27-LCDMY medium under 20% O_2 and 5% CO_2 at 37°C. The N2B27 medium was prepared as follows: 1:1 mixture of DMEM/F12 (Thermo Fisher Scientific, 11330-032) and Neurobasal (Thermo Fisher Scientific, 21103-049), 0.5X N2 supplement (Thermo Fisher Scientific, 17502-048), 0.5X B27 supplement (Thermo Fisher Scientific, 12587-010), 1% nonessential amino acids (Thermo Fisher Scientific, 11140-050), 1% GlutaMAX™ (Thermo Fisher Scientific, 35050-061), 0.1 mM β-mercaptoethanol (Thermo Fisher Scientific, 21985-023), and 1% penicillin–streptomycin (Thermo Fisher Scientific, 15140-122). To prepare the N2B27-LCDMY medium, small molecules and cytokines were added to the N2B27 medium as follows: 10 ng/ml recombinant human LIF (L, 10 ng/ml; StemImmune, EST-LIF-0100), CHIR 99021 (C, 1 μM; MCE, HY-10182), (S)-(+)-dimethindene maleate (D, 2 μM; Tocris, 1425), minocycline hydrochloride (M, 2 μM; Santa Cruz Biotechnology, sc-203339), and Y-27632 (2 μM; Hanxiang, 17109). Human EPS cells were cultured on mitomycin C (Sigma-Aldrich, M4287)-inactivated mouse embryonic fibroblast (MEF) feeder cells (3×10^4 cells per cm^2). The N2B27-LCDMY medium was changed daily. Human EPS cells were passaged by single-cell trypsin digestion (0.05% trypsin-EDTA, Thermo Fisher Scientific, 25300-062) every 3 days (normally at a split ratio of 1:3–1:6).

Establishment of Doxycycline-Inducible WDR36-Modified hEPS Cell Lines

We first generated the WDR36 short hairpin RNA (shRNA) vector from a PiggyBac TetR puro vector by inserting multiple cloning sites (MCSs) and further cloned human WDR36 shRNA oligonucleotides into MCS (EcoR I and Age I sites), and then these plasmids were confirmed by sequencing. The mutually priming oligonucleotides used in this study are listed in **Supplementary Table S3**. The WDR36 overexpression vector was produced from the PiggyBac TetOn3G tdTomato puro vector by inserting multiple cloning sites (MCSs) and further cloned human WDR36 cDNA into MCS (EcoR I and Age I sites) to generate the TRE3G-WDR36-T2A-tdTomato plasmid.

By using the Lipofectamine Stem Transfection Reagent (Thermo Scientific, STEM00001) according to the manufacturer's instruction, the WDR36 shRNA and pCyL43 (PiggyBac transposase) vectors were co-transfected into hEPS1 and hEPS2 cells separately to make the inducible shWDR36-hEPS cell lines. To generate the inducible overexpression WDR36-hEPS cell line, TRE3G-WDR36-T2A-tdTomato and pCyL43 vectors were co-transfected into hEPS1 cells. Puromycin (0.25 μg/ml) was added for positive clone selection for 7 days, and the expanded clones were designated as the overexpressed or knockdown hEPS cells, respectively. The gene expression level of WDR36 was verified by quantitative RT-PCR (qRT-PCR) and Western blot analysis.

Immunofluorescence Staining

Cells were fixed with 4% paraformaldehyde for 20 min at room temperature (RT) and then permeabilized and blocked with PBS containing (vol/vol) 0.25% Triton X-100 (Beyotime, P0096) together with (vol/vol) 2.5% donkey serum (Jackson ImmunoResearch, 017-000-121) for 50 min at RT. The primary antibody (information about antibodies is listed in **Supplementary Table S1**) with the blocking solution was diluted, and the sample was incubated for at least 12 h at 4°C. Subsequently, the samples were washed five times with PBS for 3 min each and then incubated at RT with the secondary antibodies diluted in 2.5% donkey serum (information about antibodies is listed in **Supplementary Table S1**). After 1 h, the samples were washed three times with PBS, and the nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI) (YIFEIXUE BIO TECH, YD0020-10). We observed the results using confocal microscopy (ZEISS LSM700, Germany).

RNA Isolation and qRT-PCR Analysis

Total RNA was extracted from cells using the TRIzol Reagent (Invitrogen 15596-026) as described by the manufacturer's instructions, and then RNA was reverse-transcribed to cDNA using a PrimeScript™ RT reagent kit with gDNA Eraser (Takara, Dalian, China). To analyze the relative expression levels of genes in the cultured hEPS1 and hEPS2 cells, the qRT-PCR was performed in a Step One Plus Real-Time PCR System (Applied Biosystems, CA, United States), using the FastStart Universal SYBR Green Master kit (Vazyme, Nanjing, China). The primers were designed with Primer 5 software and are listed in **Supplementary Table S2**. The melting curve of each mRNA was used to evaluate the amplification quality. The expression data were assessed by the $2^{-\triangle\triangle CT}$ method, and the expression level of glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) was used as an endogenous normalization control.

Western Blot Analysis

Total protein was extracted from the cultured hEPS1 and hEPS2 cells with the addition of RIPA buffer (Beyotime, Beijing, China) containing 1 mM phenylmethylsulfonyl fluoride and quantified *via* the BCA method. The protein samples were boiled in water for 10 min and then 30 μg of total protein was electrophoresed for each sample in 8% SDS-polyacrylamide gels, followed by being transferred onto polyvinylidene difluoride membranes (Millipore, MA, United States). After being blocked with 5% non-fat dried milk for 2 h, we incubated the membranes overnight with primary antibodies at 4°C. Then, they were incubated with the secondary antibodies for 1 h at RT. Details of the antibodies are provided in **Supplementary Table S1**. After being washed three times, bands were detected with the enhanced chemiluminescence (ECL) detection kit (TransGen Biotech, Beijing, China). We used ImageJ software (Wayne Rasband, MD, United States) to calculate grayscale values, and GAPDH played as a reference. The experiments were repeated three times.

Embryoid Body (EB) Formation Assay

The embryoid body (EB) formation assay was based on the previous study by Liu et al. (2021). hEPS1 cells were rinsed once and digested with 0.05% trypsin-EDTA (Thermo Fisher Scientific, 25300-062) for 3 min in a 37°C incubator. Then, the cell suspension was collected and centrifuged at 900 rpm for 4 min. After being rinsed with DMEM/F12 and centrifugated, the supernatant was removed, and the colonies were resuspended in the EB medium (DMEM/F12 containing 20% KSR (knockdown serum replacement), 1% GlutaMAX™, 1% nonessential amino acids, and 55 μM β-mercaptoethanol and cultured in low-adherent cell culture dishes. Y27632 was also added into the EB medium during the first 24 h to ensure high cell viability. The EB medium was changed daily and EBs samples were collected on days 2, 4, and 6 for qRT-PCR analysis. For plating, the cell clumps were transferred into DMEM/F12 containing 20% FBS on the gelatin-coated 24-well plate for 5–7 days and fixed for further analysis.

Committed Mesodermal Differentiation

Committed mesodermal differentiation was performed according to the reported protocol by Lian et al. (2013). Approximately 30% confluent of hEPS1 and hEPS2 cells were treated with 5 μM Y-27632 (Hanxiang, 17109) and 6 μM CHIR99021 (MCE, HY-10182) in the mesoderm differentiation induction media containing RPMI1640, 100X B27 (Thermo Fisher Scientific, A1895601) for 4 days. Later, the differentiated cells were fixed and stained for markers of the mesodermal germ layer.

Statistical Analysis

All data were presented as mean values ±standard error of the mean (SEM). Statistical analysis was performed using the two-tailed Student's *t*-test or one-way analysis of variance (ANOVA) and was defined as statistically significant when $p < 0.05$.

RESULTS

WDR36 Expression Was Downregulated During hEPS Cell Differentiation

To understand the role of WDR36 in hEPS cells' maintenance, we first detected the expression of WDR36 in hEPS1 and

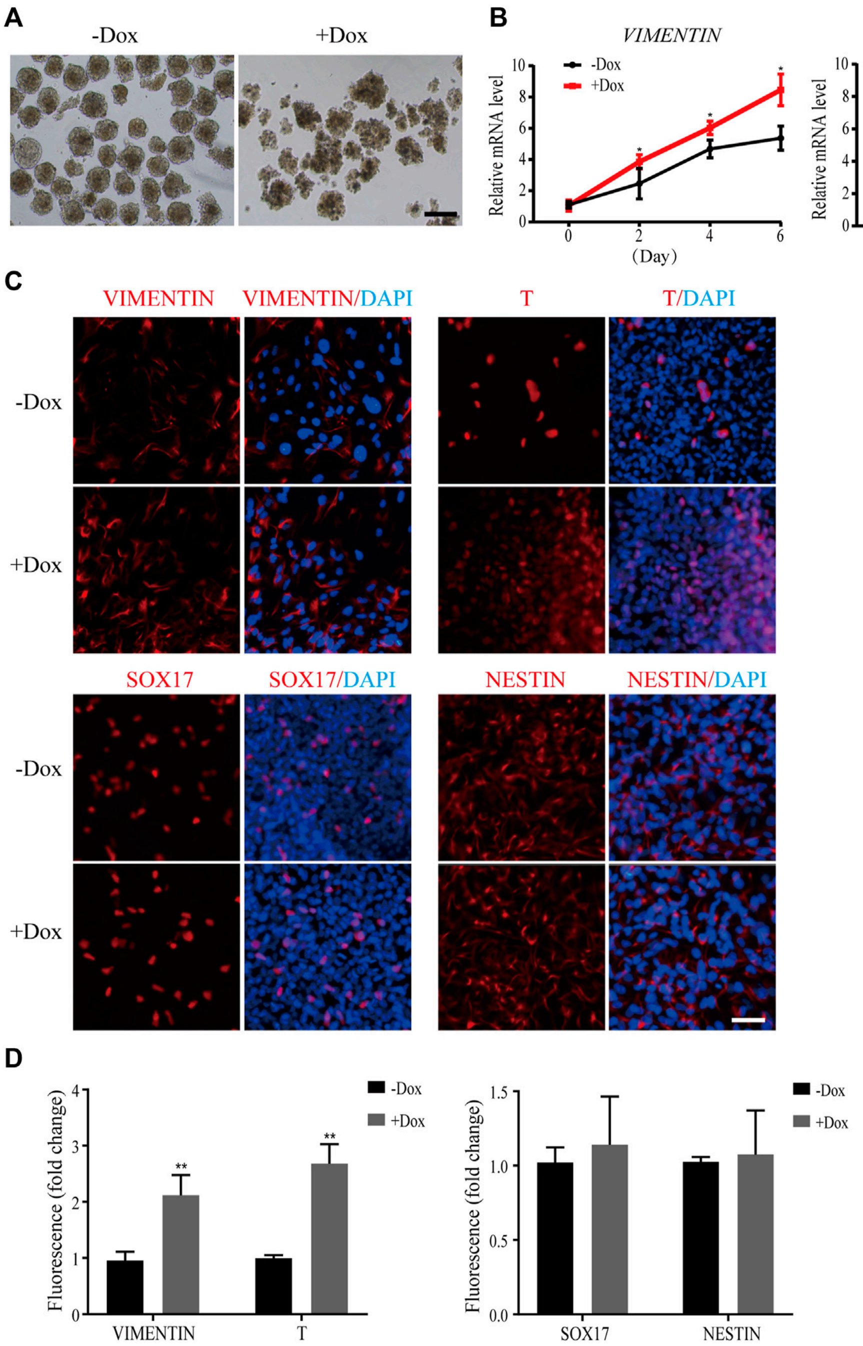


FIGURE 3 | Dox-induced WDR36 knockdown elevated differentiation potential of hEPS1 cells in an EB formation assay. **(A)** Representative images of WDR36-modified hEPS1 cells forming EBs in the EB medium. On day 7, EBs were collected and observed under the microscope. Bar = 200 μm. **(B)** Expression of mesodermal genes in WDR36-modified EBs treated with or without Dox. **(C,D)** Immunofluorescent images **(C)** and quantification intensity for immunofluorescence **(D)** of WDR36-modified EBs treated with or without Dox (2 μg/ml). Bar = 50 μm; $n = 3$ experiments; mean ± S. D; two-tailed Student's *t*-test. *, $0.01 < p < 0.05$; **, $p < 0.01$; no labeling indicates no statistical significance.

hEPS2 cells cultured in the N2B27-LCDMY medium (undifferentiated state). We found that WDR36 is mainly distributed both in the cytoplasm and nucleus, while a few are only in the nucleus of cells (**Figure 1A** and **Supplementary Figure S1A**). When LCDMY was withdrawn, hEPS1 cells initiated to differentiate, losing their dome-shaped clone morphology gradually (**Figures 1B,C**), and WDR36 tended to locate mainly in the cytoplasm of cells, especially at passage 1 (**Figure 1D**). After being exposed to the condition without LCDMY cocktail for two passages, all the hEPS1 cells lost their

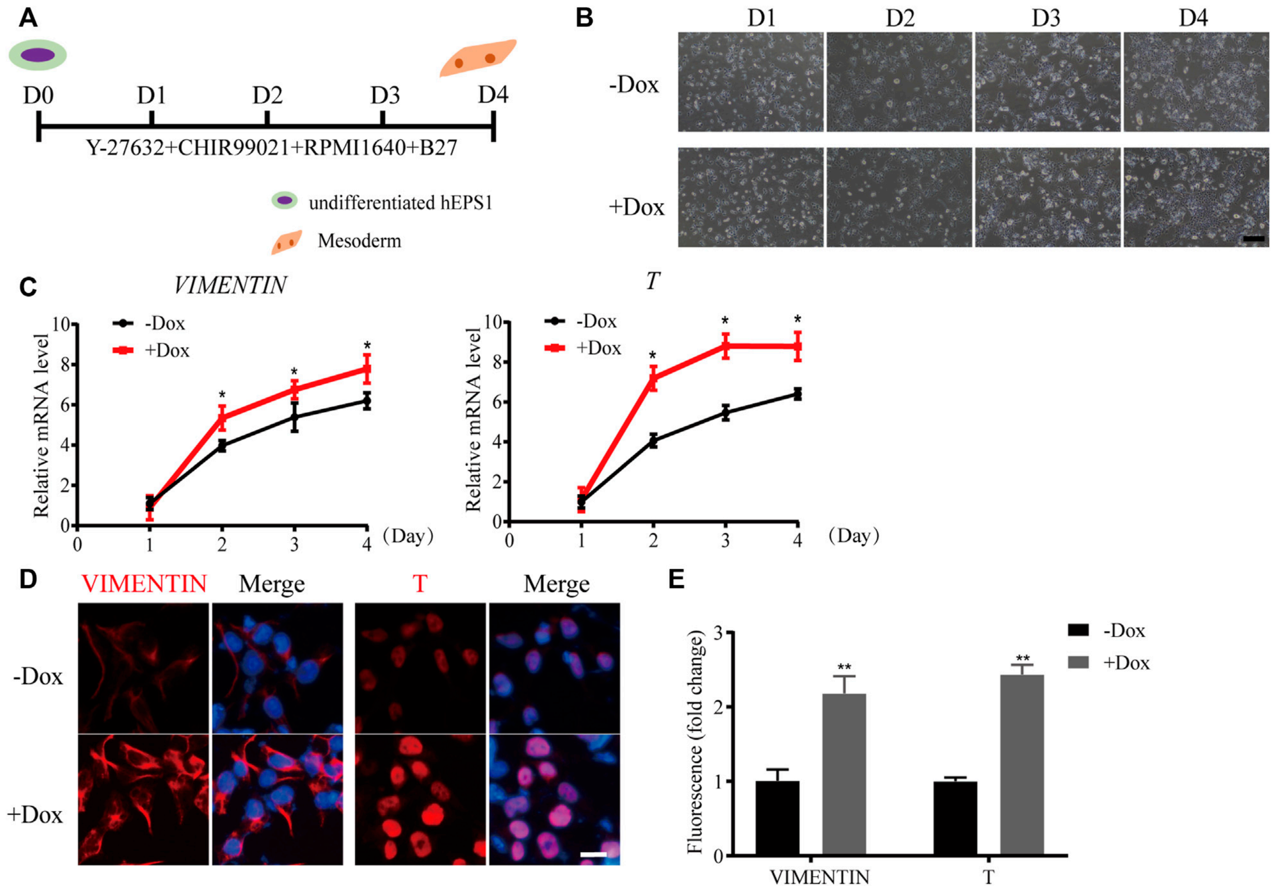


FIGURE 4 | WDR36 knockdown promoted mesodermal differentiation potential of hEPS1 cells in a mesodermal-committed differentiation assay. **(A)** Time scheme of the mesodermal-committed differentiation assay. **(B)** Representative images of Dox-induced shWDR36 hEPS1 cells in mesoderm differentiation medium with or without Dox. Bar = 200 μm. **(C,D)** mRNA **(C)** and protein expression **(D)** of T and vimentin in sh1-hEPS1 cells treated with or without Dox. Bar = 50 μm. **(E)** Quantification intensity levels for immunofluorescence of T and vimentin in sh1-hEPS1 cells treated with or without Dox. *n* = 3 experiments; mean ± S. D; two-tailed Student's *t*-test. *, 0.01 < *p* < 0.05; **, *p* < 0.01; no labeling indicates no statistical significance.

EPS identity with no OCT4 or WDR36 expression (**Figure 1D**). To be similar but slightly different, hEPS2 differentiated more quickly than hEPS1, completely losing self-renewal after passage 1 (**Supplementary Figures S1B,C**). All the hEPS2 cells expressed WDR36 only in their cytoplasm as early as LCDMY being removed for 3 and 4 days (passage 0) and were OCT4- and WDR36-negative at passage 1 (**Supplementary Figures S1C**). Furthermore, data indicated that the expressions of *WDR36*, *NANOG*, *OCT4,* and *SOX2* were decreased during the differentiated process of hEPS1 and hEPS2 at transcriptional levels (**Figure 1E** and **Supplementary Figures S1D**), while the expression of three embryonic germ layer markers—endoderm (*SOX17* and *FOXA2*), mesoderm (*T* and *vimentin*), and ectoderm (*nestin* and *β-tubulin*)—increased (**Figure 1F** and **Supplementary Figures S1E**). Considering hEPS can efficiently contribute to both embryonic and extraembryonic lineages both *in vitro* and *in vivo*, we also detected the expression of extraembryonic differentiation-related genes. The results showed that *GATA3*, *KRT8*, *KRT18*, and *TFAP2C* were significantly increased during hEPS1 and hEPS2 differentiation (**Figure 1G** and **Supplementary Figures S1F**). The aforementioned results showed that the WDR36 expression was downregulated during hEPS cell differentiation.

Inhibition of WDR36 Impaired hEPS Cells' Self-Renewal

Mechanistic studies have identified WDR36 as a functional homolog of yeast Utp21 (Skarie and Link, 2008), which is related to 18S rRNA maturation, and rRNA synthesis plays a role in self-renewal regulation in PSCs. We hypothesized that WDR36 played an impotent role in hEPS cells' self-renewal. The specific small-interference RNAs (siRNAs) for WDR36 (NCBI accession number: NM_139281.2), and siRNA-control (**Supplementary Table S3**) was synthesized and purified at Tsingke (Nanjing, China), and the most efficient two siRNAs were selected as candidates and further confirmed by qRT-PCR and Western blotting experiments (**Figures 2A,B**, and

Supplementary Figure S2A). We noticed robustly reduced expression of *OCT4*, *NANOG*, and *SOX2* in siRNA-WDR36 hEPS1 cells compared with siRNA-control (**Figure 2C**), consistent with the results of immunofluorescence (**Figure 2D**). This assay was performed by using hEPS2, and a similar phenomenon was also observed (**Supplementary Figure S2B**). Moreover, as shown in **Figure 2E**, embryonic germ layer markers (*SOX17*, *FOXA2*, *T*, *vimentin*, *nestin*, and *β-tubulin*) all significantly enhanced in the siRNA-WDR36 groups. In addition, we found that WDR36 inhibition induced an increase in *GATA3* expression, while the expressions of *KRT8*, *KRT18*, and *TFAP2C* did not undergo pronounced changes (**Figure 2F**). Collectively, our findings suggested that WDR36 impaired the self-renewal of hEPS.

Dox-Induced WDR36 Knockdown Elevated Mesodermal Differentiation Potential of hEPS Cells *In Vitro*

To further explore what role WDR36 played in differentiation potency of hEPS, we constructed two Dox-induced WDR36 knockdown sub-cell lines based on hEPS1 (**Supplementary Figure S3A**), namely, sh1-hEPS1 and sh2-hEPS1. After adding Dox, qRT-PCR and protein blot analysis confirmed successful construction; thus, sh1-hEPS1 was selected for subsequent experiments due to the higher silencing efficiency (**Supplementary Figures S3B,C**).

Next, we simulated early human embryonic development by EB formation assay. As shown in **Figure 3A**, the morphology of EBs became irregular when WDR36 was knocked down, indicating that the spontaneous differentiation potential of hEPS1 cells was disrupted. In addition, we collected EBs on days 2, 4, and 6 for mRNA level confirmation by qRT-PCR. We observed a significant decrease in the expressions of *OCT4* and *NANOG* as spontaneous differentiation proceeded (**Supplementary Figure S4A**). Furthermore, the results suggested that WDR36 knockdown promoted mesodermal differentiation potential of hEPS1 cells (**Figure 3B**), whereas no significant change in ectodermal or endodermal differentiation was exhibited, and so did the expressions of extraembryonic specific genes *GATA3* and *KRT8* (**Supplementary Figure S4A**). Then, EBs adhered to Matrigel for 5–7 days for further analysis, and the immunofluorescent results also showed a similar tendency, in comparison to the untreated group (**Figures 3C,D**). The aforementioned evidence indicated that WDR36 had an important role in germ layer differentiation of hEPS1 cells.

Since mesodermal lineage genes increased obviously during sh1-hEPS1 forming EBs, we hypothesized that WDR36 inhibition would promote mesodermal differentiation of hEPS. To address this, we carried out a mesodermal-committed differentiation assay (**Figure 4A**). We found no significant morphological difference between Dox-treated and -untreated sh1-hEPS1 cells during differentiation (**Figure 4B**). However, robustly higher expressions of *T* and vimentin were noted in the Dox-induced group (**Figure 4C**), consistent with the results of immunofluorescence (**Figures 4D,E**). In order to confirm this in hEPS2, we constructed a Dox-induced WDR36 knockdown sub-cell line called sh1-hEPS2 and successfully confirmed the Dox-induced inhibition effect by qRT-PCR and protein blot (**Supplementary Figures S5A,B**). As expected, WDR36 inhibition boosted mesodermal differentiation of hEPS2 (**Supplementary Figures S5C–F**).

WDR36-Overexpression Hardly Affected Self-Renewal and Differentiation of hEPS1 Cells

In addition to Dox-induced shRNA cell lines, we also constructed a Dox-induced WDR36-overexpression hEPS1 cell line to evaluate the effect of WDR36 overexpression on hEPS self-renewal and differentiation (**Supplementary Figure S6A**). Following Dox treatment, both qRT-PCR and Western blot results confirmed the successful construction (**Supplementary Figures S6B,C**). We then asked whether WDR36-overexpression affected the self-renewal and differentiation of hEPS1. We observed slight upregulation of *OCT4* and *NANOG*, but no statistical difference was observed (**Supplementary Figure S6D**). Additionally, immunofluorescent results showed that there was no difference in the protein levels (**Supplementary Figure S6E**).

We further examined whether WDR36-overexpression would affect the differentiation potential of hEPS1 by EB formation assay. As shown in **Supplementary Figure S7A**, no significant difference in EB morphology was observed in Dox-induced WDR36-overexpressing hEPS1 cells. The qRT-PCR results showed that as spontaneous differentiation proceeded, pluripotent genes *OCT4* and *NANOG* decreased (**Supplementary Figure S7B**), while differentiated genes enhanced (**Supplementary Figure S7C**). During this procedure, no apparent difference was found between Dox-treated and -untreated groups (**Supplementary Figures S7C,D**). Thus, we indicated that neither self-renewal nor differentiation potential would be affected by overexpression of WDR36.

P53 Inhibition Could Reverse the Effect of WDR36 Inhibition on hEPS1 Cells

Multiple studies have reported that tumor suppressor P53 plays an active role in promoting differentiation and opposing self-renewal of human ESCs (Jain et al., 2012), and loss of WDR36 causes an activation of the P53 stress-response pathway (Skarie and Link, 2008). To determine whether P53 inhibition could recover the effect of WDR36 inhibition on hEPS1 cells, we first checked whether P53 had similar effects on hEPS1 cells compared to traditionally primed human PSCs. As expected, *P53* was activated when LCDMY was withdrawn from the culture condition, which resulted in differentiation. Meanwhile, the expressions of *P21*, *MDM2*, and *BAX* were also increased but not so significant (**Figure 5A**), which was in accordance with Lin et al. (2005). Furthermore, the expressions of *P53*, *P21*, *MDM2*, and *BAX* in siRNA-WDR36 hEPS1 cells were significantly increased compared with those of siRNA-control cells

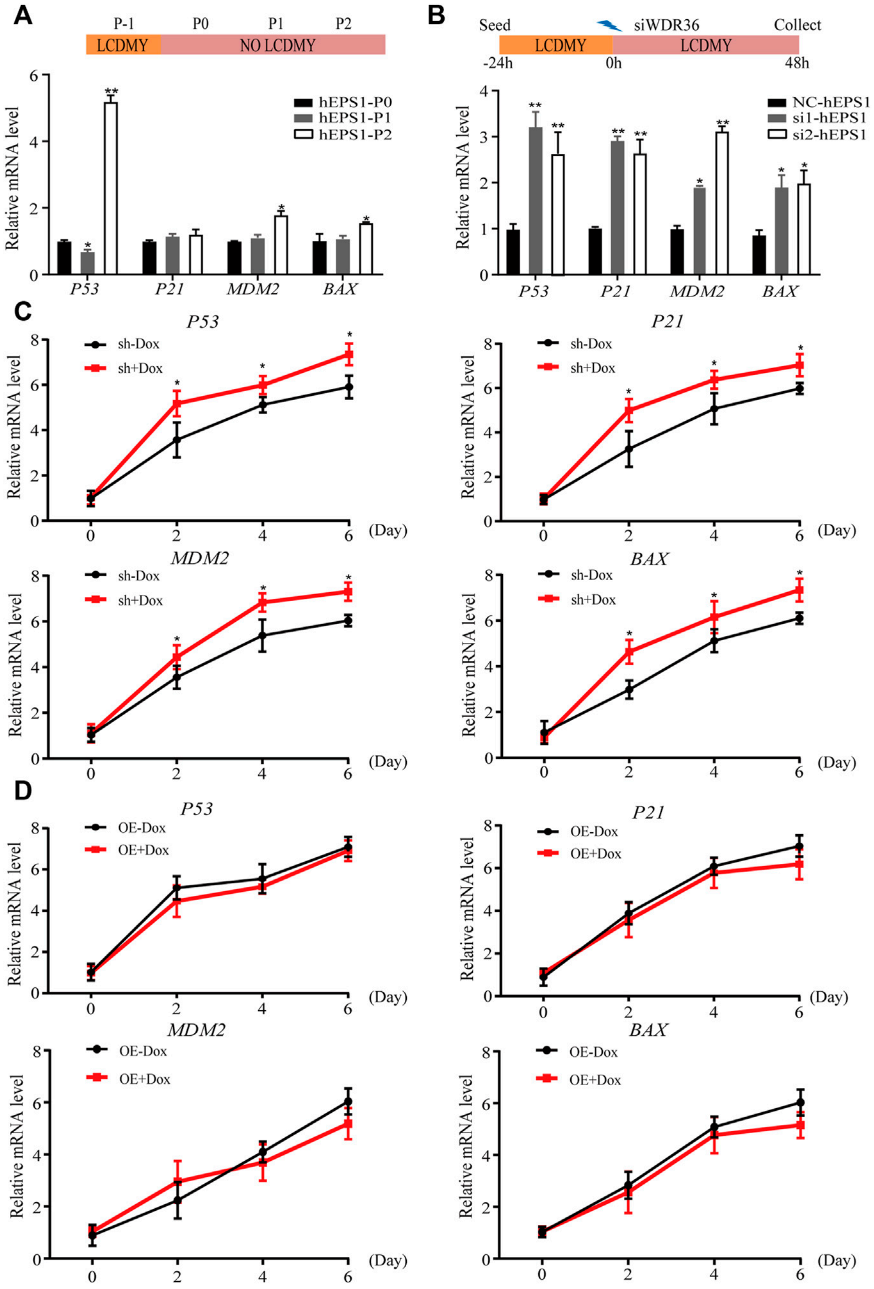


FIGURE 5 | Expression of P53 pathway-related genes during the spontaneous differentiation procedure of hEPS1 cells. **(A)** Expression of *P53*, *P21*, *MDM2*, and *BAX* in hEPS1 cells which were cultured in the N2B27 medium at P0, P1, and P2. **(B)** Expression of *P53*, *P21*, *MDM2*, and *BAX* in hEPS1 cells with WDR36 siRNA. NC: siRNA-scramble; si1: small interference RNA 1 of WDR36; si2: small interference RNA 2 of WDR36. **(C)** Expressions of *P53*, *P21*, *MDM2*, and *BAX* in Dox-induced WDR36 knockdown EBs treated with or without Dox. **(D)** Expressions of *P53*, *P21*, *MDM2*, and *BAX* in Dox-induced WDR36 overexpression EBs treated with or without Dox. OE: overexpression of WDR36. $n = 3$ experiments; mean ± S. D; two-tailed Student's t-test. *, $0.01 < p < 0.05$; **, $p < 0.01$; no labeling indicates no statistical significance.

(**Figure 5B**). We then asked whether similar results were present during the EB differentiation. We observed a significant upward reversion in *P53* expression as expected, and the expressions of *P21*, *MDM2*, and *BAX* also showed an upward trend in WDR36-knockdown hEPS1 cells compared with those of Dox-untreated cells (**Figure 5C**). However, the expressions of *P53*, *P21*, *MDM2*, and *BAX* did not drop down obviously after overexpression of WDR36 (**Figure 5D**).

Thus, we wondered whether inhibition of the P53 pathway could recover the inhibitory effect of WDR36 knockdown by treating hEPS1 cells with pifithrin-alpha (PFT-α), a canonical inhibitor of the P53 pathway (Ulmke et al., 2021). First of all,

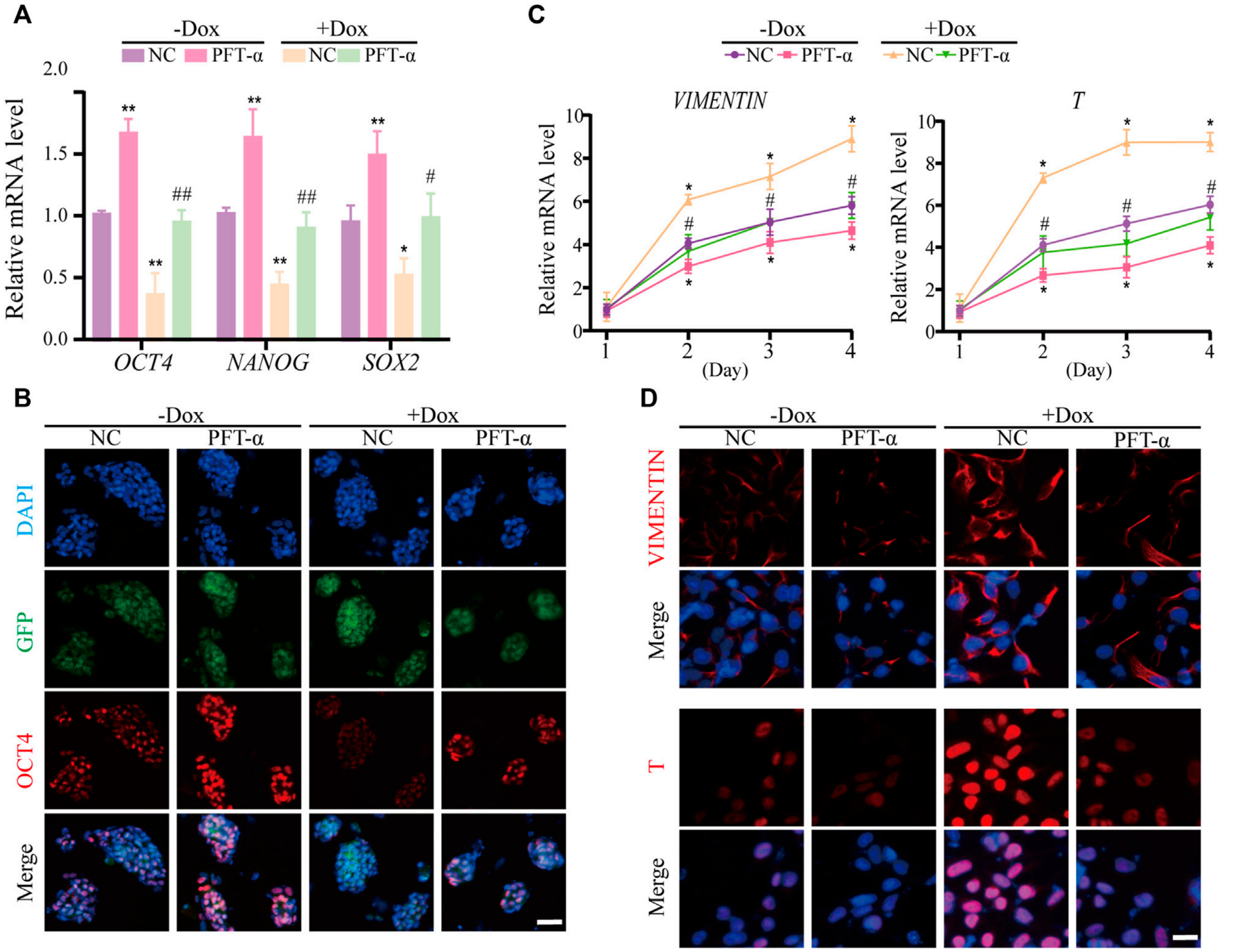


FIGURE 6 | Inhibition of P53 could restore the influences of WDR36 knockdown on hEPS1 cells. **(A)** qRT-PCR analysis of the Dox-induced shWDR36 hEPS1 cells cultured under the condition with or without Dox. **(B)** Immunofluorescent images of OCT4 in Dox-induced shWDR36 hEPS1 cells. **(C)** Expression of T and vimentin in Dox-induced shWDR36 hEPS1 cells. **(D)** Immunofluorescent images of T and vimentin in Dox-induced shWDR36 hEPS1 cells. Bars = 50 μm. NC, culture condition did not contain 10 μM PFT-α. *n* = 3 experiments; mean ± S. D; one-way analysis of variance (ANOVA). *, $0.01 < p < 0.05$; **, $p < 0.01$; #, $0.01 < p < 0.05$; ##, $p < 0.01$; no labeling indicates no statistical significance.

we verified the inhibitory effect of PFT-α on the expression of P53 and P21 in hEPS1 cells at different concentrations. The qRT-PCR and Western blot results indicated that the expressions of P53 and P21 were effectively inhibited in 10 μM of PFT-α treatment and did not affect the survival of hEPS1 cells (**Supplementary Figures S8A,B**). Therefore, we chose 10 μM PFT-α for subsequent experiments. As shown in **Supplementary Figures S8C,D**, PFT-α treatment induced notable improvement in the expressions of pluripotent genes *OCT4*, *NANOG*, and *SOX2*, which was consistent with the immunofluorescent results. Furthermore, we also found that PFT-α treatment could impair the commitment of hEPS1 cells into mesodermal lineage (**Supplementary Figures S8E,F**). These results implied that P53 played an important role in the differentiation of hEPS1 cells.

Previous studies have shown that P53 inhibits NANOG expression and downregulates stemness of stem cells (Lin et al., 2005), and loss of WDR36 leads to activation of the P53 stress-response pathway (Skarie and Link, 2008), suggesting that defects in the P53 pathway may affect the regulation of hEPS1 cells by WDR36. Thus, we examined whether PFT-α could reverse the effects of WDR36 knockdown on the self-renewal and mesoderm differentiation potential of hEPS1 cells. The qRT-PCR and immunofluorescence staining revealed that WDR36 knockdown significantly inhibited the self-renewal of hEPS1 cells, while PFT-α treatment could rescue the effects (**Figures 6A,B**). Furthermore, we reconfirmed the effect of PFT-α in a committed mesodermal differentiation assay. The data showed that WDR36 knockdown significantly promoted the mesoderm differentiation potential of hEPS1 cells. However, when hEPS1 cells were co-treated with WDR36 shRNA and PFT-α, the mesoderm differentiation potential was attenuated (**Figures 6C,D**). These results showed that knocking-down WDR36 significantly regulated hEPS1 cells' self-renewal and differentiation potential, and inhibition of P53, in turn, restored these effects.

DISCUSSION

PSCs are capable of self-renewal and can differentiate into different cell types under certain differentiation conditions. Their self-renewal and differentiation ability is precisely regulated by various regulatory factors at the transcriptional, translational, and post-translational levels (Liou et al., 2017). The maintenance of pluripotency, the commitment to a specific lineage fate, and the switch to cell differentiation depend on the tight regulation of protein synthesis and ribosome biogenesis (Saba et al., 2021). The same is true of hEPS cells' self-renewal and differentiation regulation. WDR36 is a nucleolar protein involved in the maturation of 18S rRNA. The primary structure of the WDR36 protein is similar to that of Utp21 (Footz et al., 2009), an essential nucleolar ribonucleoprotein. It is also essential for early mouse embryonic development, as homozygous WDR36-deficient mouse embryos die before reaching the blastocyst stage (Gallenberger et al., 2011). In addition, WDR5, which has a similar structure to WDR36, can directly interact with the pluripotent transcription factor OCT4 to regulate ESC self-renewal and differentiation (Ang et al., 2011). However, whether WDR36-mediated ribosome biogenesis or other biological events can regulate the self-renewal and differentiation of hEPS cells remain unknown.

In this study, the WDR36 protein localized in the nucleus and cytoplasm in undifferentiated hEPS1 and hEPS2 cells (**Figure 1A**, **Supplementary Figure S1A**), which was in line with previous research that WDR36 protein was ubiquitously expressed in the cytoplasm and nucleus of zebrafish embryos. In addition, WDR36 fusions first localized to nucleoli and then became enriched in both the cytoplasm and nucleus of zebrafish embryos as protein levels increased (Skarie and Link, 2008). For WDR36 localized in the cytoplasm, it may be related to the interaction of WDR36 with G protein-coupled receptors (GPCRs). WDR36 acts as a scaffold protein tethering G-protein-coupled receptors, Gαq and PLCβ, in a signaling complex (Cartier et al., 2011; Blankenbach et al., 2020). GPCRs are the largest class of cell surface receptors and therefore mediate a variety of biological processes (Wootten et al., 2018). For example, they are developmental regulators of cell morphology, polarity, and migration (Wettschureck and Offermanns, 2005; Cotton and Claing, 2009). Most importantly, the role of GPCRs in stem cell maintenance is indubitably important, and they contribute to ESCs' self-renewal, pluripotency, and the maintenance of the clonal morphology (Layden et al., 2010). In addition, GPCRs also regulate the mesoderm differentiation of stem cells through lipid-mediated signaling pathways and Wnt signaling pathways (Kobayashi et al., 2010).

To broaden the effect of WDR36 on hEPS cells, we removed the LCDMY cocktail which was essential for hEPS cell maintenance from the N2B27-LCDMY culture condition and cultured cells for two to three passages. In hEPS1, we observed almost no morphological change on day 3 of P0, but the doomed colonies gradually flattened and collapsed later, whereas in hEPS2, we observed morphological changes as early as day 3 at P0, and the doomed colonies gradually flattened and collapsed afterward. After being exposed to the condition without LCDMY cocktail for one or two passages, all the hEPS cells lost their EPS identity with no OCT4 or WDR36 expression. Moreover, the expression of WDR36 and pluripotent transcription factors decreased, along with the embryonic and extraembryonic lineages differentiation genes being activated (**Figures 1E–G** and **Supplementary Figures S1D,F**). In addition, both siRNA-WDR36 and shRNA-WDR36 impaired the self-renewal of hEPS cells (**Figures 2C,D** and **Supplementary Figure S1**), reminding a pivotal role of WDR36 in hEPS cells' self-renewal. Notably, studies have shown that *OCT4* and *NANOG* were significantly downregulated in feeder-free EPS cells, compared with primed ESC, and WDR36 was also downregulated in feeder-free EPS cells, although not significant (Zheng et al., 2021), which was consistent with our results.

Furthermore, the expression of *P53*, *P21*, *MDM2*, and *BAX* increased during hEPS1 cell differentiation, and so they did in WDR36 knockdown hEPS1 cells (**Figures 5A–C**). Previous studies have shown that WD40 repeat proteins are involved in regulating a variety of cellular functions, such as cell division, cell fate determination, and transmembrane signaling (Li and Roberts, 2001). Disruption of WDR36 in HTM-N cells causes apoptosis and upregulation of *P53*, *P21*, and *BAX* expressions (Gallenberger et al., 2011). The same phenomenon also occurs during embryonic development, that is, the loss of WDR36 expression disrupts nucleolus homeostasis and results in significant upregulation of P53 (Skarie and Link, 2008). Indeed, P53 plays a key role in actively promoting the differentiation of human ESCs. Before ESC differentiation, the expression level of P53 is very low in human ESCs, and HDM2 and TRIM24, two negative regulators of P53, trigger P53 degradation. Active P53 in turn promotes the expression of a cell cycle regulator P21, which slows down human ESC by prolonging the cell cycle gap (G1) phase (Jain et al., 2012; Ghatak et al., 2021). In addition, many studies have shown that P53 has many downstream target genes that play regulatory roles, including some non-coding RNAs, such as miR-34a and miR-145, currently being identified as regulated by P53 (Hermeking, 2007; Jain et al., 2012). Recently, Dox-inducible exogenous expression of P21 in human ESCs was shown to induce cell cycle arrest and substantial human ESC differentiation (Ruiz et al., 2011), further supporting that induced expression of P21 is required for human ESCs to differentiate. Moreover, the ubiquitin ligase MDM2 is best known for balancing the activity of the tumor suppressor P53, but it can also promote adipocyte differentiation in a P53-independent manner (Hallenborg et al., 2016), and hESC differentiation can change the status of active BAX and sensitivity to DNA damage (Dumitru et al., 2012).

Evidence suggests that the P53 family is critical for mesodermal specification during exit from pluripotency in the embryo and culture. Wnt3 and its receptor Fzd1 are direct target genes of the P53 family, and induction of Wnt signaling by P53 is critical for the activation of mesodermal differentiation genes (Wang et al., 2017). In addition, transient WDR5 inhibition can stimulate ESC differentiation toward mesodermal fate *via* P53 as well as allow global chromatin accessibility to the landscape of mesodermal differentiation (Li et al., 2020). In the current study, WDR36 knockdown significantly promoted the differentiation of hEPS cells to the mesoderm lineage (**Figures 3,4**, **Supplementary**

Figure S5), indicating that WDR36 plays a key role in committing the mesodermal fate of hEPS cells. Therefore, we detected whether P53 signaling inhibition would reverse the effect of WDR36 knockdown on hEPS. Several works have proven that supplementation of PFT-α in the culture medium can effectively inhibit P53 (Garbern et al., 2020; Samarasinghe et al., 2021). In the current study, we treated hEPS1 cells with 2, 4, 6, 8, and 10 μM PFT-α, and the expressions of P53 and P21 were significantly inhibited at 10 μM, without affecting the survival of hEPS1 cells (**Supplementary Figures S8A,B**). As expected, inhibition of p53 signaling reversed the phenomenon that WDR36 knockdown attenuated self-renewal and enhanced mesoderm differentiation (**Supplementary Figures S8C,F**).

Collectively, our research showed that WDR36 was downregulated during hEPS differentiation. By constructing hEPS cell lines with Dox-induced WDR36 overexpression and WDR36 silencing, we further found that WDR36 knockdown disrupted self-renewal but promoted differentiation, especially toward the mesodermal lineage. Moreover, P53 signaling inhibition would reverse the effect of WDR36 knockdown on hEPS, in which the pluripotency and differentiation genes and the P53 signaling pathway-related factors were all altered. Taken together, our study provides a new insight into how WDR36 influences hEPS cells' self-renewal and differentiation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committee of Nanjing Medical University.

AUTHOR CONTRIBUTIONS

SA designed and performed the experiments, prepared the figures, and wrote the manuscript. DY performed the experiments, contributed to data interpretation, and helped in writing the manuscript. WZ performed the experiments and analyzed the data. HS performed the experiments and edited the manuscript. TY performed the experiments and edited the manuscript. RJ supervised the experiments and revised the manuscript. YY conceived, designed, and supervised the experiments and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Program of China (2019YFA0802700) and the National Natural Science Foundation of China (Nos 32070804, 82101791, 81971393, and 81571444). This work was also supported by the Jiangsu Provincial Program of Innovation and Entrepreneurship, China (No. (2020)10527) and the Six Talent Peaks Project in Jiangsu Province, China (No. WSW-035).

ACKNOWLEDGMENTS

We are grateful to Hongkui Deng for providing hEPS1 and hEPS2 cell lines.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.905395/full#supplementary-material

REFERENCES

Abu-Dawud, R., Graffmann, N., Ferber, S., Wruck, W., and Adjaye, J. (2018). Pluripotent Stem Cells: Induction and Self-Renewal. *Phil. Trans. R. Soc. B* 373, 20170213. doi:10.1098/rstb.2017.0213

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., et al. (2011). Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. *Cell.* 145, 183–197. doi:10.1016/j.cell.2011.03.003

Blankenbach, K. V., Bruno, G., Wondra, E., Spohner, A. K., Aster, N. J., Vienken, H., et al. (2020). The WD40 Repeat Protein, WDR36, Orchestrates Sphingosine Kinase-1 Recruitment and Phospholipase C-β Activation by Gq-Coupled Receptors. *Biochimica Biophysica Acta (BBA) - Mol. Cell. Biol. Lipids* 1865, 158704. doi:10.1016/j.bbalip.2020.158704

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell.* 122, 947–956. doi:10.1016/j.cell.2005.08.020

Cartier, A., Parent, A., Labrecque, P., Laroche, G., and Parent, J.-L. (2011). WDR36 Acts as a Scaffold Protein Tethering a G-Protein-Coupled Receptor, Gαq and Phospholipase Cβ in a Signalling Complex. *J. Cell. Sci.* 124, 3292–3304. doi:10.1242/jcs.085795

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., et al. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell.* 113, 643–655. doi:10.1016/s0092-8674(03)00392-1

Chen, J. (2016). The Cell-Cycle Arrest and Apoptotic Functions of P53 in Tumor Initiation and Progression. *Cold Spring Harb. Perspect. Med.* 6, a026104. doi:10.1101/cshperspect.a026104

Cotton, M., and Claing, A. (2009). G Protein-Coupled Receptors Stimulation and the Control of Cell Migration. *Cell. Signal.* 21, 1045–1053. doi:10.1016/j.cellsig.2009.02.008

Darr, H., Mayshar, Y., and Benvenisty, N. (2006). Overexpression of NANOG in Human ES Cells Enables Feeder-free Growth while

Inducing Primitive Ectoderm Features. *Development* 133, 1193–1201. doi:10.1242/dev.02286

Dumitru, R., Gama, V., Fagan, B. M., Bower, J. J., Swahari, V., Pevny, L. H., et al. (2012). Human Embryonic Stem Cells Have Constitutively Active Bax at the Golgi and Are Primed to Undergo Rapid Apoptosis. *Mol. Cell.* 46, 573–583. doi:10.1016/j.molcel.2012.04.002

Fan, Y., Min, Z., Alsolami, S., Ma, Z., Zhang, E., Chen, W., et al. (2021). Generation of Human Blastocyst-like Structures from Pluripotent Stem Cells. *Cell. Discov.* 7, 81. doi:10.1038/s41421-021-00316-8

Footz, T. K., Johnson, J. L., Dubois, S., Boivin, N., Raymond, V., and Walter, M. A. (2009). Glaucoma-associated WDR36 Variants Encode Functional Defects in a Yeast Model System. *Hum. Mol. Genet.* 18, 1276–1287. doi:10.1093/hmg/ddp027

Gallenberger, M., Meinel, D. M., Kroeber, M., Wegner, M., Milkereit, P., Bösl, M. R., et al. (2011). Lack of WDR36 Leads to Preimplantation Embryonic Lethality in Mice and Delays the Formation of Small Subunit Ribosomal RNA in Human Cells *In Vitro*. *Hum. Mol. Genet.* 20, 422–435. doi:10.1093/hmg/ddq478

Garbern, J. C., Helman, A., Sereda, R., Sarikhani, M., Ahmed, A., Escalante, G. O., et al. (2020). Inhibition of mTOR Signaling Enhances Maturation of Cardiomyocytes Derived from Human-Induced Pluripotent Stem Cells via P53-Induced Quiescence. *Circulation* 141, 285–300. doi:10.1161/CIRCULATIONAHA.119.044205

Ghatak, D., Das Ghosh, D., and Roychoudhury, S. (2021). Cancer Stemness: P53 at the Wheel. *Front. Oncol.* 10, 604124. doi:10.3389/fonc.2020.604124

Hallenborg, P., Siersbæk, M., Barrio-Hernandez, I., Nielsen, R., Kristiansen, K., Mandrup, S., et al. (2016). MDM2 Facilitates Adipocyte Differentiation through CRTC-Mediated Activation of STAT3. *Cell. Death Dis.* 7, e2289. doi:10.1038/cddis.2016.188

Hermeking, H. (2007). p53 Enters the microRNA World. *Cancer Cell.* 12, 414–418. doi:10.1016/j.ccr.2007.10.028

Huang, G., Ye, S., Zhou, X., Liu, D., and Ying, Q.-L. (2015). Molecular Basis of Embryonic Stem Cell Self-Renewal: from Signaling Pathways to Pluripotency Network. *Cell. Mol. Life Sci.* 72, 1741–1757. doi:10.1007/s00018-015-1833-2

Jain, A. K., Allton, K., Iacovino, M., Mahen, E., Milczarek, R. J., Zwaka, T. P., et al. (2012). p53 Regulates Cell Cycle and microRNAs to Promote Differentiation of Human Embryonic Stem Cells. *PLoS Biol.* 10, e1001268. doi:10.1371/journal.pbio.1001268

Kobayashi, N. R., Hawes, S. M., Crook, J. M., and Pébay, A. (2010). G-protein Coupled Receptors in Stem Cell Self-Renewal and Differentiation. *Stem Cell. Rev Rep* 6, 351–366. doi:10.1007/s12015-010-9167-9

Koifman, G., Aloni-Grinstein, R., and Rotter, V. (2019). Corrigendum to 'p53 Balances between Tissue Hierarchy and Anarchy'. *J. Mol. Cell. Biol.* 11, 1106. doi:10.1093/jmcb/mjz111

Kressler, D., Hurt, E., and Baβler, J. (2010). Driving Ribosome Assembly. *Biochimica Biophysica Acta (BBA) - Mol. Cell. Res.* 1803, 673–683. doi:10.1016/j.bbamcr.2009.10.009

Layden, B. T., Newman, M., Chen, F., Fisher, A., and Lowe, W. L., Jr. (2010). G Protein Coupled Receptors in Embryonic Stem Cells: a Role for Gs-Alpha Signaling. *PLoS One* 5, e9105. doi:10.1371/journal.pone.0009105

Li, D., and Roberts, R. (2001). Human Genome and Diseases: WD-Repeat Proteins: Structure Characteristics, Biological Function, and Their Involvement in Human Diseases. *CMLS, Cell. Mol. Life Sci.* 58, 2085–2097. doi:10.1007/pl00000838

Li, Q., Mao, F., Zhou, B., Huang, Y., Zou, Z., Dendekker, A. D., et al. (2020). p53 Integrates Temporal WDR5 Inputs during Neuroectoderm and Mesoderm Differentiation of Mouse Embryonic Stem Cells. *Cell. Rep.* 30, 465–480. e466. doi:10.1016/j.celrep.2019.12.039

Li, V. C., and Kirschner, M. W. (2014). Molecular Ties between the Cell Cycle and Differentiation in Embryonic Stem Cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9503–9508. doi:10.1073/pnas.1408638111

Lian, X., Zhang, J., Azarin, S. M., Zhu, K., Hazeltine, L. B., Bao, X., et al. (2013). Directed Cardiomyocyte Differentiation from Human Pluripotent Stem Cells by Modulating Wnt/β-Catenin Signaling under Fully Defined Conditions. *Nat. Protoc.* 8, 162–175. doi:10.1038/nprot.2012.150

Lin, T., Chao, C., Saito, S. i., Mazur, S. J., Murphy, M. E., Appella, E., et al. (2005). p53 Induces Differentiation of Mouse Embryonic Stem Cells by Suppressing Nanog Expression. *Nat. Cell. Biol.* 7, 165–171. doi:10.1038/ncb1211

Liou, J.-Y., Ko, B.-S., and Chang, T.-C. (2017). An Efficient Transfection Method for Differentiation and Cell Proliferation of Mouse Embryonic Stem Cells. *Methods Mol. Biol.* 1622, 139–147. doi:10.1007/978-1-4939-7108-4_11

Liu, Y., Zhu, D., Zhao, Z., Zhou, Q., Pan, Y., Shi, W., et al. (2021). Comparative Cytotoxicity Studies of Halophenolic Disinfection Byproducts Using Human Extended Pluripotent Stem Cells. *Chemosphere* 263, 127899. doi:10.1016/j.chemosphere.2020.127899

Marión, R. M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., et al. (2009). A P53-Mediated DNA Damage Response Limits Reprogramming to Ensure iPS Cell Genomic Integrity. *Nature* 460, 1149–1153. doi:10.1038/nature08287

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., et al. (2007). Pluripotency Governed by Sox2 via Regulation of Oct3/4 Expression in Mouse Embryonic Stem Cells. *Nat. Cell. Biol.* 9, 625–635. doi:10.1038/ncb1589

Monemi, S., Spaeth, G., Dasilva, A., Popinchalk, S., Ilitchev, E., Liebmann, J., et al. (2005). Identification of a Novel Adult-Onset Primary Open-Angle Glaucoma (POAG) Gene on 5q22.1. *Hum. Mol. Genet.* 14, 725–733. doi:10.1093/hmg/ddi068

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., et al. (1998). Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell.* 95, 379–391. doi:10.1016/s0092-8674(00)81769-9

Niwa, H., Miyazaki, J.-i., and Smith, A. G. (2000). Quantitative Expression of Oct-3/4 Defines Differentiation, Dedifferentiation or Self-Renewal of ES Cells. *Nat. Genet.* 24, 372–376. doi:10.1038/74199

Qin, H., Yu, T., Qing, T., Liu, Y., Zhao, Y., Cai, J., et al. (2007). Regulation of Apoptosis and Differentiation by P53 in Human Embryonic Stem Cells. *J. Biol. Chem.* 282, 5842–5852. doi:10.1074/jbc.M610464200

Ruiz, S., Panopoulos, A. D., Herrerías, A., Bissig, K.-D., Lutz, M., Berggren, W. T., et al. (2011). A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity. *Curr. Biol.* 21, 45–52. doi:10.1016/j.cub.2010.11.049

Saba, J. A., Liakath-Ali, K., Green, R., and Watt, F. M. (2021). Translational Control of Stem Cell Function. *Nat. Rev. Mol. Cell. Biol.* 22, 671–690. doi:10.1038/s41580-021-00386-2

Samarasinghe, R. A., Miranda, O. A., Buth, J. E., Mitchell, S., Ferando, I., Watanabe, M., et al. (2021). Identification of Neural Oscillations and Epileptiform Changes in Human Brain Organoids. *Nat. Neurosci.* 24, 1488–1500. doi:10.1038/s41593-021-00906-5

Schapira, M., Tyers, M., Torrent, M., and Arrowsmith, C. H. (2017). WD40 Repeat Domain Proteins: a Novel Target Class? *Nat. Rev. Drug Discov.* 16, 773–786. doi:10.1038/nrd.2017.179

Skarie, J. M., and Link, B. A. (2008). The Primary Open-Angle Glaucoma Gene WDR36 Functions in Ribosomal RNA Processing and Interacts with the P53 Stress-Response Pathway. *Hum. Mol. Genet.* 17, 2474–2485. doi:10.1093/hmg/ddn147

Sozen, B., Jorgensen, V., Weatherbee, B. A. T., Chen, S., Zhu, M., and Zernicka-Goetz, M. (2021). Reconstructing Aspects of Human Embryogenesis with Pluripotent Stem Cells. *Nat. Commun.* 12, 5550. doi:10.1038/s41467-021-25853-4

Ulmke, P. A., Xie, Y., Sokpor, G., Pham, L., Shomroni, O., Berulava, T., et al. (2021). Post-transcriptional Regulation by the Exosome Complex Is Required for Cell Survival and Forebrain Development via Repression of P53 Signaling. *Development* 148, dev188276. doi:10.1242/dev.188276

Utikal, J., Polo, J. M., Stadtfeld, M., Maherali, N., Kulalert, W., Walsh, R. M., et al. (2009). Immortalization Eliminates a Roadblock during Cellular Reprogramming into iPS Cells. *Nature* 460, 1145–1148. doi:10.1038/nature08285

Wang, Q., Zou, Y., Nowotschin, S., Kim, S. Y., Li, Q. V., Soh, C.-L., et al. (2017). The P53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell. Stem Cell.* 20, 70–86. doi:10.1016/j.stem.2016.10.002

Wettschureck, N., and Offermanns, S. (2005). Mammalian G Proteins and Their Cell Type Specific Functions. *Physiol. Rev.* 85, 1159–1204. doi:10.1152/physrev.00003.2005

Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M., and Sexton, P. M. (2018). Mechanisms of Signalling and Biased Agonism in G Protein-Coupled Receptors. *Nat. Rev. Mol. Cell. Biol.* 19, 638–653. doi:10.1038/s41580-018-0049-3

Yang, Y., Liu, B., Xu, J., Wang, J., Wu, J., Shi, C., et al. (2017). Derivation of Pluripotent Stem Cells with *In Vivo* Embryonic and Extraembryonic Potency. *Cell.* 169, 243–257. e225. doi:10.1016/j.cell.2017.02.005

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., et al. (2008). The Ground State of Embryonic Stem Cell Self-Renewal. *Nature* 453, 519–523. doi:10.1038/nature06968

Zheng, R., Geng, T., Wu, D.-Y., Zhang, T., He, H.-N., Du, H.-N., et al. (2021). Derivation of Feeder-free Human Extended Pluripotent Stem Cells. *Stem Cell. Rep.* 16, 2410–2414. doi:10.1016/j.stemcr.2021.07.019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fgene.2022.910439

Completing Single-Cell DNA Methylome Profiles *via* Transfer Learning Together With KL-Divergence

Sanjeeva Dodlapati[1], Zongliang Jiang[2] and Jiangwen Sun[1]*

[1]Department of Computer Science, Old Dominion University, Norfolk, VA, United States, [2]School of Animal Sciences, AgCenter, Louisiana State University, Baton Rouge, LA, United States

The high level of sparsity in methylome profiles obtained using whole-genome bisulfite sequencing in the case of low biological material amount limits its value in the study of systems in which large samples are difficult to assemble, such as mammalian preimplantation embryonic development. The recently developed computational methods for addressing the sparsity by imputing missing have their limits when the required minimum data coverage or profiles of the same tissue in other modalities are not available. In this study, we explored the use of transfer learning together with Kullback-Leibler (KL) divergence to train predictive models for completing methylome profiles with very low coverage (below 2%). Transfer learning was used to leverage less sparse profiles that are typically available for different tissues for the same species, while KL divergence was employed to maximize the usage of information carried in the input data. A deep neural network was adopted to extract both DNA sequence and local methylation patterns for imputation. Our study of training models for completing methylome profiles of bovine oocytes and early embryos demonstrates the effectiveness of transfer learning and KL divergence, with individual increase of 29.98 and 29.43%, respectively, in prediction performance and 38.70% increase when the two were used together. The drastically increased data coverage (43.80–73.6%) after imputation powers downstream analyses involving methylomes that cannot be effectively done using the very low coverage profiles (0.06–1.47%) before imputation.

Keywords: DNA methylation, single cell WGBS, embryo methylome, methylation imputation, transfer learning, KL divergence

Edited by:
Xi Wang,
Nanjing Medical University, China

Reviewed by:
Jianxiong Tang,
University of Electronic Science and Technology of China, China
Hao Li,
Institute of Health Service and Transfusion Medicine, China

****Correspondence:***
Jiangwen Sun
jsun@cs.odu.edu

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *01 April 2022*
Accepted: *25 May 2022*
Published: *22 July 2022*

Citation:
Dodlapati S, Jiang Z and Sun J (2022) Completing Single-Cell DNA Methylome Profiles via Transfer Learning Together With KL-Divergence.
Front. Genet. 13:910439.
doi: 10.3389/fgene.2022.910439

1 INTRODUCTION

DNA methylation, a process of adding a methyl group to the fifth carbon of cytosines, is ubiquitous in genome of all kingdoms of life from bacteria to eukaryotes (Zemach et al., 2010). Although there exist methylated cytosines in other contexts, methylation in the context of CpG dinucleotides (i.e., a cytosine nucleotide being immediately followed by a guanine nucleotide along the 5’ → 3’ direction of a sequence) is the most common form (Feng et al., 2010) and is the subject of this study. DNA methylation plays critical roles in the regulation of both genome stability and gene expression (Greenberg and Bourc’his, 2019), involved in many important biological processes such as embryonic development (Zhu et al., 2018; Duan et al., 2019), X-chromosome inactivation

(Grant et al., 1992), genomic imprinting (Proudhon et al., 2012), and aging (Xiao et al., 2019). Alterations in the usual methylation patterns may lead to disruption of normal cellular functions and disease conditions. Disrupted DNA methylation has been linked to several diseases such as cancer (Ko et al., 2010; Russler-Germain et al., 2014), immunological disorders (Rajshekar et al., 2018), and neurological disorders (Sun et al., 2014; Kernohan et al., 2016).

Due to its importance, obtaining DNA methylome profiles for varying biological systems has attracted considerable attentions (Abascal et al., 2020). Several techniques have been developed for profiling DNA methylation genome-wide, including methylated DNA immunoprecipitation sequencing (MeDIP-Seq) (Taiwo et al., 2012), whole-genome bisulfite sequencing (WGBS) (Clark et al., 2017), reduced representation bisulfite sequencing (RRBS) (Gu et al., 2011), and nanopore sequencing Clarke et al. (2009) followed by methylation detection. Since MeDIP-Seq relies on a methyl-cytosine antibody to pull down methylated DNA fragments followed by sequencing, the obtained profiles, even though genome-wide, are in low resolution (100–300 bp) and biased, with substantial underrepresentation of CpG poor regions (Rauluseviciute et al., 2019), limiting its application in biological studies. Nanopore sequencing, one of the emerging third-generation sequencing techniques, is capable of producing reads of much longer length (in tens to hundreds of thousands bases) compared to their short-read sequencing counterparts. Several computational approaches have been developed to predict DNA methylation from nanopore sequencing reads (Yuen et al., 2021). However, due to limited accuracy in both sequencing and subsequent methylation prediction (Liu et al., 2021), nanopore sequencing has yet become a widely used approach for methylome profiling.

Both RRBS and WGBS are based on bisulfite conversion and capable of producing methylome profiles at single-base pair resolution. Without the bias of RRBS for CpG dense regions, WGBS is currently the most popularly used methylome profiling technology and has been used to obtain profiles for a wide range of tissues in varying organisms (Roadmap Epigenomics Consortium et al., 2015; Abascal et al., 2020). However, to obtain a profile with high data coverage rate (defined as the proportion of CpG sites with profiled methylation state out of the total in the entire genome) using WGBS, large amount of genetic input coupled with high sequencing depth is required. Single-cell WGBS is well known for its very low coverage rate. When excluding CpG sites with low amount (below 5) of overlapping reads, the data coverage rate in single-cell methylomes can get down to just a little over 1% (Zhu et al., 2018) or even well below 1% (Smallwood et al., 2014). In applications, such as the study of mammalian preimplantation of embryos where genetic material is precious, the coverage rate can go extremely low after rigors data cleaning (see Materials and Methods), only 0.06–1.47% (all but one below 0.3%) in a recent study of bovine embryonic development (Duan et al., 2019). Sparsity in methylome profiles hinders the downstream analyses, limiting their value in efforts to understand the dynamics and regulation of biological processes.

To address the sparsity in DNA methylome profiles, many computational approaches have been developed in the past to impute missing data by training machine learning models to predict methylation state. With the advancement of technologies for assessing DNA methylation, the computational approaches have shifted from predicting the overall methylation level of a DNA fragment such as a CpG island (Bock et al., 2006) to the methylation state of individual CpG sites (Zhang et al., 2015). Varying types of data have been explored to use as input to predict DNA methylation, including a variety of DNA sequence patterns, methylation state of neighboring CpGs, profiles of other functional genomic events such as histone modifications in the same sample, and epigenetic profiles of other related samples. By leveraging a diverse of genomic profiles, several methods achieved very high prediction accuracy. For example, BoostMe (Zou et al., 2018) obtained an accuracy that is above 0.96 with using profiles of 7 histone markers, predicted binding sites of 608 transcriptional factors, predictions for 13 chromatin states, and chromatin accessibility profiles by assay for transposase-accessible chromatin with sequencing (ATAC-Seq). However, these methods have limited usage in the study of biological systems for which a wealth of additional data are not available.

Earlier methods used hand-crafted features derived from DNA sequence, which is limited by the understanding of the biology at the time and leads to suboptimal results. With seeing the successful applications of deep neural networks (DNNs) in many other domains, especially computer vision (Krizhevsky et al., 2012) and natural language processing (Otter et al., 2021), several recent studies have attempted to use DNNs to learn unbiased DNA sequence and/or local methylation patterns (Sharma et al., 2017; Zeng and Gifford, 2017; De Waele et al., 2022). Even though with success to some extent, these methods are limited by the availability of sufficient amount of data for training the DNNs. Transfer learning performs well in various low amount data scenarios by transferring knowledge learned on a large dataset that is different but related to the target learning problem (Zhuang et al., 2020). Several methods utilizing transfer learning have been developed recently in genomic data contexts where limited data are available, such as in the prediction of cancer survival using gene expression data (López-García et al., 2020), molecular cancer classification (Sevakula et al., 2018), denoising single-cell transcriptomics data (Wang et al., 2019), and imputing missing data in gene expression profiling with the input from DNA methylation profiles (Zhou X. et al., 2020). To the best of our knowledge, transfer learning has not been explored to leverage profiles with higher coverage rate in training predictive models for much sparser methylation profiles.

Due to allelic methylation, intercellular variability, or clusters of interspersed methylated and unmethylated CpGs within each cell, the intermediate DNA methylation (represented by a value in between 0 and 1) is widespread in the genome (Elliott et al., 2015). It has been indicated that intermediate methylation states may be functional and are dynamically regulated (Stadler et al., 2011). Moreover, large amount of methylome profiles that are available in public repositories were obtained by averaging across a group of cells that may be heterogeneous. Therefore, the variation in (intermediate) methylation level among CpG sites is indicative of difference in the context that regulates their methylation, such as the surrounding DNA sequence. However, when training models for predicting methylation by gradient descent to optimize a concrete objective, previous works chose to convert methylation level to the binary on or off state followed by employing a binary classification loss function such as logistic loss (Painsky and Wornell, 2018). Such an binary conversion results in loss of information and may lead to suboptimally trained models. Technically, to avoid binary conversion, the learning problem can be modeled as a general regression problem, where mean squared error (MSE) can be applied as the loss function with or without a final sigmoid mapping to ensure the model prediction within [0,1]. However, sigmoid mapping drives model outputs towards either 0 or 1, likely leading to suboptimal models; and, if without the sigmoid mapping, the model can output values beyond [0,1], making the prediction difficult to interpret. Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) that measures the difference between two distributions can be a better choice as a loss function for training classifiers without binary conversion, but so far has not been exploited in DNA methylation prediction.

Here in this article, we report the results from the exploration of using transfer learning together with KL divergence to train DNNs for completing DNA methylome profiles with extremely low coverage rate by leveraging those with higher coverage. We employed a hybrid network architecture adapted from DeepGpG (Sharma et al., 2017), a mixture of convolutional neural network (CNN) and recurrent neural network (RNN). The CNN learns predictive DNA sequence patterns and the RNN exploits known methylation state of neighboring CpGs in the target profile to complete and across others. To obtain pretrained network components (i.e., subnetworks), we used bovine methylome profiles of varying somatic tissues downloaded from NCBI GEO under accession numbers: GSE106538 and GSE147087. The majority of these profiles have a data coverage rate greater than 5% after cleaning (see Materials and Methods). The pretrained subnetworks were then transferred for the training of models to complete profiles of bovine oocytes and early embryos, which was also obtained from NCBI GEO (GSE121758). All of these profiles except one have a data coverage rate below 0.3%. The results from our empirical study indicate both model transferring and the use of KL divergence help to improve the performance of trained DNNs. Specifically, on average, there is about 22.45% increase in the performance measured in F1 score with model transferring and about 29.43% increase when using KL divergence. The use of both leads to even higher increase (about 38.70%), which suggests that the contributions of the two are in different nature and can be combined. The subsequent imputation using the trained DNNs increased the data coverage rate to 43.80–73.65% from the initial 0.06–1.47% for profiles of bovine oocytes and early embryos. The expanded data enable the methylation quantification for substantially more genomic features, such as genome bins, promoters, and CpG islands (CGIs). This could in turn lead to more insights into the dynamics in methylomes of bovine oocytes and early embryos across different stages and the understanding of roles of DNA methylation in regulating varying biological functions.

2 BACKGROUND AND RELATED WORK

There has been a wealth of research work on building computational models to predict DNA methylation since the pioneer work in 2005 that trained a support vector machine (SVM) for predicting methylation level of short DNA fragments (Bhasin et al., 2005). Limited by the lack of technologies for obtaining data in high resolution, the majority of earlier works focused on the prediction of methylation level of CpG islands, genomic regions that are rich of CpG sites (Bock et al., 2006; Das et al., 2006; Fang et al., 2006; Fan et al., 2008; Zheng et al., 2013). The input used in the prediction comprised varying sequence features derived from DNA fragment in initial works (Bock et al., 2006; Das et al., 2006; Fang et al., 2006) and later was expanded to consider chromatin state of histone modifications including both methylation (Fan et al., 2008) and acetylation (Zheng et al., 2013). The used DNA sequence features typically included characteristics of CpG islands such as G + C content and CpG ratio, evolutionary conservation, count of k-mers, and occurrence of (predicted) transcription factor binding sites and repetitive elements such as AluY. Due to the small size of available data, machine learning algorithms that work well on small datasets were typically employed, including SVM, linear discriminant analysis, and logistic regression. Among them, SVM was used most often and frequently led to models that had the best performance. Even though these earlier approaches predict accurately the methylation level of CpG islands, they offer limited view of the involvement of DNA methylation in biological functions, because many functional elements such as enhancers are frequently located outside of CpG islands (Li et al., 2021).

Thanks to the rapid advancement of high-throughput sequencing technologies, profiling genome-wide DNA methylation at single base resolution has become possible and with increasingly low cost. Large numbers of genome-wide DNA methylation profiles of a wide range of tissues and cell lines for varying organisms have been deposited in public accessible data repositories such as ENCODE (Dunham et al., 2012), Roadmap (Roadmap Epigenomics Consortium et al., 2015), and NCBI GEO. The availability of these high-resolution genome-wide profiles enables the training of machine learning models that predict DNA methylation at individual CpG sites, which has become the primary target of recently developed approaches for methylation prediction.

Depending on the type of input, the methods for methylation prediction at individual base resolution can be generally classified into three categories. The first category includes methods that predict from coarse profiles obtained with MeDIP-Seq and Methylation-sensitive Restriction Enzyme sequencing (MRE-Seq) (Stevens et al., 2013), or methylation state of neighboring CpGs and methylation profile of other (related) samples (Ma et al., 2014; Kapourani and Sanguinetti, 2019; Yu et al., 2020; Tang et al., 2021), or additionally with the help from profiles for

TABLE 1 | Summary of used bovine WGBS profiles.

Profile	Tissue	Accession number	Source	Breed	Data coverage rate (%)	Methylation rate (%)
Sperm	Sperm	GSE106538	Gamete	Holstein	21.34	74.11
MamGl	Mammary gland	GSE106538	Somatic	Holstein	12.6	73.59
PreCor	Prefrontal cortex	GSE106538	Somatic	Holstein	9.94	84.16
WBC1	White blood cell	GSE106538	Somatic	Holstein	15.38	81.22
WBC2	White blood cell	GSE147087	Somatic	Holstein	13.04	86.53
Adip1	Adipose	GSE147087	Somatic	Holstein	7.98	82.5
Adip2	Adipose	GSE147087	Somatic	Hereford	1.09	94.04
Muscle	Muscle	GSE147087	Somatic	Holstein	7.99	79.1
Heart1	Heart	GSE147087	Somatic	Holstein	1.32	80.44
Heart2	Heart	GSE147087	Somatic	Hereford	0.85	92.47
Lung	Lung	GSE147087	Somatic	Holstein	7.09	78.53
Spleen	Spleen	GSE147087	Somatic	Holstein	11.5	83.85
Liver1	Liver	GSE147087	Somatic	Holstein	5.46	83.27
Liver2	Liver	GSE147087	Somatic	Hereford	0.88	88.6
Ileum	Ileum	GSE147087	Somatic	Holstein	8.69	79.6
Rumen	Rumen	GSE147087	Somatic	Holstein	5.18	59.92
Jejun	Jejun	GSE147087	Somatic	Hereford	1.17	81.84
Kidn1	Kidney	GSE147087	Somatic	Hereford	1.28	88.31
Kidn2	Kidney	GSE147087	Somatic	Holstein	5.69	84.03
Uterus	Uterus	GSE147087	Somatic	Holstein	6.09	84.91
Ovary	Ovary	GSE147087	Somatic	Holstein	11.1	73.01
Placenta	Placenta	GSE147087	Somatic	Hereford	0.98	40.81
GVO	GV Oocyte	GSE121758	Gamete	Holstein	0.16	4.3
MIIO1	MII Oocyte	GSE121758	Gamete	Holstein	0.06	8.39
MIIO2	MII Oocyte[a]	GSE121758	Gamete	Holstein	0.13	5.05
2-Cell	Embryo	GSE121758	Embryo	Holstein	0.26	2.41
4-Cell	Embryo	GSE121758	Embryo	Holstein	0.21	3.32
8-Cell	Embryo	GSE121758	Embryo	Holstein	0.18	1.97
16-Cell	Embryo	GSE121758	Embryo	Holstein	1.47	5.94

[a]*In vitro oocyte. All other oocytes and embryos are* in vivo.

other epigenetic markers, such as histone modifications (Ernst and Kellis, 2015; Zou et al., 2018). Due to the availability of large amount of data for training, the most popularly used machine learning algorithm by these approaches is ensemble trees, either random forest or gradient boosting machines. To make accurate prediction using these approaches, either relative high data coverage or the availability of profiles of many other epigenetic markers is needed. The second category consists of methods that employ only the sequence features derived from the DNA fragment centered at the target CpG site to predict for. These methods vary mainly in the length of input DNA fragment and ways of deriving sequence features that include simply treating the input sequence as structured data (in other words each position is taken as an individual input variable) (Kim et al., 2008), counting of k-mers (Lu et al., 2010; Zhou et al., 2012), and using a CNN (Zeng and Gifford, 2017). The models obtained using these methods generally make less accurate prediction than those from the first category. The third category consists of methods that leverage both sequence features and functional chromatin states to varying extent, including methylation state of neighboring CpGs. There are methods relying on hand-crafted DNA sequence features similar to those approaches developed for predicting methylation level of CpG islands (Zhang et al., 2015; Jiang et al., 2019), but with the majority employing DNNs to derive features that are unbiased (Wang et al., 2016; Sharma et al., 2017; Fu et al., 2019; Levy-Jurgenson et al., 2019; De Waele et al., 2022). Notably, methods using DNNs generally perform better than those not when there is no additional input beyond the methylation profile of the target sample (Sharma et al., 2017; De Waele et al., 2022). However, it is well known that training DNNs is difficult, requiring large amount of training data. Therefore, the success in the application of existing DNN-based methods is limited when methylation profiles are extremely sparse.

Transfer learning is able to mitigate data scarcity problems of target domain by learning model priors on larger data in a source domain related to the target domain but with different data distribution. It has been shown with effectiveness in the learning for various low data scenarios (Zhuang et al., 2020). Different transferring strategies have been developed, among which instance-based, mapping-based, network-based, and adversarial-based are more prominent approaches (Tan et al., 2018). It has been reported that logistic loss is not effective in learning features for transferring (Islam et al., 2021), since it results in hard class separation and hence leads to less adaptability of the source model while transfer it to the target domain. This problem is acute when very few examples for training are available in target domain. Recently, transfer learning has been applied to impute incomplete RNA-sequencing data by transferring features learned during predicting DNA methylation (Zhou X. et al., 2020). To the best of our knowledge, transfer learning has not been explored to train DNN-based models for predicting DNA methylation to impute sparse methylomes.

3 MATERIALS AND METHODS

To enhance downstream analyses, such as gene expression regulation, we train DNNs to impute missing methylation data in methylome profiles with the consideration of both DNA sequence patterns and methylation state of neighboring CpG sites. It is known that well-performing DNNs require large data in their training. However, methylome profiles of oocytes and mammalian preimplantation embryos are typically very sparse due to low amount of genetic material available for sequencing, limiting the amount of data for training DNNs. As a result, it is a challenging problem to obtain trained DNNs that make accurate predictions for missing CpGs in these profiles. To improve prediction accuracy of DNNs, we 1) employ the Kullback-Leibler (KL) divergence as the loss function in training to maximize usage of information carried in the data and 2) leverage transfer learning to make use of the much denser methylation profiles that are available for other tissues. Specifically, in this study, we trained DNNs for imputing missing CpG sites in methylome profiles of oocytes and preimplantation embryos of bovine.

3.1 Datasets

The methylome profiles of bovine oocytes and preimplantation embyros were obtained by downloading from NCBI GEO repository with accession number GSE121758. These profiles were produced via WGBS in a recent study of mythylome dynamics of oocytes and *in vivo* early embryos of bovine (Duan et al., 2019). More specifically, there are profiles for three types of oocytes, including two *in vivo* at different developmental stages, that is, germinal vesicle (GV) oocyte and metaphase II (MII) oocyte, and one *in vitro* MII oocyte. The dataset includes profiles for *in vivo* embryos at four different developmental stages: 2-cell, 4-cell, 8-cell, and 16-cell. The data coverage rate, that is, the proportion of CpG sites in the whole genome with known state in a methylation profile, is very low among these profiles, ranging from 0.06% for in vio MII oocyte to 1.47% for 16-cell embryo with all but one below 0.3% (**Table 1**).

To enhance the training for oocytes and early embryos, we identified two bovine WGBS datasets in the NCBI GEO repository with accession numbers: GSE106538 and GSE147087, respectively. Both datasets provide methylome profiles for somatic tissues for which large amount of genetic materials are available for sequencing. Specifically, GSE106538 provides profiles for sperm in addition to three different somatic tissues of Holstein cattle: mammary gland, prefrontal cortex, and white blood cell (Zhou et al., 2018), while GSE147087 provides methylome profiles with varying availability for cattle of two different breeds: Holstein and Hereford for a total of 14 tissues, including lung, heart, spleen, kidney, liver, rumen, jejun, ileum, ovary, uterus, placenta, white blood cell, muscle, and adipose (Zhou Y. et al., 2020). Profiles included in GSE106538 have high data coverage rate, ranging from 9.94% for prefrontal cortex to 21.34% for sperm (**Table 1**). Compared to these profiles, the data coverage rate of profiles from GSE147087 is much lower, ranging from 0.95 to 13.04% with the majority above 5% (**Table 1**), which is still significantly higher than that in profiles of oocytes and early embryos.

To prepare data for network training and subsequent imputation, downloaded datasets underwent a sequence of preprocessing steps. First, the profiles that are replicates of the same tissue were merged within the same data source. Following the consolidation, we excluded CpGs from a profile that have limited support for their profiled methylation state, that is, with a number of overlapping sequencing reads no greater than 3. DNA methylation is known to be stable during replication and remains symmetric, meaning that the copy of the cytosine on one strand at a CpG site is expected to have the same methylation state as the copy on the other strand (Vandiver et al., 2015; Petryk et al., 2020). In other words, hemi-methylated (unsymmetrical) CpG sites are rare and the existence of such CpGs is high likely due to errors in the methylation profiling. To ensure high data quality and avoid causing confusion during the network training by ambiguous labeling, we excluded from all profiles the hemi-methylated CpG sites and those with data for only one strand. The data coverage rate in each profile after going through all the preprocessing steps, together with the methylation rate calculated using the remaining CpG sites, is provided in **Table 1**.

3.2 Network Architecture

To leverage both DNA sequence patterns and correlation in methylation state among neighboring CpGs, we employed networks with the architecture adapted from one that has been utilized for predicting DNA methylation in human and mouse genome (Angermueller et al., 2017). As illustrated in **Figure 1**, three feature learning subnetworks: Sequence, Methylation, and Joint were used to extract features from the input. Specifically, the Sequence subnetwork learns DNA sequence patterns that are predictive to methylation; the Methylation subnetwork learns correlation in the methylation state among neighboring CpGs; and the Joint subnetwork fuses the features extracted by the Sequence and Methylation subnetworks.

To learn sequence features, the Sequence subnetwork takes in one-hot coded DNA fragment of 1,001 base pair (bp) long, centered at the CpG to predict for and propagates the data through two consecutive convolution blocks and one fully connected layer. As in Angermueller et al. (2017), each convolution block consists of a convolutional layer followed by a max pooling layer. The size of filters and their amounts are indicated in **Figure 1**. Data normalization is known to facilitate the training by both speeding up the training process and making it less sensitive to different choices of hyperparameters, such as learning rate. Thus, following the max pooling layer in each convolution block, we added a batch normalization layer.

A bi-directional gated recurrent unit (GRU) network (Cho et al., 2014) was used to exploit the methylation correlation among neighboring CpGs and learn such correlation from multiple methylome profiles. The input to a GRU is a vector

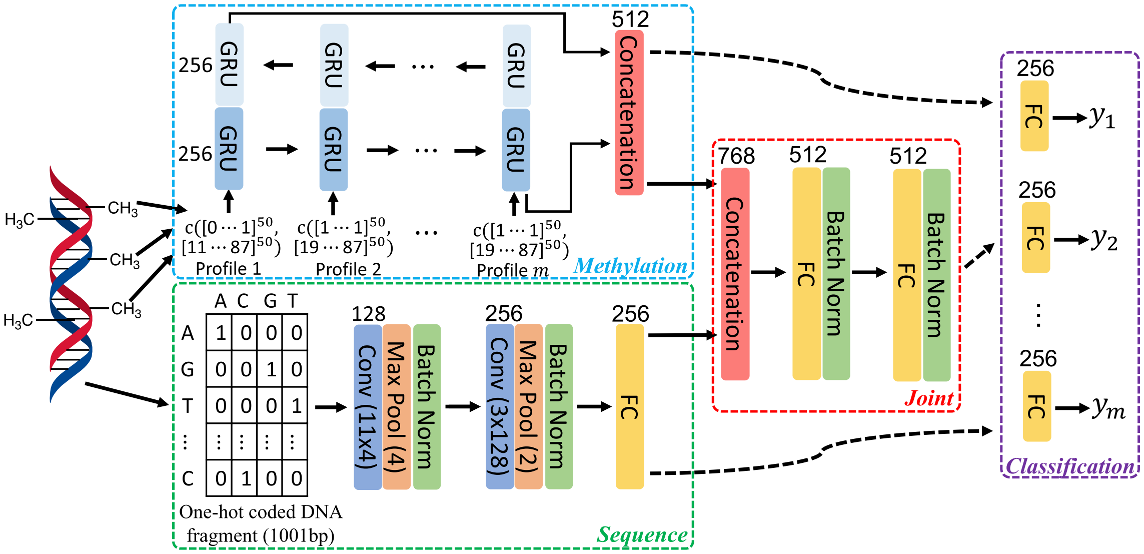


FIGURE 1 | Architecture of network components in our study. Each colored bar represents a layer of operation in the network as indicated by the enclosed description. The numbers on the top or the side of bars specify the number of filters or hidden units in the corresponding layers. Activation functions following convolutional or fully connected layers and that in GRU are not shown. The former uses rectified linear function, while in the latter, hyperbolic tangent function is employed. $[d_1 \cdots d_{50}]^{50}$ represents a numerical vector of size of 50 and $c\ (\mathbf{v}_1, \mathbf{v}_2)$ denotes the concatenation of two vectors: $\mathbf{v}_1$ and $\mathbf{v}_2$. y_i's represent methylation state of the single CpG to predict in multiple profiles. When transferring trained models from source to target, components of Sequence, Methylation, Joint, or their combinations are transferred. Components that are not transferred are trained from scratch (i.e., random initialization). Conv, convolution; Max Pool, max pooling; Batch Norm, batch normalization; FC, fully connected.

of size of 100, composed of concatenating two vectors. One of them contains methylation level of 50 CpG sites surrounding the one to predict for, 25 on each side. The other vector includes the base pair distance from the corresponding surrounding CpG sites to the one to predict for. The features learned from passing through the sequential input in two opposite directions were combined by simple concatenation to produce the final representation of learned methylation correlation among neighboring CpG sites (**Figure 1**).

To fuse sequence features and methylation patterns, the Joint subnetwork propagates the combined representation (by concatenation) from Sequence and Methylation subnetworks through two fully connected layers. Like in Sequence subnetwork, batch normalization is used following each fully connected layer to facilitate network training.

With features extracted by the three feature learning subnetworks, the methylation state prediction for a targeted CpG can be made with a Classification subnetwork head. The input to this classification head is determined by the data to consider in the prediction. Specifically, the output of the Sequence subnetwork is used when only surrounding DNA sequence patterns are utilized. Similarly, when only local methylation patterns are considered, the output of the Methylation subnetwork should be used. If to take into account both the sequence and methylation patterns, the output of the Joint subnetwork is used. The classification head includes a fully connected layer followed by a Softmax layer (not shown in **Figure 1**). Multi-task learning has been widely used to improve model performance in many applications, including prediction for functional genomics events (Zhou and Troyanskaya, 2015; Avsec et al., 2021). In multi-task learning, multiple models are jointly trained with sharing certain components of the models, allowing mutual learning among tasks to improve performance. In this work, we also leverage multi-task learning to jointly train networks for multiple methylome profiles, with predicting for each profile being a separate learning task. All tasks share the same feature extraction subnetworks, but with task-specific classification head as illustrated in **Figure 1**.

3.3 Loss Function

To train DNNs for predicting varying functional genomic events including DNA methylation, the logistic loss has been the primary loss function utilized so far in the literature. Let $\mathbf{y} \in \{0,1\}^N$ denote the vector containing true labels and $\tilde{\mathbf{y}} \in [0,1]^N$ represent the corresponding predicted probabilities, and the logistic loss (LL) is calculated as shown below:

$$LL(\mathbf{y}, \tilde{\mathbf{y}}) = \sum_{i=1}^{N} -y_i \log(\tilde{y}_i) - (1 - y_i)\log(1 - \tilde{y}_i). \qquad (1)$$

In the obtained methylome profiles, the methylation state of a CpG is characterized by the fraction number of reads that contain methylated cytosine out of the total number of reads that overlap with the CpG. In other words, the methylation state of any CpG is a value in [0, 1] and a CpG (s_1) with a value of 0.51 is expected to be in very different methylation state compared to another CpG (s_2) with a value of 0.99. However, to compute the logistic loss as in **Eq. 1**, the methylation state needs to be converted to a binary value (as $y_i \in \{0, 1\}$) by comparing to pre-defined threshold, typically 0.5. More specifically, CpGs with an assessed methylation state in a profile above 0.5 would be considered as methylated and labeled with 1 in the profile; while CpGs would be considered as unmethylated and labeled with 0 if their assessed state is below 0.5. Such a conversion results in no difference at all in the methylation state between s_1 and s_2, as both would be labeled with 1 (i.e., methylated). The information loss during this process may lead to suboptimal models.

To make use of the most information carried in the profiles for training, we propose to utilize KL divergence score (D_{KL}) which needs no binary conversion of the methylation state. KL divergence measures the difference between two distributions. In our problem of predicting for the methylation state of a CpG (i), the empirically assessed (true) state ($y_i \in [0, 1]$) and the predicted state ($\tilde{y}_i \in [0,1]$) can be seen as two distributions. KL divergence, as calculated in **Eq. 2**, can be used to measure the difference between true and predicted states. The optimization goal here is to find a network producing a prediction that minimizes the $D_{KL}(y_i, \tilde{y}_i)$.

$$D_{KL}(y_i, \tilde{y}_i) = y_i \log\left(\frac{y_i}{\tilde{y}_i}\right) + (1 - y_i)\log\left(\frac{1 - y_i}{1 - \tilde{y}_i}\right). \tag{2}$$

Let $\mathbf{w}$ be a vector that contains all learnable parameters in the network and $\mathbf{X}$ denote the network input. Considering all CpGs for training in a profile with simultaneously learning for multiple (m) profiles, the following is the overall loss function to minimize by finding the optimal $\hat{\mathbf{w}}$ during the network training.

$$\ell\left(\mathbf{w}; \mathbf{X}, \mathbf{y}^{j:\ j=1\ldots m}\right) = \sum_{j=1}^{m} \alpha_j \sum_{i=1}^{N_i} \beta_i^j D_{KL}\left(y_i^j, \tilde{y}_i^j\right), \tag{3}$$

Where $\mathbf{y}^j$ represents the true methylation state of CpGs in j-th profile and N_j is the number of CpGs in j-th profile for training. There are two sets of hyperparameters involved in this loss function: α_j's and β_i^j's. The former balances the contribution of each individual task to the overall loss; while the latter specifies that of each individual CpG in every profile.

3.4 Transfer Learning

To obtain models for completing methylome profiles of oocytes and early embryos (target profiles), we started from training feature extraction subnetworks: Sequence, Methylation, and Joint, leveraging profiles of somatic tissues and sperm (**Table 1**) using multi-task learning as illustrated in **Figure 1**. The trained subnetworks, referred as source models, were subsequently used as pretrained ones to train networks (target models) for target profiles.

3.4.1 Source Model

To study the contributions of DNA sequence and local methylation patterns to the prediction of methylation, we trained models that uses DNA sequence only, methylation state of neighboring CpGs only, or the combination of the two (full model). In addition, we studied three different ways of training to obtain the best performing full model for transferring. The trained models are summarized in below:

Seq: Model that predicts from DNA sequence only, consisting of the Sequence subnetwork followed by the Classification head. The two subnetworks were trained from scratch with randomly initialized network weights.

Met: Model that predicts from methylation state of neighboring CpGs only, consisting of the Methylation subnetwork followed by the Classification head. Same as in Seq model, the two subnetworks were trained from scratch with random initialization.

The following three are all full models that predict from both DNA sequence and methylation state of neighboring CpGs, consisting of all three feature extraction subnetworks followed by the Classification head. They differ in how the full model was built.

Full1: All four subnetworks were trained from scratch with random initialization.

Full2: The Sequence subnetwork in the Seq model and Methylation subnetwork in the Met model were utilized as pretrained subnetworks. The full model was built by training the Joint subnetwork and the classification head from scratch with the two pretrained subnetworks remaining fixed.

Full3: The full model was built in the same way as for Full2 except that the two pretrained subnetworks were fine-tuned during the training.

The three feature extraction subnetworks from the best performing full model were transferred for subsequent model training to predict DNA methylation in target profiles.

3.4.2 Target Model

Given their distinct nature, there is likely variation among the three feature extraction subnetworks in their contribution to the improvement of target models through transferring. To study such differential impact, we trained models with/without transferring for predicting from DNA sequence only, and methylation state of neighboring CpGs only, and both. The detailed description of the explored settings is provided in below.

3.4.2.1 Predicting From DNA Sequence Only

SeqN: The Sequence subnetwork was trained from scratch (i.e., without transferring) together with the Classification head.

SeqT1: The Sequence subnetwork was initialized using the transferred source model and remained fixed during the training for the Classification head.

SeqT2: The Sequence subnetwork was initialized with the transferred source model and fine-tuned while training for the Classification head.

3.4.2.2 Predicting From Methylation State of Neighboring CpGs Only

MetN: The Methylation subnetwork was trained from scratch together with the Classification head.

MetT1: The Methylation subnetwork was initialized with the transferred source model and remained fixed during the training for the Classification head.

MetT2: The Methylation subnetwork was initialized with the transferred source model and fine-tuned while training for the Classification head.

3.4.2.3 Predicting From Both DNA Sequence and Methylation State of Neighboring CpGs

FullN: All three feature extraction subnetworks were trained without transferring together with the Classification head.

FullTS1: Sequence subnetwork was transferred but remained fixed during the target model training. The other two feature extraction subnetworks were trained without transferring together with the Classification head.

FullTS2: Identical to FullTS1 except that the transferred Sequence subnetwork was fine tuned.

FullTM1: Similar to FullTS1 but with Methylation subnetwork being the only transferred subnetwork.

FullTM2: Identical to FullTM1 except that the transferred Methylation subnetwork was fine tuned.

FullTB1: Both Sequence and Methylation subnetworks were transferred but remained fixed during the target model training. The Joint subnetwork was trained without transferring together with the Classification head.

FullTB2: Identical to FullTB1 except that the two transferred subnetworks were fine tuned.

FullTA1: All three feature extraction subnetworks were transferred but remained fixed during the training for the Classification head.

FullTA2: Identical to FullTA1 except that all transferred subnetworks were fine tuned.

3.5 Network Training and Evaluation

The networks were implemented and the experiments were carried out using TensorFlow framework in Python, a popular open-source software library in deep learning research. To train and evaluate all the networks, we partitioned the methylome profile into three parts by chromosomes that were used for training, validation, and testing, respectively. More specifically, data from chromosomes 1, 4, 7, 10, 13, 16, 19, 22, 25, and 28 were used for training to optimize network weights. Data from chromosomes 3, 6, 9, 12, 15, 18, 21, 24, and 27 were used for validation to identify optimal setting for hyperparameters, such as learning rate. Data from chromosomes 2, 5, 8, 11, 14, 17, 20, 23, 26, and 29 were used for testing to evaluate the performance of all trained networks. Adam optimizer (Kingma and Ba, 2015) was used to optimize network weights with weight decay and early stopping. All networks were trained with applying both ℓ_1 and ℓ_2 regularizers and with a mini-batch size of 128. The hyperparameter β_i^j's (individual sample weights) in **Eq. 3** were specified according to the class label distribution in individual profiles. To simplify, all α_j's (task weights) were set to 1 in this study. For all settings, we fine-tuned the learning rate with grid search from {0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001}. If not specified otherwise, the performance of the best model among the different choices of learning rate was reported for each setting.

There are different metrics that can be used to evaluate the performance of classification models, such as accuracy, area under curve of receiver operating characteristic (AUC-ROC), area under precision recall curve (AUPRC), and F1 score. Most of existing works on functional genomics events prediction used varying combinations of AUC-ROC, AUPRC, and accuracy Zhang et al. (2015); Angermueller et al. (2017); Liu et al. (2018); Zhang and Hamada (2018). The AUC-ROC and AUPRC take into account the uncertainty in prediction and are not metrics to evaluate the performance of models in making specific binary classification. In addition, these two metrics and accuracy tend to overestimate model performance when there is large imbalance in the class label distribution, which is the case in our study (**Table 1**). To avoid this problem, we used F1 Score based on the minor class as the primary metric for evaluating models in making specific binary classification.

4 RESULTS AND DISCUSSION

4.1 Comparison of Evaluation Metrics

As indicated in **Table 1**, the methylation rate, the proportion of methylated CpGs (with a methylation level above 0.5) out of the total in the genome, in target profiles is very low, ranging from 1.97 to 8.39%. This leads to datasets with large imbalance in the class label distribution when labeling methylated CpGs as positive examples and unmethylated as negative ones. In contrast, profiles from GSE106538 and GSE147087 (source profiles) have a methylation rate in a range of 40.81–94.04% with the majority around 80%, which results in a dataset that has much less imbalance in class label distribution. To show the difference among different metrics including AUC (i.e., AUC-ROC), accuracy, and F1 score in cases of large class label imbalance, besides the five models described in the above section (Materials and Methods) trained on source profiles, models using the exact same settings were also trained on target profiles.

Performance of all models evaluated by three metrics (AUC, accuracy, and F1 score) is provided in **Figure 2**. According to accuracy and AUC, all five models for target profiles perform better than the corresponding models for source profiles. However, by F1 score the comparison indicates a completely different story, the performance of target models being substantially worse. The reason that accuracy and AUC associated with target profiles are high is the high level of class label imbalance that resulted from extensive low methylation rate. In an extreme case, a classifier that does not learn any intrinsic patterns in the data that are predictive of methylation and simply predicts every example to be negative after just learning the class label distribution can achieve an accuracy above 91%. Therefore, in the presence of large class imbalance, F1 score, specifically the F1-score calculated with labeling the minor class as positive, is a better metric to use for evaluating how well a classifier learning intrinsic patterns from the data.

4.2 Models for Source Profiles

To demonstrate the advantage of using KL divergence as the training objective over logistic loss and MSE with/without sigmoid mapping, we trained models in all five settings (see Materials and Methods) using all losses on the source profiles. The performance of obtained models using KL divergence and

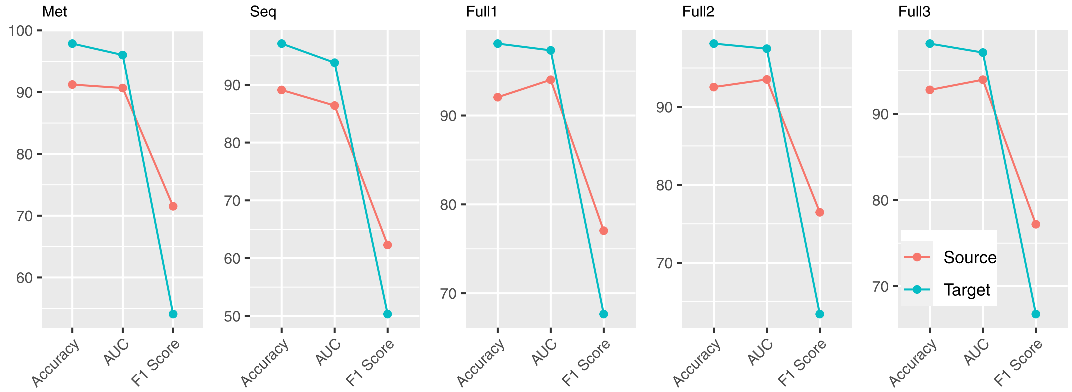


FIGURE 2 | Comparison of accuracy, AUC, and F1 score using models obtained on source and target profiles.

TABLE 2 | F1 score of models obtained on source profiles. Models trained using both logistic loss and KL divergence as the objective function are included for comparison.

Profile	Logistic loss					KL divergence				
	Met	**Seq**	**Full1**	**Full2**	**Full3**	**Met**	**Seq**	**Full1**	**Full2**	**Full3**
Sperm	0.9560	0.8919	0.9585	0.9557	0.9588	0.9584	0.8964	0.9593	0.9595	0.9593
MamGl	0.8768	0.7965	0.8940	0.8917	0.8952	0.8830	0.8070	0.8986	0.8964	0.8984
PreCor	0.7910	0.6898	0.8130	0.8080	0.8119	0.8010	0.6922	0.8198	0.8156	0.8207
WBC1	0.8603	0.7760	0.8913	0.8851	0.8915	0.8753	0.7786	0.8947	0.8917	0.8934
WBC2	0.7188	0.6006	0.7767	0.7499	0.7740	0.7434	0.5916	0.7867	0.7788	0.7870
Adip1	0.7556	0.6321	0.7864	0.7824	0.7866	0.7683	0.6323	0.7944	0.7896	0.7928
Adip2	0.5783	0.5023	0.7137	0.5595	0.6908	0.6540	0.4588	0.7227	0.7190	0.7341
Muscle	0.7459	0.6206	0.7893	0.7803	0.7873	0.7569	0.6379	0.7982	0.7905	0.7959
Heart1	0.5475	0.2910	0.6145	0.5984	0.6155	0.5426	0.3607	0.6024	0.6114	0.6058
Heart2	0.5650	0.4650	0.6976	0.6147	0.6841	0.6140	0.4636	0.6954	0.6945	0.7099
Lung	0.7641	0.6509	0.8000	0.7927	0.7982	0.7729	0.6653	0.8053	0.8005	0.8039
Spleen	0.7463	0.6623	0.7907	0.7785	0.7891	0.7598	0.6555	0.7928	0.7892	0.7957
Liver1	0.5832	0.4178	0.6490	0.6298	0.6492	0.5674	0.4426	0.6473	0.6393	0.6532
Liver2	0.4855	0.3829	0.6116	0.5319	0.5900	0.5024	0.3968	0.6219	0.6008	0.6252
Ileum	0.6702	0.6022	0.7721	0.7512	0.7654	0.6678	0.6461	0.7813	0.7710	0.7830
Rumen	0.7493	0.6120	0.7963	0.7909	0.7999	0.7586	0.6918	0.8052	0.8005	0.8059
Jejun	0.6173	0.5453	0.7780	0.7221	0.7756	0.6238	0.6180	0.7913	0.7756	0.7841
Kidn1	0.6690	0.5564	0.7584	0.7044	0.7395	0.6999	0.5688	0.7602	0.7581	0.7719
Kidn2	0.5741	0.4029	0.6239	0.6065	0.6223	0.5579	0.4169	0.6246	0.6109	0.6168
Uterus	0.6181	0.4919	0.6775	0.6579	0.6754	0.6291	0.4840	0.6840	0.6806	0.6821
Ovary	0.8088	0.7026	0.8549	0.8463	0.8544	0.8166	0.7459	0.8609	0.8565	0.8610
Placenta	0.7000	0.5323	0.7405	0.7361	0.7329	0.7194	0.6265	0.7523	0.7488	0.7592
Average	0.6991	0.5830	0.7631	0.7352	0.7585	0.7124	0.6035	0.7681	0.7627	0.7700

logistic loss measured by F1 score is provided in **Table 2** (**Supplementary Table S1** for corresponding results when MSE was used). KL divergence outperforms logistic loss in all settings. Specifically, the average F1 scores of models trained to predict from DNA sequence only, neighboring CpG methylation states only, and both are 0.7124, 0.6035, and 0.7700, respectively with KL divergence compared to 0.6991, 0.5830, and 0.7585 with logistic loss. Models trained with KL divergence also have better performance than those trained using MSE with or without the sigmoid mapping (**Supplementary Table S1**).

Among the five models, there are three that were trained to predict methylation from both sequence and neighboring CpGs in different settings (see Materials and Methods). The results in **Table 2** indicate fine-tuning the separately trained methylation and sequence subnetworks while training the joint subnetwork (Full3) leads to the full model that has the best average performance. This is consistent between the use of different objective functions, even though with logistic loss Full3 is only minimally better than training all subnetworks from scratch (Full1). The result from comparing Full3 to Full2 suggests fine-tuning the pretrained methylation and sequence subnetworks is necessary to obtain models with better performance. The three subnetworks obtained in Full3 were subsequently used as pretrained networks to obtain models for target profiles.

Compared models trained to predict methylation from sequence only (Seq), from neighboring CpG methylation states

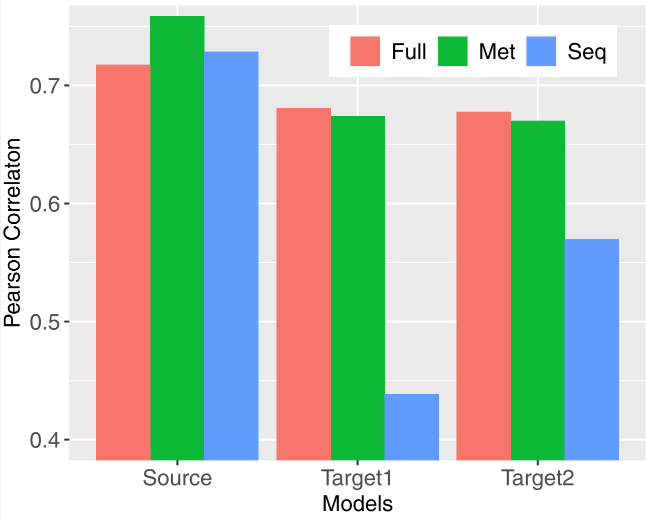


FIGURE 3 | Pearson correlation between model performance and the data coverage rate in corresponding profiles. Source indicates the models trained for source profiles; Target1 (Target2) labels models trained on target profiles without (with) transferring.

only (Met), and from both (Full), the Full model always has the best prediction performance regardless the objective function being used (**Table 2**). Models predicting from neighboring CpGs perform much better (average F1 score: 0.7124) than that predicting from sequence (average F1 score: 0.6035). The performance of all three model variants for individual profiles closely correlates with the data coverage rate in corresponding profiles with a Pearson correlation ranging from 0.72 to 0.76 (**Figure 3**). In other words, the higher the coverage rate is in a profile, the better performance the corresponding models are likely to have. Among the three scenarios, the performance of models predicting from methylation state of neighboring CpGs is the one that mostly correlates the data coverage ($cor = 0.76$). This makes sense because the higher the data coverage rate is, the local CpG methylation pattern is more informative to the prediction for the target CpG.

4.3 Models for Target Profiles

To find out how transfer learning helps with obtaining models for target profiles, we trained networks in varying settings (see Materials and Methods) using KL divergence as the objective function. The subnetworks that were transferred are those in Full3 trained for source profiles also with KL divergence as the objective function. The performance of all models assessed using F1 score is provided in **Table 3** (**Supplementary Table S2** for other performance metrics, including accuracy, AUC-ROC, precision, and recall). Similar to the case with source profiles, models predicting from both sequence and neighboring CpGs perform better than those predicting from anyone of them only with one exception: 4-cell, for which predicting from CpG only (F1 score: 0.8367) is slightly better than predicting from both (F1 score: 0.8357). Models trained with transferring all three subnetworks together with subsequent fine-tuning (FullTA2) achieved the overall best performance across profiles (average F1 score: 0.6828). Aligning with the observations in training for source profiles, the performance of models for individual profiles also well positively correlates with the data coverage rate, but with reduced correlation, especially in the case of predicting from sequence only (**Figure 3**). Such much-reduced correlation is likely due to the extremely high sparsity in several profiles that leads to overfitting. This is evidenced by the much-improved correlation (from 0.44 to 0.57) when more data were considered via model transferring. For profiles that have extremely low coverage rate, including GVO, MIIO1, and MIIO2, predicting from neighboring CpGs only does not perform well with F1 score ranging from only 0.1754 to 0.2868, much worse than predicting from sequence only.

Model transferring helped to obtain models with significantly improved performance to predict from neighboring CpGs only or from both sequence and neighboring CpGs. However, there is no gain to be seen in training models predicting from sequence only, except for profiles from 2-cell and 16-cell stages that have the highest data coverage rate among all target profiles. Such lack of improvement is likely due to the extremely low data coverage, causing the learning to arrive in a local minimal that is difficult to reach when training starting from a pretrained sequence subnetwork. The results in **Table 3** also indicate that fine-tuning the transferred models always helped, with just very few exceptions. In the case of predicting from sequence only,

TABLE 3 | F1 score of models trained for target profiles in varying transfer settings.

Setting	GVO	MIIO1	MIIO2	2-Cell	4-Cell	8-Cell	16-Cell	Average
SeqN	0.4997	0.4817	0.5214	0.4918	0.5407	0.3546	0.5564	0.4923
SeqT1	0.0815	0.0074	0.0410	0.0071	0.5500	0	0.3749	0.0810
SeqT2	0.4756	0.4553	0.5045	0.4968	0.5003	0.3347	0.5643	0.4759
MetN	0.1338	0.1070	0.1807	0.7432	0.8126	0.4369	0.8004	0.4592
MetT1	0.2625	0.1332	0.1989	0.7634	0.8204	0.4924	0.7995	0.4958
MetT2	0.2868	0.1754	0.2211	0.7435	0.8367	0.5036	0.8062	0.5104
FullN	0.5101	0.5107	0.5668	0.7649	0.7955	0.5165	0.7950	0.6372
FullTS1	0.2922	0.1108	0.2905	0.7334	0.8180	0.5086	0.7809	0.5049
FullTS2	0.5791	0.5483	0.5987	0.7177	0.7791	0.5540	0.7690	0.6494
FullTB1	0.3641	0.2954	0.2244	0.7463	0.8037	0.4944	0.8118	0.5343
FullTB2	0.5733	0.5316	0.5858	0.7806	0.8085	0.5933	0.8143	0.6696
FullTA1	0.4094	0.3345	0.4017	0.7364	0.8004	0.4945	0.8027	0.5685
FullTA2	0.6275	0.5279	0.6380	0.7770	0.8357	0.5630	0.8102	0.6828

TABLE 4 | F1 score of models trained for target profiles using different objective functions and with/without transfer.

Profile	Baseline	KLD	TLR	KLD + TLR
GVO	0.3073	0.5101	0.5840	0.6275
MIIO1	0.5169	0.5107	0.5299	0.5279
MIIO2	0.3750	0.5668	0.5635	0.6380
2-Cell	0.7473	0.7650	0.7311	0.7770
4-Cell	0.8154	0.7955	0.8148	0.8357
8-Cell	0.5169	0.5165	0.4437	0.5630
16-Cell	0.7095	0.7950	0.8124	0.8102
Average	0.4923	0.6372	0.6399	0.6828
SD	0.1931	0.1401	0.1461	0.1237

Baseline: models trained using logistic loss without transfer, corresponding to DeepCpG (Angermueller et al., 2017) with the exception of the addition of batch normalization layers to facilitate training; KLD: KL divergence, models trained using KL divergence without transfer; Trn: transfer, models trained using logistic loss with transfer; KLD + Trn: models trained using KL divergence with transfer; SD: standard deviation.

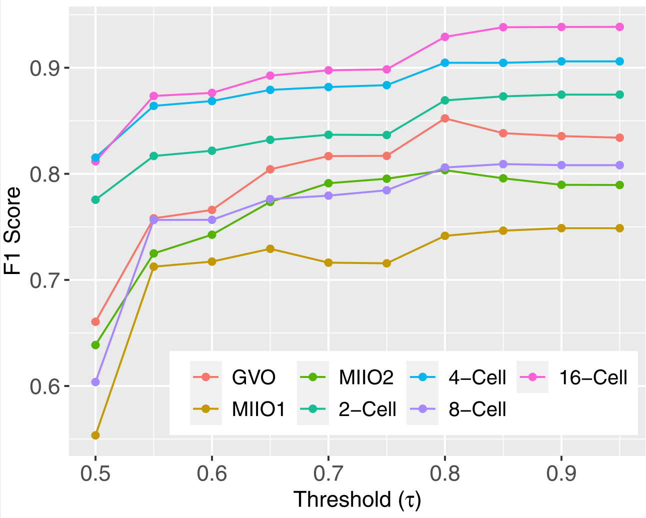


FIGURE 4 | F1 score from the use of varying threshold τ.

without fine-tuning, the obtained models almost completely failed to perform for all profiles except those at 4-cell and 16-cell stages, which have relatively higher data coverage rate (0.21 and 1.47%, respectively). Fine-tuning has the least impact on training models to predict from neighboring CpGs only, which suggests that the methylation subnetwork trained using one dataset is ready for using in models for another dataset.

To differentiate the impact of model transferring and KL divergence on models for target profiles, we trained models predicting from both sequence and neighboring CpGs in two additional settings that are using logistic loss as the training objective with and without model transferring. The performance (in F1 score) of these models, together with those trained using KL divergence with/without transferring, is presented in **Table 4** (**Supplementary Table S3** for performance by other metrics). The results indicate that both transferring and the use of KL divergence helped to improve the performance, importantly in distinct ways that are complementary to each other, since the combination of the two leads to the best-performing models. The improvement from using KL divergence by 29.43% in average F1 score (from 0.4923 to 0.6372) is similar to that from model transferring and much more significant than the similar improvement seen in the model training for source profiles. This again indicates that KL divergence is a more effective objective function to use when training models for DNA methylation prediction. It is also worth noting that both the use of KL divergence and model transferring lead to reduced variance in the performance across profiles (**Table 4**), with higher reduction seen with transferring. This suggests that the initial worse performing models gained more improvement when leveraging either KL divergence or model transferring.

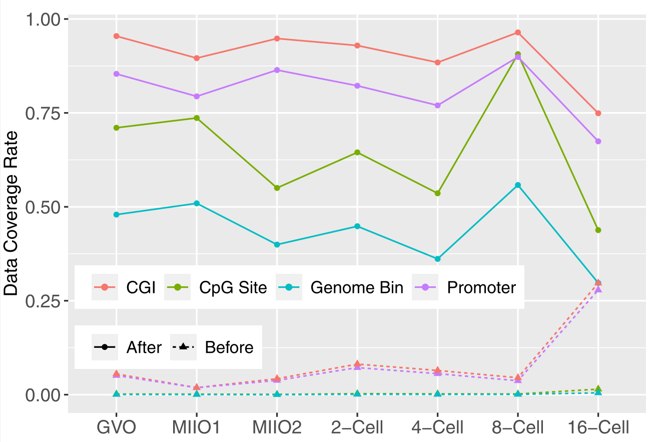


FIGURE 5 | Before and after imputation, the data coverage rate for all CpGs in the genome and three categories of genomic features: promoter, CGI, and 300bp genome bin in methylome profiles of bovine oocytes and early-stage embryos.

4.4 Imputation for Methylome Profiles of Oocytes and Early-Stage Embryos

The best-performing full models on target profiles, that is, those obtained with setting FullTA2 (**Table 2**) were used to complete the target profiles by imputing the methylation state for CpG sites that do not have experimental data. The models output probabilities of a CpG being methylated in individual profiles. To have the highest possible quality, we used a threshold τ and only kept imputed results for CpGs with a predicted probability either above τ or below $1 - \tau$. The test data used before for evaluating model performance were leveraged to find the best τ to use. For a given τ, there was no prediction being made for CpGs in the test set with a predicted probability in between $1 - \tau$ and τ. These CpGs were not considered in the subsequent F1 score calculation, leading to variation in the F1 score among different choices of τ. Intuitively, higher the τ is, more certain the prediction is and higher the calculated F1 score is. **Figure 4** shows how F1 score varies along with different choices of τ, indicating that the improvement in the F1 score becomes minimal for all

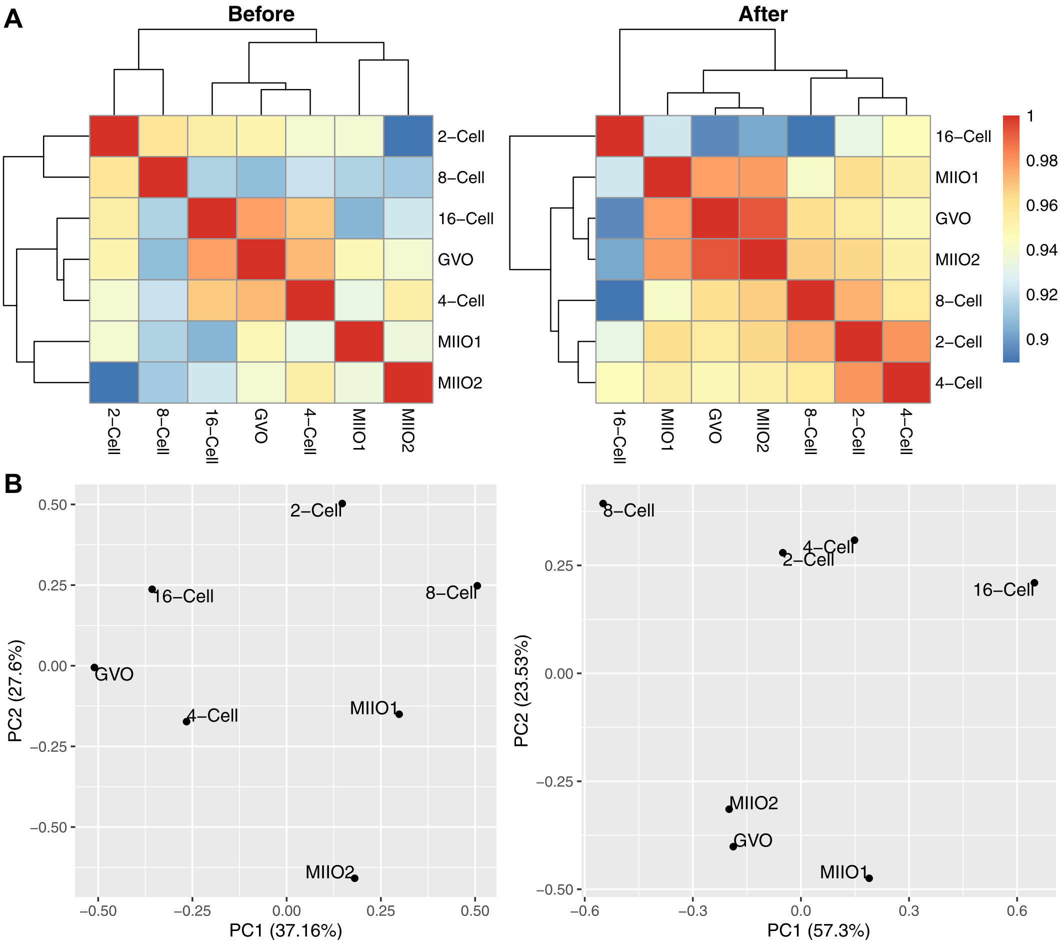


FIGURE 6 | Comparison of DNA methylation profiles of bovine oocytes and early-stage embryos before and after imputation. **(A)**: Pearson correlations among profiles; **(B)**: profiles embedded in the space spanned by the first two principal components.

profiles starting $\tau = 0.8$. As a result, 0.8 was used as the threshold in subsequent imputation for CpG sites with missing data.

Large number of missing CpGs had imputed data in each individual profile, leading to a drastic increase in the data coverage rate from the initial range of 0.06–1.47% to that of 43.80–73.65% (**Figure 5**). To demonstrate the impact of imputed data on subsequent analyses of functional genomics, we compared the number of genomic features that are considered to have data before and after the imputation. Three categories of genomic features were considered: genome bin, promoter of gene, and CGI. Genome bins were obtained by tiling the reference genome to produce equal-sized and nonoverlapping bins of 300 bp long each. Promoters were defined by 1001bp regions centered at annotated transcription starting site of genes, which were obtained from Ensembl Genome Browser. The CGI annotations were downloaded from UCSC Genome Browser. A genome bin was considered to have data when there were at least three CpG sites with known methylation state within the bin; while given its longer length, a promoter (or a CGI) was considered to have data when there were at least 10 CpG sites with known state within the promoter (or CGI) region. The percentages of genome bins, promoters, and CGIs that were considered to have data out of total 8,869,705, 22,118, and 37,226, respectively, before and after imputation in individual profiles are shown in **Figure 5**. As for individual CpG sites, substantial increase in the data coverage can be seen for all three categories of genomic features. Specifically, the coverage rate was increased to 29.74–55.80% from 0.02 to 0.48% for genome bins, to 67.44–89.90% from 1.85 to 27.86% for promoters, and to 74.92–96.42% from 1.87 to 29.70% for CGIs. The expanded data will greatly enhance the analyses to understand the mechanisms underlying DNA methylation and its role in regulating various biological functions.

To demonstrate the impact of imputation on downstream analyses, we calculated the Pearson correlation between each pair of profiles before and after imputation, followed by hierarchical clustering to group profiles. In addition, we performed principal component analysis (PCA) on profiles before and after imputation. The methylation level of 300 bp genome bins (assessed by the average methylation level of CpGs within each bin) was used as input data for these analyses with excluding bins that have missing data in any of the profiles. The results are provided in **Figure 6**, indicating that imputation helped to obtain profile groupings that better align with existing biological understandings. Specifically, the grouping of the three oocytes profiles and that of 2-cell and 4-cell profiles followed by grouping with 8-cell and 16-cell profiles after imputation (top right, **Figure 6**) align well with the natural reproductive phases. In contrast, the groupings obtained before imputation (top left, **Figure 6**) lack clear biological interpretation. The PCA plots with profiles embedded in the two-dimensional space spanned by the first two principal components (bottom panel, **Figure 6**) also indicate the same story.

5 CONCLUSION

Here, we reported our exploration of utilizing transfer learning and KL divergence in training DNNs to impute for DNA methylome profiles with very low data coverage. The target profiles to complete in our study are those of bovine oocytes and early embryos by WGBS with a data coverage rate ranging from 0.06 to 1.47% after cleaning. To obtain pre-trained models for transferring, WGBS profiles of sperm and a wide range of somatic tissues (coverage rate: 0.85–21.34%) were utilized. The results of our analyses indicate that both model transferring and KL divergence improve the prediction performance of the target models.

Our study demonstrated that KL divergence is a more effective objective function to use than the commonly used logistic loss for training models to prediction DNA methylation. Compared to logistic loss, the use of KL divergence led to models with improved performance in the training for both source and target profiles. Note that KL divergence helps to boost the average F1 score to 0.6372 from 0.4923 across target profiles, which is a much larger increase compared to that seen in source model training (from 0.7585 to 0.7700). This suggests that the use of KL divergence is especially beneficial when the data coverage rate is low, which makes sense as the ability of utilizing as much information carried in the data as possible is of greater importance in the case of limited training set size. Our results also demonstrated that the transferring of models built for profiles with relatively high coverage greatly improves training for those that are in low coverage, with increased average F1 score 0.6399 (from 0.4923). Importantly, model transferring and KL divergence enhance the training of target models in two distinctive ways that are additive, evidenced by the further improved performance (average F1 score: 0.6828) when both were exploited simultaneously. Moreover, our exploration further into the different components of the adopted DNN indicates that local methylation patterns are more transferable across datasets than learned DNA sequence patterns. Finally, to obtain the best models for target profiles, fine-tuning is necessary regardless of which components of the source model are transferred.

The results from the subsequent application of trained models for imputation demonstrated the high effectiveness of our approach in completing DNA methylome profiles that have very low data coverage. Drastic increase in data coverage rate after imputation were seen at both individual CpG sites and varying genomic features, including genome bins, gene promoters, and CGIs. The imputed data would greatly strengthen analyses toward the understanding of biological mechanisms and functional roles of DNA methylation. One of our future works will be to link the methylation level of genomic features to transcriptomic profiles to understand how DNA methylation regulates gene expression as a *cis* regulator. The results from such an analysis will allow more accurate reconstruction of gene regulatory networks underlying a biological system, which is also our future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found at: NCBI GEO: GSE106538, GSE147087, GSE121758. The code, including the implementation of network architecture and the training and evaluating the models is available at the following URL: https://github.com/ODU-CSM/Pub-Met-TL.

ETHICS STATEMENT

Ethical review and approval were not required for the animal study because the data were downloaded from the public data repository.

AUTHOR CONTRIBUTIONS

JS and ZJ conceived the study. JS designed the study. JS and SD implemented the proposed approach, conducted experiments, and analyzed the results. All authors participated in the development of the manuscript.

FUNDING

This study was supported by the NIH National Heart, Lung, and Blood Institute with grant no. R01HL157519. Additional support for JZ was provided by the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD102533) and USDA National Institute of Food and Agriculture (2019-67016-29863).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.910439/full#supplementary-material

REFERENCES

Abascal, F., Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., et al. (2020). Expanded Encyclopaedias of DNA Elements in the Human and Mouse Genomes. *Nature* 583, 699–710. doi:10.1038/s41586-020-2493-4

Angermueller, C., Lee, H. J., Reik, W., and Stegle, O. (2017). Erratum to: DeepCpG: Accurate Prediction of Single-Cell DNA Methylation States Using Deep Learning. *Genome Biol.* 18, 90–13. doi:10.1186/s13059-017-1233-z

Angermueller, C., Lee, H. J., Reik, W., and Stegle, O. (2017). DeepCpG: Accurate Prediction of Single-Cell DNA Methylation States Using Deep Learning. *Genome Biol.* 18, 67. doi:10.1186/s13059-017-1189-z

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., et al. (2021). Effective Gene Expression Prediction from Sequence by Integrating Long-Range Interactions. *Nat. Methods* 18, 1196–1203. doi:10.1038/s41592-021-01252-x

Bhasin, M., Zhang, H., Reinherz, E. L., and Reche, P. A. (2005). Prediction of Methylated CpGs in DNA Sequences Using a Support Vector Machine. *FEBS Lett.* 579, 4302–4308. doi:10.1016/j.febslet.2005.07.002

Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T., and Walter, J. (2006). CpG Island Methylation in Human Lymphocytes Is Highly Correlated with DNA Sequence, Repeats, and Predicted DNA Structure. *PLoS Genet.* 2, e26–0252. doi:10.1371/journal.pgen.0020026

Cho, K., van Merrienboer, B., Bahdanau, D., and Bengio, Y. (2014). "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," in *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, 2014. doi:10.3115/v1/w14-4012

Clark, S. J., Smallwood, S. A., Lee, H. J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide Base-Resolution Mapping of DNA Methylation in Single Cells Using Single-Cell Bisulfite Sequencing (scBS-Seq). *Nat. Protoc.* 12, 534–547. doi:10.1038/nprot.2016.187

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing. *Nat. Nanotech* 4, 265–270. doi:10.1038/nnano.2009.12

Das, R., Dimitrova, N., Xuan, Z., Rollins, R. A., Haghighi, F., Edwards, J. R., et al. (2006). Computational Prediction of Methylation Status in Human Genomic Sequences. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10713–10716. doi:10.1073/pnas.0602949103

De Waele, G., Clauwaert, J., Menschaert, G., and Waegeman, W. (2022). CpG Transformer for Imputation of Single-Cell Methylomes. *Bioinformatics* 38, 597–603. doi:10.1093/bioinformatics/btab746

Duan, J. E., Jiang, Z. C., Alqahtani, F., Mandoiu, I., Dong, H., Zheng, X., et al. (2019). Methylome Dynamics of Bovine Gametes and *In Vivo* Early Embryos. *Front. Genet.* 10, 512. doi:10.3389/fgene.2019.00512

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., et al. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* 489, 57–74. doi:10.1038/nature11247

Elliott, G., Hong, C., Xing, X., Zhou, X., Li, D., Coarfa, C., et al. (2015). Intermediate DNA Methylation Is a Conserved Signature of Genome Regulation. *Nat. Commun.* 6. doi:10.1038/ncomms7363

Ernst, J., and Kellis, M. (2015). Large-scale Imputation of Epigenomic Datasets for Systematic Annotation of Diverse Human Tissues. *Nat. Biotechnol.* 33, 364–376. doi:10.1038/nbt.3157

Fan, S., Zhang, M. Q., and Zhang, X. (2008). Histone Methylation Marks Play Important Roles in Predicting the Methylation Status of CpG Islands. *Biochem. Biophysical Res. Commun.* 374, 559–564. doi:10.1016/j.bbrc.2008.07.077

Fang, F., Fan, S., Zhang, X., and Zhang, M. Q. (2006). Predicting Methylation Status of CpG Islands in the Human Brain. *Bioinformatics* 22, 2204–2209. doi:10.1093/bioinformatics/btl377

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., et al. (2010). Conservation and Divergence of Methylation Patterning in Plants and Animals. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8689–8694. doi:10.1073/pnas.1002720107

Fu, L., Peng, Q., and Chai, L. (2019). Predicting DNA Methylation States with Hybrid Information Based Deep-Learning Model. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 17, 1. doi:10.1109/tcbb.2019.2909237

Grant, M., Zuccotti, M., and Monk, M. (1992). Methylation of CpG Sites of Two X-Linked Genes Coincides with X-Inactivation in the Female Mouse Embryo but Not in the Germ Line. *Nat. Genet.* 2, 161–166. doi:10.1038/ng1092-161

Greenberg, M. V. C., and Bourc'his, D. (2019). The Diverse Roles of Dna Methylation in Mammalian Development and Disease. *Nat. Rev. Mol. Cell. Biol.* 20, 590–607. doi:10.1038/s41580-019-0159-6

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of Reduced Representation Bisulfite Sequencing Libraries for Genome-Scale DNA Methylation Profiling. *Nat. Protoc.* 6, 468–481. doi:10.1038/nprot.2010.190

Islam, A., Chen, C.-F., Panda, R., Karlinsky, L., Radke, R., and Feris, R. (2021). A Broad Study on the Transferability of Visual Representations with Contrastive Learning. *arXiv Prepr. arXiv:2103.13517*. doi:10.1109/iccv48922.2021.00872

Jiang, L., Wang, C., Tang, J., and Guo, F. (2019). LightCpG: A Multi-View CpG Sites Detection on Single-Cell Whole Genome Sequence Data. *BMC Genomics* 20, 1–17. doi:10.1186/s12864-019-5654-9

Kapourani, C.-A., and Sanguinetti, G. (2019). Melissa: Bayesian Clustering and Imputation of Single-Cell Methylomes. *Genome Biol.* 20, 1–15. doi:10.1186/s13059-019-1665-8

Kernohan, K. D., Cigana Schenkel, L., Cigana Schenkel, L., Huang, L., Smith, A., Pare, G., et al. (2016). Identification of a Methylation Profile for DNMT1-Associated Autosomal Dominant Cerebellar Ataxia, Deafness, and Narcolepsy. *Clin. Epigenet* 8, 4–9. doi:10.1186/s13148-016-0254-x

Kim, S., Li, M., Pair, H., Nephew, K., Shi, H., Kramer, R., et al. (2007). Predicting DNA Methylation Susceptibility Using CpG Flanking Sequences. *Pac. Symposium Biocomput. Pac. Symposium Biocomput.* 2008, 315–326. doi:10.1142/9789812776136_0031

Kingma, D. P., and Ba, J. (2015). "Adam: A Method for Stochastic Optimization," in International Conference for Learning Representations, 1–15.

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., et al. (2010). Impaired Hydroxylation of 5-methylcytosine in Myeloid Cancers with Mutant TET2. *Nature* 468, 839–843. doi:10.1038/nature09586

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks," in *NIPS*. doi:10.1061/(ASCE)GT.1943-5606.0001284

Kullback, S., and Leibler, R. A. (1951). On Information and Sufficiency. *Ann. Math. Stat.* 22, 79–86. doi:10.1214/aoms/1177729694

Roadmap Epigenomics Consortium., Kundaje, A., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative Analysis of 111 Reference Human Epigenomes. *Nature* 518, 317–330. doi:10.1038/nature14248

Levy-Jurgenson, A., Tekpli, X., Kristensen, V. N., and Yakhini, Z. (2019). Predicting Methylation from Sequence and Gene Expression Using Deep Learning with Attention. *Algorithms Comput. Biol.* 2019, 179–190. doi:10.1007/978-3-030-18174-1_13

Li, C., Sun, J., Liu, Q., Dodlapati, S., Ming, H., Wang, L., et al. (2021). The Landscape of Accessible Chromatin in Quiescent Cardiac Fibroblasts and Cardiac Fibroblasts Activated after Myocardial Infarction. *Epigenetics* 2021, 1–20. doi:10.1080/15592294.2021.1982158

Liu, Q., Xia, F., Yin, Q., and Jiang, R. (2018). Chromatin Accessibility Prediction via a Hybrid Deep Convolutional Neural Network. *Bioinformatics* 34, 732–738. doi:10.1093/bioinformatics/btx679

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., et al. (2021). DNA Methylation-Calling Tools for Oxford Nanopore Sequencing: a Survey and Human Epigenome-wide Evaluation. *Genome Biol.* 22. doi:10.1186/s13059-021-02510-z

López-García, G., Jerez, J. M., Franco, L., and Veredas, F. J. (2020). Transfer Learning with Convolutional Neural Networks for Cancer Survival Prediction Using Gene-Expression Data. *PloS one* 15, e0230536. doi:10.1371/journal.pone.0230536

Lu, L., Lin, K., Qian, Z., Li, H., Cai, Y., and Li, Y. (2010). Predicting DNA Methylation Status Using Word Composition. *JBiSE* 03, 672–676. doi:10.4236/jbise.2010.37091

Ma, B., Wilker, E. H., Willis-Owen, S. A. G., Byun, H.-M., Wong, K. C. C., Motta, V., et al. (2014). Predicting DNA Methylation Level across Human Tissues. *Nucleic Acids Res.* 42, 3515–3528. doi:10.1093/nar/gkt1380

Otter, D. W., Medina, J. R., and Kalita, J. K. (2021). A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 604–624. doi:10.1109/TNNLS.2020.2979670

Painsky, A., and Wornell, G. (2018). On the Universality of the Logistic Loss Function. *IEEE Int. Symposium Inf. Theory - Proc.* 2018, 936–940. doi:10.1109/ISIT.2018.8437786

Petryk, N., Bultmann, S., Bartke, T., and Defossez, P.-A. (2020). Staying True to Yourself: Mechanisms of DNA Methylation Maintenance in Mammals. *Nucleic Acids Res.* 49, 3020–3032. doi:10.1093/nar/gkaa1154

Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., et al. (2012). Protection against De Novo Methylation Is Instrumental in Maintaining Parent-Of-Origin Methylation Inherited from the Gametes. *Mol. Cell.* 47, 909–920. doi:10.1016/j.molcel.2012.07.010

Rajshekar, S., Yao, J., Arnold, P. K., Payne, S. G., Zhang, Y., Bowman, T. V., et al. (2018). Pericentromeric Hypomethylation Elicits an Interferon Response in an Animal Model of ICF Syndrome. *eLife* 7, e39658. doi:10.7554/eLife.39658

Rauluseviciute, I., Drabløs, F., and Rye, M. B. (2019). DNA Methylation Data by Sequencing: Experimental Approaches and Recommendations for Tools and Pipelines for Data Analysis. *Clin. Epigenet* 11, 1–13. doi:10.1186/s13148-019-0795-x

Russler-Germain, D. A., Spencer, D. H., Young, M. A., Lamprecht, T. L., Miller, C. A., Fulton, R., et al. (2014). The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-type DNMT3A by Blocking its Ability to Form Active Tetramers. *Cancer Cell.* 25, 442–454. doi:10.1016/j.ccr.2014.02.010

Sevakula, R. K., Singh, V., Verma, N. K., Kumar, C., and Cui, Y. (2018). Transfer Learning for Molecular Cancer Classification Using Deep Neural Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform* 16, 2089–2100. doi:10.1109/TCBB.2018.2822803

Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., et al. (2014). Single-cell Genome-wide Bisulfite Sequencing for Assessing Epigenetic Heterogeneity. *Nat. Methods* 11, 817–820. doi:10.1038/nmeth.3035

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., et al. (2011). DNA-binding Factors Shape the Mouse Methylome at Distal Regulatory Regions. *Nature* 480, 490–495. doi:10.1038/nature10716

Stevens, M., Cheng, J. B., Li, D., Xie, M., Hong, C., Maire, C. L., et al. (2013). Estimating Absolute Methylation Levels at Single-CpG Resolution from Methylation Enrichment and Restriction Enzyme Sequencing Methods. *Genome Res.* 23, 1541–1553. doi:10.1101/gr.152231.112

Sun, Z., Wu, Y., Ordog, T., Baheti, S., Nie, J., Duan, X., et al. (2014). Aberrant Signature Methylome by DNMT1 Hot Spot Mutation in Hereditary Sensory and Autonomic Neuropathy 1E. *Epigenetics* 9, 1184–1193. doi:10.4161/epi.29676

Taiwo, O., Wilson, G. A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., et al. (2012). Methylome Analysis Using MeDIP-Seq with Low DNA Concentrations. *Nat. Protoc.* 7, 617–636. doi:10.1038/nprot.2012.012

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). "A Survey on Deep Transfer Learning," in *International Conference on Artificial Neural Networks* (Berlin, Germany: Springer), 270–279. doi:10.1007/978-3-030-01424-7_27

Tang, J., Zou, J., Fan, M., Tian, Q., Zhang, J., and Fan, S. (2021). CaMelia: Imputation in Single-Cell Methylomes Based on Local Similarities between Cells. *Bioinformatics* 37, 1814–1820. doi:10.1093/bioinformatics/btab029

Vandiver, A. R., Idrizi, A., Rizzardi, L., Feinberg, A. P., and Hansen, K. D. (2015). DNA Methylation Is Stable during Replication and Cell Cycle Arrest. *Sci. Rep.* 5, 17911. doi:10.1038/srep17911

Wang, J., Agarwal, D., Huang, M., Hu, G., Zhou, Z., Ye, C., et al. (2019). Data Denoising with Transfer Learning in Single-Cell Transcriptomics. *Nat. Methods* 16, 875–878. doi:10.1038/s41592-019-0537-1

Wang, Y., Liu, T., Xu, D., Shi, H., Zhang, C., Mo, Y.-Y., et al. (2016). Predicting DNA Methylation State of CpG Dinucleotide Using Genome Topological Features and Deep Networks. *Sci. Rep.* 6, 1–15. doi:10.1038/srep19598

Xiao, F.-H., Wang, H.-T., and Kong, Q.-P. (2019). Dynamic DNA Methylation during Aging: A "Prophet" of Age-Related Outcomes. *Front. Genet.* 10, 1–8. doi:10.3389/fgene.2019.00107

Yu, F., Xu, C., Deng, H.-W., and Shen, H. (2020). A Novel Computational Strategy for DNA Methylation Imputation Using Mixture Regression Model (MRM). *BMC Bioinforma.* 21, 1–17. doi:10.1186/s12859-020-03865-z

Yuen, Z. W.-S., Srivastava, A., Daniel, R., McNevin, D., Jack, C., and Eyras, E. (2021). Systematic Benchmarking of Tools for CpG Methylation Detection from Nanopore Sequencing. *Nat. Commun.* 12, 1–12. doi:10.1038/s41467-021-23778-6

Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* 328, 916–919. doi:10.1126/science.1186366

Zeng, H., and Gifford, D. K. (2017). Predicting the Impact of Non-coding Variants on DNA Methylation. *Nucleic acids Res.* 45, e99. doi:10.1093/nar/gkx177

Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T., and Engelhardt, B. E. (2015). Predicting Genome-wide DNA Methylation Using Methylation Marks, Genomic Position, and DNA Regulatory Elements. *Genome Biol.* 16, 1–20. doi:10.1186/s13059-015-0581-9

Zhang, Y., and Hamada, M. (2018). DeepM6ASeq: Prediction and Characterization of m6A-Containing Sequences Using Deep Learning. *BMC Bioinforma.* 19, 1–11. doi:10.1186/s12859-018-2516-4

Zheng, H., Wu, H., Li, J., and Jiang, S.-W. (2013). CpGIMethPred: Computational Model for Predicting Methylation Status of CpG Islands in Human Genome. *BMC Med. Genomics* 6, 1–12. doi:10.1186/1755-8794-6-S1-S13

Zhou, J., and Troyanskaya, O. G. (2015). Predicting Effects of Noncoding Variants with Deep Learning-Based Sequence Model. *Nat. Methods* 12, 931–934. doi:10.1038/nmeth.3547

Zhou, X., Chai, H., Zhao, H., Luo, C. H., and Yang, Y. (2020a). Imputing Missing RNA-Sequencing Data from DNA Methylation by Using a Transfer Learning-Based Neural Network. *GigaScience* 9, giaa076. doi:10.1093/gigascience/giaa076

Zhou, X., Li, Z., Dai, Z., and Zou, X. (2012). Prediction of Methylation CpGs and Their Methylation Degrees in Human DNA Sequences. *Comput. Biol. Med.* 42, 408–413. doi:10.1016/j.compbiomed.2011.12.008

Zhou, Y., Connor, E. E., Bickhart, D. M., Li, C., Baldwin, R. L., Schroeder, S. G., et al. (2018). Comparative Whole Genome DNA Methylation Profiling of Cattle Sperm and Somatic Tissues Reveals Striking Hypomethylated Patterns in Sperm. *GigaScience* 7, 1–13. doi:10.1093/gigascience/giy039

Zhou, Y., Liu, S., Hu, Y., Fang, L., Gao, Y., Xia, H., et al. (2020b). Comparative Whole Genome DNA Methylation Profiling across Cattle Tissues Reveals Global and Tissue-specific Methylation Patterns. *BMC Biol.* 18, 85. doi:10.1186/s12915-020-00793-5

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., et al. (2018). Single-cell DNA Methylome Sequencing of Human Preimplantation Embryos. *Nat. Genet.* 50, 12–19. doi:10.1038/s41588-017-0007-6

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A Comprehensive Survey on Transfer Learning. *Proc. IEEE* 109, 43–76.

Zou, L. S., Erdos, M. R., Erdos, M. R., Taylor, D. L., Chines, P. S., Varshney, A., et al. (2018). BoostMe Accurately Predicts DNA Methylation Values in Whole-Genome Bisulfite Sequencing of Multiple Human Tissues. *BMC Genomics* 19, 390. doi:10.1186/s12864-018-4766-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The code, including the implementation of network architecture and the training and evaluating the models is available at the following URL: https://github.com/ODU-CSM/Pub-Met-TL.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 22 August 2022
doi: 10.3389/fgene.2022.943058

Clinical Utility of Medical Exome Sequencing: Expanded Carrier Screening for Patients Seeking Assisted Reproductive Technology in China

Keya Tong[1,2†], **Wenbin He**[3,4,5†], **Yao He**[1,2,6], **Xiurong Li**[3], **Liang Hu**[3], **Hao Hu**[3], **Guangxiu Lu**[3,4,5], **Ge Lin**[3,4,5], **Chang Dong**[7], **Victor Wei Zhang**[7], **Juan Du**[3,4,5]* and **Dongyun Liu**[1,2,6]*

[1]Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University, Chongqing, China, [2]Chongqing Key Laboratory of Human Embryo Engineering, Women and Children's Hospital of Chongqing Medical University, Chongqing, China, [3]National Engineering and Research Center of Human Stem Cells, Changsha, China, [4]School of Basic Medical Science, Institute of Reproductive and Stem Cell Engineering, Central South University, Changsha, China, [5]Genetics Centre, Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha, China, [6]Chongqing Clinical Research Center for Reproductive Medicine, Women and Children's Hospital of Chongqing Medical University, Chongqing, China, [7]AmCare Genomics Lab, Guangzhou, China

OPEN ACCESS

Edited by:
Mengcheng Luo,
Wuhan University, China

Reviewed by:
Jin Haixia,
First Affiliated Hospital of Zhengzhou University, China
Yongjuan Guan,
University of Pennsylvania, United States

***Correspondence:**
Juan Du
tandujuan@csu.edu.cn
Dongyun Liu
dongyunliu@163.com

†These authors have contributed equally to this work and share first authorship.

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *13 May 2022*
Accepted: *20 June 2022*
Published: *22 August 2022*

Citation:
Tong K, He W, He Y, Li X, Hu L, Hu H, Lu G, Lin G, Dong C, Zhang VW, Du J and Liu D (2022) Clinical Utility of Medical Exome Sequencing: Expanded Carrier Screening for Patients Seeking Assisted Reproductive Technology in China. Front. Genet. 13:943058. doi: 10.3389/fgene.2022.943058

Purpose: Expanded carrier screening (ECS) is an effective method to identify at-risk couples (ARCs) and avoid birth defects. This study aimed to reveal the carrier spectrum in the Chinese population and to delineate an expanded carrier gene panel suitable in China.

Methods: Medical exome sequencing (MES), including 4,158 disease-causing genes, was offered to couples at two reproductive centers. It was initially used as a diagnostic yield for potential patients and then used for ECS. Clinical information and ECS results were retrospectively collected.

Results: A total of 2,234 couples, representing 4,468 individuals, underwent MES. In total, 254 individuals showed genetic disease symptoms, and 56 of them were diagnosed with genetic diseases by MES. Overall, 94.5% of them were carriers of at least one disease-causing variant. The most prevalent genes were *GJB2* for autosomal recessive disorders and *G6PD* for X-linked diseases. The ARC rate was 9.80%, and couples were inclined to undergo preimplantation genetic testing when diseases were classified as "profound" or "severe."

Conclusion: This study provided insight to establish a suitable ECS gene panel for the Chinese population. Disease severity significantly influenced reproductive decision-making. The results highlighted the importance of conducting ECS for couples before undergoing assisted reproductive technology.

Keywords: expanded carrier screening, Han Chinese ethnicity, assisted reproductive technology, preimplantation genetic testing, clinical utility

1 INTRODUCTION

In the past few decades, the development of sequencing technology and the increased awareness of rare inherited diseases have led to the elucidation of a significant number of hereditary diseases. The Online Mendelian Inheritance in Man (OMIM) database has recorded more than 8,000 monogenic diseases to date. In mainland China, the incidence rate of birth defects is approximately 5.6%, and genetic factors account for 40–50% of them. (NHC, 2018)

The objective of carrier screening is to identify asymptomatic individuals who are carrying heterozygous disease-causing variants to avoid the possibility of conceiving offspring with birth defects. It was first proposed to screen Tay–Sachs disease carriers in the Ashkenazi Jewish population. (Kaback, 2000) The strategies for carrier screening markedly reduced the incidence of certain diseases in at-risk populations. (Antonarakis, 2019) The emergence of next-generation sequencing (NGS) has made it possible to simultaneously detect thousands of conditions at an affordable cost and rapid turnaround time. (Tucker et al., 2009) Benefiting from high-throughput sequencing, carrier screening has transitioned from limited ethnicity to general population implementation and from a few diseases to multiple heritable disorders. In 2011, Bell et al. (2011) screened 448 selected severe recessive childhood diseases and found an average genomic carrier burden of 2.8%, and it was the first time NGS was utilized in carrier screening.

Multiple facilities conducted ECS using designed panels for different populations, such as early pregnancy women, ART patients, or the general population. Martin et al. 2015 tested 549 autosomal recessive and X-linked genes for ART-seeking couples in Spain and found that 84% of these individuals were carriers of at least one condition. The investigator estimated that carrier screening prevented 1.25% of birth defects. Franasiak et al. (2016) found that the ECS outcomes affected reproductive decisions in 0.21% of cases after screening 117 conditions in patients in an infertility care center. Haquel et al. (2016) screened 417 pathogenic (P) variants in 94 genes and estimated that the risk of having an autosomal recessive disorder-affected child was at a minimum of 1 in 628 pregnancies in Northern European couples.

In 2015, the American College of Medical Genetics (ACMG) along with other professional societies launched a joint statement focused on principles of expanded carrier screening (ECS) for prenatal or preconception and criteria for panel design. (Edwards et al., 2015) Afterward, the American College of Obstetricians and Gynecologists (ACOG) published recommendations, suggesting that conditions suitable for the ECS panel should meet the following criteria: 1) carrier frequencies equal to or greater than 1/100, 2) clear and definite genotype–phenotype relationships, 3) early onset in life, 4) shortened lifespan, 5) cognitive or physical disability that affects quality of life, and 6) surgical or medical intervention requested. In addition, the conditions included in ECS should be accessed for prenatal diagnosis, and early intervention should be affordable and improve infant outcomes. (Opinion No, 2017) In the latest published ACMG recommendations in 2021, Gregg et al. (2021) proposed 113 autosomal recessive and X-linked genes that were appropriate for ECS. The recommendations have taken into consideration autosomal recessive carrier frequencies equal to or greater than 1/200 and X-linked conditions.

Although the ACMG and the ACOG have made recommendations and guidelines for carrier screening, it is noticed that such announcements are current conclusions drawn from the population in the United States and Europe, rather than extending the investigation to other countries. In particular, carrier screening studies that focused on the Chinese population are limited. The carrier frequencies and ARC rates detected in ECS highly depend on the panel selection. Guo and Gregg (2019) used an exome sequencing database to estimate the carrier frequencies in six different ancestries for 415 genes associated with severe hereditary diseases. They found that the cumulative carrier rates (CCR) were up to 62.9% in the Ashkenazi Jewish population while only 32.6% in the East Asian population. The result implied that the current expanded carrier sequencing panel cannot fully cover the variant spectrum in the East Asian population.

This study aimed to reveal the carrier spectrum in the Chinese population and delineate a gene set suitable for ECS in China. The study retrospectively analyzed the results of medical exome sequencing (MES) for couples seeking advice at reproductive centers.

2 MATERIALS AND METHODS

2.1 Study Design

As presented in **Figure 1**, we retrospectively gathered couples from reproductive medicine centers at the Chongqing Health Center for Women and Children and the Genetics Center of Reproductive and Genetic Hospital in CITIC-Xiangya between January 2019 and June 2021. The clinical features of the couples were assessed, and necessary examinations were performed by physicians. ECS was offered to couples who were planning to undergo assisted reproductive technology (ART) or seeking preconception genetic counseling. In particular, MES was first used to identify the variants related to genetic disorders for patients, followed by ECS. The usage and possible outcomes of ECS were explained by physicians to patients before the test. The couples self-reported their age, ethnicity, family history, and gestational history. The following sequencing results were reported to the couples: 1) variants that can explain one's clinical symptoms (if detected), 2) at-risk couples (ARCs) (see definition in **Section 2.2**), and 3) carrier status and secondary findings (SFs) (if detected). Post-test counseling was offered to all couples, and ARCs would receive additional ART recommendations from their physicians. Information on medical history, clinical features, ECS results, and post-test interventions were collected. To estimate carrier frequency, P and likely pathogenic (LP) variants from each individual were collected, whereas variants associated with clinical features were excluded from the analysis. Informed consent was signed by all patients.

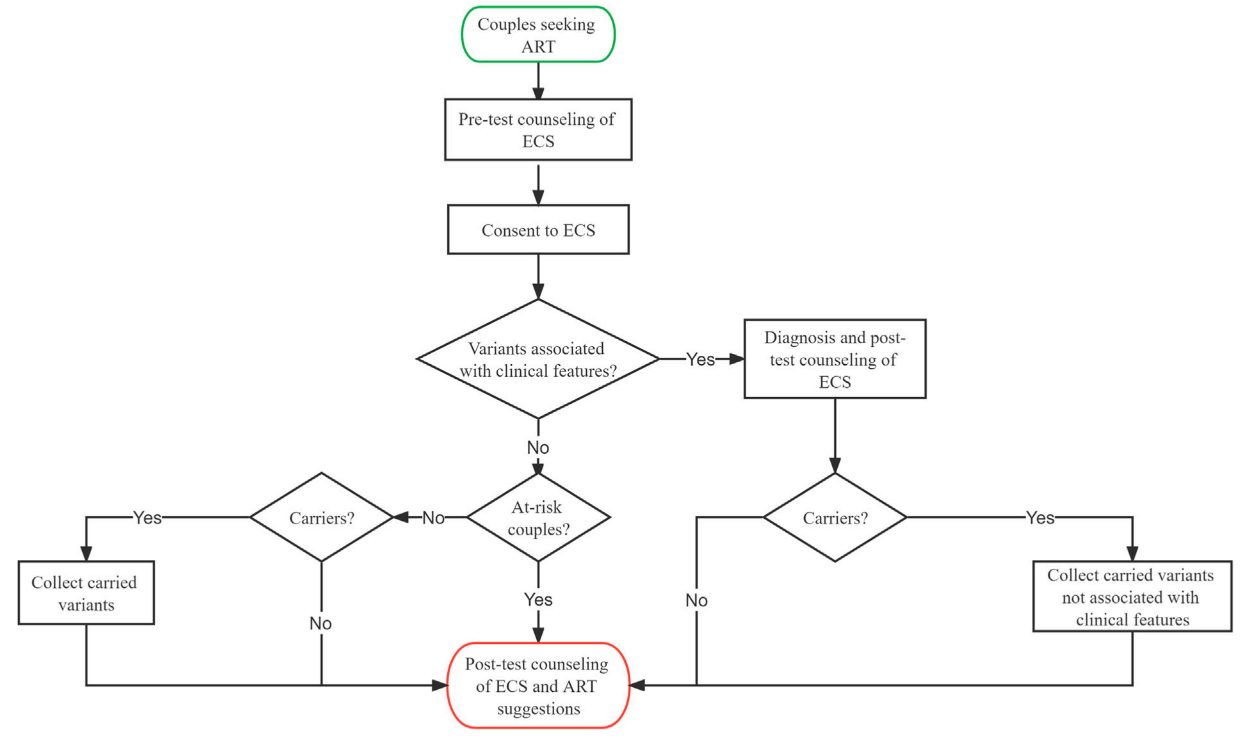


FIGURE 1 | | Clinical workflow of expanded carrier screening in the reproductive center.

2.2 At-Risk Couple Identification

At-risk couple is defined as follows: both partners carrying a P or LP variant of the same autosomal recessive gene, the female partner carrying a P or LP variant in an X-linked gene, one of the partners carrying a P or LP variant while the other is carrying a variant of the same gene but classified as variant of uncertain significance (VUS), or both partners carrying a VUS of the same gene. The family history of the couples was carefully reviewed. If the proband was available, Sanger sequencing validation was performed on the proband, and the couples were considered ARC when the identified variants segregated in a Mendelian manner. In contrast, if no proband was available, software prediction was used to predict the structure, conservation domain, and function domain for VUS. When the prediction results indicated that the VUS may compromise protein function, the couples were also considered as ARCs.

There are several reasons indicated for ECS, and screened couples were divided into six subgroups: 1) infertility, couples are unable to get pregnant naturally; 2) recurrent miscarriage/ stillborn, couples experienced recurrent miscarriage or stillborn; 3) birth defect/ultrasound anomaly, couples gave birth to affected children of known or suspicious genetic diseases or in whom fetal abnormalities were found in prenatal ultrasound; 4) patients, one or both of the partners had clinical manifestations suspected to be genetic diseases; 5) consanguineous marriage, couples were lineal or collateral blood relatives within three generations; and 6) routine screening, both partners were healthy individuals with no family history.

2.3 Medical Exome Sequencing

We utilized custom-designed NimbleGen SeqCap probes (Roche NimbleGen, Madison, WI, United States) for in-solution hybridization to enrich target sequences. Target genes included 4,158 genes collected from the OMIM database (updated on June 2021) with definite corresponding diseases, comprising AR, X-linked, and AD genes. Only genes with well-defined genotype–phenotype relationships were included in the panel. The sequencing covered coding exons, and known P variants reported in deep introns or non-coding regions were also included. In terms of gene inclusion, this is the largest carrier screening panel for ECS to date.

2.4 Next-Generation Sequencing and Data Analysis

Genomic DNA was extracted from peripheral blood using the Solpure Blood DNA Kit (Magen). To augment target sequences, we used custom-designed NimbleGen SeqCap probes (Roche NimbleGen, Madison, WI, United States) for in-solution hybridization. DNA samples were indexed and sequenced on the AmCareSeq 2000 (Amcare, Guangzhou, China). The average coverage depth was about 200× with over 98% of the target regions covered by at least 20 reads. The sequenced reads were compared with the reference human genome version (GRCh37/ hg19). Nucleotide changes found in aligned reads were pulled and analyzed using the NextGENe software (Version 2.4.2) (SoftGenetics, State College, PA, United States). Online

TABLE 1 | Demographic and carrier status of 4,468 individuals.

Subgroup	Screened	Percentage (%)	Age		Carrier Status		ARC rate
	n		Female	Male	Positive	Negative	
All	4,468	100	31.4 ± 4.7	33.1 ± 5.7	4,242	226	9.80%
Recurrent miscarriage/stillborn	1756	39	32.2 ± 4.6	34.2 ± 5.4	1,653	103	6.49%
Infertility	1,174	26	29.6 ± 4.0	31.3 ± 5.1	1,119	55	9.71%
Birth defect/ultrasound anomaly	894	20	31.6 ± 4.8	33.5 ± 5.0	852	42	10.24%
Patients	254	6	28.4 ± 5.8	31.7 ± 6.1	239	15	15.66%
Routine screening	340	8	31.2 ± 4.7	31.5 ± 4.2	331	9	8.24%
Consanguineous marriage	50	1	25.3 ± 2.1	26.7 ± 3.2	48	2	32.00%

software programs PolyPhen-2, SIFT, PROVEAN, MutationTaster, and GeneSplicer were used for *in silico* analysis. The software programs and their corresponding cutoff values were as follows: SIFT (score < 0.05, deleterious; score ≥ 0.05, tolerated), PolyPhen-2 (score ≥ 0.909, probably damaging; score: 0.447–0.909, possibly damaging; score ≤ 0.446, B: benign), PROVEAN (score:−14–2.5, deleterious; score: −2.5–14, neutral), and MutationTaster (A: disease-causing automatic, D: disease-causing, N: polymorphis, P: polymorphism automatic).

Population and literature databases, including gnomAD r2.0.2 (http://gnomad.broadinstitute.org), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar), and OMIM (https://omim.org/), were used to annotate variants. Variants were classified as "P," "LP," "VUS," "likely benign (LB)," or "Benign (B)" according to the ACMG guidelines. (Richards et al., 2015; Riggs et al., 2020)

The Shapiro–Wilk test was utilized for data distribution. Continuous variables were expressed as mean ± SD. Comparison of numeric variables was performed using Student's *t*-tests. A chi-squared test was used to compare categorical variables and rates. A *p*-value < 0.05 was considered statistically significant.

3 RESULTS

3.1 Population Demographics

A total of 2,234 couples (4,468 individuals) underwent ECS. They were divided into six subgroups as described in **Section 2.2**. As shown in **Table 1**, the average age of men was 33.13 ± 5.37 years, and the average age of women was 31.38 ± 4.72 years. Among them, 94.9% were positive carriers of at least one condition. In addition, **Supplementary Table S1** shows that patients carrying two or three variants account for 44.8% of the population, whereas only three (0.13%) men carried a maximum of 10 variants. There were no statistical differences in the carrier rate between women and men ($p = 0.88$).

3.2 Patients

Of the 254 individuals in the patient subgroup, 56 were diagnosed as having genetic diseases. The result showed that 22% of the patients were diagnosed as having genetic diseases by MES. The majority of the diagnosed patients were women (82.14%). The most common clinical characteristics were hearing impairment (including hearing loss, deafness, and surdimutism), intellectual disability (ID), eye diseases, and skeletal system diseases. The patients' information is shown in **Supplementary Table S2.**

3.3 Carrier Frequencies and Recurrent Variants

A total of 1833 A genes and 44 X-linked genes were detected in ECS, among which 1435 and 17 were recurrent, respectively (**Supplementary Table S3, S4**). The top 10 A genes in all individuals and the top 3 X-linked genes in women are shown in **Table 2**. The most common disease carried by individuals was *GJB2* (OMIM: 121,011), with a 20.17% carrier frequency, which is associated with *GJB2*-related hearing loss. The most frequent X-linked genes carried by women were *G6PD* (OMIM: 305,900), *DMD* (OMIM: 300,377), and *CACNA1F* (OMIM: 300,110). The top three recurrent variants were 109G > A (p.V37I) of *GJB2* with a frequency of 1 in 6, c.1210−11T > G of *CFTR* with a frequency of 1 in 13, and c.115+6T > C of *IL36RN* with a frequency of 1 in 35 (**Table 3**).

3.3 Secondary Findings

Different from the primary findings that were relevant to the diagnostic indication, the SFs in exome and genome sequencing were defined as P or LP variants that were not related to the patients' clinical characteristics. Such findings may not contribute to diagnostic yield but have the potential medical value for patient care. In this study, we found a total of 265 individuals who had positive SFs (**Supplementary Table S5**). The most frequent SF genes are listed in **Table 4**: *TTN* (OMIM: 188,840) accounts for 9.56% of positive cases, related to dilated cardiomyopathy. All of the listed genes are autosomal dominant.

ACMG launched a 73-gene list for reporting SFs in clinical exome and whole-genome sequencing. (Miller et al., 2021) In the SFs, 4 ACMG SF genes were detected, including 26 cases for *TTN*, 5 cases for *FLNC*, 1 case for *TMEM127* (OMIM: 613,403), and 1 case for *BRCA2* (OMIM: 600,185).

3.4 At-Risk Couples

Through ECS, 219 (9.80%) ARCs with 87 different diseases out of 2,234 couples were identified. The comprehensive characteristics of ECS are shown in **Table 5**. There were 204 couples that carried the same autosomal recessive gene, and the female partner in 1 couple carried an X-linked gene; another 15 couples carried two

TABLE 2 | Carrier frequencies of at-risk and X-linked genes.

Inheritance	Gene	Associated disease	Positive case	Carrier frequency (%)
AR	*GJB2*	*GJB2*-related hearing loss	901	20.17
	CFTR	Cystic fibrosis	455	10.18
	DUOX2	Thyroid dyshormonogenesis 6	321	7.18
	SERPINB7	Palmoplantar keratoderma, nagashima type	167	3.74
	IL36RN	Pustular psoriasis, generalized	127	2.84
	GALC	Krabbe disease	94	2.10
	CD36	Platelet glycoprotein IV deficiency	84	1.88
	HBB	ß-Thalassemia	81	1.81
	USH2A	Usher syndrome type 2A	77	1.72
	MYORG	Idiopathic basal ganglia calcification-7	74	1.66
	G6PD	Glucose 6 phosphate dehydrogenase deficiency	40	0.90
XL	*DMD*	Duchenne muscular dystrophy	10	0.22
	CACNA1F	Congenital stationary night blindness, type 2A	5	0.11

Note: AR, autosomal recessive; XL, X-linked.

TABLE 3 | Top 10 recurrent variants detected in expanded carrier screening (ECS).

Gene	Variant location	Recurrent variant	Allele count (N)	1 in _
GJB2	NM_004,004.6	c.109G > A (p.V37I)	797	6
CFTR	NM_000,492.4	c.1210-11T > G	352	13
IL36RN	NM_012,275.3	c.115+6T > C	126	35
SERPINB7	NM_003,784.4	c.796C > T (p.R266*)	123	36
DUOX2	NM_001363711.2	c.1588A > T (p.K530*)	121	37
GALC	NM_000,153.4	c.1901T > C (p.L634S)	91	49
MYORG	NM_020,702.5	c.40dupC	74	60
GJB2	NM_004,004.6	c.235delC	73	61
DUOX2	NM_001363711.2	c.2654G > T	72	62
C9	NM_001,737.5	c.346C > T	70	64

TABLE 4 | Secondary findings detected in ECS.

Gene	Inheritance	Associated disease	Positive case	Percentage (%)
				Total positive *N* = 272 (%)
Top 10 genes				
TTN	AD	Dilated cardiomyopathy-1G	26	9.56
PDE11A	AD	Primary pigmented nodular adrenocortical disease-2	10	3.68
MYH6	AD	Dilated cardiomyopathy-1EE, familial hypertrophic cardiomyopathy-14	8	2.94
LAMA4	AD	Dilated cardiomyopathy-1JJ	6	2.21
ABCA7	AD	Susceptibility to late-onset Alzheimer’s disease-9	5	1.84
COMP	AD	Multiple epiphyseal dysplasia-1	5	1.84
SLC39A5	AD	Myopia-24	5	1.84
SYNE2	AD	Emery–Dreifuss muscular dystrophy-5	5	1.84
TUBB1	AD	Macrothrombocytopenia	5	1.84
FLNC	AD	Dilated cardiomyopathy	5	1.84
ACMG recommended secondary finding genes				
TTN	AD	TTN-related myopathies	26	9.56
FLNC	AD	Dilated cardiomyopathy	5	1.84
TMEM127	AD	Hereditary paraganglioma–pheochromocytoma syndrome	1	0.37
BRCA2	AD	Hereditary breast and/or ovarian cancer	1	0.37

Note: AD, autosomal dominant.

at-risk genes. Couples in the patient subgroup who were diagnosed with genetic diseases were not considered as ARCs, unless they satisfied the ARC definition (**Section 2.2**). In addition, among the 219 ARCs, 17 carried at least 1 VUS. PGT was an option offered for these couples. At-risk genes and their related diseases were classified into four groups, 1) profound, 2) severe, 3) moderate, and 4) mild, using a disease severity classification method developed by Lazarin et al. (2014) (**Supplementary**

TABLE 5 | Reasons and circumstances for ECS.

Characteristic	Categories	*N* = 219 (pair) n (pair)	Percentage (%)
Reason for screening	Repeated miscarriage/stillborn	57	26.03
	Infertility	57	26.03
	Patients	13	5.94
	Birth defect/ultrasound anomaly	70	31.96
	Routine screening	14	6.39
	Consanguineous marriage	8	3.65
Disease classification[a]	Profound	13	5.94
	Severe	93	42.47
	Moderate	100	45.66
	Mild	13	5.94
Measures taken for at-risk gene(s)	Yes	71	32.42
	No	143	65.30
	Loss to follow-up	5	2.28
Variants classification	Both P or LP	202	92.24
	Both or one VUS	17	7.76

[a] *If couples carried more than one at-risk genes, the classification was consistent with the higher rank.*

Note: ECS, expanded carrier screening; P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance.

TABLE 6 | Disease severity classification and measures taken.

	Taken measures	No measures taken	Total	χ^2 (*p*-value)
Profound/severe	48 (45.28%)	58 (54.72%)	106	15.514 ($p < 0.001$)
Moderate/mild	23 (20.35%)	90 (79.65%)	113	
Total	71 (32.42%)	148 (67.58%)	219	

Table S6). The classifications of the ARCs in each group were as follows: profound ($n = 13$ pairs), severe ($n = 93$ pairs), moderate ($n = 100$ pairs), and mild ($n = 13$ pairs). Furthermore, we compared the ARCs' attitude towards taking measures between profound/severe and moderate/mild groups. As shown in **Table 6**, a total of 71 ARCs decided to take measures to specifically target the at-risk gene. Those ARCs found to carry profound/severe disease were more inclined to take measures to avoid having an affected child, which indicated that the reproductive decisions were significantly affected by disease severity ($p < 0.001$). The measures to prevent birth of an affected child included a plan to undergo PGT-M, gamete donation, and pregnancy termination.

The detailed characteristics of ARCs are shown in **Supplementary Table S7**. *GJB2*-related hearing loss ($n = 75$ pairs), cystic fibrosis ($n = 15$ pairs), thyroid dyshormonogenesis ($n = 12$ pairs), and thalassemia ($n = 16$ pairs) were the most common conditions found in the ARCs, which correspond with the related gene carrier frequencies identified in individuals.

When ARCs were identified, not only P or LP variants with high carrier frequencies but also some complex conditions were detected. For example, ARC 152 had a fetus with hydrocephalus and callosal dysplasia detected by prenatal ultrasound, and they chose pregnancy termination. ECS was performed on the couple before their next conception. The result showed that both partners carried VUS of the *CCDC88C* gene that is related to congenital hydrocephalus. The phenotype caused by the *CDDC88C* variants includes fetal enlarged ventricles due to the accumulation of cerebrospinal fluid, and neurological impairment in live births. Considering the symptoms observed *via* prenatal ultrasound and the disease phenotype, the couple was classified as ARC, and the ECS result was reported to them. They had a healthy live birth after undergoing PGT-M. In parallel, ARC 218 underwent IVF and gave birth to an SMA-affected child previously. ECS was used as a routine test before they underwent ART again. The result revealed that they both carried not only a P variant of *SMN1* but also an LP variant of *STRC*. *STRC* is associated with autosomal recessive deafness-16, which did not show symptoms in their firstborn. They decided to undergo PGT-A and PGT-M that targeted both *SMN1* and *STRC*. ARC 219 also received ECS as a routine test before undergoing ART because of infertility. The female partner suffered from aniridia and bilateral horizontal nystagmus. The ECS results showed that they were classified as ARC for *GJB2*, whereas the female partner had a P variant of autosomal dominant condition *PAX6* related to aniridia. They chose to undergo PGT-M that targeted *PAX6* instead of IVF alone. ARC 204 was previously diagnosed with in-born 21-hydroxylase deficiency, and MLPA confirmed the presence of the *CYP21A2* variant. The couple decided to pursue PGT-M to have a healthy child. Therefore, although they were found to be carriers of *UVSSA* related to a mild condition UV-sensitive syndrome, the PGT-M was targeted against *CYP21A2*.

4 DISCUSSION

To search for genes suitable for ECS in the Chinese population, this study retrospectively analyzed the results of MES in 2,234 couples. The results showed that 94.9% of individuals were carriers of at least one P or LP variant related to AR or X-linked genes and revealed the highly prevalent genes and their hotspot variants. Disease severity had a significant impact on reproductive decisions made by ARCs.

The carriers of at least one P or LP variant related to AR or X-linked genes rate was 94.4%, similar to the results reported in 78% postive for being a carrier individuals after screening 728 gene disorder pairs in the United States, (Punj et al., 2018) implying the possibility that most individuals were carriers of disease-causing variants. Expanded carrier screening has been provided in many countries for couples at their preconception or early pregnancy. Previous carrier screening for certain diseases in the risk population has reduced the incidence of genetic disorders such as Tay–Sachs disease in the ASJ Ashkenazi Jewish population. Unlike Europe or the United States, ECS in China was only proposed recently and currently lack guidelines for panel design for the Chinese population and post-test counseling. There were previous studies that attempted to perform ECS in China. Zhao et al. (2019) screened 11 genes related to 12 Mendelian conditions in a diverse number of couples. In parallel, Shi et al. (2021) screened 11 recessive conditions in women at early pregnancy in Hong Kong. These two studies drew similar conclusions that α-thalassemia and ß-thalassemia had a high prevalence in the Chinese population. Both of these studies established the significance of ECS in reducing the risk of genetic diseases in newborns. However, these two studies selected panels with limited conditions, which may miss potential ARCs. Xi et al. (2020) used 201 gene panels for couples seeking ART. They reported the CCR (45.91%) of the 187 autosomal recessive genes, which was lower than that of over 60% of 100 gene panels in the ASJ population. (Guo and Gregg, 2019) The CCR results suggested that their panel may not cover the prevalent genetic diseases in China. Thus, genes suitable for ECS in the Chinese population needed further investigation and selection.

In this study, the most common gene in ECS was *GJB2* associated with *GJB2*-related hearing loss. The hotspot variants included c.109G > A (p.V37I) and c.235delC (p.L79Cfs*3). In mainland China, the incidence of hearing loss in newborns was reported to be between 1% and 3.47%, and genetic factors accounted for 50–60% of the patients. We found the carrier frequency of *GJB2* to be 20.17%, which was much higher than previously reported (15%) in all hearing loss-related genes in the Chinese population. (Kim et al., 2013) Moreover, congenital hypothyroidism with an incidence of 0.04% was considered to be the most common neonatal endocrine disease in Chinese newborns. (Deng et al., 2018) *DUOX2* accounted for 44% of the congenital hypothyroidism patients in newborn screening. (Huang et al., 2021) In this study, the results showed that the carrier frequency of *DUOX2* was 7.18%. In addition, *CD36* associated with platelet glycoprotein IV deficiency and *SERPINB7* associated with Nagashima-type palmoplantar keratoderma have been unreported in China, and their prevalence is unclear. These two genes were classified as moderate and mild phenotypes, respectively, and were not included in most ECS panels. Furthermore, *GALC* associated with Krabbe disease causes ID and early death. (Shin et al., 2016) The incidence varies in ethnic groups ranging from 1 in 100 live births in the Druze population (Rafi et al., 1996) to 2 in 1,000,0000 live births in Japan. (Foss et al., 2013) The prevalence in China remains unclear. This study reported that *GALC* had a rather high carrier frequency of 2.10% in China, along with the hotspot mutation c.1901T > C (p.L634S), which accounted for 96.8% (91/94) of the detected variants. This indicated that although *GALC* was not recommended in the gene set according to the ACMG, it should be included in carrier screening panels for the Chinese population. In addition, it has been reported that *CYP21A2* and *FMR1* have a high prevalence in some populations. (Hernandez-Nieto et al., 2020) Yet, such highly homologous genes and GC-rich regions may easily lead to false-positive or false-negative results in NGS. Screening for these genes should be compensated for by other molecular methods. (Li et al., 2015)

The awareness of SFs is important for disease prevention and early intervention. This study also analyzed SFs uncovered in ECS. *TTN* had a high carrier frequency (8.24%) in the Chinese population, which is associated with dilated cardiomyopathy. *TTN* truncating variants were responsible for about 25% of familial dilated cardiomyopathy and in 18% of sporadic cases. *TTN*-related dilated cardiomyopathy has a high penetrance after 40 years of age (>95%). (Kellermayer et al., 2019) There were no *TTN*-related patients identified possibly because the individuals were in their fertile age, were in their 30s, and did not meet the onset age for dilated cardiomyopathy. Screening for *TTN* allows for the prevention or early diagnosis and therapy of *TTN*-related dilated cardiomyopathy. The SFs were mostly associated with late onset diseases (*MYH6*, *FLNC*, *LAMA4*, *ABCA7*, and *SYNE2*) or clinical heterogeneous phenotypes (*TUBB1*, *SLC39A5*, and *COMP*). The deleterious variants were sporadic without hotspot regions.

The ARC rate in this study was 9.80%. Analysis of the usage of PGT in ARCs showed that the severity of the disease had a significant impact on reproductive decision-making. Couples that carried diseases classified as profound or severe were inclined to take measures compared with couples that carried moderate or mild diseases. The severity classification of genetic diseases was discussed by Lazarin et al. (2014), and they focused on the characteristics of the diseases. Consistent with the carrier frequencies in individuals, 75/219 ARCs were *GJB2* carriers, and 15/75 ARCs were *CFTR* carriers. Despite the high carrier frequency of *GJB2*, only 8 in 75 ARCs underwent PGT-M for the targeted *GJB2* variant. The possible reasons were that despite c.109G > A (p.V37I) being P, this variant is a hypomorph and mostly causes mild to moderate hearing loss with late onset and low penetrance (17%). (Kim et al., 2013) A total of 20 individuals with homozygous c.109G > A (p.V37I) variant were detected in this study, but only 6 of them experienced hearing loss, which increased the penetrance to 30% (**Supplementary Table S8**). However, a consensus interpretation of the p. V37I variant of *GJB2* was launched by the ClinGen Hearing Loss Expert Panel in 2019; they reviewed case–control studies and functional, computational, allelic, and segregation data regarding the variant. The panel concluded that the p. V37I variant of *GJB2* is pathogenic for autosomal recessive nonsyndromic hearing loss with variable expressivity and incomplete penetrance. (Shen et al., 2019) A recent study also found that homozygous p. V37I *GJB2* is associated with progressive hearing loss in adults, especially over 60 years old. (Chen et al., 2022) Therefore, knowing the ECS results allowed the couples to consider neonatal hearing screening and avoid the hearing loss-inducing factors in adulthood. (Chai et al., 2015) In parallel, variant c.1210-11T > G, also known as IVS8-5T allele, of *CFTR* is the most common allele worldwide. (Chillón et al., 1995) Many genotype–phenotype studies demonstrated that the IVS8-5T allele is pathogenic for congenital bilateral absence of the vas deferens with incomplete penetrance and is a genetic modifier for CF. (Bombieri et al., 2011) In this study, the IVS8-5T allele was mostly found in men with infertility issues. After explaining the disease risk in the genetic counseling, none of the 15 ARCs underwent PGT-M that targeted *CFTR*. Instead of identifying ARCs after giving birth to an affected offspring, ECS helps determine ARCs during preconception and enables the couples to make alternative reproductive decisions. (Ghiossi et al., 2018) PGT allows for the detection of embryos with genetic disorders before implantation to prevent delivering infants with a birth defect. It is also designed to detect monogenic disorders. The utility of PGT-M relies on the confirmation of P genes, which can be achieved through ECS. (Kuliev and Rechitsky, 2017) In addition, for those ARCs who do not take part in PGT-M as an option, the ECS results still provide them with clues to undergo prenatal diagnosis or early intervention for affected newborns. On the other hand, in clinical practice, the ECS results may be challenging to interpret. For example, limited by the current knowledge of genes and variants, detection of a large number of variants in ECS that are classified as VUS provides uncertain information, and couples may be confused in terms of what steps to follow after receiving this information. For couples in whom one partner carried a P or LP variant while the other carried a VUS, it was a great challenge to decide whether to report a positive result to the patients. The clinical validity of the ECS panel relied on the ARC detection rate. Although the ARC rate of 9.80% was considerably higher than those in previous studies, using such a large-scale panel may pose an excessive information overload to both patients and physicians.

The ideal situation for ECS is to identify carriers to provide information regarding the possibility of having an affected child before conception or during early pregnancy. Taken together with the results in this study, we make the following suggestions: 1) Couples who intend to receive ART should undergo ECS regardless of family history and ethnicity. 2) Current ECS panel design is mostly based on carrier frequency and severity. However, due to the complexity of genotype–phenotype correlations, some highly prevalent variants only cause mild phenotypes. The inclusion of such variants should be evaluated to avoid excessive burden on patients and counseling. 3) Adult-onset conditions with high prevalence and severe phenotypes should be considered and included in the carrier screening panel to prevent birth defects and aid them in early management.

In summary, this study performed clinical exome sequencing on couples seeking ART in China. The results revealed that 94.9% of the individuals were carriers of at least one deleterious variant. We preliminary demonstrated a set of highly prevalent genes along with their hotspot variants, providing insight to further establish a suitable ECS gene panel for the Chinese population. The total ARC rate was 9.80% in this study. PGT was offered to the ARCs. Furthermore, the severity of related diseases had a tremendous influence on fertility planning. The study emphasized the significance of couples receiving ECS before undergoing ART.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Chongqing Health Center for Women and Children and ethics committee of the Reproductive and Genetic Hospital of CITIC-Xiangya. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

DL and JD created the research design and Funding acquisition. KT and WH performed formal analysis, manuscript formulation and writing. XL, LH, HH, GL, GL, and CD performed data collection, data curation and data analysis. WZ reviewed and edited the manuscript.

FUNDING

This study is funded by the Research Grant of CITIC-Xiangya (YNXM-202002); China Postdoctoral Science Foundation (2022M711119); the General Project of Chongqing Natural Science foundation of China(cstc2021jcyj-msxmX0877).

ACKNOWLEDGMENTS

The authors thank the patients who volunteered to participate in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.943058/full#supplementary-material

REFERENCES

Antonarakis, S. E. (2019). Carrier Screening for Recessive Disorders. *Nat. Rev. Genet.* 20 (9), 549–561. doi:10.1038/s41576-019-0134-2

Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., et al. (2011). Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing. *Sci. Transl. Med.* 3 (65), 65ra4. doi:10.1126/scitranslmed.3001756

Bombieri, C., Claustres, M., De Boeck, K., Derichs, N., Dodge, J., Girodon, E., et al. (2011). Recommendations for the Classification of Diseases as CFTR-Related Disorders. *J. Cyst. Fibros.* 10 (Suppl. 2), S86–S102. doi:10.1016/s1569-1993(11)60014-3

Chai, Y., Chen, D., Sun, L., Li, L., Chen, Y., Pang, X., et al. (2015). The Homozygous p.V37I Variant ofGJB2is Associated with Diverse Hearing Phenotypes. *Clin. Genet.* 87 (4), 350–355. doi:10.1111/cge.12387

Chen, Y., Wang, Z., Jiang, Y., Lin, Y., Wang, X., Wang, Z., et al. (2022). Biallelic p.V37I Variant in GJB2 Is Associated with Increasing Incidence of Hearing Loss with Age. *Genet. Med.* 24 (4), 915–923. doi:10.1016/j.gim.2021.12.007

Chillón, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., et al. (1995). Mutations in the Cystic Fibrosis Gene in Patients with Congenital Absence of the Vas Deferens. *N. Engl. J. Med.* 332 (22), 1475–1480. doi:10.1056/nejm199506013322204

Deng, K., He, C., Zhu, J., Liang, J., Li, X., Xie, X., et al. (2018). Incidence of Congenital Hypothyroidism in China: Data from the National Newborn Screening Program, 2013-2015. *J. Pediatr. Endocrinol. Metab.* 31 (6), 601–608. doi:10.1515/jpem-2017-0361

Edwards, J. G., Feldman, G., Goldberg, J., Gregg, A. R., Norton, M. E., Rose, N. C., et al. (2015). Expanded Carrier Screening in Reproductive Medicine-Points to Consider. *Obstet. Gynecol.* 125 (3), 653–662. doi:10.1097/aog.0000000000000666

Foss, A. H., Duffner, P. K., and Carter, R. L. (2013). Lifetime Risk Estimators in Epidemiological Studies of Krabbe Disease. *Rare Dis.* 1, e25212. doi:10.4161/rdis.25212

Franasiak, J. M., Olcha, M., Bergh, P. A., Hong, K. H., Werner, M. D., Forman, E. J., et al. (2016). Expanded Carrier Screening in an Infertile Population: How Often Is Clinical Decision Making Affected? *Genet. Med.* 18 (11), 1097–1101. doi:10.1038/gim.2016.8

Ghiossi, C. E., Goldberg, J. D., Haque, I. S., Lazarin, G. A., and Wong, K. K. (2018). Clinical Utility of Expanded Carrier Screening: Reproductive Behaviors of At-Risk Couples. *J. Genet. Couns.* 27 (3), 616–625. doi:10.1007/s10897-017-0160-1

Gregg, A. R., Aarabi, M., Klugman, S., Leach, N. T., Bashford, M. T., Goldwaser, T., et al. (2021). Screening for Autosomal Recessive and X-Linked Conditions during Pregnancy and Preconception: a Practice Resource of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 23 (10), 1793–1806. doi:10.1038/s41436-021-01203-z

Guo, M. H., and Gregg, A. R. (2019). Estimating Yields of Prenatal Carrier Screening and Implications for Design of Expanded Carrier Screening Panels. *Genet. Med.* 21 (9), 1940–1947. doi:10.1038/s41436-019-0472-7

Haque, I. S., Lazarin, G. A., Kang, H. P., Evans, E. A., Goldberg, J. D., and Wapner, R. J. (2016). Modeled Fetal Risk of Genetic Diseases Identified by Expanded Carrier Screening. *JAMA* 316 (7), 734–742. doi:10.1001/jama.2016.11139

Hernandez-Nieto, C., Alkon-Meadows, T., Lee, J., Cacchione, T., Iyune-Cojab, E., Garza-Galvan, M., et al. (2020). Expanded Carrier Screening for Preconception Reproductive Risk Assessment: Prevalence of Carrier Status in a Mexican Population. *Prenat. Diagn* 40 (5), 635–643. doi:10.1002/pd.5656

Huang, M., Lu, X., Dong, G., Li, J., Chen, C., Yu, Q., et al. (2021). Analysis of Mutation Spectra of 28 Pathogenic Genes Associated With Congenital Hypothyroidism in the Chinese Han Population. *Front. Endocrinol.* 12, 695426. doi:10.3389/fendo.2021.695426

Kaback, M. M. (2000). Population-based Genetic Screening for Reproductive Counseling: the Tay-Sachs Disease Model. *Eur. J. Pediatr.* 159, S192–S195. doi:10.1007/pl00014401

Kellermayer, D., Smith, J. E., 3rd, and Granzier, H. (2019). Titin Mutations and Muscle Disease. *Pflugers Arch. - Eur. J. Physiol.* 471 (5), 673–682. doi:10.1007/s00424-019-02272-5

Kim, S. Y., Park, G., Han, K.-H., Kim, A., Koo, J.-W., Chang, S. O., et al. (2013). Prevalence of p.V37I Variant of GJB2 in Mild or Moderate Hearing Loss in a Pediatric Population and the Interpretation of its Pathogenicity. *PLoS One* 8 (4), e61592. doi:10.1371/journal.pone.0061592

Kuliev, A., and Rechitsky, S. (2017). Preimplantation Genetic Testing: Current Challenges and Future Prospects. *Expert Rev. Mol. Diagnostics* 17 (12), 1071–1088. doi:10.1080/14737159.2017.1394186

Lazarin, G. A., Hawthorne, F., Collins, N. S., Platt, E. A., Evans, E. A., and Haque, I. S. (2014). Systematic Classification of Disease Severity for Evaluation of Expanded Carrier Screening Panels. *PLoS One* 9 (12), e114391. doi:10.1371/journal.pone.0114391

Li, J., Dai, H., Feng, Y., Tang, J., Chen, S., Tian, X., et al. (2015). A Comprehensive Strategy for Accurate Mutation Detection of the Highly Homologous PMS2. *J. Mol. Diagnostics* 17 (5), 545–553. doi:10.1016/j.jmoldx.2015.04.001

Martin, J., Asan, fnm, Yi, Y., Alberola, T., Rodríguez-Iglesias, B., Jiménez-Almazán, J., et al. (2015). Comprehensive Carrier Genetic Test Using Next-Generation Deoxyribonucleic Acid Sequencing in Infertile Couples Wishing to Conceive through Assisted Reproductive Technology. *Fertil. Steril.* 104 (5), 1286–1293. doi:10.1016/j.fertnstert.2015.07.1166

Miller, D. T., Lee, K., Chung, W. K., Gordon, A. S., Herman, G. E., Klein, T. E., et al. (2021). ACMG SF v3.0 List for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing: a Policy Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 23 (8), 1381–1390. doi:10.1038/s41436-021-01172-3

NHC (2018). *National Comprehensive Prevention and Control Programme for Birth Defects*. Beijing: National Health Commission of the People's Republic of China. Available from http://www.nhc.gov.cn/fys/s3589/201809/9644ce7d265342779099d54b6962a4e0.shtml.

Opinion No, Committee (2017). Committee Opinion No. 690: Carrier Screening in the Age of Genomic Medicine. *Obstet. Gynecol.* 129 (3), e35–e40. doi:10.1097/AOG.0000000000001951

Punj, S., Akkari, Y., Huang, J., Yang, F., Creason, A., Pak, C., et al. (2018). Preconception Carrier Screening by Genome Sequencing: Results from the Clinical Laboratory. *Am. J. Hum. Genet.* 102 (6), 1078–1089. doi:10.1016/j.ajhg.2018.04.004

Rafi, M. A., Luzi, P., Zlotogora, J., and Wenger, D. A. (1996). Two Different Mutations Are Responsible for Krabbe Disease in the Druze and Moslem Arab Populations in Israel. *Hum. Genet.* 97 (3), 304–308. doi:10.1007/bf02185759

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). 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.* 17 (5), 405–424. doi:10.1038/gim.2015.30

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., et al. (2020). Technical Standards for the Interpretation and Reporting of Constitutional Copy-Number Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet. Med.* 22 (2), 245–257. doi:10.1038/s41436-019-0686-8

Shen, J., Oza, A. M., del Castillo, I., Duzkale, H., Matsunaga, T., Pandya, A., et al. (2019). Consensus Interpretation of the p.Met34Thr and p.Val37Ile Variants in GJB2 by the ClinGen Hearing Loss Expert Panel. *Genet. Med.* 21 (11), 2442–2452. doi:10.1038/s41436-019-0535-9

Shi, M., Liauw, A. L., Tong, S., Zheng, Y., Leung, T. Y., Chong, S. C., et al. (2021). Clinical Implementation of Expanded Carrier Screening in Pregnant Women at Early Gestational Weeks: A Chinese Cohort Study. *Genes (Basel)* 12 (4), 496. doi:10.3390/genes12040496

Shin, D., Feltri, M. L., and Wrabetz, L. (2016). Altered Trafficking and Processing of GALC Mutants Correlates with Globoid Cell Leukodystrophy Severity. *J. Neurosci.* 36 (6), 1858–1870. doi:10.1523/jneurosci.3095-15.2016

Tucker, T., Marra, M., and Friedman, J. M. (2009). Massively Parallel Sequencing: the Next Big Thing in Genetic Medicine. *Am. J. Hum. Genet.* 85 (2), 142–154. doi:10.1016/j.ajhg.2009.06.022

Xi, Y., Chen, G., Lei, C., Wu, J., Zhang, S., Xiao, M., et al. (2020). Expanded Carrier Screening in Chinese Patients Seeking the Help of Assisted Reproductive Technology. *Mol. Genet. Genomic Med.* 8 (9), e1340. doi:10.1002/mgg3.1340

Zhao, S., Xiang, J., Fan, C., Asan, fnm, Shang, X., Zhang, X., et al. (2019). Pilot Study of Expanded Carrier Screening for 11 Recessive Diseases in China: Results from 10,476 Ethnically Diverse Couples. *Eur. J. Hum. Genet.* 27 (2), 254–262. doi:10.1038/s41431-018-0253-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership