Yan Zhao

KU LEUVEN

UNIVERSITY OF TARTU · UNIVERSITAS TARTUENSIS · 1632

Unravelling the mechanisms of chromosomal instability during early embryogenesis

Development of single cell genomic methods to explore the spectrum, origin, and developmental compatibility of chromosomal abnormalities

Yan Zhao

Supervisory Committee:
Prof. Dr. Joris Vermeesch
Prof. Dr. Ants Kurg
Prof. Dr. Thierry Voet

Dissertation presented in partial fulfillment of the requirements for the double degree of Doctor of Biomedical Sciences (KU Leuven) and Doctor of Gene Technology (University of Tartu)

March 2025

UNRAVELLING THE MECHANISMS OF CHROMOSOMAL INSTABILITY DURING EARLY EMBRYOGENESIS

March 2025

KU Leuven
Biomedical Sciences Group
Faculty of Medicine
Department of Human Genetics

DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

University of Tartu
Faculty of Science and Technology
Institute of Molecular and Cell Biology

UNRAVELLING THE MECHANISMS OF CHROMOSOMAL INSTABILITY DURING EARLY EMBRYOGENESIS

Development of single cell genomic methods to explore the spectrum, origin, and developmental compatibility of chromosomal abnormalities

YAN ZHAO

Jury:

Supervisor KU Leuven: Prof. Dr. Joris Vermeesch
Co-supervisors: Prof. Dr. Ants Kurg University of Tartu
Prof. Dr. Thierry Voet KU Leuven
Chair examining committee: Prof. Dr. Peter Claes
Chair public defence: Prof. Dr. Patrick Callaerts
Jury members: Prof. Dr. Alejandro Sifrim
Prof. Dr. Sophie Debrock
Prof. Dr. Hilde Van de Velde
Prof. Dr. Christian Ottolini
Prof. Dr. Osamu Shimmi
Prof. Dr. Tambet Tõnissoo

Dissertation presented in partial fulfilment of the requirements for the double degree of Doctor of Biomedical Sciences (KU Leuven) and Doctor of Gene Technology (University of Tartu)

March 2025

ACKNOWLEDGEMENTS

As I reach the end of my doctoral training, I reflect on my journey that began with my application for a PhD position from China in 2020. I want to extend my heartfelt thanks to Prof. Joris Vermeesch for selecting me from the CV pool, leading to my acceptance as a PhD student focusing on embryo studies as part of the MATER program. Joris, you've been an exceptional mentor. I admire not only your conscientious attitude toward work but also your optimism toward science and life. Your positive demeanor keeps everyone in good spirits, and your dedication—often working weekends and late nights on students' projects—made me feel that my work was taken seriously. I truly appreciate your supervision style. You provided valuable guidance and suggestions while allowing me the freedom to pursue my own path. I am immensely grateful to have had such a supportive supervisor guiding me through my PhD journey. I also want to thank Prof. Ants Kurg, my co-supervisor at the University of Tartu. Your kindness during my time in Tartu was invaluable. You supported my scientific progress and helped me navigate daily life, making my 10 months there a cherished memory. I appreciated your invitations to department and university events, as well as your practical advice on surviving the freezing winter—such a warm gesture for a newcomer. Your prompt and thorough responses to my emails made it easy for me to complete my academic requirements smoothly and stay on track. Thank you, Prof. Thierry Voet, my co-supervisor at KU Leuven, for closely overseeing my progress throughout my PhD journey. Your guidance was crucial in helping me maintain attention to detail and consistently move forward.

Thank you to everyone in Joris's lab for your support and collaboration. Tine, Margot, Olga, and Tanja, you have been wonderful coworkers. Your willingness to help and our teamwork ensured the smooth completion of our tasks. Olga, your valuable insights on embryo studies have been incredibly helpful, and I've learned so much from your expertise. Tanja, I truly appreciate your steady support with everything, from schematic figures to experiments. I also enjoyed our dinner at the Chinese restaurant—it was a pleasure to find out that you all love Chinese food! A special thanks to Greet for your help with the experiments; your professionalism shines through in the clean, high-quality data you consistently provide. To Erika and Benjamin, thank you for your insightful suggestions on data analysis and bioinformatics. Benjamin, your humor and frankness have always been appreciated. Kate, I feel fortunate that we are both part of the MATER program. Traveling together to various meetings and courses has been a joy. I've cherished our conversations, where we've shared our worries, happiness, and thoughts. You're a person with a cool character and a talented scientist. Despite our differences, we always enjoy our time together. Stefania and Dhanya,

it's been a pleasure working with you—whether we're chatting about life's little things, solving scientific problems, or just venting about challenges. Stefania, thank you for inviting me to your wedding—it was an unforgettable experience. And Dhanya, I appreciate your help in picking out the perfect Italian-style dress for the occasion. Thank you, Mathilde, for your invaluable assistance with my project and for frequently checking in to see how I'm doing. Angelica, I truly enjoy our regular conversations where we share our life experiences and project status. Tom, I genuinely value your support and encouragement, which is always frank and sincere. Thank you to Qiang, Marta, and Anja as well for always being around. We're a diverse group of individuals from different backgrounds, and I've really enjoyed our time together. It's left me with many great memories. Lastly, to Huiwen and Jia, thank you for being such good friends. I've loved our outings—whether it's for concerts, shopping, dinners, or simply chatting about life. Your support and suggestions have truly made me feel at home.

Many thanks to the group in Tartu. Tonis and Merle, thank you for including me in teaching activities and inviting me to events at the institute. Your warmth helped me get to know the local staff better and made my experience in Tartu much more enjoyable. Baris, it's great to be in the same MATER program with you. I appreciate how we can always share information and support one another; your kindness and valuable suggestions mean a lot.

I would like to express my appreciation to the EU MATER program. Through this program, I've not only received valuable training as a PhD researcher but also taken courses on a variety of topics. I truly enjoyed working and connecting with PhD students and PIs from different universities and diverse backgrounds. This experience has broadened my perspective and allowed me to form meaningful friendships with my fellow MATER students.

Finally, I want to thank my husband, Shuang, for always being by my side. You are not only my partner but also my best friend, and I feel fortunate to share both our happiness and sadness together.

The end of my PhD journey marks the conclusion of a significant chapter in my life. I will always cherish the memories of my four years in Leuven and the city's quiet, academic atmosphere. Wishing us all the best as we move forward and may tomorrow bring new opportunities and joy.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADO	Allele dropout
AF	Allele frequency
AI	Artificial intelligence
BAF	B-allele frequency
CIN	Chromosomal instability
CNV	Copy number variation
DNM	*De novo* mutation
DP	Depth
dpf	Days post fertilization
EGA	Embryonic genome activation
FISH	Fluorescent in situ hybridization
G&T-seq	Genome and transcriptome sequencing
GBS	Genotyping-by-sequencing
GIAB	Genome in a Bottle
GRN	Gene regulatory network
GW	Genome-wide
GWAS	Genome-wide association studies
HiFi	High-fidelity
hG&T-seq	Haplotype-based G&T-seq
ICM	Inner cell mass
ICSI	Intracytoplasmic sperm injection
IVEP	*In vitro* embryo production
IVF	*In vitro* fertilization
IVM	*In vitro* maturation
lrWGS	Long-read whole-genome sequencing
MAR	Medically assisted reproduction
MDA	Multiple displacement amplification
NGS	Next-generation sequencing
OMIA	Online Mendelian Inheritance in Animals
ONT	Oxford Nanopore Technologies
PacBio	Pacific Bioscience
PCR	Polymerase chain reaction

PGT	Preimplantation genetic testing
PGT-A	Preimplantation genetic testing for aneuploidy
PGT-M	Preimplantation genetic testing for monogenic disorders
PGT-P	Preimplantation genetic testing for polygenic disorders
PGT-SR	Preimplantation genetic testing for structural rearrangements
PN	Pronuclei
PRS	Polygenic risk score
PTA	Primary template-directed amplification
qPCR	Quantitative polymerase chain reaction
RCT	Randomized controlled trial
SAC	Spindle assembly checkpoint
scRNA-seq	Single-cell RNA sequencing
SMRT	Single molecule real-time
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SR	Structural chromosomal rearrangement
T2T	Telomere-to-Telomere
TE	Trophectoderm
TF	Transcription factor
UAF	Unique allele fraction
UMAP	Unified Manifold Approximation and Projection
WG	Whole-genome

SUMMARY OF WORK

Chromosomal instability (CIN) is a hallmark of preimplantation embryos for mammals such as human and bovine. These chromosomal abnormalities are known to be selected against during development and cause embryo loss, spontaneous abortion, or lead to congenital abnormalities if compatible with live birth. **In the first project**, we explored whether CIN observed in human preimplantation embryos also occurs in equine preimplantation embryos. We adapted the haplarithmisis algorithm to be compatible for equine analysis and uncovered frequent chromosomal abnormalities in equine preimplantation embryos, with higher chromosomal errors in arrested cleavage stage embryos compared to blastocysts, similar to findings in human and bovine. The method can be used to select equine embryos devoid of genetic errors and pathogenic variants, while carrying variants of interest. In addition to aneuploidies, whole-genome (WG) aberrations are also detected in preimplantation embryos. For example, triploidy involves the presence of an additional haploid set of chromosomes from one parent. Notably, we recently observed androgenetic, gynogenetic, and polyploid blastomeres coexisting within the same embryo. **In the second project**, we explored the origin and developmental potential of these chimeric and mixoploid embryos. We confirmed that parental genomes segregate into distinct blastomeres during the first zygotic division, resulting in the co-occurrence of WG abnormal cells with normal diploid cells or other WG abnormal cells, a process we previously discovered and coined “heterogoneic division”. Stress responses in gene expression contribute to developmental impairment in WG abnormal cells, and their differing developmental fates can explain the formation of androgenotes, gynogenotes, triploidy, chimerism and mixoploidy observed later during development. We therefore recommend haplotype-based preimplantation genetic testing (PGT) for WG abnormalities to enhance baby-take home rates during IVF programs. With the rapid development of long-read sequencing technology, we explored the performance of long-read whole-genome sequencing (lrWGS)-based comprehensive PGT **in the third project.** A benchmark study using the Genome in a Bottle (GIAB) Ashkenazi trio demonstrated the high performance of lrWGS data from single cells for variant calling and phasing. Testing lrWGS-based PGT on human embryos showed 100% consistency with array-based comprehensive PGT for single nucleotide variations (SNVs), indels, and aneuploidies, highlighting lrWGS-based PGT as a promising alternative to current methods. The three projects within this PhD program introduced innovative approaches for embryo study and selection, significantly advancing our understanding of the origins and developmental impacts of chromosomal abnormalities in preimplantation mammalian embryos.

KOKKUVÕTE

KROMOSOMAALSE EBASTABIILSUSE TEKKEMEHHANISMID VARASES EMBRÜOGENEESIS

Kromosomaalne ebastabiilsus (ingl. *Chromosomal INstability*, CIN) on imetajate, näiteks inimeste ja veiste implantatsioonieelsete embrüote küllaltki sage tunnus. Enamasti eemaldatakse kromosoomianomaaliatega embrüod juba varase arengu käigus iseeneslike abortide kaudu, kuid elussünnini jõudmise korral võib kromosomaalne ebastabiilsus viia ka kaasasündinud väärarendite tekkele. Esimeses projektis uurisime, kas inimese implantatsioonieelsetes embrüodes täheldatud CIN esineb ka hobuste implantatsioonieelsetes embrüodes. Selleks kohandasime haplaritmisis (ingl. *haplarithmisis*) algoritmi nii, et see ühilduks hobuste analüüsiga ja avastasime, et sagedased kromosomaalsed kõrvalekalded esinevad ka hobuste implantatsioonieelsetes embrüodes. Hobuste lõigustuvates embrüodes avastatud kromosomaalsed vead olid sarnased nii inimeste kui veiste leidudele võrrelduna kromosomaalselt normaalsete blastotsüstidega ehk lootepõiekestega. Antud meetodit saaks edukalt kasutada tõuaretuses - huvipakkuvate geneetiliste variantidega kuid ilma geneetiliste vigade ja patogeensete variantideta hobuste embrüote valimiseks. Lisaks aneuploidiatele tuvastati implantatsioonieelsetes embrüodes ka kogu genoomi aberratsioone. Näiteks triploidiaid (kolmekordne kromosoomide arv), ühe vanema täiendava haploidse kromosoomi-komplekti olemasolul. Samuti avastasime samas embrüos koos eksisteerivaid androgeneetilisi, günogeneetilisi ja polüploidseid blastomeere ehk lõigustusrakke. Teises projektis uurisime kimäärsete ja miksoploidsete embrüote päritolu ja arengupotentsiaali. Avastasime, et vanemate genoomid eralduvad viljastatud munaraku ehk sügoodi esimese jagunemise ajal eraldiseisvateks blastomeerideks, mille tulemuseks on ebanormaalse genoomiga rakkude samaaegne esinemine koos normaalsete diploidsete (kahekordne kromosoomide arv) rakkude või muude ebanormaalsete rakkudega. Selle protsessi kirjeldamiseks lõime termini "heterogoonne jagunemine". Geenieksрessiooni muutvad stressireaktsioonid soodustavad ülegenoomsete muutuste tagajärjel tekkinud arenguhäireid ebanormaalsetes rakkudes ja nende erinev arengusaatus võib seletada androgenootide, günogenootide, triploidsuse, kimäärsuse ja miksoploidsuse teket, mida täheldati hilisema arengu käigus. Seetõttu soovitame haplotüüpidel põhinevat preimplantatsioonilist geneetilist testimist (ingl. *Preimplantation Genetic Testing*, PGT) ülegenoomsete kõrvalekallete tuvastamiseks, et suurendada munaraku kehavälise viljastamise (IVF) programmide edu. Kolmandas projektis uurisime kolmanda põlvkonna sekveneerimistehnoloogiate kiire arenguga seotud pikkadel lugemitel põhineva kogu genoomi järjestamise (lrWGS) PGT tehnoloogiat. Võrdlusuuring, milles kasutati nn. *Genome in a Bottle* (GIAB) Aškenazi triot, näitas üksikute rakkude lrWGS-i

andmete sobivust geenivariantide eristamiseks ja faasimiseks. lrWGS-põhise PGT meetodi kasutamine inimese embrüotel näitas 100% vastavust DNA mikrokiibi põhise PGT-ga ühe nukleotiidi variatsioonide (SNV), indelite ja aneuploidiate tuvastamiseks, tõestades et lrWGS-põhine PGT on sobiv alternatiiv praegustele meetoditele.

Antud doktoritöös käsitletud neli projekti tutvustasid uuenduslikke lähenemisviise embrüote uurimiseks ja valikuks, aidates oluliselt kaasa meie arusaamisele kromosomaalsete kõrvalekallete päritolu ja mõjude kohta imetajate implantatsioonieelsete embrüote arengus.

CHAPTER 1 – INTRODUCTION

Mammalian embryo development

Mammalian embryo development begins with fertilization, a critical process for sexual reproduction and the continuation of life. During fertilization, a sperm cell fuses with a metaphase II oocyte, triggering the completion of maternal meiosis II. The haploid maternal and paternal chromosomes in the zygote then replicate, followed by a series of cleavage divisions. In humans, by the fourth day of development, the embryo compacts into a morula and then begins its first distinct cellular differentiations. Within the first 5 to 6 days post fertilization (dpf), the zygote develops into a blastocyst, consisting of an outer layer of trophectoderm (TE) surrounding the blastocyst cavity, where the inner cell mass (ICM) is asymmetrically positioned on one side. This blastocyst then implants into the uterine endometrium and gradually develops into a full-term baby. Early embryo development involves substantial transcriptomic changes as control shifts from the maternal to the zygotic genome. Furthermore, both parental genomes are essential for successful embryogenesis.

Maternal to zygotic genome transition

Mammalian preimplantation development is a complex process involving significant changes in transcriptional architecture. Early embryonic development is controlled by the maternal genome, with maternal mRNAs and proteins driving essential functions. As development progresses, control shifts from maternal gene products to those produced by the zygotic genome. During this process, maternal transcripts are degraded, and the transcription of the embryonic genome, known as embryonic genome activation (EGA), begins (Tadros and Lipshitz, 2009). Both EGA-dependent and -independent processes are involved in maternal mRNA clearance. Defects in this mRNA decay process have been associated with early developmental arrest in *in vitro* fertilized human embryos (Sha *et al.*, 2020). EGA occurs in multiple waves, with the timing of major EGA varying by species: in mouse embryos, major EGA occurs at the two-cell stage (Sirard, 2012); in human embryos, it occurs between the four- to eight-cell stages (Braude, Bolton and Moore, 1988; Leng *et al.*, 2019); and in bovine embryos, it happens between the eight- to 16-cell stages (Sirard, 2012). EGA is also coordinated with other embryonic events such as changes in the cell cycle, chromatin state, and nuclear-to-cytoplasmic ratios (Tadros and Lipshitz, 2009). Furthermore, technological advancements have significantly deepened our understanding of EGA. Recently, Asami et al. revealed that gene expression begins at the one-cell stage in human embryos using high-resolution single-cell RNA sequencing (scRNA-seq). Specifically, fertilization triggers EGA by activating maternal transcription factors. While initial gene expression is modest, it remains

elevated until the eight-cell stage (Asami *et al.*, 2022).

Both parental genomes are required for mammalian development

Mammals have relinquished parthenogenesis, the process of producing offspring solely from maternal germ cells. Successful mammalian embryogenesis requires contributions from both maternal and paternal genomes. Although both parents contribute equal nuclear material to the zygote during sexual reproduction, this genetic information is not functionally equivalent. As a result, a diploid genome derived solely from one parent cannot fully support embryogenesis and leads to poor development after implantation. The development of uniparental embryos varies depending on the origin of the pronuclei. Embryos with two male pronuclei show better trophoblast development, while those with two female pronuclei exhibit superior embryonic development. This indicates that the paternal genome is critical for the formation of extraembryonic tissues, whereas the maternal genome is essential for the development of the embryo itself (Barton, Surani and Norris, 1984; Surani, Barton and Norris, 1984). The functional differences between parental genomes arise from the parent-of-origin-specific expression of imprinted genes. Maternally imprinted genes express only the paternal copy, while paternally imprinted genes express only the maternal copy. This expression pattern is regulated by heritable epigenetic modifications, mainly DNA methylation, established in the germ line and carried from gametes to the developing embryo and through to adulthood (Swales and Spears, 2005). Thurston et al. demonstrated that monoallelic expression of most imprinted genes does not occur until after the blastocyst stage (Thurston *et al.*, 2008), indicating that this expression is not crucial for blastocyst development but becomes more significant as the fetus and placenta form. Despite this, parental genome-specific effects have been reported in preimplantation human embryos (Leng *et al.*, 2019). After implantation, a conceptus with only maternal genomic material generally leads to an ovarian teratoma, while one with only paternal genomic material tends to form a complete hydatidiform mole (Jose de Carli and Campos Pereira, 2017).

The origin and impact of chromosomal abnormalities in preimplantation mammalian embryos

Human reproduction is inefficient, with only 30% of conceptions resulting in live births and ~60% of conceptions are lost pre-clinical (Macklon, Geraedts and Fauser, 2002). A major cause of this inefficiency are chromosomal abnormalities in embryos, which can originate from gametogenesis, fertilization, or early development (Delhanty, 2001) and result in embryonic arrest, pregnancy loss, or congenital disorders. Since human embryos are less accessible, model organisms are frequently used for studying embryonic development. Bovine embryos, for instance, serve as a valuable model due to their similarities with human embryos. First,

chromosome instability (CIN) in bovine cleavage-stage embryos is comparable to that in human embryos following *in vitro* fertilization (IVF) (Destouni *et al.*, 2016). Second, the regulatory mechanisms of EGA in bovine embryos are similar to those in human embryos (Halstead *et al.*, 2020). Third, both human (Amargant *et al.*, 2021) and bovine (Uzbekov *et al.*, 2023) zygotes inherit centrioles from the sperm. Differences between the two species have also been observed. For example, while a comparison of *in vivo* and *in vitro* human embryos showed similar euploidy rates (Munné *et al.,* 2020), research in bovine models demonstrated significantly higher chromosomal instability in *in vitro*-cultured embryos compared to those derived in vivo (Tšuiko *et al.,* 2017).

Chromosomal abnormalities caused by meiotic errors during gametogenesis

Human fertility exhibits a U-shaped curve, with reduced fertility rates in both teenagers and women over 35 (Hawkes and Smith, 2010), primarily due to chromosomal abnormalities in oocytes (Gruhn *et al.*, 2019). In young girls, whole-chromosome nondisjunction in meiosis I is a common cause of increased aneuploidy, while in older women, loss of chromosomal cohesion impacts fertility (Gruhn *et al.*, 2019). Furthermore, the spindle assembly checkpoint (SAC), which ensures proper chromosome alignment and kinetochore attachment before anaphase (Lara-Gonzalez, Westhorpe and Taylor, 2012), is less effective during maternal meiosis I in mice. This inefficiency may also account for the inherent error-proneness of human oocytes and their increased susceptibility to chromosomal errors with age (Nagaoka *et al.*, 2011). In contrast, male meiosis is much less error-prone (Bell *et al.*, 2020).

Post-zygotic chromosomal instability

CIN is characterized by an elevated rate of gains or losses of complete chromosomes or segments of chromosomes per cell cycle resulting in cell-to-cell variability. Post-zygotic CIN is characteristic of human cleavage-stage embryos. The results are mosaic embryos caused by mitotic errors, characterized by coexisting cell populations with diverse genome composition. Mosaicism occurs for both whole-chromosome aneuploidies and segmental chromosomal aneuploidies (Vanneste *et al.*, 2009, 2012). The incidence of mosaicism varies widely, ranging from 15% to over 90%, depending on the embryonic stage examined and the method of analysis used (Mantikou *et al.*, 2012). In a later comprehensive study of 28,052 dpf 3 IVF embryos, more than 25% were found to harbor mitotic-origin aneuploidies. These aneuploidies are constant with maternal age, exhibit weak tendencies toward specific chromosomes, and frequently affect multiple chromosomes simultaneously (McCoy, 2017). Tšuiko et al. also confirmed that aneuploidies resulting from postzygotic mitotic errors do not show a preference for either maternal or paternal homologs (Tšuiko *et al.*, 2021). Several mechanisms have been proposed to explain why early cell divisions in mammalian life are so error-prone. Spindle

abnormalities are a key factor underlying CIN at the cleavage stage. Chatzimeletiou et al. observed abnormalities such as abnormal spindle shapes, chromosome loss, and multipolar spindles in *in vitro* preimplantation human embryos. The rate of these abnormalities is significantly higher at the cleavage stages in normally developing embryos but decreased by the blastocyst stage. (Chatzimeletiou *et al.*, 2005). A later study by McCoy et al. identified tripolar chromosome segregation in normally fertilized diploid human IVF zygotes as a key mechanism generating complex aneuploidy in cleavage-stage embryos, strongly associated with common maternal genetic variants involving the centrosomal regulator PLK4 (McCoy *et al.*, 2018). Additionally, a compromised SAC further contributes to the chromosome segregation errors observed in early mammalian embryos. Studies in mouse cleavage-stage embryos revealed that the SAC, which is essential for regulating mitotic cell cycle progression (Wei *et al.*, 2011), is not fully functional. Although misaligned chromosomes can recruit SAC components to initiate a checkpoint signal, this signal often fails to prevent anaphase onset, resulting in high levels of chromosome segregation errors. (Vázquez-Diez, Paim and FitzHarris, 2019). Finally, since CIN is more frequent in *in vitro*-produced embryos compared to *in vivo*-derived embryos, as shown in bovine models (Tšuiko *et al.*, 2017), the *in vitro* process itself may predispose embryos to chromosomal errors.

Fertilization error-induced chromosomal abnormalities

Errors during fertilization, such as polyspermy, are another cause of chromosomal segregation abnormalities in zygotes and contribute to aneuploidies in cleavage-stage embryos. In conventional IVF, polyspermy is the predominant cause of tripronuclear human zygotes. These zygotes often divide directly into three cells, each with a severely abnormal chromosomal complement. Alternatively, they may cleave into two diploid cells plus an extrusion or into two triploid cells (Kola *et al.*, 1987). Subsequent research investigated chromatin and microtubule morphologies during the first cell cycle in dispermic bovine IVF zygotes. These zygotes exhibited multiple sperm asters originating from active paternal centrosomes, demonstrating significant variability in both number and morphology, along with varying degrees of association with the chromatin (Long *et al.*, 1993). Intracytoplasmic sperm injection (ICSI) mitigates the risk of polyspermy; however, tripronuclear digynic zygotes can still arise due to the failure of second polar body extrusion. A recent study compared spindle morphology in diandric human zygotes from conventional IVF with digynic zygotes from ICSI. The study observed that in humans, the first mitotic spindle is formed from the sperm centrosome. Diandric zygotes exhibited multipolar spindles, while digynic zygotes displayed bipolar spindles. The formation of the first mitotic spindle from the sperm centrosome may therefore contribute to a high incidence of zygotic division errors in human tripronuclear zygotes (Kai, Kawano and Yamashita, 2021).

The first zygotic cleavage is inherently error-prone.

Recent studies have highlighted the vulnerability of the first division in mammals, revealing its susceptibility to errors. Contrary to the traditional belief that a single microtubule spindle combines and segregates the two parental genomes to form a two-cell embryo, a recent study revealed that two separate bipolar spindles form in the mouse zygote through centrosome-independent pathways. These two spindles align their poles before anaphase but keep the parental genomes apart during the first cleavage. If the alignment of the two spindles is disrupted, the parental genomes can be pulled in different directions, causing significant mitotic errors (Reichmann *et al.*, 2018). However, extrapolating this observation to other mammals is challenging, as rodent zygotes, unlike those of most other mammals, lack canonical centrosomes (Woolley and Fawcett, 1973; Manandhar *et al.*, 1998). Interestingly, dual spindles were also observed in a human zygote (Xu *et al.*, 2019), suggesting a similar mechanism may be at play in humans. More importantly, Schneider et al. later reported conserved dual spindle assembly in bovine zygotes, which inherit two paternally derived centrioles at fertilization, just like in humans. During pro-metaphase, the two adjacent spindles align their longitudinal axes, giving the appearance of a single fused structure at the metaphase stage. Centrosomes have a limited role in shaping the architecture of the metaphase spindle (Schneider *et al.*, 2021). Overall, the formation and fusion of distinct paternal and maternal spindles during the first cleavage division distinguish this mitotic event from later divisions, which may explain its increased susceptibility to errors. For example, if the two spindles do not fuse properly, it could lead to the formation of multipolar spindles. Although centrosomes are not essential for spindle assembly, they perform other crucial functions in the zygote. Cavazza et al. observed that centrosomes drive the clustering of parental genomes at the pronuclear interface in human and bovine zygotes. This clustering is crucial for the rapid fusion of parental genomes during nuclear envelope breakdown. However, clustering fails frequently, resulting in errors in chromosome segregation (Cavazza *et al.*, 2021). Additionally, live-cell imaging of chromosome segregation also showed that the first mitotic division is a major source of mitotic-origin aneuploidies in human embryos (Currie *et al.*, 2022).

Heterogoneic zygotic division results in mixoploid and chimeric embryos

Although CIN in mammalian cleavage-stage embryos has long been recognized, analyzing the haplotype composition of individual blastomeres has only become feasible with the development of current haplotyping and copy-number profiling methods for single-cells. Using one such methods called haplarithmisis (Zamani Esteki *et al.*, 2015), it was unexpectedly discovered, for the first time, the coexistence of androgenetic, gynogenetic, biparental, and triploid blastomeres within individual dpf 2-3 bovine embryos. These diverse blastomeres result from the spontaneous segregation of entire parental genomes into distinct cell lineages during

the first cleavage division, a phenomenon we refer to as “heterogoneic cell division”. Heterogoneic divisions occur primarily in dispermic embryos but also occur in normally fertilized monospermic embryos. Notably, in all dispermic embryos, one of the two paternal genomes consistently segregates into a distinct cell lineage **(Figure 1)** (Destouni *et al.*, 2016). Since early cleavage-stage embryos display a more comprehensive and unfiltered spectrum of chromosomal abnormalities, it is proposed that mixoploidy following fertilization may be more common than previously recognized. Additionally, the differential developmental fates of blastomeres with varying haplotype compositions likely contribute to chimerism and mixoploidy observed in later stages of development. Heterogoneic divisions were later confirmed in cleavage-stage *in vivo* bovine embryos (Tšuiko *et al.*, 2017), as well as in other IVF bovine embryos (Middelkamp *et al.*, 2020) and IVF rhesus macaque embryos (Daughtry *et al.*, 2019). Zygotic intrinsic asymmetries are speculated to serve as underlying hypothetical models of heterogoneic division, involving mechanisms such as tripolar spindle formation and parental “private” spindle assembly (Destouni and Vermeesch, 2017).

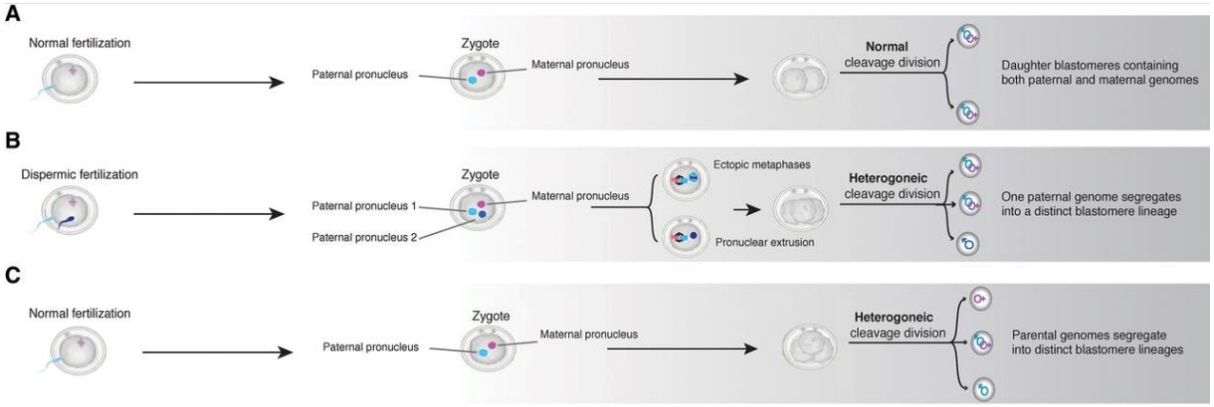


Figure 1. Sequence of events and their outcomes following the first post-zygotic division. A. Normal fertilization followed by post-zygote mitosis, resulting in two diploid biparental daughter blastomeres. **B.** Dispermic fertilization followed by ectopic metaphases or the extrusion of the paternal pronucleus, leading to the segregation of one paternal genome in an androgenetic cell line. **C.** Normal fertilization followed by the segregation of the maternal and the paternal genomes into distinct cell lineages. Figure retrieved with permission from Destouni *et al.*, 2016.

Additionally, frequent chromosomal abnormalities, including WG abnormalities, were observed in dpf 3 human embryos (Dimitriadou *et al.*, 2017). Heterogoneic division may be the underlying mechanistic cause of these abnormalities.

Heterogoneic zygotic division as the origin of mixoploidy and chimerism in human

WG abnormalities are also observed in spontaneous abortions, abnormal pregnancies, and live-born individuals. Approximately 1% of all human conceptuses are triploid, most of which fail to progress beyond the early stages, contributing to ~10% of all spontaneous abortions

(Hassold *et al.*, 1980). Roughly 1 in 1,000 pregnancies results in a complete hydatidiform mole, which carries a paternally derived diploid genome (Seckl, Sebire and Berkowitz, 2010). Additionally, prenatal cases of androgenetic/biparental chimeras have been reported with abnormal development of the placenta and fetus (Morales *et al.*, 2009). Live-born cases of gynogenetic/biparental and androgenetic/biparental chimeras, as well as triploid/diploid mixoploids, have also been observed with congenital abnormalities (Madan, 2020). For each of these studies, researchers proposed mechanisms to explain the origins of the observed abnormalities, including fertilization of an "empty" egg, dispermic fertilization, incorporation of a polar body, fusion of two zygotes, and missegregation of the paternal genome, etc. However, it's important to note that these proposed mechanisms are largely speculative and are based on cell lines detected in late pregnancy or in live-born individuals. Chromosomal errors are known to be selected against during embryo development, meaning that the cell lines observed at later stages may not fully represent the abnormalities present during the initial stages of development. Heterogoneic division, however, could give rise to a variety of uniparental and polyploid cells within the same embryo. Depending on their developmental trajectories, these cells could lead to uniparental, polyploid, or chimeric outcomes, providing a more comprehensive and unbiased view of the origins of these abnormalities (Masset, Tšuiko and Vermeesch, 2021).

Effects of chromosomal errors on embryo development

Most chromosomal abnormalities are incompatible with survival to an advanced stage of pregnancy and are gradually selected against. Approximately 50% of abortions are chromosomally abnormal, compared to 5% of stillbirths and 0.5% of live births (Hassold, 1986). This selection against chromosomally abnormal human embryos begins as early as the preimplantation stage of development. Rubio et al. observed that normal euploid embryos on dpf 3 showed significantly higher blastocyst formation rates on dpf 5 compared to chromosomally abnormal and mosaic embryos (Rubio *et al.*, 2007). Tšuiko et al. similarly observed that the incidence of genomic aberrations was significantly higher in cleavage-stage embryos (43.2%) compared to blastocysts (24.9%) (Tšuiko *et al.*, 2021). McCoy et al. found that after EGA, whole-chromosome abnormalities are heavily filtered, with complex aneuploidies from catastrophic mitotic errors being the primary cause of embryo losses before blastocyst formation. In contrast, embryos with generally milder meiotic errors are more likely to progress to the blastocyst stage (McCoy *et al.*, 2015). However, when accounting for the number of affected chromosomes the type of aneuploidy (mitotic *versus* meiotic) does not significantly influence the likelihood of embryo arrest. Additionally, aneuploidies were found to delay blastocyst formation and expansion, and to adversely affect blastocyst morphology (McCoy *et al.*, 2023).

Mitotic errors create two or more karyotypically distinct cell lineages within a single embryo, a phenomenon known as chromosomal mosaicism. During preimplantation development, the frequency of mosaicism increases from 49% at the cleavage stage to 91% at the blastocyst stage. However, the proportion of aneuploid cells within mosaic embryos decreases over this period, from 40% at the cleavage stage to 22% in the blastocyst. Additionally, the number of chromosomes affected by chaotic abnormalities is lower at the blastocyst stage compared to the cleavage stage (McCoy, 2017). Three models explain the increasing frequency of diploid cells from the cleavage to the blastocyst stage: (1) early mortality, which selects against embryos with high levels of mosaicism; (2) clonal depletion, involving the active elimination or ineffective propagation of aneuploid cells within mosaic embryos; and (3) trisomic/monosomic rescue, in which aneuploid cell lines are corrected by mitotic errors that restore diploidy through chromosome gain or loss **(Figure 2)** (McCoy, 2017).

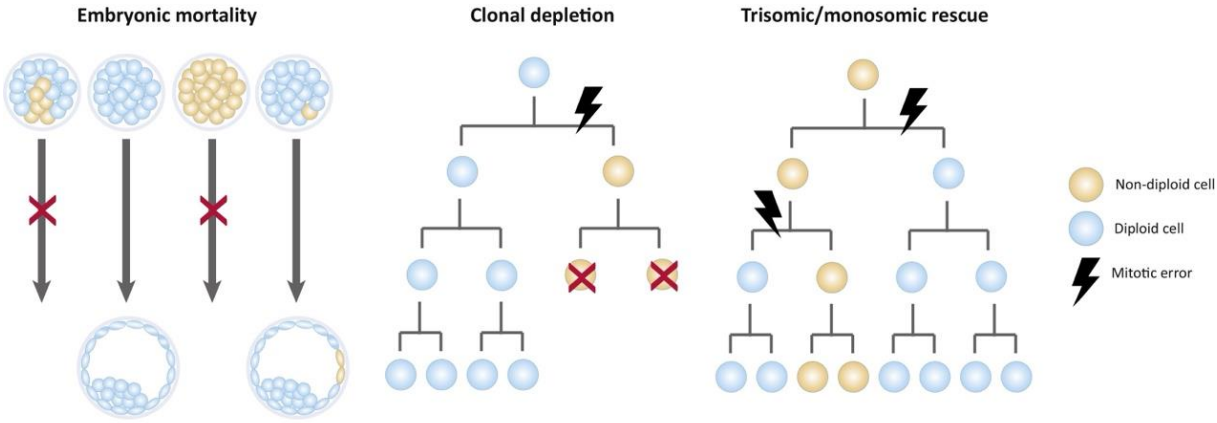


Figure 2. Models explaining the increasing frequency of diploid cells from the cleavage to the blastocyst stage. The embryonic mortality model invokes selection against embryos based on the proportion of aneuploid cells. The clonal depletion model describes apoptosis or reduced propagation of aneuploid cells within mosaic embryos. Monosomic and trisomic rescue are proposed mechanisms by which aneuploid cells can give rise to diploid cells through mitotic chromosome gain or loss, respectively. Figure retrieved with permission from McCoy, 2017.

The most common pattern of mosaicism is euploid–aneuploid mosaicism, where the embryo contains both normal and abnormal cells (van Echten-Arends *et al.*, 2011). Detailed mechanisms by which aneuploid cells are eliminated from euploid–aneuploid mosaic embryos have been elucidated. Consistent with the “clonal depletion” hypothesis, Bolton et al. demonstrated the progressive, lineage-specific depletion of aneuploid cells from the preimplantation embryo. Aneuploid cells are first noticeably eliminated from the embryo during blastocyst maturation, just before implantation. They induce apoptosis in the ICM and cause cell cycle delays in the TE (Bolton *et al.*, 2016). Later, the same group observed that autophagy-mediated apoptosis further eliminates aneuploid cells during implantation and early

post-implantation development. Proteotoxic stress activates autophagy, and p53-induced autophagy is essential for removing aneuploid cells. Normal diploid cells then increase their proliferation rate to compensate for the elimination of aneuploid cells (Singla *et al.*, 2020).

The gradual elimination of aneuploid cells during embryonic development aligns with the findings of Zamani Esteki et al. (Zamani Esteki *et al.*, 2019). They profiled the genomic landscape of fetal and placental tissues postpartum from both IVF and naturally conceived children, revealing that the high level of chromosomal mosaicism observed in IVF cleavage-stage embryos does not persist into later stages of prenatal development. Similarly, Capalbo et al. observed that low-to-moderate mosaicism (<50%) has the same developmental potential as fully euploid embryos and do not appear to affect live birth (Capalbo *et al.*, 2021).

Technologies for preimplantation embryo selection and analysis

Since the birth of the first baby conceived through IVF in 1978 (Steptoe and Edwards, 1978), IVF has become a vital method for treating infertility. Embryo quality is widely recognized as the most critical factor for achieving successful implantation and the birth of a healthy, full-term baby. To improve embryo selection, various technologies have been applied or developed. Morphological assessments using microscopes or time-lapse imaging help identify embryos with optimal morphology and morphokinetics. Genome analysis through preimplantation genetic testing (PGT) enables the selection of euploid embryos free from parental disease mutations. Additionally, we are witnessing rapid advancements in transcriptomics and other omics technologies, which are enhancing our understanding of the complex biological processes underlying preimplantation embryo development.

Visual assessment of embryo morphology

Visual assessment of embryo morphology using microscopy is the most traditional and straightforward method of embryo selection, and it was the first parameter extensively evaluated in the early days of IVF. Historically, the strategy focused on selecting "good quality" embryos at the cleavage stage, primarily based on cell number, fragmentation, and multinucleation. Today, with improved culture conditions, extended culture to the blastocyst stage and selecting "good quality" blastocysts is increasingly practiced. The current practice evaluates blastocysts using three parameters: the stage of development, the quality of the ICM, and the quality of the TE. These factors are used to select the most viable blastocysts for embryo transfer (Gardner *et al.*, 2000).

With a traditional microscope, only limited static images are obtained, which fail to accurately reflect the dynamic process of embryonic growth, leading to incorrect scoring and improper ranking of embryo quality. For instance, there is a greater than 25% chance that a poor-quality dpf 3 embryo will develop into a high-quality blastocyst with similar implantation potential to

that of a good-quality embryo (Zhang *et al.*, 2011). The introduction of time-lapse monitoring systems has transformed workflows in IVF laboratories by capturing morphokinetic variables—such as pronuclear appearance and disappearance, cell divisions, and blastocyst formation—at precise time points during embryo development. Despite these apparent benefits, a recent Cochrane meta-analysis of nine randomized controlled trials (n = 2,955 couples) found insufficient outcome differences between embryos cultured or selected using time-lapse systems compared to traditional methods (Armstrong *et al.*, 2019). While the clinical advantages of time-lapse technology for embryo selection remain inconclusive, it can be utilized to deselect embryos with atypical cleavage patterns that negatively impact implantation, such as multipolar cleavage (Athayde Wirka *et al.*, 2014).

To enhance consistency and accuracy in embryo selection based on embryo morphology, the Istanbul Consensus Workshop on Embryo Assessment established standardized guidelines for evaluating oocyte and embryo morphology (Balaban et al., 2011). These recommendations provide a common language for embryologists worldwide, including standardized timing for observing fertilized oocytes and embryos with expected stage of development at each time point, as well as consensus scoring systems for pronuclei, cleavage-stage embryos, dpf 4 embryos, and blastocysts. All laboratories performing medically assisted reproduction (MAR) should be capable of providing this information. Serving as the minimum standards for oocyte and embryo morphology scoring, these guidelines should be regularly reviewed to ensure their continued relevance.

Preimplantation genetic testing

Morphologic assessment has been the primary method for prioritizing IVF embryos, but it cannot accurately determine chromosome status. Since embryos with genetic anomalies are less likely to develop into healthy live births, PGT was developed and applied in IVF clinics to assess and select genetically normal embryos for transfer. PGT consists of three primary forms: PGT-A, PGT-M, and PGT-SR, where PGT-A screens embryos for chromosomal abnormalities, PGT-M targets monogenic disorders, and PGT-SR focuses on structural rearrangements. During PGT, a biopsy is taken from the developing embryo for genetic analysis. Initially, a single blastomere was obtained on dpf 3, when pre-compacted embryos typically contain 6 to 8 cells. Currently, PGT generally involves taking 5 to 10 cells from the TE of a day 5 or 6 blastocyst. Embryos that are free of mutations for the tested diseases or possess a balanced genetic constitution after PGT can then be transferred to the woman's uterus (Vermeesch, Voet and Devriendt, 2016).

The high incidence of chromosomal aneuploidy in human gametes and embryos contributes to IVF failure and miscarriage. To address this challenge, PGT-A is employed to prioritize

chromosomally normal embryos for transfer, potentially increasing pregnancy rates and reducing miscarriage rates. Initially, fluorescent in situ hybridization (FISH) analysis was used, employing a limited number of probes to target specific chromosomes (Gianaroli *et al.*, 1997). While FISH allowed for the determination of individual chromosome copy numbers, its reliance on a limited number of probes restricted its effectiveness. Consequently, newer methods, such as multiplex quantitative polymerase chain reaction (qPCR) (Treff *et al.*, 2012) and single-cell microarrays (Le Caignec *et al.*, 2006; Treff *et al.*, 2010), have emerged to assess all chromosomes. With the development of next-generation sequencing (NGS), low-pass whole genome sequencing is now frequently used for aneuploidy screening (Wells *et al.*, 2014). More recently, long-read sequencing has been developed to enable long-read based PGT-A (Tan *et al.*, 2023). Although PGT-A is commonly performed, the true clinical benefits remain a topic of discussion, particularly due to the high mosaicism rates in preimplantation embryos and the ability of mosaic embryos to result in healthy live births.

In addition to aneuploidy, structural chromosomal rearrangements can lead to infertility, repeated implantation failure, pregnancy loss, and the birth of congenitally affected children, even though balanced carrier parents often show no apparent phenotypic abnormalities. The risk of couples with balanced chromosomal rearrangements producing chromosomally unbalanced gametes and embryos highlights the importance of PGT-SR, which aims to select embryos that are chromosomally normal or balanced like the carrier parent. PGT-SR primarily employs the same technologies as PGT-A, initially utilizing FISH and now incorporating array-based (Vanneste *et al.*, 2011) and sequencing-based methods (Voet *et al.*, 2013; Griffin and Ogur, 2023).

PGT-M is an essential tool for couples who are known carriers of mutant alleles. This testing enables the identification of embryos affected by these mutations, allowing healthcare providers to transfer only unaffected embryos, thereby preventing the transmission of genetic disorders to offspring. Despite its importance, PGT-M faces inherent challenges due to the limited starting material available from embryo biopsies, which can lead to issues such as allele dropout and preferential amplification, resulting in misdiagnosis. To mitigate the risk of misdiagnosis, PGT-M often incorporates simultaneous haplotype linkage analysis (Rechitsky *et al.*, 1998). This approach ensures that only embryos without the haplotype containing the disease-causing allele are selected for transfer. Additionally, advancements in whole genome amplification (WGA) methods and the development of genome-wide analysis technologies such as SNP array and NGS, have facilitated a shift from targeted PGT-M to WGA-based genome-wide PGT-M. This transition eliminates the need for personalized design for each individual case (de Rycke *et al.*, 2020).

An ideal comprehensive PGT approach would enable simultaneous, genome-wide PGT-A,

PGT-M, and PGT-SR. This is made possible by genome-wide SNP genotype calls, which allow for the reconstruction of genome-wide haplotype. Currently, several primary single-cell haplotyping algorithms utilizing SNP-array data are available, including karyomapping (Handyside *et al.*, 2010) and haplarithmisis (Zamani Esteki *et al.*, 2015). These algorithms have been successfully implemented in clinical practice, enabling PGT-M and the simultaneous detection of a wide range of chromosomal aberrations across the genome (Ben-Nagi *et al.*, 2017; Dimitriadou *et al.*, 2017). With the development of NGS, read-based solutions for comprehensive PGT that utilize the principles of haplarithmisis, including genotyping-by-sequencing (GBS) (Masset *et al.*, 2019) and whole genome sequencing (WGS)-based comprehensive PGT (Janssen *et al.*, 2024), demonstrate good alternative to the traditional array-based method. Additionally, another NGS-based method, Haploseek, has been developed to enable comprehensive PGT using economical low-coverage sequencing (Backenroth *et al.*, 2018).

Polygenic risk scores (PRSs), which capture the contributions of thousands of small allelic effects on complex traits, have become increasingly powerful due to the availability of results from large-scale genome-wide association studies (GWAS) for polygenic disorders (Visscher *et al.*, 2017). As a result of this advancement and the development of comprehensive PGT methods, PGT for polygenic disorders (PGT-P) is now possible. PRSs are calculated for each embryo to estimate the likelihood of developing common diseases, such as diabetes, with embryos selected for implantation based on their polygenic scores. Although PGT-P has been available since 2019 (Treff *et al.*, 2019; Kumar *et al.*, 2022), the effectiveness of PRSs in reducing disease risk in preimplantation embryos remains unproven, and significant ethical concerns persist, making it a topic of ongoing debate (Lázaro-Muñoz *et al.*, 2021; Siermann *et al.*, 2024).

Recently, there has been a growing focus on noninvasive method for preimplantation embryo screening through the analysis of spent culture media. Huang et al. examined the efficacy of noninvasive PGT-A (niPGT-A) by analyzing the cell-free DNA in spent culture media of human blastocysts. Their findings suggest that niPGT-A is more reliable than traditional TE biopsy PGT-A (Huang *et al.*, 2019). Additionally, HLA-G has been recognized as a promising biomarker for embryo selection, as its levels in dpf 3 and dpf 5 media correlate with blastocyst implantation potential (Radwan *et al.*, 2022). Furthermore, Ferrick et al. demonstrated that embryos with better morphological classification and those that successfully established pregnancy consumed more glucose, based on their analysis of spent culture media (Ferrick, Lee and Gardner, 2020). This indicates that glucose consumption could serve as a potential indicator for embryo selection. Despite these advancements in noninvasive preimplantation embryo screening, no commercial device is currently available for routine clinical use.

Whole genome amplification

The advent of WGA protocols enabled the acquisition of sufficient input material for genome-wide platforms from as little as 6 to 7 pg of DNA from a single cell, which marked a significant turning point in PGT. By accessing an almost limitless supply of amplified DNA from a single or a few cells, multilocus analysis became feasible, making PGT more versatile and safer.

Present-day WGA techniques include those based on PCR, multiple displacement amplification (MDA), and other innovative methods. Among the earliest of these was primer extension pre-amplification PCR (PEP-PCR), introduced by Zhang et al. in 1992, which allowed for the amplification of DNA from single haploid cells (Zhang *et al.*, 1992). Although PEP-PCR is less commonly used today, its principles have influenced subsequent WGA developments. In the same year, the more widely adopted degenerate oligonucleotide-primed PCR (DOP-PCR) was developed (Telenius *et al.*, 1992). Following this, other PCR-based methods emerged, including multiple annealing and looping-based amplification cycles (MALBAC). MALBAC introduces quasi-linear preamplification, which helps reduce the bias typically associated with non-linear amplification (Zong *et al.*, 2012). Generally, PCR-based methods yield relatively short products, typically no longer than a few kilobases, and preferentially amplify fragments that are a few hundred base pairs in length (Del Rey *et al.*, 2018). The concept of MDA, which relies on strand displacement and rolling circular amplification, was first proposed in 1998 (Lizardi *et al.*, 1998). Dean et al. later refined this into a robust WGA method in 2002 (Dean *et al.*, 2002) **(Figure 3)**. MDA primarily relies on a mild isothermal reaction to produce high-quality DNA products using phi29 DNA polymerase and random hexamer primers. This method typically yields DNA fragments ranging from 50 to 100 kb (Huang *et al.*, 2015). Subsequently, Gonzalez-Pena et al. introduced the primary template-directed amplification (PTA) approach in 2021 (Gonzalez-Pena *et al.*, 2021). This new MDA-based method incorporates an extension terminator, resulting in a quasi-linear amplification process that reduces errors and significantly enhances genome coverage.

In terms of applications, PCR-based WGA methods offer better uniformity of coverage, making them more suitable for copy number variation (CNV) studies. In contrast, MDA-based WGA systems are more appropriate for SNV analysis due to better genome coverage (Zhou *et al.*, 2020). Despite this differentiation, MDA remains the most popular WGA method (Gawad, Koh and Quake, 2016) due to its high DNA yield, fidelity, and low amplification bias (Dean *et al.*, 2002). When it comes to WGA of DNA from single cells for long-read sequencing, many traditional methods fall short because they produce limited fragment lengths after amplification. MDA stands out as a promising option, as it can generate longer fragments. Hård et al. further advanced this technique by employing droplet-based MDA, where MDA reactions are performed in individual droplets. They demonstrated that this approach enables long-read

whole-genome analysis of single human cells (Hård *et al.*, 2023).

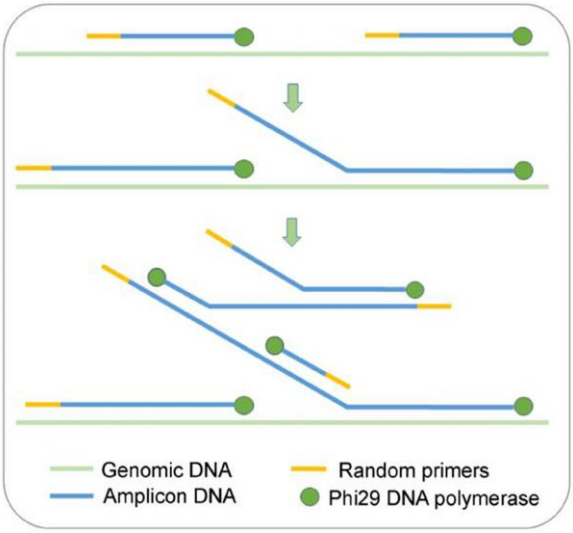


Figure 3. Multiple displacement amplification (MDA) process diagram. Figure retrieved with permission from Long *et al.*, 2020.

Although MDA is frequently used, no WGA method is currently considered the gold standard in the ever-evolving field of PGT. WGA methods have certain drawbacks, including incomplete genomic coverage, amplification bias, and the formation of chimeras (Sabina and Leamon, 2015). Incomplete genomic coverage becomes more apparent at very low initial DNA concentrations, where stochastic effects dictate whether specific genomic regions will be amplified. Amplification bias arises when certain regions or amplicons within a multitemplate reaction are preferentially amplified over the entire pool of potential templates, with the effect becoming more pronounced at lower initial template concentrations. Additionally, WGA can produce chimera amplicons, which are artificial constructs that map to different parts of the genome that are not physically linked. These errors can affect downstream analysis and should be carefully managed.

Single-cell omics technologies

Recent advancements in single-cell omics technologies have transformed our understanding of preimplantation embryo development. One of the most significant tools in this field is scRNA-seq, which provides valuable insights into early embryonic development. For example, with scRNA-seq, Petropoulos et al. unveiled the gene expression dynamics during early cell-lineage decisions of human embryos (Petropoulos *et al.*, 2016). Additionally, Sha et al. utilized scRNA-seq to explore the dynamics, functional importance, and pathological relevance of maternal mRNA decay in human preimplantation embryos. Their findings highlighted the critical role of this decay process during early development (Sha *et al.*, 2020). Furthermore, Asami et al. employed high-resolution scRNA-seq to uncover previously inaccessible gene

expression changes from oocyte to embryo. Their research demonstrated that human embryonic transcription begins at the one-cell stage, providing new insights into the initiation of development (Asami *et al.*, 2022).

The zygote experiences remarkable epigenetic changes in addition to transcriptional modifications (Eckersley-Maslin, Alda-Catalinas and Reik, 2018). Besides the development and application of scRNA-seq, new technologies have emerged to measure these epigenetic signatures at the single-cell level (Wen and Tang, 2018). Utilizing these advanced techniques has led to significant discoveries in preimplantation epigenetic changes. For instance, single-cell DNA methylome sequencing analysis of preimplantation embryos has shown that tens of thousands of genomic loci undergo *de novo* DNA methylation. This observation suggests that genome-wide DNA methylation reprogramming during preimplantation development is characterized by a dynamic balance between extensive global demethylation and targeted remethylation (Zhu *et al.*, 2017). Such insights are crucial for understanding the complexities of DNA methylation reprogramming in early human embryos.

Single-cell multi-omics analysis offers a comprehensive view of the biological processes within a cell, surpassing the insights provided by any single omics dataset. Several multi-omics technologies have been developed to facilitate this integrated analysis. For instance, G&T-seq enables the simultaneous analysis of genomic and transcriptomic data from same single cells (Macaulay *et al.*, 2015), while scM&T-seq allows for parallel examination of the methylome and transcriptome at the single-cell level (Angermueller *et al.*, 2016). Additionally, single-cell COOL-seq technology can simultaneously assess chromatin state, nucleosome positioning, DNA methylation, CNV, and ploidy from individual cells (Guo *et al.*, 2017). The application of single-cell multi-omics tools provides crucial mechanistic insights into the interrelated molecular events underlying human pre-implantation development. For instance, Liu et al. utilized LiCAT-seq (low-input chromatin accessibility and transcriptome sequencing) to profile chromatin structure and gene expression dynamics in human pre-implantation embryos. They demonstrated the key regulatory mechanisms for genes activated during EGA (Liu *et al.*, 2019). Fernandez Gallardo et al. investigated gene expression perturbations instigated by aneuploidy using G&T-seq, enhancing our understanding of how chromosomal abnormalities affect the transcriptome of preimplantation embryos (Fernandez Gallardo *et al.*, 2023).

References

Amargant, F. *et al.* (2021) 'The human sperm basal body is a complex centrosome important for embryo preimplantation development', *Molecular Human Reproduction*, 27(11). Available at: https://doi.org/10.1093/MOLEHR/GAAB062.

Angermueller, C. *et al.* (2016) 'Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity', *Nature methods*, 13(3), pp. 229–232. Available at: https://doi.org/10.1038/NMETH.3728.

Armstrong, S. *et al.* (2019) 'Time-lapse systems for embryo incubation and assessment in assisted reproduction', *The Cochrane database of systematic reviews*, 5(5). Available at: https://doi.org/10.1002/14651858.CD011320.PUB4.

Asami, M. *et al.* (2022) 'Human embryonic genome activation initiates at the one-cell stage', *Cell Stem Cell*, 29(2), pp. 209-216.e4. Available at: https://doi.org/10.1016/j.stem.2021.11.012.

Athayde Wirka, K. *et al.* (2014) 'Atypical embryo phenotypes identified by time-lapse microscopy: High prevalence and association with embryo development', *Fertility and Sterility*, 101(6), pp. 1637-1648.e5. Available at: https://doi.org/10.1016/j.fertnstert.2014.02.050.

Backenroth, D. *et al.* (2018) 'Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (PGD) of monogenic disease and aneuploidy', *Genetics in Medicine 2018 21:6*, 21(6), pp. 1390–1399. Available at: https://doi.org/10.1038/s41436-018-0351-7.

Balaban, B. et al. (2011) 'The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting', *Human reproduction (Oxford, England)*, 26(6), pp. 1270–1283. Available at: https://doi.org/10.1093/HUMREP/DER037.

Barton, S.C., Surani, M.A.H. and Norris, M.L. (1984) 'Role of paternal and maternal genomes in mouse development', *Nature 1984 311:5984*, 311(5984), pp. 374–376. Available at: https://doi.org/10.1038/311374a0.

Bell, A.D. *et al.* (2020) 'Insights into variation in meiosis from 31,228 human sperm genomes', *Nature 2020 583:7815*, 583(7815), pp. 259–264. Available at: https://doi.org/10.1038/s41586-020-2347-0.

Ben-Nagi, J. *et al.* (2017) 'Karyomapping: a single centre's experience from application of methodology to ongoing pregnancy and live-birth rates', *Reproductive biomedicine online*, 35(3), pp. 264–271. Available at: https://doi.org/10.1016/J.RBMO.2017.06.004.

Bolton, H. *et al.* (2016) 'Mouse model of chromosome mosaicism reveals lineage-specific depletion of aneuploid cells and normal developmental potential', *Nature Communications 2016 7:1*, 7(1), pp. 1–12. Available at: https://doi.org/10.1038/ncomms11165.

Braude, P., Bolton, V. and Moore, S. (1988) 'Human gene expression first occurs between the four- and eight-cell stages of preimplantation development', *Nature 1988 332:6163*, 332(6163), pp. 459–461. Available at: https://doi.org/10.1038/332459a0.

Le Caignec, C. *et al.* (2006) 'Single-cell chromosomal imbalances detection by array CGH', *Nucleic Acids Research*, 34(9), p. e68. Available at: https://doi.org/10.1093/NAR/GKL336.

Capalbo, A. *et al.* (2021) 'Mosaic human preimplantation embryos and their developmental potential in a prospective, non-selection clinical trial', *American Journal of Human Genetics*, 108(12), pp. 2238–2247. Available at: https://doi.org/10.1016/j.ajhg.2021.11.002.

Cavazza, T. *et al.* (2021) 'Parental genome unification is highly error-prone in mammalian embryos', *Cell*, 184(11), pp. 2860-2877.e22. Available at: https://doi.org/10.1016/J.CELL.2021.04.013/ATTACHMENT/E77CCF69-14E8-4F43-9711-31598ADFA3B5/MMC6.MP4.

Chatzimeletiou, K. *et al.* (2005) 'Spindle abnormalities in normally developing and arrested human preimplantation embryos in vitro identified by confocal laser scanning microscopy', *Human Reproduction*, 20(3), pp. 672–682. Available at: https://doi.org/10.1093/HUMREP/DEH652.

Currie, C.E. *et al.* (2022) 'The first mitotic division of human embryos is highly error prone', *Nature Communications 2022 13:1*, 13(1), pp. 1–13. Available at: https://doi.org/10.1038/s41467-022-34294-6.

Daughtry, B.L. *et al.* (2019) 'Single-cell sequencing of primate preimplantation embryos reveals chromosome elimination via cellular fragmentation and blastomere exclusion', *Genome Research*, 29(3), pp. 367–382. Available at: https://doi.org/10.1101/GR.239830.118.

Dean, F.B. *et al.* (2002) 'Comprehensive human genome amplification using multiple displacement amplification', *Proceedings of the National Academy of Sciences*, 99(8), pp. 5261–5266. Available at: https://doi.org/10.1073/PNAS.082089499.

Delhanty, J.D.A. (2001) 'Preimplantation genetics: an explanation for poor human fertility?', *Annals of human genetics*, 65(Pt 4), pp. 331–338. Available at: https://doi.org/10.1017/S0003480001008739.

Destouni, A. *et al.* (2016) 'Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy', *Genome Research*, 26(5), pp. 567–578. Available at: https://doi.org/10.1101/GR.200527.115.

Destouni, A. and Vermeesch, J.R. (2017) 'How can zygotes segregate entire parental genomes into distinct blastomeres? The zygote metaphase revisited', *BioEssays : news and reviews in molecular, cellular and developmental biology*, 39(4). Available at: https://doi.org/10.1002/BIES.201600226.

Dimitriadou, E. *et al.* (2017) 'Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos', *Human Reproduction*, 32(3), pp. 687–697. Available at: https://doi.org/10.1093/HUMREP/DEX011.

van Echten-Arends, J. *et al.* (2011) 'Chromosomal mosaicism in human preimplantation embryos: a systematic review', *Human Reproduction Update*, 17(5), pp. 620–627. Available at: https://doi.org/10.1093/humupd/dmr014.

Eckersley-Maslin, M.A., Alda-Catalinas, C. and Reik, W. (2018) 'Dynamics of the epigenetic landscape during the maternal-to-zygotic transition', *Nature Reviews Molecular Cell Biology 2018 19:7*, 19(7), pp. 436–450. Available at: https://doi.org/10.1038/s41580-018-0008-z.

Fernandez Gallardo, E. *et al.* (2023) 'A multi-omics genome-and-transcriptome single-cell atlas of human preimplantation embryogenesis reveals the cellular and molecular impact of chromosome instability', *bioRxiv*, p. 2023.03.08.530586. Available at: https://doi.org/10.1101/2023.03.08.530586.

Ferrick, L., Lee, Y.S.L. and Gardner, D.K. (2020) 'Metabolic activity of human blastocysts correlates with their morphokinetics, morphological grade, KIDScore and artificial intelligence ranking', *Human reproduction (Oxford, England)*, 35(9), pp. 2004–2016. Available at: https://doi.org/10.1093/HUMREP/DEAA181.

Gardner, D.K. *et al.* (2000) 'Blastocyst score affects implantation and pregnancy outcome: Towards a single blastocyst transfer', *Fertility and Sterility*, 73(6), pp. 1155–1158. Available at: https://doi.org/10.1016/S0015-0282(00)00518-5.

Gawad, C., Koh, W. and Quake, S.R. (2016) 'Single-cell genome sequencing: current state of the science', *Nature Reviews Genetics 2016 17:3*, 17(3), pp. 175–188. Available at: https://doi.org/10.1038/nrg.2015.16.

Gianaroli, L. *et al.* (1997) 'Preimplantation genetic diagnosis increases the implantation rate in human in vitro fertilization by avoiding the transfer of chromosomally abnormal embryos', *Fertility and Sterility*, 68(6), pp. 1128–1131. Available at: https://doi.org/10.1016/S0015-0282(97)00412-3.

Gonzalez-Pena, V. *et al.* (2021) 'Accurate genomic variant detection in single cells with primary template-directed amplification', *Proceedings of the National Academy of Sciences of the United States of America*, 118(24). Available at: https://doi.org/10.1073/PNAS.2024176118/-/DCSUPPLEMENTAL.

Griffin, D.K. and Ogur, C. (2023) 'PGT-SR: A Comprehensive Overview and a Requiem for the Interchromosomal Effect', *DNA 2023, Vol. 3, Pages 41-64*, 3(1), pp. 41–64. Available at: https://doi.org/10.3390/DNA3010004.

Gruhn, J.R. *et al.* (2019) 'Chromosome errors in human eggs shape natural fertility over reproductive life span', *Science*, 365(6460), pp. 1466–1469. Available at: https://doi.org/10.1126/SCIENCE.AAV7321/SUPPL_FILE/AAV7321S1.MOV.

Guo, F. *et al.* (2017) 'Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells', *Cell Research 2017 27:8*, 27(8), pp. 967–988. Available at: https://doi.org/10.1038/cr.2017.82.

Halstead, M.M. *et al.* (2020) 'Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation', *Nature Communications 2020 11:1*, 11(1), pp. 1–16. Available at: https://doi.org/10.1038/s41467-020-18508-3.

Handyside, A.H. *et al.* (2010) 'Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes', *Journal of Medical Genetics*, 47(10), pp. 651–658. Available at: https://doi.org/10.1136/JMG.2009.069971.

Hård, J. *et al.* (2023) 'Long-read whole-genome analysis of human single cells', *Nature Communications 2023 14:1*, 14(1), pp. 1–12. Available at: https://doi.org/10.1038/s41467-023-40898-3.

Hassold, T. *et al.* (1980) 'A cytogenetic study of 1000 spontaneous abortions', *Annals of Human Genetics*, 44(2), pp. 151–164. Available at: https://doi.org/10.1111/J.1469-1809.1980.TB00955.X.

Hassold, T.J. (1986) 'Chromosome abnormalities in human reproductive wastage', *Trends in Genetics*, 2(C), pp. 105–110. Available at: https://doi.org/10.1016/0168-9525(86)90194-0.

Hawkes, K. and Smith, K.R. (2010) 'Do women stop early? Similarities in fertility decline in humans and chimpanzees', *Annals of the New York Academy of Sciences*, 1204, pp. 43–53. Available at: https://doi.org/10.1111/J.1749-6632.2010.05527.X.

Huang, L. *et al.* (2015) 'Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications', *Annual review of genomics and human genetics*, 16, pp. 79–102. Available at: https://doi.org/10.1146/ANNUREV-GENOM-090413-025352.

Huang, L. *et al.* (2019) 'Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy', *Proceedings of the National Academy of Sciences of the United States of America*, 116(28), pp. 14105–14112. Available at: https://doi.org/10.1073/PNAS.1907472116/SUPPL_FILE/PNAS.1907472116.SAPP.PDF.

Janssen, A.E.J. *et al.* (2024) 'Clinical-grade whole genome sequencing-based haplarithmisis enables all forms of preimplantation genetic testing', *Nature Communications 2024 15:1*, 15(1), pp. 1–15. Available at: https://doi.org/10.1038/s41467-024-51508-1.

Jose de Carli, G. and Campos Pereira, T. (2017) 'On human parthenogenesis', *Medical Hypotheses*, 106, pp. 57–60. Available at: https://doi.org/10.1016/J.MEHY.2017.07.008.

Kai, Y., Kawano, H. and Yamashita, N. (2021) 'First mitotic spindle formation is led by sperm centrosome-dependent MTOCs in humans', *Reproduction*, 161(5), pp. V19–V22. Available at: https://doi.org/10.1530/REP-21-0061/VIDEO-3.

Kola, I. *et al.* (1987) 'Tripronuclear Human Oocytes: Altered Cleavage Patterns and Subsequent Karyotypic Analysis of Embryos', *Biology of Reproduction*, 37(2), pp. 395–401. Available at: https://doi.org/10.1095/BIOLREPROD37.2.395.

Kumar, A. *et al.* (2022) 'Whole-genome risk prediction of common diseases in human preimplantation embryos', *Nature Medicine 2022 28:3*, 28(3), pp. 513–516. Available at: https://doi.org/10.1038/s41591-022-01735-0.

Lara-Gonzalez, P., Westhorpe, F.G. and Taylor, S.S. (2012) 'The Spindle Assembly Checkpoint', *Current Biology*, 22(22), pp. R966–R980. Available at: https://doi.org/10.1016/J.CUB.2012.10.006.

Lázaro-Muñoz, G. *et al.* (2021) 'Screening embryos for polygenic conditions and traits: ethical considerations for an emerging technology', *Genetics in Medicine*, 23(3), pp. 432–434. Available at: https://doi.org/10.1038/s41436-020-01019-3.

Leng, L. *et al.* (2019) 'Single-Cell Transcriptome Analysis of Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation Development', *Cell Stem Cell*, 25(5), pp. 697-712.e6. Available at: https://doi.org/10.1016/j.stem.2019.09.004.

Liu, L. *et al.* (2019) 'An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos', *Nature Communications 2019 10:1*, 10(1), pp. 1–11. Available at: https://doi.org/10.1038/s41467-018-08244-0.

Lizardi, P.M. *et al.* (1998) 'Mutation detection and single-molecule counting using isothermal rolling-circle amplification', *Nature Genetics 1998 19:3*, 19(3), pp. 225–232. Available at: https://doi.org/10.1038/898.

Long, C.R. *et al.* (1993) 'Chromatin and microtubule morphology during the first cell cycle in bovine zygotes', *Molecular Reproduction and Development*, 36(1), pp. 23–32. Available at: https://doi.org/10.1002/MRD.1080360105.

Long, N. et al. (2020) 'Recent advances and application in whole-genome multiple displacement amplification', *Quantitative Biology*, 8(4), pp. 279–294. Available at: https://doi.org/10.1007/S40484-020-0217-2.

Macaulay, I.C. *et al.* (2015) 'G&T-seq: parallel sequencing of single-cell genomes and transcriptomes', *Nature methods*, 12(6), pp. 519–522. Available at: https://doi.org/10.1038/NMETH.3370.

Macklon, N.S., Geraedts, J.P.M. and Fauser, B.C.J.M. (2002) 'Conception to ongoing pregnancy: the "black box" of early pregnancy loss', *Human Reproduction Update*, 8(4), pp. 333–343. Available at: https://doi.org/10.1093/HUMUPD/8.4.333.

Madan, K. (2020) 'Natural human chimeras: A review', *European Journal of Medical Genetics*, 63(9), p. 103971. Available at: https://doi.org/10.1016/J.EJMG.2020.103971.

Manandhar, G. *et al.* (1998) 'Centrosome Reduction during Mouse Spermiogenesis', *Developmental Biology*, 203(2), pp. 424–434. Available at: https://doi.org/10.1006/DBIO.1998.8947.

Mantikou, E. *et al.* (2012) 'Molecular origin of mitotic aneuploidies in preimplantation embryos', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1822(12), pp. 1921–1930. Available at: https://doi.org/10.1016/J.BBADIS.2012.06.013.

Masset, H. *et al.* (2019) 'Multi-centre evaluation of a comprehensive preimplantation genetic test through haplotyping-by-sequencing', *Human Reproduction*, 34(8), pp. 1608–1619. Available at: https://doi.org/10.1093/HUMREP/DEZ106.

Masset, H., Tšuiko, O. and Vermeesch, J.R. (2021) 'Genome-wide abnormalities in embryos: Origins and clinical consequences', *Prenatal diagnosis*, 41(5), pp. 554–563. Available at: https://doi.org/10.1002/PD.5895.

McCoy, R.C. *et al.* (2015) 'Evidence of Selection against Complex Mitotic-Origin Aneuploidy during Preimplantation Development', *PLOS Genetics*, 11(10), p. e1005601. Available at: https://doi.org/10.1371/JOURNAL.PGEN.1005601.

McCoy, R.C. (2017) 'Mosaicism in Preimplantation Human Embryos: When Chromosomal Abnormalities Are the Norm', *Trends in Genetics*, 33(7), pp. 448–463. Available at: https://doi.org/10.1016/J.TIG.2017.04.001.

McCoy, R.C. *et al.* (2018) 'Tripolar chromosome segregation drives the association between maternal genotype at variants spanning PLK4 and aneuploidy in human preimplantation embryos', *Human Molecular Genetics*, 27(14), pp. 2573–2585. Available at: https://doi.org/10.1093/HMG/DDY147.

McCoy, R.C. *et al.* (2023) 'Meiotic and mitotic aneuploidies drive arrest of in vitro fertilized human preimplantation embryos', *Genome Medicine*, 15(1), pp. 1–15. Available at: https://doi.org/10.1186/S13073-023-01231-1/FIGURES/5.

Middelkamp, S. *et al.* (2020) 'Sperm DNA damage causes genomic instability in early embryonic development', *Science Advances*, 6(16). Available at: https://doi.org/10.1126/SCIADV.AAZ7602/SUPPL_FILE/AAZ7602_TABLE_S1.XLSX.

Morales, C. *et al.* (2009) 'Reproductive consequences of genome-wide paternal uniparental disomy mosaicism: description of two cases with different mechanisms of origin and pregnancy outcomes', *Fertility and Sterility*, 92(1), pp. 393.e5-393.e9. Available at: https://doi.org/10.1016/J.FERTNSTERT.2009.03.090.

Munné, S. et al. (2020) 'First PGT-A using human in vivo blastocysts recovered by uterine lavage: comparison with matched IVF embryo controls', *Human Reproduction,* 35(1), pp. 70–80. Available at: https://doi.org/10.1093/HUMREP/DEZ242.

Nagaoka, S.I. *et al.* (2011) 'Oocyte-specific differences in cell-cycle control create an innate susceptibility to meiotic errors', *Current Biology*, 21(8), pp. 651–657. Available at: https://doi.org/10.1016/j.cub.2011.03.003.

Petropoulos, S. *et al.* (2016) 'Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos', *Cell*, 165(4), pp. 1012–1026. Available at: https://doi.org/10.1016/j.cell.2016.03.023.

Radwan, P. *et al.* (2022) 'The impact of soluble HLA-G in IVF/ICSI embryo culture medium on implantation success', *Frontiers in Immunology*, 13, p. 982518. Available at: https://doi.org/10.3389/FIMMU.2022.982518/BIBTEX.

Rechitsky, S. *et al.* (1998) 'Allele Dropout in Polar Bodies and Blastomeres', *Journal of Assisted Reproduction and Genetics*, 15(5), p. 253. Available at: https://doi.org/10.1023/A:1022532108472.

Reichmann, J. *et al.* (2018) 'Dual-spindle formation in zygotes keeps parental genomes apart in early mammalian embryos', *Science*, 361(6398), pp. 189–193. Available at: https://doi.org/10.1126/SCIENCE.AAR7462/SUPPL_FILE/AAR7462S9.MP4.

Del Rey, J. *et al.* (2018) 'Novel Double Factor PGT strategy analyzing blastocyst stage embryos in a single NGS procedure', *PLOS ONE*, 13(10), p. e0205692. Available at: https://doi.org/10.1371/JOURNAL.PONE.0205692.

Rubio, C. *et al.* (2007) 'Impact of chromosomal abnormalities on preimplantation embryo development', *Prenatal Diagnosis*, 27(8), pp. 748–756. Available at: https://doi.org/10.1002/PD.1773.

de Rycke, M. *et al.* (2020) 'PREIMPLANTATION GENETIC TESTING: Clinical experience of preimplantation genetic testing', *Reproduction*, 160(5), pp. A45–A58. Available at: https://doi.org/10.1530/REP-20-0082.

Sabina, J. and Leamon, J.H. (2015) 'Bias in Whole Genome Amplification: Causes and Considerations', *Methods in Molecular Biology*, 1347, pp. 15–41. Available at: https://doi.org/10.1007/978-1-4939-2990-0_2.

Schneider, I. *et al.* (2021) 'Dual spindles assemble in bovine zygotes despite the presence of paternal centrosomes', *Journal of Cell Biology*, 220(11). Available at: https://doi.org/10.1083/JCB.202010106/VIDEO-5.

Seckl, M.J., Sebire, N.J. and Berkowitz, R.S. (2010) 'Gestational trophoblastic disease', *Lancet (London, England)*, 376(9742), pp. 717–729. Available at: https://doi.org/10.1016/S0140-6736(10)60280-2.

Sha, Q.Q. *et al.* (2020) 'Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans', *Nature Communications 2020 11:1*, 11(1), pp. 1–16. Available at: https://doi.org/10.1038/s41467-020-18680-6.

Siermann, M. *et al.* (2024) 'Polygenic embryo screening: quo vadis?', *Journal of Assisted Reproduction and Genetics*, 41(7), pp. 1719–1726. Available at: https://doi.org/10.1007/S10815-024-03169-8/METRICS.

Singla, S. *et al.* (2020) 'Autophagy-mediated apoptosis eliminates aneuploid cells in a mouse model of chromosome mosaicism', *Nature Communications 2020 11:1*, 11(1), pp. 1–15. Available at: https://doi.org/10.1038/s41467-020-16796-3.

Sirard, M.A. (2012) 'Factors Affecting Oocyte and Embryo Transcriptomes', *Reproduction in Domestic Animals*, 47(SUPPL.4), pp. 148–155. Available at: https://doi.org/10.1111/J.1439-0531.2012.02069.X.

Steptoe, P.C. and Edwards, R.G. (1978) 'Birth after the reimplantation of a human embryo', *Lancet (London, England)*, 2(8085), p. 366. Available at: https://doi.org/10.1016/S0140-6736(78)92957-4.

Surani, M.A.H., Barton, S.C. and Norris, M.L. (1984) 'Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis', *Nature 1984 308:5959*, 308(5959), pp. 548–550. Available at: https://doi.org/10.1038/308548a0.

Swales, A.K.E. and Spears, N. (2005) 'Genomic imprinting and reproduction', *Reproduction*, 130(4), pp. 389–399. Available at: https://doi.org/10.1530/REP.1.00395.

Tadros, W. and Lipshitz, H.D. (2009) 'The maternal-to-zygotic transition: a play in two acts', *Development*, 136(18), pp. 3033–3042. Available at: https://doi.org/10.1242/DEV.033183.

Tan, V.J. *et al.* (2023) 'Third-Generation Single-Molecule Sequencing for Preimplantation Genetic Testing of Aneuploidy and Segmental Imbalances', *Clinical Chemistry*, 69(8), pp. 881–889. Available at: https://doi.org/10.1093/CLINCHEM/HVAD062.

Telenius, H. *et al.* (1992) 'Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer', *Genomics*, 13(3), pp. 718–725. Available at: https://doi.org/10.1016/0888-7543(92)90147-K.

Thurston, A. *et al.* (2008) 'Monoallelic expression of nine imprinted genes in the sheep embryo occurs after the blastocyst stage', *Reproduction*, 135(1), pp. 29–40. Available at: https://doi.org/10.1530/REP-07-0211.

Treff, N.R. *et al.* (2010) 'SNP microarray-based 24 chromosome aneuploidy screening is significantly more consistent than FISH', *Molecular human reproduction*, 16(8), pp. 583–589. Available at: https://doi.org/10.1093/MOLEHR/GAQ039.

Treff, N.R. *et al.* (2012) 'Development and validation of an accurate quantitative real-time polymerase chain reaction–based assay for human blastocyst comprehensive chromosomal aneuploidy screening', *Fertility and Sterility*, 97(4), pp. 819-824.e2. Available at: https://doi.org/10.1016/J.FERTNSTERT.2012.01.115.

Treff, N.R. *et al.* (2019) 'Validation of concurrent preimplantation genetic testing for polygenic and monogenic disorders, structural rearrangements, and whole and segmental chromosome aneuploidy with a single universal platform', *European Journal of Medical Genetics*, 62(8), p. 103647. Available at: https://doi.org/10.1016/J.EJMG.2019.04.004.

Tšuiko, O. *et al.* (2017) 'Genome stability of bovine in vivo-conceived cleavage-stage embryos is higher compared to in vitro-produced embryos', *Human Reproduction*, 32(11), pp. 2348–2357. Available at: https://doi.org/10.1093/HUMREP/DEX286.

Tšuiko, O. *et al.* (2021) 'Haplotyping-based preimplantation genetic testing reveals parent-of-origin specific mechanisms of aneuploidy formation', *NPJ genomic medicine*, 6(1). Available at: https://doi.org/10.1038/S41525-021-00246-0.

Uzbekov, R. *et al.* (2023) 'Centrosome Formation in the Bovine Early Embryo', *Cells*, 12(9), p. 1335. Available at: https://doi.org/10.3390/CELLS12091335/S1.

Vanneste, E. *et al.* (2009) 'Chromosome instability is common in human cleavage-stage embryos', *Nature Medicine 2009 15:5*, 15(5), pp. 577–583. Available at: https://doi.org/10.1038/nm.1924.

Vanneste, E. *et al.* (2011) 'PGD for a complex chromosomal rearrangement by array comparative genomic hybridization', *Human reproduction (Oxford, England)*, 26(4), pp. 941–949. Available at: https://doi.org/10.1093/HUMREP/DER004.

Vanneste, E. *et al.* (2012) 'Aneuploidy and copy number variation in early human development', *Seminars in reproductive medicine*, 30(4), pp. 302–308. Available at: https://doi.org/10.1055/S-0032-1313909.

Vázquez-Diez, C., Paim, L.M.G. and FitzHarris, G. (2019) 'Cell-Size-Independent Spindle Checkpoint Failure Underlies Chromosome Segregation Error in Mouse Embryos', *Current Biology*, 29(5), pp. 865-873.e3. Available at: https://doi.org/10.1016/j.cub.2018.12.042.

Vermeesch, J.R., Voet, T. and Devriendt, K. (2016) 'Prenatal and pre-implantation genetic diagnosis', *Nature Reviews Genetics 2016 17:10*, 17(10), pp. 643–656. Available at: https://doi.org/10.1038/nrg.2016.97.

Visscher, P.M. *et al.* (2017) '10 Years of GWAS Discovery: Biology, Function, and Translation', *The American Journal of Human Genetics*, 101(1), pp. 5–22. Available at: https://doi.org/10.1016/J.AJHG.2017.06.005.

Voet, T. *et al.* (2013) 'Single-cell paired-end genome sequencing reveals structural variation per cell cycle', *Nucleic acids research*, 41(12), pp. 6119–6138. Available at: https://doi.org/10.1093/NAR/GKT345.

Wei, Y. *et al.* (2011) 'Spindle Assembly Checkpoint Regulates Mitotic Cell Cycle Progression during Preimplantation Embryo Development', *PLOS ONE*, 6(6), p. e21557. Available at: https://doi.org/10.1371/JOURNAL.PONE.0021557.

Wells, D. *et al.* (2014) 'Clinical utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of aneuploidy in human embryos prior to implantation', *Journal of medical genetics*, 51(8), pp. 553–562. Available at: https://doi.org/10.1136/JMEDGENET-2014-102497.

Wen, L. and Tang, F. (2018) 'Single cell epigenome sequencing technologies', *Molecular Aspects of Medicine*, 59, pp. 62–69. Available at: https://doi.org/10.1016/J.MAM.2017.09.002.

Woolley, D.M. and Fawcett, D.W. (1973) 'The degeneration and disappearance of the centrioles during the development of the rat spermatozoon', *The Anatomical record*, 177(2), pp. 289–301. Available at: https://doi.org/10.1002/AR.1091770209.

Xu, X. *et al.* (2019) 'Observation of two separate bipolar spindles in the human zygote', *Journal of Assisted Reproduction and Genetics*, 36(4), pp. 601–602. Available at: https://doi.org/10.1007/S10815-019-01440-X/FIGURES/1.

Zamani Esteki, M. *et al.* (2015) 'Concurrent whole-genome haplotyping and copy-number profiling of single cells', *American journal of human genetics*, 96(6), pp. 894–912. Available at: https://doi.org/10.1016/J.AJHG.2015.04.011.

Zamani Esteki, M. *et al.* (2019) 'In vitro fertilization does not increase the incidence of de novo copy number alterations in fetal and placental lineages', *Nature Medicine 2019 25:11*, 25(11), pp. 1699–1705. Available at: https://doi.org/10.1038/s41591-019-0620-2.

Zhang, L. *et al.* (1992) 'Whole genome amplification from a single cell: implications for genetic analysis', *Proceedings of the National Academy of Sciences of the United States of America*, 89(13), pp. 5847–5851. Available at: https://doi.org/10.1073/PNAS.89.13.5847.

Zhang, X. *et al.* (2011) 'Successful pregnancy following the transfer of vitrified blastocyst which developed from poor quality embryos on day 3', *Iranian Journal of Reproductive Medicine*, 9(3), p. 203. Available at: /pmc/articles/PMC4575755/ (Accessed: 25 August 2024).

Zhou, X. *et al.* (2020) 'Comparison of Multiple Displacement Amplification (MDA) and Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) in Limited DNA Sequencing Based on Tube and Droplet', *Micromachines 2020, Vol. 11, Page 645*, 11(7), p. 645. Available at: https://doi.org/10.3390/MI11070645.

Zhu, P. *et al.* (2017) 'Single-cell DNA methylome sequencing of human preimplantation embryos', *Nature Genetics 2017 50:1*, 50(1), pp. 12–19. Available at: https://doi.org/10.1038/s41588-017-0007-6.

Zong, C. *et al.* (2012) 'Genome-wide detection of single-nucleotide and copy-number variations of a single human cell', *Science (New York, N.Y.)*, 338(6114), pp. 1622–1626. Available at: https://doi.org/10.1126/SCIENCE.1229164.

CHAPTER 2 – OBJECTIVES

This PhD study aims to develop novel methodologies for analyzing single blastomeres and to apply these methods to investigate the origins and developmental fates of various chromosomal abnormalities in preimplantation embryos. The study evolved into three main projects, each with specific objectives.

Human preimplantation embryos are known to be chromosomally unstable and a majority of embryos at 3 dpf harbor blastomeres with chromosomal abnormalities. To improve embryo selection, our laboratory developed a method enabling concurrent haplotyping and aneuploidy profiling in the embryo, coined haplarithmisis. Using this method, we scanned the incidence and spectrum of chromosomal anomalies in human and bovine embryos. However, whether similar chromosomal aberrations also occur in other mammals remained largely unknown. **In the first project**, we aimed to determine the full spectrum of chromosomal anomalies in equine embryos. To do this, we first investigated the feasibility of concurrent genome-wide PGT-A and PGT-M in horses, which we then utilized to identify chromosomal anomalies in equine embryos.

During our genomic screenings of human and bovine preimplantation embryos, we observed not only various WG abnormalities, such as haploidy and triploidy, but also instances of WG abnormal blastomeres coexisting with biparental diploid blastomeres or other WG abnormal blastomeres within the same embryo. In **the second project**, by investigating bovine preimplantation embryos, we aimed to: (1) confirm the co-occurrence of WG segregation errors with multi-polar first zygotic division and identify different types of WG separation errors; (2) explore the developmental potential of blastomeres with different chromosome constitution (androgenetic, gynogenetic, biparental, etc.) following heterogoneic division, with the hypothesis of selective elimination of blastomeres containing WG abnormalities during preimplantation development; and (3) investigate the developmental program of androgenetic, gynogenetic, and polyploid blastomeres post-heterogoneic division by comparing their developmental trajectories to those of blastomeres with a normal chromosomal constitution. To enable these studies, we combined GBS-based haplarithmisis with transcriptome profiling, creating a novel adaptation of G&T-seq, termed haplotype-based G&T-seq (hG&T-seq), which allows for simultaneous haplotyping and transcriptome profiling of the same single cell.

One major limitation of the above described PGT methodologies is the requirement for DNA from additional family members to deduce haplotypes. With the advent of lrWGS, we aimed to determine whether this limitation could be overcome, enabling direct phasing. Hence, **in the third project,** we explored the variant calling and phasing performance of lrWGS data from single cells and conducted a proof-of-concept study to demonstrate the clinical validity of

lrWGS-based concurrent haplotyping and aneuploidy profiling in human embryos.

CHAPTER 3 – GENOME-WIDE EQUINE PREIMPLANTATION GENETIC TESTING ENABLED BY SIMULTANEOUS HAPLOTYPING AND COPY NUMBER DETECTION

This chapter is based on the following paper:

De Coster, T. *, **Zhao, Y. ***, Tšuiko, O., Demyda-Peyrás, S., Van Soom, A., Vermeesch, J. R., & Smits, K. (2024). Genome-wide equine preimplantation genetic testing enabled by simultaneous haplotyping and copy number detection. ***Scientific Reports*** 2024 14:1, 14(1), 1–12. https://doi.org/10.1038/s41598-023-48103-7. (*Co-first authors)

Genome-wide equine preimplantation genetic testing enabled by simultaneous haplotyping and copy number detection

T De Coster[1,2*#], Y Zhao[2#], O Tšuiko[2], S Demyda-Peyrás[3,4], A Van Soom[1], JR Vermeesch[2] and K Smits[1*]

[1]Department of Internal Medicine, Reproduction and Population Medicine, Ghent University, Merelbeke, Belgium

[2]Department of Human Genetics, KU Leuven, Leuven, Belgium

[3]Department of Genetics, University of Cordoba, Spain

[4]Department of Animal Production, Veterinary School, National University of La Plata, La Plata, Argentina.

#These authors contributed equally

*Corresponding author: Katrien Smits; katrien.smits@ugent.be

Abstract

In different species, embryonic aneuploidies and genome-wide errors are a major cause of developmental failure. The increasing number of equine embryos being produced worldwide provides the opportunity to characterize and rank or select embryos based on their genetic profile prior to transfer. Here, we explored the possibility of generic, genome-wide preimplantation genetic testing concurrently for aneuploidies (PGT-A) and monogenic (PGT-M) traits and diseases in the horse, meanwhile assessing the incidence and spectrum of chromosomal and genome-wide errors in *in vitro*-produced equine embryos. To this end, over 70,000 single nucleotide polymorphism positions were genotyped in 14 trophectoderm biopsies and corresponding biopsied blastocysts, and in 26 individual blastomeres from six arrested cleavage-stage embryos. Subsequently, concurrent genome-wide copy number detection and haplotyping by haplarithmisis was performed and the presence of aneuploidies and genome-wide errors and the inherited parental haplotypes for four common disease-associated genes with high carrier frequency in different horse breeds (*GBE1, PLOD1, B3GALNT2, MUTYH*), and for one color coat-associated gene (*STX17*) were compared in biopsy-blastocyst combinations. The euploid (n = 12) or fully aneuploid (n = 2) state and the inherited parental haplotypes for 42/45 loci of interest of the biopsied blastocysts were predicted by the biopsy samples in all successfully analyzed biopsy-blastocyst combinations (n = 9). Two biopsies showed a loss of maternal chromosome 28 and 31, respectively, which were confirmed in the corresponding blastocysts. In one of those biopsies, additional complex

aneuploidies not present in the blastocyst were found. Five out of six arrested embryos contained chromosomal and/or genome-wide errors in most of their blastomeres, demonstrating their contribution to equine embryonic arrest *in vitro*. The application of the described PGT strategy would allow to select equine embryos devoid of genetic errors and pathogenetic variants, and with the variants of interest, which will improve foaling rate and horse quality. We believe this approach will be a gamechanger in horse breeding.

Introduction

While traditional breeding in the horse was based on natural mating, modern breeding strategies are influenced by the rapid development of medically assisted reproduction (MAR). The introduction of artificial insemination at the end of the 19th century allowed storage and distribution of male genetics [1]. Nowadays, also valuable female individuals can produce more than one foal per year through the collection of oocytes and the production of embryos, both *in vivo* and *in vitro*. The number of equine embryos being produced is increasing annually, with 37 094 embryos produced in 2021, an increase of 10% as compared to 2020 [2]. This increase is mainly caused by the rapid growth of equine *in vitro* embryo production (IVEP) and recent progress in embryo cryopreservation [2,3]. The popularity of IVEP results from the possibility to increase the genetic gain from high quality and/or subfertile mares and from economic benefits related to the full exploitation of expensive straws of semen from high quality stallions. Moreover, in contrast to flushed embryos, *in vitro*-produced embryos can be generated outside the reproductive season and are routinely cryopreserved and traded. As a consequence, modern breeding strategies gave rise to a new market in which equine embryos are sold based upon pedigree's phenotypes or performances. On the 4th of October 2022, a record price of € 124 000 euro was paid for an equine embryo with a promising pedigree [4]. However, it remains unknown if transfer of such an embryo will result in a healthy foal with the aspired phenotypical characteristics.

Embryonic and fetal loss remain one of the greatest challenges in equine breeding. Indeed, only ~18-26% of injected oocytes reach the transferrable blastocyst stage during IVEP [3,5]. Moreover, only 76-83% of the artificial inseminations with fresh semen [6,7] and ~85% of fresh *in vivo* and ~70% of frozen-thawed day 6-9 *in vitro*-produced embryo transfers [3,8–10] generate a clinical pregnancy. When a pregnancy is clinically detected after *in vivo* fertilization, 5-10% of day 14- and 5 to 10% of day 70-pregnancies additionally fail to produce a viable foal (reviewed elsewhere [11]). Following IVEP, a doubled incidence of early pregnancy losses [9], of which a higher rate seems to present as anembryonic vesicles [12,13] has been reported. Due to the seasonal breeding and the formation of endometrial cup cells, pregnancy losses in the horses are complicated, as those occurring after day 35 will usually leave the mare barren for that season, which has economic repercussions.

Similar to what has been reported in human and other species [14–19], chromosomal gains or losses, also called aneuploidies, are a major cause of equine pregnancy loss, occurring at a rate of 22% in early lost conceptuses [20]. In human and cattle, aneuploidy occurs frequently throughout early embryo development *in vivo* and *in vitro* [21–28], and contributes also to pregnancy loss before the clinical detection of pregnancy. The latter is known from transfers of aneuploid bovine and human blastocysts [25,29–31] and the analysis of *in vitro*-produced human embryos that arrest before reaching the blastocyst stage [32,33]. In addition, aneuploidy may present itself also on a genome-wide level, resulting in ploidy errors and IVEP may increase the incidence of genomic instability [22,23,24,26,30,31,27,34,61]. In horses, genomic instability throughout embryo development has only been assessed in few studies. Two studies showed the presence of chromosome-containing micronuclei in cleavage- and blastocyst- stage *in vitro*-produced embryos, suggesting chromosome segregation errors [35,36]. Another one described embryonic aneuploidies on chromosome two and four only in both *in vitro*- and *in vivo*- produced morulas and blastocysts and suggested that IVEP increases the likelihood of aneuploidy, although data were not significant [37]. Adding to this suggestion, a study investigating the effect of *in vitro* maturation (IVM) on the rate of aneuploidy by immunostaining found that equine *in vitro*-matured oocytes were significantly more affected by aneuploidy than *in vivo* matured oocytes (45.5% vs 0%, respectively) [38]. In all equine studies thus far, genome-wide errors have not been recovered due to technical limitations and embryos were sacrificed for the analysis and thus, not transferred. By consequence, the incidence and nature of aneuploidies and genome-wide errors throughout early embryonic development, and the effect on early embryo viability remains undetermined in the horse.

Besides aneuploidies or genome-wide errors, also the inherited alleles determine the viability of the conceptus, and additionally blueprint the phenotype of the foal. Domestication and centuries of horse breeding, selecting (related) mares and stallions with the desired traits, have led to the evolution of modern horses. These traits used to include phenotypical characteristics facilitating transportation, farming, and warfare purposes, but breeding goals have shifted since the industrialization towards traits related to the exterior of the specific breed standards and sports performances. Intensive human selection during the last 250 years has resulted in homogeneity within and a substantial variation among different horse breeds [39]. This resulted in the accumulation of both alleles for the desired traits and deleterious mutations in today's horse genome [40,41]. As a consequence, several hereditary diseases have come up alongside desirable traits characterizing these horse breeds and some of them additionally contribute to pregnancy loss or lethality at or close to birth. One of those, occurring with a carrier frequency of 7 and 8% in quarter and paint horses, respectively [42], is the glycogen branching enzyme deficiency (GBED). This autosomal recessive disease is caused by a mutation in the glycogen branching enzyme 1 (*GBE1)* gene that results in non-functional glycogen storage [43,44]. It is

responsible for 3% of spontaneous abortions, and cardiac or respiratory failure, seizures, muscle weakness and death of foals within 18 weeks of age following homozygous inheritance [45]. Another one, called warmblood fragile foal syndrome (WFFS) is caused by a mutation in the procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 *(PLOD1)* gene, and is important to warmblood horses in which it occurs with a frequency of up to 17%, depending on the specific breed [46,47]. The *PLOD1* mutation causes deviant collagen formation, and homozygous inheritance presents as severe skin lesions in foals, which require euthanasia just after birth [48]. In turn, 5-9% of Arabian horses are burdened by the occurrence of cerebellar abiotrophy (CA), a neurodegenerative, autosomal recessive disease caused by a mutation of mutY DNA glycosylase *(MUTYH)* gene that results in the loss of Purkinje neurons and causes ataxia in foals [49–51]. Also in Friesian horses, 13.3-17.3% are carrying a mutation of the beta-1,3-N-acetylgalactosaminyltransferase 2 (*B3GALNT2)* gene, responsible for the autosomal recessive congenital hydrocephalus (CH), which results in stillbirth, dystocia or postnatal euthanasia [52]. In closed studbooks, carriers of known recessive diseases are excluded from breeding, giving rise to a further increase of inbreeding in these small populations.

The production of equine embryos provides the opportunity for characterization of embryos by preimplantation genetic testing (PGT), which involves the sampling of embryonic material for genetic profiling prior to embryo transfer. Application of PGT for aneuploidies (PGT-A) would allow to prioritize chromosomally normal embryos for transfer, potentially increasing pregnancy rates and reducing abortion rates. On the other hand, PGT for monogenic (PGT-M), or polygenic (PGT-P) traits and diseases and structural chromosomal errors (PGT-SR) would provide the possibility to select only healthy embryos with desirable traits for transfer. While testing of the mating stallion and mare for monogenic traits and diseases can currently be used for informing mating choices when breeding with animals carrying recessive alleles, allowing for non-affected offspring, implementation of PGT-M would allow to include animals carrying both recessive and dominant alleles in breeding programs and , prevent the birth of both carrier and affected foals by selective transfer of embryos to recipient mares, thus avoiding inbreeding while eliminating unwanted alleles from the population. PGT-P has the further ability to increase the accuracy of selection for desired polygenic traits and diseases and to reduce the generation interval, improving selective breeding in horses. In general, applying PGT would also increase our understanding of the origin of aneuploidy and genome-wide errors in horses and the contribution of aneuploidy, genome-wide errors and mutations to pregnancy loss before and after clinical pregnancy detection. However, PGT for horses remains underdeveloped as compared to humans or cattle with no publications on PGT-A, and few studies on PGT-M, limited to the interrogation of a maximum of 33 loci related to diseases or traits, including sex, ID markers, coat color, based on (multiplex) targeted PCR [3,53–56]. A generic method allowing for concurrent, genome-wide PGT-A, PGT-M and PGT-SR with the potential

for PGT-P [57–59], as successfully applied for humans [24,31,60] and cattle [26,30], would increase the potential of PGT for both clinical and research-based equine applications.

Here, we explored the possibility of concurrent, generic, genome-wide PGT-A and PGT-M in horses using the 'single-cell haplotyping and imputation of linked disease variants' (siCHILD)/haplarithmisis technique previously applied for humans and cattle [24,57,61,62]. Simultaneously, we assessed the incidence and spectrum of chromosomal and genome-wide aberrations in equine preimplantation development *in vitro* by analysis of arrested and blastocyst-stage embryos.

Results

Following IVEP, fourteen trophectoderm (TE) biopsies and corresponding biopsied whole blastocysts and 26 individual blastomeres from six arrested cleavage-stage embryos were analyzed for their genome-wide copy number and haplotype. All sample characteristics, results of the genetic analysis, and genome-wide haplarithm plots obtained from this study are provided in C3-Additional file 1, Table 1 and Figure. 1 (TE biopsies and corresponding biopsied blastocysts) and C3-Additional file 2, Table 1 and Figure. 1 (blastomeres from arrested embryos). The number of informative single nucleotide polymorphisms (SNPs) was around 20,000 for samples for which parental genotypes were phased with sibling embryos and 7229 for samples of Mare03, for which parental genotypes were phased with paternal grandparents. Five biopsies and seven single blastomeres failed to give a result. The average coverage rate of successfully analyzed biopsied blastocysts, biopsies and single blastomeres was 70.54, 55.83 and 45.19%, respectively (C3-Additional file 1, Table 1 and C3-Additional file 2, Table 1). The average rate of Mendelian inconsistency of the called SNP genotypes in successfully analyzed blastocysts, biopsies and single blastomeres was 0.91, 3.14 and 18.94%, respectively. Nine and five of the blastocyst-stage embryos and one and five of the arrested embryos were found to be male and female, respectively, which was confirmed by qPCR.

Trophectoderm biopsies and corresponding biopsied whole blastocysts

PGT-A

As a proof of principle for PGT-A, the presence of aneuploidies and genome-wide errors was compared in biopsy-blastocyst combinations. The euploid or fully aneuploid state of the biopsied blastocysts was predicted by the analysis of the corresponding TE biopsy samples in all nine successfully analyzed biopsy-blastocyst combinations. Two out of 14 analyzed biopsied blastocysts were aneuploid. Among them, Mare01_ Embryo01 contained a maternal chromosomal loss of chromosome 31 and Mare02_Embryo07 contained a maternal chromosomal loss of chromosome 28 **(Figure 1)**. The presence of a maternal chromosomal loss in both the biopsies and the biopsied blastocysts indicates the presence of a chromosomal error in all cells (full aneuploidy) and by consequence, the occurrence of a chromosomal error

of meiotic origin. For Mare02_Embryo07, additional aneuploidies, not detected in the blastocyst were found in the TE biopsy. These included maternal losses of chromosomes 1, 5, 14, 20 and 28 and paternal losses of chromosomes 7, 12 and 25. Additionally, paternal trisomy was noted on chromosome 4 and maternal trisomy on chromosome 17 and 23. Furthermore, a partial nullisomy was found in chromosome 11. Because of the complex aneuploidy, the collection of multiple cells and fluctuations of the LogR and parental BAF plots, deviating from values expected for full chromosome aberrations in one cell (i.e., LogR should be at -1, 0.58 and 1 for a monosomy, trisomy or tetrasomy, respectively (LogR = Log2(rate detected/rate expected), and parental BAF values should be at 0, and 0 and 1 for a monosomy; 0 and 0.33, and 1 and 0.67 for a trisomy; and 0 and 0.25, and 1 and 0.75 for a tetrasomy (BAF = number of B-alleles/number of total alleles), the described and possibly additional aberrations seemed to be present in a mosaic state (i.e., differently in different cells) in the biopsy sample. The presence of additional aneuploidies in a putative mosaic state in the biopsy sample indicate the presence of distinct chromosomal errors in the blastocyst too. However, as these were not detected, they were likely present only in a small fraction of cells (low grade mosaic aneuploidy), resulting from chromosomal segregation errors during the late embryonic divisions.

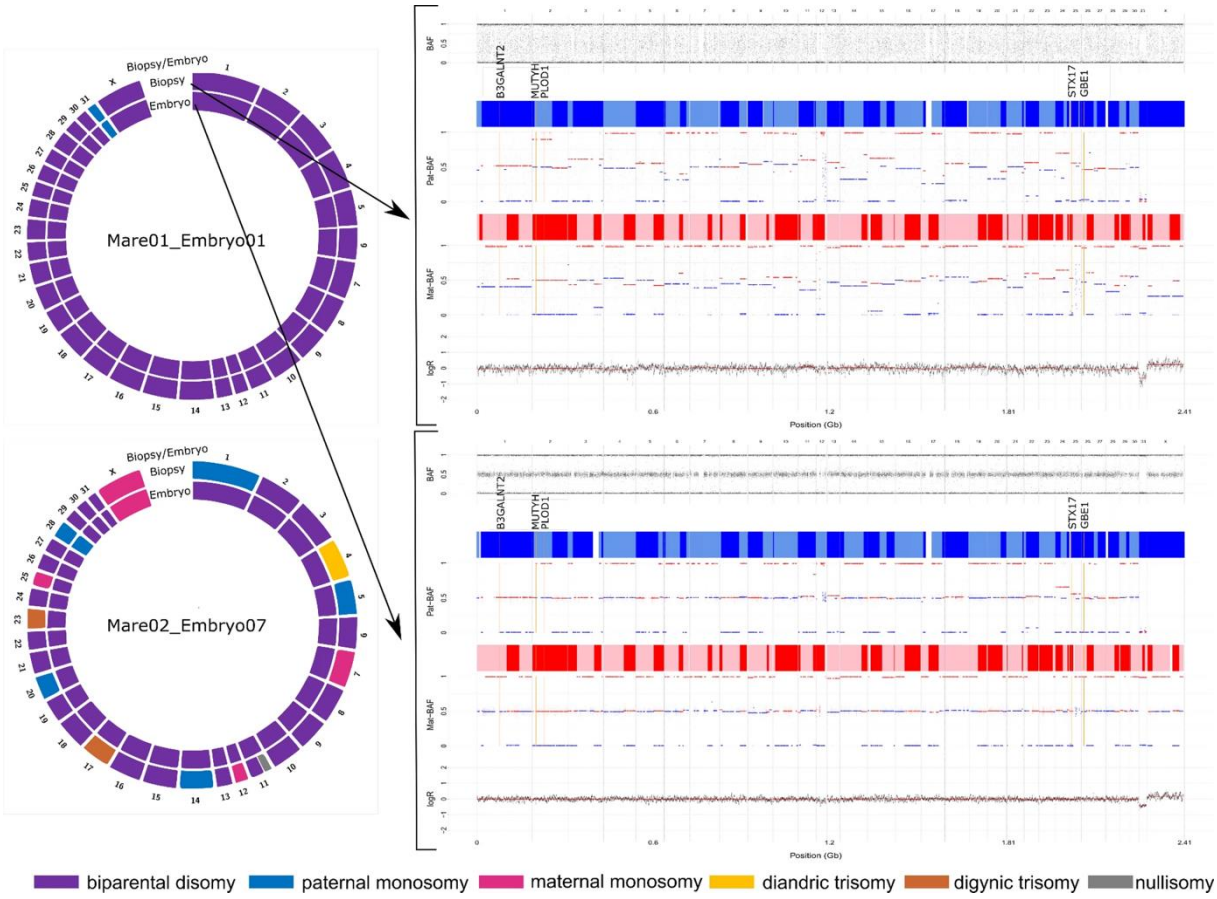


Figure 1. PGT analysis of blastocyst-stage equine embryos. On the left side, the circos plots are depicted for two out of 14 blastocyst-stage equine embryos that were aneuploid. The outer and inner circles represent the genome constitution per chromosome (1-X) of the biopsy sample and the corresponding biopsied blastocyst sample, respectively. Different colors represent different types of aneuploidies. On the right, the parental haplarithm plots of the biopsy (upper plot) and the corresponding embryo sample (lower plot) of Mare01_Embryo01 are depicted containing the genome-wide LogR value (measure for the chromosome copy number), the maternal B-allele-frequency (Mat-BAF), the maternally inherited haplotype blocks (pink/red colors represent regions inherited from maternal homologue 1 or 2), the paternal B-allele-frequency (Pat-BAF) and the paternally inherited haplotype blocks (light/dark blue colors represent regions inherited from maternal homologue 1 or 2) from the bottom to the top, respectively. The inherited parental haplotypes of four disease-causing genes (*B3GALNT2* for congenital hydrocephalus, *MUTYH* for cerebellar abiotrophy, *PLOD1* for warmblood fragile foal syndrome, *GBE1* for glycogen branching enzyme deficiency) and the gene causative for grey color (*STX17*) coat are indicated by the yellow line. Genome-wide haplarithm plots and detailed chromosome-wise haplarithm plots of the chromosomes carrying a gene of interest of all samples can be retrieved in C3-Additional file 1, Figs. 1 and 2, respectively.

For the TE biopsy of Mare2_Embryo5_biopt, no clear evidence of a full chromosomal loss exists. Yet, the LogR and the distance between the parental BAF values fluctuate between 0 and -1 and deviate from 0.5 respectively, which can be caused by the sampling of mosaic cells.

However, also the plot quality could be affected by the lower coverage of SNPs (50.90%) compared to the average SNP coverage of TE biopsy samples (55.83%).

PGT-M

As a proof of principle for PGT-M, we compared the inherited maternal and paternal haplotypes for four common disease-associated genes with an autosomal recessive inheritance pattern and a high carrier frequency in different horse breeds as described in the introduction, and for one phenotypical trait with an autosomal dominant inheritance pattern in all successfully analyzed biopsy-blastocyst combinations. These included the *B3GALNT2* gene on chromosome 1, mutated in case of CH, the *PLOD1* and the *MUTYH* genes on chromosome 2, mutated in case of WFFS and CA, respectively, the *GBE1* gene on chromosome 26, mutated in case of GBED and the syntaxin 17(*STX17)* gene on chromosome 25, causative for grey color coat. An example for Mare01_ Embryo01 can be found in **Figure 1**. Detailed haplarithm plots of the chromosomes containing the genes of interest in all analyzed samples can be found in C3-Additional file 1, Figure. 2. The parental haplotypes were inconclusive for four out of 90 analyzed loci, all of which involved samples of Mare02_Embryo07. The first two included the paternal haplotype of *MUTYH* in both the biopsy and embryo sample due to the presence of a recombination site and absence of informative SNPs at the gene location. The other two included the maternal haplotype of *B3GALNT2* and the paternal haplotype of *STX17* in the biopsy sample due to the loss of the maternal homologue and paternal homologue, respectively. At the other loci, the parental haplotype in the biopsy sample was confirmed by the analysis of the corresponding biopsied embryo.

Arrested cleavage-stage embryos

From six arrested embryos, only Mare02_ Embryo02 contained a normal diploid profile in all four blastomeres. All other five contained chromosomal and/or genome-wide errors in most of their blastomeres **(Figure 2)**. A minority of chromosomal losses were only seen in the LogR plot, but not confirmed by the parental haplarithm plots. For these errors, depicted with a green color in **Figure 2**, the parental origin remained undetermined.

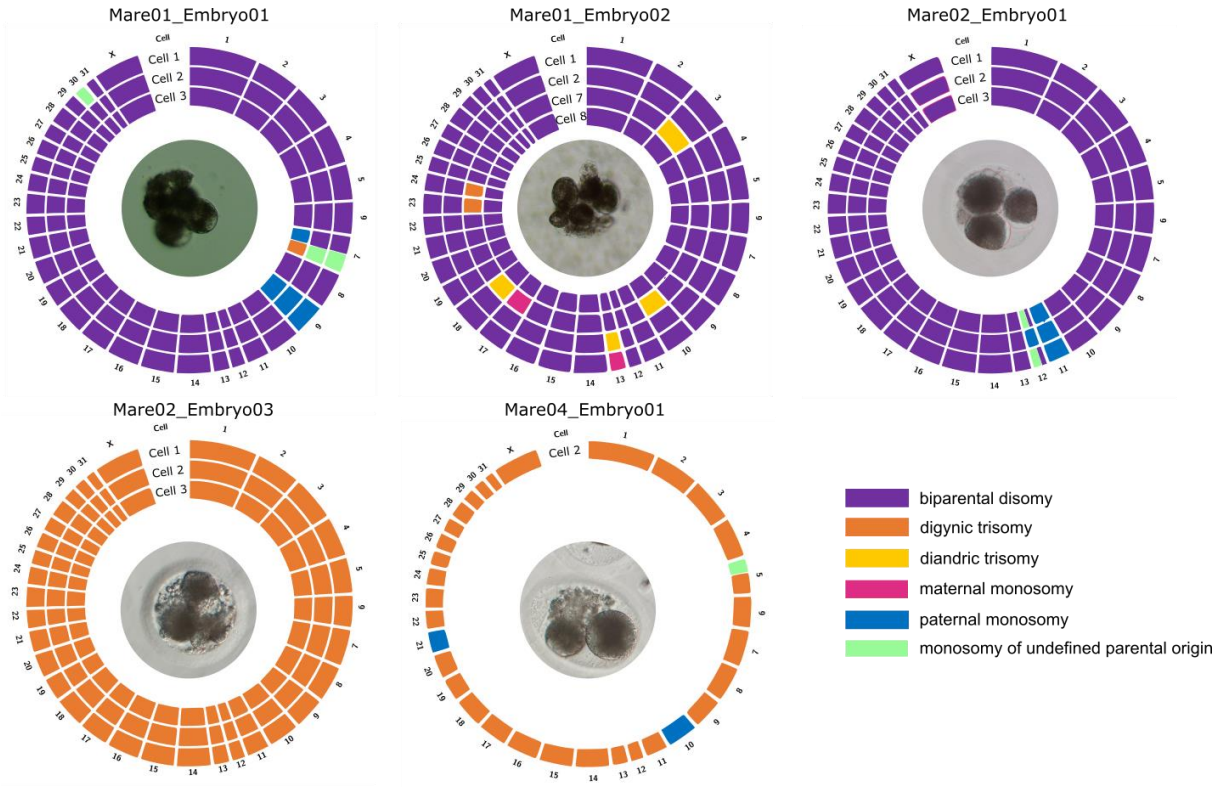


Figure 2. Circos plots of the five arrested equine embryos that contained chromosomal or genome-wide errors, with each circle representing the genome constitution per chromosome (1-X) of a single blastomere. Different colors represent different types of aneuploidies. Signatures of digynic trisomy across (almost all) chromosomes in Mare02_Embryo03 and Mare04_Embryo01 indicate digynic triploidy. Chromosomal errors of undefined parental origin (green) represent those that could be seen in the LogR plot, but were not confirmed by the parental haplarithm plots. Mare02_ Embryo02 was the only embryo that contained a normal, biparental diploid profile in all blastomeres and is not depicted. Lost blastomeres or blastomeres for which genetic analysis failed are not included, also blastomere 5 of Mare01_Embryo02 was not included due to low genome coverage and complex aneuploid profile. Haplarithm plots of all blastomeres can be retrieved in C3-Additional file 2, Figure. 1. In the middle of each circus plot, a picture of each embryo prior to dissociation is shown.

For Mare01_Embryo01, one blastomere was lost during the collection. The four remaining blastomeres all contained a maternal loss of chromosome nine, which thus most likely resulted from a meiotic error. In addition, a variety of other chromosomal errors was retrieved in the distinct blastomeres, pointing to errors of mitotic origin. Chromosome seven was affected differently in all blastomeres. In blastomere three, a partial chromosomal loss of the first half and a partial chromosomal gain of the second half of maternal chromosome 7 was found. Reciprocal to what was seen in blastomere three, blastomeres one and two presented with a partial, chromosomal loss of unknown parental origin of the last half of chromosome 7. Blastomere one additionally presented with a loss of chromosome 30 of unknown parental origin. For blastomere five, a lower number of SNPs were analyzed (coverage of 35.80%) and by consequence, the quality of the haplarithm plot was poorer. Nonetheless, a complex

aneuploid profile was observed, with (partial) chromosomal losses and gains of both paternal and paternal origin including at least chromosomes 7, 9, 24, 27, 28 and 30.

For Mare01_ Embryo02, one blastomere was lost during collection and three other blastomeres failed analysis. From the remaining four blastomeres, blastomere eight was normal diploid and three other blastomeres were found with distinct chromosomal errors. By consequence, these are of mitotic origin. Blastomere one and two contained a paternal chromosomal loss and the reciprocal gain on chromosome 13, respectively. Blastomere seven showed a complex aneuploid profile with paternal chromosomal gains on chromosome 3 ,10 and 18, a paternal chromosomal loss on chromosome 17 and maternal chromosomal gains on chromosomes 23 and 24.

In Mare02_ Embryo01, all three blastomeres presented with a maternal loss of chromosome 11, pointing to a meiotic origin. Maternal chromosome 12 was lost in blastomere two and lost partially in blastomeres one and three. For the last two blastomeres, a loss of the second half of chromosome 12 of unknown parental origin was found.

In Mare02_ Embryo03, three blastomeres failed analysis and three blastomeres showed an identical triploid profile with an additional maternal genome copy. Genome-wide regions of heterodisomy point out the presence of two distinct maternal genomes which must have resulted from a genome-wide meiotic error.

Finally, Mare04_ Embryo01, contained one blastomere which failed analysis and one blastomere with a digynic triploid profile with only one maternal genotype, pointing to a genome-wide mitotic error. Additionally, two out of three maternal copies were lost on chromosomes 10 and 21 and a partial chromosomal loss of chromosome five of unknown parental origin occurred.

Discussion

Here, we provide proof of concept for concurrent genome-wide copy number profiling and haplotyping in the horse enabling combined PGT-A and PGT-M in a generic way. Like in human and cattle, a similar analysis could additionally enable PGT-SR and PGT-P in the horse. Research-based application of this strategy has the potential to determine the contribution of genetic aberrations to embryonic or fetal arrest before or after clinical detection of pregnancy. Clinically, this method allows the selection of embryos devoid of both aneuploidies and pathogenic variants, and with the genetic variants of interest. We envision this application will revolutionize horse breeding.

The simultaneous detection of the genome-wide copy number and the inherited parental haplotypes allows to determine the presence, the type (gain/loss) and the origin (maternal/paternal and meiotic/mitotic) of (segmental) aneuploidies and genome-wide errors, also called ‘ploidy’ aberrations[24,57,61]. Ploidy aberrations include triploidy and haploidy, in which

case a whole parental genome of maternal or paternal origin is additionally present (digynic and diandric triploidy, respectively) or missing (androgenetic or gynogenetic haploidy, respectively), or androgenetic/gynogenetic genome-wide uniparental disomy when two genomes derived from the same parent are present. This method enabled to chart for the first time the full extent and spectrum of chromosomal and genome-wide aberrations in arrested cleavage-stage and transferrable blastocyst-stage *in vitro*-produced equine embryos. As such, we could provide the first solid evidence of the occurrence of genome-wide errors in equine embryos. These included two digynic triploidies, one of meiotic and one of mitotic origin. Considering the almost exclusive application of intracytoplasmic sperm injection (ICSI) for *in vitro* fertilization (IVF) in horses, the occurrence of predominantly maternal genome-wide errors can be expected as diandry via dispermic fertilization is prevented. Furthermore, a staggering five out of six (83%) arrested cleavage-stage embryos were found presenting with chromosomal and genome-wide errors. Considering this is a bit lower compared to the frequencies of human arrested cleavage-stage embryos (93-94% in [32,33]), this could represent a realistic estimate. At the blastocyst-stage, errors were detected in two out of 14 samples (14%) and suspected in a third. This is lower as compared to the incidence reported in biopsies from human blastocyst-stage embryos (25-69% in [24,28,29,32]) and cattle (14-31% in [26,30]). A higher incidence was also reported in the one study on *in vitro*-produced equine samples based on cytogenetic evaluation (40% in blastocysts [37]). This discrepancy might be related to either sample size or species-specific differences in gene content, effect of aneuploidies on embryonic viability and/or IVEP conditions. The high incidence of chromosomal and genome-wide errors in arrested equine cleavage-stage embryos as compared to transferrable blastocyst-stage embryos supports the view that these errors are a cause of *in vitro* embryo arrest in mammals.

With the exception of one, all arrested cleavage-stage embryos were characterized by (complex) aneuploidy and/or digynic triploidy of mitotic origin affecting all or most of their blastomeres, sometimes in combination with meiotic aneuploidy of one of the larger chromosomes or, meiotic triploidy, affecting all cells per definition. In contrast, the two aneuploid blastocysts contained errors of meiotic origin on smaller chromosomes, and one of them additionally contained mitotic aneuploidies in a small proportion of the embryo, as they were detected in the TE biopsy only. These findings corroborate the human observation that especially genomic aberrations that affect a large part of the genome in a large proportion of the embryo, such as (complex) aneuploidy resulting from mitotic errors at the earliest embryonic divisions, combinations of meiotic and mitotic aneuploidies, or ploidy aberrations seem detrimental to early embryonic development while meiotic errors or low-grade mosaicism do not necessarily prohibit embryo development to the blastocyst stage [24,28,32,33,67,68]. In horses, meiotic monosomies of smaller autosomal chromosomes have been reported in aborted

equine specimens (e.g., chromosome 26, 27 and 31), but never in live born horses [11,20]. Here, we demonstrate the presence of meiotic monosomies of larger chromosomes (e.g., chromosome 9 in Mare01_Embryo01 and chromosome 11 in Mare02_Embryo03) in arrested embryos but not in blastocysts. Together, these data point out that monosomies of the larger chromosomes affecting all embryonic cells cause gestational loss before clinical pregnancy detection[11], which can be due to gene-dosage imbalances resulting in embryo arrest before embryonic genome activation. In general, it should be taken into consideration that only a limited number of embryos from a limited number of mare-stallion combinations were analyzed here and therefore, analysis of more samples will be required to draw valid and generalizable conclusions.

When successful, the analysis of the TE biopsy could predict the euploid or full aneuploid state of the embryo, highlighting its potential for PGT-A. Monosomies on small chromosomes, here detected in two embryos on chromosome 28 and 31, are associated with fetal demise by 65 days of gestation [11]. Therefore, clinical application and selection against such embryos could increase success rates per embryo transfer as well as the live-birth rates, as was shown for *in vitro*-produced cow embryos [30]. Yet, such application remains controversial. In human for example, over 70 retrospective single and multi-center studies attest to the efficacy of PGT-A [69], but a lack of evidence for its beneficial effect in randomized-controlled multicenter studies trials remains, except in older age groups [70]. The risk of PGT-A being performed on a few cells obtained by a biopsy procedure, is that the analysis result does not necessarily reflect the status of the whole embryo in case of mosaic aneuploidy, the latter of which is not necessarily incompatible with healthy live birth [71–73]. In the study presented here, additional aneuploidies detected in the TE biopsy of one embryo, but not in the corresponding biopsied blastocyst, demonstrated the presence of putative low-grade mosaicism. Although mosaicism was indicated by the genetic profile of the TE profile of that and one other TE profile, the presence and extent cannot be accurately predicted by a TE biopsy alone. By consequence, whilst fully aneuploid embryos are unviable or severely diseased, the application of PGT-A may lead to misdiagnosis of mosaic embryos [74] and additionally, to damage to the embryo due to the biopsy procedure, leading to wastage of viable embryos. In regard of the biopsy procedure, both the needle aspiration biopsy of *in vitro*- and *in vivo*-generated embryos and the collection of herniating cells from *in vitro*-generated embryos have shown similar early pregnancy results as compared to non-biopsied controls [3,65].

In contrast to cattle but similar to humans, horse embryos have a high individual value, from both an emotional and economic perspective. For that reason, the clinical application of PGT-A in horses is expected to result rather in the selection against embryos with a meiotic trisomy (which are unviable and can be discriminated by haplarithmisis) and the further ranking of

euploid embryos and embryos presenting with other chromosomal errors, genome-wide errors or traces of mosaicism based on the type and number of chromosomes affected than in the selection against all abnormal embryos, as performed in cows [30]. Such ranking could impact both pregnancy rates and embryo prizes. The clinical application of PGT-A may particularly benefit the subpopulation of slowly developing *in vitro*-produced equine embryos. These are currently discarded by some centers since transfer results in decreased pregnancy and foaling rates [3,36,75–77]. Yet, since female embryos develop slower [75], such practice may also underlie an increased disposal of desired female embryos resulting in the observed skewed sex ratio following transfer of equine *in vitro*-produced embryos [8,75]. PGT-A by haplarithmisis has the potential to select against fully aneuploid embryos while salvaging the euploid and potential or presumed mosaic embryos, which might increase embryo viability and (partly) reinstate a more balanced sex ratio within this cohort. A subset of TE biopsies analyzed here, failed to give a result following haplarithmisis, which may be due to failed tubing or low fidelity of the whole genome amplification (WGA) process due to inhibition [65]. The same samples also failed analysis when analyzed by qPCR to verify the sex of the embryo, which may thus provide a method for quality assessment of the TE biopsy. The chance of a successful analysis could be increased by dividing the initial TE biopsy of 10 to 20 cells into two samples, generating a spare sample, and the upfront selection of the best sample for analysis by haplarithmisis.

We also demonstrate the ability to detect the inherited parental haplotypes at the location of five genes related to traits and diseases based on a biopsy sample, enabling PGT-M when real-case pedigree information regarding the inheritance of the allele of interest is available. Haplotyping-based PGT-M has a number of advantages as compared to other types of PGT-M. First, it minimizes misdiagnosis due to upfront WGA, which is a necessary evil to obtain sufficient input DNA when genotyping multiple targets or performing multiple analysis on one or a few cells, but results in erroneous and biased amplification. PCR-based PGT-M methods rely on the direct genotyping of the allele of interest and therefore, they are at risk to pick up the bias created by the WGA process. Especially the misdiagnosis of heterozygous embryos in horses [53–55] and humans [78,79] occurs frequently in PCR-based PGT-M, because of allele drop-out, resulting from the failure WGA process to amplify one of both alleles, to which especially at GC-rich locations, like the *GBE1* locus, are prone. PGT-M by haplotyping counters this problem, because the presence of the alleles of interest is not determined directly but instead, inferred from parental haplotypes determined from polymorphic markers flanking the loci of interest, most of which will not undergo allele drop-out. Second, the analysis is generic and performed on a genome-wide scale, making the simultaneous screening for many Mendelian or polygenic disorders and traits possible. This is in contrast to PCR-based methods, which could only detect a limited number of alleles of interest and had to be designed per family and/or location or of interest. Whilst this study looked for the inherited parental

haplotypes limited to five loci related to color coat and diseases, haplotype-based PGT-M can in the short term be applied for each pedigree to determine the presence of all alleles for which the causative mutation is currently known. In horses, Mendelian inheritance has been detected for 62 equine traits and diseases, but only for 49, the (likely) causal variants is known (https://www.omia.org/home/). Further research will continue to identify the underlying genetic mutation of other Mendelian disorders and traits.

Concurrent genome-wide copy number detection and haplotyping of embryos requires pedigree genotypes to enable phasing of the genotypes into haplotypes. These include genotypes from both parents, together with a sibling (affected or unaffected for the monogenic disease or trait) or at least one of the grandparents (from the parental side(s) at risk of transmitting the allele of interest in case of PGT-M). For PGT-M, pedigree information regarding the inheritance pattern of the allele of interest is additionally required. These requirements bring along opportunities and hurdles. On the one hand, clinical application of haplotype-based PGT would enable the development of large genomic databases. On a longer term, these databases may lead to the annotation of genes for multifactorial/polygenic traits or disorders and additionally enable future ranking of embryos for traits based on polygenic risk analysis. While the initiation of human embryo selection based on PGT-P poses several ethical and technical problems [80–84], the development of PGT-P in horses can be valuable for assessing important equine polygenic disorders like osteochondrosis dissecans [85–87] or insect bite hypersensitivity [88] and traits like performance [89,90], parasite susceptibility [91] or fertility [92,93], for which genome-wide association studies in horses are increasing the knowledge on the heritability and associated SNPs. On the other hand, obtaining materials from family members is straightforward in humans but poses a hurdle for the application of haplotyping-based PGT in the horse industry. While the DNA of the mare can be obtained via a blood sample during ovum-pick up or embryo transfer, DNA collection from the stallion is less straightforward as the sale of stallions' semen is exploited by external stud farms which send expensive small doses of semen for which DNA extraction is inherently more challenging as compared to blood. As such, the enrollment of equine PGT by the method presented here in the current setting would require the collaboration of stud farms. The necessity of grandparental or sibling genotype information from live horses can be avoided by genotyping an arrested or blastocyst-stage sibling embryo from the same mare-stallion combination, as we demonstrated here too. On the longer term, the development of a large equine genotype database would enable to reuse genotype information obtained from previous PGT sessions, especially because of the popular use of certain stallions in the equine breeding programs. Such database would also allow the increased use of parental siblings for phasing of the parental haplotypes [94].

The resolution of haplotyping via haplarithmisis depends on the number and distribution of total analysed and fraction of informative SNPs. Here, a medium density, 70 K SNP array was applied, analyzing only a limited number of SNPs. As a consequence, some of the chromosomal copy number errors shown by the LogR plot were not confirmed by the parental haplarithm plots, resulting in the failure to determine the parental origin. Furthermore, the paternal haplotype at one locus of interest in one biopsy-blastocyst combination could not be conclusively determined due to the co-occurrence of a recombination site and lack of informative SNPs. In human haplarithmisis-based PGT, parameters based on call rate and Mendelian inconsistency are applied for general quality assessment of the samples and for PGT-M specifically, the distance to the flanking homologous recombination site and number and accuracy of informative SNPs within this area are applied for conclusive determination of the parental haplotypes at loci of interest [95]. The determination of the resolution and threshold values for equine haplotype calling requires further research, including genomic samples with known haplotypes. Applying a higher density SNP array would result in more data points for subsequent genome-wide haplotyping and thus higher resolution PGT-A and PGT-M. However, for horses, only a 671K SNP array for concurrent analysis of 96 samples is commercially available (Axiom MNEC670; Thermo Fisher Scientific, Waltham, MA, USA [96]), which is not cost-efficient for application of PGT in horses. Alternatively, the rapid development and increased cost-efficiency of sequencing-based technologies may provide a solution to increase the number of data points. Genotyping-by-sequencing, for example, scalable in terms of both data points and sample throughput, can provide a cost-efficient alternative for arrays to provide input genotypes with an increased resolution for haplarithmisis, as demonstrated for human and bovine samples [62].

The number of equine embryos being produced and traded worldwide is expected to grow further, especially since the recent development of repeatable IVF in horses [97]. Similar to cattle and humans, we postulate that the selection and ranking of these embryos by a reliable method for PGT as demonstrated here, will enhance and predict embryo transfer success rates, and will enable selection for traits and against diseases, accelerating breeding programs. In cattle, the use of SNP-arrays for genomic selection of liveborn and embryos via a biopsy has revolutionized breeding. The integration of genomic selection has doubled the rate of genetic progress for important economic traits and increased selection accuracy, while decreasing the generation interval and avoiding undesirable recessive conditions [98,99]. While we generated new insights in the incidence and nature of chromosomal and genome-wide errors in equine *in vitro*-produced embryos until the blastocyst stage and their contribution to developmental arrest, further clinical application of PGT will determine the contribution of chromosomal errors, genome-wide errors and mutations to pregnancy failure before or after clinical pregnancy detection following transfer of those blastocysts. Furthermore, since horses provide the

opportunity to study embryos from a population that is less biased towards age and infertility as compared to human, and because *in vivo*-generated blastocysts are frequently flushed out of the uterus and transferred to a recipient mare, they provide an excellent model to determine the contribution of chromosomal an genome-wide errors to pregnancy loss before or after the clinical detection of pregnancy, and to study the *in vivo* and the *in vitro* situation (both IVF and ICSI) and assess and optimize *in vitro* culture conditions with regard to genome instability.

Material and methods

Equine in vitro embryo production

Embryos were produced *in vitro* as described previously [63,64], separating oocytes and embryos per donor. Briefly, oocytes retrieved from slaughterhouse-derived ovaries from four mares of unknown identity were kept for ~15 to 18 h in holding medium (Emcare, Manhattan, KS, USA) at room temperature (22°C) prior to IVM. Maturation of cumulus-oocyte-complexes was performed in a humidified atmosphere per donor for 31-33 h at 38.5˚C in 5% CO_2 in 500 µL of maturation medium consisting of tissue culture medium (TCM)-199 with Earle's salts containing 10% fetal bovine serum (FBS) (31150-022 and 10082147; Thermo Fisher Scientific, Waltham, MA, USA), 9.4 µg/mL follicle-stimulating hormone, 1.88 µg/mL luteinizing hormone (BE-V157997; Stimufol; Reprobiol, Ouffet, Belgium) and 50 µg/mL gentamycin (15710049; Thermo Fisher Scientific, Waltham, MA, USA). A tissue sample was collected for each donor animal from the ovary and stored at -80°C. Following maturation, oocytes were denuded by successive pipetting in 50 µL droplets of 0.1% bovine hyaluronidase (H3506-1G; Sigma-Aldrich, Saint Louis, MO, USA) in TCM-199 with Hank's salts (22350-029; Thermo Fisher Scientific, Waltham, MA, USA) with 10% FBS and 50 µg/mL gentamycin (15710049; Thermo Fisher Scientific, Waltham, MA, USA) and TCM199 with Hank's salts with 10% FBS using a STRIPPER pipettor and 170 µm and 135 µm capillaries (MXL3-STR-CGR, MXL3-175 and MXL3-135; Cooper Surgical, Trumbull, CT, USA). Frozen-thawed semen from one stallion was used for fertilization and prepared as described previously [63]. Mature oocytes were fertilized by intracytoplasmic sperm injection (ICSI) as described previously [64] and presumptive zygotes were cultured at 38.2°C in a humidified atmosphere of 5% O_2, 5% CO_2, and 90% N_2 in 20 µL droplets of Dulbecco's Modified Eagle Medium Nutrient Mixture F-12 (DMEM/F-12; 21331-020; Thermo Fisher Scientific, Waltham, MA, USA) with 10% FBS and 50 µg/mL gentamycin under oil (ART-4008-5PA; Cooper Surgical, Trumbull, CT, USA). Blastocyst development was monitored daily from day six onwards until day 12.

Embryo biopsy and single-cell collection

All but two embryos reaching the blastocyst stage were biopsied at the day of TE delineation and the remaining biopsied embryos were collected immediately afterwards. Embryos not collected included two embryos of Mare 4, one of which presented as an apparent twin embryo

and the other of which developed at day 13. Collection of the TE biopsies was performed as described in detail by De Coster et al. (2023) [65]. Briefly, embryos were placed in 5 µL droplets of TCM-199 with Hank's salts containing 10% FBS under oil and immobilized by suction of a holding pipette (MPH-MED-35; Cooper Surgical, Cooper Surgical, Trumbull, CT, USA) on a heated stage (39°C) of an inverted microscope (Olympus IX73; Olympus Shinjuku, Tokyo, Japan) equipped with a Research Instruments micromanipulation system (Cooper Surgical, Trumbull, CT, USA). To avoid the collection of presumably degenerated cells, embryos were positioned such way that visual cell debris was not aspirated. Ten to 20 cells were aspirated from the TE layer following piercing of the embryo with a beveled biopsy micropipette (MPB-BS-30; Cooper Surgical, Trumbull, CT, USA) and placed in a 2.5 µL droplet of Ca^{+2}/Mg^{+2}-free phosphate buffered saline (PBS; 14190-144; Thermo Fisher Scientific, Waltham, MA, USA) with 1% polyvinylpyrrolidone (PVP; ART-4005; Cooper Surgical, Trumbull, CT, USA), of which 2 µL containing the biopsy was subsequently transferred to a DNase-free 0.2 mL tube on ice using a 135 µm capillary (MXL3-STR-CGR; MXL3-135; Cooper Surgical, Trumbull, CT, USA). The remaining biopsied embryos were individually removed from their zona pellucida, by incubation in warm 0.1% pronase (P8811; Merck KGaA, Darmstadt, Germany; with Hank's salts) dissolved in TCM-199 with Hank's salts, and subsequent washing in TCM-199 with Hank's salts containing 10% FBS and Ca^{+2}/Mg^{+2}-PBS with 1% PVP, while removing the zona by pipetting with a 200 µm capillary. Zona-free embryos were subsequently transferred individually to a DNase-free 0.2 mL tube on ice containing 2 µL Ca^{+2}/Mg^{+2}-free PBS% with 1%PVP using a new 200 µm capillary. For embryos that arrested at the cleavage-stage, individual blastomeres were isolated and collected at day 11 and 12 of embryo culture as described elsewhere [22,61]. When individual blastomeres could not be discriminated due to degeneration and/or fragmentation, embryos were not analyzed. All samples were immediately stored at -80°C. To avoid cross-contamination, every embryo was washed in TCM-199 with Hank's salts with 10% FBS before sample collection and manipulated with a new set of micropipettes and/or capillaries.

Genetic analysis

Whole-genome amplification and DNA extraction

DNA of single blastomeres from arrested embryos, TE biopsies and their corresponding biopsied embryos was extracted and amplified by multiple displacement whole-genome amplification (WGA) using the REPLI-g Single Cell Kit (150345; Qiagen, Hilden, Germany) according to the manufacturer's instructions with minor modifications. Full (TE biopsies and biopsied whole blastocysts) or half (single blastomeres from arrested embryos) reaction volumes were applied and the incubation reaction was reduced to 3 h. The concentration of WGA DNA was determined by Qubit Broad Range Assay (Q33266; Thermo Fisher Scientific,

Waltham, MA, USA), according to the manufacturer's protocol. Ovarian tissue from the donor mares (i.e., mothers of the respective embryos), semen from the stallion (i.e., fathers of the respective embryos) and blood from the parents of the stallion (i.e., paternal grandparents of the respective embryos) were used to extract bulk DNA (DNeasy Blood and Tissue kit; 69504; Qiagen, Hilden, Germany; User-developed protocol for semen extraction). Purity and concentration of the bulk DNA was measured with Nanodrop (Isogen, Utrecht, The Netherlands), according to the manufacturer's protocol.

SNP genotyping and haplarithmisis

Over 70,000 SNPs were genotyped in all samples on GeneSeek Genomic Profiler (GGP) equine SNP arrays (outsourced to Neogen, Ayrshire, UK). Subsequently, discrete genotypes, B-allele frequency (BAF) values, and LogR values were exported using the GenomeStudio software Genotyping Module v. 2.0.5 (Illumina, San Diego, USA). SNP genotypes were called by setting the GenCall score at 0.75. Next the computational workflow siCHILD, which includes haplarithmisis [57], modified for analysis of equine samples according to equine reference genome EquCab3.0 was applied. Briefly, haplarithmisis uses the phased parental genotypes and SNP BAF-values of the embryo sample to determine genome-wide haplotypes, copy-number of the haplotypes, as well as the parental and segregational origin of anomalies. Parental genotypes were phased with the SNP genotype calls derived from the paternal grandparents, when no sibling blastocyst was available (Mare03_Embryo01) or from a sibling biopsied whole blastocyst (other embryos). Next, for specific combinations of phased parental genotypes, corresponding SNP BAF values of the embryo sample were retrieved. Consequently, these values were plotted on paternal and maternal haplarithms, Visualization of genome-wide raw BAF values, paternal and maternal haplarithms, and LogR values using siCHILD rendered genome-wide haplarithm plots.

A visual overview on the interpretation of the haplarithm plots on chromosomal and genome-wide level can be consulted in [57] and [61], respectively. Failed samples were those samples resulting in inconclusive haplarithm plots. The sex of the embryo was determined by the copy number of the X-chromosome as determined by the combined interpretation of the parental haplarithm plots and the LogR value, and verified by qPCR, using 1 µL of 1 ng/µL diluted WGA material as a template [65].

PGT-M

The coordinates of the five selected genes on the equine reference genome EquCab3.0 were extracted from Online Mendelian Inheritance in Animals (OMIA) (https://omia.org/). The gene regions were then indicated by orange lines on whole genome haplarithm plots. Maternally inherited haplotype blocks (regions inherited from homologue 1 or 2 of the mother) were visualized through pink/red colors and paternally inherited haplotype blocks (regions inherited

from homologue 1 or 2 of the father) were visualized with light/dark blue colors. Haplotype inheritance patterns for the five gene regions were inferred according to the color of haplotype block. The haplotype block at loci of interest were called inconclusive in the absence of heterozygosity (e.g., maternal or paternal monosomy), or informative SNP positions or, when co-localized with a meiotic recombination site.

Data visualization

Circos plots were drawn using the Circos software [66].

Ethical approval

Ethical approval was waived for this study since slaughterhouse-obtained material and embryonic samples are no subject of ethical approval.

Acknowledgments

Funding was received from the Research Foundation Flanders (FWO) under grant agreement No. 1139820N to T. De Coster, PICT-A-2021-00063 (ANPCyT, Argentina) to S. Demyda-Peyrás, European Union's Horizon 2020 research and innovation program under grant agreement No 824110 – EASI-Genomics to J. R. Vermeesch and the Marie Skłodowska-Curie grant agreement No 813707 (MATER) to Y. Zhao. Institutional support was received from the KU Leuven, C1- C14/18/092 and C14/22/125 to J. R. Vermeesch and from UGent, BOF/STA/202202/008, to K. Smits.

Author contributions

Conceptualization: TDC, AVS, JV, KS; Embryo production and biopsies: TDC; SNP-arrays: TDC, SDP; bioinformatics: YZ; data interpretation: TDC, YZ, OT; writing: TDC, KS; review and editing: all authors

Data availability

All SNP-array data will be available on EGA https://ega-archive.org/.

Competing interests

The authors declare no competing interests.

References

1. Aurich, J. E. Artificial Insemination in Horses—More than a Century of Practice and Research. *Journal of Equine Veterinary Science* **32**, 458–463 (2012).
2. Viana, J. H. 2020 Statistics of embryo production and transfer in domestic farm animals. In: Embryo Technology Newsletter, v.40, n.4, 2022. vol. 39 https://www.iets.org/Portals/0/Documents/Public/Committees/DRC/IETS_Data_Retrieval_Report_2020.pdf (2022).
3. Lazzari, G. *et al.* Laboratory Production of Equine Embryos. *Journal of Equine Veterinary Science* **89**, 103097 (2020).
4. VRT nws. World record paid for horse embryo: "You're buying hope!". https://www.vrt.be/vrtnws/en/2023/09/14/world-record-paid-for-horse-embryo-youre-buying-hope/(2023).
5. Meyers, S. *et al.* Equine non-invasive time-lapse imaging and blastocyst development. *Reproduction, Fertility, and Development* **31**, 1874–1884 (2019).
6. Fanelli, D. *et al.* Interspecific and Intraspecific Artificial Insemination in Domestic Equids. *Animals* **13**, 582 (2023).
7. McKinnon, A. O., Squires, E. L., Vaala, W. E. & Varner, D. D (eds). In: *Equine Reproduction*. (Wiley-Blackwell, 2011).
8. Claes, A. & Stout, T. A. E. Success rate in a clinical equine in vitro embryo production program. *Theriogenology* **187**, 215–218 (2022).
9. Cuervo-Arango, J., Claes, A. N. & Stout, T. A. A retrospective comparison of the efficiency of different assisted reproductive techniques in the horse, emphasizing the impact of maternal age. *Theriogenology* **132**, 36–44 (2019).
10. Cuervo-Arango, J., Claes, A. N. & Stout, T. A. E. Small day 8 equine embryos cannot be rescued by a less advanced recipient mare uterus. *Theriogenology* **126**, 36–40 (2019).
11. Shilton, C. A., Kahler, A., Roach, J. M., Raudsepp, T. & de Mestre, A. M. Lethal variants of equine pregnancy : is it the placenta or foetus leading the conceptus in the wrong direction? *Reproduction, Fertility and Development* **35**, 51–69 (2023).
12. Claes, A. *et al.* Factors affecting the likelihood of pregnancy and embryonic loss after transfer of cryopreserved in vitro produced equine embryos. *Equine Veterinary Journal* **51**, 446–450 (2019).
13. Vanderwall, D. K., Squires, E. L., Brinsko, S. P. & McCue, P. M. Diagnosis and management of abnormal embryonic development characterized by formation of an embryonic vesicle without an embryo in mares. *Journal of the American Veterinary Medical Association* **217**, 58–63 (2000).
14. Schmutz, S. M., Moker, J. S., Clark, E. G. & Orr, J. P. Chromosomal aneuploidy associated with spontaneous abortions and neonatal losses in cattle. *Journal of Veterinary Diagnostic Investigation* **8**, 91–95 (1996).
15. Jia, C.-W. *et al.* Aneuploidy in Early Miscarriage and its Related Factors. *Chinese Medical Journal* **128**, 2772–2776 (2015).
16. Chen, S. *et al.* A copy number variation genotyping method for aneuploidy detection in spontaneous abortion specimens. *Prenatal Diagnosis* **37**, 176–183 (2017).
17. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics* **2**, 280–291 (2001).
18. Coates, J. W., Schmutz, S. M. & Rousseaux, C. G. A survey of malformed aborted bovine fetuses, stillbirths and nonviable neonates for abnormal karyotypes. *Canadian Journal of Veterinary Research* **52**, 258–263 (1988).
19. Nikitina, T. V *et al.* Comparative Cytogenetic Analysis of Spontaneous Abortions in Recurrent and Sporadic Pregnancy Losses. *Biomedicine Hub* **1**, 1–11 (2016).
20. Shilton, C. A. *et al.* Whole genome analysis reveals aneuploidies in early pregnancy loss in the horse. *Scientific Reports* **10**, 13314 (2020).
21. Vanneste, E. *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nature Medicine* **15**, 577–583 (2009).
22. Tšuiko, O. *et al.* Genome stability of bovine in vivo conceived cleavage-stage embryos is higher compared to in vitro-produced embryos. *Human Reproduction* **32**, 2348–2357 (2017).
23. Destouni, A. *et al.* Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy. *Genome Research* **26**, 567–578 (2016).
24. Tšuiko, O. *et al.* Haplotyping-based preimplantation genetic testing reveals parent-of-origin specific mechanisms of aneuploidy formation. *npj Genomic Medicine* **6**, 81 (2021).
25. Fragouli, E., Munne, S. & Wells, D. The cytogenetic constitution of human blastocysts: Insights from comprehensive chromosome screening strategies. *Human Reproduction Update* **25**, 15–33 (2019).

26. Turner, K. J. *et al.* Karyomapping for simultaneous genomic evaluation and aneuploidy screening of preimplantation bovine embryos: The first live-born calves. *Theriogenology* **125**, 249–258 (2019).
27. Tutt, D. A. R. *et al.* Analysis of bovine blastocysts indicates ovarian stimulation does not induce chromosome errors, nor discordance between inner-cell mass and trophectoderm lineages. *Theriogenology* **161**, 108–119 (2021).
28. McCoy, R. C. *et al.* Evidence of Selection against Complex Mitotic-Origin Aneuploidy during Preimplantation Development. *PLoS Genetics* **11**, e1005601 (2015).
29. Fragouli, E. *et al.* The origin and impact of embryonic aneuploidy. *Human Genetics* **132**, 1001–1013 (2013).
30. Silvestri, G. *et al.* Preimplantation Genetic Testing for Aneuploidy Improves Live Birth Rates with In Vitro Produced Bovine Embryos: A Blind Retrospective Study. *Cells* **10**, 2284 (2021).
31. Dimitriadou, E. *et al.* Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos. *Human Reproduction* **32**, 687–697 (2017).
32. McCoy, R. C. *et al.* Meiotic and mitotic aneuploidies drive arrest of in vitro fertilized human preimplantation embryos. *Genome Medicine*(2023).
33. Ottolini, C. S. *et al.* Tripolar mitosis and partitioning of the genome arrests human preimplantation development in vitro. *Scientific Reports* **7**, 9744 (2017).
34. Coppola, G. *et al.* Use of cross-species in-situ hybridization (ZOO-FISH) to assess chromosome abnormalities in day-6 in-vivo- or in-vitro-produced sheep embryos. *Chromosome Research* **15**, 399–408 (2007).
35. Brooks, K. E. *et al.* Assessing equine embryo developmental competency by time-lapse image analysis. *Reproduction, Fertility, and Development* **31**, 1840–1850 (2019).
36. Ducheyne, K. D. *et al.* In vitro production of horse embryos predisposes to micronucleus formation, whereas time to blastocyst formation affects likelihood of pregnancy. *Reproduction, Fertility, and Development* **31**, 1830–1839 (2019).
37. Rambags, B. P. B. *et al.* Numerical chromosomal abnormalities in equine embryos produced in vivo and in vitro. *Molecular Reproduction and Development* **72**, 77–87 (2005).
38. Franciosi, F. *et al.* Analysis of Chromosome Segregation, Histone Acetylation, and Spindle Morphology in Horse Oocytes. *Journal of Visualized Experiments* 55242 (2017) doi:10.3791/55242.
39. Petersen, J. L. *et al.* Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genetics* **9**, e1003211 (2013).
40. Fages, A. *et al.* Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell* **177**, 1419-1435.e31 (2019).
41. Raudsepp, T., Finno, C. J., Bellone, R. R. & Petersen, J. L. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Animal Genetics* **50**, 569–597 (2019).
42. Wagner, M. L. *et al.* Allele frequency and likely impact of the glycogen branching enzyme deficiency gene in Quarter Horse and Paint Horse populations. *Journal of Veterinary Internal Medicine* **20**, 1207–1211 (2006).
43. Ward, T. L. *et al.* Glycogen branching enzyme (GBE1) mutation causing equine glycogen storage disease IV. *Mammalian Genome* **15**, 570–577 (2004).
44. Tryon, R. C. *et al.* Evaluation of allele frequencies of inherited disease genes in subgroups of American Quarter Horses. *Journal of the American Veterinary Medical Association* **234**, 120–125 (2009).
45. Valberg, S. J. *et al.* Glycogen branching enzyme deficiency in quarter horse foals. *Journal of Veterinary Internal Medicine* **15**, 572–580 (2001).
46. Reiter, S. *et al.* Distribution of the Warmblood Fragile Foal Syndrome Type 1 Mutation (PLOD1 c.2032G>A) in Different Horse Breeds from Europe and the United States. *Genes* **11**, 1518 (2020).
47. Winand, N. J. Identification of the Causative Mutation for Inherited Connective Tissue Disorders in Equines. (2011).
48. Monthoux, C., de Brot, S., Jackson, M., Bleul, U. & Walter, J. Skin malformations in a neonatal foal tested homozygous positive for Warmblood Fragile Foal Syndrome. *BMC Veterinary Research* **11**, 12 (2015).
49. Tarr, C. J., Thompson, P. N., Guthrie, A. J. & Harper, C. K. The carrier prevalence of severe combined immunodeficiency, lavender foal syndrome and cerebellar abiotrophy in Arabian horses in South Africa. *Equine Veterinary Journal* **46**, 512–514 (2014).
50. Bugno-Poniewierska, M. *et al.* Genetic screening for cerebellar abiotrophy, severe combined

immunodeficiency and lavender foal syndrome in Arabian horses in Poland. *The Veterinary Journal* **248**, 71–73 (2019).
51. Brault, L. S., Cooper, C. A., Famula, T. R., Murray, J. D. & Penedo, M. C. T. Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics* **97**, 121–129 (2011).
52. Ducro, B. J. *et al.* A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses. *BMC Genomics* **16**, 761 (2015).
53. Guignot, F. *et al.* Preimplantation genetic diagnosis in Welsh pony embryos after biopsy and cryopreservation. *Journal of Animal Science* **93**, 5222–5231 (2015).
54. Choi, Y. H., Penedo, M. C. T., Daftari, P., Velez, I. C. & Hinrichs, K. Accuracy of preimplantation genetic diagnosis in equine in vivo-recovered and in vitro-produced blastocysts. *Reproduction, Fertility, and Development* **28**, 1382–1389 (2015).
55. Choi, Y. H. *et al.* Viability of equine embryos after puncture of the capsule and biopsy for preimplantation genetic diagnosis. *Reproduction* **140**, 893–902 (2010).
56. Barandalla, M. *et al.* 73 Preimplantation genetic diagnosis of glycogen branching enzyme deficiency and sex determination in equine in vitro-produced embryos. *Reproduction, Fertility and Development* **34**, 272–273 (2022).
57. Zamani Esteki, M. *et al.* Concurrent Whole-Genome Haplotyping and Copy-Number Profiling of Single Cells. *American Journal of Human Genetics* **96**, 894–912 (2015).
58. Handyside, A. H. *et al.* Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. *Journal of Medical Genetics* **47**, 651–658 (2010).
59. De Witte, L. *et al.* GENType: all-in-one preimplantation genetic testing by pedigree haplotyping and copy number profiling suitable for third-party reproduction. *Human Reproduction* **37**, 1678–1691 (2022).
60. Ben-Nagi, J. *et al.* Karyomapping: a single centre's experience from application of methodology to ongoing pregnancy and live-birth rates. *Reproductive Biomedicine Online* **35**, 264–271 (2017).
61. De Coster, T. *et al.* Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. *Genome Biology* **23**, 201 (2022).
62. Masset, H. *et al.* Single-cell genome-wide concurrent haplotyping and copy-number profiling through genotyping-by-sequencing. *Nucleic Acids Research* **52**, e63 (2022).
63. Angel-Velez, D. *et al.* New Alternative Mixtures of Cryoprotectants for Equine Immature Oocyte Vitrification. *Animals* vol. 11(11) 3077 (2021).
64. Papas, M. *et al.* Anti-Müllerian Hormone and OPU-ICSI Outcome in the Mare. *Animals* **11**, 2004 (2021).
65. De Coster, T.,Van Poucke, M. *et al.* Single closed-tube quantitative real-time PCR assay with dual-labeled probes for improved sex determination of equine embryos. *Animal*, 100952 (2023).
66. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Research* **19**, 1639–1645 (2009).
67. Vega, M., Breborowicz, A., Moshier, E. L., McGovern, P. G. & Keltz, M. D. Blastulation rates decline in a linear fashion from euploid to aneuploid embryos with single versus multiple chromosomal errors. *Fertility and Sterility* **102**, 394–398 (2014).
68. McCoy, R. C. Mosaicism in Preimplantation Human Embryos: When Chromosomal Abnormalities Are the Norm. *Trends in Genetics* **33**, 448–463 (2017).
69. Griffin, D. K. & Ogur, C. Chromosomal analysis in IVF: just how useful is it? *Reproduction* **156**, F29–F50 (2018).
70. Munné, S. *et al.* 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. *Fertility and Sterility* **112**, 1071-1079.e7 (2019).
71. Fragouli, E. *et al.* Analysis of implantation and ongoing pregnancy rates following the transfer of mosaic diploid-aneuploid blastocysts. *Human Genetics* **136**, 805–819 (2017).
72. Munné, S. *et al.* Clinical outcomes after the transfer of blastocysts characterized as mosaic by high resolution Next Generation Sequencing- further insights. *European Journal of Medical Genetics* **63**, 103741 (2020).
73. Wang, Y. *et al.* Clinical outcomes of subtypes of mosaic single aneuploid embryos after preimplantation genetic testing for aneuploidy. *Journal of Assisted reproduction and Genetics* **40**, 639–652 (2023).
74. Gleicher, N., Patrizio, P., Mochizuki, L. & Barad, D. H. Previously reported and here added cases demonstrate euploid pregnancies followed by PGT-A as 'mosaic' as well as 'aneuploid' designated embryos. *Reproductive Biology and Endocrinology* **21**, 25 (2023).

75. Claes, A. *et al.* Speed of in vitro embryo development affects the likelihood of foaling and the foal sex ratio. *Reproduction, Fertility, and Development* **32**, 468–473 (2020).
76. Galli, C., Duchi, R., Colleoni, S., Lagutina, I. & Lazzari, G. Ovum pick up, intracytoplasmic sperm injection and somatic cell nuclear transfer in cattle, buffalo and horses: From the research laboratory to clinical practice. *Theriogenology* **81**, 138–151 (2014).
77. Lewis, N. *et al.* Equine in vitro produced blastocysts: relationship of embryo morphology, stage and speed of development to foaling rate. *Reproduction, Fertility, and Development* **35**, 338–351 (2023).
78. Spits, C. & Sermon, K. PGD for monogenic disorders: aspects of molecular biology. *Prenatal Diagnosis* **29**, 50–56 (2009).
79. Piyamongkol, W., Bermúdez, M. G., Harper, J. C. & Wells, D. Detailed investigation of factors influencing amplification efficiency and allele drop-out in single cell PCR: implications for preimplantation genetic diagnosis. *Molecular Human Reproduction* **9**, 411–420 (2003).
80. Treff, N. R. *et al.* Preimplantation Genetic Testing for Polygenic Disease Relative Risk Reduction: Evaluation of Genomic Index Performance in 11,883 Adult Sibling Pairs. *Genes* **11**, 648 (2020).
81. Treff, N. R. *et al.* Validation of concurrent preimplantation genetic testing for polygenic and monogenic disorders, structural rearrangements, and whole and segmental chromosome aneuploidy with a single universal platform. *European Journal of Medical Genetics* **62**, 103647 (2019).
82. Tellier, L. C. A. M. *et al.* Embryo Screening for Polygenic Disease Risk: Recent Advances and Ethical Considerations. *Genes* **12**, 1105 (2021).
83. Siermann, M. *et al.* Limitations, concerns and potential: attitudes of healthcare professionals toward preimplantation genetic testing using polygenic risk scores. *European Journal of Human Genetics* (2023) doi:10.1038/s41431-023-01333-9.
84. Turley, P. *et al.* Problems with Using Polygenic Scores to Select Embryos. *The New England Journal of Medicine* **385**, 78–86 (2021).
85. McCoy, A. M. *et al.* Identification and validation of risk loci for osteochondrosis in standardbreds. *BMC Genomics* **17**, 41 (2016).
86. Zimmermann, E. & Distl, O. SNP-Based Heritability of Osteochondrosis Dissecans in Hanoverian Warmblood Horses. *Animals* **13**, 1462 (2023).
87. Drabbe, A. *et al.* Genome-Wide Association Analyses of Osteochondrosis in Belgian Warmbloods Reveal Candidate Genes Associated With Chondrocyte Development. *Journal of Equine Veterinary Science* **111**, 103870 (2022).
88. Velie, B. D. *et al.* Using an Inbred Horse Breed in a High Density Genome-Wide Scan for Genetic Risk Factors of Insect Bite Hypersensitivity (IBH). *PLoS ONE* **11**, e0152966 (2016).
89. Bailey, E., Petersen, J. L. & Kalbfleisch, T. S. Genetics of Thoroughbred Racehorse Performance. *Annual Review of Animal Biosciences* **10**, 131–150 (2022).
90. Han, H. *et al.* Common protein-coding variants influence the racing phenotype in galloping racehorse breeds. *Communications Biology* **5**, 1320 (2022).
91. Dias De Castro, L. L. *et al.* Genome-wide association study in thoroughbred horses naturally infected with cyathostomins. *Animal Biotechnology* 1–13 (2022) doi:10.1080/10495398.2022.2099880.
92. Laseca, N. *et al.* A genome-wide association study of mare fertility in the Pura Raza Español horse. *Animal* **16**, 100476 (2022).
93. Nikitkina, E. V *et al.* Genome-wide association study for frozen-thawed sperm motility in stallions across various horse breeds. *Animal Bioscience* **35**, 1827–1838 (2022).
94. Ding, J. *et al.* Identity-by-state-based haplotyping expands the application of comprehensive preimplantation genetic testing. *Human Reproduction* **35**, 718–726 (2020).
95. Masset, H. *et al.* Multi-centre evaluation of a comprehensive preimplantation genetic test through haplotyping-by-sequencing. *Human Reproduction* **34**, 1608–1619 (2019).
96. Schaefer, R. J. & McCue, M. E. Equine Genotyping Arrays. *Veterinary Clinics of North America: Equine Practice* **36**, 183–193 (2020).
97. Felix, M. R., Turner, R. M., Dobbie, T. & Hinrichs, K. Successful in vitro fertilization in the horse: production of blastocysts and birth of foals after prolonged sperm incubation for capacitation. *Biology of Reproduction* **107**, 1551–1564 (2022).
98. Wiggans, G. R., Cole, J. B., Hubbard, S. M. & Sonstegard, T. S. Genomic Selection in Dairy Cattle: The USDA Experience. *Annual Review of Animal Biosciences* **5**, 309–327 (2017).
99. Wiggans, G. R. & Carrillo, J. A. Genomic selection in United States dairy cattle. *Frontiers in Genetics* **13**, 994466 (2022).

CHAPTER 4 – ORIGIN AND DEVELOPMENT OF BOVINE UNIPARENTAL AND POLYPLOID BLASTOMERES

This chapter is based on the following papers:

De Coster, T., Masset, H., Tšuiko, O., Catteeuw, M., **Zhao, Y.**, Dierckxsens, N., Aparicio, A. L., Dimitriadou, E., Debrock, S., Peeraer, K., de Ruijter-Villani, M., Smits, K., Van Soom, A., & Vermeesch, J. R. (2022). Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. ***Genome Biology*** 2022 23:1, 23(1), 1–29. https://doi.org/10.1186/S13059-022-02763-2.

Zhao, Y. *, Fernández-Montoro, A. *, Peeters, G., Jatsenko, T., Coster, T. De, Angel-Velez, D., Lefevre, T., Voet, T., Tšuiko, O., Kurg, A., Smits, K., Soom, A. Van, & Vermeesch, J. R. (2024). Origin and development of uniparental and polyploid blastomeres. *BioRxiv*, 2024.07.30.605883. https://doi.org/10.1101/2024.07.30.605883. Manuscript peer review completed and accepted by *iScience*. (*Co-first authors)

Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts

The following is a summary of the paper: De Coster, T., Masset, H., Tšuiko, O., Catteeuw, M., Zhao, Y., Dierckxsens, N., Aparicio, A. L., Dimitriadou, E., Debrock, S., Peeraer, K., de Ruijter-Villani, M., Smits, K., Van Soom, A., & Vermeesch, J. R. (2022). Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. *Genome Biology*, 23(1), 1–29. https://doi.org/10.1186/S13059-022-02763-2. In this study, I contributed to genome-wide haplotyping and copy number profiling. The full paper is provided as C4-Additional file 1.

Background

Chromosomal mosaicism, marked by the coexistence of genetically distinct cell populations within a single organism, is well-documented in human (Vanneste *et al.*, 2009; Fragouli, Munne and Wells, 2019) and bovine (Destouni *et al.*, 2016; Tšuiko *et al.*, 2017) early embryonic development. This phenomenon contributes to embryonic arrest, pregnancy loss, and congenital disorders (Biesecker and Spinner, 2013; Ottolini *et al.*, 2017; Zhang *et al.*, 2019). A peculiar form of mosaicism involves the presence of cells with different genome ploidy or distinct parental genotypes, known as mixoploidy and chimerism, respectively. Instances where one parental genome contributes exclusively to a cell lineage are termed mosaic genome-wide (GW) uniparental disomy. Chimerism, mixoploidy, and mosaic GW uniparental disomy have been associated with rare developmental disorders and placental abnormalities in humans and bovines (Jarvela *et al.*, 1993; Strain *et al.*, 1995; Meinecke *et al.*, 2006; Robinson *et al.*, 2007).

Despite the recognized clinical associations, the mechanistic origins of mixoploidy and chimerism remain poorly understood. Proposed mechanisms include zygotic and polar body aggregation, parthenogenetic activation, and fertilization errors (Madan, 2020). However, these models are derived from studies performed on patients and abnormal placentae and rely on the cytogenetic or molecular genetic detection of cells that have undergone rigorous prenatal selection. Hence, these models remain speculative and largely unsupported by molecular or cell biological data.

Recent advances in genome-wide single-cell haplotyping and copy number profiling have revealed androgenetic, gynogenetic, biparental, and triploid cells coexisting within individual day-2 and day-3 bovine embryos (Destouni *et al.*, 2016; Tšuiko *et al.*, 2017). The existence of mixoploidy and chimerism was subsequently confirmed in other bovine (Middelkamp *et al.*, 2020), and non-human primate (Daughtry *et al.*, 2019) *in vitro*-fertilized (IVF) cleavage-stage embryos. These discoveries imply that mixoploidy and chimerism may result from the

segregation of parental genomes into different daughter cells during the zygotic division. We coined this phenomenon “heterogoneic division”—Greek for a different parental origin—and reasoned that it might be enriched in embryos cleaving into more than two blastomeres directly (multipolar zygotic division) (Destouni *et al.*, 2016; Masset, Tšuiko and Vermeesch, 2021).

Methods and results

Genome-wide mosaicism exists in human blastocysts

Considering that bovine and non-human primate embryogenesis parallel human embryogenesis, including the chromosomal instability (Destouni *et al.*, 2016; Daughtry *et al.*, 2019), we hypothesized mixoploidy and/or chimerism to be traceable in human cleavage- and blastocyst-stage embryos. In two human embryos, a gynogenetic blastomere at day-3 biopsy was identified through comprehensive haplotyping-based preimplantation genetic testing (PGT). In order to understand if the detected GW aberrations were indicative of a mixoploid and/or chimeric embryo, we dissociated the resulting blastocysts for single-cell analysis. This analysis revealed a mixture of gynogenetic, biparental diploid, and polyploid cells. These analyses point to the occurrence of human parental genome segregation errors and the persistence of these lineages to the blastocyst stage resulting into chimeric and/or mixoploid human blastocysts.

Multipolar zygotic divisions are characterized by whole-genome segregation errors

Because human embryo research is ethically and numerically restricted, we used a bovine model to pinpoint the origin of parental genome segregation errors and to test whether multipolar zygotic divisions coincide with parental genome segregation errors and gain insights in the mechanisms. We analyzed the GW haplotype architecture and ploidy state of all cells derived from 25 *in vitro*-produced bovine zygotes undergoing multipolar division into three or four blastomeres. Haplarithmisis was performed on 82 single blastomeres and two cellular fragments from these embryos. Remarkably, all embryos (n=25) exhibited GW abnormalities in at least one blastomere, including polyploidy with additional maternal or paternal genomes, uniparental signatures (androgenetic or gynogenetic), the GW presence of complex aneuploidies, or the apparent absence of DNA. Heterogoneic division, characterized by the separate segregation of whole parental genomes into distinct blastomeres, was observed in 18 out of 25 embryos, 17 of which were polyspermic and one monospermic. Among the seven embryos without heterogoneic division, only one showed polyspermic fertilization. These observations suggest that polyspermic fertilization is the primary driver of multipolar division and is associated with whole-genome segregation errors.

Moreover, to pinpoint how these mis-segregations arise, we followed the zygotic division, in real-time, in 55 bovine zygotes transiently expressing live fluorescent markers for

chromosomes and microtubules. Out of 55 imaged embryo , three exhibited multipolar division into three cells, while one divided into four cells **(Figure 1)**. Heterogoneic division occurred in three of these four embryos **(Figure 1C-E)**, with observed mechanisms including a lack of synchrony between pronuclei, the formation of private parental spindles and the presence of separate spindles.

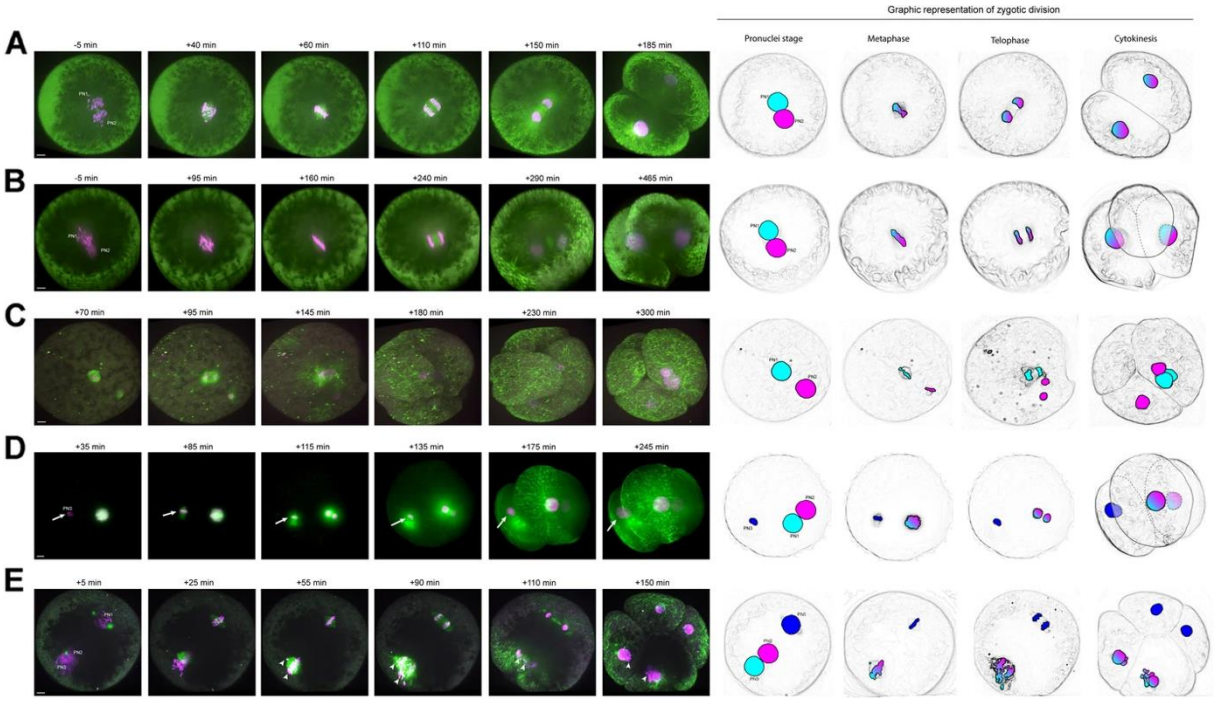


Figure 1. Spindle assembly and cytokinesis dynamics in live-imaged bovine zygotes. Spindle assembly and cytokinesis into two (**A**) or multiple blastomeres (**B–E**) in live-imaged bovine zygotes containing two (**A**, **C**) or three (**D**, **E**) pronuclei. **A–E** Bovine zygotes expressing transient microtubule markers (EGFP-MAP4, green) and chromatin marker (H2B-mCherry, magenta) were imaged by light sheet (**D**) or spinning disk (**A–C**, **D**) microscopy every 10 min throughout mitosis and for up to 12 h in total. Graphical representation of the zygotic division on the embryos imaged illustrating the segregation dynamics of the different genomes. Timings are respective to synchronous pro-nuclear evelope breadown (NEBD) (**A–C**) or to NEBD of the leading PN (PN1) in case of asynchrony (**D**, **E**). PN3, asynchronous lagging PN. Arrows indicate the extrusion of a parental genome into a separate blastomere. Arrowheads indicate misaligned and lagging chromosomes resulting in micronuclei formation in the daughter blastomeres. Projected scale bars, 10 µm.

Whole parental genome segregation errors occur via different mechanisms

Mapping of the segregational origin of the genomic content revealed distinct mechanisms that lead to separate segregation of whole parental genomes into distinct blastomeres in the 18 embryos with heterogoneic division. Three embryos consisted of an androgenetic, a biparental diploid, and a diandric triploid (i.e., containing two distinct paternal set of haplotypes) blastomere. Thirteen embryos consisted of a combination of biparental and androgenetic

blastomeres **(Figure 2A, B)**. The extra paternal genomes were either extruded into a separate blastomere **(Figure 2A)** or segregated by the operation of an additional paternal spindle **(Figure 2B)**. In two zygotes, the parental genomes were segregated into three or four blastomeres, each containing a maternal or paternal genome **(Figure 2C)**.

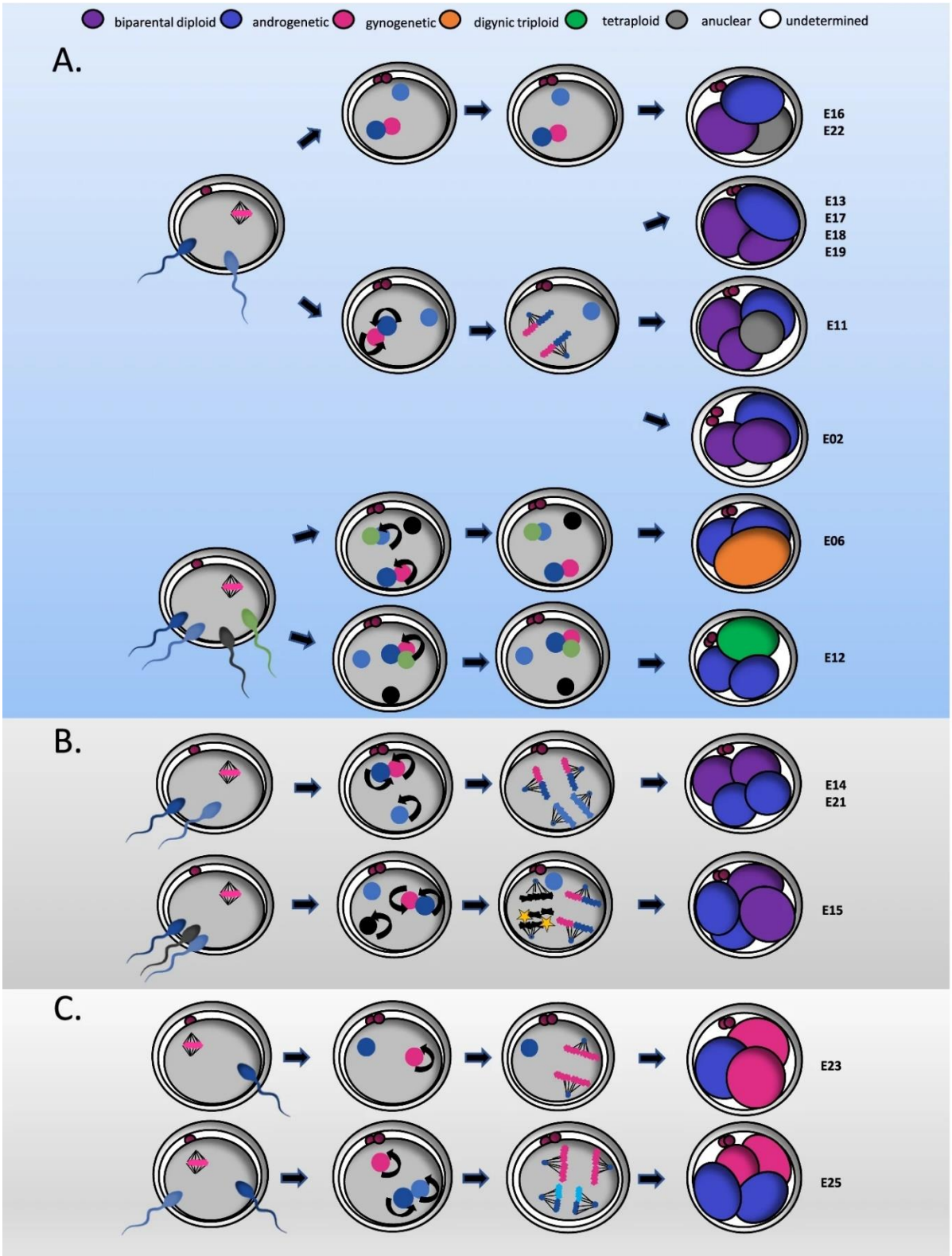


Figure 2. Schematic representation of events that best explain the segregation of a zygote into

biparental and androgenetic (A, B) or androgenetic and gynogenetic blastomeres (C). Curved arrows depict replication of the genome. In **A**, extrusion of paternal genome(s) is depicted following polyspermic fertilization. In some embryos, this event occurred in parallel with the replication and karyokinesis of the primary maternal and paternal genome. In E06 and E12, some of the parental genomes were replicated but failed to undergo karyokinesis. In **B**, segregation of an additional paternal genome occurs through a second, paternally organized spindle. In E15, a third paternal genome was protruded to one of the androgenetic blastomeres and genome-wide chromosomal losses occurred on one paternal genome (yellow stars). In **C**, private parental spindles are established around the maternal and the paternal genomes (E25) following polyspermic fertilization, or only around the maternal genome following normal fertilization (E23). In E23, the paternal genome is extruded in a separate blastomere.

Genome-wide abnormalities resulting from heterogoneic division contribute to embryonic arrest and persist in the blastocyst-stage bovine embryo

To prove that blastomeres containing GW abnormalities can propagate and persist following heterogoneic division, we cultured an additional cohort of bovine embryos to the blastocyst stage. Of the 190 zygotes cultured, 92.6% cleaved, and 41.5% developed to the blastocyst stage. The blastocyst formation rate from bovine zygotes that underwent multipolar division (30.1 ± 8.1%) was reduced compared to zygotes that underwent bipolar division (49.3 ± 7.1%) (P = 0.03), which shows that embryos with GW abnormalities resulting from multipolar zygotic division are more prone to embryonic arrest. Seven blastocysts which formed from zygotes that underwent multipolar zygotic division were dissociated and six to twelve cells were genotyped individually. The subgroup of sampled cells presented a biparental diploid constitution in two blastocysts and a GW mosaic constitution in five blastocysts. In the five GW mosaic embryos, cells of the same blastocyst showed distinct paternal haplotypes or GW paternal heterodisomy, confirming polyspermy to instigate multipolar zygotic division.

Conclusions

This study demonstrates that multipolar zygotic division, a deviation from the typical bipolar cleavage, can lead to whole-genome segregation errors, resulting in blastomeres with distinct parental genomes. The persistence of these genomic anomalies into the blastocyst stage underscores the potential for mixoploid or chimeric development in both human and bovine embryos. While multipolar divisions reduce the likelihood of successful blastocyst formation, they offer a mechanistic explanation for developmental disorders associated with mixoploidy and chimerism. Our study lays the groundwork for future investigations into how these early genomic mis-segregations influence long-term developmental outcomes.

References

Biesecker, L.G. and Spinner, N.B. (2013) 'A genomic view of mosaicism and human disease', *Nature Reviews Genetics 2013 14:5*, 14(5), pp. 307–320. Available at: https://doi.org/10.1038/nrg3424.

Daughtry, B.L. *et al.* (2019) 'Single-cell sequencing of primate preimplantation embryos reveals chromosome elimination via cellular fragmentation and blastomere exclusion', *Genome Research*, 29(3), pp. 367–382. Available at: https://doi.org/10.1101/GR.239830.118.

Destouni, A. *et al.* (2016) 'Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy', *Genome Research*, 26(5), pp. 567–578. Available at: https://doi.org/10.1101/GR.200527.115.

Fragouli, E., Munne, S. and Wells, D. (2019) 'The cytogenetic constitution of human blastocysts: insights from comprehensive chromosome screening strategies', *Human Reproduction Update*, 25(1), pp. 15–33. Available at: https://doi.org/10.1093/HUMUPD/DMY036.

Jarvela, I.E. *et al.* (1993) '46,XX/69,XXX diploid-triploid mixoploidy with hypothyroidism and precocious puberty.', *Journal of Medical Genetics*, 30(11), pp. 966–967. Available at: https://doi.org/10.1136/JMG.30.11.966.

Madan, K. (2020) 'Natural human chimeras: A review', *European Journal of Medical Genetics*, 63(9), p. 103971. Available at: https://doi.org/10.1016/J.EJMG.2020.103971.

Masset, H., Tšuiko, O. and Vermeesch, J.R. (2021) 'Genome-wide abnormalities in embryos: Origins and clinical consequences', *Prenatal diagnosis*, 41(5), pp. 554–563. Available at: https://doi.org/10.1002/PD.5895.

Meinecke, B. *et al.* (2006) 'A Diploid-Triploid (60,XX/90,XXY) Intersex in a Holstein Heifer', *Sexual Development*, 1(1), pp. 59–65. Available at: https://doi.org/10.1159/000096239.

Middelkamp, S. *et al.* (2020) 'Sperm DNA damage causes genomic instability in early embryonic development', *Science Advances*, 6(16). Available at: https://doi.org/10.1126/SCIADV.AAZ7602/SUPPL_FILE/AAZ7602_TABLE_S1.XLSX.

Ottolini, C.S. *et al.* (2017) 'Tripolar mitosis and partitioning of the genome arrests human preimplantation development in vitro', *Scientific Reports 2017 7:1*, 7(1), pp. 1–10. Available at: https://doi.org/10.1038/s41598-017-09693-1.

Robinson, W.P. *et al.* (2007) 'Origin and outcome of pregnancies affected by androgenetic/biparental chimerism', *Human Reproduction*, 22(4), pp. 1114–1122. Available at: https://doi.org/10.1093/HUMREP/DEL462.

Strain, L. *et al.* (1995) 'A human parthenogenetic chimaera', *Nature Genetics 1995 11:2*, 11(2), pp. 164–169. Available at: https://doi.org/10.1038/ng1095-164.

Tšuiko, O. *et al.* (2017) 'Genome stability of bovine in vivo-conceived cleavage-stage embryos is higher compared to in vitro-produced embryos', *Human Reproduction*, 32(11), pp. 2348–2357. Available at: https://doi.org/10.1093/HUMREP/DEX286.

Vanneste, E. *et al.* (2009) 'Chromosome instability is common in human cleavage-stage embryos', *Nature Medicine 2009 15:5*, 15(5), pp. 577–583. Available at: https://doi.org/10.1038/nm.1924.

Zhang, L. *et al.* (2019) 'Rates of live birth after mosaic embryo transfer compared with euploid embryo transfer', *Journal of Assisted Reproduction and Genetics*, 36(1), pp. 165–172. Available at: https://doi.org/10.1007/S10815-018-1322-2/TABLES/4.

Origin and development of uniparental and polyploid blastomeres

Yan Zhao[1,6,7], Andrea Fernández-Montoro[2,7], Greet Peeters[1], Tatjana Jatsenko[1], Tine De Coster[2], Daniel Angel-Velez[2,3], Thomas Lefevre[4], Thierry Voet[4,5], Olga Tšuiko[1], Ants Kurg[6], Katrien Smits[2], Ann Van Soom[2], Joris Robert Vermeesch[1]

[1]Laboratory for Cytogenetics and Genome Research, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium

[2]Department of Internal Medicine, Reproduction, and Population Medicine - Ghent University, 9820, Merelbeke, Belgium.

[3]Research Group in Animal Sciences – INCA-CES, Universidad CES, 050021 Medellin, Colombia

[4]Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium

[5]KU Leuven Institute for Single Cell Omics (LISCO), University of Leuven, KU Leuven, Leuven, Belgium

[6]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, 51010, Tartu, Estonia

[7]These authors contributed equally

Corresponding author: Joris.Vermeesch@KULeuven.be

Abstract

Whole-genome (WG) abnormalities, such as uniparental diploidy and triploidy, cause fetal death. Occasionally, they coexist with biparental diploid cells in live births. Understanding the origin and early development of WG abnormal blastomeres is crucial for explaining the formation of androgenotes, gynogenotes, triploidy, chimerism, and mixoploidy. By haplotyping 118 bovine blastomeres from the first cleavages, we identified that heterogoneic division occurs in both multipolar and bipolar cleaving zygotes. During heterogoneic division, parental genomes segregate into distinct blastomeres, resulting in the coexistence of uniparental and biparental diploid or polyploid cells. Mechanisms underlying such heterogoneic division, as inferred from haplotyping results, include the formation of additional paternal or private parental spindles, the occurrence of tripolar spindles, and the extrusion of extra paternal genomes. After culturing the totipotent blastomeres to three preimplantation stages and exploring transcriptomes of 446 cells, we discovered that stress responses contribute to developmental impairment in WG abnormal cells, resulting in either cell arrest or blastocyst formation. Some first-cleavage-derived WG abnormal blastomeres can survive early development and progress to blastocysts. Their dominance in preimplantation embryos represents an overlooked cause

of abnormal development. Haplotype based screening could improve in vitro fertilization outcomes.

Introduction

Chromosomal abnormalities are common during human development, especially in the preimplantation phase (Hassold et al., 1980; Hassold and Hunt, 2001; Vanneste et al., 2009). One notable category of chromosome abnormalities are whole-genome (WG) anomalies (Masset et al., 2021). Embryos bearing WG abnormalities contain abnormal levels of parental genomes rather than the normal diploid constellation with one maternal and one paternal haploid genome. For instance, uniparental haploid or diploid embryos carry only chromosomes from a single parent, while triploid embryos contain an additional haploid set of chromosomes from one of the parents. Constitutional WG anomalies are generally lethal during the embryonic stage. Triploidy occurs in approximately 1% of all conceptuses, contributing to around 10% of all spontaneous abortions (Hassold et al., 1980; Zaragoza et al., 2000) with extremely rare live born cases surviving less than a year (Sherard et al., 1986). Complete and partial hydatidiform moles, occurring at rates of 1 in 1000 and 3 in 1000 pregnancies respectively, are primarily androgenetic diploid and diandric triploid, respectively (Seckl et al., 2010). Mosaic forms with coexistence of normal cells can result in rare cases of live birth. For example, gynogenetic and androgenetic chimerism (coexistence of uniparental and normal cells) and diploid/triploid mixoploidy (coexistence of triploid and normal cells) have been reported in live-born individuals with congenital abnormalities (Madan, 2020). As a model organism with a developmental process and an aneuploidy profile similar to that of humans (Hansen, 2010; Santos et al., 2014; Tšuiko et al., 2017), WG abnormalities have also been identified at various stages during bovine development (De Coster et al., 2022; Destouni et al., 2016; Dunn et al., 1970; Meinecke et al., 2003; Tšuiko et al., 2017).

Several models have been proposed to explain the mechanistic origins of WG abnormalities in embryos (Masset et al., 2021). These include dispermy or diploid oocytes leading to triploidy, as well as parthenogenetic division of the oocyte or fertilization of an 'empty' egg resulting in uniparental embryos. While these models explain the origin of triploid or uniparental embryos, they are limited by the deduction from surviving cells at later developmental stages. Additionally, the mechanistic origins of chimerism and mixoploidy remain speculative (Madan, 2020; Masset et al., 2021). Since all body cells stem from a single zygote, diverse WG abnormalities observed during the late developmental stages are likely the result of irregular zygotic cleavages. Our recent study with bovine *in vitro* fertilized embryos provides direct evidence of a non-canonical first zygotic division as the mechanistic origin of cell lines exhibiting WG abnormalities (De Coster et al., 2022), which we termed "heterogoneic division". This specialized division involves the segregation of entire parental genomes into distinct

blastomeres, often triggered by polyspermy and occurring concurrently with multipolar zygotic division. Embryos following heterogoneic division typically contain blastomeres with diverse chromosomal constitutions, including biparental diploid, polyploid, and uniparental blastomeres, resembling those identified in chimeric and mixoploid individuals. Such cleavage involving WG segregation errors has also been reported in *in vitro* human and non-human primate embryos (Daughtry et al., 2019; Ottolini et al., 2017).

Heterogoneic division, as the natural origin of blastomeres with WG abnormalities, represents a fundamental yet underexplored aspect of embryo development. While research on preimplantation embryo development often focuses on aneuploidies and their implications (Fernandez Gallardo et al., 2023; Fragouli et al., 2013; Tšuiko et al., 2021), there has been a lack of investigation into the developmental trajectory of spontaneously arising blastomeres carrying WG abnormalities. Examining the subsequent development of blastomeres following both heterogoneic and normal first zygotic divisions will provide valuable insights into the developmental disparities between blastomeres with WG abnormalities and normal blastomeres. Such insights are essential for advancing our understanding of the mechanistic origins of triploidy, moles, chimerism, and mixoploidy identified late during development, and will guide genetic testing and embryo selection strategies regarding WG abnormalities.

Here, we explored the genome constitution and developmental potential of individual blastomeres resulting from both bi- and multipolar first zygotic divisions. With haplotyping results of 118 blastomere outgrowths collected at three preimplantation stages, we demonstrated that uniparental and polyploid blastomeres arise from both multipolar and bipolar cleaving zygotes and exhibit a reduced blastocyst rate. Along with single-cell transcriptome analysis of 446 transcriptomes from 124 blastomere outgrowths, we observed impaired transcriptomic development in blastomere outgrowths carrying WG abnormalities, starting from major embryonic genome activation (EGA). This impaired development is attributed to stress responses induced in blastomere outgrowths with WG abnormalities during EGA. Despite these challenges, some blastomere outgrowths with WG abnormalities successfully managed this stress period and progressed to the blastocyst stage, underscoring their potential significant role in abnormal embryo development.

Results

Whole-genome segregation errors occur in multipolar and bipolar first cleavages

To investigate the developmental program of gynogenetic/androgenetic and polyploid blastomeres following spontaneous heterogoneic division, we exploited the totipotency of blastomeres from early cleavage-stage embryos (Johnson et al., 1995; Willadsen, 1980). Following the first cleavage, 148 blastomeres from 52 bovine embryos (17 with bipolar and 35

with multipolar zygotic cleavage) were dissected and individually cultured to either day 2 (T1, 72 hours post fertilization (hpf), 4-6 cell stage), day 4 (T2, 121 hpf, 4-12 cell stage), or day 6 (T3, 170 hpf, blastocyst stage) post blastomere splitting, respectively **(Figure 1)**.

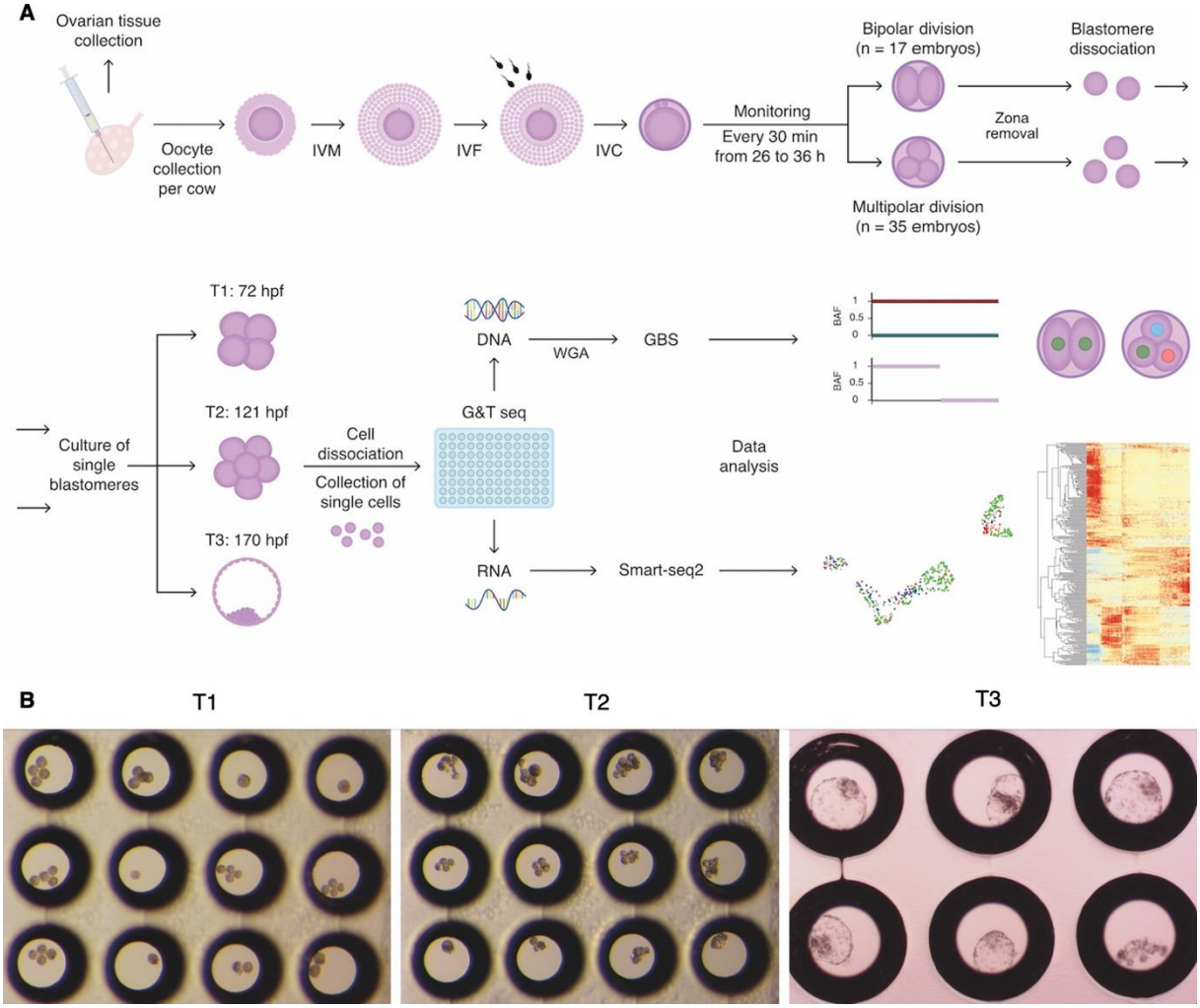


Figure 1. Experimental setup. (A) Diagram illustrating the experimental design. Bovine embryos were created and cultured *in vitro*. After the first zygotic division, blastomeres were dissected and cultured to three preimplantation timepoints. Blastomere outgrowths were subsequently collected and dissected into single cells. Each dissected cell underwent genome and transcriptome separation followed by sequencing. Genome data were used for haplotyping and copy number profiling, while transcriptome data were used for gene expression analysis. Genome and transcriptome sequencing (G&T-seq): Protocol for isolating poly(A) RNA and genomic DNA from single cells. Genotyping-by-sequencing (GBS): Sequencing-based protocol for genome-wide haplotyping and copy number profiling. **(B)** Example images of blastomere outgrowths collected at T1, T2, and T3, respectively. IVM: *In vitro* maturation; IVF: *In vitro* fertilization; IVC: *In vitro* culture; hpf: hours post-fertilization.

To determine the chromosomal constitution of the blastomeres in culture, we performed genome-wide haplotyping and copy-number profiling on cells derived from blastomere outgrowths using genotyping-by-sequencing (GBS) (Masset et al., 2022) **(Figure 1)**. Of 148

blastomeres in culture from the 52 embryos, we successfully determined the chromosomal constitution of 118 blastomeres in culture from 43 embryos through haplotyping. Specifically, the chromosomal constitution was determined for all cultured blastomeres in 31 embryos, for a subset of cultured blastomeres in 12 embryos, and for none of the cultured blastomeres in 9 embryos **(Figure 2A-B)**. Among these, 45% (n=53) were identified as biparental diploid, while 41% (n=48) exhibited WG abnormalities, including 35 androgenetic, 9 gynogenetic, 3 diandric triploid, and 1 triandric tetraploid blastomeres. Additionally, 14% (n=17) of the blastomeres in culture with inferred chromosomal constitution were found to be anuclear. All anuclear blastomeres and the majority (96%, 46 out of 48) of blastomeres with WG abnormalities originated from embryos undergoing multipolar first cleavage **(Figure 2B)**.

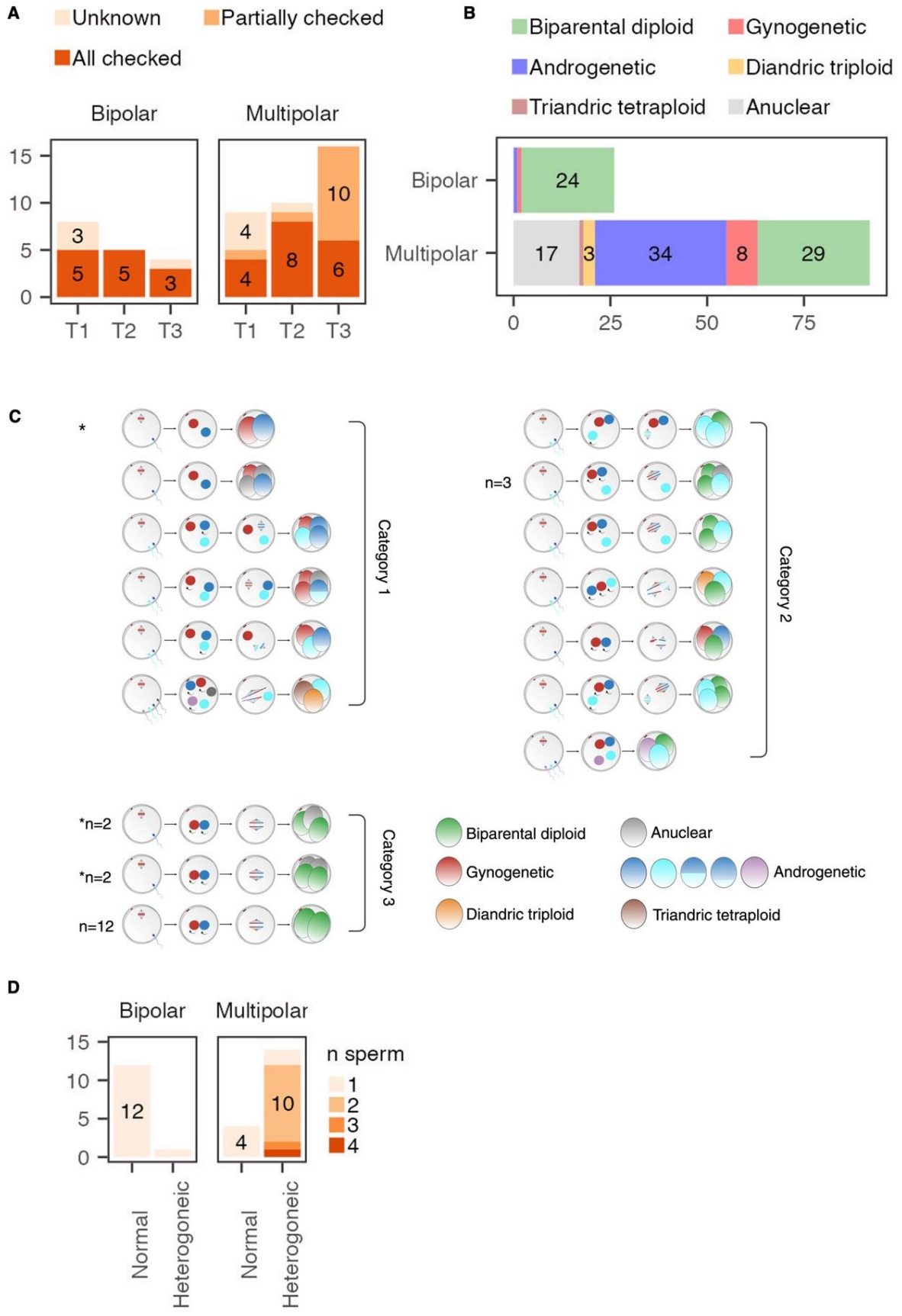


Figure 2. Whole-genome segregation errors occur in multipolar and bipolar first cleavages. (A) Numbers of bipolar and multipolar cleaving embryos used for each blastomere culturing timepoint,

categorized based on the completeness of embryo-derived blastomere outgrowth haplotyping: fully haplotyped (All checked), partially haplotyped (Partial), or not haplotyped (Unknown). **(B)** The inferred chromosomal constitution of 118 blastomeres in culture. **(C)** Three categories of genome segregation patterns identified during the first cleavage. In the oocyte being fertilized, the second meiotic division is depicted, along with the likely number of sperms fertilizing the embryo as deduced from the haplotypes. The arrows surrounding the pronuclei indicate replication occurred. During the first cleavage, spindles segregating the sister chromatids are shown. In the absence of a spindle, the genome is extruded. Numbers are given for patterns that occurred more than once. * in category 1 indicates the one embryo with bipolar first cleavage and WG segregation error. * in category 3 indicates the four embryos with multipolar first cleavage and without WG segregation error. For androgenetic blastomeres, different colors indicate genetic material from different sperm. **(D)** Chromosomal segregation patterns for embryos with bipolar or multipolar first cleavage and the inferred number of sperm that fertilized the egg. Normal: Zygote cleaved without WG segregation error. Heterogoneic: Zygote cleaved with WG segregation error. Bipolar and multipolar refer to embryos undergoing either bipolar or multipolar first cleavage, respectively. Raw GBS results, including the haplarithm plots and aneuploidy profiles from this study, are provided in C4-Additional file 2.

For the 31 fully checked embryos, genome composition of all cultured blastomeres allowed us to reconstruct genome segregation patterns during the first cleavage. We observed three broad categories of genome segregation patterns during the first cleavage **(Figure 2C)**. Category 1 contains 6 embryos with all blastomeres harboring WG abnormalities. Category 2 comprises 9 embryos with a mixture of blastomeres, some with WG abnormalities and others being biparental diploid. Category 3 consists of 16 embryos with only biparental diploid blastomeres. The four multipolar cleaving zygotes in this category also produced anuclear cells. Among the 31 fully examined embryos, 78% (14 out of 18) of those with multipolar first cleavage demonstrated heterogoneic division, while only 8% (1 out of 13) of embryos with bipolar first cleavage showed WG segregation errors (p=0.00048) **(Figure 2C and 2D)**, confirming the enrichment of heterogoneic division in multipolar cleaving zygotes (De Coster et al., 2022). Furthermore, the high frequency of polyspermic conceptions in embryos undergoing multipolar first zygotic division and exhibiting WG segregation errors (De Coster et al., 2022) was validated, with 86% (12 out of 14) of these embryos inferred to be fertilized by more than one sperm. In contrast, all other embryos with bipolar first cleavage or with multipolar first cleavage but without WG segregation errors were inferred to be fertilized by a single sperm **(Figure 2C and 2D)**. Only 10% (1 out of 10) of dispermic embryos give rise to triploid blastomeres, which is consistent with previous observations that dispermic fertilization seldom leads to triploid development (De Coster et al., 2022; Kola et al., 1987). Most of the observed mechanisms driving heterogoneic division align with those from our previous study (De Coster et al., 2022; Destouni et al., 2016). Among the 15 embryos undergoing heterogoneic division, 10 were inferred to involve the extrusion of paternal pronuclei into separate cells, 4 likely involved an

independent spindle separating paternal or maternal chromosomes, 2 likely involved a tripolar spindle, and 1 likely involved double spindles **(Figure 2C)**.

Interestingly, two novel genome segregation mechanisms were identified. Firstly, within category 1, we observed an embryo undergoing bipolar first cleavage yet exhibiting heterogoneic division **(Figure 2C)**. The most plausible explanation is that the two parental zygotic spindles failed to align and instead segregated into distinct blastomeres. Hence, genome-wide segregation errors are not confined to multipolar cleavers. Additionally, within category 3, we identified four embryos undergoing multipolar first cleavage yet exhibiting normal genome segregation. Specifically, two biparental diploid blastomeres were generated alongside one or two anuclear cells **(Figure 2C)**. These findings suggest that bipolar first cleavage does not always ensure normal genome segregation, and multipolar first cleavage does not inevitably result in WG or chromosomal segregation errors.

Blastomeres with whole-genome abnormalities can reach blastocyst stage despite impaired developmental potential

To investigate the developmental potential of heterogoneic division-derived blastomeres with WG abnormalities, we compared the developmental stage of the 53 biparental diploid blastomere outgrowths and 48 blastomere outgrowths with WG abnormalities at the time of outgrowth collection **(Figure 3A-B)**. We observed impaired development for blastomeres with WG abnormalities compared to biparental diploid blastomeres at T3, with 80% (12 out of 15) of biparental diploid blastomere outgrowths developing to blastocysts, while only 21% (6 out of 28) of blastomere outgrowths with WG abnormalities reached this stage (p=0.0016) **(Figure 3B-C)**.

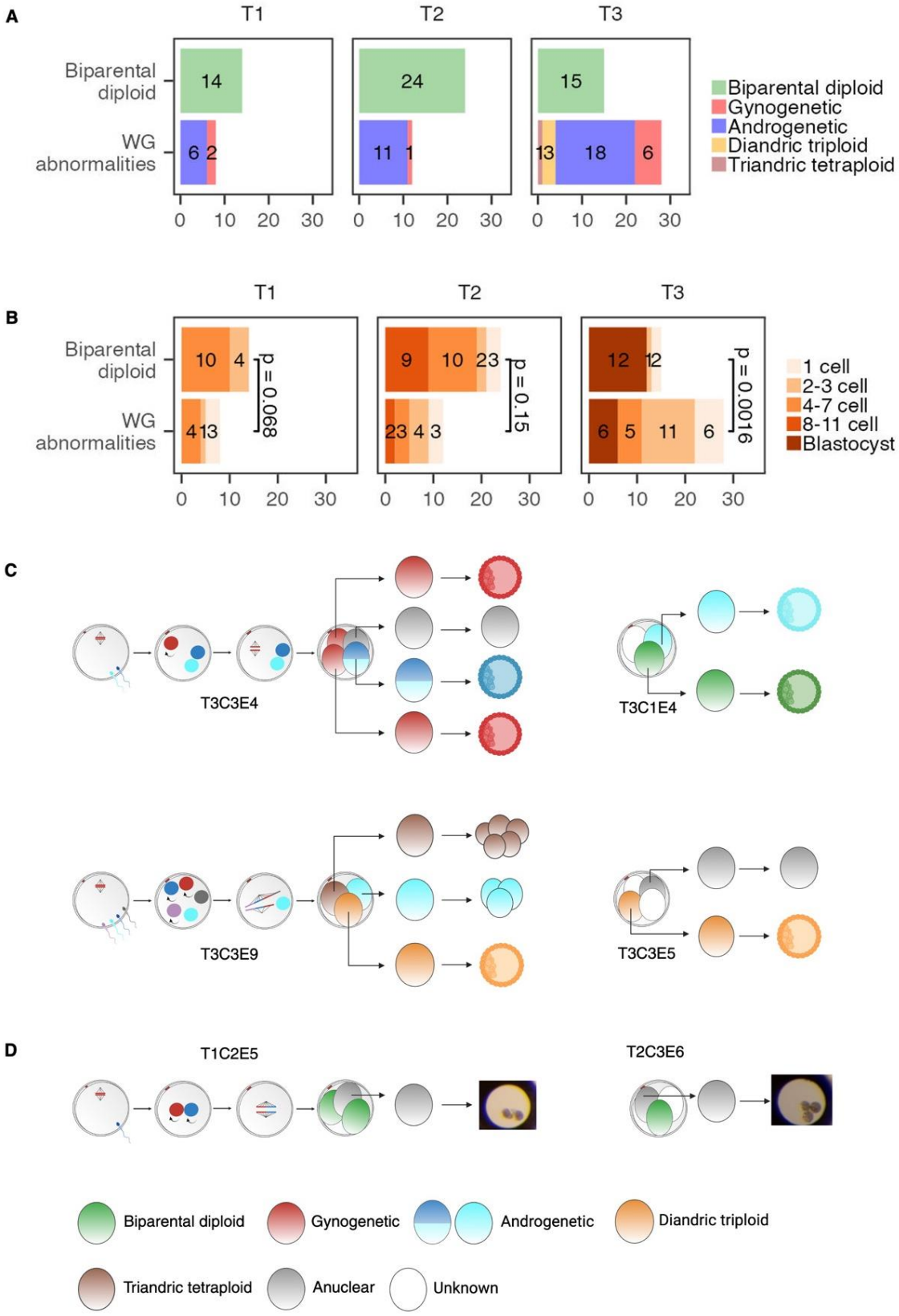


Figure 3. Blastomeres with whole-genome abnormalities can reach blastocyst stage despite impaired developmental potential. (A) Distribution of 53 biparental diploid blastomere outgrowths and 48 blastomere outgrowths with WG abnormalities across different timepoints of outgrowth collection. **(B)** Developmental stage comparison of biparental diploid blastomere outgrowths and blastomeres

outgrowths with WG abnormalities in (A). **(C)** Schematic representation of the six blastomere outgrowths with WG abnormalities that reached the blastocyst stage at T3, along with their sibling blastomere outgrowths and original embryos. **(D)** Images of the two cleaved anuclear blastomere outgrowths and schematic diagrams illustrating the embryos from which they originated. In (C) and (D), the same symbols as in Figure 2C are employed. Fertilization and the first cleavage are not depicted for embryos containing blastomeres with an unknown haplotype (due to sample loss or haplotype failure).

Since aneuploidy is known to affect embryo development, we next evaluated the impact of heterogoneic division on aneuploidy levels in daughter blastomeres. By comparing the aneuploidy spectrum of blastomeres from both normal genome segregation and heterogoneic division, we observed a lower proportion of euploid blastomeres and a higher proportion of blastomeres with high-level aneuploidy (defined as >10% of autosome regions being aneuploid, with 10% corresponding to two large chromosomes) in blastomeres from heterogoneic division, compared to blastomeres originating from zygotes with normal genome segregation. Specifically, for blastomeres with WG abnormalities resulting from heterogoneic division, 32% (10 out of 31) were euploid, while 55% (17 out of 31) displayed high level aneuploidy. In contrast, for blastomeres from normal genome segregation, 56% (18 out of 32) were euploid, and only 3% (1 out of 32) exhibited high level aneuploidy (p=6.3×10^{-6}). Similarly, other studies demonstrated the increased rate of complex chromosomal abnormalities for embryos with multipolar division (Daughtry et al., 2019; Zhan et al., 2016). The association between heterogoneic division and vulnerability to chaotic chromosomal missegregations might be caused by the spindle abnormalities implicated in this process (Chatzimeletiou et al., 2005) **(Figure 2C)**. To control for the effect of aneuploidy on blastomere development, we compared the development of only euploid blastomeres. At T3, all biparental diploid blastomere outgrowths (100%,12 out of 12) reached blastocyst stage, while a much lower proportion (36%, 4 out of 11) of blastomere outgrowths with WG abnormalities reached blastocyst stage (p=0.0013). Additionally, for androgenetic blastomere outgrowths collected at T3, 2 out of 15 containing the X chromosome reached the blastocyst stage, while the 3 without the X chromosome did not, indicating that the Y chromosome alone is insufficient for blastocyst formation.

The majority of the 17 anuclear blastomeres arrested at the 1-cell stage. However, we identified 2 exceptions: one reached the 2-cell stage at T1, and the other reached the 3-cell stage at T2 **(Figure 3D)**, suggesting that the cytoplasmic divisions could occur without a nucleus. Additionally, we identified 16 more anuclear biopsies from 14 non-anuclear blastomere outgrowths at various stages of development, with the majority of the original blastomeres (12 out of 14) from embryos underwent multipolar first cleavage. Hence, the formation of anuclear blastomeres is not restricted to the first zygotic division. To confirm the origin of these empty outgrowths/biopsies, we examined the mitochondrial genotypes for empty biopsies obtained

from the same cows. Maternal inheritance of the mitochondrial genome was corroborated by similar mitochondrial genotypes.

To conclude, blastomeres with WG abnormalities from heterogoneic division show diminished development. Despite this, certain uniparental/polyploid blastomeres still progress to the blastocyst stage **(Figure 3C)**, indicating their viability during the preimplantation stage and raising concerns about their further development and potential contribution to abnormal pregnancies.

Whole-genome abnormalities do not alter the preimplantation developmental program but hinder transcriptomic development

We hypothesized that the impaired development of blastomere outgrowths with WG abnormalities would be reflected in the transcriptomic profiles of constituent cells, with WG abnormalities potentially causing deviations in the preimplantation developmental program. To investigate the effects of WG abnormalities on transcriptomic development, we conducted single-cell transcriptome sequencing on 677 cells derived from collected blastomere outgrowths. Among these, 446 transcriptomes met our quality control criteria, 397 of which had corresponding haplotyping information **(Methods)**. Before quality control, we noted significantly higher fractions of mitochondrial transcripts in cells from blastocyst stage outgrowths compared to cells from other stages, except the 1-cell stage **(Supplementary Figure 1)**. The blastocyst stage cells with high fraction of mitochondrial transcripts were not enriched for cells with WG abnormalities. This finding suggests active mitochondrial gene expression in blastocysts and potential cytoplasmic RNA leakage in arrested 1-cell stage outgrowths. Consistent with our findings, previous studies utilizing PCR analysis have demonstrated a significant increase in mitochondrial transcripts at morula and blastocyst stages across various mammalian species (Ma et al., 2008; May-Panloup et al., 2005; Thundathil et al., 2005).

Following dimensionality reduction with Unified Manifold Approximation and Projection (UMAP), single-cell gene expression profiles exhibited a developmental trajectory from T1 to T3 **(Figure 4A)**. Cells with WG abnormalities did not form distinct clusters **(Figure 4B).** Instead, by checking the expression of marker genes for essential preimplantation stages, we identified clusters corresponding to minor EGA, major EGA, inner cell mass (ICM), and trophectoderm (TE) stages, which we define as the molecular developmental stages of the cells. The cell cluster without EGA (pre-EGA) contained cells arrested before EGA **(Figure 4C)**. Pseudotime analysis with single-cell transcriptome data inferred developmental trajectories pointing to ICM and TE **(Figure 4D)**. These results indicate that transcriptomic changes are mainly driven by the natural developmental program rather than variations in genome constitution, aligning with

previous single-cell transcriptome analyses of uniparental (Leng et al., 2019) and aneuploid (Fernandez Gallardo et al., 2023) cells from human preimplantation embryos.

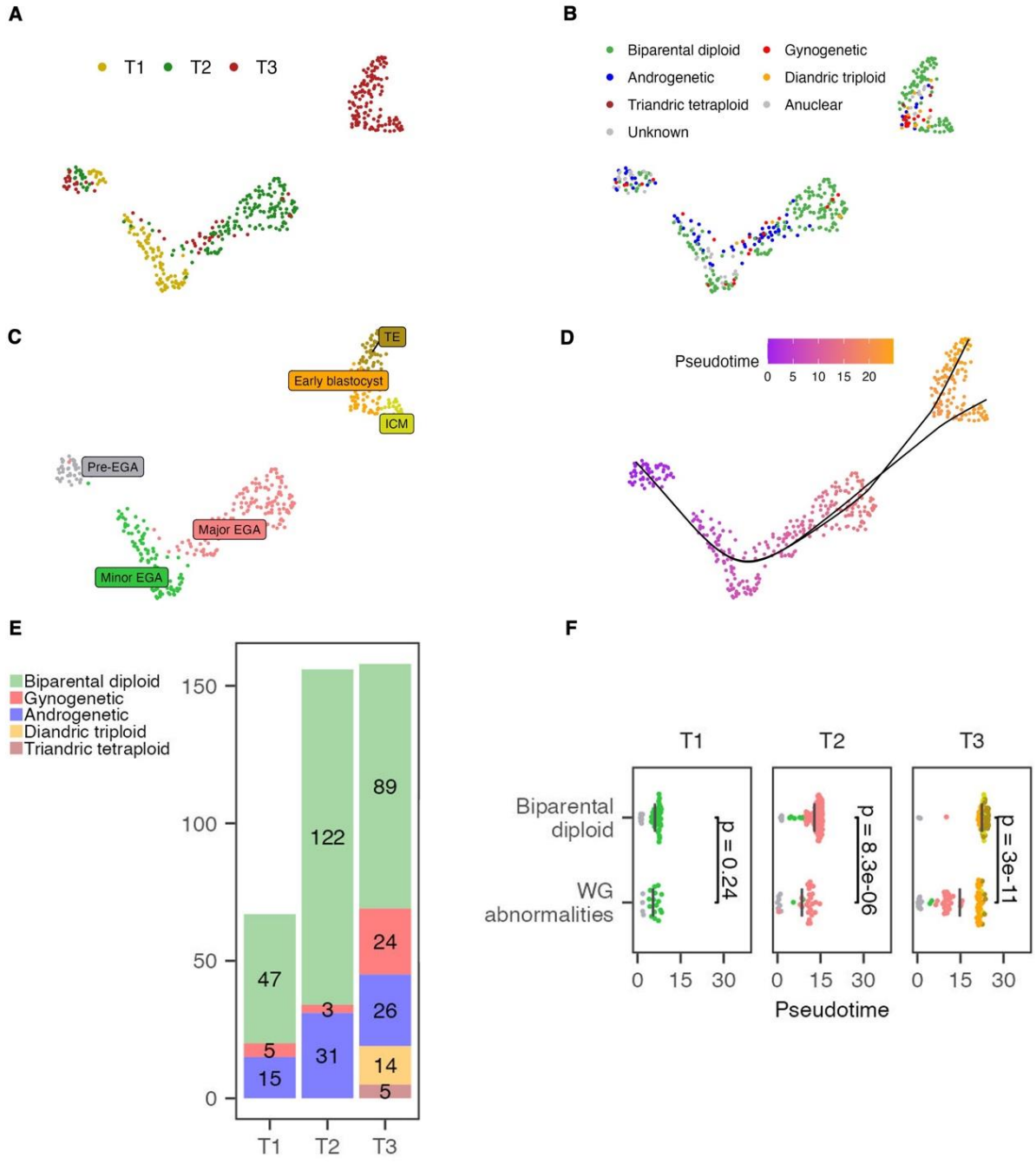


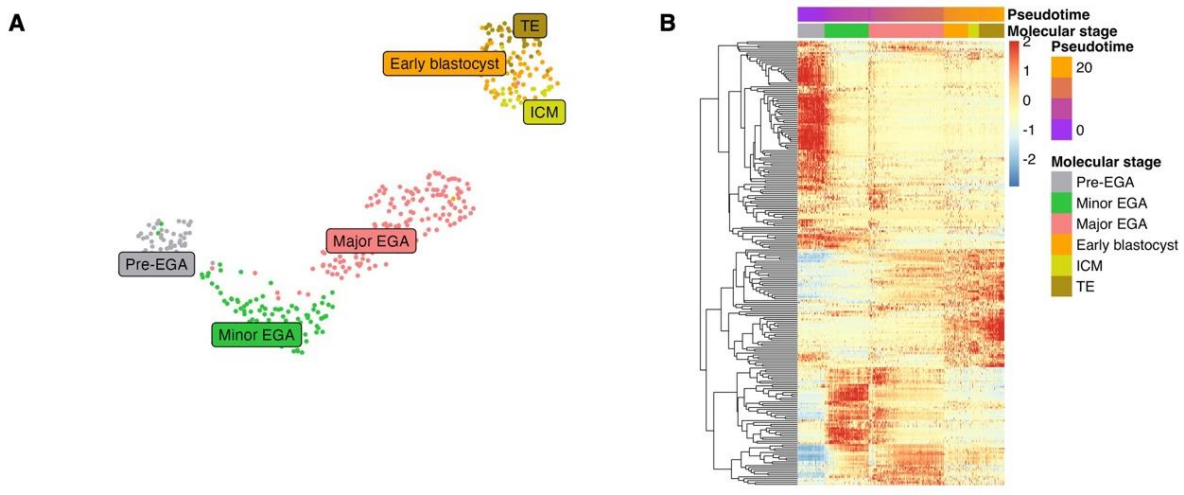
Figure 4. Whole-genome abnormalities do not alter the preimplantation developmental program but hinder transcriptomic development. UMAP for single-cell transcriptome data with cells colored by **(A)** time of outgrowth collection **(B)** genome constitution **(C)** molecular developmental stage indicated by marker genes expression **(D)** pseudotime value inferred from trajectory analysis. The inferred pseudotime trajectories are indicated by black curved lines on the UMAP. **(E)** Number of biparental diploid cells and cells with WG abnormalities used for transcriptome analysis for each timepoint. **(F)** Pseudotime comparison of biparental diploid cells and cells with WG abnormalities at each outgrowth collection timepoint. Each dot represents one cell, with the x-axis indicating its pseudotime value

extracted from (D). Cells are colored according to their corresponding molecular developmental stage in (C) using the same color scheme. Vertical grey bars indicate the mean pseudotime value of each group.

We then explored whether the impaired development of blastomeres exhibiting WG abnormalities was reflected in single-cell transcriptome profiles. We characterized the physical age of the cells based on the time of outgrowth collection, while their molecular age was assessed using pseudotime values along the transcriptional trajectory. Compared to cells from biparental diploid outgrowths, those from outgrowths displaying WG abnormalities exhibited impaired development at T2 and T3, as evidenced by their smaller transcriptome-pseudotime ages (**Figure 4E-F**; $p=4.5\times10^{-5}$ for T2 and $p=4\times10^{-10}$ for T3). This trend persisted when considering only cells inferred to be euploid ($p=0.00026$ for T2 and $p=0.039$ for T3). Our observation indicates the cell-level impaired transcriptomic development of blastomeres with WG abnormalities starting form T2, coinciding with major EGA.

Some blastomere outgrowths collected at T2 or T3 did not have the expected number of cells and appeared to be cleaving slower or blocked in their development. Surprisingly, when mapping the transcriptome profiles on the developmental trajectories, we observed that some cells from 1 cell and 2-3 cell stage outgrowths at T3 displayed profiles resembling major EGA, while some cells from 4-7 cell outgrowths exhibited profiles akin to TE or ICM **(Supplementary Figure 2)**. These observations suggest that transcriptomic development can continue in the absence of cell divisions.

Distinct preimplantation cellular states are governed by specific key transcription factors



C

Molecular stage	Selected key regulons
Pre-EGA	SIX3, PRDM5, ZBTB43, SMC3, EBF1, HES7
Minor EGA	KLF17, MAX, ZNF143, THAP11, CHD1, IRF3
Major EGA	MITF, ARNT, RARB, HSF1, USF1, BRCA1
Early blastocyst	GATA2, TBX3
ICM	SOX15, NELFB, SREBF1, GATA6, ERF, MAZ
TE	CDX2, TCF7L1, JDP2, JUN, GATA2, GATA3

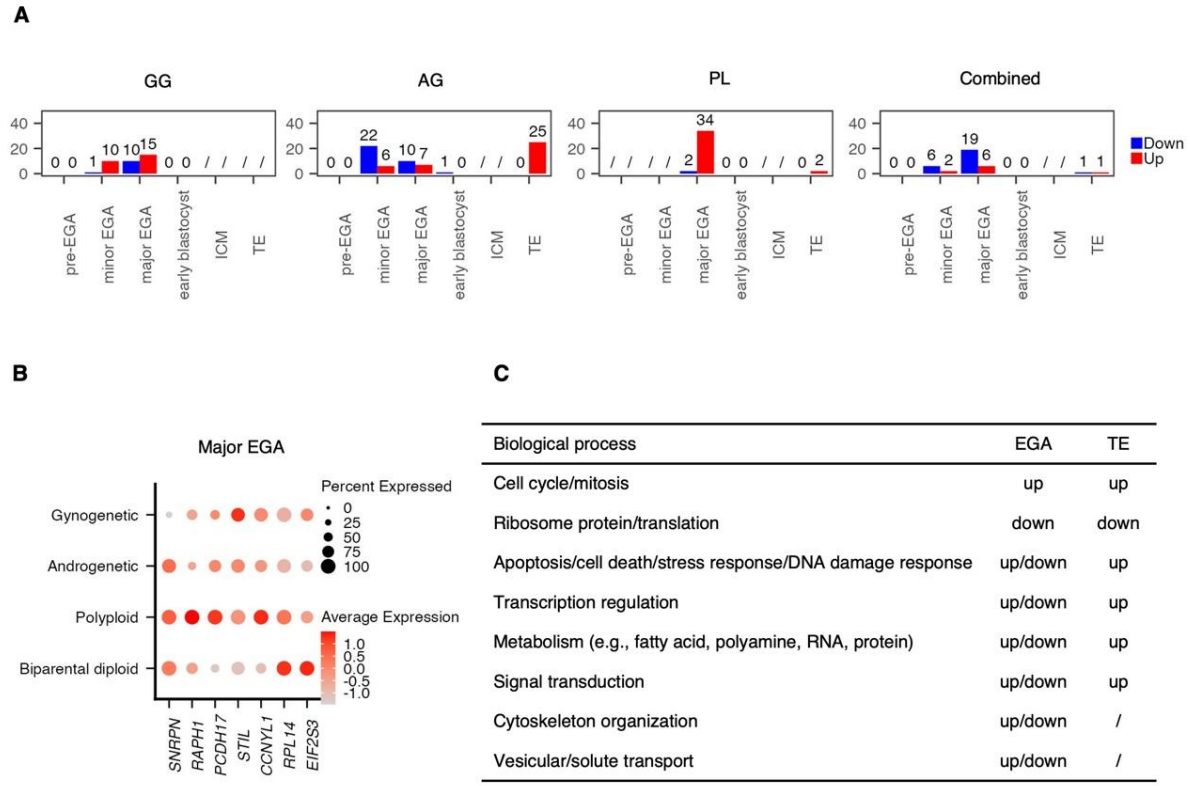
Figure 5. Distinct preimplantation cellular states are governed by specific key transcription factors. (A) UMAP displaying cell clustering based on regulon activity, with cells colored according to their macular developmental stage as depicted in Figure 4C. **(B)** Heat map of regulon activity. Each row represents one regulon and each column represent one cell. The plot is generated using rescaled Area Under the Curve (AUC) values, with cells ordered according to their molecular developmental stage and pseudotime values. **(C)** Key regulons selected for each stage.

Since cells were clustered according to their molecular developmental stage regardless of genome composition, we next aimed to uncover the specific gene regulatory networks (GRNs) shaping cellular identity at each stage. We used pySCENIC to infer GRNs from single-cell RNA-seq data, based on co-expression of TFs and their target genes as well as cis-regulatory motif analysis. This approach identified 237 regulons, each comprising a transcription factor (TF) alongside its predicted target genes. Subsequently, we generated a UMAP visualization based on regulon activity, partitioning cells into distinct regulatory states **(Figure 5A)**. As anticipated, cells belonging to same molecular stage exhibited clustering, indicative of stage-specific regulatory dynamics. To further investigate the coactivity of TF combinations within each stage, we ordered cells based on their molecular stage and pseudotime values on regulon activity heatmap. The resulting plot revealed stage-specific regulatory states characterized by the activation or repression of specific TFs **(Figure 5B)**. We then selected key regulons with both high activity and specificity for each stage **(Figure 5C;**

Methods). Most of the TFs for selected key regulons have been shown to be crucial for embryonic development. For instance, GATA6 has demonstrated significance in the ICM of mouse, bovine, and human embryos (Marsico et al., 2023; Niakan and Eggan, 2013). Similarly, CDX2, GATA2 and GATA3 have been implicated in the TE of mouse, bovine and human embryos (Fernandez Gallardo et al., 2023; Gerri et al., 2020; Home et al., 2017; Nagatomo et al., 2013; Negrón-Pérez et al., 2017; Niakan and Eggan, 2013). Interestingly, for the pre-EGA stage, PRDM5 was identified as a key regulon, reflecting the arrested characteristics of these cells, as this TF is involved in G2/M arrest and apoptosis through suppressing the expression of several oncogenes and antagonizing WNT/β-catenin signaling (Deng and Huang, 2004; Shu et al., 2011). These findings shed light on the combinations of TFs that underlie cell state transitions and cellular identity during bovine preimplantation embryogenesis.

Whole-genome abnormalities induce stress responses during embryonic genome activation

Although WG abnormalities did not result in obvious deviations in the preimplantation developmental program, we hypothesized that the imbalanced parental genome compositions in cells with WG abnormalities could lead to parent-specific gene expression patterns for genes predominantly expressed from only one parent's genome. Furthermore, these WG abnormalities may function as stress stimuli, inducing global transcriptomic changes that ultimately impair the overall developmental potential of cells. To investigate the parental genome-specific effects on gene expression, we compared the gene expression profiles of gynogenetic, androgenetic, and polyploid (diandric triploid and triandric tetraploid) cells to those of biparental diploid cells within each molecular developmental stage. Additionally, to capture the overall effects of WG abnormalities on gene expression, we aggregated all cells with WG abnormalities for combined analyses.



C

Biological process	EGA	TE
Cell cycle/mitosis	up	up
Ribosome protein/translation	down	down
Apoptosis/cell death/stress response/DNA damage response	up/down	up
Transcription regulation	up/down	up
Metabolism (e.g., fatty acid, polyamine, RNA, protein)	up/down	up
Signal transduction	up/down	up
Cytoskeleton organization	up/down	/
Vesicular/solute transport	up/down	/

Figure 6. Whole-genome abnormalities induce stress responses during embryonic genome activation. (A) Numbers of DE genes identified in gynogenetic (GG), androgenetic (AG), polyploid (PL) and combined WG-aberrant cells compared to biparental diploid cells for each stage. The "/" symbol denotes comparisons not conducted due to insufficient cells. **(B)** Gene expression dot plot for example DE genes during major EGA. **(C)** Overall alterations in biological processes reflected by deviations in gene expression. "/" if no DE genes observed.

Differentially expressed (DE) genes were identified from the onset of EGA **(Figure 6A)**, suggesting that WG abnormalities lead to gene expression changes during EGA. Only one DE gene was identified at the early blastocyst stage, indicating a high degree of similarity between normal cells and cells with WG abnormalities just before the first lineage specification. DE genes found in the combined analysis partially overlap with those identified in separate analyses. Manual functional annotation revealed some parental genome specific effects. We noted the downregulation of the maternally imprinted gene *SNRPN* in gynogenetic cells during major EGA, aligning with its known paternal-specific expression (Lucifero et al., 2006). Furthermore, we identified the up regulation of genes involved in cell adhesion (e.g., *RAPH1, PCDH17*) in polyploid cells during major EGA and at the TE stage **(Figure 6B)**. This indicates active cell adhesion-related processes, such as compaction in these cells, possibly caused by the additional paternal genome contribution.

Other DE genes observed in uniparental/polyploid cells, as well as those identified through combined analysis, are mainly involved in similar biological processes, indicating global effects

of WG abnormalities **(Figure 6C)**. We observed upregulation of cell cycle/mitosis-related genes (e.g., *STIL*, *CCNYL1*), indicating active preparation for cell cycle progression in cells with WG abnormalities. Meanwhile, we noted downregulation of ribosomal protein genes (e.g., *RPL14*) and translation initiation factor genes (e.g., *EIF2S3*). It is known that during cellular adaptation to stress stimuli, growth-related genes, including those involved in ribosomes and translation, are typically repressed, prioritizing resources towards stress protection over rapid proliferation (López-Maury et al., 2008). Our observation indicates reduced growth in cells with WG abnormalities as a consequence of the stress response triggered by WG abnormalities. Interestingly, for certain biological processes, DE genes were identified during EGA, leading to either upregulation or downregulation of these processes. No relevant DE genes were found for the early blastocyst stage, while in the TE, no DE genes or only DE genes resulting in upregulated activities of these processes were observed. The biological processes involved include apoptosis, cell death, stress response, DNA damage response, transcription regulation, metabolism, single transduction, cytoskeleton organization and transport **(Figure 6C).** This indicates that cells with WG abnormalities adapt to stimuli caused by genome imbalance through transcriptional responses during EGA, involving transient upregulation or downregulation of relevant genes (López-Maury et al., 2008). A quiescent period follows during the early blastocyst stage, while the upregulation of specific genes in TE reflects establishment of new steady-state levels. Additionally, we conducted differential regulon analysis and DE regulons were identified for cells with WG abnormalities only during EGA. Most DE regulons are essential for embryonic development and their bidirectional changes mirror bidirectional gene expression alterations observed in differential gene expression analysis.

In summary, we observed parental genome-specific effects induced by WG abnormalities, suggesting distinct roles of the maternal and paternal genomes. Moreover, we identified universal cellular responses to WG abnormalities, marked by active preparation for cell cycle progression alongside reduced cell growth. Bidirectional alterations in specific biological processes and regulons during EGA, and their eventual balance, may dictate whether cells arrest or advance towards the blastocyst stage. Overall, these observed stress responses provide insights into the impaired development cause by WG abnormalities.

Discussion

WG abnormalities are a type of chromosomal error frequently observed in spontaneous abortions (Hassold et al., 1980) and on very rare occasions in a chimeric state in live-born individuals (Madan, 2020). Contrary to its implications in abnormal pregnancy and congenital abnormalities, the proposed origins of such anomalies are deductions from surviving cells at later stages or remain speculative, and their effect on early development remains largely unknown. Here, we provide fundamental insights into the natural emergence and early

developmental biology of gynogenetic, androgenetic, and polyploid blastomeres using haplotyping and single-cell transcriptome analysis. We demonstrated the spontaneous occurrence of uniparental and polyploid blastomeres through a non-canonical division during the first zygotic cleavage, which segregates parental genomes into distinct blastomeres. We not only confirmed our previous observation (De Coster et al., 2022) that heterogoneic division mostly coincides with multipolar first cleavage in polyspermic embryos, but also identified, for the first time, heterogoneic division occurring with bipolar first cleavage in normally fertilized embryos. The reduced developmental potential of uniparental and polyploid blastomeres derived from heterogoneic division was evidenced by a lower blastocyst rate and a delay in pseudotime values beginning from T2, coinciding with major EGA. Stress responses observed during EGA suggest the cells' efforts to combat fitness challenges induced by aberrant genome composition, with the outcome determining whether they arrest or progress to the blastocyst stage.

Although to our knowledge we are the first to explore the development of first-cleavage-derived uniparental and polyploid blastomeres, some studies have examined the preimplantation development of mammalian uniparental embryos. These embryos were generated through various methods, including nuclear manipulation (Hoppe and Illmensee, 1977; Leng et al., 2019; Xu et al., 2021), *in vitro* induction (Hao et al., 2004; Hwang et al., 2020; Ito et al., 1991; Leng et al., 2019; Liu et al., 2002) or natural occurrence (Hardy et al., 1989) of parthenogenetic development in unfertilized oocytes, as well as bisection of one-cell fertilized eggs (Tarkowski, 1977). Aligned with the impaired developmental potential observed in blastomeres with WG abnormalities in our study, those studies on preimplantation uniparental embryos have shown a decrease in the morula/blastocyst rate compared to biparental diploid embryos., alongside poorer morphology and reduced cell count in resulting blastocysts (Hardy et al., 1989; Leng et al., 2019). Several studies have monitored post-implantation development and affirmed the compatibility of diploid uniparental genome constitution with later development. For instance, porcine parthenogenetic fetuses can develop until day 35 with abnormal organ morphologies (Hwang et al., 2020). Additionally, live-born fertile mice have been created via both androgenesis and gynogenesis (Hoppe and Illmensee, 1977). Moreover, diploid gynogenetic cells have been identified in various organs of live-born chimeric mice developed from aggregated chimeric embryos (Ito et al., 1991). These findings are consistent with observations in humans, suggesting that uniparental cell lines are typically selected against during development but can persist until late pregnancy and even be compatible with live birth, raising various concerns. Our study demonstrates that heterogoneic division could serve as a spontaneous origin of uniparental and polyploid cell lines observed during late development.

The parental genome-specific effects identified in this study correlate with findings from a prior

investigation of uniparental human embryos (Leng et al., 2019). As for the stress response observed solely in our study, there are possible causes aside from species-specific effects and sample variations (Xu et al., 2021). The referenced study created diploid gynogenetic and androgenetic embryos through nuclear manipulation and induction of parthenogenesis, respectively. Only embryos demonstrating good morphology and developmental speed were utilized. Only XX and XY androgenetic embryos were included. And their analysis focused exclusively on genes with highly biased expression in gynogenetic or androgenetic embryos. In contrast, our study focused on a broad spectrum of spontaneously occurring WG abnormalities, which frequently exhibit complex aneuploidy and contain either haploid or diploid genome constitutions. While haploidy has been demonstrated to convert frequently to diploidy during development (Ito et al., 1991; Leng et al., 2017; Tarkowski, 1977), both complex aneuploidy and haploidy are known to impact preimplantation development (Liu et al., 2002; McCoy et al., 2015). Additionally, our cohort included androgenetic blastomeres lacking the X chromosome, a condition known to cause embryonic arrest. Finally, rather than focusing solely on genes with highly biased expression in gynogenetic or androgenetic embryos, our study examined the broader impact of WG abnormalities on global gene expression changes. These factors led us to observe EGA as a critical period for embryos with WG abnormalities. During this stage, their constituent cells exhibit stress responses in gene expression, which could determine whether they arrest or successfully progress to the blastocyst stage.

We identified anuclear cells from outgrowth at different stages, which is in agreement with previous observations in rhesus (Daughtry et al., 2019) and human embryos (Hardy et al., 1993). Consistent with our previous analysis (De Coster et al., 2022), most of the anuclear blastomeres were identified from blastomere outgrowths following multipolar first cleavage, suggesting a potential association between the first cleavage pattern and the origin of these anuclear cells. In addition, we demonstrate that some of the anuclear cells can cleave without nuclei. Cytoplasmic divisions are generally thought to rely on nuclear divisions and mitotic signals. However, a recent study by Bakshi et al. observed that cytoplasm can compartmentalize, and a rare fraction of the compartments can divide repeatedly without nuclei and independent of mitotic CDK/cyclin complexes during normal drosophila embryogenesis (Bakshi et al., 2023). They showed that this phenomenon is relevant to mitotically delayed nuclei. We hypothesize that a similar mechanism may apply here, with the anuclear cells derived from mother cells containing nuclei with delayed mitosis. These findings collectively underscore the need for further investigation into the detailed mechanisms underlying the presence of anuclear cells in mammalian embryos and their potential roles in embryo development.

Insights from this study raise clinical concerns. Firstly, in IVF laboratories, human embryos with

two pronuclei (PN) are preferentially selected for transfer (Balaban et al., 2011). We demonstrate here that zygotes with three PN may result in a mixture of biparental diploid and uniparental/polyploid blastomeres. Since the latter are developmentally compromised, the resulting embryo may still develop normally. Secondly, there is a growing trend in many laboratories to monitor embryonic development using time-lapse microscopy and rank embryos based on cleavage kinetics (Balaban et al., 2011; Sigalos et al., 2016; Sugimura et al., 2017, 2012). Embryos with multipolar first cleavage are typically given a low rank. We showed here that 22% of embryos with a multipolar first cleavage exhibited normal, equal division of the genome with additional anuclear cells. It would be unjust to low rank these embryos since they have the potential to lead to healthy babies. Additionally, even though uniparental/polyploid cells are normally generated with multipolar first cleavage, it is not uncommon that biparental diploid blastomeres are generated together, which may gain developmental advantage and lead to normal development. Thirdly, embryos typically ranked high, characterized by a normal pronuclei count and bipolar first cleavage, may still experience WG segregation errors, leading to the generation of solely uniparental cells, posing a potential risk of abnormal pregnancy. Rather than removing or keeping embryos based on apparent abnormalities, it might be better to evaluate the genome constitution of preimplantation embryos and deselect those with WG abnormalities, while retaining those with a normal chromosome count and parental haplotype constitution. Preimplantation genetic testing for aneuploidy (PGT-A) has become a common practice to select embryos with a normal number of chromosomes for transfer. However, the majority of PGT-A tests only measure aneuploidies, with only some measuring the parental haplotype constitution (Coonen et al., 2020). We suggest the use of genotyping/haplotyping together with PGT-A (Masset et al., 2022; Zamani Esteki et al., 2015) to check both chromosome count and parental haplotype constitution and select against WG abnormalities. This approach is likely to further improve the overall IVF success rate (Caroselli et al., 2023).

It is important to note that our study used bovine embryos, and caution is needed when extrapolating the results to humans. Tripolar mitoses in monospermic human embryos have been observed to cause genome partitioning, resulting in daughter cells with sub-diploid profiles containing mixed maternal and paternal chromosomes (Ottolini et al., 2017). In contrast, the two monospermic embryos with multipolar first cleavage in our study resulted in uniparental and anuclear cells, or uniparental and biparental cells. The uniparental cells observed suggest heterogoneic division rather than random genome partitioning. Although similar to Ottolini et al., we observed a high rate of concurrent aneuploidies. Further studies in human preimplantation embryos are necessary to understand the extent of heterogoneic division in human embryos and its developmental effects.

In conclusion, this study provides comprehensive insights into the natural origin and early development of uniparental/polyploid blastomeres. We demonstrated their overall impaired developmental potential, primarily due to stress responses triggered by WG abnormalities during EGA. Blastomeres that successfully navigate through this phase and progress to the blastocyst stage may constitute a significant fraction of embryos and contribute to abnormal development. Our findings underscore the importance of routine screening against embryos bearing WG abnormalities in IVF labs.

Materials and methods

Single-cell collection from bovine blastomere outgrowths

In vitro embryo production

Prior ethical agreement was not necessary. Ovaries from cows raging 4 to 5 years were collected post-mortem in a commercial slaughterhouse. The sperm used for *in vitro* fertilization was collected from a 5-year-bull. Standard *in vitro* procedures were followed to produce bovine embryos (Wydooghe et al., 2014). Briefly, bovine (*Bos taurus*) ovaries were collected and processed within 2 h. The ovaries were rinsed three times in warm physiological saline supplemented with 0.25µg/mL kanamycin. Using an 18-G needle and a 10-mL syringe, antral follicles (2 – 8 mm diameter) were punctured and kept separately per ovary in 2.5 mL Hepes-Tyrode's albumin-pyruvate-lactate (Hepes-TALP). Using a stereomicroscope, cumulus-oocyte complexes were collected and washed in Hepes-TALP and then in maturation medium (modified bicarbonate-buffered TCM-199 supplemented with 50 ppm gentamicin and 20 ng/mL epidermal growth factor). Cumulus-oocyte complexes were *in vitro* matured per donor in four-well dishes (NuncTM) in 500 µL maturation medium for 22 h at 38.5°C in 5% CO_2 in humidified air. *In vitro* fertilization was performed with frozen-thawed semen from a Holstein-Friesian bull (*Bos taurus*) after selection over a discontinuous 45/90% Percoll® gradient (GE Healthcare Biosciences, Uppsala, Sweden). The mature oocytes were fertilized by incubating them with spermatozoa at a concentration of 1×10^6 spermatozoa/mL in IVF–TALP medium enriched with BSA (Sigma A8806; 6 mg/mL) and heparin (20 µg/mL) for 21 hours at 38.5°C, 5% CO_2 in humidified air. After fertilization, the presumed zygotes were vortexed in 2.5 mL HEPES-TALP for 3 minutes to remove the cumulus and sperm cells adhered to the *zona pellucida* and subsequently transferred to 50 µL droplets of synthetic oviductal fluid (SOF), enriched with 4 mg/mL BSA (Sigma A9647), non-essential and essential amino acids (SOFaa), 5 µg/mL insulin, 5 µg/mL transferrin, and 5 ng/mL selenium. The droplets were covered with 900 µL paraffin oil (SAGE oil for tissue culture, ART-4008-5P, Cooper Surgical Company). *In vitro* culture was performed at 38.5 °C in 5% CO_2, 5% O_2, and 90% N_2.

Blastomere dissociation and culture

The presumptive zygotes were monitored from 26 to 36 hours post-fertilization (hpf) every 30 minutes to identify a direct cleavage of the zygote into three or four blastomeres (multipolar division) or into two blastomeres (bipolar division). Blastomeres from embryos that cleaved into more than four cells or presented multiple fragments were not processed further. Immediately upon the first division, embryos were washed in Hepes-TALP and treated with 0.1% pronase (protease from *S. griseus*) in TCM-199 to dissolve the *zona pellucida*. Next, the embryos were washed in TCM-199 with 10% FBS and then transferred to Ca^{+2}/Mg^{+2}-free PBS with 0.05% BSA to enhance blastomere dissociation, which was performed using a STRIPPER pipet holder and 170 µm and 135 µm capillaries (Origio, Cooper Surgical, CT, US) in Ca^{+2}/Mg^{+2}-free PBS supplemented with 0.1% polyvinylpyrrolidone (PVP). Subsequently, single blastomeres were washed in culture medium and transferred individually to one well of a Primo VisionTM micro well group culture dish (Vitrolife, Göteborg, Sweden), which contained a total of 16 small wells covered by a 40 µL droplet of culture medium and 3.5 mL paraffin oil. Single blastomeres were cultured at 38.5°C in 5% CO_2, 5% O_2, and 90% N_2 until 72, 121 or 170 hpf according to the experimental design.

Single-cell isolation

For isolation of single cells at 72 hpf, 121 hpf or from blastomeres that did not reach the blastocyst stage at 170 hpf, the outgrowths were transferred to Ca^{+2}/Mg^{+2}-free PBS with 0.1% PVP, and the cells were dissociated mechanically with a STRIPPER pipet holder and 170 µm and 135 µm capillaries (Origio, Cooper Surgical, CT, US). For blastomeres that reached the blastocyst stage at 170 hpf, single-cell isolation was done by incubating the embryos in trypsine-EDTA at 38.5°C, followed by washing in Ca^{+2}/Mg^{+2}-free PBS with 0.1% PVP, pipetting with a STRIPPER pipet holder, and 135 µm and 70 µm capillaries (Origio, Cooper Surgical, CT, US) and micromanipulation with holding (MPH-MED-35, Origio, Cooper Surgical, CT, US) and biopsy (MBB-BP-M-30, Origio, Cooper Surgical, CT, US) pipettes. When single-cell dissociation was not possible, clusters of cells were collected. Single cells from all time points were washed in Ca^{+2}/Mg^{+2}-free PBS with 0.1% PVP before transferring them into a well of a skirted 96-well plate (4ti-0960/C, AZENTA Life Sciences, Bioké, Leiden, The Netherlands) containing 2.5 µL of RLT lysis buffer with a 70 µm capillary (Origio, Cooper Surgical, CT, US) and a STRIPPER pipet holder screwed on 0.5 µL. The collected samples were kept on ice during the whole procedure and then stored at −80°C.

DNA and RNA separation, library preparation and sequencing

Separation of DNA and RNA from single cells

The DNA and mRNA from single cells were separated following the genome and transcriptome sequencing (G&T-seq) protocol (Macaulay et al., 2016) on a robotic liquid-handling platform

(Microlab STAR Plus, Hamilton). Specifically, the 96-well sample plate was positioned on the robot deck alongside (a) a plate for capturing poly-A mRNAs, containing 10µl per well of Dynabeads® MyOneTM Streptavidin C1 (Thermo Fisher Scientific) bound to biotinylated poly-dT oligos containing the SmartSeq2 primer sequence '5BioTinTEG/ -AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN' (IDT), (b) a plate for washing away DNA from the cell lysate supernatant, containing 25µl per well of G&T-wash buffer, and (c) an empty DNA destination plate. First, the cell lysate was mixed with biotinylated poly-dT beads and incubated for 20 min. Subsequently, beads bound to mRNA were pulled down using a low elution magnet (Alpaqua) for 2 min, and the supernatant was transferred to the DNA destination plate. Following this, the beads were subjected to two washes, each with 10µl of G&T wash buffer, and the supernatant was transferred to the DNA destination plate. The DNA destination plate, containing 37.5µl of G&T wash buffer, was then centrifuged for 1 min at 1000g and stored at -80°C.

RNA amplification, library preparation and sequencing

The plate containing poly-A mRNA-bound beads was processed using an adapted SmartSeq2 protocol with 20 PCR cycles. Subsequently, the amplified single-cell cDNA was purified using a 0.8:1 ratio of Agencourt AMPure XP beads (Analis), followed by washing with 80% ethanol and elution in water. Libraries were generated from the amplified cDNA according to the Nextera XT (Illumina) protocol with quarter volumes. These libraries were pooled and sequenced using single-end 50 sequencing on a HiSeq4000 Illumina sequencer, aiming at 1 million reads per sample. For a subset of blastocyst stage single cells, the libraries were resequenced paired-end 150 on a NovaSeq 6000 Illumina sequencer for more reads.

DNA amplification, GBS library preparation and sequencing

For each blastomere outgrowth, DNA from either individual single cell or several cells was used for GBS (with double enzyme restriction) processing. A sibling blastocyst was collected and processed together when available. Whole-genome amplification (WGA) was carried out using the REPLI-g SC kit (Qiagen, Hilden, Germany) following the manufacturer's guidelines, with variations in reaction volumes (full or half) and an incubation period of 2-3 hours. The DNA separated from the single cells stored in G&T wash buffer (see before), was thawed on ice, centrifuged for 1 min at 1000g, purified with Agencourt AMPure XP beads (Beckman Coulter, USA), eluted in 4 µl of scPBS and processed following the REPLI-g SC kit (Qiagen, Hilden, Germany) manufacturer's guidelines with an incubation time of 2h. Additionally, bulk DNA was extracted from ovarian tissue of the donor cows (mothers of the respective embryos), semen from the bull (father of the embryos), and blood from the parents of the bull (paternal grandparents of the embryos) using the DNeasy Blood and Tissue kit (Qiagen, Hilden,

Germany).

The whole-genome amplified DNA and bulk DNA were subjected to double restriction digestion with 8 units (U) of PstI-HF (R3140S, NEB) and 4U of CviAII (R0640L, NEB) enzymes (New England Biolabs, NEB, USA) combined with adapter ligation using 200-300 ng as an input DNA for each sample. The restriction-ligation reaction was performed in a total reaction volume of 15 µl, with a final concentration of 1X rCutSmart buffer (B6004S, NEB), 1mM ATP (P0756S, NEB) and 160U of T4 DNA Ligase (M0202L, NEB). Next, the double-sided size-selection was performed with Agencourt AMPure XP beads (Beckman Coulter, CA, USA) and 7 cycles of PCR to amplify and barcode size-selected adapter-ligated fragments using Q5 High-Fidelity 2X Master Mix (NEB) and 0.5 µM primer mix with the following program. Subsequently, the libraries were purified with Agencourt AMPure XP beads (Beckman Coulter, CA, USA), equimolarly pooled and sequenced paired-end 150 on NovaSeq 6000 Illumina sequencer, with a target of 20-30 million reads per sample.

Data analysis

GBS data processing, haplotyping and aneuploidy profiling

The raw reads were processed using fastp (v0.23.2) (Chen et al., 2018) to trim adaptor sequences, filter out low-quality reads, remove UMI sequences and append them to read names. Next, cutadapt (v 1.18) (Martin, 2011) was applied to select reads that start with the internal barcode and trim off the barcode. Reads were then mapped to bovine reference genome bosTau9 (ARS-UCD1.2) using BWA-MEM (v0.7.17) (Li, 2013). Next, UMICollapse (Liu, 2019) was used for collapsing duplicated reads with the same UMI, while accounting for sequencing/PCR errors. Across all samples, we obtained a median of 64 million mapped reads. After UMI deduplication, a median of 35 million mapped reads were retained. Variant calling was performed with freebayes (v 1.3.2) (Garrison and Marth, 2012). Initially, variants were called for the parents and phasing reference(s) (parental grandparents or a sibling blastocyst). Subsequently, the identified variants were utilized as input to call variants for each single/multi-cell sample. The called SNVs were converted into bi-allelic calls (e.g., AA, AB, and BB) with B-allele frequency (BAF) values calculated based on allele-specific depth of coverage using a custom R script. The genotype calls and BAFs were then used as input for the siCHILD pipeline (details in the method paper (Zamani Esteki et al., 2015)). In brief, siCHILD performs pedigree-based haplotyping analysis. The parents were first phased using phasing reference(s). Subsequently, for specific combinations of phased parental genotypes, corresponding SNP BAF values of the sample of interest were retrieved and plotted on paternal and maternal haplarithms. Genome-wide parental haplotype inheritance can then be inferred through visual inspection of the genome-wide haplarithm plots.

In addition to haplotype information, the R package QDNAseq and custom scripts were applied to BAM files post-deduplication for aneuploidy profiling with a fixed bin size of 100kb. To quantify the extent of aneuploidy, the percentage of aneuploid autosomal regions was calculated for each sample. Specifically, bins with a segmentation logR value within $log2(1/2)$) and $\log2(3/2)$ were classified as normal, while all other bins were classified as abnormal. The percentage of aneuploid autosomal regions was calculated by summing up the length of all abnormal bins and dividing it by the size of all autosomes. Samples were categorized based on the percentage of aneuploid autosomal regions: $\leq \%1$ as euploid, 1-5% as low-level aneuploidy, 5-10% as medium-level aneuploidy, and >10% as high-level aneuploidy. To determine chromosome X nullisomy status, the lengths of all bins on chromosome X with a segmentation logR value ≤ -3 were aggregated. The fraction of the chromosome X region exhibiting nullisomy was determined by dividing this sum by the size of chromosome X. Samples with >95% nullisomy in the chromosome X region were classified as chromosome X nullisomy.

The haplotype composition and aneuploidy status of each blastomere outgrowth were determined using GBS results from a single-cell/multi-cell sample. A sample was considered anuclear if >60% and >10% of the reads mapped to the mitochondrial genome before and after UMI deduplication, respectively, with empty haplarithm and copy number plots. To infer an outgrowth to be anuclear, each cell in the outgrowth was individually checked and confirmed to be anuclear. To verify the presence of mitochondrial genomes in inferred anuclear biopsies and ascertain their maternal inheritance, we exclusively called variants on the mitochondrial genome across all inferred anuclear biopsies using Freebayes (v1.3.2) (Garrison and Marth, 2012) and compared the called genotypes.

Single-cell RNA sequencing data preprocessing

The raw reads were processed with fastp (v0.23.2) (Chen et al., 2018) to trim adaptor sequences and filter out low-quality reads. Subsequently, the reads were mapped to the bovine reference genome bosTau9 (ARS-UCD1.2) using STAR aligner (v 2.7.3) (Dobin et al., 2013) with Ensembl annotations ARS-UCD1.2.105. Next, raw counts per cell were obtained using HTSeq (v 0.9.1) (Anders et al., 2015). Cells with fewer than 200,000 detected molecules or fewer than 2,000 expressed genes were excluded from further analysis, resulting in 446 cells out of 677 passing these quality control thresholds. The retained cells exhibited a median of 862,384 molecules and 7,272 genes detected per cell. Counts were then normalized and scaled using R package Seurat (v5.0.1) (Hao et al., 2023).

Cell clustering, cell type identification and trajectory analysis

Cell clustering was performed based on the normalized and scaled transcriptome profiles using

UMAP implemented in Seurat (v5.0.1). Subsequently, we assigned molecular cell stage labels to each cluster based on the expression of (a) known marker genes for ICM and TF (e.g. *FN1*, *KDM2B* for ICM, *GATA3* for trophectoderm, etc.(Nagatomo et al., 2013; Negrón-Pérez et al., 2017)); (b) genes first expressed at 4-cell stage (minor EGA) or 8-16 cell stage (major EGA) in bovine embryos (Graf et al., 2014). For trajectory analysis, we employed the R package Slingshot (v2.10.0) to infer developmental trajectories based on transcriptome profiles.

Regulon analysis

Regulon analysis was performed on the transcriptome data using pySCENIC (v0.12.1) (Van de Sande et al., 2020). First, for each motif in the public collection (https://resources.aertslab.org/cistarget/motif_collections/), genes in the bovine reference genome bosTau9 (ARS-UCD1.2) with orthologs in humans were prioritized. This prioritization was determined by a scoring system that assessed the presence of the motif within a 10kb window upstream and downstream of the transcription start site. Consequently, a motif x gene matrix was generated. Next, this matrix served as input alongside the raw count matrix to pySCENIC (v0.12.1) to detect active regulons. pySCENIC was run ten separate times, and only regulons appearing two or more times were retained. The final AUC matrix was constructed using the maximum AUC values for each regulon among all runs in which they were present. The AUC heatmap was created with AUC values scaled for each regulon. To rank the identified regulons according to their specificity in each cell lineage, a regulon specificity score was calculated using “calcRSS” function within R package SCENIC (v 1.3.1). To identify the key regulons for each molecular stage, we considered only regulons where corresponding TFs were expressed in more than 50% of the cells. From these, for each stage, we selected the top 6 regulons with the most frequent TF expression among those ranked in the top 20 for both activity and specificity.

Differential expression and regulon analysis

Differential expression analysis was performed by comparing androgenetic/gynogenetic/polyploid cells to biparental diploid cells within each molecular cell stage using the FindMarkers function in Seurat (v5.0.1), employing the “MAST” method with a minimum percentage threshold of 0.5. Pseudotime values and aneuploid status were incorporated as latent variables in the model. Genes were considered differentially expressed if they met the criteria of an adjusted p-value < 0.05 and an absolute log2 fold change greater than 1. Differential regulon analysis was performed between cells exhibiting WG abnormalities and biparental diploid cells within each molecular cell stage by Wilcoxon ranked-sum test on the AUC values. Regulons were considered to have significantly different activity if they demonstrated an adjusted p-value < 0.05 and an absolute log2 fold change greater than 1.

The functions of the differentially expressed genes and regulons were curated through manual online searches and literature reviews.

Statistical analysis

Categorical data were compared using the chi-squared test, with Fisher's exact test applied when cell counts were too small. Pseudotime values were compared using the t-test. Statistical significance was determined using a two-tailed approach with a significance threshold of $p < 0.05$. All tests were conducted in R (v4.3.2).

Data availability

The GBS and single-cell RNA sequencing data reported in this paper are available at the European Nucleotide Archive (ENA) under project number PRJEB76932 (https://www.ebi.ac.uk/ena/browser/view/PRJEB76932).

Acknowledgements

Funding was received from the Marie Skłodowska-Curie grant agreement No 813707 (MATER) and from the KU Leuven, C1-C14/22/125 to J.R.V. and T.V. T.L. and T.V. were supported by the Research Foundation Flanders (FWO: G0C6120N, G088621N and I001818N). Y.Z. was supported by the Marie Skłodowska-Curie grant agreement No 813707 (MATER). A.F. is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860960 and BOF22/ITN/036.

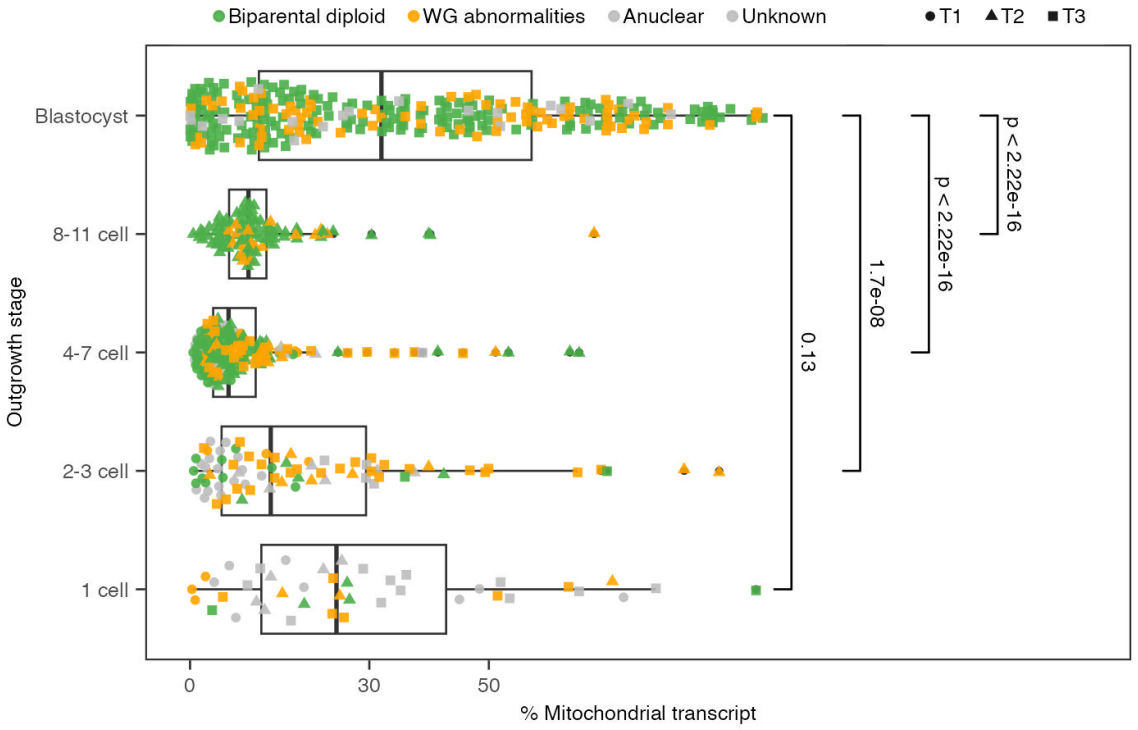
Author contributions

J.R.V., A.V.S., and K.S. conceptualized and designed the study. A.F. conducted embryo experiments, with assistance from T.D.C. and D.A.V., T.L. and T.V. performed G&T-seq. G.P. and T.J. conducted GBS processing and coordinated sequencing. Y.Z. conducted bioinformatics analysis. Y.Z., J.R.V., and O.T. interpreted the data. Y.Z., J.R.V. A.F. and A.K. wrote the manuscript. All co-authors reviewed and approved the manuscript.

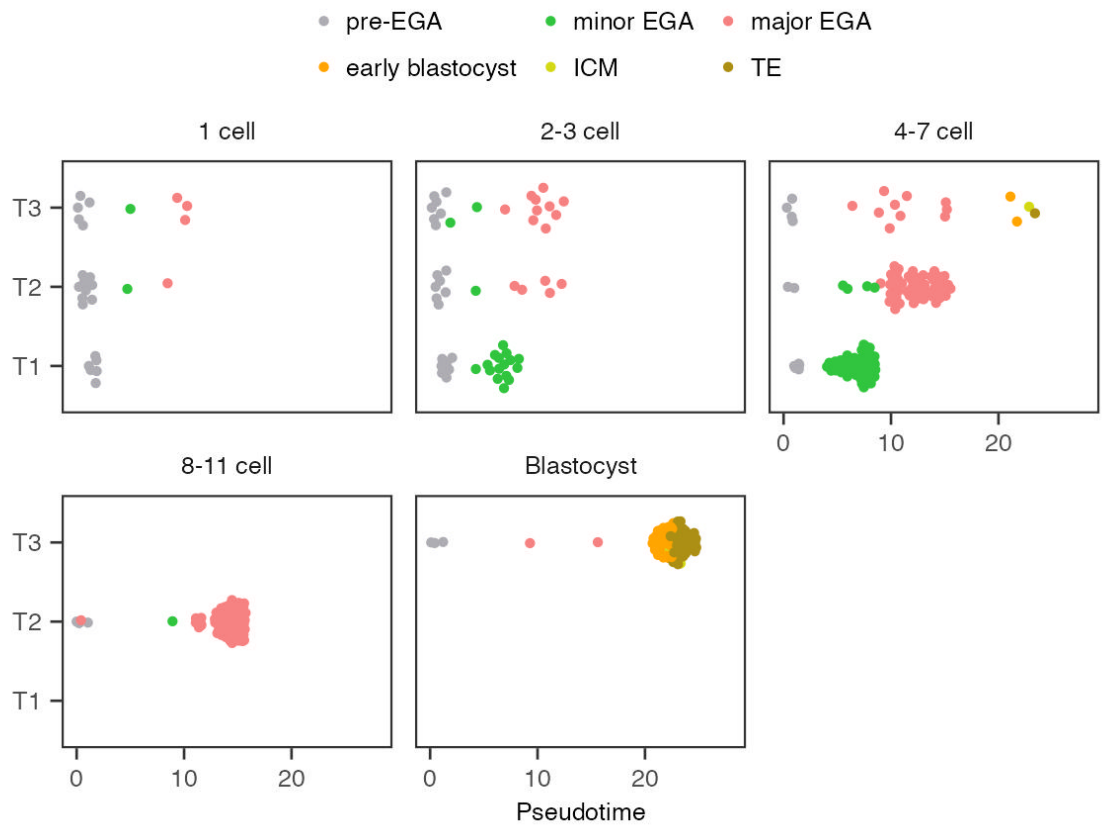
Competing interests

T.V. and J.R.V. are co-inventors on licensed patents WO/2011/157846 (Methods for haplotyping single cells), WO/2014/053664 (High-throughput genotyping by sequencing low amounts of genetic material) and WO/2015/028576 (Haplotyping and copy number typing using polymorphic variant allelic frequencies).

Supplementary figures



Supplementary Figure 1. High percentages of mitochondrial transcripts in cells from blastocysts.



Supplementary Figure 2. Comparison of molecular age across collection timepoints for cells from outgrowths with comparable cell numbers.

References

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166. doi:10.1093/BIOINFORMATICS/BTU638

Bakshi A, Iturra FE, Alamban A, Rosas-Salvans M, Dumont S, Aydogan MG. 2023. Cytoplasmic division cycles without the nucleus and mitotic CDK/cyclin complexes. *Cell* **186**:4694-4709.e16. doi:10.1016/J.CELL.2023.09.010/ATTACHMENT/CD4A0F61-45ED-48F1-B6DD-C5F5C35AA432/MMC3

Balaban B, Brison D, Calderón G, Catt J, Conaghan J, Cowan L, Ebner T, Gardner D, Hardarson T, Lundin K, Cristina Magli M, Mortimer D, Mortimer S, Munné S, Royere D, Scott L, Smitz J, Thornhill A, Van Blerkom J, Van Den Abbeel E. 2011. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Human Reproduction* **26**:1270–1283. doi:10.1093/HUMREP/DER037

Caroselli S, Figliuzzi M, Picchetta L, Cogo F, Zambon P, Pergher I, Girardi L, Patassini C, Poli M, Bakalova D, Cimadomo D, Findikli N, Coban O, Serdarogullari M, Favero F, Bortolato S, Anastasi A, Capodanno F, Gallinelli A, Brancati F, Rienzi L, Ubaldi FM, Jimenez-Almazán J, Blesa-Jarque D, Miravet-Valenciano J, Rubio C, Simòn C, Capalbo A. 2023. Improved clinical utility of preimplantation genetic testing through the integration of ploidy and common pathogenic microdeletions analyses. *Human Reproduction* **38**:762–775. doi:10.1093/HUMREP/DEAD033

Chatzimeletiou K, Morrison EE, Prapas N, Prapas Y, Handyside AH. 2005. Spindle abnormalities in normally developing and arrested human preimplantation embryos in vitro identified by confocal laser scanning microscopy. *Human Reproduction* **20**:672–682. doi:10.1093/HUMREP/DEH652

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. doi:10.1093/BIOINFORMATICS/BTY560

Coonen E, Rubio C, Christopikou Di, DImitriadou E, Gontar J, Goossens V, Maurer M, Spinella F, Vermeulen N, De Rycke M. 2020. ESHRE PGT Consortium good practice recommendations for the detection of structural and numerical chromosomal aberrations. *Hum Reprod Open* **2020**:1–20. doi:10.1093/HROPEN/HOAA017

Daughtry BL, Rosenkrantz JL, Lazar NH, Fei SS, Redmayne N, Torkenczy KA, Adey A, Yan M, Gao L, Park B, Nevonen KA, Carbone L, Chavez SL. 2019. Single-cell sequencing of primate preimplantation embryos reveals chromosome elimination via cellular fragmentation and blastomere exclusion. *Genome Res* **29**:367–382. doi:10.1101/GR.239830.118

De Coster T, Masset H, Tšuiko O, Catteeuw M, Zhao Y, Dierckxsens N, Aparicio AL, Dimitriadou E, Debrock S, Peeraer K, de Ruijter-Villani M, Smits K, Van Soom A, Vermeesch JR. 2022. Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. *Genome Biology 2022 23:1* **23**:1–29. doi:10.1186/S13059-022-02763-2

Deng Q, Huang S. 2004. PRDM5 is silenced in human cancers and has growth suppressive activities. *Oncogene 2004 23:28* **23**:4903–4910. doi:10.1038/sj.onc.1207615

Destouni A, Esteki MZ, Catteeuw M, Tšuiko O, Dimitriadou E, Smits K, Kurg A, Salumets A, Van Soom A, Voet T, Vermeesch JR. 2016. Zygotes segregate entire parental genomes in distinct blastomere lineages causing cleavage-stage chimerism and mixoploidy. *Genome Res* **26**:567–578. doi:10.1101/GR.200527.115

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/BIOINFORMATICS/BTS635

Dunn HO, McEntee K, Hansel W. 1970. Diploid-triploid chimerism in a bovine true hermaphrodite. *Cytogenet Genome Res* **9**:245–259. doi:10.1159/000130095

Fernandez Gallardo E, Sifrim A, Chappell J, Demeulemeester J, Herrmann JC, Vermotte R, Kerremans A, Van Der Haegen M, Herck J Van, Vanuytven S, Vandereyken K, Macaulay IC, Vermeesch JR, Peeraer K, Debrock S, Pasque V, Voet T, Affiliations #. 2023. A multi-omics genome-and-transcriptome single-cell atlas of human preimplantation embryogenesis reveals the cellular and molecular impact of chromosome instability. *bioRxiv* 2023.03.08.530586. doi:10.1101/2023.03.08.530586

Fragouli E, Alfarawati S, Spath K, Jaroudi S, Sarasa J, Enciso M, Wells D. 2013. The origin and impact of embryonic aneuploidy. *Hum Genet* **132**:1001–1013. doi:10.1007/S00439-013-1309-0/TABLES/4

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing.

Gerri C, McCarthy A, Alanis-Lobato G, Demtschenko A, Bruneau A, Loubersac S, Fogarty NME, Hampshire D, Elder K, Snell P, Christie L, David L, Van de Velde H, Fouladi-Nashta AA, Niakan KK. 2020. Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature 2020 587:7834* **587**:443–447. doi:10.1038/s41586-020-2759-x

Graf A, Krebs S, Zakhartchenko V, Schwalb B, Blum H, Wolf E. 2014. Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc Natl Acad Sci U S A* **111**:4139–4144. doi:10.1073/PNAS.1321569111/-/DCSUPPLEMENTAL/PNAS.201321569SI.PDF

Hansen PJ. 2010. Medawar redux - an overview on the use of farm animal models to elucidate principles of reproductive immunology. *Am J Reprod Immunol* **64**:225–230. doi:10.1111/J.1600-0897.2010.00900.X

Hao Y, Lai L, Mao J, Im GS, Bonk A, Prather RS. 2004. Apoptosis in Parthenogenetic Preimplantation Porcine Embryos. *Biol Reprod* **70**:1644–1649. doi:10.1095/BIOLREPROD.103.026005

Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, Srivastava A, Molla G, Madad S, Fernandez-Granda C, Satija R. 2023. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology 2023 42:2* **42**:293–304. doi:10.1038/s41587-023-01767-y

Hardy K, Handyside AH, Winston RML. 1989. The human blastocyst: cell number, death and allocation during late preimplantation development in vitro. *Development* **107**:597–604. doi:10.1242/DEV.107.3.597

Hardy K, Winston RML, Handyside AH. 1993. Binucleate blastomeres in preimplantation human embryos in vitro: failure of cytokinesis during early cleavage. *Reproduction* **98**:549–558. doi:10.1530/JRF.0.0980549

Hassold T, Chen N, Funkhouser J, Jooss T, Manuel B, Matsuura J, Matsuyama A, Wilson C, Yamane JA, Jacobs PA. 1980. A cytogenetic study of 1000 spontaneous abortions. *Ann Hum Genet* **44**:151–164. doi:10.1111/J.1469-1809.1980.TB00955.X

Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics 2001 2:4* **2**:280–291. doi:10.1038/35066065

Home P, Kumar RP, Ganguly A, Saha B, Milano-Foster J, Bhattacharya B, Ray S, Gunewardena S, Paul A, Camper SA, Fields PE, Paul S. 2017. Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. *Development (Cambridge)* **144**:876–888. doi:10.1242/DEV.145318/264332/AM/GENETIC-REDUNDANCY-OF-GATA-FACTORS-IN

Hoppe PC, Illmensee K. 1977. Microsurgically produced homozygous-diploid uniparental mice. *Proceedings of the National Academy of Sciences* **74**:5657–5661. doi:10.1073/PNAS.74.12.5657

Hwang IS, Park MR, Lee HS, Kwak TU, Son HY, Kang JK, Lee JW, Lee K, Park EW, Hwang S. 2020. Developmental and Degenerative Characterization of Porcine Parthenogenetic Fetuses during Early Pregnancy. *Animals 2020, Vol 10, Page 622* **10**:622. doi:10.3390/ANI10040622

Ito M, Kaneko-Ishino T, Ishino F, Matsuhashi M, Yokoyama M, Katsuki M. 1991. Fate of haploid parthenogenetic cells in mouse chimeras during development. *Journal of Experimental Zoology* **257**:178–183. doi:10.1002/JEZ.1402570206

Johnson WH, Loskutoff NM, Plante Y, Betteridge KJ. 1995. Production of four identical calves by the separation of blastomeres from an in vitro derived four-cell embryo. *Vet Rec* **137**:15–16. doi:10.1136/VR.137.1.15

Kola I, Trounson A, Dawson G, Rogers P. 1987. Tripronuclear human oocytes: altered cleavage patterns and subsequent karyotypic analysis of embryos. *Biol Reprod* **37**:395–401. doi:10.1095/BIOLREPROD37.2.395

Leng L, Ouyang Q, Kong X, Gong F, Lu C, Zhao L, Shi Y, Cheng D, Hu L, Lu G, Lin G. 2017. Self-diploidization of human haploid parthenogenetic embryos through the Rho pathway regulates endomitosis and failed cytokinesis. *Scientific Reports 2017 7:1* **7**:1–10. doi:10.1038/s41598-017-04602-y

Leng L, Sun J, Huang J, Gong F, Yang L, Zhang S, Yuan X, Fang F, Xu X, Luo Y, Bolund L, Peters BA, Lu G, Jiang T, Xu F, Lin G. 2019. Single-Cell Transcriptome Analysis of Uniparental Embryos Reveals Parent-of-Origin Effects on Human Preimplantation Development. *Cell Stem Cell* **25**:697-712.e6. doi:10.1016/j.stem.2019.09.004

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*. doi:10.6084/M9.FIGSHARE.963153.V1

Liu D. 2019. Algorithms for efficiently collapsing reads with Unique Molecular Identifiers. *PeerJ* **7**. doi:10.7717/PEERJ.8275

Liu L, Trimarchi JR, Keefe DL. 2002. Haploidy but Not Parthenogenetic Activation Leads to Increased Incidence of Apoptosis in Mouse Embryos. *Biol Reprod* **66**:204–210. doi:10.1095/BIOLREPROD66.1.204

López-Maury L, Marguerat S, Bähler J. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics 2008 9:8* **9**:583–593. doi:10.1038/nrg2398

Lucifero D, Suzuki J, Bordignon V, Martel J, Vigneault C, Therrien J, Filion F, Smith LC, Trasler JM. 2006. Bovine SNRPN Methylation Imprint in Oocytes and Day 17 In Vitro-Produced and Somatic Cell Nuclear Transfer Embryos. *Biol Reprod* **75**:531–538. doi:10.1095/BIOLREPROD.106.051722

Ma LB, Yang L, Zhang Y, Cao JW, Hua S, Li JX. 2008. Quantitative analysis of mitochondrial RNA in goat–sheep cloned embryos. *Mol Reprod Dev* **75**:33–39. doi:10.1002/MRD.20736

Macaulay IC, Teng MJ, Haerty W, Kumar P, Ponting CP, Voet T. 2016. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nature Protocols 2016 11:11* **11**:2081–2103. doi:10.1038/nprot.2016.138

Madan K. 2020. Natural human chimeras: A review. *Eur J Med Genet* **63**:103971. doi:10.1016/J.EJMG.2020.103971

Marsico TV, Valente RS, Annes K, Oliveira AM, Silva MV, Sudano MJ. 2023. Species-specific molecular differentiation of embryonic inner cell mass and trophectoderm: A systematic review. *Anim Reprod Sci* **252**:107229. doi:10.1016/J.ANIREPROSCI.2023.107229

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**:10–12. doi:10.14806/EJ.17.1.200

Masset H, Ding J, Dimitriadou E, Debrock S, Tšuiko O, Smits K, Peeraer K, Voet T, Esteki MZ, Vermeesch JR. 2022. Single-cell genome-wide concurrent haplotyping and copy-number profiling through genotyping-by-sequencing. *Nucleic Acids Res* **50**:e63–e63. doi:10.1093/NAR/GKAC134

Masset H, Tšuiko O, Vermeesch JR. 2021. Genome-wide abnormalities in embryos: Origins and clinical consequences. *Prenat Diagn* **41**:554–563. doi:10.1002/PD.5895

May-Panloup P, Vignon X, Chrétien MF, Heyman Y, Tamassia M, Mathièry Y, Reynier P. 2005. Increase of mitochondrial DNA content and transcripts in early bovine embryogenesis associated with upregulation of mtTFA and NRF1 transcription factors. *Reproductive Biology and Endocrinology* **3**:1–8. doi:10.1186/1477-7827-3-65/FIGURES/3

McCoy RC, Demko ZP, Ryan A, Banjevic M, Hill M, Sigurjonsson S, Rabinowitz M, Petrov DA. 2015. Evidence of Selection against Complex Mitotic-Origin Aneuploidy during Preimplantation Development. *PLoS Genet* **11**. doi:10.1371/JOURNAL.PGEN.1005601

Meinecke B, Kuiper H, Drögemüller C, Leeb T, Meinecke-Tillman S. 2003. A Mola Hydatidosa Coexistent with a Foetus in a Bovine Freemartin Pregnancy. *Placenta* **24**:107–112. doi:10.1053/PLAC.2002.0872

Nagatomo H, Kagawa S, Kishi Y, Takuma T, Sada A, Yamanaka K ichi, Abe Y, Wada Y, Takahashi M, Kono T, Kawahara M. 2013. Transcriptional Wiring for Establishing Cell Lineage Specification at the Blastocyst Stage in Cattle1. *Biol Reprod* **88**:158–159. doi:10.1095/BIOLREPROD.113.108993/2514370

Negrón-Pérez VM, Zhang Y, Hansen PJ. 2017. Single-cell gene expression of the bovine blastocyst. *Reproduction* **154**:627–644. doi:10.1530/REP-17-0345

Niakan KK, Eggan K. 2013. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev Biol* **375**:54–64. doi:10.1016/J.YDBIO.2012.12.008

Ottolini CS, Kitchen J, Xanthopoulou L, Gordon T, Summers MC, Handyside AH. 2017. Tripolar mitosis and partitioning of the genome arrests human preimplantation development in vitro. *Scientific Reports 2017 7:1* **7**:1–10. doi:10.1038/s41598-017-09693-1

Santos RR, Schoevers EJ, Roelen BAJ. 2014. Usefulness of bovine and porcine IVM/IVF models for reproductive toxicology. *Reprod Biol Endocrinol* **12**:1–12. doi:10.1186/1477-7827-12-117

Seckl MJ, Sebire NJ, Berkowitz RS. 2010. Gestational trophoblastic disease. *The Lancet* **376**:717–729. doi:10.1016/S0140-6736(10)60280-2

Sherard J, Bean C, Bove B, DelDuca V, Esterly KL, Karcsh HJ, Munshi G, Reamer JF, Suazo G, Wilmoth D. 1986. Long survival in a 69,XXY triploid male. *Am J Med Genet* **25**:307–312. doi:10.1002/AJMG.1320250216

Shu X sheng, Geng H, Li L, Ying J, Ma C, Wang Y, Poon FF, Wang X, Ying Y, Yeo W, Srivastava G, Tsao SW, Yu J, Sung JJY, Huang S, Chan ATC, Tao Q. 2011. The Epigenetic Modifier PRDM5 Functions as a Tumor Suppressor through Modulating WNT/β-Catenin Signaling and Is Frequently Silenced in Multiple Tumors. *PLoS One* **6**:e27346. doi:10.1371/JOURNAL.PONE.0027346

Sigalos G, Triantafyllidou O, Vlahos NF. 2016. Novel embryo selection techniques to increase embryo implantation in IVF attempts. *Arch Gynecol Obstet* **294**:1117–1124. doi:10.1007/S00404-016-4196-5/TABLES/1

Sugimura S, Akai T, Hashiyada Y, Somfai T, Inaba Y, Hirayama M, Yamanouchi T, Matsuda H, Kobayashi S, Aikawa Y, Ohtake M, Kobayashi E, Konishi K, Imai K. 2012. Promising System for Selecting Healthy In Vitro–Fertilized Embryos in Cattle. *PLoS One* **7**:e36627. doi:10.1371/JOURNAL.PONE.0036627

Sugimura S, Akai T, Imai K. 2017. Selection of viable in vitro-fertilized bovine embryos using time-lapse monitoring in microwell culture dishes. *Journal of Reproduction and Development* **63**:353–357. doi:10.1262/JRD.2017-041

Tarkowski AK. 1977. In vitro development of haploid mouse embryos produced by bisection of one-cell fertilized eggs. *Development* **38**:187–202. doi:10.1242/DEV.38.1.187

Thundathil J, Filion F, Smith LC. 2005. Molecular control of mitochondrial function in preimplantation mouse embryos. *Mol Reprod Dev* **71**:405–413. doi:10.1002/MRD.20260

Tšuiko O, Catteeuw M, Esteki MZ, Destouni A, Pascottini OB, Besenfelder U, Havlicek V, Smits K, Kurg A, Salumets A, D'Hooghe T, Voet T, Van Soom A, Vermeesch JR. 2017. Genome stability of bovine in vivo-conceived cleavage-stage embryos is higher compared to in vitro-produced embryos. *Human Reproduction* **32**:2348–2357. doi:10.1093/HUMREP/DEX286

Tšuiko O, Vanneste M, Melotte C, Ding J, Debrock S, Masset H, Peters M, Salumets A, De Leener A, Pirard C, Kluyskens C, Hostens K, van de Vijver A, Peeraer K, Denayer E, Vermeesch JR, Dimitriadou E. 2021. Haplotyping-based preimplantation genetic testing reveals parent-of-origin specific mechanisms of aneuploidy formation. *NPJ Genom Med* **6**. doi:10.1038/S41525-021-00246-0

Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols 2020 15:7* **15**:2247–2276. doi:10.1038/s41596-020-0336-2

Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, Debrock S, Amyere M, Vikkula M, Schuit F, Fryns JP, Verbeke G, D'Hooghe T, Moreau Y, Vermeesch JR. 2009. Chromosome instability is common in human cleavage-stage embryos. *Nature Medicine 2009 15:5* **15**:577–583. doi:10.1038/nm.1924

Willadsen SM. 1980. The viability of early cleavage stages containing half the normal number of blastomeres in the sheep. *Reproduction* **59**:357–362. doi:10.1530/JRF.0.0590357

Wydooghe E, Vaele L, Piepers S, Dewulf J, Van Abbeel E Den, De Sutter P, Van Soom A. 2014. Individual commitment to a group effect: strengths and weaknesses of bovine embryo group culture. *Reproduction* **148**:519–529. doi:10.1530/REP-14-0213

Xu J, Shu Y, Yao G, Zhang Yu, Niu W, Zhang Yile, Ma X, Jin H, Zhang F, Shi S, Wang Y, Song W, Dai S, Cheng L, Zhang X, Xie W, Hsueh AJ, Sun Y. 2021. Parental methylome reprogramming in human uniparental blastocysts reveals germline memory transition. *Genome Res* **31**:1519. doi:10.1101/GR.273318.120/-/DC1

Zamani Esteki M, Dimitriadou E, Mateiu L, Melotte C, Van der Aa N, Kumar P, Das R, Theunis K, Cheng J, Legius E, Moreau Y, Debrock S, D'Hooghe T, Verdyck P, De Rycke M, Sermon K, Vermeesch JR, Voet T. 2015. Concurrent Whole-Genome Haplotyping and Copy-Number Profiling of Single Cells. *Am J Hum Genet* **96**:894. doi:10.1016/J.AJHG.2015.04.011

Zaragoza M V., Surti U, Redline RW, Millie E, Chakravarti A, Hassold TJ. 2000. Parental origin and phenotype of triploidy in spontaneous abortions: Predominance of diandry and association with the partial hydatidiform mole. *Am J Hum Genet* **66**:1807–1820. doi:10.1086/302951

Zhan Q, Ye Z, Clarke R, Rosenwaks Z, Zaninovic N. 2016. Direct Unequal Cleavages: Embryo Developmental Competence, Genetic Constitution and Clinical Outcome. *PLoS One* **11**:e0166398. doi:10.1371/JOURNAL.PONE.0166398

CHAPTER 5 – LONG-READ WHOLE-GENOME SEQUENCING-BASED CONCURRENT HAPLOTYPING AND ANEUPLOIDY PROFILING OF HUMNAN SINGLE CELLS

This chapter is based on the following paper:

Zhao, Y., Tsuiko, O., Jatsenko, T., Peeters, G., Souche, E., Geysens, M., Dimitriadou, E., Vanhie, A., Peeraer, K., Debrock, S., Esch, H. Van, & Vermeesch, J. R. (2024). Long-read whole-genome sequencing-based concurrent haplotyping and aneuploidy profiling of single cells. BioRxiv, 2024.09.24.614469. https://doi.org/10.1101/2024.09.24.614469. Manuscript under revision in *Nucleic Acid Research*.

Long-read whole-genome sequencing-based concurrent haplotyping and aneuploidy profiling of single cells

Yan Zhao[1], Olga Tsuiko[1], Tatjana Jatsenko[1], Greet Peeters[1], Erika Souche[1], Mathilde Geysens[1], Eftychia Dimitriadou[1], Arne Vanhie[2], Karen Peeraer[2], Sophie Debrock[2], Hilde Van Esch[3], Joris Robert Vermeesch[1*]

[1]Laboratory for Cytogenetics and Genome Research, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium

[2]Leuven University Fertility Center, University Hospitals Leuven, Leuven 3000, Belgium

[3]Centre for Human Genetics, University Hospitals Leuven, Leuven 3000, Belgium

*To whom correspondence should be addressed. Email: joris.vermeesch@kuleuven.be

Abstract

Long-read whole-genome sequencing (lrWGS) enhances haplotyping by providing more phasing information per read compared to short-read sequencing. However, its use for single-cell haplotype phasing remains underexplored. This proof-of-concept study examines lrWGS data from single cells for small variant (SNV and indel) calling and haplotyping using the Genome in a Bottle (GIAB) Ashkenazi trio. lrWGS was performed on single-cell (1 cell) and multi-cell (10 cells) samples from the offspring. Chromosome-length haplotypes were obtained by leveraging both long reads and pedigree information. These haplotypes were further refined by replacing them with matched parental haplotypes. In single-cell and multi-cell samples, 92% and 98% of heterozygous SNVs, and 74% and 78% of heterozygous indels were accurately haplotyped. Applied to human embryos for preimplantation genetic testing (PGT), lrWGS demonstrated 100% consistency with array-based methods for detecting monogenic disorders, without requiring phasing references. Aneuploidies were accurately detected, with insights into the mechanistic origins of chromosomal abnormalities inferred from the parental unique allele fractions. We show that lrWGS-based concurrent haplotyping and aneuploidy profiling of single cells provides an alternative to current PGT methods, with applications potential in areas such as cell-based prenatal diagnosis and animal and plant breeding.

Introduction

Most mammalian genomes are diploid, consisting of one haploid set of chromosomes from each parent. Haplotyping reconstructs the unique nucleotide content of the two homologous chromosome sets, known as haplotypes. This process is crucial because the haplotypes can have different functional roles (1). Recent advancements in long-read sequencing technologies provide read lengths over 10 kilobases and accuracy comparable to next-generation sequencing (NGS) (2). These improvements significantly enhance genome haplotyping, as

individual long reads cover more heterozygous SNVs and provide haplotype information across extensive genomic regions, surpassing the capabilities of traditional SNP arrays and short-read data. The main long-read sequencing technologies are the single molecule real-time (SMRT) sequencing from Pacific Bioscience (PacBio) (3) and nanopore sequencing from Oxford Nanopore Technologies (ONT) (4). Studies have highlighted the effectiveness of long-read sequencing in variant identification, haplotyping, and genome assembly. Wenger et al. demonstrated the ability of ~28x PacBio high-fidelity (HiFi) reads for high-performance small variant (SNV and indel) calling and phased 99.64% of called variants. (5). More recently, the Telomere-to-Telomere (T2T) Consortium created a complete human reference genome, T2T-CHM13, using PacBio HiFi reads and ONT ultralong reads (6).

In addition to enabling haplotyping and genome assembly for bulk samples, long-read sequencing also holds potential for facilitating haplotyping at the single-cell level. Genetic analysis for single cells is challenging due to the necessity of whole genome amplification (WGA), typically using techniques like multiple displacement amplification (MDA). WGA can introduce technical errors, including allele dropout (ADO) due to the failure to amplify one allele, false-positive errors resulting from polymerase infidelity, and coverage nonuniformity caused by uneven amplification (7). Due to WGA artifacts, haplotype-based analysis of single cells is crucial for areas like preimplantation genetic testing for monogenic disorders (PGT-M). In this context, embryos from couples at risk of transmitting genetic disorders to their offspring are tested for the inheritance of disease alleles using DNA from trophectoderm (TE) biopsies containing 5 to 10 cells or from single blastomere biopsies. Several genome-wide haplotyping methods for single cells have been developed, including karyomapping (8), siCHILD (9), One PGT (10) Haploseek (11), and scGBS (12). These methods utilize SNP arrays or NGS data for genotyping, which provide minimal or no haplotype information. As a result, genetic phasing is applied, requiring DNA samples from prospective parents and first-degree relative(s) (Carvalho *et al.*, 2020), which are not always available. Furthermore, in cases of *de novo* mutations (DNMs) in prospective parents, the variant loci cannot be phased through genetic phasing. Long-read data has the potential to directly phase both parents and embryo biopsies, including DNMs in the parents, without requiring relatives. Initial explorations focused on targeted long-read sequencing. For instance, Wu et al. conducted haplotype linkage analysis for the HBB gene by phasing the parents and TE biopsies using SMRT reads (13). Similarly, Tsuiko et al. explored both SMRT and ONT data in preclinical workup to infer the parental origin of DNMs (14). More recently, long-read whole-genome sequencing (lrWGS) has been employed for phasing parental genomes. Zhang et al. utilized ~30x PacBio long-read data for phasing the parents and conducted reference-free PGT-M for three monogenetic diseases (15). However, the effectiveness of lrWGS for phasing single cells and its application in generic PGT remains unexplored. Hård et al. assessed variant calling and genome assembly with lrWGS data from

single cells (16), but the limited sequencing depth of ~ 5x HiFi reads constrains a thorough exploration of its potential for clinical applications.

Beyond haplotyping, SNP arrays and NGS-based single-cell haplotyping methods enable concurrent haplotype-aware aneuploidy profiling, facilitating the identification of chromosomal abnormalities in single cells and determining their mechanistic origins. This has significant clinical implications, as chromosomal abnormalities can arise during human gametogenesis and are common in early embryogenesis (17, 18). PGT for aneuploidy (PGT-A) prevents the transmission of chromosomally abnormal embryos and enhances the *in vitro* fertilization (IVF) success rate (19). Long-read sequencing has proven valuable for detecting aneuploidies (20) and segmental imbalances (21) in embryo biopsies. However, to our knowledge, the potential of long-read data to infer the mechanistic origins of aneuploidies in embryo biopsies has not yet been explored.

Here, we present the first comprehensive analysis of lrWGS data from human single cells at an adequate depth of ~24x for SNV and indel calling, as well as haplotyping. Using a Genome in a Bottle (GIAB) trio consisting of HG002 (offspring), HG003 (father), and HG004 (mother) for benchmarking, we demonstrate the feasibility of lrWGS data for concurrent haplotyping and aneuploidy profiling of single cells without requiring additional phasing references. The clinical proof-of-concept application was validated in two PGT families, achieving 100% diagnostic concordance with SNP array-based PGT results **(Figure 1)**. This lrWGS-based PGT approach surpasses current methods with reduced clinical work-up, fewer family members involved, and a more comprehensive genomic analysis that integrates direct variant detection, haplotyping, and aneuploidy assessment. Furthermore, our data analysis strategy for concurrent haplotyping and aneuploidy profiling of single cells can be applied to other areas of single-cell genome analysis, such as cell-based prenatal diagnosis and animal and plant breeding.

Materials and methods

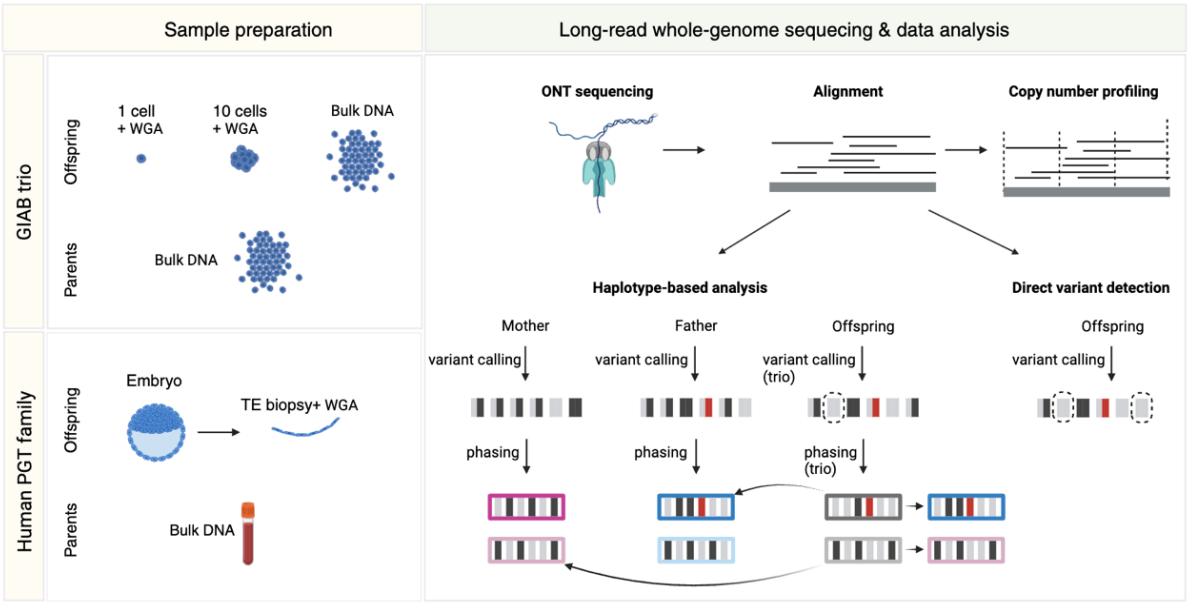


Figure 1. lrWGS-based concurrent haplotyping and aneuploidy profiling workflow. The benchmarking study utilized a Genome in a Bottle (GIAB) trio, consisting of HG002 (offspring), HG003 (father), and HG004 (mother). Single-cell (1 cell), multi-cell (10 cells), and bulk samples were collected from the HG002 cell line. The single-cell and multi-cell samples underwent WGA before lrWGS using nanopore technology. The long reads were then aligned to the human reference genome. Publicly available aligned lrWGS data for the parents were used in subsequent analyses. After alignment, aneuploidy profiling was performed on each sample. Two approaches were applied to determine whether a disease allele present in the parents was inherited by the offspring: haplotype-based analysis and direct variant detection. In the haplotype-based analysis, the parents underwent individual variant calling and phasing, while the offspring was analyzed in a trio setting, which incorporates parental information for better results. By comparing the offspring's haplotypes with those of the parents, the inherited parental haplotypes were inferred, which were subsequently used for diagnosis. For direct variant detection, the offspring underwent independent variant calling without parental information, and the results were utilized for diagnosis. In variant calling results, different alleles are depicted as black or grey rectangles, with the "disease allele" highlighted in red. Each pair of rectangles represents a genotype, and those affected by ADO are indicated by dashed circles. After phasing, the two alleles in a genotype were organized into distinct haplotypes. In this example, the disease allele inherited by the offspring was identified by both haplotype-based analysis and direct variant detection. The utility of lrWGS-based haplotyping and aneuploidy profiling was further evaluated in families undergoing PGT-M, where TE biopsies from embryos and bulk DNA from the prospective parents were analyzed.

Long-read whole-genome sequencing for the GIAB trio

Lymphoblastoid cell line of the offspring (HG002) from the GIAB Ashkenazi trio, consisting of HG002 (offspring), HG003 (father), and HG004 (mother) (Coriell Institute), was cultured in DMEM/F12, complemented with 10% FBS (Gibco). Single-cell (1 cell) and multi-cell (10 cells) samples were manually collected into 0.2ml tubes and DNA from these samples was amplified

by MDA using the REPLI-g single cell kit (Qiagen). ONT libraries were prepared on 3 µg of amplified material using the SQK-LSK114 kit, following ONT's recommendations for WGA library preparations. A bulk sample was collected in parallel, and DNA was extracted using the genomic DNA purification kit (Monarch). Libraries were then prepared on 3 µg of bulk DNA using the SQK-LSK114 kit. All libraries were loaded (30 fmol) on the PromethION device for sequencing with R10.4.1 flow cells. lrWGS datasets for the father (HG003) and mother (HG004) were obtained from the Oxford Nanopore Open Datasets. Specifically, BAM files were downloaded from the AWS storage bucket at s3://ont-open-data/giab_lsk114_2022.12/ and downsampled to ~24x, which is comparable to that of the offspring, using samtools (v 1.9) (22).

Long-read whole-genome sequencing for human PGT families

This study was approved by the Ethical Committee of UZ/KU Leuven (S68291). The parents consented to the use of residuary Human Bodily Material (HBM) for scientific research at the start of their IVF treatment. IVF and universal PGT-related procedures have been performed according to the standard operating procedures (SOPs) of UZ Leuven. Specifically, for universal PGT, parental bulk DNA was extracted from whole blood. TE biopsy was performed on blastocyst-stage embryos, obtaining an average of five TE cells, which were subjected to WGA using the MDA method with the REPLI-g SC kit (Qiagen, Germany). Residual DNA from both parents and corresponding embryo biopsies (MDA amplified) is available at the hospital, and an aliquot of excess DNA material were utilized for this study. ONT library preparation and sequencing were done using the same procedures as for cell line samples.

Read preprocessing and mapping

Read quality was assessed using NanoPlot (v1.39.0) (23) and FastQC (v0.11.7). Reads with an average quality score below 9 or a length shorter than 500 bp were filtered out using NanoFilt (v2.8.0) (23). Subsequently, the processed reads were aligned to human reference genome hg38 using minimap2 (v2.12) (24). Mapping statistics were generated using samtools flagstat (v1.9) (22). Chimeric read count was determined by counting the number of unique reads for all alignments with 0x800 SAM flag. Depth of coverage for each genomic position was computed using samtools depth with -aa flag (v1.9) (22), and average depth of coverage was determined by dividing the sum of depths across all positions by the size of the genome.

Variant calling

Clair3 (v0.1) (25) was used for variant calling with BAM file of a single individual. Clair3-Trio (v0.3) (26) was used for trio variant calling with BAM files of the parents and the offspring. For both strategies, default parameters were used and SNVs and short indels (≤50 base pairs) were called. GIAB benchmark data (v4.2.1) was used to assess variant calling performance on high-confidence regions. VCF files containing high confident small variants (SNVs and short

indels) on autosomes and corresponding BED files containing high confident regions were obtained for each trio member from https://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/AshkenazimTrio/. RTG Tools (v3.12.1) (27) was used for benchmarking analysis with the vcfeval command.

To include supplementary alignments in variant calling (omitted by default), a customized shell script was employed to modify the SAM flags of the supplementary alignments (2048 and 2064) as normal flags (0 and 16, respectively). Subsequently, the modified BAM files were utilized for variant calling with Clair3 (v0.1).

Haplotype phasing

The default phasing mode of WhatsHap (v1.0) (28), which uses data from a single individual, was employed to phase variants identified by Clair3. The pedigree phasing mode of WhatsHap (v1.0) (28), utilizing data from both parents and the offspring, was used to phase parental variants called by Clair3 and offspring's variants called by Clair3-Trio. For both modes, default parameters were applied to phase only SNVs, while the –indels flag was added to phase both SNVs and indels. Only variants with GQ values higher than 2 in VCF files were retained for phasing. Preliminary conservative paternal|maternal phasing data from GIAB (available at https://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/release/AshkenazimTrio/HG002_NA24385_son/NISTv4.2.1/GRCh38/SupplementaryFiles/HG002_GRCh38_1_22_v4.2.1_benchmark_phased_MHCassembly_StrandSeqANDTrio.vcf.gz) was used as a benchmark for evaluating phasing performance with WhatsHap Compare (v1.0) (28).

Inference of inherited parental haplotypes

To deduce the parental haplotypes inherited by the offspring, haplotypes of the offspring obtained from the pedigree phasing mode were compared with parental haplotypes from the default phasing mode. Only biallelic loci were retained for comparison, excluding those with unphased genotypes in either parent or identical homozygous genotypes in both parents. Each chromosome was divided into 1 Mb consecutive segments, referred to as comparison units, and haplotype comparisons were performed within each segment. Since in the resulting VCF file from pedigree phasing mode, haplotype alleles of the offspring are given as paternal|maternal, for each comparison unit, we compared the first haplotype of the offspring to the two paternal haplotypes and the second haplotype of the offspring to the two maternal haplotypes. During comparison, loci with unphased genotypes or genotypes that violated Mendelian inheritance rules in the offspring were disregarded. The inherited parental haplotypes were identified as those exhibiting the highest number of matched SNVs and were used to replace the original haplotype information in the offspring. For each locus of interest,

the offspring's genotype was determined from these inherited parental haplotypes.

Performance evaluation for haplotype linkage analysis and direct variant detection

To evaluate the performance of both haplotype linkage analysis and direct variant detection, familial high-confidence regions were first obtained by intersecting the high-confidence regions of each trio member using bedtools multiinter (v2.27.1) (29). The familial high-confidence regions were then intersected with protein-coding gene regions (extracted from the genome annotation file downloaded from https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_42/gencode.v42.annotation.gtf.gz) using bedtools intersect (v2.27.1) (29). Within the resulting intersected regions, loci with parental genotype combinations of heterozygous (0/1) and homozygous reference (0/0) were selected. The performances of the two methods were then assessed by evaluating whether the genotypes of the offspring at these loci could be accurately identified from the inferred parental haplotypes (haplotype linkage analysis) or from default variant calling results (direct variant detection).

De novo mutation screening

Genome-wide DNM screening was restricted to biallelic SNV loci within familial high confidence gene regions as detailed above. DNMs were identified by comparing the genotypes of the trio members. A locus was classified as DNM if both parental genotypes were homozygous reference while the offspring showed a different genotype. Further detailed analysis of the identified DNMs were done with a customized R script.

Aneuploidy profiling

Aneuploidy profiling was performed using Nano-GLADIATOR (v1.0) (30) with a window size of 1,000,000 base pairs.

To determine parental haplotype contributions across the genome, we selected SNV loci where the parents exhibited differing homozygous genotypes (homozygous reference (0/0) for one parent and homozygous alternate (1/1) for the other). Only loci with GQ > 2 and depth (DP) between 5 and 50 in the offspring were retained. For each locus, we computed the paternal and maternal allele fractions for the offspring. If a parent's genotype was 1/1, the allele frequency (AF) value for the offspring was used as the allele fraction for that parent. Conversely, 1 minus AF was used as the allele fraction if the parent's genotype was 0/0. Subsequently, we grouped loci within fixed bin sizes of 1 MB and calculated the average paternal and maternal allele fractions within each bin. Bins containing fewer than 30,000 loci were excluded. The calculated average paternal and maternal allele fraction values then underwent Circular Binary Segmentation using the R package PSCBS. Finally, we plotted the mean parental allele fraction values for each bin along with the segmentations across the genome. The parental

haplotype fraction for each chromosome was determined through visual inspection of the plot, enabling the identification of the parental origin of any aneuploidies.

To determine the mitotic or meiotic origin of an aneuploidy with identified parental origin, we analyzed SNV loci that were heterozygous in the parent causing the aneuploidy and, in the offspring, but homozygous in the other parent. For each locus in the offspring, the unique allele fraction (UAF) was inferred from the AF value, where the unique allele refers to the distinct allele among the four parental alleles. Loci were grouped within fixed bins of 1 MB, and the average UAF was calculated for each bin. Bins with fewer than 30,000 loci were excluded from the analysis. The average UAF was subjected to Circular Binary Segmentation using the R package PSCBS. Finally, the average UAF for each bin, along with segmentations, was plotted across the genome. The mitotic or meiotic origin of the aneuploidy was then inferred through visual inspection of the plot.

Results

SNV and indel calling with long-read whole-genome sequencing data from single cells

To evaluate the potential of lrWGS-based haplotyping for single cells, we took one single-cell and one multi-cell (10 cells) sample from the offspring of the GIAB trio, mimicking single blastomere and TE biopsy, respectively. Additionally, a bulk sample was included for comparison **(Figure 1)**. We obtained 22-24x lrWGS data for the bulk, multi-cell, and single-cell samples, covering 95%, 94%, and 88% of the human genome, respectively.

Given that SNVs and indels are the primary focus of most PGT-M cases and serve as genetic markers for haplotype construction, we first performed SNV and indel calling using lrWGS data. Variant calling performance was assessed by comparing with GIAB benchmark data. Across all sample types, we observed better variant calling performance for SNVs than for indels **(Figure 2A)**. Among different sample types, single-cell data showed the lowest performance, while multi-cell data was more similar to bulk data. Specifically, multi-cell data exhibited an SNV F-measure of 0.9605, comparable to bulk data at 0.9925, whereas the F-measure for single-cell data decreased significantly to 0.6305 **(Figure 2A)**. Not surprisingly, for single-cell data, the sensitivity for heterozygous SNVs was notably lower (0.3844) compared to homozygous SNVs (0.8422) **(Supplementary Figure 1)**, suggesting a high rate of ADO.

Both single- and multi-cell data contain a high percentage of chimeric reads (55% and 48% for single- and multi-cell data, respectively) **(Supplementary Figure 2A)**. This is expected due to the nature of MDA amplification which is characterized by the formation of chimeric DNA rearrangements. During mapping, each chimeric read was fragmented into multiple smaller segments and mapped to their original positions within the genome. Among these alignments, one was selected as the representative alignment, while all others were classified as

supplementary alignments **(Supplementary Figure 2B)**. By default, supplementary alignments were not utilized for variant calling. We hypothesized that incorporating supplementary alignments might enhance coverage in specific genomic regions and improve SNV calling performance. Hence, we conducted tests that deliberately included supplementary alignments for variant calling. In contrast to expectation, we noted slightly reduced SNV F-measures compared to results obtained without including supplementary alignments **(Supplementary Figure 2C)**. Supplementary alignments were thus not utilized for variant calling throughout this study.

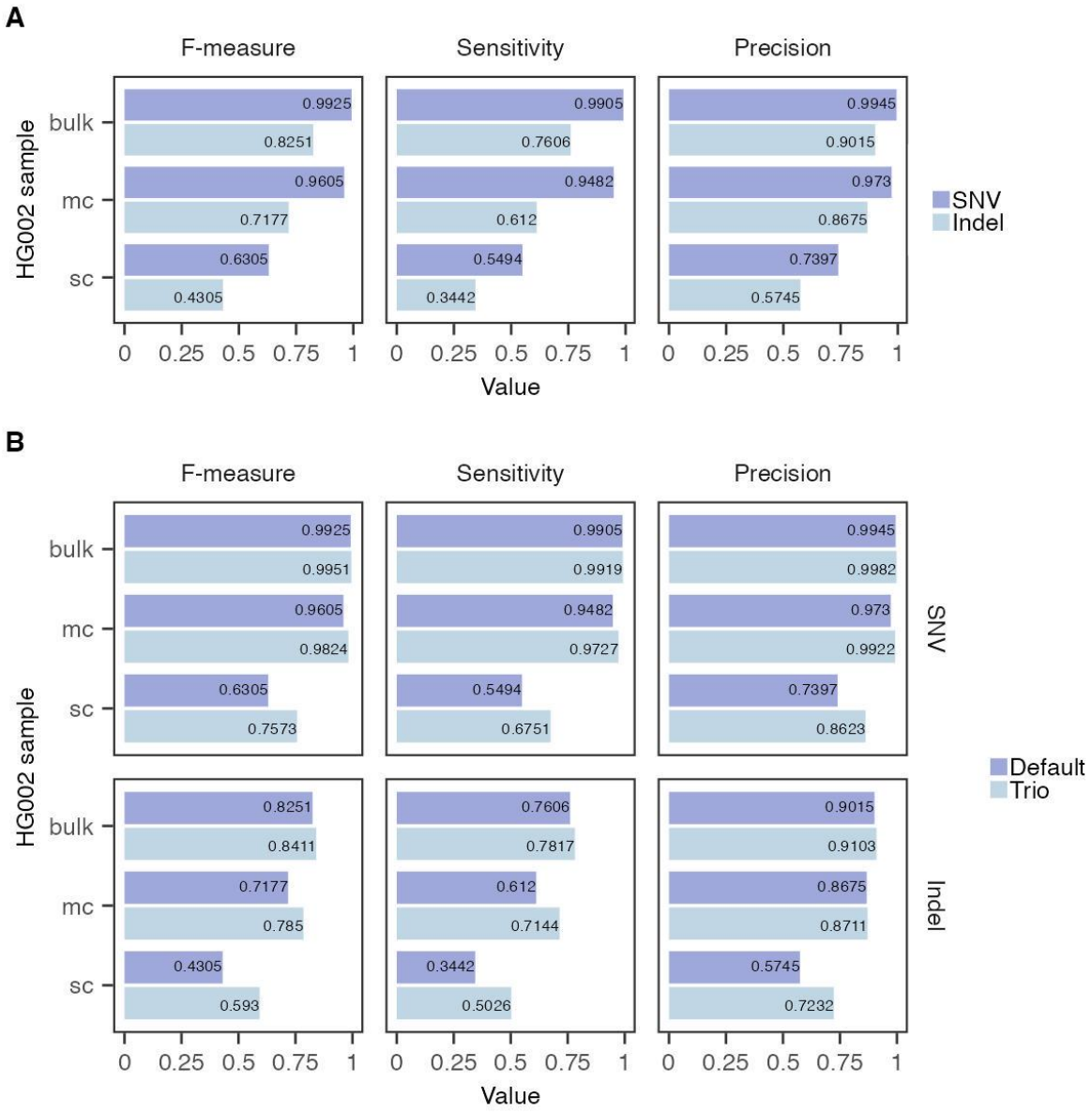


Figure 2. Small variant (SNV and indel) calling performance for lrWGS data from single-cell (sc), multi-cell (mc), and bulk samples of the offspring. (**A**) Variant calling performance in individual samples. (**B**) Trio variant calling yields varying degrees of improvements in variant calling performance for both SNVs and indels. F-measure represents the harmonic mean of precision and sensitivity.

High-quality SNVs and indels are required to enable haplotype phasing. Since both single- and multi-cell data showed lower SNV and indel calling performance compared to bulk data **(Figure 2A)**, we aimed to improve variant calling performance by incorporating parental data (publicly available lrWGS ONT data with ~24x coverage, see methods) to enable trio information-aware variant calling and reduce Mendelian inheritance violation variants. Trio variant calling was performed using Clair3-Trio (26), resulting in varying degrees of improvements in variant calling performance for the offspring, with the most substantial increase observed for single-cell data **(Figure 2B)**. Parental variant detection performance did not yield obvious benefits from trio variant calling **(Supplementary Figure 3)**.

Long-read whole-genome sequencing of single cells enables accurate phasing of the human genome

The next step is to achieve accurate phasing. We used the pedigree phasing mode of WhatsHap (28) to achieve high phasing performance by incorporating parental data, combining both read-based and genetic phasing. Meanwhile, we also tested the default phasing mode of WhatsHap (28), which relies solely on read-based phasing. We then compared the trio-based variant calling and phasing results with those obtained using the default settings. We tested phasing with only SNVs and with both SNVs and indels, as these two variant types demonstrated different variant calling performances. When phasing only SNVs, 95-98% of heterozygous SNVs were phased into 2,350, 59,763, and 121,116 blocks, with corresponding block NG50 of 2,964,293, 29,851, and 0 bp, covering 90%, 64%, and 40% of the autosomal regions for bulk, multi-cell, and single-cell data respectively. Adding indels for phasing achieved similar statistics (**Figure 3A-E**). When performing pedigree phasing the outcome was spectacularly improved, especially for multi- and single-cell data (**Figure 3A-E**). The most significant improvement was the generation of chromosome-long haplotypes with one block per autosome, covering 98% of the autosomal region for all data types (**Figure 3C and 3E**)**.** We assessed the accuracy of the phased blocks by comparing them with the phased GIAB benchmark data and obtained switch error rates as an indication of phasing accuracy. A lower switch error rate indicates higher accuracy. We observed higher switch error rates in the phased blocks when phasing both SNVs and indels compared to phasing only SNVs, likely due to lower indel calling performance. Compared to default phasing, pedigree phasing resulted in decreased accuracy for bulk data but improved accuracy for multi- and single-cell data. For bulk data, the switch error rates increased from 0.09% to 0.11% when phasing only SNVs and from 0.22% to 0.49% when phasing both SNVs and indels. In contrast, for multi-cell data, the switch error rates decreased from 0.40% to 0.11% when phasing only SNVs and from 0.54% to 0.44% when phasing both SNVs and indels. For single-cell data, the rates decreased from 1.31% to 0.19% when phasing only SNVs and from 1.46% to 0.50% when phasing both

SNVs and indels (**Figure 3F**).

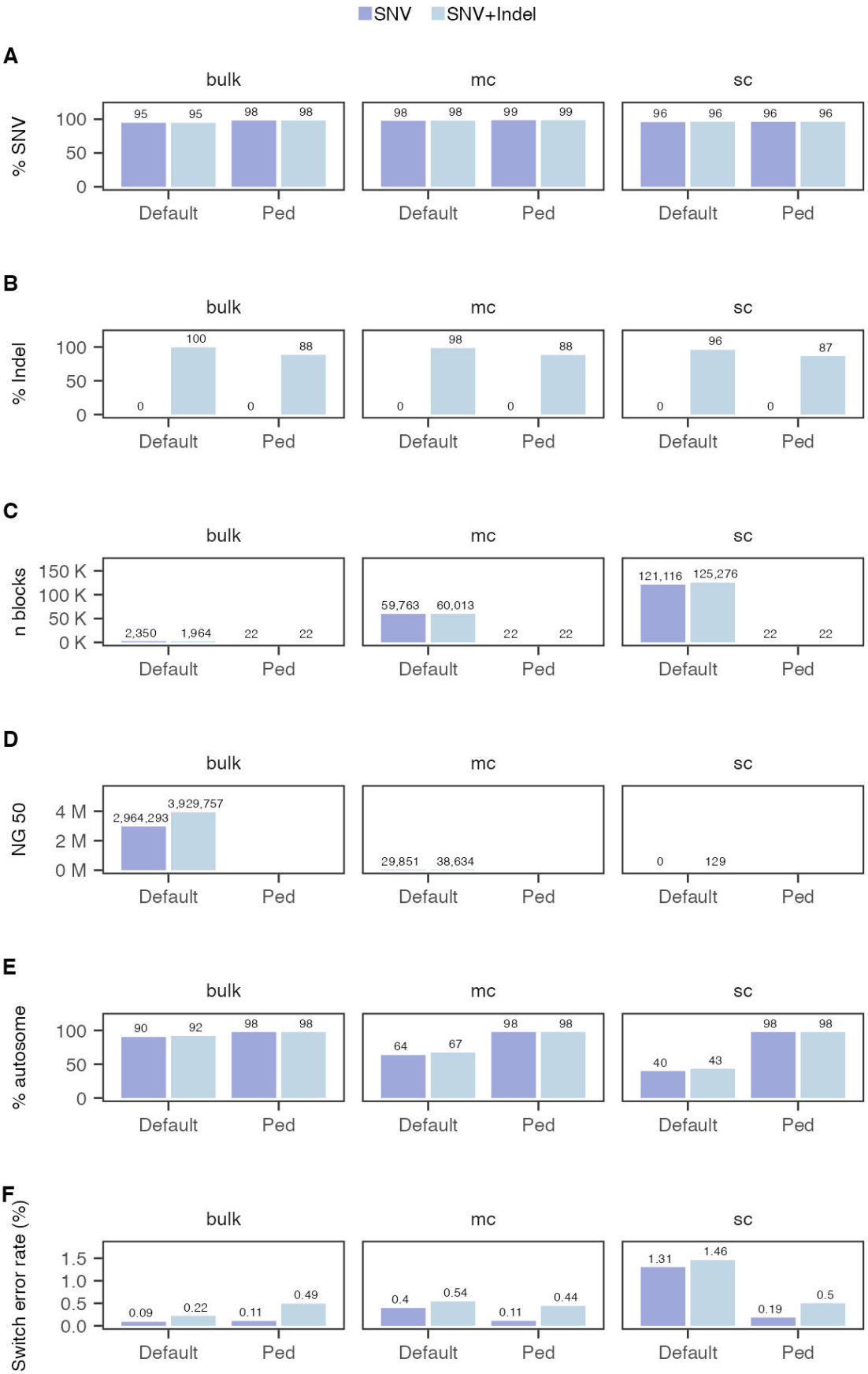


Figure 3. Phasing performance is superior with the pedigree phasing mode (ped) compared to the default phasing mode (default) for single-cell (sc) and multi-cell (mc) data of the offspring. **(A)** Percentage of phased heterozygous SNVs. (**B)** Percentage of phased heterozygous indels. (**C)** Total number of phased blocks. **(D)** NG50 of phased blocks. (**E)** Percentage of autosomal region covered by phased blocks. **(F)** Switch error rate. Shown are statistics for autosomes. For NG50, values are calculated per chromosome and averaged across autosomes. NG50 values are not applicable for pedigree phasing results because each chromosome has only one phased block. For each phasing mode, purple bars represent results when phasing only SNVs, while blue bars represent results when phasing both SNVs and indels.

Performance comparison of direct variant detection and haplotype-based variant inference

Given the promising quality of trio-based variant calling and phasing from single- and multi-cell data, we hypothesized that the resulting haplotypes could be used to infer parental haplotypes inherited by the offspring, enabling genotype extraction based on these inferred parental haplotypes. To infer the transmitted parental haplotypes, we compared parental haplotypes from default phasing mode with haplotypes of the offspring from pedigree phasing mode. Haplotype comparison was conducted within 1Mb segments across the genome (methods; **Figure 1**). Compared to the transmitted parental haplotypes inferred from the offspring's bulk data phasing results, those inferred from the offspring's multi- and single-cell data phasing results demonstrated consistencies of 95% and 82% when phasing only SNVs, and 93% and 82% when phasing both SNVs and indels, respectively **(Supplementary Figure 4)**.

We then assessed the specific added value of using long reads for phasing. With short-read or SNP arrays data, genetic phasing is typically applied within a pedigree. This method can phase variant loci that are heterozygous in one parent and homozygous in the other, as well as loci exhibiting different homozygous genotypes in the parents. However, loci that are heterozygous across all trio members remain unresolved. In contrast, long-read data have the potential to phase these loci. We evaluated the proportions of these loci among all loci used for haplotype comparison (**Supplementary Figure 5**). The results showed that 15-19% of informative SNVs and 14-17% of informative indels were heterozygous across trio members and could only be phased with long reads (**Supplementary Figure 5**). This indicates that lrWGS data enable more loci to be phased and utilized for subsequent analyses.

Next, we compared the performance of two variant detection approaches: direct variant detection and haplotype-based analysis. In direct variant detection, variants are identified from default variant calling results, whereas in haplotype-based analysis, variants are inferred from the transmitted parental haplotypes **(Figure 1)**. Using GIAB benchmark data, we selected loci within high-confidence protein-coding gene regions where one parent is homozygous reference (0/0) and the other is heterozygous (0/1). The offspring can be either heterozygous

(0/1) or homozygous reference (0/0), allowing us to evaluate the performance of both approaches in detecting transmitted coding variants. In total, we selected 812,973 SNV loci, with 407,232 heterozygous and 405,741 homozygous reference in the offspring, and 102,465 indel loci, with 50,650 heterozygous and 51,815 homozygous reference in the offspring. Using direct variant detection, we correctly identified 94% of heterozygous SNVs in multi-cell data and 43% in single-cell data, with no false positives for homozygous reference loci **(Figure 4A)**. Performance for indel loci was lower, with 68% of heterozygous indels detected in multi-cell data and 26% in single-cell data, along with 1% false positives for homozygous reference loci **(Figure 4B)**. In contrast, haplotype-based analysis demonstrated superior performance over direct variant detection, accurately inferring 98% of heterozygous SNVs in multi-cell data and 92% in single-cell data, with 2-3% false positives for homozygous reference loci **(Figure 4A)**. The performance for indel loci was still lower, identifying 78% of heterozygous indels in multi-cell data and 74% in single-cell data, with 4-5% false positives for homozygous reference loci **(Figure 4B)**. Multi-cell data demonstrated direct variant detection performance comparable to bulk data and also benefited from haplotype-based analysis, identifying more heterozygous loci despite a small rise in false positives for homozygous reference loci **(Figure 4)**.

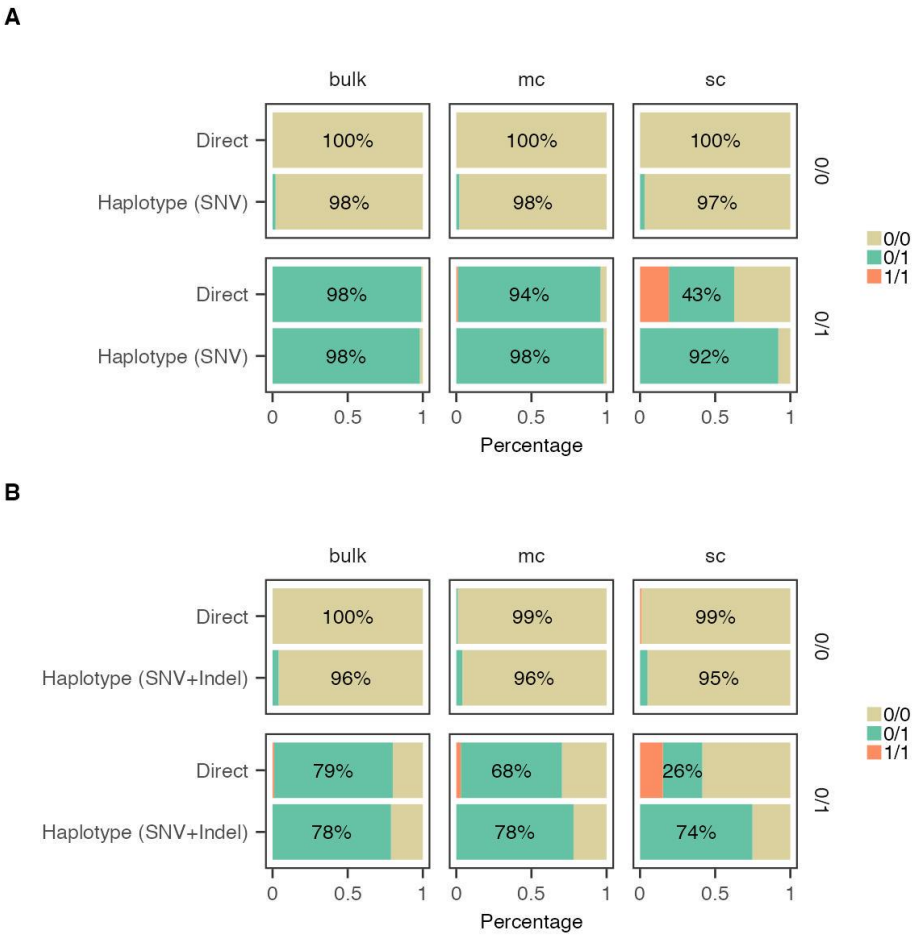


Figure 4. Performance comparison of direct variant detection and haplotype-based variant inference on selected (A) SNV and (B) indel loci using single-cell (sc), multi-cell (mc), and bulk data of the offspring. Direct: direct variant detection. Haplotype (SNV): haplotype-based analysis with only SNVs phased in the haplotypes. Haplotype (SNV + Indel): haplotype-based analysis with both SNVs and indels phased in the haplotypes. The true genotypes from GIAB benchmark data are indicated on the right. The proportions of inferred genotypes are color-coded. 0/0 for homozygous reference, 0/1 for heterozygous, and 1/1 for homozygous alternate.

Exploration of de novo mutation screening

An essential aspect of single-cell genomics is exploring new mutations that arise in individual cells during processes such as cell division, cancer genesis and tissue differentiation. Understanding DNM rates at the single-cell level will deepen our knowledge of the heterogeneity of both cancerous and normal cells, as well as the mechanisms driving cancer progression. Additionally, a large proportion of common and rare genetic disorders are a consequence of DNMs (Veltman and Brunner, 2012). Therefore, we assessed the efficacy of lrWGS data for genome-wide DNM screening, focusing on SNVs. We identified 452 heterozygous SNVs in the offspring as “true DNMs” within high-confidence protein-coding regions of the GIAB benchmark data, with both parents exhibiting homozygous reference

genotypes at the corresponding loci. This is significantly higher than expected, since the number of new point mutations present in human offspring is on average 60 (30–90 depending on parental age at conception), with around one DNM per exome (31–34). Hence, most are likely false positives or mosaic variants introduced by cell culture. Of the 452 DNMs detected in the benchmark data, 407 (90%) were identified using multi-cell data, while 167 (37%) were detected with single-cell data from the offspring. However, the false positive rates were 87,434 out of 87,841 (99.5%) for multi-cell data and 321,201 out of 321,368 (99.9%) for single-cell data. In summary, the abundance of false positives complicates DNM screening when using single-cell or multi-cell lrWGS data of the offspring for DNM screening.

Direct variant detection, haplotype-based variant inference, and haplotype-aware aneuploidy profiling in human embryos using lrWGS data

Since the proof-of-concept study with GIAB cell lines demonstrated promising performance, we proceeded to test lrWGS-based PGT on TE biopsies from five human embryos derived from two different couples. Each TE biopsy contains 5-10 cells, corresponding to the multi-cell sample in above benchmark study. The DNA was previously analyzed using clinically accredited SNP array-based comprehensive PGT, and the results were used as a reference. For family ONT1 the father carried a pathogenic SNV (c.1384C>T) in the *MSH2* gene causing Lynch syndrome, an autosomal dominant cancer predisposition syndrome. For family ONT2, the mother carried an indel (c.2955delG), and the father carried a pathogenic SNV (c.1133-708A>G) in the *LAMB3* gene, responsible for autosomal recessive junctional epidermolysis bullosa. Two embryos from family ONT1 (ONT1-E02, ONT1-E03) and three from family ONT2 (ONT2-E04, ONT2-E06, ONT2-E20) were tested using lrWGS-based PGT. We obtained 21-31x coverage following lrWGS of the parents and embryos, covering 93-95% of the human genome. With direct mutation detection, the carrier status of the variant alleles was accurately determined, except for one indel that was missed in ONT2-E20. However, we identified 3 out of 5 reads supporting the presence of this deletion **(Supplementary Table 1)**. Haplotype-based analysis yielded 100% concordance with SNP array-based PGT-M results **(Figure 5)**. SNVs were phased to identify the risk parental haplotypes carrying the disease alleles and to infer the parental haplotypes inherited by the embryo, from which the embryo's carrier status can be determined **(Figure 5)**. For the indel in the mother of family ONT2, the risk haplotype carrying the indel was identified by phasing both the indel and the SNVs. The indel locus was then incorporated into the SNV phasing results, with each allele assigned to the corresponding maternal haplotype **(Figure 5B)**.

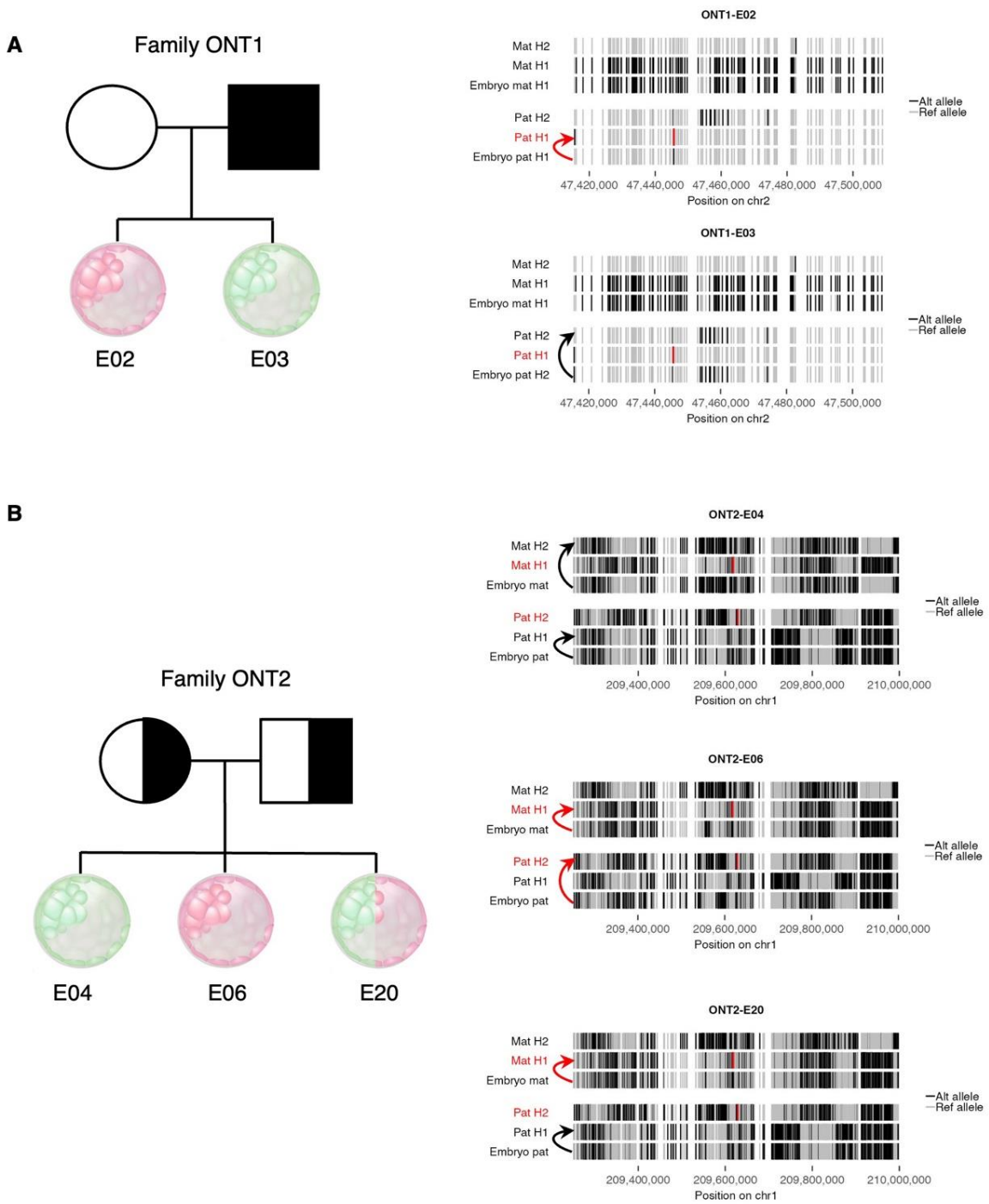


Figure 5. Visualization of haplotype-based PGT-M in human embryos. Pedigree plots for the two PGT families are shown on the left. Embryos are color-coded based on their carrier status as determined by haplotype-based PGT-M: green indicates that the disease allele was not inherited, red indicates that the dominant mutation was inherited or both recessive mutations were inherited, and green/red mosaic indicates the inheritance of a single recessive mutation. SNV haplotyping results for the chromosomal region linked to the disease loci are shown on the right. In the maternal (Mat H1, Mat H2) and paternal (Pat H1, Pat H2) haplotypes, each vertical line represents an allele from an informative locus: grey indicates the reference allele, black indicates the alternate allele, and disease alleles are shown in red. Parental haplotypes that carry disease alleles are labeled in red text. Arrows link the embryonic

haplotypes to their corresponding maternal and paternal haplotypes, with red arrows used when the corresponding parental haplotypes carry the disease alleles. For the indel in the mother of family ONT2, both SNVs and indels were first phased to obtain phasing information for the indel locus. This locus was then added to the SNV phasing results, with each allele assigned to its corresponding haplotype. **(A)** In Family ONT1, ONT1-E02 inherited the paternal haplotype carrying the disease allele, while ONT1-E03 inherited the normal paternal haplotype. **(B)** In Family ONT2, ONT2-E04 inherited normal haplotypes from both parents, ONT2-E06 inherited pathogenic allele-carrying haplotypes from both parents, and ONT2-E20 inherited the normal haplotype from the father and the pathogenic allele-carrying haplotype from the mother.

Since SNP array-based comprehensive PGT allows for the concurrent analysis of aneuploidy (PGT-A) and enables the identification of mosaic aneuploidy and the origin of aneuploidy, we explored whether these can also be detected by lrWGS data. We identified all aneuploidies and their mechanistic origins using lrWGS data, achieving 100% consistency with SNP array-based PGT. Specifically, in ONT1-E02 we observed complex segmental aberration on chromosome 2, with putative mosaic duplication of short arm and mosaic deletion of the long arm. In addition, the same embryo exhibited complex segmental aberrations on chromosome 14 **(Supplementary Figure 6)**. Additionally, we detected trisomy on chromosomes 9 and 16 in ONT2-E20 **(Figure 6A)**, with parental allele fractions indicating the presence of an extra maternal chromosome **(Figure 6B)**. To determine the mitotic or meiotic origin of the extra maternal chromosomes, we analyzed loci with homozygous genotypes in the father and heterozygous genotypes in the mother and the embryo. For each locus, we calculated the unique allele fraction (UAF) in the embryo, with the unique allele being the distinct allele among the four parental alleles. The distribution of the unique allele fraction across the chromosome helps identify the mechanistic origin of the trisomy **(Figure 6C)**. Both chromosomes 9 and 16 were inferred to originate from maternal meiotic I nondisjunction **(Figure 6D)**, consistent with SNP array results.

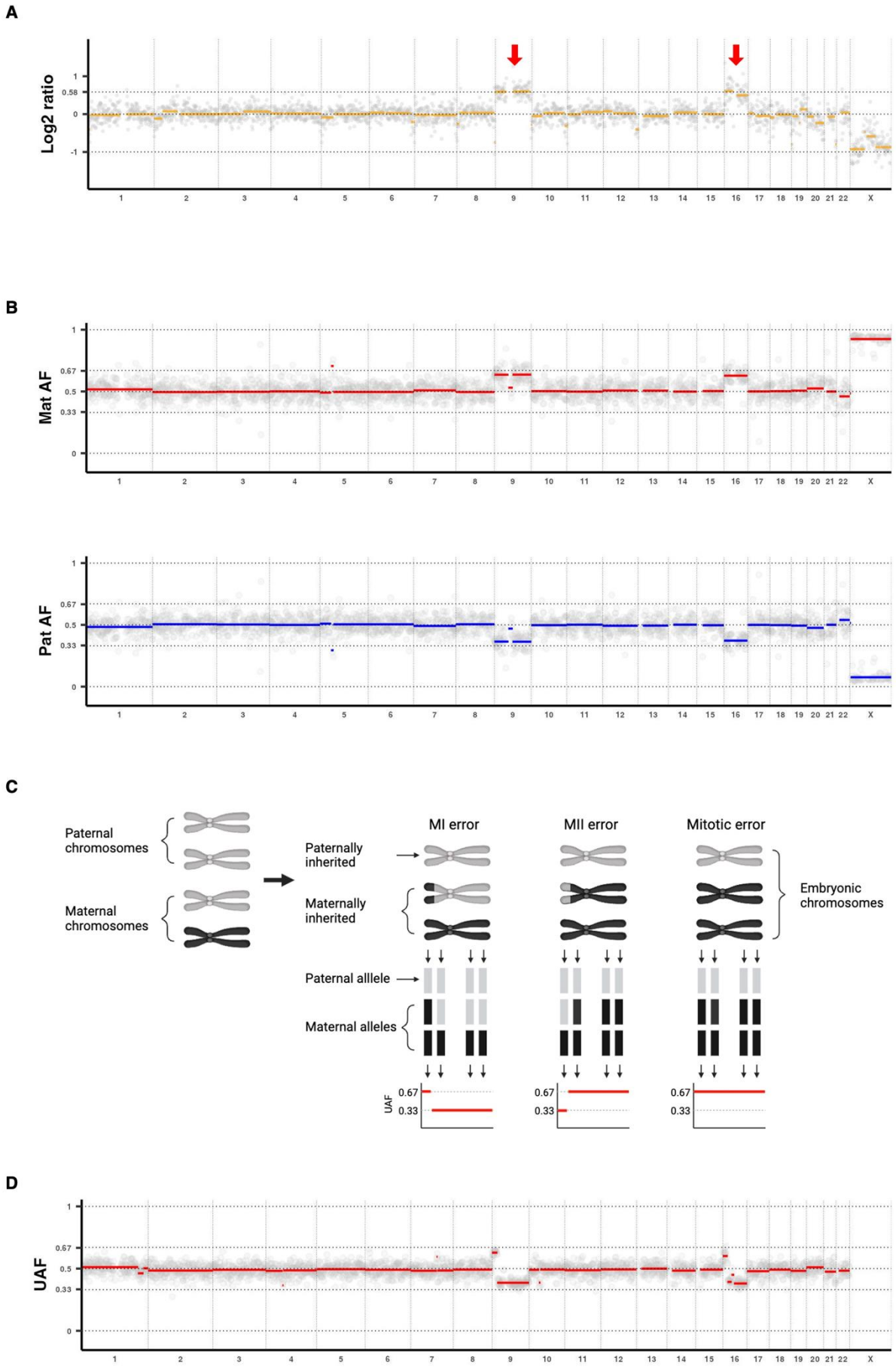

A
Log2 ratio
1
0.58
0
-1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
B
Mat AF
1
0.67
0.5
0.33
0
Pat AF
C
Paternal chromosomes
Maternal chromosomes
MI error
MII error
Mitotic error
Paternally inherited
Maternally inherited
Embryonic chromosomes
Paternal alllele
Maternal alleles
UAF
0.67
0.33
D
UAF

Figure 6. PGT-A analysis using lrWGS data identified aneuploidies and their mechanistic origins in ONT2-E20. (A) The copy number plot shows trisomy on chromosomes 9 and 16. **(B)** Paternal allele fractions (Pat AF) and maternal allele fractions (Mat AF) were assessed across the genome for loci homozygous in one parent and heterozygous in the other, reflecting parental genome contributions across the embryonic genome. In the trisomic regions, Mat AF = 0.67 and Pat AF = 0.33, indicating one additional chromosome from the maternal side for chromosomes 9 and 16. **(C)** A schematic illustrates the reasoning used to infer the mechanistic origin of a maternally derived trisomy. Loci with homozygous genotypes in the father and heterozygous genotypes in both the mother and the embryo are selected. For each shown heterozygous locus of the embryo, grey bar(s) represent the allele matching the homozygous paternal allele, while black bar(s) represent the unique allele from the heterozygous maternal locus that differs from the homozygous paternal allele. The unique allele fraction (UAF)—the fraction of this unique allele in the embryo—is inferred from the allele frequency (AF) value in variant calling results, which is expected to be 0.33 if two different maternal alleles were inherited and 0.67 if two same unique maternal alleles were inherited. The pattern of UAF values across the trisomic chromosome indicates the mechanistic origin of the trisomy. MI: meiotic I; MII: meiotic II. **(D)** UAF values across the genome. For trisomic chromosomes 9 and 16, the UAF is 0.67 at the beginning and decreases to 0.33 across the remaining chromosomal regions. This pattern indicates that the origin of these two trisomies is maternal meiotic I nondisjunction.

In summary, we confirmed in human embryos that lrWGS can be used for concurrent PGT-M and PGT-A analysis. For PGT-M, while direct variant detection missed an inherited indel in one of the embryos, inclusion of haplotype-based analysis mitigated this drawback, resulting in 100% concordance with SNP array-based PGT results. For PGT-A, lrWGS data enables not only the detection of aneuploidies but also their parental and mechanistic origins.

Discussion

Here, we explored the characteristics of lrWGS data from single cells and evaluated its performance for variant calling and reference-free haplotyping. Our benchmark analysis revealed that with lrWGS data from TE biopsies (10 cells), direct mutation analysis has a 94% probability of identifying an inherited SNV. For haplotype-based analysis, the probabilities are 98% for TE biopsies (10 cells) and 92% for single blastomere biopsies (1 cell). Considering that TE biopsy is becoming the golden standard, lrWGS-based PGT can thus enable direct variant detection coupled with haplotype-based analysis for increased diagnostic accuracy. Using human embryos, we validated the high performance of lrWGS-based PGT-M and concurrent PGT-A, with the mechanistic origins of aneuploidies correctly identified.

Correctly matching the haplotypes of the offspring to those of their parents is essential for effective haplotype linkage analysis. Key factors contributing to successful matching include: 1) Attainment of high-quality variant calling and phasing outcomes for the parents, which played a pivotal role in identifying high-risk and low-risk haplotypes. 2) Trio variant calling for

the offspring predicted significantly fewer Mendelian inheritance violation loci. 3) Pedigree-based phasing for the offspring generated chromosome-spanning haplotypes with increased phasing accuracy by combining read based phasing and genetic phasing. 4) The increased density of informative loci in our study, primarily due to loci that could only be phased using long reads, enabled successful haplotype matching over relatively short genome regions. 5) Exclusion of loci violating Mendelian rules from haplotype comparison. 6) Flexibility in the haplotype matching process, allowing for discrepancies between the offspring's haplotype and the inferred parental haplotype.

The main advantage of using parental lrWGS data for phasing is that it allows for direct phasing of the parents, thereby enabling genetic phasing of single cells from the offspring without requiring additional family members. However, genetic phasing within a trio cannot resolve loci where all trio members are heterozygous (15). By incorporating lrWGS data from single cells of the offspring, we found that 14-19% of informative loci fall into this category. These additional informative loci increase the probability that loci affected by ADO in single-cell data can be correctly inferred from inherited parental haplotypes by enhancing the density of informative loci available for haplotype matching. The broad genome coverage of lrWGS data from single cells also enables direct variant detection. These advantages make lrWGS-based PGT a superior option compared to current PGT methods. Firstly, there is no need to obtain DNA from close relatives, which often causes delays, increases costs, and may not always be available. Secondly, lrWGS-based PGT enables the direct phasing of DNMs present in prospective parents, eliminating the need for additional steps such as analyzing single sperm, polar bodies (35–37) or affected sibling embryos (38). Thirdly, it allows for the direct detection of pathogenic variants in embryos, complementing haplotype-based analysis by resolving uncertain or inconclusive findings and addressing cases where the embryo is at risk of carrying a pathogenic mutation due to parental germline mosaicism. Fourthly, using a higher density of informative loci for analysis enhances the identification of inherited parental haplotypes within relatively short regions. Additionally, lrWGS data from single cells allows for the detection of both the parental origin and the mitotic or meiotic nature of chromosomal anomalies, providing valuable insights into the etiology of aneuploidies. Such information is crucial in clinical practice, as aneuploidies resulting from meiotic chromosome segregation errors rarely survive to term and often lead to adverse pregnancy outcomes; thus, selecting against these embryos could improve IVF success rates (39). The potential applications of lrWGS-based single-cell haplotyping and aneuploidy profiling go beyond human PGT and can be adapted for other species, such as equine and bovine, to improve reproductive outcomes. It also holds promise for cell-based noninvasive prenatal diagnosis by analyzing single fetal cells present in maternal blood (40).

DNMs arise during various biological and pathological processes, such as cell division and cancer development. Additionally, DNMs are a major cause of rare human disorders. Genome-wide DNM screening would be valuable for identifying these mutations. Using multi- and single-cell lrWGS data of the offspring, we identified 90% and 37% of DNMs in benchmark data, respectively. However, true DNMs represented only 0.5% and 0.1% of all identified DNM candidates, highlighting the abundance of false positives. These results demonstrate that whole-genome DNM screening with ONT lrWGS data remains a challenging task at present However, with increasing sequencing accuracy and methodological improvements, this is likely to be possible in the future. A previous pilot study attempted to use variant annotation databases and functional prediction algorithms to identify real pathogenic DNMs among numerous DNM candidates (41). Such strategies and additional quality metrics could be integrated into DNM screening to enhance detection accuracy. It is worth noting that we used cell line samples for DNM detection. Since DNMs arise during each cell division and increase with each passage of culturing (Londin et al., 2011), the culture process may have influenced the observed high incidence of DNMs.

In this study, we achieved high-quality phasing results with lrWGS data from the bulk sample of the offspring. Compared to a previous study that phased the autosomes of the same individual into 19,215 blocks with a switch error rate of 0.37% for SNVs and indels using 28x PacBio HiFi reads (5), we obtained 1,964 phased blocks with a lower switch error rate of 0.22% using ~24x ONT reads. This improvement is notable despite the lower variant calling performance of ONT reads compared to HiFi reads (F-measures: 0.9991 for SNVs and 0.9598 for indels with HiFi reads; 0.9925 for SNVs and 0.8251 for indels with ONT reads). The improvement in phasing performance could be attributed to advancements in the reference genome, updates in bioinformatics software, and enhancements in the GIAB benchmark data, as newer versions were utilized for this study.

A constant recombination rate of 1.26 cM/Mb was used during phasing, which is a suitable assumption for the human genome. However, the inherent errors in lrWGS data, especially those derived from single-cell and multi-cell samples, may hinder the accurate detection of actual recombination sites. Consequently, unidentified recombination events may affect haplotype comparison. Moreover, imperfect phasing results contain switch errors that may influence haplotype matching. To mitigate the impact of recombination events and switch errors on the inference of inherited parental haplotypes, we manually constrained the maximum comparison block length to 1 Mb in this study. Additionally, visual inspection of haplotype blocks can help identify recombination events and switch errors, further reducing the risk of misdiagnosis. With ongoing improvements in sequencing accuracy, read length, and bioinformatics algorithms, this constraint on block length may eventually become unnecessary,

making accurate phasing of the entire genome feasible.

This study has limitations and areas for potential improvement. First, we utilized ONT sequencing, a cost-effective long-read sequencing technology that is rapidly advancing in read length and accuracy. These improvements will enhance variation discovery and phasing performance, making it essential to conduct updated benchmark studies regularly. Second, more data from clinical applications are needed to further validate the practical utility of lrWGS-based haplotyping and aneuploidy profiling in single cells. As lrWGS becomes more widely adopted in clinical settings, performance evaluations from large-scale clinical PGT cycles could provide additional insights into the efficiency of lrWGS-based PGT in a broader range of clinical scenarios.

To summarize, we developed a bioinformatics pipeline that enables genome-wide concurrent haplotyping and aneuploidy profiling of single cells using lrWGS data, and validated its effectiveness for genome-wide, reference-free comprehensive PGT. Additionally, we evaluated the performance of lrWGS data from single cells for direct variant detection and DNM screening. Beyond PGT for human embryos, our bioinformatics pipeline has potential applications in other areas of single-cell genomics. For instance, it can be adapted for PGT in animal species like bovine and equine to improve reproductive outcomes, and for cell-based noninvasive prenatal testing by analyzing single circulating trophoblast cells in maternal blood (40).

Data availability

Raw data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAD50000000787. It is available to academic users upon request to the Data Access Committee (DAC) of KU Leuven via the corresponding author (JRV). We have provided the bioinformatical scripts via the following link: https://github.com/JorisVermeeschLab/ONT_PGT.git

Author contributions

J.R.V. conceptualized and designed the study. G.P. prepared cell line samples. O.T. and E.D. selected and prepared human PGT samples. G.P., T.J. and M.G. conducted library preparation and nanopore sequencing. Y.Z. did bioinformatics analysis, Y.Z., J.R.V., and O.T. interpreted the data. Y.Z., J.R.V. and E.S. wrote the manuscript. All co-authors reviewed and approved the manuscript.

Acknowledgements

The authors would like to thank the couples who participated in the study.

Funding

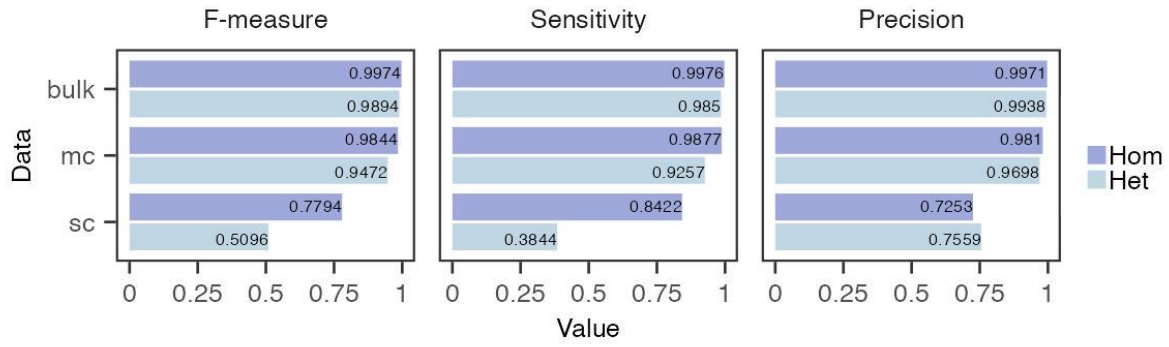
Funding was received from the Marie Skłodowska-Curie grant agreement No 813707 (MATER)

and from the KU Leuven, C1-C14/22/125 to J.R.V. Y.Z. was supported by the Marie Skłodowska-Curie grant agreement No 813707 (MATER).

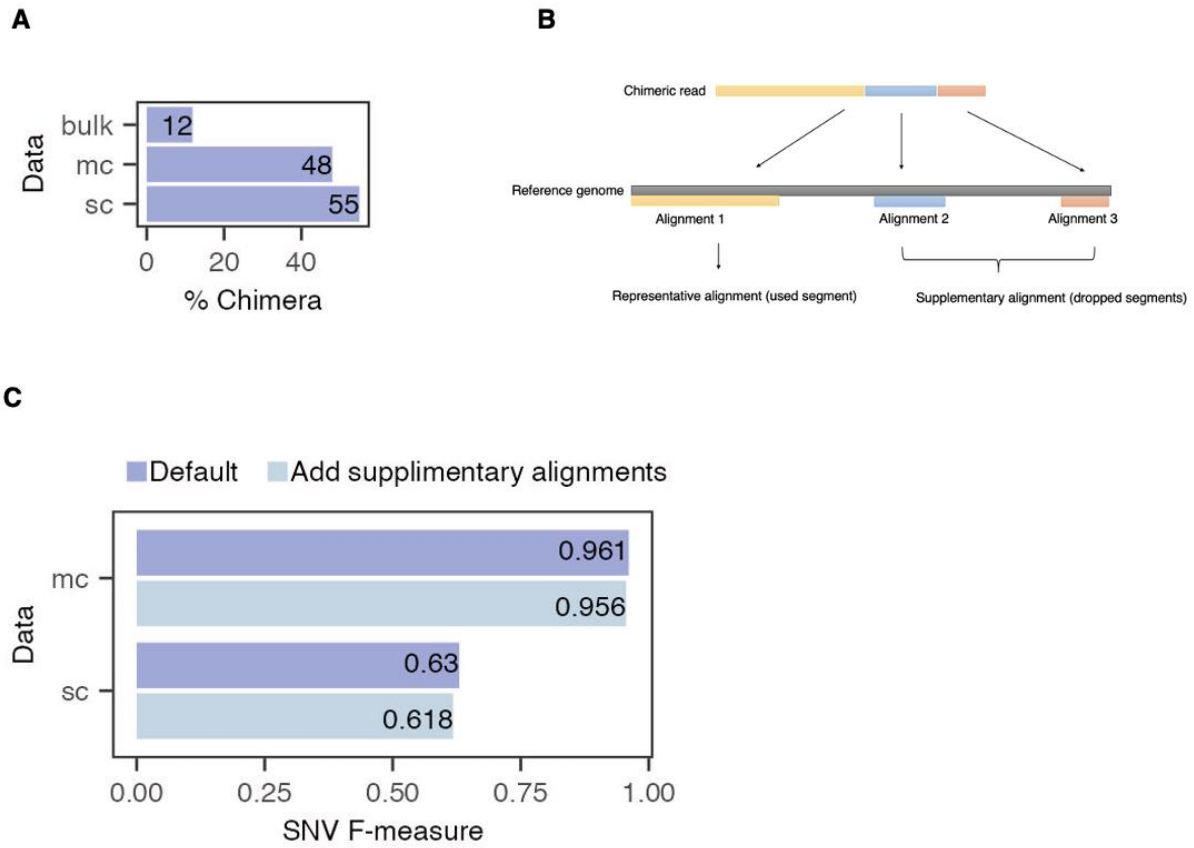
Conflict of Interest Disclosure

The authors declare no conflict of interest.

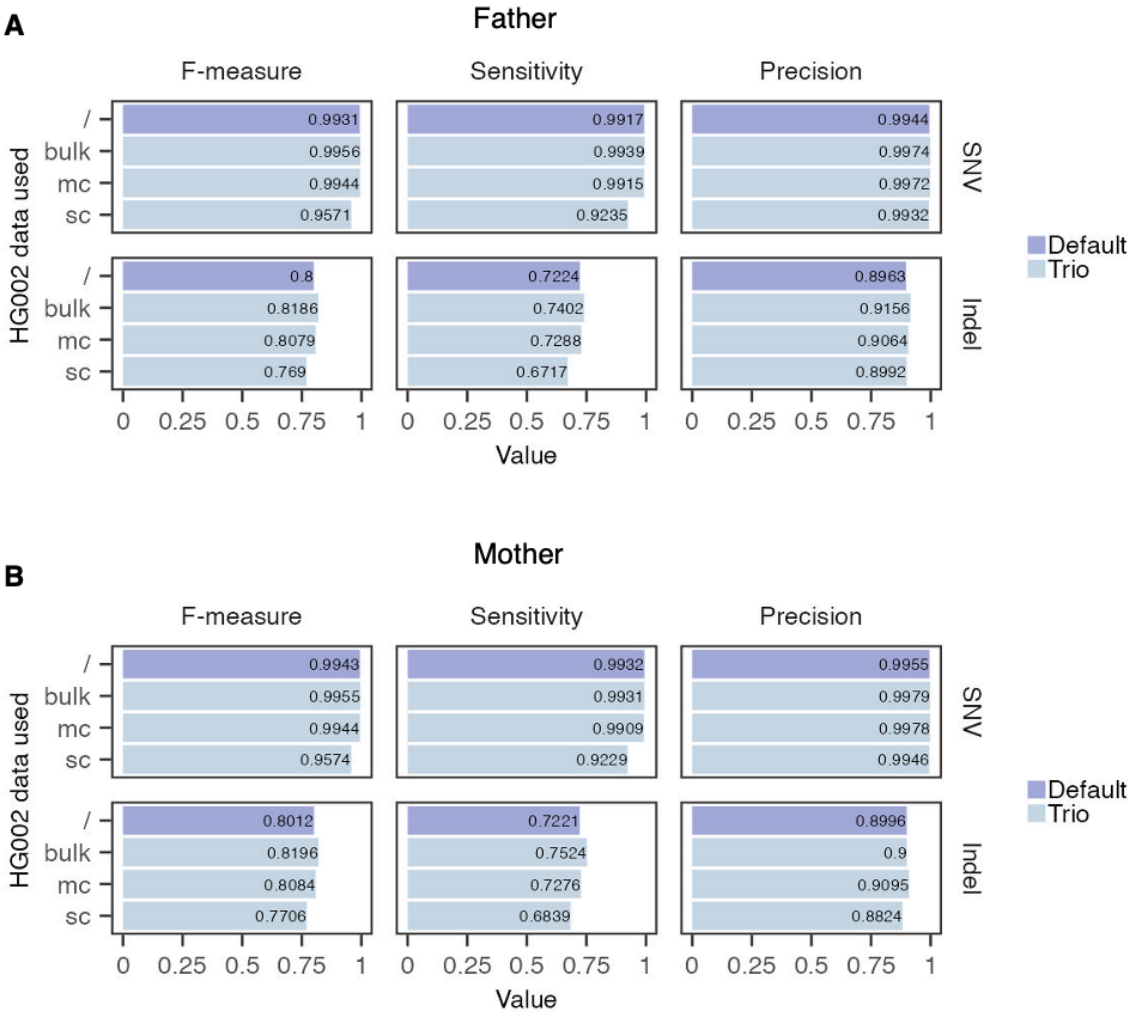
Supplementary data



Supplementary Figure 1. Heterozygous and homozygous SNV calling performance for lrWGS data from single-cell (sc), multi-cell (mc), and bulk samples of the offspring.



Supplementary Figure 2. Supplementary alignments did not improve SNV calling performance. **(A)** Percentage of chimeric reads in bulk, multi-cell (mc) and single-cell (sc) lrWGS data of the offspring. **(B)** Schematic demonstration of chimeric reads and supplementary alignments. **(C)** Comparison of default SNV calling F-measures with those obtained after including supplementary alignments for variant calling.

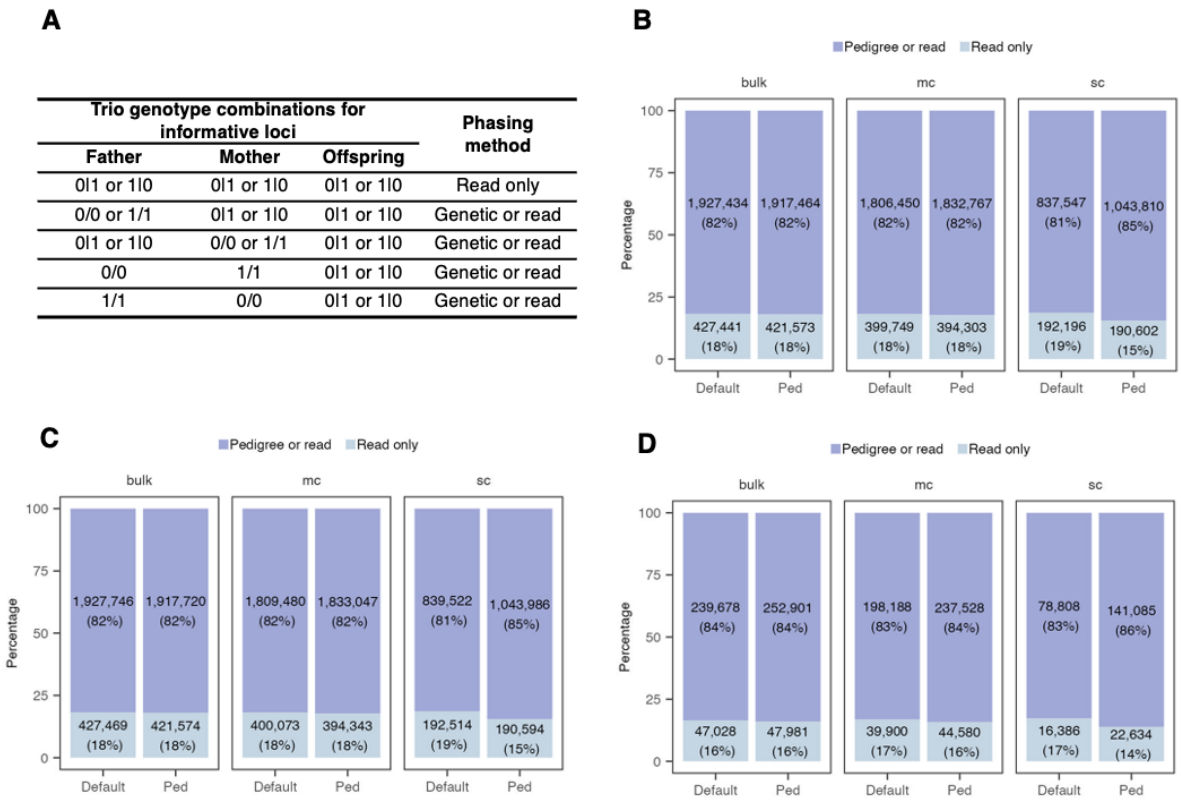


Supplementary Figure 3. Small variant (SNV and indel) calling performance obtained from default and trio variant calling for (A) the father and (B) the mother. On the y-axis, "bulk," "mc," and "sc" indicate bulk, multi-cell (mc), or single-cell (sc) data of the offspring were used for trio variant calling together with parental data. The "/" denotes only paternal or maternal data were utilized for variant calling.

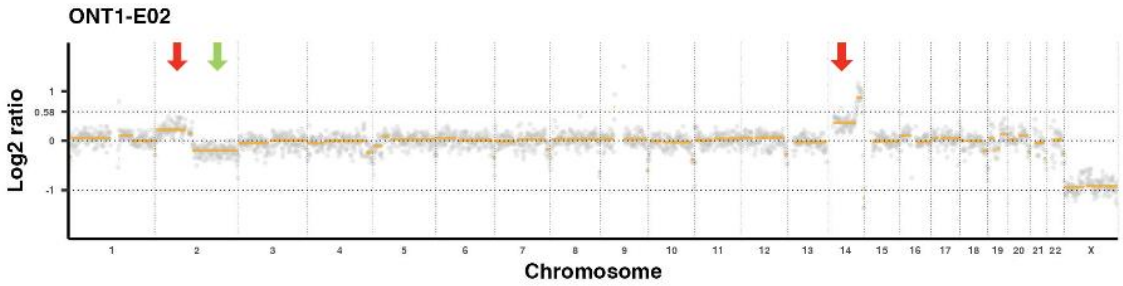
Supplementary Figure 4. Consistency of transmitted parental haplotypes inferred from the offspring's bulk, multi-cell (mc) and single-cell (sc) data phasing results when phasing (A) only SNVs (B) SNVs and indels. For each panel, vertical colored lines indicate inferred paternal (pat) or maternal (mat) haplotypes for specific regions. Different parental haplotypes are distinguished by different colors.

A

Trio genotype combinations for informative loci			Phasing method
Father	Mother	Offspring	
0|1 or 1|0	0|1 or 1|0	0|1 or 1|0	Read only
0/0 or 1/1	0|1 or 1|0	0|1 or 1|0	Genetic or read
0|1 or 1|0	0/0 or 1/1	0|1 or 1|0	Genetic or read
0/0	1/1	0|1 or 1|0	Genetic or read
1/1	0/0	0|1 or 1|0	Genetic or read



Supplementary Figure 5. Proportion of informative loci that can only be phased by long reads in pedigree (ped) or default phasing (default) results of bulk, multi-cell (mc) and single-cell (sc) data. **(A)** Trio genotype combinations for loci retained as informative loci for haplotype comparison. The "Phasing method" column indicates method(s) that could be used to phase variants loci of the offspring. Read only: variant loci could only be phased through read-based phasing. Genetic or read: variant loci could be phased by either genetic phasing or read-based phasing. **(B-D)** Informative **(B)** SNVs from SNV phasing results, **(C)** SNVs from SNV and indel phasing results, and **(D)** indels from SNV and indel phasing results retained for haplotype comparison, categorized by phasing method(s) that could be used to phase the variant loci.



Supplementary Figure 6. PGT-A analysis using lrWGS data identified putative mosaic chromosomal abnormalities in ONT1-E02. The copy number plot reveals mosaic duplication of the short arm of chromosome 2 (indicated by a green arrow) and mosaic deletion of the long arm (indicated by a red arrow). Additionally, mosaic duplications are observed on chromosome 4 (indicated by a red arrow).

Supplementary Table 1. Concordance between SNP array-based PGM results and lrWGS-based direct mutation detection

Family	Embryo	Pat/mat	Genetic lesion	Mutation position (hg38)	Inheritance pattern	PGT-M (SNP array)	PGT-M (lrWGS direct)
ONT1	ONT1-E02	Pat	*MSH2*: c.1384C>T	chr2:47445655	Autosomal dominant	inherited	inherited (22:38)
	ONT1-E03	Pat	*MSH2*: c.1384C>T	chr2:47445655	Autosomal dominant	not inherited	not inherited (32:1)
ONT2	ONT2-E04	Pat	*LAMB3*: c.1133-708A>G	chr1:209628898	Autosomal recessive	not inherited	not inherited (7:0)
		Mat	*LAMB3*: c.2955delG	chr1:209618003	Autosomal recessive	not inherited	not inherited (6:0)
	ONT2-E06	Pat	*LAMB3*: c.1133-708A>G	chr1:209628898	Autosomal recessive	inherited	inherited (7:18)
		Mat	*LAMB3*: c.2955delG	chr1:209618003	Autosomal recessive	inherited	inherited (11:7)
	ONT2-E20	Pat	*LAMB3*: c.1133-708A>G	chr1:209628898	Autosomal recessive	not inherited	not inherited (11:0)
		Mat	*LAMB3*: c.2955delG	chr1:209618003	Autosomal recessive	inherited	not inherited (2:3)

The "Pat/mat" column indicates whether the indicated mutation is from the paternal (Pat) or maternal (Mat) side. In the "PGT-M (lrWGS direct)" column, variant inheritance status is listed first, followed by the number of reads supporting each allele, presented in brackets as "n reads for reference allele : n reads for alternate allele". Any direct variant calling result discordant with SNP array results is in red.

References

1. Tewhey,R., Bansal,V., Torkamani,A., Topol,E.J. and Schork,N.J. (2011) The importance of phase information for human genomics. *Nature Reviews Genetics 2011 12:3*, **12**, 215–223.

2. Logsdon,G.A., Vollger,M.R. and Eichler,E.E. (2020) Long-read human genome sequencing and its applications. *Nature Reviews Genetics 2020 21:10*, **21**, 597–614.

3. Eid,J., Fehr,A., Gray,J., Luong,K., Lyle,J., Otto,G., Peluso,P., Rank,D., Baybayan,P., Bettman,B., *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science (1979)*, **323**, 133–138.

4. Mikheyev,A.S. and Tin,M.M.Y. (2014) A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour*, **14**, 1097–1102.

5. Wenger,A.M., Peluso,P., Rowell,W.J., Chang,P.C., Hall,R.J., Concepcion,G.T., Ebler,J., Fungtammasan,A., Kolesnikov,A., Olson,N.D., *et al.* (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology 2019 37:10*, **37**, 1155–1162.

6. Nurk,S., Koren,S., Rhie,A., Rautiainen,M., Bzikadze,A. V., Mikheenko,A., Vollger,M.R., Altemose,N., Uralsky,L., Gershman,A., *et al.* (2022) The complete sequence of a human genome. *Science (1979)*, **376**, 44–53.

7. Zhang,C.Z., Adalsteinsson,V.A., Francis,J., Cornils,H., Jung,J., Maire,C., Ligon,K.L., Meyerson,M. and Love,J.C. (2015) Calibrating genomic and allelic coverage bias in single-cell sequencing. *Nature Communications 2015 6:1*, **6**, 1–10.

8. Natesan,S.A., Bladon,A.J., Coskun,S., Qubbaj,W., Prates,R., Munne,S., Coonen,E., Dreesen,J.C.F.M., Stevens,S.J.C., Paulussen,A.D.C., *et al.* (2014) Genome-wide karyomapping accurately identifies the inheritance of single-gene defects in human preimplantation embryos in vitro. *Genet Med*, **16**, 838–845.

9. Zamani Esteki,M., Dimitriadou,E., Mateiu,L., Melotte,C., Van der Aa,N., Kumar,P., Das,R., Theunis,K., Cheng,J., Legius,E., *et al.* (2015) Concurrent whole-genome haplotyping and copy-number profiling of single cells. *Am J Hum Genet*, **96**, 894–912.

10. Esteki,Masoud.Z., Melotte,C., Coonen,E., Dreesen,J., Paulussen,A., Dumoulin,J., Van Uum,C., Drusedau,M., Engelen,J., Derks,K., *et al.* (2019) Agilent Technologies OnePGT solution: External verification on both blastomere and trophectoderm biopsies. *Reprod Biomed Online*, **38**, e11–e12.

11. Backenroth,D., Zahdeh,F., Kling,Y., Peretz,A., Rosen,T., Kort,D., Zeligson,S., Dror,T., Kirshberg,S., Burak,E., *et al.* (2019) Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (PGD) of monogenic disease and aneuploidy. *Genetics in Medicine*, **21**, 1390–1399.

12. Masset,H., Ding,J., Dimitriadou,E., Debrock,S., Tšuiko,O., Smits,K., Peeraer,K., Voet,T., Esteki,M.Z. and Vermeesch,J.R. (2022) Single-cell genome-wide concurrent haplotyping and copy-number profiling through genotyping-by-sequencing. *Nucleic Acids Res*, **50**, e63–e63.

13. Wu,H., Chen,D., Zhao,Q., Shen,X., Liao,Y., Li,P., Chiu,P.C.N. and Zhou,C. (2022) Long-read sequencing on the SMRT platform enables efficient haplotype linkage analysis in preimplantation genetic testing for β-thalassemia. *J Assist Reprod Genet*, **39**, 739–746.

14. Tsuiko,O., El Ayeb,Y., Jatsenko,T., Allemeersch,J., Melotte,C., Ding,J., Debrock,S., Peeraer,K., Vanhie,A., De Leener,A., *et al.* (2023) Preclinical workup using long-read amplicon sequencing provides families with de novo pathogenic variants access to universal preimplantation genetic testing. *Human Reproduction*, **38**, 511–519.

15. Zhang,P., Zhao,X., Li,Q., Xu,Y., Cheng,Z., Yang,L., Wang,H., Tao,Y., Huang,G., Wu,R., *et al.* (2024) Proband-independent haplotyping based on NGS-based long-read sequencing for detecting pathogenic variant carrier status in preimplantation genetic testing for monogenic diseases. *Front Mol Biosci*, **11**, 1329580.

16. Hård,J., Mold,J.E., Eisfeldt,J., Tellgren-Roth,C., Häggqvist,S., Bunikis,I., Contreras-Lopez,O., Chin,C.S., Nordlund,J., Rubin,C.J., *et al.* (2023) Long-read whole-genome analysis of human single cells. *Nature Communications 2023 14:1*, **14**, 1–12.

17. Hassold,T. and Hunt,P. (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics 2001 2:4*, **2**, 280–291.

18. Vanneste,E., Voet,T., Le Caignec,C., Ampe,M., Konings,P., Melotte,C., Debrock,S., Amyere,M., Vikkula,M., Schuit,F., *et al.* (2009) Chromosome instability is common in human cleavage-stage embryos. *Nat Med*, **15**, 577–583.

19. Bhatt,S.J., Marchetto,N.M., Roy,J., Morelli,S.S. and McGovern,P.G. (2021) Pregnancy outcomes following in vitro fertilization frozen embryo transfer (IVF-FET) with or without preimplantation genetic testing for aneuploidy (PGT-A) in women with recurrent pregnancy loss (RPL): a SART-CORS study. *Hum Reprod*, **36**, 2339–2344.

20. Madjunkova,S., Sundaravadanam,Y., Antes,R., Abramov,R., Chen,S., Yin,Y., Zuzarte,P.C., Moskovtsev,S.I., Jorgensen,L.G.T., Baratz,A., *et al.* (2020) Detection of Structural Rearrangements in Embryos. *New England Journal of Medicine*, **382**, 2472–2474.

21. Tan,V.J., Liu,T., Arifin,Z., Pak,B., Tan,A.S.C., Wong,S., Khor,C.C., Yang,H., Lee,C.G., Huang,Z., *et al.* (2023) Third-Generation Single-Molecule Sequencing for Preimplantation Genetic Testing of Aneuploidy and Segmental Imbalances. *Clin Chem*, **69**, 881–889.

22. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

23. De Coster,W. and Rademakers,R. (2023) NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, **39**.

24. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

25. Zheng,Z., Li,S., Su,J., Leung,A.W.S., Lam,T.W. and Luo,R. (2022) Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science 2022 2:12*, **2**, 797–803.

26. Su,J., Zheng,Z., Ahmed,S.S., Lam,T.W. and Luo,R. (2022) Clair3-trio: high-performance Nanopore long-read variant calling in family trios with trio-to-trio deep neural networks. *Brief Bioinform*, **23**, 1–12.

27. Cleary,J.G., Braithwaite,R., Gaastra,K., Hilbush,B.S., Inglis,S., Irvine,S.A., Jackson,A., Littin,R., Rathod,M., Ware,D., *et al.* (2015) Comparing Variant Call Files for Performance Benchmarking of Next-Generation Sequencing Variant Calling Pipelines. *bioRxiv*, 10.1101/023754.

28. Martin,M., Patterson,M., Garg,S., Fischer,S.O., Pisanti,N., Klau,G.W., Schöenhuth,A. and Marschall,T. (2016) WhatsHap: fast and accurate read-based phasing. *bioRxiv*, 10.1101/085050.

29. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

30. Magi,A., Bolognini,D., Bartalucci,N., Mingrino,A., Semeraro,R., Giovannini,L., Bonifacio,S., Parrini,D., Pelo,E., Mannelli,F., *et al.* (2019) Nano-GLADIATOR: real-time detection of copy number alterations from nanopore sequencing data. *Bioinformatics*, **35**, 4213–4221.

31. Jónsson,H., Sulem,P., Kehr,B., Kristmundsdottir,S., Zink,F., Hjartarson,E., Hardarson,M.T., Hjorleifsson,K.E., Eggertsson,H.P., Gudjonsson,S.A., *et al.* (2017) Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature 2017 549:7673*, **549**, 519–522.

32. Goldmann,J.M., Wong,W.S.W., Pinelli,M., Farrah,T., Bodian,D., Stittrich,A.B., Glusman,G., Vissers,L.E.L.M., Hoischen,A., Roach,J.C., *et al.* (2016) Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics 2016 48:8*, **48**, 935–939.

33. Sasani,T.A., Pedersen,B.S., Gao,Z., Baird,L., Przeworski,M., Jorde,L.B. and Quinlan,A.R. (2019) Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife*, **8**.

34. Veltman,J.A. and Brunner,H.G. (2012) De novo mutations in human genetic disease. *Nature Reviews Genetics 2012 13:8*, **13**, 565–575.

35. Altarescu,G., Eldar-Geva,T., Varshower,I., Brooks,B., Haran,E.Z., Margalioth,E.J., Levy-Lahad,E. and Renbaum,P. (2009) Real-time reverse linkage using polar body analysis for preimplantation genetic diagnosis in female carriers of de novo mutations. *Human Reproduction*, **24**, 3225–3229.

36. Rechitsky,S., Pomerantseva,E., Pakhalchuk,T., Pauling,D., Verlinsky,O. and Kuliev,A. (2011) First systematic experience of preimplantation genetic diagnosis for de-novo mutations. *Reprod Biomed Online*, **22**, 350–361.

37. Crugnola,E., Gobbetti,A., Fiandanese,N. and Filippini,G. (2021) P–562 PGT-M with de novo mutations : how to deal with it? *Human Reproduction*, **36**.

38. Yuan,P., Xia,J., Ou,S., Liu,P., Du,T., Zheng,L., Yin,X., Xie,L., Zhang,S., Yan,H., *et al.* (2020) A whole-genome sequencing–based novel preimplantation genetic testing method for de novo mutations combined with chromosomal balanced translocations. *J Assist Reprod Genet*, **37**, 2525–2533.

39. Dimitriadou,E., Melotte,C., Debrock,S., Esteki,M.Z., Dierickx,K., Voet,T., Devriendt,K., De Ravel,T., Legius,E., Peeraer,K., *et al.* (2017) Principles guiding embryo selection following genome-wide haplotyping of preimplantation embryos. *Human Reproduction*, **32**, 687–697.

40. Vossaert,L., Wang,Q., Salman,R., McCombs,A.K., Patel,V., Qu,C., Mancini,M.A., Edwards,D.P., Malovannaya,A., Liu,P., *et al.* (2019) Validation Studies for Single Circulating Trophoblast Genetic Testing as a Form of Noninvasive Prenatal Diagnosis. *The American Journal of Human Genetics*, **105**, 1262–1273.

41. Murphy,N.M., Samarasekera,T.S., Macaskill,L., Mullen,J. and Rombauts,L.J.F. (2020) Genome sequencing of human in vitro fertilisation embryos for pathogenic variation screening. *Scientific Reports 2020 10:1*, **10**, 1–10.

CHAPTER 6 - INTEGRATIVE DISCUSSION AND FUTURE PERSPECTIVES

This PhD study advanced existing embryo analysis techniques and developed novel methods for embryo selection. We expanded the application of a comprehensive PGT method, haplarithmisis (Zamani Esteki *et al.*, 2015), to equine embryos. Additionally, we combined GBS-based haplarithmisis (Masset *et al.*, 2022) with transcriptome profiling for the first time, representing a novel adaptation of G&T-seq (Macaulay *et al.*, 2015), referred to as hG&T-seq, which enables simultaneous haplotyping and transcriptome profiling of individual blastomeres. Furthermore, we validated the effectiveness of lrWGS-based, reference-free comprehensive PGT. Using these methods, we investigated WG abnormalities and aneuploidies in preimplantation mammalian embryos and revealed their mechanistic origins and developmental implications.

Impact of novel embryo analysis technologies on PGT and prospects for advanced embryo selection methods

The embryo analysis technologies developed in this PhD program hold significant importance for the field of PGT. Current state-of-the-art PGT methods, such as SNP array and NGS-based genome-wide comprehensive PGT (Zamani Esteki *et al.*, 2015; Masset *et al.*, 2022), enable all forms of PGT and eliminate the need for family and disease-specific setups. However, for PGT-M, these methods still require additional family members for phasing, and they cannot phase pathogenic DNMs in prospective parents or determine the carrier status of pathogenic mutations in cases of parental germline mosaicism. Since reads from long-read sequencing are significantly longer than those from short-read sequencing and can easily span two or more heterozygous SNVs (Logsdon, Vollger and Eichler, 2020), lrWGS data holds the potential to overcome these limitations by enabling direct phasing of both the parents and the embryo. In our study, we demonstrated the high performance of lrWGS for single-cell haplotyping and aneuploidy profiling, showing its promise for comprehensive PGT of human embryos. With direct and accurate phasing of the parents using lrWGS data, high-risk haplotypes carrying pathogenic mutations in prospective parents can be identified without requiring additional family members. Furthermore, when combined with lrWGS data obtained from a TE biopsy, loci that are heterozygous in both the parents and the embryo can be phased, allowing for the inclusion of more loci in haplotype-based analysis for better performance. In addition to these specific advantages over short-read or SNP array-based comprehensive PGT, the lrWGS-based method also enables direct mutation analysis, further complementing haplotype-based analysis and helping to identify whether pathogenic variants are transmitted to the embryo in

cases of parental germline mosaicism.

We used hG&T-seq, which combines G&T seq (Macaulay *et al.*, 2015) with GBS-based haplarithmisis (Masset *et al.*, 2022), to investigate the origin and developmental effects of WG abnormalities. The ability to simultaneously assess both genomic and transcriptomic information from the same single blastomere using such approach not only advances embryo research but also represents a promising avenue for better embryo selection. State-of-the-art comprehensive PGT methods rely on genomic information to select embryos free of genetic abnormalities for transfer (Zamani Esteki *et al.*, 2015; Masset *et al.*, 2022). In contrast, the G&T-seq-based approach, which enables the concurrent assessment of the transcriptome, has the potential to provide insights into gene expression and molecular homeostasis in embryos. In IVF practice, this could offer valuable information about embryo competence, thereby enhancing the selection of viable embryos. Interestingly, Groff et al. demonstrated that the gene expression profile of the TE can reflect embryo competence and highlighted the potential utility of RNA-seq for embryo evaluation in IVF clinics (Groff *et al.*, 2019). Wang et al. further showed that using the TE transcriptome for selecting viable embryos was more accurate than routine PGT (Wang *et al.*, 2024). Further exploration is needed in G&T-seq-based embryo selection, particularly to identify gene expression patterns that could serve as criteria for embryo selection at the clinical level.

In addition to invasive embryo selection methods, morphological grading remains the most widely used noninvasive approach for assessing the quality and developmental potential of preimplantation embryos in both clinical practice and research. In our study, we observed whole-genome segregation errors during the first zygotic cleavage by combining microscopy images and whole-genome haplotyping. Furthermore, the impaired development of WG abnormal blastomeres was evident from their delayed developmental stages compared to normal diploid blastomeres, as observed under the microscope. In recent years, artificial intelligence (AI) has been increasingly incorporated into imaging-based embryo ranking, marking a promising direction for better morphological assessment and selection of embryos. Unlike conventional methods, which depend on technicians and are subject to considerable variation across individuals and fertility clinics, AI-driven approaches, combined with time-lapse technology improve both the accuracy and consistency of embryo selection. Several companies are developing sophisticated software that can predict embryo euploidy with around 80% accuracy (Huang *et al.*, 2021; Liao *et al.*, 2021), and it is likely that this will approach near 100% in the coming years. Additionally, some AI-driven time-lapse incubators assign scores to embryo to identify those with the highest implantation potential. Many studies have demonstrated that this score-based selection improves not only implantation rates but also live birth rates (Bori *et al.*, 2022; Ueno *et al.*, 2022). In its present form, AI complements

and enhances embryologists' decision-making rather than replacing it (Glatstein, Chavez-Badiola and Curchoe, 2023). The ultimate aim is to create a robust AI system capable of accurately predicting the ploidy status and implantation potential of embryos in culture, while being adaptable enough to incorporate omics data for further optimization.

In addition to WG abnormalities, we observed frequent aneuploidies in equine preimplantation embryos, similar to those observed in humans and bovine (Vanneste *et al.*, 2009, 2012; Tšuiko *et al.*, 2017). Founded on the premise that most aneuploid embryos are either non-viable or likely to result in severe birth defects, PGT-A has seen widespread implementation. However, its effectiveness in improving IVF outcomes remains a subject of debate. Two primary concerns challenge its routine use. First, mosaicism is frequently found in preimplantation embryos, typically originated from post-zygotic mitotic errors (Vanneste *et al.*, 2009, 2012; Levy *et al.*, 2021). Gleicher et al. indicated through mathematical models that, even when assuming an even distribution of mosaicism in the trophectoderm, a trophectoderm biopsy cannot reliably classify an embryo as aneuploid or euploid (Gleicher *et al.*, 2017). Second, embryos have a natural ability to self-correct aneuploidies (Yang *et al.*, 2021), meaning initial biopsy results may not reflect the embryo's final genetic status. This is evidenced by cases where healthy babies were born from embryos initially deemed aneuploid or mosaic (Greco, Minasi and Fiorentino, 2015). In addition to these concerns, the clinical effects of PGT-A have been extensively explored. Randomized controlled trials (RCTs) have shown that PGT-A leads to increased ongoing pregnancy rates and live birth rates per embryo transferred (Scott *et al.*, 2013; Rubio *et al.*, 2017; Rubio, 2019). However, some studies found these benefits only in women aged 35–40 years, with no improvements seen in younger women (Munné *et al.*, 2019; Ozgur *et al.*, 2019). Furthermore, Yan et al. demonstrated that PGT-A does not improve cumulative live birth rates (Yan *et al.*, 2021). Wang et al. observed that trials failing to demonstrate positive outcomes with PGT-A might have neglected critical factors such as individual clinic performance and the impact of biopsy techniques, which can influence PGT-A results (Wang *et al.*, 2022). Given the ongoing debate, PGT-A may not offer universal benefits for all IVF patients. Instead, each clinic is advised to evaluate the efficacy of PGT-A based on its specific patient population, expertise, and limitations. Couples should be informed about the potential benefits and risks of PGT-A, customized to their individual clinical context and needs.

Significance and biological consequences of heterogoneic division

By utilizing hG&T-seq, we demonstrated that heterogoneic first zygotic division, during which maternal and paternal genomes segregate into distinct blastomeres, can occur in multiple ways, leading to various WG abnormalities among the daughter blastomeres. This heterogoneic division offers a more comprehensive explanation for the origin of WG anomalies compared to previously proposed mechanisms, which were inferred indirectly from surviving

cells that had already undergone developmental selection and, therefore, did not fully reflect the original genomic heterogeneity of the embryo (Masset, Tšuiko and Vermeesch, 2021). Further development of a subset of blastomeres following heterogoneic division could result in triploid, mixoploid, chimeric, uniparental diploid, or biparental diploid embryos, depending on the selective processes acting on these blastomeres. Typically, diploid blastomeres exhibit greater fitness compared to those with WG abnormalities, enabling them to outgrow the abnormal blastomeres. However, some uniparental and polyploid blastomeres still retain the ability to develop into blastocyst. In the present study, we used bovine as the model organism. Previous research has identified frequent occurrences of various WG abnormalities in human preimplantation embryos, suggesting a conserved presence of heterogoneic division in human (Tšuiko *et al.*, 2021). Replicating our hG&T-seq-based study on heterogoneic division in humans and other mammalian species would allow for a more detailed examination of the frequency of heterogoneic division, as well as the developmental consequences of the resulting WG abnormal blastomeres in human and across different species. Such studies would significantly enhance our understanding of the conservation of heterogoneic division and its impact on mammalian embryo development. Furthermore, additional studies in model organisms could investigate the developmental effects of heterogoneic division-derived WG abnormalities after implantation. When integrated with other omics data, these analyses could provide a comprehensive understanding of the clinical consequences of these abnormalities and yield detailed insights into the underlying molecular mechanisms, such as the role of imprinting. In this study, we unveiled that WG abnormalities cause differential expression of both imprinted genes and stress response-related genes during preimplantation development. Further investigation into the role and effects of imprinting in the development of heterogoneic division-derived uniparental and polyploid blastomeres after implantation would provide valuable insights into the specific function of parental genomes.

Conserved chromosomal instability in mammalian preimplantation embryos

It has been well studied in both humans and bovines that chromosomal abnormalities are common in preimplantation embryos (Vanneste *et al.*, 2009, 2012; Tšuiko *et al.*, 2017) and are selected against during development, particularly toward complex abnormalities (McCoy *et al.*, 2015, 2023; Tšuiko *et al.*, 2021). By expanding the application of a comprehensive PGT approach, haplarithmisis, in equine and exploring the full range of chromosomal abnormalities, we observed a conserved and frequent occurrence of both aneuploidies and WG abnormalities in preimplantation equine embryos. Additionally, we confirmed the selective elimination of complex abnormalities during preimplantation development. These results demonstrate the conservation of CIN in preimplantation embryos across mammalian species. This PGT approach can in principle be applied to other mammals to investigate early embryo genome

abnormalities, allowing for comparisons of their frequency, spectrum, and developmental effects across species. Comparing CIN in *in vitro* and *in vivo*-produced embryos could offer insights into how IVF procedures influence CIN levels. While a comparison of *in vivo* and *in vitro* human embryos showed similar euploidy rates (Munné *et al.*, 2020), research in bovine models demonstrated significantly higher chromosomal instability in *in vitro*-cultured embryos compared to those derived *in vivo* (Tšuiko *et al.*, 2017), highlighting species differences.

Clinical impact and potential further applications of the results

The results of this PhD program have significant practical and clinical implications. By performing hG&T-seq, we demonstrated that various WG-abnormal blastomeres and normal diploid blastomeres can coexist in the same embryo following heterogoneic first zygotic division. The WG-abnormal blastomeres exhibit impaired development compared to the normal diploid ones, suggesting that the latter are likely to dominate the blastocyst. However, WG-abnormal blastomeres can still develop into blastocysts, resulting in mixoploid or chimeric embryos, which may lead to abnormal development or congenital disorders. Given that embryos can develop into normal diploid, WG-abnormal, or mixoploid/chimeric blastocysts, implementing haplotype-aware PGT approaches (Zamani Esteki *et al.*, 2015; Masset *et al.*, 2022) is recommended to screen for WG abnormalities and improve IVF outcomes. Furthermore, we demonstrated the applicability of haplarithmisis (Zamani Esteki *et al.*, 2015) in equine to select biparental diploid embryos that are free from mutations associated with undesired traits while carrying mutations for desirable characteristics. This approach not only enhances equine breeding but also has broader potential applications in MAR for various other mammalian species, aiding livestock breeders in producing high-quality embryos and improving reproductive outcomes.

As sequencing technologies improve and costs decrease, the adoption of WGS-based PGT is expected to increase, bringing important ethical concerns to the forefront. WGS-based PGT not only identifies variants associated with specific disorders but also reveals the entire genetic makeup, including variants related to complex disorders, mild effects, and non-health-related traits. This additional information raises concerns about using genetic data to select embryos that carry variants for desired traits, often referred to as "designer babies." It also sparks discussions about whether prospective parents have the right to choose the "best" embryo and how "best" should be defined in the context of genome-wide analysis. Addressing these challenges entails increased complexity in genetic counseling, data interpretation, and embryo transfer policies. Further research and consensus guidelines are essential to help clinicians and families navigate the intricate information available from WGS-based PGT, ensuring a balance between technological innovation and ethical considerations for the optimal use of PGT.

In summary, this PhD study has significantly advanced current embryo study and selection methods and enhanced our understanding of the origins and developmental impacts of chromosomal abnormalities in preimplantation mammalian embryos. Combined with imaging-based methods, we envision a more accurate system for selecting embryos with the highest developmental potential and free from genetic abnormalities, which will substantially improve reproductive outcomes. The accompanying ethical considerations will need to be addressed as WGS-based embryo selection becomes more widely adopted.

References

Bori, L. *et al.* (2022) 'The higher the score, the better the clinical outcome: retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories', *Human reproduction (Oxford, England)*, 37(6), pp. 1148–1160. Available at: https://doi.org/10.1093/HUMREP/DEAC066.

Glatstein, I., Chavez-Badiola, A. and Curchoe, C.L. (2023) 'New frontiers in embryo selection', *Journal of Assisted Reproduction and Genetics*, 40(2), pp. 223–234. Available at: https://doi.org/10.1007/S10815-022-02708-5/FIGURES/4.

Gleicher, N. *et al.* (2017) 'A single trophectoderm biopsy at blastocyst stage is mathematically unable to determine embryo ploidy accurately enough for clinical use', *Reproductive Biology and Endocrinology*, 15(1), pp. 1–8. Available at: https://doi.org/10.1186/S12958-017-0251-8/FIGURES/3.

Greco, E., Minasi, M.G. and Fiorentino, F. (2015) 'Healthy Babies after Intrauterine Transfer of Mosaic Aneuploid Blastocysts', *New England Journal of Medicine*, 373(21), pp. 2089–2090. Available at: https://doi.org/10.1056/NEJMC1500421/SUPPL_FILE/NEJMC1500421_DISCLOSURES.PDF.

Groff, A.F. *et al.* (2019) 'RNA-seq as a tool for evaluating human embryo competence', *Genome Research*, 29(10), pp. 1705–1718. Available at: https://doi.org/10.1101/GR.252981.119.

Huang, B. *et al.* (2021) 'An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data', *Reproductive Biology and Endocrinology : RB&E*, 19(1), p. 185. Available at: https://doi.org/10.1186/S12958-021-00864-4.

Levy, B. *et al.* (2021) 'Chromosomal mosaicism: Origins and clinical implications in preimplantation and prenatal diagnosis', *Prenatal Diagnosis*, 41(5), pp. 631–641. Available at: https://doi.org/10.1002/PD.5931.

Liao, Q. *et al.* (2021) 'Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring', *Communications Biology 2021 4:1*, 4(1), pp. 1–9. Available at: https://doi.org/10.1038/s42003-021-01937-1.

Logsdon, G.A., Vollger, M.R. and Eichler, E.E. (2020) 'Long-read human genome sequencing and its applications', *Nature Reviews Genetics 2020 21:10*, 21(10), pp. 597–614. Available at: https://doi.org/10.1038/s41576-020-0236-x.

Macaulay, I.C. *et al.* (2015) 'G&T-seq: parallel sequencing of single-cell genomes and transcriptomes', *Nature methods*, 12(6), pp. 519–522. Available at: https://doi.org/10.1038/NMETH.3370.

Masset, H. *et al.* (2022) 'Single-cell genome-wide concurrent haplotyping and copy-number profiling through genotyping-by-sequencing', *Nucleic Acids Research*, 50(11), pp. e63–e63. Available at: https://doi.org/10.1093/NAR/GKAC134.

Masset, H., Tšuiko, O. and Vermeesch, J.R. (2021) 'Genome-wide abnormalities in embryos: Origins and clinical consequences', *Prenatal diagnosis*, 41(5), pp. 554–563. Available at: https://doi.org/10.1002/PD.5895.

McCoy, R.C. *et al.* (2015) 'Evidence of Selection against Complex Mitotic-Origin Aneuploidy during Preimplantation Development', *PLOS Genetics*, 11(10), p. e1005601. Available at: https://doi.org/10.1371/JOURNAL.PGEN.1005601.

McCoy, R.C. *et al.* (2023) 'Meiotic and mitotic aneuploidies drive arrest of in vitro fertilized human preimplantation embryos', *Genome Medicine*, 15(1), pp. 1–15. Available at: https://doi.org/10.1186/S13073-023-01231-1/FIGURES/5.

Munné, S. *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', *Fertility and Sterility*, 112(6), pp. 1071-1079.e7. Available at: https://doi.org/10.1016/J.FERTNSTERT.2019.07.1346.

Munné, S. et al. (2020) 'First PGT-A using human in vivo blastocysts recovered by uterine lavage: comparison with matched IVF embryo controls', *Human Reproduction*, 35(1), pp. 70–80. Available at: https://doi.org/10.1093/HUMREP/DEZ242.

Ozgur, K. *et al.* (2019) 'Single best euploid versus single best unknown-ploidy blastocyst frozen embryo transfers: a randomized controlled trial', *Journal of Assisted Reproduction and Genetics*, 36(4), pp. 629–

636. Available at: https://doi.org/10.1007/S10815-018-01399-1/TABLES/2.

Rubio, C. *et al.* (2017) 'In vitro fertilization with preimplantation genetic diagnosis for aneuploidies in advanced maternal age: a randomized, controlled study', *Fertility and Sterility*, 107(5), pp. 1122–1129. Available at: https://doi.org/10.1016/J.FERTNSTERT.2017.03.011.

Rubio, C. (2019) 'PGT-A and RCT proof in AMA and SMF couples', *Reproductive BioMedicine Online*, 38, p. e8. Available at: https://doi.org/10.1016/J.RBMO.2019.03.016.

Scott, R.T. *et al.* (2013) 'Blastocyst biopsy with comprehensive chromosome screening and fresh embryo transfer significantly increases in vitro fertilization implantation and delivery rates: a randomized controlled trial', *Fertility and Sterility*, 100(3), pp. 697–703. Available at: https://doi.org/10.1016/J.FERTNSTERT.2013.04.035.

Tšuiko, O. *et al.* (2017) 'Genome stability of bovine in vivo-conceived cleavage-stage embryos is higher compared to in vitro-produced embryos', *Human Reproduction*, 32(11), pp. 2348–2357. Available at: https://doi.org/10.1093/HUMREP/DEX286.

Tšuiko, O. *et al.* (2021) 'Haplotyping-based preimplantation genetic testing reveals parent-of-origin specific mechanisms of aneuploidy formation', *NPJ genomic medicine*, 6(1). Available at: https://doi.org/10.1038/S41525-021-00246-0.

Ueno, S. *et al.* (2022) 'Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: a single-centre, large-cohort retrospective study', *Journal of assisted reproduction and genetics*, 39(9), pp. 2089–2099. Available at: https://doi.org/10.1007/S10815-022-02562-5.

Vanneste, E. *et al.* (2009) 'Chromosome instability is common in human cleavage-stage embryos', *Nature Medicine 2009 15:5*, 15(5), pp. 577–583. Available at: https://doi.org/10.1038/nm.1924.

Vanneste, E. *et al.* (2012) 'Aneuploidy and copy number variation in early human development', *Seminars in reproductive medicine*, 30(4), pp. 302–308. Available at: https://doi.org/10.1055/S-0032-1313909.

Wang, L. *et al.* (2022) 'PGT-A: The biology and hidden failures of randomized control trials', *Prenatal Diagnosis*, 42(9), pp. 1211–1221. Available at: https://doi.org/10.1002/PD.6199.

Wang, Yuqian *et al.* (2024) 'Concurrent Preimplantation Genetic Testing and Competence Assessment of Human Embryos by Transcriptome Sequencing', *Advanced Science*, 11(32), p. 2309817. Available at: https://doi.org/10.1002/ADVS.202309817.

Yan, J. *et al.* (2021) 'Live Birth with or without Preimplantation Genetic Testing for Aneuploidy', *New England Journal of Medicine*, 385(22), pp. 2047–2058. Available at: https://doi.org/10.1056/NEJMOA2103613/SUPPL_FILE/NEJMOA2103613_DATA-SHARING.PDF.

Yang, M. *et al.* (2021) 'Depletion of aneuploid cells in human embryos and gastruloids', *Nature Cell Biology 2021 23:4*, 23(4), pp. 314–321. Available at: https://doi.org/10.1038/s41556-021-00660-7.

Zamani Esteki, M. *et al.* (2015) 'Concurrent whole-genome haplotyping and copy-number profiling of single cells', *American journal of human genetics*, 96(6), pp. 894–912. Available at: https://doi.org/10.1016/J.AJHG.2015.04.011.

CURRICULUM VITAE

Personal Details

Name: Yan Zhao

Email: yan.zhao1@kuleuven.be

Education

2020-Present: PhD in Biomedical Sciences, KU Leuven, Belgium (Joint PhD student, co-host university: University of Tartu, Estonia)

2017-2019: Master in Public Health, Xi'an Jiaotong University, China

2012-2017: Bachelor of Clinical Medicine (Preventive Medicine), Xi'an Jiaotong University, China

Conference Presentations

- MIMICRY Symposium in Leuven, Belgium on June 7th, 2024

  Title: "Non-canonical zygotic genome segregations as a cause of embryonic lethality and birth defects" (Oral Presentation)

- ESHG Conference in Berlin, Germany from June 1 to 4, 2024

  Title: "The potential of long-read whole genome sequencing based preimplantation genetic testing" (Oral Presentation)

- Karl Ernst von Baer's Seminar in Tartu, Estonia from October 19 to 21, 2023

  Title: "Unraveling the mechanisms of chromosomal instability during early embryogenesis" (Oral Presentation)

- PGDIS Conference in Paris, France from April 17 to 19, 2023

  Title: "Evaluating the potential of long-read whole-genome analysis for preimplantation genetic testing" (Poster Presentation)

Practical Skills

- Shell scripting
- R programming
- Single-cell G&T seq data analysis
- Analysis of short-read and long-read whole-genome sequencing data from single-cell and

bulk samples, including detailed exploration of SNVs, indels, CNVs, and haplotype phasing

Teaching Activities

- Assisted in practical training for the biotechnology part of the Introductory Laboratory Course at University of Tartu (2023)
- Engaged in writing assignments correction at KU Leuven (2024)

List of Publications

- **Zhao, Y. ***, Fernández-Montoro, A. *****, Peeters, G., Jatsenko, T., Coster, T. De, Angel-Velez, D., Lefevre, T., Voet, T., Tšuiko, O., Kurg, A., Smits, K., Soom, A. Van, & Vermeesch, J. R. (2024). Origin and development of uniparental and polyploid blastomeres. *BioRxiv*, 2024.07.30.605883. https://doi.org/10.1101/2024.07.30.605883 Manuscript peer review completed and accepted by *iScience*. (*Co-first authors)
- **Zhao, Y.**, Tsuiko, O., Jatsenko, T., Peeters, G., Souche, E., Geysens, M., Dimitriadou, E., Vanhie, A., Peeraer, K., Debrock, S., Esch, H. Van, & Vermeesch, J. R. (2024). Long-read whole-genome sequencing-based concurrent haplotyping and aneuploidy profiling of single cells. *BioRxiv*, 2024.09.24.614469. https://doi.org/10.1101/2024.09.24.614469 Manuscript under revision in *Nucleic Acid Research*.
- De Coster, T. *, **Zhao, Y. ***, Tšuiko, O., Demyda-Peyrás, S., Van Soom, A., Vermeesch, J. R., Smits, K., & Coster, D. (2024). Genome-wide equine preimplantation genetic testing enabled by simultaneous haplotyping and copy number detection. *Scientific Report*s 2024 14:1, 14(1), 1–12. https://doi.org/10.1038/s41598-023-48103-7 (*Co-first authors)
- van Riel, M., **Zhao, Y.**, Jatsenko, T., Lannoo, L., Timmerman, D., & Vermeesch, J. R. (2023). Ultrasensitive Microhaplotype-Based Detection of Minor Alleles in Cervical Samples of Pregnant Women Identifies Sperm Rather Than Fetal DNA. *Clinical Chemistry*, 69(9), 1089–1091. https://doi.org/10.1093/CLINCHEM/HVAD107
- De Coster, T., Masset, H., Tšuiko, O., Catteeuw, M., **Zhao, Y**., Dierckxsens, N., Aparicio, A. L., Dimitriadou, E., Debrock, S., Peeraer, K., de Ruijter-Villani, M., Smits, K., Van Soom, A., & Vermeesch, J. R. (2022). Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. *Genome Biology* 2022 23:1, 23(1), 1–29. https://doi.org/10.1186/S13059-022-02763-2
- **Zhao, Y.** *, Ning, Y. *, Zhang, F., Ding, M., Wen, Y., Shi, L., Wang, K., Lu, M., Sun, J., Wu, M., Cheng, B., Ma, M., Zhang, L., Cheng, S., Shen, H., Tian, Q., Guo, X., & Deng, H. W.

(2019). PCA-based GRS analysis enhances the effectiveness for genetic correlation detection. *Briefings in Bioinformatics*, 20(6), 2291–2298. https://doi.org/10.1093/BIB/BBY075 (*Co-first authors)

⑩ **Zhao, Y.** *, He, A. *, Zhu, F., Ding, M., Hao, J., Fan, Q., Li, P., Liu, L., Du, Y., Liang, X., Guo, X., Zhang, F., & Ma, X. (2018). Integrating genome-wide association study and expression quantitative trait locus study identifies multiple genes and gene sets associated with schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 81, 50–54. https://doi.org/10.1016/J.PNPBP.2017.10.003 (*Co-first authors)

⑩ **Zhao, Y.** *, Liang, X. *, Zhu, F., Wen, Y., Xu, J., Yang, J., Ding, M., Cheng, B., Ma, M., Zhang, L., Cheng, S., Wu, C., Wang, S., Wang, X., Ning, Y., Guo, X., & Zhang, F. (2018). A large-scale integrative analysis of GWAS and common meQTLs across whole life course identifies genes, pathways and tissue/cell types for three major psychiatric disorders. *Neuroscience & Biobehavioral Reviews*, 95, 347–352. https://doi.org/10.1016/J.NEUBIOREV.2018.10.005 (*Co-first authors)

UNRAVELLING THE MECHANISMS OF CHROMOSOMAL INSTABILITY DURING EARLY EMBRYOGENESIS