REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

PREIMPLANTATION GENETIC DIAGNOSIS IN BALANCED REARRANGEMENT CARRIERS AND INVESTIGATION OF INTER CHROMOSOMAL EFFECT

Çağrı OĞUR

DOCTOR OF PHILOSOPHY THESIS
Department of Bioengineering
Bioengineering Program

Supervisor
Assoc.Prof.Dr. Alper YILMAZ

Co-supervisor
Prof. Dr. Şehime Gülsün TEMEL

March, 2023

REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

PREIMPLANTATION GENETIC DIAGNOSIS IN BALANCED REARRANGEMENT CARRIERS AND INVESTIGATION OF INTER CHROMOSOMAL EFFECT

A thesis submitted by Çağrı OĞUR in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY** is approved by the committee on 16 March 2023 in Department of Bioengineering, Bioengineering Program.

Assoc.Prof.Dr. Alper YILMAZ
Yildiz Technical University
Supervisor

Prof. Dr. Şehime Gülsün TEMEL
Uludag University
Co-supervisor

Approved By the Examining Committee

Assoc.Prof.Dr. Alper YILMAZ, Supervisor
Yildiz Technical University _____

Prof.Dr. Dilek BALIK, Member
Yildiz Technical University _____

Prof.Dr. Tülay İREZ, Member
Yeni Yuzyil University _____

Prof.Dr. Nehir ÖZDEMİR ÖZGENTÜRK, Member
Yildiz Technical University _____

Prof.Dr. Ahmet İlter GÜNEY, Member
Marmara University _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of PREIMPLANTATION GENETIC DIAGNOSIS IN BALANCED REARRANGEMENT CARRIERS AND INVESTIGATION OF INTER CHROMOSOMAL EFFECT supervised by my supervisor, Assoc.Prof.Dr. Alper YILMAZ. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Çağrı OĞUR

Signature

Dedicated to my daugther Talya

ACKNOWLEDGEMENTS

First, I would like to thank to Prof. Dr. Semra Kahraman, MD for her support in my academical career. Without her confidence in me and permission to use the invaluable data which is the result of more than ten years of effort of her team full of dedicated health care specialists, this study could not be performed. I am honored to have analyzed the data of one of the most successful and a reference clinic for IVF and genetic diagnosis in the world.

I would like to thank to my advisor Dr. Alper Yılmaz, PhD for his continuous support and belief in me during these years. He was always there to help me facilitate the work in every occasion with his data analysis skills and magical commands in R environment.

I would like to thank to Prof Dr Sehime Gülsün Temel, MD, PhD for her accelerating force upon me from the beginning of this project until the submission of the thesis. Without her this study would not have been finished.

I would very much like to thank to Prof. Dr. Darren Karl Griffin, PhD, DSs for his generous mentoring, teaching, writing skills and being an academic role model for me.

I would like to thank to Prof. Dr. Tülay İrez, PhD and Prof. Dr. Dilek Balık, PhD for their guidance throughout this journey with beneficial meetings and monitoring of my thesis research.

I would like to thank to Prof. Dr. İlker Ercan for teaching me the advanced concepts of statistics and helping me get the most out of this data.

And last but not the least, I would like to thank to my family Talya Beyazyürek and Nursel İnal for their continuous support, belief and their patience.

Çağrı OĞUR

TABLE OF CONTENTS

LIST OF SYMBOLS

A	Adenine
°C	Celcius
Cy3	Cyanine-3 which is a fluorescent compound with an excitation peak at 555 nm (green)
Cy5	Cyanine-5 which is a fluorescent compound with an excitation peak at 683 nm (red)
C	Cytosine
δ	delta
der	Derivative
g	Gravitational force
G	Guanine
h+	Heterochromatin region
ins	Insertion
inv	Inversion
Mb	Mega Base
μl	Microliter
ml	Mililiter
nm	Nanometer
p	p arm (short arm) of the chromosome
ps	p terminal satellite
pter	p terminal of the chromosome
ϕ	phi
pH	Potential of hydrogen
q	q arm (long arm) of the chromosome

qter	q terminal of the chromosome
r	Ring
T	Thymine
t	Translocation
X	X-chromosome
Y	Y-chromosome

LIST OF ABBREVIATIONS

AMA	Advanced Maternal Age
AMH	Anti-Mullerian hormone
AUC	Area Under the Curve
aCGH	Array Comperative Genomic Hybridization
ANN	Artificial Neural Network
ART	Assisted Reproduction Technologies
BAC	Bacterial Artificial Chromosome
BMI	Body Mass Index
CS	Centric Segment
CPR	Clinical Pregnancy Rate
CGH	Comperative Genomic Hybridization
CCR	Complex Chromosomal Rearrangements
COMP	Complex Translocation
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
DS	Dextran sulphate
DSB	Double Stranded Breaks
ET	Embryo Transfer
ESHRE	European Society of Human Reproduction and Embryology
FISH	Fluorescence in situ Hybridization
FSH	Follicle Stimulating Hormone
GnRH	Gonadotrophin Releasing Hormone

HS	High Sensitivity
hCG	Human Chorionic Gonadotrophin
hMG	Human Menopausal Gonadotropins
HEPES	Hydroxyethylpiperazine ethane sulfonic acid
IR	Implantation Rate
IVF	*in vitro* Fertilisation
ICSI	Intracytoplasmic Sperm Injection
ISCN	International System for Human Cytogenomic Nomenclature
INV	Inversion
KNN	K-Nearest Neighbor
KAE	Kromozomlar Arası Etki
LIMS	Laboratory Information Management Systems
LBR	Live Birth Rate
LH	Luteinizing Hormone
MALBAC	Multiple Annealing and Looping Based Amplification Cycles
NGS	Next Generation Sequencing
NAHR	Non Allelic Homologous Recombination
NA	Not available
OR	Odds Ratio
OPU	Ovum Pick-Up
PAI	Paracentric Inversion
PEI	Pericentric Inversion
PBS	Phosphate Buffer Saline
PMT	Photo Multiplier Tube
PB	Polar Body
PCR	Polymerase Chain Reaction
PGD	Preimplantation Genetic Diagnosis
PGT	Preimplantation Genetic Testing
PGT-A	Preimplantation Genetic Testing for Aneuploidy

PGT-M	Preimplantation Genetic Testing for Monogenic Disorders
PGT-SR	Preimplantation Genetic Testing for Structural Rearrangements
RF	Random Forest
ROC	Receiver Operating Characteristic
RECT	Reciprocal Translocation
RPM	Revolutions Per Minute
ROBT	Robertsonian Translocation
SNP	Single Nucleotide Polymorphisms
SPSS	Statistical Package for the Social Sciences
SR	Structural Rearrangement
THBR	Take Home Baby Rate
TS	Translocated Segment
TE	Trophectoderm
WGA	Whole Genome Amplification
YKA	Yapısal Kromozom Anomalisi
ZP	Zona Pellucida

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

PREIMPLANTATION GENETIC DIAGNOSIS IN BALANCED REARRANGEMENT CARRIERS AND INVESTIGATION OF INTER CHROMOSOMAL EFFECT

Çağrı OĞUR

Department of Bioengineering
Doctor of Philosophy Thesis

Supervisor: Assoc.Prof.Dr. Alper YILMAZ
Co-supervisor: Prof. Dr. Şehime Gülsün TEMEL

Balanced structural chromosome rearrangement (SR) carriers are poor prognosis patients in IVF due to the production of a high proportion of unbalanced gametes and a correspondingly low number of transferable embryos. Genetic and clinical outcomes are highly variable and not easy to predict due to a high number of factors. There is an urgent need to identify these factors in order to offer these patients a more personalized treatment and improve their chances of having a baby. Since 1963, a hypothesis entitled the "inter-chromosomal effect" (ICE) claimed that SRs induce further chromosomal abnormalities among non-rearranged chromosomes in gametes/embryos. This hypothesis is mostly based on observation, lacks statistical evidence and needs to be tested with analytical tools. The purpose of this study was therefore to investigate factors that affect the proportion of chromosomally balanced embryos available for transfer and test the "ICE" hypothesis. The end goal was to establish predictive models using machine learning algorithms to improve personalized treatments for SR carriers seeking preimplantation genetic testing (PGT-SR). Data analysis comprised genetic and clinical outcomes from 300 couples referred to Şişli Memorial Hospital in vitro Fertilization (IVF) unit between 2011-2019. A well-matched control group was selected from the same database. 1835 embryos and 117,033 chromosome pairs were analyzed by array comperative genomic hybridization (aCGH) and next generation sequencing (NGS). Statistical analysis was performed using SPSS and R-software. Rearrangement type, maternal

age and sex of the carrier were found to have significant impacts on the proportion of transferable embryos. Results did not support any evidence for an ICE and the hypothesis was therefore rejected. This study helped to provide a predictive model with the use of advanced statistical and machine learning tools to reveal parameters to provide personalized treatment and better genetic counselling for PGT-SR couples in the future.

Keywords: Inter chromosomal effect, machine learning, next generation sequencing, preimplantation genetic testing for structural rearrangements, prediction modelling via machine learning

ÖZET

DENGELİ YAPISAL KROMOZOM ANOMALİSİ TAŞIYICILARINDA PREİMPLANTASYON GENETİK TANI VE KROMOZOMLAR ARASI ETKİLERİN İNCELENMESİ

Çağrı OĞUR

Biyomühendislik Bölümü
Doktora Tezi

Danışman: Doç. Dr. Alper YILMAZ
Eş-Danışman: Prof. Dr. Şehime Gülsün TEMEL

Dengeli yapısal kromozom anomalisi (YKA) taşıyıcıları, yüksek oranda dengesiz gamet üretimi ve buna bağlı olarak transfer edilebilir embriyo sayısının düşük olması nedeniyle IVF tedavilerine başvuran hastaların içerisinde en kötü prognozlu hasta gruplarından birini oluşturmaktadır. Taşıyıcılara ait genetik ve klinik birçok özelliğin tedavi sonuçları üzerindeki etkileri ne kadar ortaya konulabilirse, tedavi verimliliğinin arttırılması da o derece mümkün olacaktır. Bu faktörlerin ortaya konması bu hastalara daha kişiselleştirilmiş bir tedavi sunulmasına ve bebek sahibi olma şanslarının artırılmasını sağlamak için gereklidir. Bu gibi konuların yanında ayrıca, 1963'te, "kromozomlar arası etki" (KAE) adlı bir hipotez ile, YKA taşıyıcılarının embriyolarında YKA dışındaki diğer kromozomların da segregasyonlarının etkileneceği iddia edilmiştir. Bu hipotez çoğunlukla gözleme dayalıdır, istatistiksel kanıtlardan yoksundur ve analitik araçlarla test edilmesi gerekmektedir. Bu çalışmanın amacı, tek bir merkezden bugüne kadarki en geniş veri setini kullanarak transfere uygun, normal/dengeli embriyoların oranını etkileyen faktörleri araştırmak ve "KAE" hipotezini test etmektir. Nihai hedef, preimplantasyon genetik testi (PGT) için başvuran hastalar için kişiselleştirilmiş tedavilerin geliştirilmesi ve makine öğrenmesi algoritmaları kullanarak tahmine dayalı modeller oluşturmaktır. Veri analizi, 2011-2019 yılları arasında Şişli Memorial Hastanesi Tüp Bebek ünitesine başvuran 300 çiftin genetik ve klinik sonuçlarını içermektedir. Kontrol grubu, birçok klinik özellik açısından bire-bir eşleştirilen

vakalardan seçilmiştir. aCGH ve NGS yöntemleriyle analiz edilen 1835 embriyo ve 117.033 kromozom çifti üzerinde SPSS ve R-yazılımı kullanılarak istatistiksel analiz yapılmıştır. Taşıyıcıların ait olduğu grubunun, kadın yaşı ve taşıyıcının cinsiyetinin transfer edilebilir embriyo oranı üzerinde önemli etkileri olduğu tespit edildi. KAE'ye dair bir kanıt bulunmadı ve bu nedenle hipotez (en azından) bu veriseti için reddedildi. Bu çalışma, gelecekte YKA hastaları için kişiselleştirilmiş tedavi ve daha iyi genetik danışmanlık sağlayabilmek için parametrelerin anlaşılmasına ve gelişmiş istatistiksel yöntemler ve makine öğrenimi/yapay zeka araçlarının kullanımıyla tahmine dayalı bir model geliştirilmesine yardımcı olmuştur.

Anahtar Kelimeler: Kromozomlar arası etki, makine öğrenmesi, yapısal kromozom anomalileri için preimplantasyon genetik test, yeni nesil dizileme, makine öğrenmesi yöntemiyle tahmin modellemesi

1
INTRODUCTION

1.1 Literature Review

1.1.1 Balanced structural rearrangements

Balanced SRs have no overall loss or gain and mainly include 4 types; reciprocal translocations, Robertsonian translocations, inversions and complex chromosomal rearrangements (Figure 1.1).

Balanced SR carriers are often not phenotypically affected, provided that the rearrangement does not disrupt any key functional genes. However, they have increased reproductive risk due to production of chromosomally unbalanced gametes (either oocytes or sperms depending on the sex of the carrier) due to abnormal segregations during gametogenesis. For this reason, the carriers often suffer from infertility, miscarriage, stillbirth, and less often, birth of a child with mental and/or physical disabilities. The estimated incidence of SRs is in the range of 0,1-1,2 % in normal population [1–3] whereas the incidence is higher (5.7%) among couples with recurrent miscarriages [4]. The severity of the reproductive risk of a carrier couple is affected mostly by the type of the rearrangement and female age [5].

1.1.1.1 Reciprocal Translocations

Reciprocal translocations (RECT) involve the exchange of segments from two nonhomologous chromosomes which give rise to the formation of two derivative chromosomes. During gametogenesis, in pachytene stage of meiosis I, this results in the formation of a quadrivalent structure which holds both the two derivative chromosomes and the original chromosomes which are partly homologous to each other (Figure 1.2). During anaphase I, two derivatives and two normal homologs segregate to the poles with alternate, adjacent-1, adjacent-2, 3:1 and 4:0 segregation types (Figure 1.3), (Table 1.1).

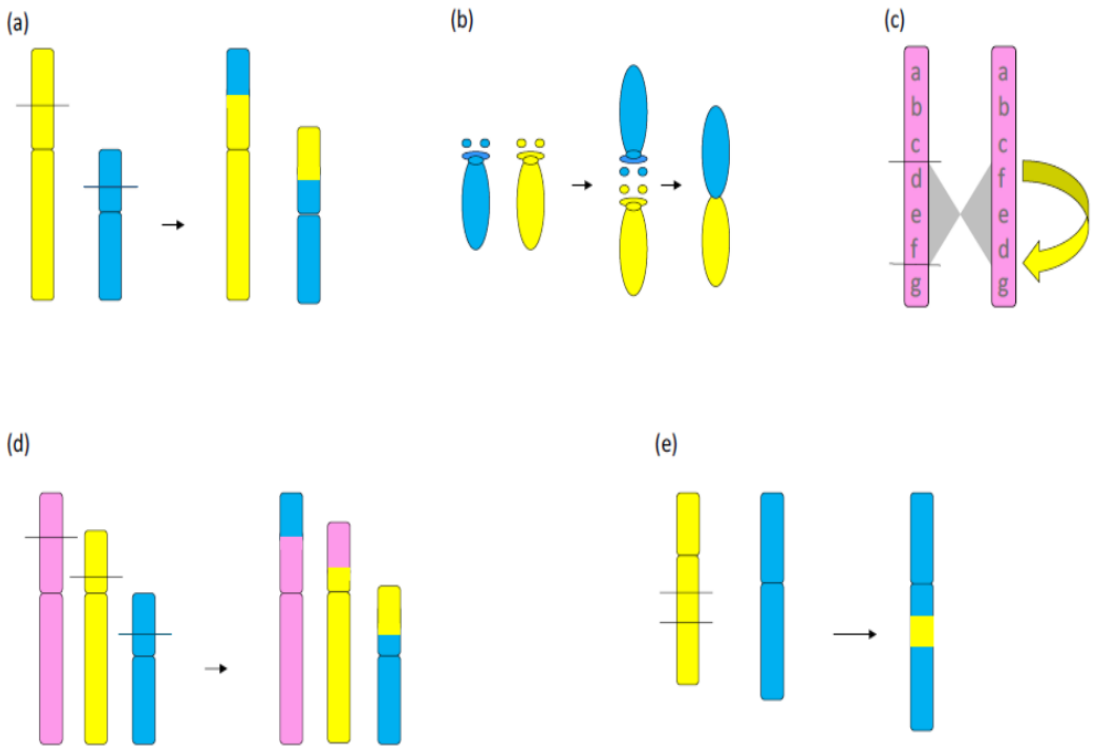


Figure 1.1 The common types of balanced rearrangements. a) Balanced reciprocal translocation between two non-homologous chromosomes b) Robertsonian translocation between acrocentric chromosomes c) inversion d) a complex chromosomal rearrangement involving three different chromosomes (e) insertional translocation by the insertion of a segment from a chromosome to a different one (with permission from Ogur and Griffin [5])

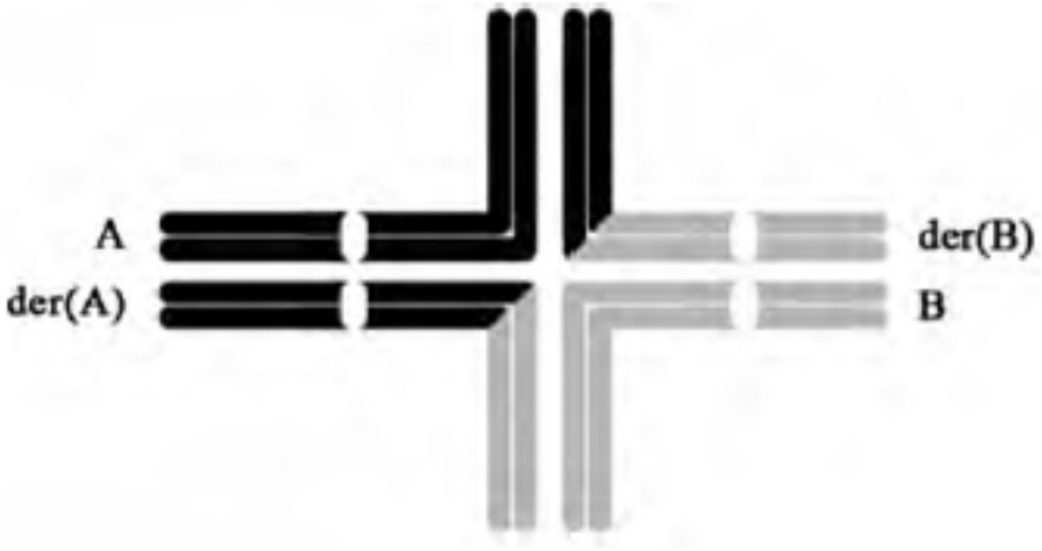


Figure 1.2 Quadrivalent structure formed in RECT carrier

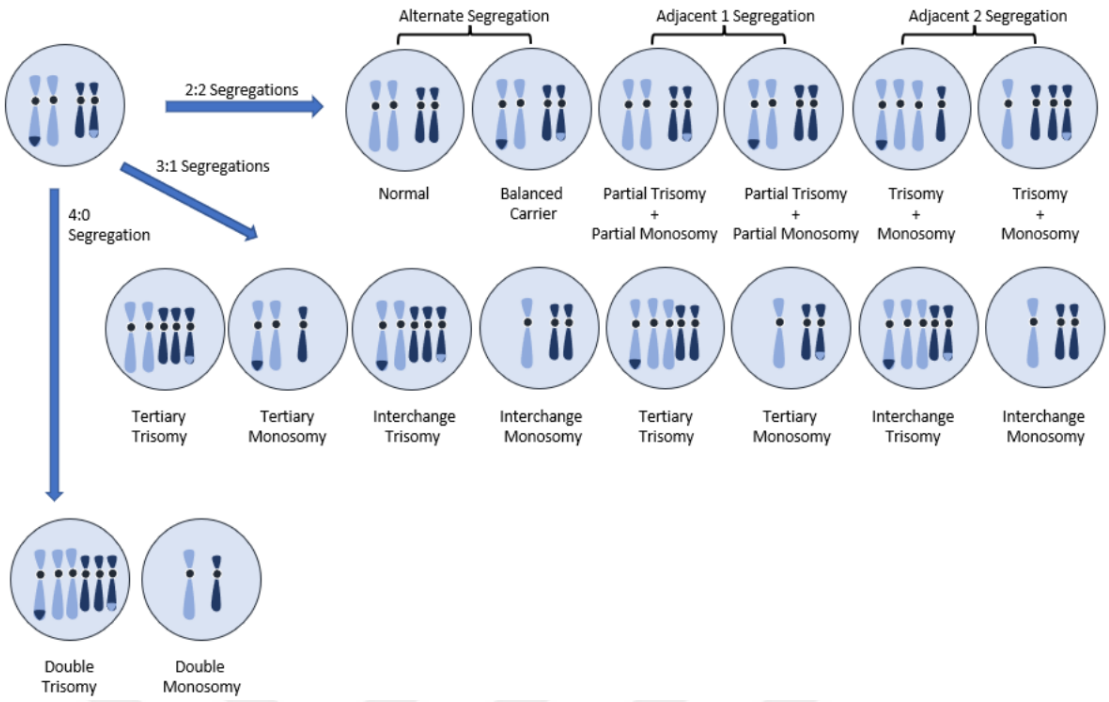


Figure 1.3 Segregation modes of a RECT carrier [5]

Table 1.1 Segregation modes and the resulting zygotes and embryos. Only major segregation modes are shown for simplicity.

Segregation mode	Segregational products/gametes	Chromosome composition in zygote	Result
2:2 alternate	A,B	A,A,B,B	Normal
	der(A),der(B)	A,der(A),B,der(B)	Balanced
2:2 adjacent-I	A,der(B)	A,A,B,der(B)	Partial trisomy (TS*-A), partial monosomy (TS of B)
	B,der(A)	A,der(A),B,B	Partial monosomy (TS of A), partial trisomy (TS of B)
2:2 adjacent-II	A,der(A)	A,A,der(A),B	Partial trisomy (CS**of A), partial monosomy (CS of B)
	B,der(B)	A,B,der(B)	Partial monosomy (CS of A), partial trisomy (CS of B)
3:1	A	A,A,B	Interchange monosomy (B)
	B,der(A),der(B)	A,der(A),B,B,der(B)	Interchange trisomy (B)
	B	A,B,B	Interchange monosomy (A)
	A,der(A),der(B)	A,A,der(A),B,der(B)	Interchange trisomy (A)
	der(A)	A,der(A),B	Tertiary monosomy (TS of A), tertiary monosomy (CS of B)
	A,B,der(B)	A,A,B,B,der(B)	Tertiary trisomy (TS of A), tertiary trisomy (CS of B)
	der(B)	A,B,der(B)	Tertiary monosomy (CS of A), tertiary monosomy (TS of B)
	A,B,der(A)	A,A,B,B,der(A)	Tertiary trisomy (CS of A), tertiary trisomy (TS of B)
4:0	A,B,der(A),der(B)	A,A,B,B,der(A),der(B)	double trisomy (A and B)
	NULL	A,B	double monosomy (A and B)

* translocated segments
** centric segments

1.1.1.2 Robertsonian Translocation

Robertsonian translocations (ROBT) result from the fusion of the long arms of two acrocentric chromosomes (13,14,15,21 and 22) with the loss of p arms which reduces the total chromosome number to 45. The incidence of ROBT is 1 in 1085 births, whereas the incidence is 6.5-9.4 times higher in infertile males than in newborn population [2].

During meiosis I, in order to allow synapsis of the homologous chromosomes, a trivalent structure is formed with 2 structurally normal and 1 derivative chromosome. The segregation results in eight different products: normal chromosomes or balanced derivative chromosome (producing a normal or carrier gamete respectively) are produced by alternate segregation; four products with nullisomy or disomy for one chromosome are produced by adjacent segregation and two segregational products with nullisomy or disomy of both chromosomes are produced via 3:0 segregation [6] (Figure 1.4).

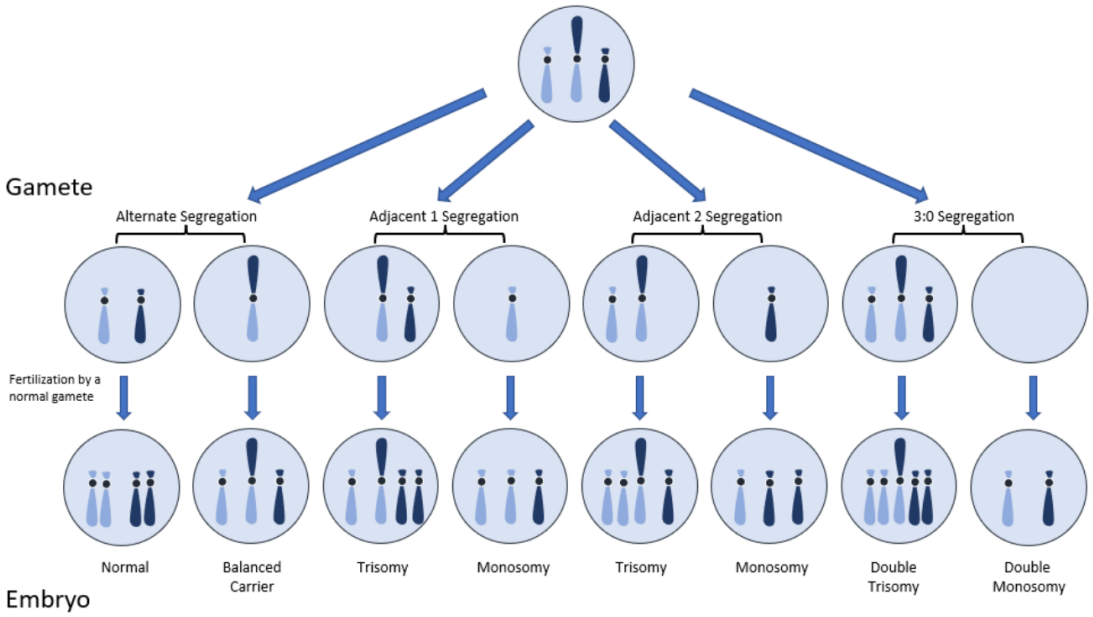


Figure 1.4 Segregation modes of a ROBT carrier [5]

1.1.1.3 Inversions

Inversions (INV) are a type of chromosomal rearrangements that result from two double stranded DNA breaks on a single chromosome and rotation of the segment before reunion. Balanced inversions do not involve a loss of genetic information, but simply the gene order has been changed. There are two forms: pericentric inversions (PEI) where the inverted segment contains the centromere, and paracentric inversions (PAI) where both breaks and reunions occur within the same arm of the chromosome.

The incidence in the general population is approximately 0.1%-2% [3].

During gametogenesis, loop formation and crossing over might result in the unbalanced segregations. During pachytene, in order to have a homologous synapse between the inverted and non-inverted regions an inversion loop might form. If an odd number of crossovers occur within the loop, acentric fragments (which are lost) and dicentric bridges are formed that will eventually be split up and produce unbalanced gametes in PAI carriers (Figure 1.5). In PEI carriers, the outcome is often partial duplication and partial deletion of the distal parts resulting in monosomy and trisomy of these chromosomal segments [5, 7].

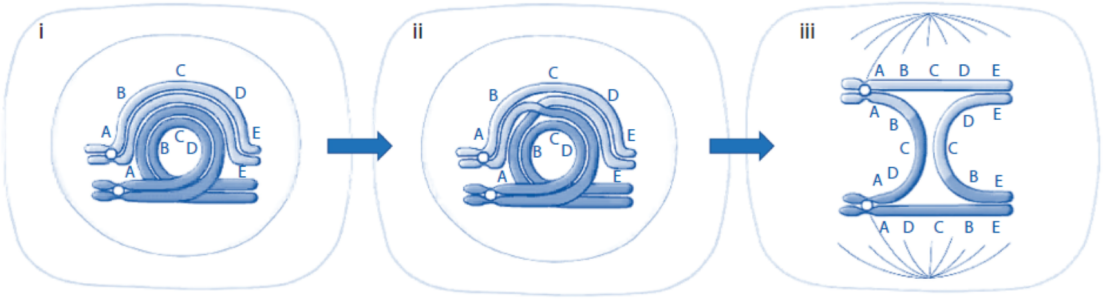


Figure 1.5 Loop formation in a PAI carrier. (i) after crossing-over event (ii); dicentric bridge and acentric fragment are formed (iii) [5]

1.1.1.4 Complex chromosomal rearrangements

Complex chromosome rearrangements (CCR) involve more than 2 chromosomal breaks and exchanges of (mostly distal) chromosomal segments. They are extremely rare and are mostly *de novo* origin (70%) [8]. CCRs are classified according to the number of chromosomes involved, the shape and the number of chromosomal breaks detected. They can be categorized into one of the following types: (i) two different rearrangements (double translocation), (ii) terminal exchanges of distal parts of three chromosomes (three-way translocation), and (iii) with more than three breakpoints and involving different types of rearrangement (terminal exchange, inversion, interstitial insertion, deletion) [9]. Balanced insertional translocations are also a type of complex chromosomal translocations. They occur by the insertion of a part of one chromosome into another non-homologous chromosome (interchromosomal) or into the same chromosome (intrachromosomal). Knowledge about segregational types is very limited. Segregational types differ with respect to the type of the complex translocations, however, hexavalent structures are thought to occur for three-way complex translocations which causes high proportion of unbalanced gametes observed in some of the previous studies (reviewed in [5]).

1.1.2 Preimplantation genetic testing for aneuploidy (PGT-A) and structural rearrangements (PGT-SR)

Preimplantation genetic testing (PGT) is a technique used to identify numerical and segmental aneuploidies in embryos that are generated via in vitro fertilization (IVF) techniques. PGT was first used clinically for an X-linked disorder with the aim of sexing the embryo in 1990 by the amplification of a Y-specific sequences using polymerase chain reaction (PCR) technique [10]. PGT techniques were evolved with the parallel development of IVF techniques such as embryo biopsy techniques and vitrification.

Aneuploidy is the presence of an abnormal number of chromosomes in a cell. Chromosomally normal human somatic cells contain 23 pairs (2n) of chromosomes. Any deviation resulting in either an excess (e.g., trisomy), missing chromosome(s) (e.g., monosomy) or a combination of different abnormalities (e.g., complex aneuploidy) can be defined as aneuploidy events (partial or whole chromosome). A further type of abnormality could affect the whole set of chromosomes resulting haploidy (one set of chromosomes [n]), triploidy (3n) or tetraplpoidy (4n) as well. Aneuploidy is very frequent in human preimplantation embryos and is the leading cause of pregnancy losses [11, 12]. Aneuploidy is generated often by chromosome disjunction errors and anaphase lagging in meiosis and mitosis. The frequency of aneuploidy is tightly related with female age. In a large study conducted with comprehensive chromosomal assessment methods it was shown that the proportion of aneuploid blastocysts increased from 30% in young women to nearly 90% in women with advancing age (44 years and above) [13].

Aneuploidy does not always affect all cells equally in a conceptus. It might also occur in a mosaic state, where there are at least two chromosomally distinct cell lineages present. Largely considered as a post-zygotic event, a variety of different mechanisms have been proposed in the formation of mosaicism such as mitotic non-disjunction, anaphase lagging and “trisomy rescue” [14].

PGT for aneuploidy (PGT-A) constitutes the majority of PGT cycles performed globally. The common indications for testing are advanced maternal age (AMA), recurrent miscarriages, recurrent implantation failures and severe male factor infertility [15]. Briefly, the aim is to reduce the risk of miscarriage and the birth of chromosomally abnormal children, decrease the time to live birth, and increase the success of the treatment per embryo transfer. However it is controversial with some recent publications claiming that it does not help to increase take home baby rate [16].

From the 1990s, PGT-SR has been a frequent indication in the context of IVF and the number is growing over the years with more carrier patients with high

reproductive risk seeking to increase their chances of a normal pregnancy. According to the multicenter analysis from the European Society of Human Reproduction and Embryology (ESHRE) Preimplantation Genetic Diagnosis (PGD) consortium, 4253 cycles had been performed for structural chromosomal abnormalities within 10 years (1997-2007) with an increasing trend upwards as the awareness of couples increases and as more cases are identified each year [15].

Although SR carriers may not have any phenotypic effects provided that the rearrangement does not disrupt any gene functions, they are at risk of infertility, miscarriages and having chromosomally abnormal children. In order to minimize these risks, prenatal and PGT techniques could be offered for carriers. However, prenatal genetic testing involves the analysis of an already established pregnancy and the couple might have to experience ethical, psychological and physiological burden of ending a pregnancy in such cases where the fetus is diagnosed as 'unbalanced'. Additionally, since a significant proportion of male carriers suffer from 'male factor infertility,' without the help of assisted reproductive techniques, the carrier couples might have to wait longer periods to achieve an implantation and a clinically recognized pregnancy. Therefore, PGT-SR, in relation to assisted reproduction techniques, is an option that offers the unique opportunity of selection and the transfer of only normal or balanced (normal/balanced) embryos [17]. In order for PGT-SR to be performed, a couple must undergo in vitro fertilization (IVF) treatments.

1.1.2.1 In vitro Fertilization (IVF) Treatment

The first applications of IVF in human embryos were performed by Steptoe and Edwards, in 1978 using Assisted Reproductive Technology (ART) which is a general term to indicate all procedures related with infertility treatments including IVF [18]. The IVF technique briefly involves the collection of egg(s) from the woman's ovaries and fertilization with sperm in laboratory [19]. The fertilized egg, called an embryo, is then transferred to the woman's uterus to implant and develop. To do this, hormonal stimulation is often required (Figure 1.8 Briefly, the pituitary gland is downregulated by administration of a gonadotrophin releasing hormone (GnRH) analogue (agonist or antagonist). Administration of follicle stimulating hormone (FSH) and/or human menopausal gonadotrophin (hMG) results in the development of multiple antral follicles. Once the follicle reaches a diameter of around 17mm, an artificial luteinizing hormone (LH) surge is usually generated by either a dose of human Chorionic Gonadotropin (hCG) or a combination of a high dose agonist/low dose hCG. After oocytes are collected, they are inseminated via intracytoplasmic sperm

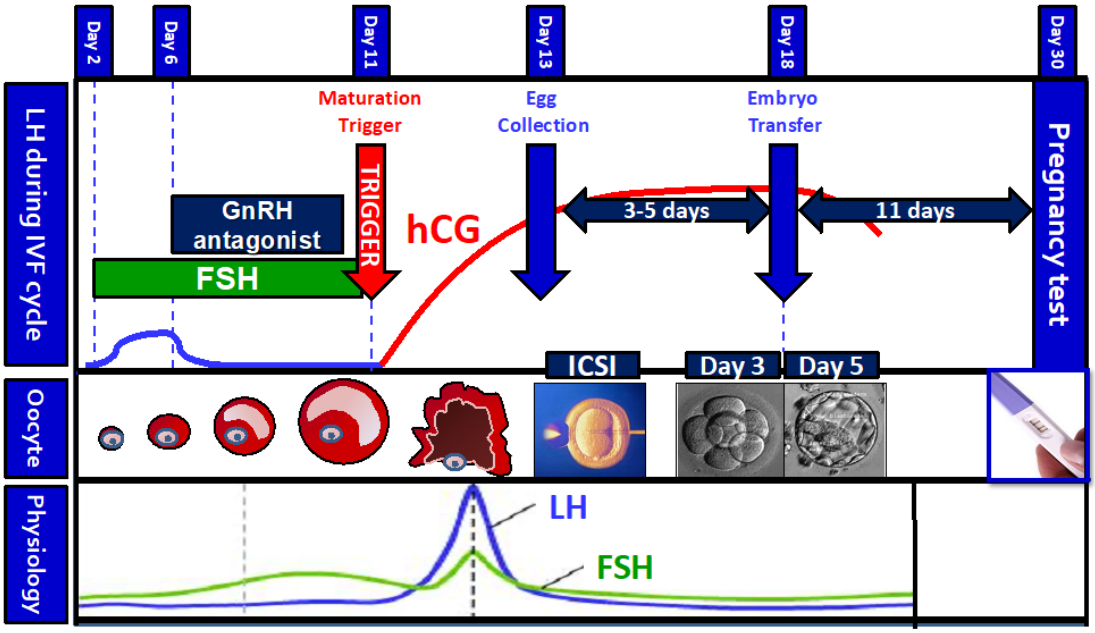


Figure 1.6 A general scheme of an IVF process [21]

injection (ICSI) technique [20]. Following fertilization, the embryos are kept in culture conditions until the day of transfer or vitrified if a frozen embryo transfer is planned.

1.1.2.2 Biopsy and Sampling methods

In order to analyze the embryo, there are three main biopsy techniques: polar bodies from oocytes, blastomeres from cleavage-stage embryos and trophectoderm from blastocysts (Figure 1.7). Recently, non-invasive sampling methods are becoming more and more popular as they do not pose any risk for the embryo.

Polar body (PB) biopsy

The polar bodies (PBs) are the by-products of meiotic division. PB biopsy is useful for maternal carriers and when embryo biopsy is restricted by the law in certain countries. PB biopsy cannot be used to diagnose abnormalities originating from paternal factors or post-meiotic errors such as mosaicism. For that reason, it is only applicable for only maternal meiotic abnormalities and translocations [22].

Cleavage-stage (blastomere) biopsy

Blastomere biopsy was once the most widely used method for PGT. Cleavage stage (blastomere) biopsy is performed by the removal of 1-2 blastomeres on day 3 embryos containing at least 6-8 blastomeres. Blastomeres express both maternal and paternal genomes, therefore this technique is both applicable to paternal and maternal rearrangement carriers. One of the major concerns of this technique is the possible damage on the further development of the biopsied embryo [23]. Another important drawback is the possible risk of misdiagnosis due to the high rate of mosaicism present at this stage of preimplantation development.

Blastocyst stage (trophectoderm) biopsy

The former techniques were replaced by blastocyst stage biopsy. In this technique, a small part of the trophectoderm (TE) is removed from a blastocyst. Since an average a blastocyst contains over 100 cells, the removal of 2-10 cells from the TE is not expected to have a detrimental effect on the blastocyst's development and particularly on the development of the fetus originating from the inner cell mass. The major advantage of this method is the amount of genetic material available for testing. The mosaicism rate is lower. However, the time for the analysis is limited by the implantation window which can be overcome by vitrification of the embryo. Blastocyst-stage biopsy became more popular after commercially available cultures were developed which support blastocyst development until day 5-6-7 in culture and with the advent of improved vitrification systems available [24].

Blastocentesis and analysis of spent culture medium

Blastocoel fluid or spent medium are alternative sources for DNA sampling. It is known that the fluid in the blastocoelic cavity contains DNA and it might be used as a source of analysis material for minimally invasive approach [25]. In this technique, fluid is aspirated using an ICSI pipette [26]. Live birth of a karyotypically normal, healthy child has been reported using spent culture medium for PGT-SR for a male ROBT (14:15) carrier. Jiao et al. used multiple annealing and looping-based amplification cycle (MALBAC) as an amplification technique which is superior when the DNA is scarce. The authors reported successful diagnosis of segmental aneuploidy, with high resolution ($\sim$ 1Mb) without maternal contamination in 41 blastocysts from 22 couples using mixed blastocoel fluid/spent medium. The diagnostic success of the combination of both blastocoel fluid and spent culture medium was higher though but still the main limitation of this method is maternal contamination [27].

1.1.2.3 Genetic analysis techniques

Mainly techniques involving fluorescence in-situ hybridization (FISH), array comperative genomic hybridization (aCGH) and next generation sequencing (NGS) techniques are used for the detection of segmental abnormalities in embryos. Since the first application of PGT, different techniques were evolved throughout the years (reviewed in Griffin and Ogur, 2018) [28]. The approach is to screen as many chromosomes as possible, and to detect unbalanced segmental abnormalities with high resolution (Table1.2).

Table 1.2 Comparison of PGT techniques

Detection parameters of techniques	FISH	aCGH	SNP-Array	NGS
Resolution of detection	50-100 Kb	5 Mb / 0.6-2.5 Mb	5-10Mb / 0.7-2.36 Mb	5-10Mb / 0,8-2Mb
Simultaneous 24 chromosome testing	-	+	+	+
Detection of de novo segmental abnormalities	-	+	+	+
Need for WGA (whole genome amplificiation)	-	+	+	+
Need for preclinical work up	+	-	+	-
Need for analysis of parental/grandparental DNA	+	-	+	-
Ability to distinguish normal from balanced	-	-	+	-

Kb: Kilo bases
Mb: Mega bases

Fluorescence in-situ Hybridization (FISH)

Briefly, this method consists of the following steps; spreading of the nucleus on the glass surface (fixation), hybridization with suitable fluorescent labeled probes, washing and analysis. FISH was initially used in 1992 using X and Y chromosomal probes for the diagnosis of embryos at risk of carrying sex-linked disorders [29].

The most pronounced disadvantage of this technique is it being highly objective in nature and being limited to only 5-9 chromosomes. FISH technique is mostly discontinued from use for these reasons, however it is still the gold standard for some rare carriers of cryptic translocations where aCGH and NGS resolutions are not sufficient.

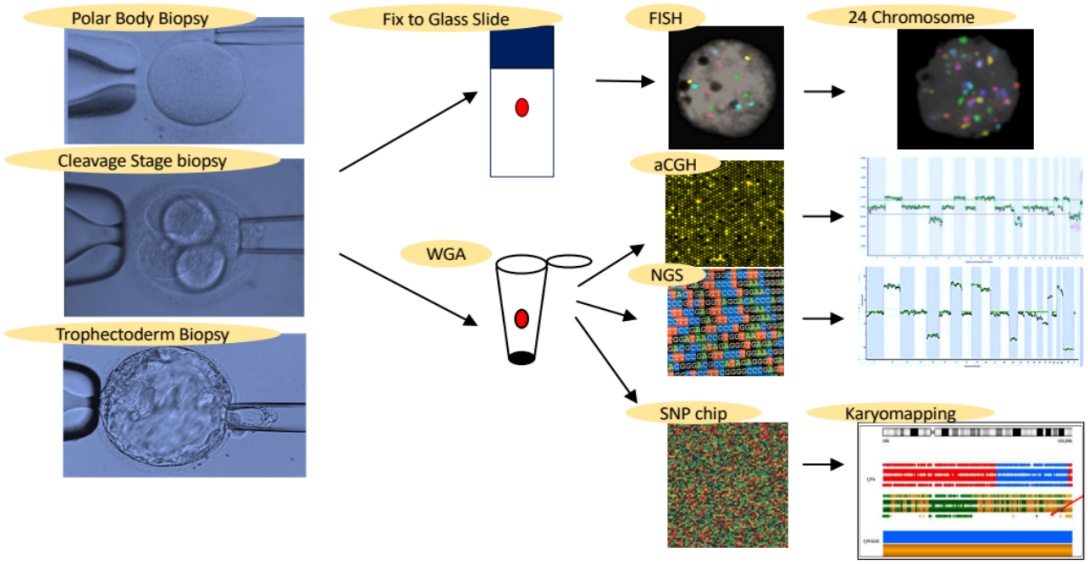


Figure 1.7 Sampling and analysis methods shown (Various approaches to PGT-SR) [5]

CGH, aCGH and NGS techniques require genome amplification methods (WGA) to increase the amount of starting DNA which is approximately 6pg-60pg depending on the number of cells taken per biopsy.

Comparative genomic hybridization (CGH)

The CGH technique is similar to FISH in terms of it also needs hybridization of probes onto glass surface. However in CGH, differentially-labelled test and reference DNAs are hybridized on normal metaphase chromosomes. Unlike FISH, CGH is a comprehensive analysis technique where all chromosomes could be analyzed; however the resolution is not optimal [30]. It was first applied to blastomeres in 1999 [31] but was described as time consuming especially in clinical settings [32]. With the help of this technique, random segmental abnormalities drew attention and give rise to the hypothesis that cleavage stage embryos are genomically unstable during in preimplantation development [33]. This technique provided a basis to develop other technologies such as aCGH and NGS.

Array-CGH (aCGH)

Array CGH technique is similar to CGH in that it requires amplification, labeling, hybridization, washing; however hybridizations which take place on a glass surface with bacterial artificial chromosomes (BACs) or oligonucleotide sequences rather than metaphase chromosomes. The analysis is performed in an automated way using software analysis program which assess the red:green fluorescent ratios on

approximately 3000 probes which are known to represent the regions throughout the genome. The losses and gains are detected by this automated analysis method, which determines the median log2 ratio for each chromosome (and each chromosomal segments). When the ratio for a specific chromosome is higher than +0.3 (.0.3), the software calls this as a "gain" and if this ratio is less than -0.3, it is called as a "loss". The resolution of BAC-arrays is small as 2.5 Mb or to 20-50 kb using oligonucleotide-arrays [34].

Next Generation Sequencing (NGS)

NGS is currently the most widely used method for PGT-SR. There are two major approaches in NGS technology: semiconductor sequencing based on detection of hydrogen ions released during DNA polymerization (Ion Torrent, Thermo-Fisher Scientific), and Illumina sequencing based on sequencing by synthesis using fluorescent (terminator) dyes(VeriSeq, Illumina). Their resolution depends on the region of the genome but roughly changes from 2Mb-10Mb. It involves the parallel sequencing of a small but representative proportion of the whole genome by using a DNA barcoding system to identify samples, multiple samples could be pooled together which also makes this method as cost effective among others. It detects all types of aneuploidy simultaneously, can incorporate mitochondrial DNA analysis and its greater dynamic range permits the diagnosis of mosaicism more effectively. In recent years, NGS has become more popular. Nakano et al, 2022 demonstrated the effectiveness of NGS in PGT-SR using aCGH and NGS in 242 blastocysts [35]. Similarly, Chen et al. 2020 combined PGT for monogenic disorders (PGT-M), PGT-A and PGT-SR in an NGS strategy, combining with haplotyping to produce a cost-effective universal PGT protocol [36].

Single Nucleotide Polymorphism (SNP)-arrays and Karyomapping

A single-nucleotide polymorphism (SNP) is defined as a DNA sequence variant that occurs every one in 1000 nucleotides. SNP arrays are designed to detect aneuploidy and single gene disorders [37, 38]. The principle is based on the polymorphic markers which are denoted by AA, AB and BB at each locus and analyzed in comparison to the human HapMap reference. SNP arrays provides higher resolution than any other technique. SNP-arrays were also able to distinguish balanced carriers from normal embryos [39]. The only disadvantage of this technique is the obligation of availability of parent DNA or at least one unbalanced embryo as a reference. Karyomapping was designed as a universal protocol for PGT-M and PGT-A and, by extension therefore, is applicable for PGT-SR. Karyomapping is based on the identification of parental/grandparental haploblocks (inherited chromosomal segments) in the

corresponding embryos and provides a universal approach to all PGT, including PGT-A, PGT-M and PGT-SR in one biopsy sample [40]. In 2019, Beyer et al. used Karyomapping in a PGT-SR setting, and successfully validated that the technique is able to distinguish normal/balanced outcomes from unbalanced [41]. As a result, SNP based approach is an effective methodology for PGT-SR with the addition of discriminating balanced and normal embryos. Whether such detection is clinically necessary or practical is a question to be answered.

1.1.3 Machine learning techniques in the prediction of PGT-SR outcomes

The recent advance in machine learning technology gave rise to the development of computer-based prediction models in medical fields, such as neurology, drug design, cardiology, and ART [42]. In the context of IVF, prediction models have long been used for embryo selection purposes via morphokinetics also known as time-lapse technology [43] or to create models which can highly predict the outcome of genetic analysis results and the outcome of treatments. There are two main categories of machine learning algorithms: supervised and unsupervised learning. Supervised learning algorithms are trained on labeled data. This means that the data used to train the model has both input features and their corresponding output or label. The goal of the algorithm is to learn a mapping function that can predict the output for new, unseen inputs. Some examples of supervised learning algorithms include linear regression, logistic regression, decision trees, random forests, and neural networks. Unsupervised learning algorithms, on the other hand, do not use labeled data. They are used to find patterns or structures in data without any prior knowledge of what the output should be. The goal of unsupervised learning is to discover hidden relationships or groupings in the data. Some examples of unsupervised learning algorithms include clustering algorithms such as k-means, hierarchical clustering, and density-based clustering, as well as dimensionality reduction techniques such as principal component analysis (PCA) and t-SNE.

Semi-supervised learning is another category of machine learning that combines both supervised and unsupervised learning techniques. It is used when only a small portion of the data is labeled, and the rest is unlabeled. The goal is to use the labeled data to guide the unsupervised learning process and improve the accuracy of the model.

In medical science, logistic regression, support vector machines, naive Bayes, random forest, artificial neural networks and decision trees are the most widely used machine learning algorithms. Each algorithm has its own strengths and weaknesses, and the choice of algorithm depends on the specific application and the data available.

1.1.3.1 Logistic Regression

Logistic regression is a statistical model that is used to analyze the relationship between a set of independent variables and a binary dependent variable. It has been used in medical science for predicting the risk of diseases, such as cancer and diabetes [44].

1.1.3.2 Support Vector Machines (SVM)

SVM is a type of supervised learning algorithm that can be used for classification or regression analysis. SVM has been used in medical science for predicting the outcome of cancer treatment and for diagnosing neurological disorders [45].

1.1.3.3 Naive Bayes

Naive Bayes is a probabilistic algorithm that is used for classification analysis. It has been used in medical science for diagnosing diseases based on symptoms and for predicting the effectiveness of treatments [46].

1.1.3.4 Decision Trees

Decision trees are a type of supervised learning algorithm that is used for classification and regression analysis. They have been used in medical science for predicting the likelihood of developing a particular disease or condition [47].

1.1.3.5 Random Forest Model

Random Forest (RF) is one of the most popular algorithms used in machine learning for both classification and regression tasks. It is known for its robustness, accuracy, and ability to handle high-dimensional data.

RF model is based on decision trees. Decision trees are a type of machine learning model that makes predictions by splitting the data into smaller subsets based on the most important features using feature selection. Random forest is an ensemble learning method that constructs a multitude of decision trees and combines their predictions. RF creates randomness first by selecting a random subset of features at each node of the decision tree, and second, by using a random subset of the training data to build each decision tree. Each decision tree in the forest independently makes a prediction, and the final prediction is based on the majority vote of all the decision trees.

RF is widely used in a variety of applications, including image recognition, natural language processing, and recommendation systems. It has been used in medical science for predicting the severity of diseases and for identifying high-risk patients [48].

1.1.3.6 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a type of machine learning algorithm inspired by the structure and function of the human brain. ANNs are composed of layers of interconnected nodes, or "neurons," that process and transmit information through the network.

ANNs consist of an input layer, one or more hidden layers, and an output layer. Each layer consists of a set of neurons. Each neuron receives input from other neurons, performs a computation, and produces an output. The computation typically involves multiplying the input by a set of weights and adding a bias term, followed by applying an activation function. the predictions are done by adjusting the weights and biases of the neurons to minimize the error between the predicted output and the actual output. This is typically done using an algorithm called backpropagation. Activation functions are used to introduce non-linearity into the computation performed by the neurons. Common activation functions include sigmoid, ReLU (rectified linear unit), and tanh (hyperbolic tangent). Overall, ANNs are a powerful machine learning technique that can learn complex patterns in data and make accurate predictions.

ANNs have been successfully applied to a wide range of problems, including image and speech recognition, natural language processing, and financial forecasting. ANNs have been used in medical science for image analysis, such as identifying tumors in medical images, and for predicting disease outcomes ([49]).

The main difference between Random Forests and ANNS algorithms is the presence of nodes in ANN, while there are decision trees in Random Forest model in between input and output (Figure1.8).

1.2 The Objective of the Thesis

The purpose of this thesis was to provide one of the largest examinations of clinical and post-zygotic cytogenetic outcomes of SR carriers from a single center to date and to identify which cycle specific parameters have impact on the genetic results and chromosomal segregations. A further objective of this thesis was to test the hypothesis that whether SRs induce further chromosome abnormalities as a result

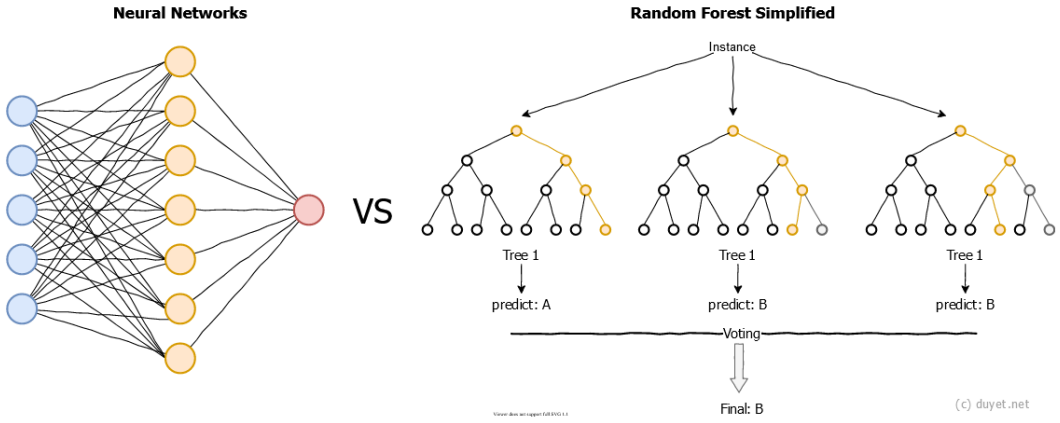


Figure 1.8 Neural network vs random forest algorithms [50]

of an ICE. By using machine learning algorithms in R software, this work aimed to create models which can highly predict the outcome of genetic analysis results and the clinical outcomes of treatment cycles based on the same input data.

1.3 The Hypothesis

The ICE hypothesis has not been questioned adequately, however this study is the first to establish an analytical model to answer the aforementioned questions. Large sample size from one center brought power and homogeneity to the statistical analysis. Establishment of a control group by the selection of cases from the same period of treatment with matching parameters such as female age and the number of analyzed embryos made this control group the best possible to date. Statistical tools such as effect size measurement was used in order not to fall into statistical errors in the interpretation of the results. As a result, this study made it possible to test this old hypothesis [51], using a novel analytical model for the first time. In addition, using machine learning algorithms predictive model was established which would be further developed into a web-based product or application for the help of patients and physicians in the planning of their treatments. This study helped to provide a predictive model with the use of advanced statistical and machine learning tools that helped to reveal parameters to provide personalized treatment and better genetic counselling for PGT-SR couples in the future.

2
MATERIALS AND METHODS

Briefly, the genetic analyses were performed with the use of FISH technique until 2011, aCGH technique between 2011-2016 and NGS between 2015-2019 years, respectively. The data analysis was performed between years 2016-2020. The development of prediction models in R software were performed during 2021-2022. The paper related with thesis work was completed and submitted in 2022 and was accepted for publication in 2022. Thesis was written using Overleaf, Online LaTeX editor in between 2022-2023 (Figure 2.1).

2.1 Patients

This study was designed as a retrospective study including the genetic and clinical outcome results of 300 couples with structural rearrangements referred to a private IVF clinic based on Istanbul Şişli Memorial Hospital between years 2011-2019. The period when FISH technique were used was excluded since this technique could not provide comprehensive information of all chromosomes. Among 300 couples, 198 were carriers of RECT; 60 were carriers of ROBT; 31 were carriers of INV and 11 were carriers of CCR. Since it is assumed to be a polymorphism, carriers of the classical pericentric inversion "inv(9)(p11q12)" were excluded from this study. A total of 31 couples with INV consisted of 17 female and 14 male carriers. Among those, 7 couples were carrying a paracentric inversion (PAI) and 24 were carriers of a pericentric inversion (PEI). The detailed data regarding karyotypes could be found in Table 2.1. Additionally, the INV carriers were further assessed by the length in megabases (Mb) and the proportion of inverted segment (%) which can be found in Table 2.1. Written informed consent was obtained from all patients. The study was approved by the ethical committee of Istanbul Memorial Hospital. Written informed consent was obtained from all patients. The study was approved by the ethical committee of Istanbul Memorial Hospital.

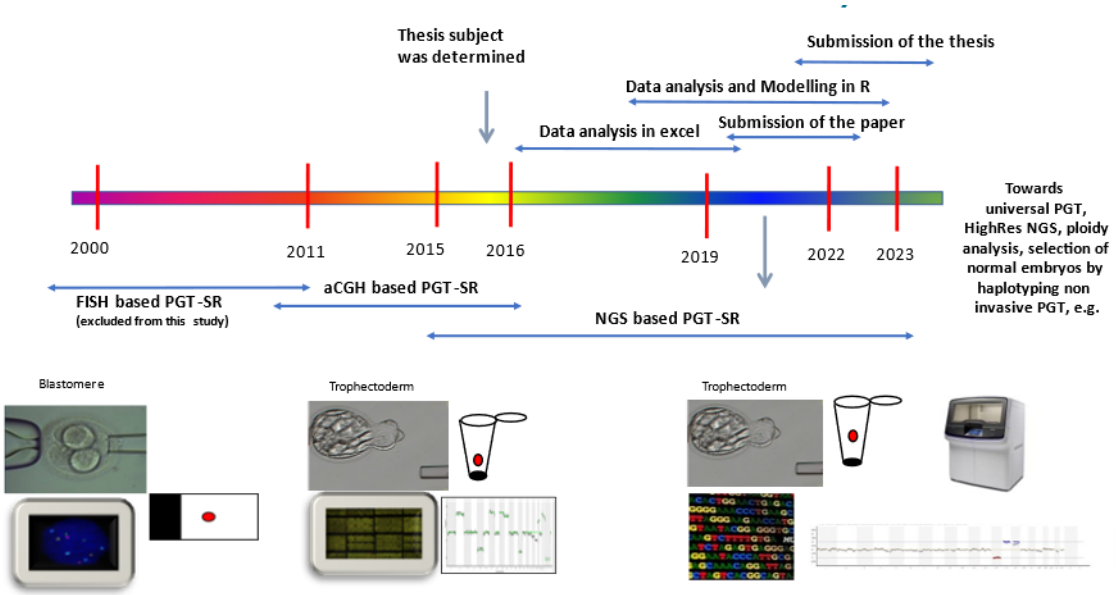


Figure 2.1 Timeline of thesis progress and the parallel evolution of PGT techniques

2.2 IVF Treatment and Biopsy protocol

All couples underwent IVF treatment in order to produce embryos to be analyzed by PGT-SR technique.

2.2.1 Ovarian Stimulation

In order to produce multiple eggs, the ovaries were stimulated with the administration of hormone medications (ovulation drugs) also known as "controlled ovarian stimulation" technique. GnRH antagonist protocol or GnRH analogue suppression with recombinant FSH or a combination of FSH and hMG were used, respectively (Gonal F®, Merck Serono, Switzerland) [52]. Oocyte retrievals were carried out by transvaginal ultrasound guidance and performed 36h after the injection of recombinant hCG, (Ovitrelle®, Merck Serono, Switzerland). ICSI was used as fertilization method [20]. Embryo biopsy was performed at blastocyst stage on day 5 or 6 after fertilization, by removal of 2-5 cells from trophectoderm tissue [24].

2.2.2 Trophectoderm biopsy procedure

1. For the embryos which were planned for biopsy and genetic analysis, an opening was made in the zona pellucide (ZP) on day 3 or 4 using laser.

2. The morphology of the blastocyst(s) were assessed and the suitable ones (decided on several parameters such as morphology and the quality of the embryo) were taken (out) from the dish on day 5.

3. A biopsy dish was prepared and pre-warmed.
4. The blastocysts were transferred into the first column of droplets containing 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)-buffered culture medium.
5. The blastocyst were washed several times to remove any residual culture medium and the embryos were moved to the second column, same row.
6. The blastocyst was held and positioned with a holding pipette at the opposite of the herniated TE cells not to mistakenly damage herniated TE cells.
7. Herniated TE cells (range 2 to 9 cells) were smoothly aspirated with the biopsy pipette.
8. A gentle suction pressure 1-2 time(s) was applied to cause collapse of the herniated TE cells.
9. A mechanical rubbing on the holding pipette could be done to promote separation if needed, or a combination of single laser pulses at TE cell junctions with mechanical stretching were used to separate the TE piece (Figure 2.2).
10. The blastocyst was released from the holding pipette.
11. The TE cells were aspirated and placed on the heated stage of the lamina flow hood
12. The biopsied blastocyst was taken out from the biopsy medium.
13. Patient name or label was checked and the embryo was placed back into the dish with routine culture media after several washings to remove any residual HEPES- buffered medium.
14. The dish was placed back into the gased incubator.
15. A PCR tube was prepared containing 2.5 ul of Phosphate buffered saline (PBS) and labeled with initials of the patient and the embryo number.
16. The biopsied trophectoderm cells were placed inside the corresponding PCR tube.

Table 2.1 Karyotypes of 300 couples with structural rearrangements. NA: not available, INV: inversion, ROBT: Robertsonian Translocation, RECT: Reciprocal translocation, CCR: Complex chromosomal rearrangements (Permission from [53])

Patient code	**Karyotype-Female**	**Karyotype-Male**
RECT-1	46,XX,t(6;10)(p21.3;q26.1)	46,XY
RECT-2	46,XX	46,XY,t(10;19)(q11.2;q13.4)
RECT-3	NA	46,XY,t(4;7)(q35.1;q31.2)
RECT-4	NA	46,XY,t(11;22)(q23;q11.2)
RECT-5	46,XX	46,XY,t(4;20)(p14;p11.2)
RECT-6	46,XX,t(9;10)(q34.1;q26.2)	NA
RECT-7	46,XX,t(4;16)(q31.1;q12.1)	46,XY
RECT-8	NA	46,XY,t(17;22)(q23;p13)
RECT-9	46,XX	46,XY,t(11:22)(q23:q11)
RECT-10	46,XX,t(8;9)(q24.1;q22.3)	46,XY
RECT-11	46,XX,t(10;12)(q25;q23)	46,XY
RECT-12	46,XX,t(17;22)(q23;q11.2)	NA
RECT-13	46,XX	46,XY,t(7;9)(q22;q32)
RECT-14	46,XX	46,XY,t(3;15)(q22.2;q26.1)
RECT-15	46,XX	46,XY,t(1;16)(q21;p11.2)
RECT-16	46,XX	46,XY,t(7;13)(p15;q21)
RECT-17	46,XX,t(1;7)(q12;q11)	46,XY,1qh+
RECT-18	46,XX,t(1;9)(p22;q21.2)	NA
RECT-19	46,XX	46,XY,t(11;22)(q23;q11)
RECT-20	46,XX,t(16;20)(q13;q13.1)	46,XY
RECT-21	46,XX,t(12;14)(p11;q22)	46,XY
RECT-22	46,XX,t(11;22)(q23.3;q11.21)	NA
RECT-23	46,XX,t(2;6)(q33;q23)	46,XY
RECT-24	46,XX	46,XY,t(1;10)(q25;26.3)
RECT-25	46,XX,t(2;22)(q31;q13.1)	46,XY
RECT-26	46,XX,t(1;3)(p35;p23)	46,XY
RECT-27	46,XX,t(3;7)(p26.3;q31.3)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	**Karyotype-Female**	**Karyotype-Male**
RECT-28	46,XX,t(9;22)(p13.1;q12.2)	NA
RECT-29	46,XX,t(9;12)(q22.3;q13.3)	46,XY
RECT-30	46,XX	46,XY,t(2;3)(q21;p11.2)
RECT-31	46,XX	46,XY,t(14;16)(q32;p11)
RECT-32	46,XX	46,XY,t(2;7)(q21;p22)
RECT-33	46,XX,t(11;22)(q25;q13.1)	NA
RECT-34	46,XX,t(10;11)(q24.3;p13)	46,XY
RECT-35	46,XX,t(4;6)(q23;p21.3)	46,XY
RECT-36	46,XX,t(1;13)(p31;q12)	46,XY
RECT-37	46,XX	45,X,t(Y,22)(p10;q10)
RECT-38	46,XX	46,XY,t(5;17)(q22;q21)
RECT-39	NA	46,XY,t(17;22)(q23.1;q11.2)
RECT-40	46,XX	46,XY,t(11;14)(q13;q24)
RECT-41	46,XX,t(4;18)(p15.2;p11.2)	46,XY
RECT-42	46,XX,t(1;15)(q12;p11.1)	46,XY
RECT-43	46,XX,t(9;10)(q32;q22)	46,XY
RECT-44	NA	46,XY,t(8;12)(q24.13;q22)
RECT-45	NA	46,XY,t(10;14)(q11.1;p11.1)
RECT-46	46,XX,t(5;17)(p35;q21)	46,XY,1qh+
RECT-47	46,XX	46,XY,t(8;11)(q24.1;q23)
RECT-48	NA	46,XY,t(1;10)(q12;q11.2)
RECT-49	46,XX,t(2;12)(q13;q15)	46,XY
RECT-50	46,XX	46,XY,t(7;19)(p21.2;q13.4)
RECT-51	NA	46,Y,(X;5)(p11,2;q22)
RECT-52	NA	46,XY,t(1;22)(q25;p11.2)
RECT-53	46,XX,t(6;21)(q23;q21)	46,XY
RECT-54	46,XX	46,XY,t(5;13)(q13;p11.2)
RECT-55	46,XX	46,XY,t(11;17)(q24.2;q21)
RECT-56	46,XX,t(1;12)(q32;q22)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
RECT-57	46,XX,t(1;9)(p22;p21)	NA
RECT-58	46,XX	46,XY,t(17;19)(q23.2;p13.3)
RECT-59	46,XX,t(4;5)(p15.1;q13)	46,XY
RECT-60	46,XX,t(11;18)(q22.2;q11.2)	46,XY
RECT-61	46,XX	46,XY,t(2;3)(q11.2;q25)
RECT-62	46,XX,t(9;20)(q22.1;p13)	46,XY
RECT-63	46,XX,t(15;16)(q11.2;p13.1)	NA
RECT-64	46,XX,t(11;15)(q11;p11)	46,XY
RECT-65	46,XX	46,XY,t(6;11)(p10;q10)
RECT-66	46,XX	46,XY,t(12;16)(q21.3;q24)
RECT-67	46,XX,t(1;13)(q23;p11.2)	46,XY
RECT-68	46,XX	46,XY,t(3;8)(q25.1;p21.3)
RECT-69	46,X,t(X;4)(q24;q31.2)	46,XY,21ps
RECT-70	46,XX,t(1;20)(q21,3;p11.21)	46,XY
RECT-71	46,XX,t(7;22)(q22;q11.2)	46,XY
RECT-72	NA	46,XY,t(1;4)(p32;q23)
RECT-73	46,XX,t(13;15)(q13.2;q15)	46,XY
RECT-74	NA	46,XY,t(7;10)(p11.2;q24.3)
RECT-75	46,XX,t(2;3),t(q31;q29)	NA
RECT-76	46,XX	46,XY,t(3;21)(q11.1;p11.1)
RECT-77	NA	46,XY,t(11;18)(p15.1;q21.1)
RECT-78	46,XX	46,XY,t(4;5)(qter;q13.2)
RECT-79	46,XX	46,XY,t(6;15)(p11;p11)
RECT-80	46,XX,t(10;12)(p14;p11.2)	46,XY
RECT-81	46,XX,t(5;15)(p15.1;q15)	46,XY
RECT-82	46,XX	46,XY,t(10;13)(p15;q12)
RECT-83	46,XX,t(1;21)(p32;q22.3)	46,XY
RECT-84	NA	46,XY,t(1;4)(q23;p12)
RECT-85	46,XX,t(11;22)(q11.2;q23.3)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
RECT-86	46,XX	46,XY,t(10;18)(q22;q21)
RECT-87	46,XX	46,XY,t(6;10)(q25.1;q22.1)
RECT-88	46,XX	46,XY,t(2;7)(q21.1;p15)
RECT-89	46,XX,t(9;11)(q21,2;q31)	NA
RECT-90	46,XX,t(11;22)(q23.3;q11.2)	46,XY
RECT-91	46,XX	46,XY,t(1;11)(q10;p15)
RECT-92	46,XX,t(1;21)(q24;q22.3)	46,XY
RECT-93	46,XX	46,XY,t(5;8)(q13;q24.1)
RECT-94	46,XX,t(9;11)(q32;q34)	46,XY
RECT-95	NA	46,XY,t(11;22)(q23;q11.2)
RECT-96	46,XX,t(6;13)(q25.1;q14.3)	46,XY
RECT-97	46,XX,t(4;14)(p15.3;q31)	46,XY
RECT-98	46,XX	46,XY,t(2;8)(q35;q22.1)
RECT-99	46,XX,t(1;4)(p32;q31,3)	46,XY
RECT-100	46,XX,t(6;11)(p21,3;p1.3)	NA
RECT-101	46,XX	46,XY,t(13;15)(q31;p11.1)
RECT-102	46,XX	46,XY,t(9;16)(p13;p13.1)
RECT-103	46,XX,t(4;5)(p12;q13)	NA
RECT-104	46,XX	46,XY,t(17;22)(q12;q13)
RECT-105	46,XX,t(5;11)(q35.3;pter)	46,XY
RECT-106	46,XX,t(8;22)(q24.1;q11.2)	46,XY
RECT-107	46,XX	46,XY,t(2;10)(q13;q11.2)
RECT-108	46,XX	46,XY,t(4;20)(p16;p11.2)
RECT-109	46,XX,t(7;13)(p22;q32)	NA
RECT-110	46,XX,t(11;22)(q23.3;q11.2)	46,XY
RECT-111	46,XX,t(3;18)(q27;q12)	46,XY
RECT-112	46,XX,t(6;12)(q27;q13.3)	46,XY
RECT-113	46,XX,t(2;3)(q11.2;q26.1)	46,XY
RECT-114	46,XX,t(8;18)(q13;q11.1)	NA

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
RECT-115	46,XX,t(4;10)(q23;q24.1)	46,XY
RECT-116	46,XX	46,XY,t(14;16)(q32;q11.1)
RECT-117	46,XX,t(2;10)(p14;q24)	46,XY
RECT-118	46,XX	46,XY,t(5;13)(q33;q32)
RECT-119	46,XX,t(8;10)(q21;q21)	46,XY,9qh+
RECT-120	46,XX,t(1;17)(q32.1;p11.2)	46,XY
RECT-121	46,XX,t(2;4)(q31;q31.3)	46,XY
RECT-122	46,XX,t(2;6)(q22;q24)	46,XY
RECT-123	46,XX,t(1;11)(p36.3;p11.2)	46,XY
RECT-124	46,XX,t(2;11)(q14;q23)	NA
RECT-125	46,XX	46,XY,t(2;7)(p22;q37)
RECT-126	46,XX,t(1;16)(p22;q12)	46,XY
RECT-127	46,XX,1gh+	46,XY,t(1;10)(q44;q24)
RECT-128	46,XX	46,XY,t(9;12)(p24;q22)
RECT-129	46,XX,t(3;6)(q26.2;p21.3)	46,XY
RECT-130	46,XX	46,XY,t(2;17)(q21;q21)
RECT-131	46,XX	46,XY,t(15;16)(q24.2;q23.1)
RECT-132	NA	46,XY,t(2;8)(q31;q21.2)
RECT-133	46,XX	46,XY,t(11;22)(q25;q13)
RECT-134	46,XX,t(7;18)(p15;p11.3)	46,XY
RECT-135	46,XX,t(14;20)(q22.3;q13.31)	NA
RECT-136	46,XX	46,XY,t(3;18)(q27;q21.1)
RECT-137	46,XX,t(5;18)(q31;q23)	46,XY
RECT-138	46,XX	46,XY,t(6;19)(q22;q13:1)
RECT-139	NA	46,XY,t(1;9)(q12;p22)
RECT-140	46,XX	46,XY,t(6;8)(p25;p22.3)
RECT-141	46,XX,t(4;14)(p12;q22)	46,XY
RECT-142	46,XX,t(1;10)(p36.1;q26)	46,XY
RECT-143	46,XX,t(7;13)(q34;q13)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
RECT-144	NA	46,XY,t(4;19)(q27;q13.3)
RECT-145	46,XX	46,XY,t(7;17)(p15;p13)
RECT-146	46,XX	46,XY,t(1;19)(p32;q13)
RECT-147	46,XX	46,XY,t(1;2)(p34.1;p23)
RECT-148	46,XX,t(4;16)(q26;q23)	46,XY
RECT-149	46,XX,t(9;19)(p12;q13.4)	46,XY
RECT-150	46,XX	46,XY,t(1;9)(p31?;p22)
RECT-151	46,XX,t(17;19)(p13.1;p13.3)	46,XY
RECT-152	46,XX	46,XY,t(1;16)(p34.1;p11.2)
RECT-153	NA	46,XY,t(5;6)(q35;q21)
RECT-154	46,XX,t(1;8)(q31.2;q21.3)	46,XY
RECT-155	46,XX	46,XY,t(6;10)(q13;q22)
RECT-156	46,XX	46,XY,t(2;5)(p11.2;q33)
RECT-157	46,XX	46,XY,t(4;11)(p16;q23)
RECT-158	46,XX	46,XY,t(5;6)(q31.2;q21)
RECT-159	46,XX,t(11;22)(q23;q12)	46,XY
RECT-160	46,XX,t(8;15)(q11.23;q15)	46,XY
RECT-161	46,XX	46,XY,t(20;22)(q13.3;q11.2)
RECT-162	46,XX,t(13;21)(q22;q21)	46,XY
RECT-163	NA	46,XY,t(1;14)(p31.2;q11.2)
RECT-164	46,XX,t(4;15)(q25;q21)	46,XY
RECT-165	NA	46,XY,t(3;17)(p21;q25)
RECT-166	46,XX,t(4;16)(q13;q24)	46,XY
RECT-167	46,XX,t(7;9)(p15:2;p22)	46,XY
RECT-168	NA	46,XY,t(7;14)(q32;p13)
RECT-169	46,XX,t(8;21)(q11.2;p11.2)	46,XY
RECT-170	46,XX	46,XY,t(1;2)(p36;p14.1)
RECT-171	46,XX,t(16;18)(p13.3;p11.2)	46,XY
RECT-172	46,XX	46,XY,t(17;20)(q25;q12)

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
RECT-173	46,XX,t(2;20)(q21;q13)	46,XY
RECT-174	46,XX,t(11;16)(q23;q24)	46,XY,inv(9)(p12q13)
RECT-175	46,XX	46,XY,t(8;11)(p23;p15)
RECT-176	46,XX,t(3;6)(p13;p21.1)	46,XY
RECT-177	46,XX,t(2;15)(q21;q13)	46,XY
RECT-178	46,XX,t(11;22)(q25;q13)	46,XY
RECT-179	NA	46,XY,t(17;22)(q23;q11.2)
RECT-180	46,XX	46,XY,t(11;16)(q23;q24)
RECT-181	46,XX,t(5;18)(q13.1;q12.2)	46,XY
RECT-182	46,XX	46,XY,t(6;21)(q25;q22)
RECT-183	46,XX	46,XY,t(10;14)(q22.3;q24.1)
RECT-184	46,XX,t(4;15)(q25;26.1)	46,XY
RECT-185	46,XX,t(16;18)(p13.1;p11.2)	46,XY
RECT-186	46,XX,t(2;3)(p13;q29)	46,XY,9qh+
RECT-187	46,XX	46,XY,t(4;19)(q32;q13.2)
RECT-188	46,XX	46,XY,t(7;10)(p12;q26)
RECT-189	46,XX,t(7;14)(q36.1;q11.2)	46,XY
RECT-190	46,XX,t(1;20)(p36.1;p12)	46,XY
RECT-191	46,XX	46,XY,t(11;21)(q11;q11)
RECT-192	NA	46,XY,t(1;11)(p31;q24.2)
RECT-193	46,XX	46,XY,t(4;9)(p12;p12)
RECT-194	46,XX,t(3:11)(p25;q25)	46,XY
RECT-195	NA	46,XY,t(8;14)(q22;p12)
RECT-196	46,XX	46,XY,t(3;13)(p13;p11.2)
RECT-197	46,XX	46,XY,t(1;16)(q11;p11)
RECT-198	NA	46,XY,t,(3;20)(p21.3;p13)
ROBT-1	NA	45,XY,der(13;14)(q10;q10)
ROBT-2	NA	45,XY,der(15;22)(q10;q10)
ROBT-3	45,XX,der(13;14)(q10;q10)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
ROBT-4	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-5	NA	45,XY,der(14;21)(q10;q10)
ROBT-6	46,XX	45,XY,der(14;21)(q10;q10)
ROBT-7	45,XX,der(14;21)(q10;q10)	46,XY
ROBT-8	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-9	46,XX	45,XY,der(14;21)(q10;q10)
ROBT-10	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-11	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-12	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-13	NA	45,XY,der(13;14)(q10;q10)
ROBT-14	45,XX,der(14;22)(q10;q10)	46,XY
ROBT-15	45,XX,der(13;15)(q10;q10)	46,XY
ROBT-16	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-17	45,XX,der(13;21)(q10;q10)	46,XY
ROBT-18	45,XX,der(14;21)(q10;q10)	46,XY
ROBT-19	46,XX	45,XY,der(14;21)(q10;q10)
ROBT-20	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-21	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-22	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-23	46,XX	45,XY,der(14;22)(q10;q10)
ROBT-24	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-25	45,XX,der(14;21)(q10;q10)	46,XY
ROBT-26	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-27	NA	45,XY,der(13;14)(q10;q10)
ROBT-28	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-29	NA	45,XY,der(13;14)(q10;q10)
ROBT-30	45,XX,der(14;15)(q10;q10)	46,XY
ROBT-31	45,XX,der(13;15)(q10;q10)	46,XY
ROBT-32	46,XX	45,XY,der(13;14)(q10;q10)

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	Karyotype-Female	Karyotype-Male
ROBT-33	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-34	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-35	45,XX,der(13;22)(q10;q10)	46,XY
ROBT-36	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-37	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-38	NA	45,XY,der(13;14)(q10;q10)
ROBT-39	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-40	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-41	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-42	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-43	46,XX	45,XY,der(13;21)(q10;q10)
ROBT-44	NA	45,XY,der(13;14)(q10;q10)
ROBT-45	46,XX	45,XY,der(13;15)(q10;q10)
ROBT-46	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-47	NA	45,XY,der(13;15)(q10;q10)
ROBT-48	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-49	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-50	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-51	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-52	45,XX,der(14;21)(q10;q10)	46,XY
ROBT-53	45,XX,der(13;14)(q10;q10)	46,XY
ROBT-54	NA	45,XY,der(13;14)(q10;q10)
ROBT-55	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-56	NA	45,XY,der(13;14)(q10;q10)
ROBT-57	46,XX	45,XY,der(13;14)(q10;q10)
ROBT-58	NA	45,XY,der(13;14)(q10;q10)
ROBT-59	NA	45,XY,der(14;21)(q10;q10)
ROBT-60	46,XX	45,XY,der(13;14)(q10;q10)
INV-1	46,XX,inv(20)(p12q13.1)	46,XY

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	**Karyotype-Female**	**Karyotype-Male**
INV-2	46,XX	46,XY,inv(1)(p31q43)
INV-3	46,XX,inv(4)(p14p16)	46,XY
INV-4	46,XX	46,XY,inv(9)(p22q22.3)
INV-5	46,XX,1qh+,inv(2)(p11.2q14.1)	NA
INV-6	46,XX,9gh+,inv(X)(p11q22)	46,XY,9gh+
INV-7	46,XX	46,XY,inv(12)(p11.2q14.1)
INV-8	46,XX	46,XY,inv(12)(p11.2q13)
INV-9	46,XX,inv(9)(p22q22)	46,XY
INV-10	46,XX	46,XY,inv(2)(p11.2q13)
INV-11	46,XX,inv(20)(p11.2p13)	46,XY
INV-12	46,XX	46,XY,inv(5)(q13q35)
INV-13	46,XX,inv(10)(p13q22.1)	46,XY
INV-14	46,XX,inv(6)(p25.2q25.3)	46,XY
INV-15	46,XX,inv(18)(p11.2q21.1)	46,XY,9hgt
INV-16	NA	46,XY,inv(16)(p11q24)
INV-17	46,XX,inv(9)(p13q21)	46,XY
INV-18	46,XX	46,XY,inv(9)(q13q32)
INV-19	46,XX,inv(7)(p15.1q21.2)	46,XY
INV-20	46,XX,inv(3)(q11.2q26)	46,XY
INV-21	46,XX,inv(21)(p11q21)	46,XY
INV-22	46,XX,inv(9)(p12q21.2)	46,XY
INV-23	46,XX	46,XY,inv(12)(p11.2q13)
INV-24	46,XX	46,XY,inv(12)(p13q21)
INV-25	46,XX	46,XY,inv(18)(q21.1q23)
INV-26	46,XX	46,XY,inv(12)(p11.2q13)
INV-27	46,XX,inv(5)(p15.3q11.2)	46,XY
INV-28	46,XX,inv(6)(p21.3q13)	46,XY
INV-29	46,XX	46,XY,inv(2)(p11q13)
INV-30	46,XX	46,XY,inv(5)(p13.3p15.3)

Table 2.1 – Karyotypes of 300 couples with structural rearrangements (Continued)

Patient code	**Karyotype-Female**	**Karyotype-Male**
INV-31	46,XX,1qh+,inv(9)(p13q21)	NA
CCR-1	46,XX,t(11;19)(p11.2;q12), t(17;19)(q21;p13.1)	46,XY
CCR-2	46,X,t(X;9;1;15) (q26;q34;p36.2;q26.1)	46,XY
CCR-3	NA	46,XY,t(3;6;8)(p12;p11.2;q13)
CCR-4	NA	45,XY,der(13;14)(q10;q10), t(11;22)(q23.3;q12)
CCR-5	46,XX	46,XY,t(4;18)(q33;q12.2), t(20;21)(q13.3;q22.1)
CCR-6*	46,XX,t(5;15)(q13;q26)	46,XY,t(5;15)(q13;q26)
CCR-7	NA	46,XY,t(5;16)(q31;p13.1), t(2;14)(q31;q32)
CCR-8	45,XX,-21(8)/46, XX,r(21)(p13q22.3)(92)	NA
CCR-9	46,XX,ins(19;10) (q13,1;p15p11.2)	NA
CCR-10	NA	46 XY,ins(2;13)(q31;q22q32)
CCR-11	46,X,ins(X)(p21.3;?)	NA

2.2.3 Vitrification and frozen embryo transfer

Blastocysts were vitrified and thawed using Cryotop® method (Kitazato, Japan) when necessary. The frozen embryo transfers were performed after proper preparation of endometrium using a modified natural cycle or with mild ovarian stimulation protocols as described previously [52].

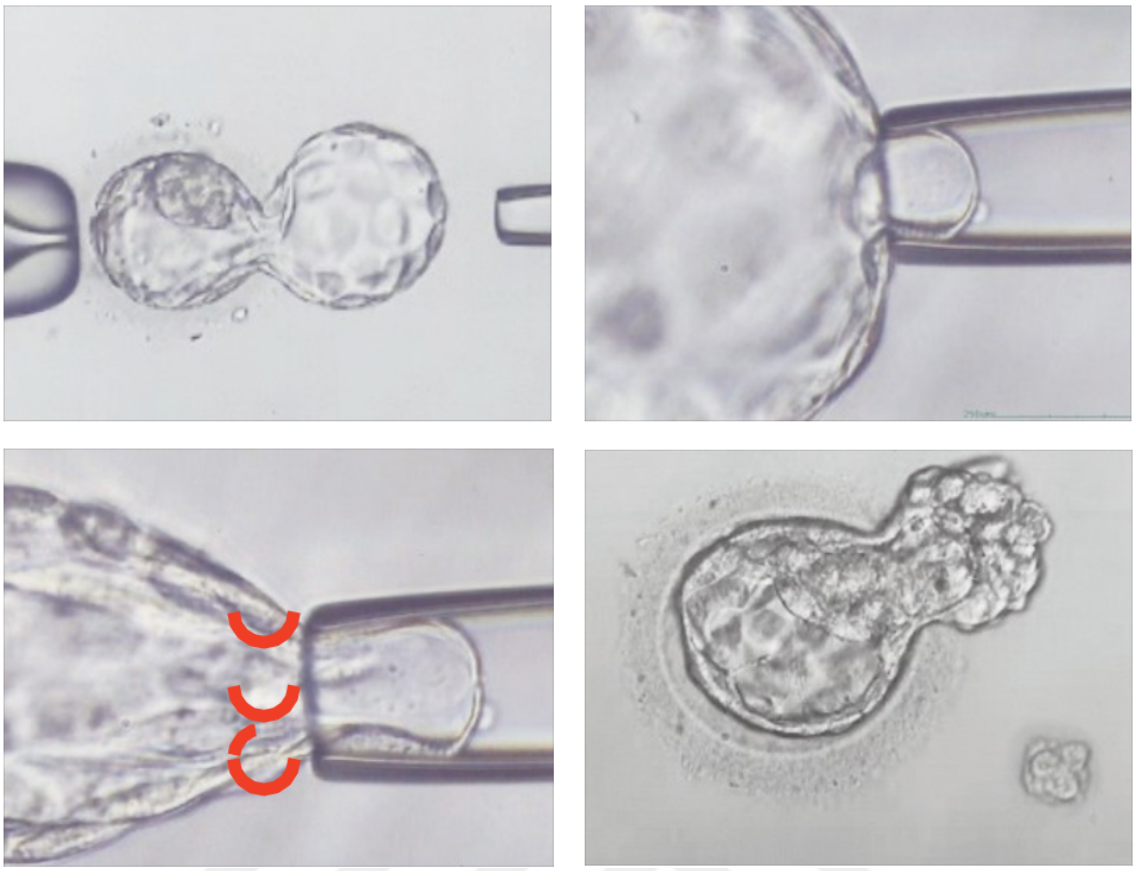


Figure 2.2 Day 5 blastocyst-stage embryo (trophectoderm) biopsy. (a) Expanding blastocyst (b) Aspiration of trophectoderm cells (c) Laser shots, represented by red circles, are applied to break down the tight junctions between trophectoderm cells. (d) Aspirated trophectoderm cells (range: 2–9 cells)
[54]

2.3 Genetic analyses

23 pairs of chromosomes (22 autosomes and XY) were assessed in each embryo either with aCGH or NGS methods. The genetic analysis protocols were outlined previously [52, 55]. Briefly, for aCGH, the samples were tested for chromosomal gain or loss using 24Sure^{TM} Cytochip system (BlueGnome Ltd, Cambridge, UK) and analysed with a commercial software (BlueFuse Multi, Illumina). The NGS procedure was carried out using the Reproseq kit using Ion Chef (Thermo Fisher Scientific, USA) and Ion GeneStudio S5 (Thermo Fisher Scientific, USA) for library preparation and sequencing. Analysis was performed using commercial software (Ion Reporter v5.6, Thermo Fisher Scientific, USA). The mosaicism threshold was set as 30%. Mosaic embryos were grouped under the category of aneuploid embryos for simplicity. The wet-lab and analysis procedures were performed in Genetics Diagnosis Centre in Istanbul Memorial Hospital. From 2011-2016 aCGH technique, from 2016-2019 NGS technique were used respectively.

2.3.1 Array Comperative Genomic Hybridization (aCGH) protocol

24sure V3 assay kit (Bluegnome, Illumina) was used for array CGH procedure (Figure 2.3). 24sure technology is based on a microarray system to detect copy number imbalance. The arrays consist of BACs spotted on a glass surface.

Figure 2.3 24sure Workflow Diagram. The whole protocol lasts for 12 hours however, can be extended to fit the procedure according to the working day [56]

2.3.1.1 Whole Genome DNA Amplification (WGA)

SurePlex Single Cell WGA Kit (BlueGnome) was used in this step. Preparation of amplification mixes were carried out in a sterile containment cabinet (vertical laminar flow) that prevents sample contamination.

Cell extraction

1. Individual tubes containing biopsy samples were centrifuged at high speed (14.000rpm)
2. 2.5 ul of cell extraction buffer (green cap) was added to each sample and negative and positive controls.
3. Extraction cocktail master mix was prepared by adding 4.8 ul from cell extraction enzyme dilution buffer and 0.2 ul cell extraction enzyme for each sample as below (Table 2.2):

Table 2.2 Extraction mix preparation

Extraction Cocktail	**Volume per single sample**	**Volume per 5 samples**
Extraction enzyme dilution buffer	4.8 µl	24 µl
Cell extraction enzyme	0.2 µl	1 µl
Total Volume	5 µl	25 µl

4. 5 ul of freshly prepared Extraction cocktail was added to the tubes containing samples.
5. Samples were centrifuged briefly to get all contents to the bottom of the tube.
6. Samples were incubated in PCR machine as follows (Table 2.3):

Table 2.3 PCR program for extraction

1 cycle	75°C	10 min
1 cycle	95°C	4 min
1 cycle	room temperature	Hold

Pre-amplification

1. Components were combined as below to prepare pre-amplification mix (Table 2.4).

Table 2.4 Pre-amp mix preparation

Pre-amp cocktail	**Volume per single sample**	**Volume per 5 samples**
SurePlex pre-amp buffer	4.8 μl	24 μl
SurePlex pre-amp enzyme	0.2 μl	1 μl
Total Volume	5 μl	25 μl

2. 5 μl of Pre-amp cocktail was added to each 10 μl of sample from previous step.
3. Samples were incubated according to thermal cycler program in the protocol.

Amplification step

1. Components were combined as below to prepare amplification mix (Table 2.5).

Table 2.5 Amplification mix preparation

Amplification cocktail	**Volume per single sample**	**Volume per 5 samples**
SurePlex amplification buffer	25 μl	125 μl
SurePlex amplification enzyme	0.8 μl	4 μl
Nuclease-free water	34.2 μl	171 μl
Total Volume	60 μl	300 μl

2. 60 μl of the freshly prepared amplification mix were added to the samples. Tubes were centrifuged briefly.
3. Samples were amplified according to thermal cycler program below (Table 2.6).
4. In order to determine the amplification efficiency, 5 μl of each amplified sample were tested in agarose gel (1.5%) electrophoresis.

Table 2.6 PCR program for amplification

1 cycle	95°C	2 min
14 cycles	95°C	15 sec
	65°C	1 min
	75°C	1 min

2.3.1.2 Labelling

Amplified sample and reference DNAs were labelled with Cy3 and Cy5 fluorophores, using random primers (Table 2.7).

1. Labelling master mixes were prepared by adding the components shown below:

Table 2.7 Labelling mix preparation

Component	Cy3 labelling mix - 1 rxn	Cy5 labelling mix - 1 rxn
reaction buffer	5 μl	5 μl
dCTP-labelling mix	5 μl	5 μl
Cy3 dCTP	1 μl	
Cy5 dCTP		1 μl
Klenow enzyme	1 μl	1 μl
Total	12 μl	12 μl

2. 5 μl of primer solution was combined with 8 μl of amplified sample DNA or reference DNA.

3. The mix was denatured in a thermal cycler for 5 minutes at 94 °C then was transferred to ice for 5 minutes.

4. 12 μl of Cy3 labelling master mix was added to the 13 μl of the sample DNA/primer solution and SureRef DNA/primer solution 12 μl of Cy5 labelling master mix was added to the other 13 μl of sample DNA/primer solution and SureRef DNA/primer solution to label them differently.

5. Tubes were incubated in a prewarmed lidded thermal cycler for 2-4 hours at 37 °C. This incubation may be increased up to 18 hours, if required.

2.3.1.3 Combination

1. In this step Cy3 and Cy5 labelled samples or references were combined together with COT Human DNA and reduced in volume with centrifugal evaporation. For this, centrifugal evaporator (LabCongo), was prewarmed to 75 °C (or high) for 30 minutes.

2. Cy3 and Cy5 labelled DNA were combined together by adding the Cy5 labelling product to the Cy3 labelling product for each hybridization area.

3. 25 μl COT Human DNA was added to each tube or PCR plate well containing combined Cy3/Cy5 labelling products.

4. The tubes were transferred to the prewarmed centrifugal evaporator with lids open. The solutions in the tubes were evaporated under centrifuge at 75 °C (or high).

2.3.1.4 Hybridization

In this step labelled DNA was resuspended in dextran sulphate hybridization buffer. Hybridization was performed using a water bath.

1. Pellets of combined labelled samples/references/COT were dissolved in 21 ul of prewarmed dextran sulphate (DS) hybridization buffer at 75 °C ensuring that pellet was completely dissolved.

2. They were denatured at 75 °C for a further 10 minutes.

3. 18 μl of labelled DNA solution were applied to each position.

4. The hybridization template was used to position cover slips and confirm labelled DNAs are loaded on to correct hybridization areas.

5. The hybridization unit was placed into the water tank for 3 to 16 hours at 47 °C.

2.3.1.5 Washing

This step was performed to remove DNA which was not hybridized.

1. Wash I solution was prepared by adding 400 ml of 2xSSC/0.05% Tween20 at room temperature.

2. The cover slips were removed from each slide by manually agitating in 2xSSC/0.05% Tween20 in a Coplin jar at room temperature.

3. The slides were dried by centrifugation at 170 xg for 3 minutes and stored in original blue box.

2.3.1.6 Scanning

In order to detect the fluorescent signals of each dot and record the resulting images of the hybridization, a laser scanner was used. The resulting images were saved in TIFF format file. This format was read by BlueFuse Multi analysis software. Green (532 nm) and red laser (635 nm) were used to excite and read Cy3 and Cy5 signals respectively.

1. The power of the laser and gain of the photomultiplier tube (PMT) were adjusted during the scan with values between 50-60% or higher when the signal intensity was low (Figure 2.4).

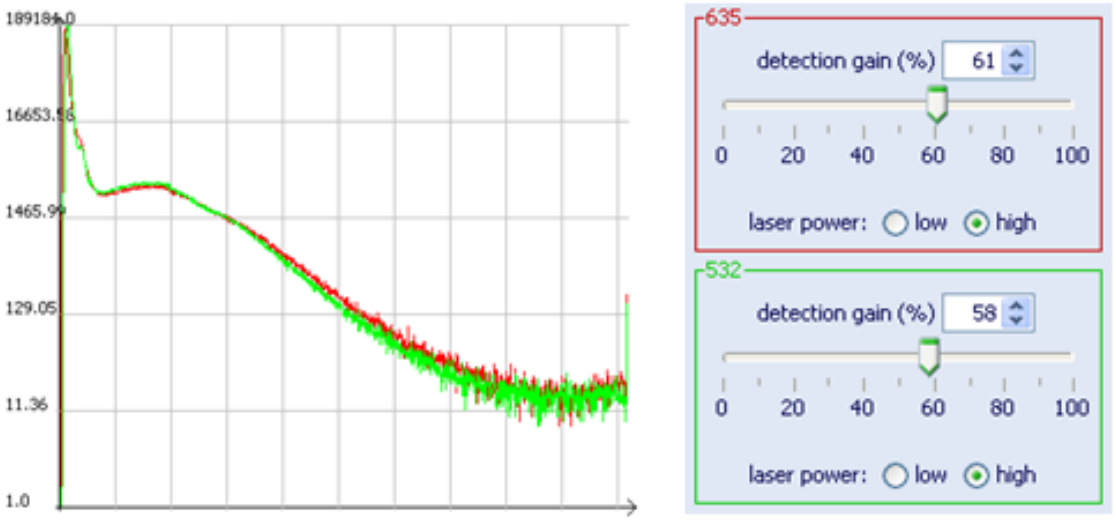


Figure 2.4 Example of a pre-scanning procedure [56]

2. The slides were scanned at 10 nm resolution (Figure 2.5)
3. The figures were saved in TIFF format.
4. Data Analysis were done by using BlueFuse Multi software (Figure 2.6) (Figure 2.7).

2.3.2 Next Generation Sequencing (NGS) protocol

Ion Torrent next-generation sequencing (NGS) technology was used. Briefly, the protocol consisted of whole genome DNA amplification, purification quantification, library preparation, sequencing and analysis [57].

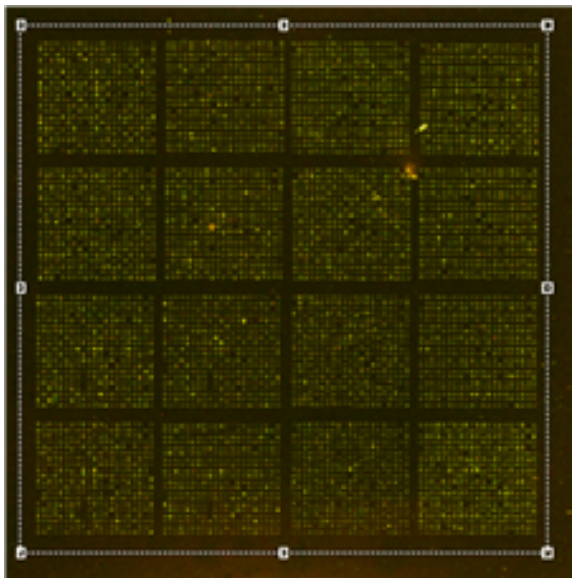


Figure 2.5 The appearance of the array after pre-scanning procedure [56]

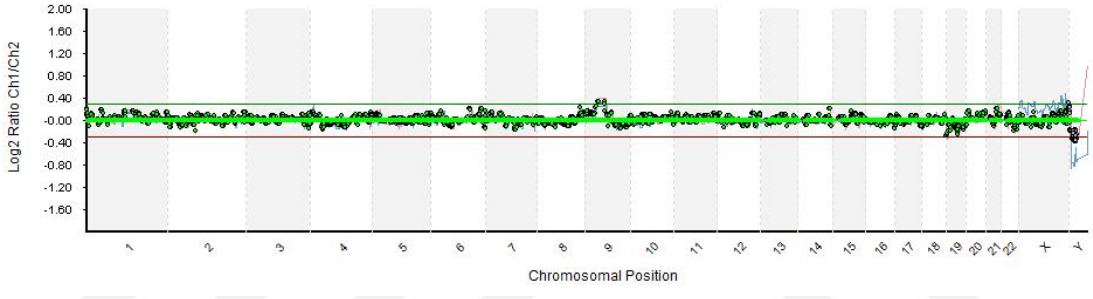


Figure 2.6 aCGH profile of a normal/balanced embryo

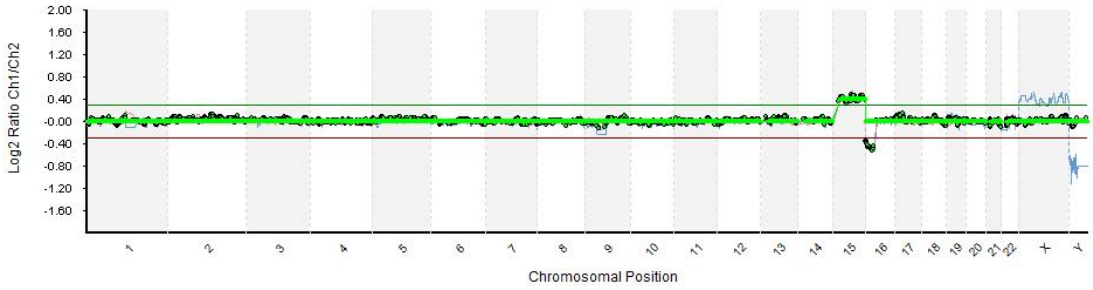


Figure 2.7 aCGH profile of a normal/balanced embryo

2.3.2.1 Whole Genome DNA Amplification

Extraction step

1. Cell Extraction Buffer (green cap) was added to each sample with 2.5 μl 1X PBS well to bring the total volume to 5 μL.

2. Extraction Enzyme master mix was prepared in a 1.5-mL tube by adding 4.8 μL of Extraction Enzyme Dilution Buffer (violet cap) and 0.2 μl Cell Extraction

Enzyme.

3. 5 μL Extraction Enzyme master mix was added to each tube.
4. The tubes were centrifuged at 1,000 $\times$ g for 30 seconds to collect liquid at the bottom of the wells.
5. The samples were incubated in a thermal cycler using the following temperature program (Table 2.8).

Table 2.8 Extraction program

Temperature	Time
75°C	10 minutes
95°C	4 minutes
22°C	Hold

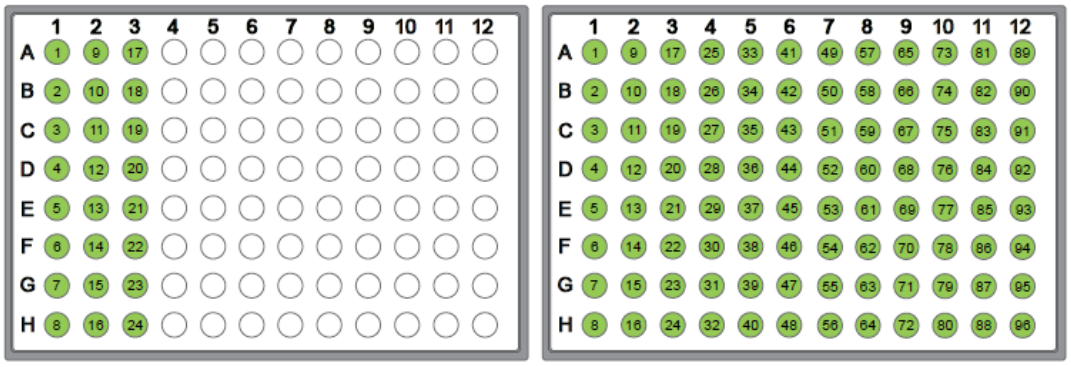
Pre-amplification step

1. Pre-amplification master mix was prepared in a 1.5-ml tube on ice by adding 4.8 μL of Pre-Amplification Buffer (red cap) and 0.2 μl (white cap) of Pre-Amplification Enzyme.
2. 5-μl of Pre-amplification master mix were added to each sample well (15-μl final volume).
3. The samples were incubated in a thermal cycler using the following temperature program (Table 2.9).

Table 2.9 Pre-amplification PCR program for NGS

Step	Temperature	Time	Number of cycles
1	95°C	2 minutes	1
	95°C	15 seconds	
	15°C	50 seconds	
2	25°C	40 seconds	12
	35°C	30 seconds	
	65°C	40 seconds	
	75°C	40 seconds	
3	4°C	Hold	1

plate handling guidelines.



Ion SingleSeq™ Barcode Plates containing Barcodes 1–24 and 1–96. Barcode Adapters are loaded in the indicated wells.

Figure 2.8 Barcode Plates contained either 24 or 96 according to the sample size and the plan of the experiment. 20 μl Barcode adapters were present in each well [57]

Amplification step

1. Ion SingleSeq™ Barcodes 1-24 or 1-96 plate were thawed.
2. Amplification mix was prepared by adding 27 μl Buffer (orange cap), (blue cap) 0.5 μl Amplification Enzyme and 2.5 μl of nuclease free water. 30 μl of amplification mix was added to each sample.
3. 5 μl of SingleSeq™ Barcode Adapter, from the Barcode Plate was added to corresponding samples (Figure 2.8).
4. The samples were incubated in a PCR machine with the following temperature program (Table 2.10).

Table 2.10 PCR program for NGS amplification

Step	Temperature	Time	Number of cycles
1	95°C	3 minutes	1
2	95°C 50°C 72°C	20 seconds 25 seconds 40 seconds	4
3	4°C	Hold	1

2.3.2.2 Pooling and purification of libraries

1. 5 μl of each library was added to a new 1.5-mL tube

2. The mix was vortexed and pulse-centrifuged to collect contents at the bottom of the tube.

3. 40 μl of the library pool was transferred to a 0.2-μl tube.

4. 40-μl of library pool was heated using the following program (Table 2.11).

Table 2.11 Pre-beads heating program

Step	Temperature	Time	Number of cycles
1	70°C	2 minutes	1
2	22°C	Hold	1

5. The heated library pool was transferred to a new 1.5-ml Eppendorf DNALoBind™ tube.

6. 40 μl (1X volume) of AMPure™ XP beads were added.

7. Incubated for 5 minutes at room temperature.

8. The tube was placed in the DynaMag™-2 magnet, for 5 minutes for beads to aggregate to the side of the tube.

9. The supernatant was aspirated carefully, then discarded.

10. Beads were washed with 250 μl of freshly prepared 70% ethanol while the tube is still on the magnet.

11. Incubated for 30 seconds.

12. The supernatant was aspirated, then discarded.

13. This wash step was repeated once more.

14. The beads were allowed to dry at room temperature for 3–4 minutes with the tube on the magnet.

15. The beads were resuspended in 40 μl of Low TE by pipetting up and down.

16. The tube was incubated at room temperature for 1 minute.

17. The tube was placed in the DynaMag™-2 magnet, then wait 2–3 minutes for beads to aggregate to the side of the tube.

18. 35 μl of the supernatant was transferred to a new 1.5-ml Eppendorf DNA LoBind™ tube.

2.3.2.3 Quantification of the library pool

1. High Sensitivity (HS) Assay Kit (Qubit dsDNA HS Assay) was used to quantify the libraries according to manufacturer's recommendations [57].
2. Qubit™ working solution was prepared by mixing the Qubit™ dsDNA HS Reagent and Qubit™ dsDNA HS Buffer 1:200 ratio.
3. For the preparation of standards 190 μl Qubit™ working solution was added to two labeled Qubit™ Assay Tubes used for standards.
4. 10 μl of each Qubit™ standard (Components C and D) were added to the appropriate tube.
5. 198 μl Qubit™ working solution was aliquoted to labeled Qubit™ Assay Tubes used for samples.
6. 2 μl of the library pool was added to the appropriate sample tube and mixed.
7. The tubes were incubated in the dark for 2 minutes.
8. First in order to generate a standard curve, the two standards were measured and recorded. When the standards are in the expected range, the pool was measured.
9. According to the concentration, the library pool was diluted to 1 nM.

2.3.2.4 Library preparation

1. Run plan was created in the Torrent Server in Torrent Suite™ Software (Figure 2.9).
2. First, with clicking the Plan tab, selecting Ion ReproSeq Aneuploidy - Ion S5 System from the list under Template Name.
3. A new Run Plan Name was added.
4. hg19(Homo sapiens) is selected from the Reference Library dropdown list.
5. Suitable settings for Target Regions and Hotspot Regions were selected as "None" from the dropdown lists.
6. The number of barcodes were entered Sample Set. Ion SingleSeq Barcode Set 1-96 (default) or Ion SingleSeq Barcode Set 1-24.
7. Flows were set to 250 flows. For details Ion ReproSeq™ PGS Kits - Ion S5™/Ion GeneStudio™ S5 Systems User Guide [57].

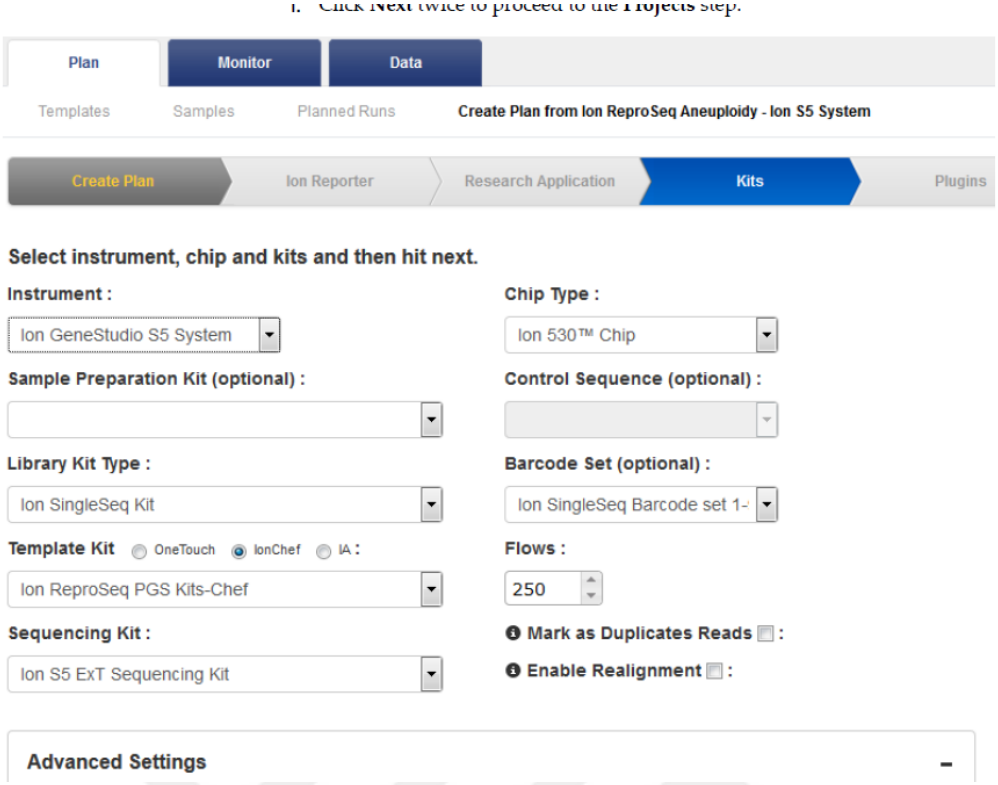


Figure 2.9 Creating the RUN plan using Ion Torrent Web Site [57]

8. The consumables in Ion Chef™ System were loaded according to manufacturer's recommendations (Figure 2.10).

9. 4 μl of the 1 nM pooled library was mixed with 46 μl Nuclease-free Water and was put into the Library Sample Tube (barcoded tube).

10. Ion Chef™ run was started by pressing Set up run.

11. Library and template preparation, including clonal amplification of final library pools were performed by Ion Chef System in a fully automated way.

12. The Run took minimum 4 hours 15 minutes and could be extended to 16 hours according to the daily workflow.

2.3.2.5 Sequencing and Analysis

1. Before Ion Chef completes, the sequencer was initialized and got ready for sequencing.

2. Initialization took 50 minutes.

3. For this, the Ion S5™ ExT Sequencing Reagents, the Ion S5™ ExT Wash Solution bottle and Ion S5™ Cleaning Solution bottles were unpacked and installed on the Ion S5™ sequencing instrument according to User Guide [57].

Figure 2.10 The Ion Chef and S5 systems

4. After initialization was completed, the loaded chip from Ion Chef was placed inside the S5 sequencer and then pressed Start run to begin the sequencing run.

5. Ion Torrent uses highly efficient DNA polymerase that readily incorporates the nucleotides and uses semi conductor sequencing technique where optical detection system is not required.

6. It took nearly 1 hours to finish the sequencing.

7. For sequencing, semiconductor sequencing method was used which was based on the release and the detection of hydrogen ions throughout the polymerization process (Figure 2.11).

   Briefly, the chemistry used was as follows; the incorporation of a deoxyribonucleoside triphosphate (dNTP) into a growing DNA strand results in the formation of a covalent bond and the release of pyrophosphate and a positively charged hydrogen ion which is specific for each nucleotide. There were many copies of one single-stranded template DNA molecule in each microwell. Each microwell is placed on a semiconductor chip that each contain are sequentially flooded with unmodified A, C, G or T dNTP. The addition of every new nucleotide leads to a voltage change if it is a counterpart of the growing chain which is detected by a silicon pH sensor.

8. The excess dNTP molecules were washed out before the next cycle with different dNTP species was introduced.

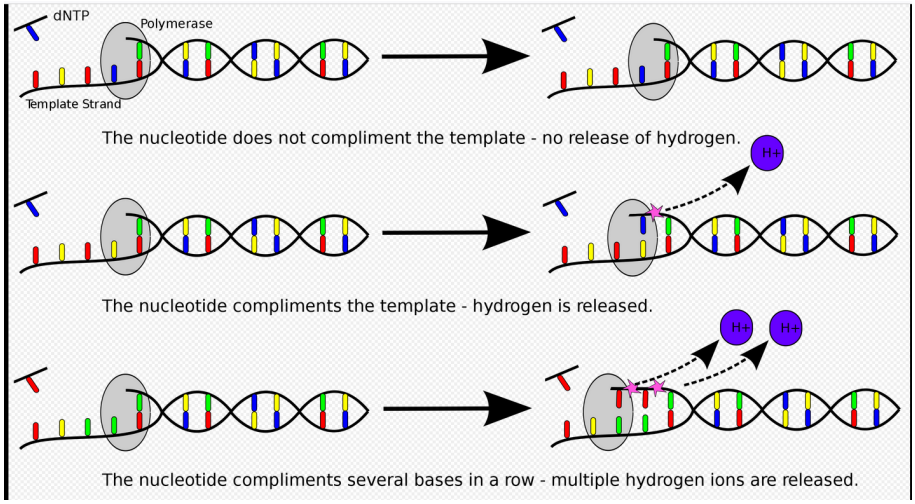


Figure 2.11 Sequencing by synthesis. Hydrogen ions are released during polymerization process [58]

9. There is a voltage sensitive sensor called semiconductor sensor beneath the layer of microwells. Each nucleotide incorporation events are measured directly by this highly specific voltage detection system (Figure 2.12).

10. Signal processing and DNA assembly were carried out in Ion Reporter™ Software. The samples were analyzed in Ion Reporter Software which has data analysis tools that support the format of the output of data from Ion S5 [57]. (Figure 2.13)

11. The profiles of all embryos were analyzed according to the previously established workflow which was validated for identification of small copy number variations (CNV) of 5-10 Mb. Whole chromosomal aneuploidies related to 24 chromosomes and segmental aneuploidies were analyzed by at least two diagnostic personnel (Figure 2.14).

12. All analysis data and the reports were generated through Aura system and the results specific to each embryo and each chromosome were saved in the database.

2.4 Outcome measures

The incidence of chromosomal abnormality per chromosome was measured by dividing the total number of errors observed for that chromosome by the total number of embryos analyzed. The total aneuploidy rate was measured by dividing the total number of aneuploidy by the total number of chromosomal pairs analyzed. Chromosome X and Y were evaluated together as a chromosomal pair. In the

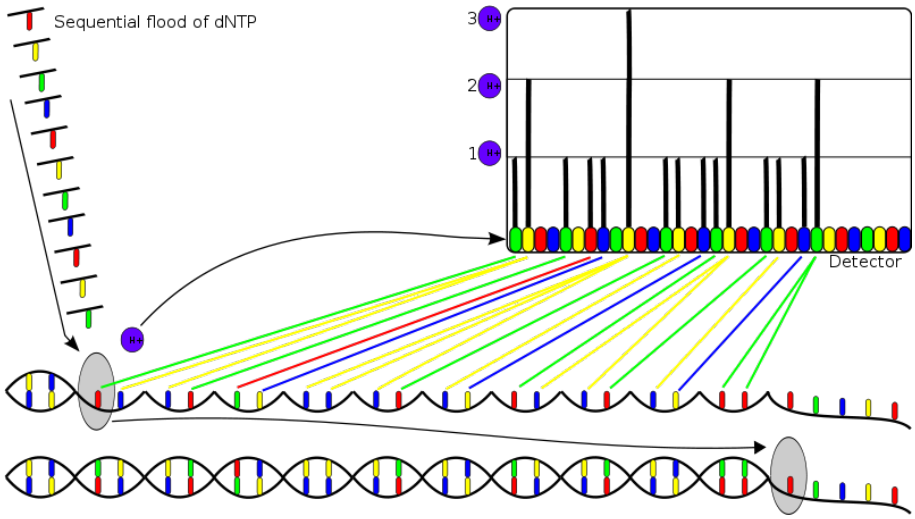


Figure 2.12 Voltage change is detected by the detectors located on semiconductor surface.During synthesis, each hydrogen ion triggers a voltage change which is specific to each base incorporation to the newly synthesized chain [58]

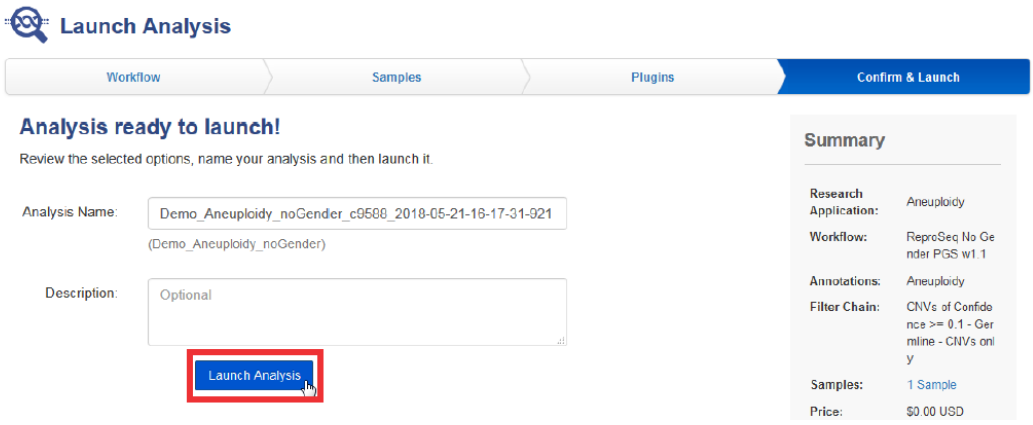


Figure 2.13 Analysis took place using Ion Reporter software [57]

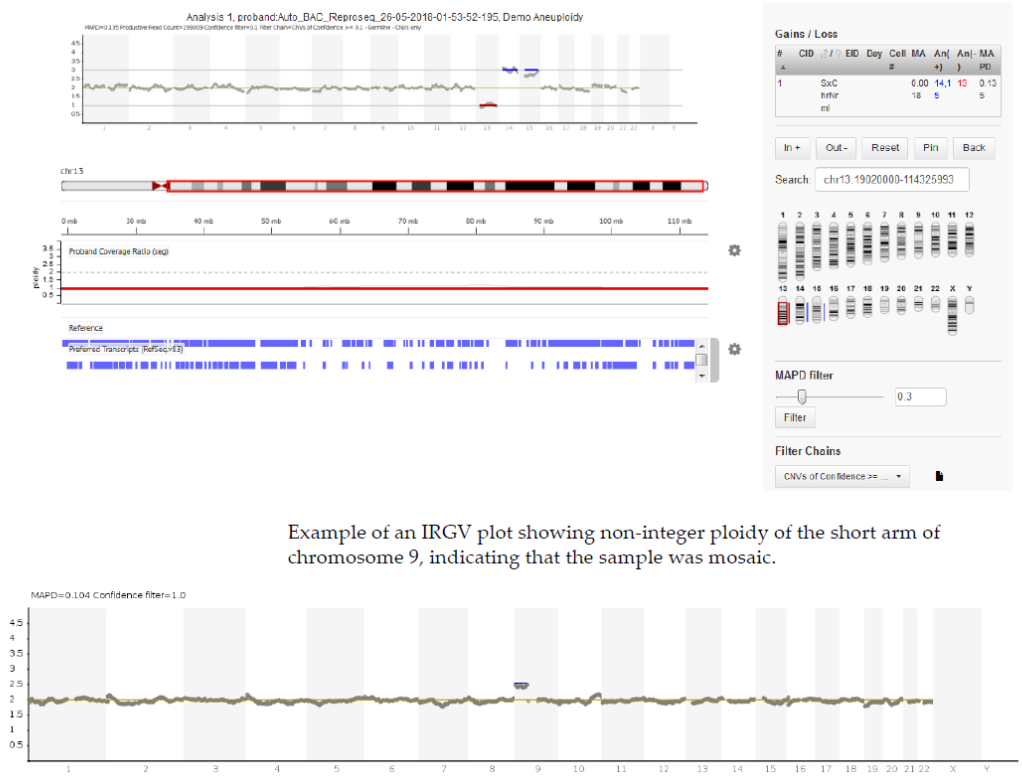


Figure 2.14 The gains and losses of each chromosomes and segments are shown in graphs and a merged profile [57]

calculation of segregational abnormalities, the chaotic (abnormalities related with more than 5 chromosomes) embryos were excluded since the underlying reason would not be clear; either segregational or a global abnormality. However, chaotic embryos were included in the calculations of de novo aneuploidy. The chromosomes involved in the rearrangements were excluded from the analysis of ICE. Abnormalities related to other chromosomes (segmental or whole chromosome aneuploidy) were grouped into "*de novo* aneuploidy" for simplification. Each embryo was grouped under five groups based on both their segregational status and their euploidy status as following: (a)balanced/euploid, (b)unbalanced/euploid, (c)balanced/aneuploid, (d)unbalanced/aneuploid, and (e)chaotic. Cumulative *de novo* aneuploidy rate was calculated with the total number of aneuploid embryos divided by the total number of embryos diagnosed (c+d+e/a+b+c+d+e), and cumulative unbalanced embryo rate was calculated by the total number of segregationally unbalanced embryos divided by total number of embryos diagnosed (b+d/a+b+c+d). International glossary of infertility and fertility care was used as a guidance for the calculation of the clinical outcome measures such as clinical pregnancy rate (CPR), implantation rate (IR) and live birth rate (LBR) [59]. According to that, CPR was defined as the percentage of clinical pregnancies with sac/foetal heartbeat divided by the number of cycles with embryo transfer. Implantation rate was defined as the number of gestational sacs

divided by the total number of embryos transferred. Cumulative CPR was defined as the total number of clinical pregnancies divided by the number of patients with transfer. Cumulative LBR was calculated by dividing the number of patients who have a healthy live birth by the number of patients with transfer. The singleton, twin, or other multiples was counted as one delivery [59]. Advanced maternal age group was defined by women aged 35 and above. This cut-off was selected in the light of previous studies [13, 60].

Another outcome measure related with prediction models were accuracy, ROC-AUC, sensitivity and specificity.

2.5 Data cleaning

All patient and treatment specific data including demographical, biochemical, genetical and clinical parameters were registered and stored in Aura® - a laboratory information management system (LIMS) which was specially designed for IVF and Reproductive Genetics Centre in Istanbul Memorial Hospital. The data was created with the use of the team specialized healthcare providers such as nurses, biologists, embryologists, clinicans. With special permission of the head of the clinic and with permission of the ethical committee of the hospital the database was requested as to only include patients who has undergone PGT between 2011-2019 years period. The database was received in excel format. The raw data was tremendously detailed with more than 200 clinical parameters. Each row represented the specific data of the individual oocytes and their corresponding embryos throughout the whole treatment such as embryo developmental parameters and the information about being transferred. Handling a data with this size was very challenging, even the computer memories were not sufficient to work. The data first should be cleaned and simplified for analysis. First; columns with a lot more than necessary information was deleted. Although most of the information was registered by automatically or using scroll down menu, some of the columns such as karyotype information was filled necessarily by handwriting which resulted in heterogeneous data entries by different users. The information in those columns were standardized by checking each cell and corrected by handwriting.

After this preliminary cleaning, R-program with different work packages which consisted of data cleaning, simplification, correction, merging of different data, standardization was used for preparation of the data before any statistical analysis on R software and SPSS took place. The data analysis involved using excel, R software and SPSS. The detailed steps could be found below:

2.5.1 Data cleaning in R software

1. There were 2 databases. First database included the data of fresh treatment outcomes and the second was related with frozen transfer outcomes.

2. There were 420 columns in each database including the detailed information of patients and their corresponding embryos. In order to simplify and prepare the data for analysis 250 columns and related information were deleted.

3. In the data, the patient specific clinical informations such as female age and BMI appeared only for the first oocyte generated on that cycle. For example, for 2593 cases with 22187 rows including all embryonic information, the case specific clinic information such as BMI, female age, karyotype, duration of infertility were present only in 2593 rows but missing in 22187-2593=19594. Since the aim of this study was to compare the clinical and genetic outcomes per embryo, this clinical information should be present for each embryo analyzed. Therefore the data was transferred into R software to standardize and fill the empty rows with corresponding information.

4. Excel data was uploaded in R software and processed with Tidyverse package with the following commands (Figure 2.15).

```
library(tidyverse)
library(writexl)
library(readxl)
excel_data <- read_excel("Cagri_Aneuploidy_28_01_2018_
   18.12.2018.xlsx", 1, guess_max = 22000)
dim(excel_data) ## [1] 22186 234.

# The column names to be filled with data were defined.
columns_to_fill <- c("Yaş (Kadın)" ,"Gercœk Kadın Yaşı" ,"BMI
   ", "İNF. TİPİ","İNF. SÜRESİ (YIL)", "DENEME SAYISI (TOTAL)
   ","KARYOTİP K", "KARYOTİP E", "ACGH ENDIKASYON", "SADE
   ENDIKASYON","MAIN INDICATION", "RECT-ROBT-INV","CARRIER-
   MALE-FEMALE","SUB GROUP")

columns_to_remove <- c("HSG","HSG YER","LAPAROTOMİ","
   LAPAROTOMİ YER","PRL","TSH")

```

Figure 2.15 Importing Excel data into R and processing

5. There was a situation that should be noted while filling in the selected columns. In some columns it was not possible to fill in simply by moving the value in the cell above to the empty cells below it. Because, in some columns, the gaps covered several groups. For example, when there is only one piece of information in a group, it is simple to move that information to sub-cells. But

since all the cells in the next group are empty, moving the information in the previous group will be an incorrect filling.

6. Therefore, in the following command, group_by command is used which helped to group according to the CYCLE ID column and the selected columns are filled only within the group with the fill function (Figure 2.16).

```
excel_data %>%
  select(-one_of(columns_to_remove)) %>%
  group_by(‘SIKLUS ID‘) %>%
  fill(one_of(columns_to_fill)) %>%
  ungroup() -> excel_data_filled

write_xlsx(excel_data_filled,
            path = "excel_data_filled_2018_12_19.xlsx")
```

Figure 2.16 Removing columns and filling rows with R

7. The rows were filled with the last patient specific information was completed.

8. The data was transferred back to excel format again.

2.5.2 Data cleaning in Excel

1. The karyotypes were checked and corrected when necessary according to 2016 International System for Cytogenetic Nomenclature (ISCN, 2016) guidelines [61]. This was performed manually by checking patient specific reports.

2. The data was filtered for structural rearrangement carriers.

3. Further deletion of unnecessary columns was done for simplification and the ease of the data analysis.

4. The Aura program was used used by 50 people and there were some parts which should be added by typing (not scroll-down menu). This has created non-standardized parameters and a lot of typos. Therefore, the typos were corrected.

5. There were missing information such as chromosomal results of some embryos and clinical outcomes due to lost to follow up. These were completed by checking the patient hard copy files and digital files requesting the information from archiving service in the hospital by permission if necessary.

6. Fresh and frozen data were merged in excel.

7. 3 excel files were created for rearrangement cases in detail, including clinical outcome data cycle number and the outcomes (1), genetic analysis results of each embryo as a embryonic diagnosis and grouped under balanced-unbalanced, *de novo* aneuploid, or euploid e.g. (2), another detailed genetic results such that it includes the specific chromosomal result of each embryo whether it is related with rearrangement or not (ICE) and inversion carriers so as to assess any effects of the length of the inversion on the incidence of unbalanced segregations. New columns, tables, graphs were prepared when needed.

8. In order to create those tables careful checking was needed in order to discriminate them into categories according to the structural rearrangement. A total of 1819 embryos and 1819x23= 41000 chromosomes were analyzed and grouped under categories such that it falls one of them; normal/balanced (having normal copy number for all chromosomes), unbalanced (abnormal copy number for the rearranged chromosomes), *de novo* aneuploid (abnormal copy number for non-rearranged chromosomes), unbalanced+*de novo* aneuploid (abnormal copy number for both rearranged and non-rearranged chromosomes).

   The data was cleaned and standardized mostly manually. After that, control group was established by selecting from the same database without any karyotypic abnormalities.

2.6 Selection of Control Group

The control cases were selected from the same database. The majority consists of age-matched patients undergoing PGT for aneuploidy testing (PGT-A) (96%) and the rest was chosen from patient undergoing PGT for monogenic disorders (PGT-M) (4%) in the similar time period. Patients with karyotypic abnormalities or polymorphisms were excluded. In order to reduce possible bias, 2 controls were selected for each case [62]. Control subjects were selected as to match study cases also in terms of the number of embryos analyzed (in the range of ±1) and the number of oocytes retrieved where available (in the range of ±2-5). When there are more than 2 controls (which was the case most of the time) qualifying those parameters, a computer based program was used to select from them randomly [63]. During the selection of control group was performed blindly without the genetic results of the embryos of the patients.

Strict criteria for matching

1. Female age should be exactly the same with cases.

2. Embryo numbers analyzed should be ±1 range with the cases.
3. Sperm source matched with the case either ejaculate or testicular.
4. Genetic analysis technique aCGH or NGS matched with the cases.
5. Infertility type was grouped into 3 categories (Primary, secondary or fertile) and was matched with the same category with cases as possible.
6. Collected oocytes were in the range of ±5 with the cases.
7. Body mass index (BMI) were grouped into 4 categories <18.5, 18.5-24.9, 25-29.9 and >30. The cases were matched with the same category as possible.
8. Anti Mullerian Hormone (AMH) were grouped into 3 categories; <1, 1-3.9 and >4 and were matched with the same category as possible with the cases.
9. Sperm count was grouped into 4 categories; <1 million/ml, 1-5 milllion, 5-20 million, >20 million.
10. The cases with karyotype abnormalities were excluded. The "variances" and the "means" of controls should not be statistically different than the cases to be compared as a good control group. In order to test this goodness of fit tests for each parameter has been performed using SPSS.

    Below an example of the SPSS analysis results of the distribution of female age, which was found to be the same across categories (cases and their corresponding controls) (Figure 2.17, Figure 2.18).

Group Statistics

	CATEGORY (INV/CNT)	N	Mean	Std. Deviation	Std. Error Mean
FEMALE AGE	INV	173	31,95	4,568	,347
	CNT	323	32,26	4,640	,258

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of FEMALE AGE is the same across categories of CATEGORY (INV/CNT).	Independent-Samples Mann-Whitney U Test	,555	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Figure 2.17 The distribution of female age of embryos belonging cases and controls

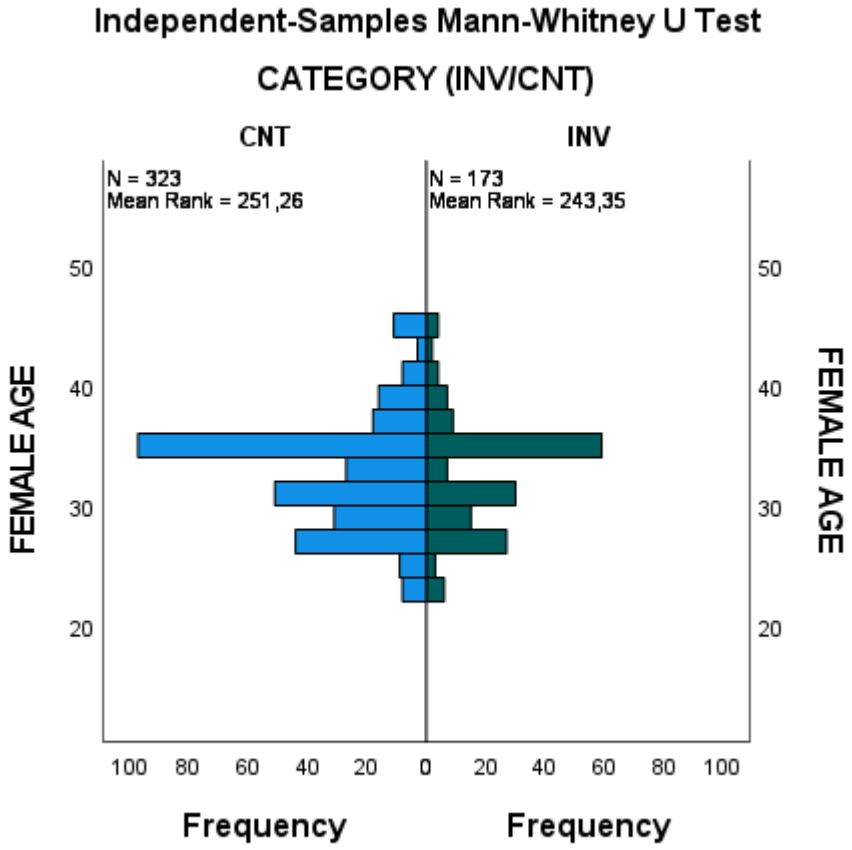


Figure 2.18 Group statistics and the histogram of female age across cases and controls

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,793[a]	1	,373		
Continuity Correction[b]	,622	1	,430		
Likelihood Ratio	,799	1	,371		
Fisher's Exact Test				,415	,216
N of Valid Cases	496				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 53,36.

b. Computed only for a 2x2 table

Figure 2.19 Group statistics of the analysis methods used across cases and controls

Below an example of the SPSS analysis results of the proportion of the techniques used across groups which was found not to be different between cases and the controls (Figure 2.19, Figure 2.20).

2.7 Statistical analysis

Shapiro Wilk test and Kolmogorov-Smirnov tests were used for the test of normality. Nonparametric tests and the descriptive statistics were used for the variables that do not have a normal distribution. For that reason, they were reported as "medians (min-max)". Mann-Whitney U or Kruskal Wallis test were used for comparisons between groups. Dunn-Bonferroni test was used for post-hoc comparisons. Fisher's exact test (for small sample size) or Fisher-Freeman-Halton test (for contingency tables larger than 2x2) were used for categorical comparisons. The type of structural abnormality (RECT, ROBT, INV), female age (young or advanced) and the sex of the carrier (either male or female) were defined as risk factors and they were assessed by using binary logistic regression analysis with two models. In the first model, segregation abnormalities were defined with two different outcomes (cumulative balanced or unbalanced) and in the second model, the outcome was categorized as either cumulative euploid or *de novo* aneuploid respectively. Statistical analyses were performed using SPSS v22.0 (IBM) and R v3.6.3 program. Significance level of α=0.05 was used in statistical analyses. When the sample size is extremely large, a statistical test (seeking for a difference between parametric or non parametric test results) always demonstrates a positive result such as a significant difference in terms of p values (indicating lower values of 0.05); although there is no practical nor clinical importance [64]. For accurate statistical interpretation, additional measurements of effect size are needed to test the practical importance of the effect [65–67]. For

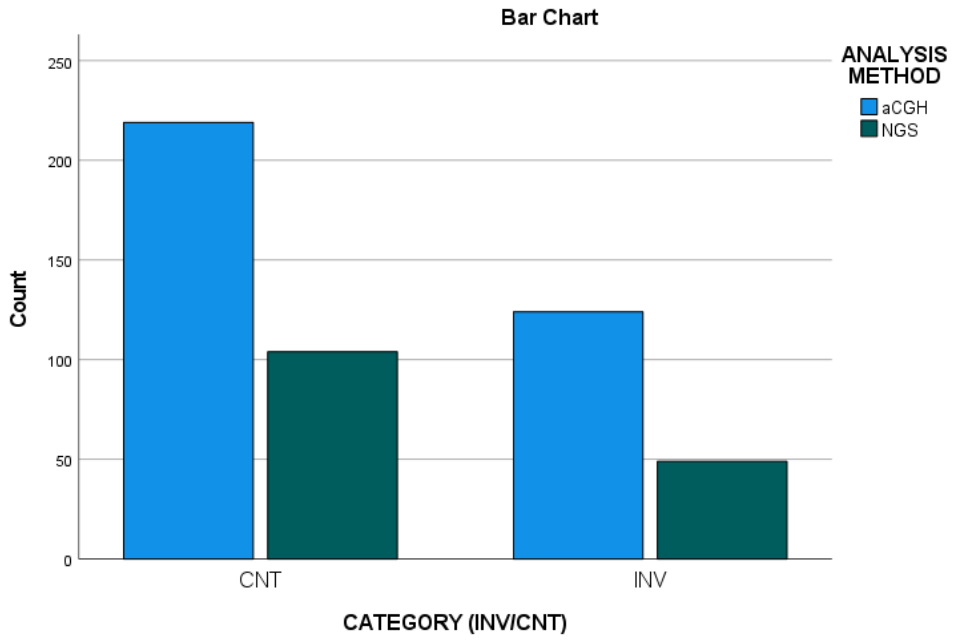


Figure 2.20 The bar chart of the analysis methods used across cases and controls

that reason, in this study, for non-parametric tests, Cliff's Delta effect size (Δ) was measured and was interpreted as small, medium, or large according to the following criteria: <0.147 (Negligible), <0.330 (Small), <0.474 (Medium), $\geq$ 0.474 (Large) [68]. For categorical variables, phi effect size (ϕ) was measured and interpreted as small, medium, or large according to the following criteria: 0.0-0.1 (Negligible association), 0.1-0.2 (Weak association), 0.2-0.4 (Moderate association), 0.4-0.6 (Relatively strong association), 0.6-0.8, (Strong association), 0.8-1.0 (Very strong association) [69].

2.8 Predictive Modelling via Machine Learning

First the parameters and their categories which will be used as input data for machine learning were defined (Table 2.12). The dataset to be used for the predictive modeling studies was the same as the original one with statistical analysis (Table 2.13).

Table 2.12 The categories of variables used for predictive modeling

Patient characteristics	Type of variable	Range
Female age (years) (min-max)	numerical	20-45
Rearrangement group	categorical	RECT,ROBT,INV,CRR
Sex	categorical	F,M
Infertility status	categorical	Primary, Secondary, Fertile
Duration of infertility (years) (min-max)	numerical	1-22
Number of previous trials (min-max)	numerical	0-13
Body mass index (kg/m^2)	categorical	4 categories <18.5, 18.5-24.9, 25-29.9 and >30.
Anti mullerian hormone (ng/ml)	categorical	3 categories <1, 1-3.9 and >4
Male age (min-max)	numerical	23-52
Sperm count (million/ml)	categorical	4 categories; <1 million/ml, 1-5 milllion, 5-20 million, >20 million.
Sperm source	categorical	Ejaculate or TESE
number of collected oocytes (min-max)	numerical	1-47
number of injected oocytes (min-max)	numerical	1-44
number of fertilized oocytes (min-max)	numerical	1-35

Table 2.13 Pre and post-treatment characteristics

Pre-treatment characteristics	**TOTAL**
No of patients	300
Total no of OPU cycles	443
Female age (y), median (min-max)	33 (20-45)
Male age (y), median (min-max)	36 (23-52)
Duration of infertility (y), median (min-max)	4 (1-22)
No of previous trials, median (min-max)	4 (0-13)
BMI, (kg/m2) (min-max)	23.6 (16.4-43.4)
AMH (ng/ml) (min-max)	2.34 (0.01-15.7)
Post-treatment characteristics	
Total collected oocytes, n (median, min-max)	12 (1-47)
Total injected oocytes, n (median, min-max)	9 (1-44)
Fertilized, n (median, min-max)	8 (1-35)
Embryos biopsied, n (median, min-max)	1960, 4 (1-17)
Embryos analyzed, n (median, min-max)	1835 4(1-17)

2.8.1 Pre-processing of dataset

In order to clean and make the data ready Tidyverse package (R software) was used. The number of empty cells, missing variables (NA) and typos were checked by the following commands. The missing variables were filled by “impute” command where necessary (Figure 2.21).

2.8.2 Model selection and training

There were two different model approaches with different inputs. The aim of the first model was to predict probability of embryo transfer using the pre-treatment clinical values; such as BMI, female age, male age which are known before the initiation of the cycle. The aim of the second model was to predict the probability of embryo transfer using both pre- and post-treatment values in this time including the data regarding the number of collected oocytes and the number of biopsied embryos and etc. For the predictions random forest and artificial neural network models were used. 75% of the data was used for training. d3 AMH values were missing in nearly 30% of the patients. These models were repeated both excluding and including AMH values in

```r
data <- readxl::read_excel("VERİ SETİ-3 - Tahmin Modelleme Calismasi_23.04.2022.xlsx") %>%
  janitor::clean_names()

good_columns <- data %>%
  summarise(across(everything(),~ sum(is.na(.)))) %>%
  pivot_longer(everything(), names_to = "na_items", values_to = "count") %>%
  filter(count < 15) %>%
  pull(na_items)

good_columns <- c(good_columns, "amh")

# Columns were categorised by 2 as numerical or categorical.
categorical_columns <- data %>%
  select(good_columns) %>%
  select(-no_of_cycles) %>%
  summarise(across(everything(),~ mean(., na.rm=T))) %>%
  pivot_longer(everything(), names_to = "items", values_to = "avg") %>%
  filter(avg < 2) %>%
  filter(items!="cycle_no") %>%
  pull(items)
```

Figure 2.21 Determining columns with empty values

order to see the additional effect. For this reason first 443 cycles without AMH values were used. Then patients with known AMH values were included only (313 cycles). As a summary different algorithms were tested with different datasets.

2.8.3 Random Forest

2.8.3.1 Random forest with pre-treatment variables

The columns having unnecessary data and missing data were cleaned. The new columns were named as "good columns". Categorical data were filtered by "<2". In order to exclude post-treatment values from the data set, step-rm command was used for values related with collected oocytes, fertilized, biopsied embryos. K-Nearest Neighbor (KNN) algorithm was used with impute command to fill the ones with missing values such as "sperm count" (Figure 2.22).

```r

library(tidymodels)
set.seed(1)
# Put 3/4 of the data into the training set
data_split <- initial_split(data_clean_amh, prop = 3/4)

# Create data frames for the two sets:
train_data <- training(data_split)
test_data  <- testing(data_split)

amh_pre_rec <-
  recipe(normal_balanced ~ ., data = train_data) %>%
  update_role(patient_code, new_role = "ID") %>%
  step_rm(normal_balanced_numeric, et_var_1_yok_0, collected_
   oocytes, injected_oocytes, fertilized, biopsied, cycle_no,
    analysis_method_a_cgh_1_ngs_2) %>%
  step_impute_knn(sperm_sayi) %>%
  step_dummy(all_nominal_predictors())

the_pre_workflow <-
  workflow() %>%
  add_model(rf_mod) %>%
  add_recipe(amh_pre_rec)

  last_mod <- rand_forest(mtry = 10, min_n = 10, trees =
   1000) %>%
  set_engine("ranger", num.threads = cores, importance = "
   impurity") %>%
  set_mode("classification")

# the last workflow
last_workflow <-
  the_pre_workflow %>%
  update_model(last_mod)
# the last fit
set.seed(345)
last_fit <-
  last_workflow %>%
  last_fit(data_split)
```

Figure 2.22 R commands for random forest with pre-treatment variables

2.8.3.2 Random forest with post-treatment variables

Treatment specific values such as collected oocytes and biopsied embryo numbers were added for this dataset.

These algorithms were repeated with and without AMH data (Figure 2.23).

```
amh_rec <-
  recipe(normal_balanced ~ ., data = train_data) %>%
  update_role(patient_code, new_role = "ID") %>%
  step_rm(normal_balanced_numeric, et_var_1_yok_0) %>%
  step_impute_knn(sperm_sayi) %>%
  step_dummy(all_nominal_predictors())
cores <- parallel::detectCores()
rf_mod <-
  rand_forest(mtry = 10, min_n = 10, trees = 1000) %>%
  set_engine("ranger", num.threads = cores) %>%
  set_mode("classification")
bt_mod <- boost_tree(
  mode = "classification",
  engine = "xgboost")
nn_mod <- mlp(mode = "classification", penalty = 0.01, engine
    = "nnet")
the_workflow <-
  workflow() %>%
  add_model(rf_mod) %>%
  add_recipe(amh_rec)
set.seed(345)
the_fit <-
  the_workflow %>%
  fit(data = train_data)
the_fit %>%
  extract_fit_parsnip() %>%
  broom::tidy()
predict(the_fit, test_data)
augment(the_fit, test_data)
augment(the_fit, test_data) %>%
  roc_curve(truth = normal_balanced, .pred_0) %>%
  autoplot()
# the last model
last_mod <-
  rand_forest(mtry = 10, min_n = 10, trees = 1000) %>%
  set_engine("ranger", num.threads = cores, importance = "
   impurity") %>%
  set_mode("classification")
last_workflow <- the_workflow %>%
  update_model(last_mod)
# the last fit
set.seed(345)
last_fit <-
  last_workflow %>%
  last_fit(data_split)
last_fit %>%
  collect_metrics()
last_fit %>%
  extract_fit_parsnip() %>%
  vip::vip(num_features = 17)
```

Figure 2.23 Commands for random forest algorithm using post-treatment variables

2.8.4 Artificial neural network

2.8.4.1 Artificial neural network with pre-treatment variables

Similar commands were repeated to clean and fill the missing the data. This time "NN(neural network)" model was used (Figure 2.24).

```
emb_pre_rec <-
  recipe(normal_balanced ~ ., data = train_data) %>%
  update_role(patient_code, new_role = "ID") %>%
  step_rm(normal_balanced_numeric, et_var_1_yok_0, collected_
   oocytes, injected_oocytes, fertilized, biopsied) %>%
  step_impute_knn(bmi, sperm_sayi) %>%
  step_dummy(all_nominal_predictors())
the_pre_workflow <-
  workflow() %>%
  add_model(nn_mod) %>%
  add_recipe(emb_pre_rec)
# the last model
last_mod <- nn_mod
# the last workflow
last_workflow <-
  the_pre_workflow %>%
  update_model(last_mod)
# the last fit
set.seed(345)
last_fit <-
  last_workflow %>%
  last_fit(data_split)
last_fit %>%
  collect_metrics()
```

Figure 2.24 Artificial neural network with pre-treatment variables

2.8.4.2 Artificial neural network with post-treatment variables

Artificial neural network model with post-treatment variables were built (Figure 2.25)

```r
library(tidymodels)
set.seed(1)
# Put 3/4 of the data into the training set
data_split <- initial_split(data_clean, prop = 3/4)
# Create data frames for the two sets:
train_data <- training(data_split)
test_data  <- testing(data_split)
The recipe
emb_rec <-
  recipe(normal_balanced ~ ., data = train_data) %>%
  update_role(patient_code, new_role = "ID") %>%
  step_rm(normal_balanced_numeric, et_var_1_yok_0) %>%
  step_impute_knn(bmi, sperm_sayi) %>%
  step_dummy(all_nominal_predictors())
cores <- parallel::detectCores()
rf_mod <-
  rand_forest(mtry = 10, min_n = 10, trees = 1000) %>%
  set_engine("ranger", num.threads = cores) %>%
  set_mode("classification")
bt_mod <- boost_tree(
  mode = "classification",
  engine = "xgboost")
nn_mod <- mlp(mode = "classification", penalty = 0.01, engine
    = "nnet")
the_workflow <-
  workflow() %>%
  add_model(nn_mod) %>%
  add_recipe(emb_rec)
set.seed(345)
the_fit <-
  the_workflow %>%
  fit(data = train_data)
the_fit %>%
  extract_fit_parsnip() %>%
  broom::tidy()
augment(the_fit, test_data) %>%
  roc_curve(truth = normal_balanced, .pred_0) %>%
  autoplot()
last_workflow <- the_workflow %>%
  update_model(last_mod)
# the last fit
set.seed(345)
last_fit <-
  last_workflow %>%
  last_fit(data_split)
last_fit %>%
  collect_metrics()
```

Figure 2.25 Artificial neural network with post-treatment variables

2.8.5 Data merging

The dataset was merged with another dataset for the purpose of increasing the success of prediction. Additional hormonal data including LH, FSH were added to the pre-existing data with matching the rows with cycle ID and embryo ID. Additionally,

some of the hormonal input parameters were categorized as 1-4 (Figure 2.26).

```
readxl::read_excel("BIG DATA SADELESTI.xlsx",3) %>%
  janitor::clean_names() %>% #colnames()
select(patient_code,cycle_id,female_age,amh_ng_ml,hcg_gunu_e2
   ,fsh_baslangic,fsh_total,d3_e2,d3_lh,d3_prog,erkek_yasi,
   ovarian_reserve_low_1_normal_2_high_3,ind_total_gnd,gunluk
   _ort_gnd) %>%
  mutate(patient_code = str_to_upper(patient_code)) %>%
  distinct() -> patient_hormone
patient_hormone <- patient_hormone %>%
  mutate(amh_cat = case_when(
    amh_ng_ml < 1 ~ "low",
    amh_ng_ml <= 4 ~ "medium",
    amh_ng_ml > 4 ~ "high",
    TRUE ~ "NA"
  )) %>%
  mutate(amh_cat = as.factor(amh_cat))
patient_hormone <- patient_hormone %>%
  mutate(female_age_cat = case_when(
    female_age <= 30 ~ 1,
    female_age <= 35 ~ 2,
    female_age <= 40 ~ 3,
    female_age > 40 ~ 4,
    TRUE ~ 0
  )) %>%
  mutate(female_age_cat = as.factor(female_age_cat))
adapter_table <- readxl::read_excel("RobT-RecT-Inv Summary
   Results & Tables_07.08.2021.xlsx",1) %>%
  select(cycle_id=‘SIKLUS ID‘, patient_code=‘Patient code‘,
   cycle_no=‘cycle no‘) %>%
  mutate(patient_code = str_replace(patient_code, "OTHER", "
   COMPLEX"))
merged_table <- readRDS("data_clean.rds") %>%
  inner_join(adapter_table, by = c("patient_code", "cycle_no"
   )) %>%
  select(patient_code, cycle_no, cycle_id, everything()) %>%
  inner_join(patient_hormone, by="cycle_id") %>%
  mutate(patient_code=patient_code.x) %>%
  select(-patient_code.y,-patient_code.x)
```

Figure 2.26 Data merging

2.8.6 Random forest on merged data for post treatment variables

After two datasets were merged, the missing variables were filled by KNN algorithm. However, for AMH column, there were 66 samples with missing data. The model was tested both with 443 cycles with AMH filled by KNN, and on the other dataset in which the missing rows were removed (Figure 2.27).

75% of the data were put into the training set. Data frames were created for the two sets: training and test.

```
library(tidymodels)
set.seed(1)
set
data_split <- initial_split(merged_table, prop = 3/4)
train_data <- training(data_split)
test_data  <- testing(data_split)
emb_rec <-
  recipe(normal_balanced ~ ., data = train_data) %>%
  update_role(patient_code, new_role = "ID") %>%
  step_rm(normal_balanced_numeric, et_var_1_yok_0, cycle_no,
   cycle_id) %>%
  step_impute_knn(bmi, sperm_sayi, amh_ng_ml, hcg_gunu_e2,
   fsh_baslangic, fsh_total, d3_e2, d3_lh, d3_prog, amh_cat,
   ind_total_gnd, gunluk_ort_gnd) %>%
  #step_normalize(all_numeric()) %>%
  #step_pca(all_numeric(), num_comp = 7) %>%
  step_dummy(all_nominal_predictors())
```

Figure 2.27 Random forest on merged data for post treatment variables

2.8.7 Random forest with hyperparameter tunning

For a better performance, "hyperparameter tunning" was applied to random forest model (Figure 2.28).

```
cores <- parallel::detectCores()
rf_mod <-
  rand_forest(mtry = tune(), min_n = tune(), trees = 1000)
   %>%
  set_engine("ranger", num.threads = cores, importance = "
   impurity") %>%
  set_mode("classification")
tune_wf <-
  workflow() %>%
  add_model(rf_mod) %>%
  add_recipe(emb_rec)
trees_folds <- vfold_cv(train_data)
#doParallel::registerDoParallel()
set.seed(345)
tune_res <- tune_grid(
  tune_wf,
  resamples = trees_folds,
  grid = 20
)
tune_res
collect_metrics() %>%
  filter(.metric == "roc_auc") %>%
  mutate(min_n = factor(min_n)) %>%
  ggplot(aes(mtry, mean, color = min_n)) +
  geom_line(alpha = 0.5, size = 1.5) +
  geom_point() +
  labs(y = "AUC")
best_auc <- select_best(regular_res, "roc_auc")
final_rf <- finalize_model(
  rf_mod,
  best_auc
)
final_rf
```

Figure 2.28 Random forest with hyperparameter tunning

3
RESULTS AND DISCUSSION

3.1 Genetic analysis results and clinical outcomes

300 couples in total underwent 443 ovarian stimulation cycles and subsequent PGT-SR. The majority of cycles (305) were performed for RECT carriers; 86 PGT cycles were performed ROBT carriers; 40 PGT cycles were performed for INV carriers; 12 PGT cycles were performed for couples with CCR. The median female age was 33 years (range 20-45). The demographic, genetic and clinical outcomes of the treatments were summarized (Table 3.1). A total of 5878 oocytes were collected of which 4049 (82.7%) were fertilized from 4896 injected oocytes. 1960 embryos were biopsied and 1835 embryos were sent for analysis. The remaining has not been requested for analysis by the patient at the time this study was conducted. Of the 1777 embryos successfully diagnosed, 423 (23.8%) were both normal/balanced for SR and euploid for non-rearranged chromosomes (a), 543 (30.6%) were unbalanced for SR but euploid for non-rearranged chromosomes (b), 362 (20.4%) were both unbalanced for SR and aneuploid for non-rearranged chromosomes (c) 385 (21.6%) were normal/balanced for SR but aneuploid for non-rearranged chromosomes (d), and 64 (3.6%) were chaotic (e). The proportion of normal/balanced euploid embryo rate among the carrier groups were statistically different; such that RECT, ROBT, and INV carriers have 19.4%, 33.2% and 35.8% of the embryos were found to be normal/balanced and euploid respectively ($p<0.001$). The balanced/euploid embryo rate was found to be 25.0% in CCR carriers. Another end point was the probability of a cycle to reach for embryo transfer. This is related with the availability of at least one transferable embryo which was highest for INV carriers (77.5%), and lowest for CCR carriers (33.3%) ($p<0.05$). In 249 cycles, embryo transfer was performed which was also termed as first embryo transfer (Table 3.1). In addition, 41 frozen-thaw ET cycles were performed when there are supernumerary transferable embryos generated in the same OPU cycle in which previous embryo transfer has not resulted in viable pregnancy and live birth. Out of 290 cycles with ET, in 88, embryo transfers were performed with fresh blastocysts, and in 202 were performed with frozen and thawed blastocysts.

According to that, the CPR per first ET was 61% (152/249) and cumulative CPR was 69.5% (173/249). The LBR per first ET was 49.4% (123/249) and cumulative LBR was 55.8% (139/249). The clinical outcome parameters such as CPR, IR, LBR were similar among different carrier groups (Table 3.1). To date 153 healthy children have been born. No clinical misdiagnoses have been reported so far as a follow up in this dataset.

3.2 Investigation of risk factors for segregation and *de novo* aneuploidy

In order to assess the effects of the parameters on both genetic and clinical results, the carriers were grouped according to the type of SR, female age and sex of the carriers. The two categories were established for female age (<35 vs. ≥ 35), and sex of the carrier (male vs female) (Table 3.2). In order to simplify the genetic results and test the hypothesis of ICE, the outcomes of embryos were categorized as "cumulative balanced" and "cumulative unbalanced" (regardless of the presence of de-novo aneuploidies). Likewise, embryos were categorized as "cumulative euploid" and "cumulative aneuploid" regardless of the segregational status of the embryos in order to investigate the factors related with de-novo aneuploidy. According to results, female age did not have any impact on segregational status for RECT carriers, but female age was found to have a strong correlation with the proportion of embryos with *de novo* aneuploidy (<0.001). Sex was not found to have any association with the segregation and the clinical outcomes for RECT carriers (Table 3.2). ROBT carriers had similar results; as RECT, female age did not have any impact on segregational status, on the proportion of embryos with *de novo* aneuploidy (<0.001). However, sex was found to have an impact on segregational status among ROBT carriers with the advantage of male carriers such that; male carriers produce higher proportion of balanced/normal embryos compared to female carriers (73.8% vs 63.0%) ($p<0.05$) (Table 3.2).

All 40 embryos were found to have balanced segregation in the older group of INV carriers (100%). This was statistically different compared to the younger group (23.3%) ($p<0.001$). However, the proportion of embryos with *de-novo* aneuploidy was higher in the older group compared to the younger group (45% vs 60%) but this difference did not reach statistical significance. Sex was not found to be associated with the segregation of re-arranged and non-rearranged chromosomes for INV carriers (Table 3.2). When the INV carriers were grouped according to the type, the segregation abnormalities of PAI, and PEI were similar: 13.5% and 18.9% respectively ($p>0.05$).

Table 3.1 The clinical characteristics and outcomes of 443 cycles with preimplantation genetic diagnosis for structural rearrangements

CLINICAL PARAMETERS	RECT (I)	ROBT (II)	INV (III)	CCR (IV)	TOTAL	p-value
No of patients, n	198	60	31	11	300	
Total no of OPU cycles, n	305	86	40	12	443	
Female age (y), median (min-max)	33 (20-45)	33 (22-42)	34 (23-44)	32 (22-41)	33 (20-45)	0.421
Total collected oocytes, median (min-max)	12 (1-43)	11.5 (1-36)	11.5 (1-47)	9 (2-27)	12 (1-47)	0.376
Total injected oocytes, median (min-max)	10 (1-42)	9 (1-30)	10 (1-44)	9 (2-21)	9 (1-44)	0.402
Embryos biopsied, n (median, min-max)	1357 (4, 1-13)	359 (3, 1-17)	190 (4, 1-16)	54(3.5,1-11)	1960, 4 (1-17)	0.240
Embryos analyzed, n (median, min-max)	1260 (4, 1-11)	342 (3, 1-17)	179 (4, 1-14)	54 (3.5, 1-11)	1835 (4, 1-17)	0.567
Embryos diagnosed, n (%)	1215 (96.4)	337 (98.5)	173 (96.6)	52 (96.3)	1777 (96.8)	0.739
Normal/Balanced & euploid embryos, (%)	236 (19.4)	112 (33.2)	62 (35.8)	13 (25.0)	423 (23.8)	<0.001
Unbalanced, n (%)	438 (36.0)	55 (16.3)	27 (15.6)	23 (44.2)	543 (30.6)	<0.001
Unbalanced & *de novo* aneuploid, n (%)	305 (25.1)	43 (12.8)	3 (1.7)	11 (21.1)	362 (20.4)	<0.001
de novo aneuploid, n (%)	194 (16.0)	112 (33.2)	77 (44.5)	2 (3.8)	385 (21.7)	<0.001
Chaotic, n (%)	42 (3.5)	15 (4.5)	4 (2.3)	3 (5.8)	64 (3.6)	0.511
Total no of cumulative ET cycles, n	183	66	37	4	290	
cCPR, n (%)	106 (66.6)	41 (74.5)	23 (74.2)	3 (75)	173 (69.5)	0.829
Clinical Pregnancy Loss, n	18	6	6	0	30	
Ectopic pregnancy, n	2	0	0	0	2	
Preterm delivery, n	1	0	0	0	1	
Stillbirth, n	1	0	0	0	1	
cLBR , n (%)	84 (52.8)	35 (63.6)	17 (54.8)	3 (75)	139 (55.8)	0.527
Implantation rate (%)	114/200 (57)	46/78 (59.0)	24/42 (57.1)	3/4 (75)	187/324 (57.7)	0.899
Delivered babies, n	92	40	18	3	153	
No of twins, n	8	5	1	0	14	

Table 3.2 Investigation of rearrangement type, female age and sex of the carrier as risk factors on segregation and de novo aneuploidy

	<35	**≥ 35**	**p-value**	**MALE**	**FEMALE**	**p-value**
RECIPROCAL TRANSLOCATION						
No of OPU cycles, n	202	103		146	159	
Female age, y (median, min-max)	30.5 (20-34)	37 (35-45)		33 (20-45)	32 (21-43)	
Cumulative unbalanced embryos , n (%)	524/836 (62.7)	219/337 (65.0)	0.455	345/567 (60.8)	398/606 (65.7)	0.097
Cumulative *de novo* aneuploid embryos, n (%)	345/862 (40.0)	196/353 (55.5)	<0.001	267/595 (44.9)	274/620 (44.2)	0.856
Embryos normal/balanced and euploid, n (%)	183/862 (21.2)	53/353 (15.0)	<0.013	111/595 (18.6)	125/620 (20)	0.554
No of OPU cycles that reach ET, n (%)	121 (59.9)	38 (36.9)	<0.001	84 (57.5)	75 (47.2)	0.090
Healthy live birth, n (%)	52 (43)	22 (57.9)	0.155	43 (51.2)	31 (41.3)	0.278
ROBERTSONIAN TRANSLOCATION						
No of OPU cycles, n	52	33		51	34	
Female age, y (median, min-max)	29.5 (22-34)	38 (35-42)		32 (22-40)	36,5 (24-42)	
Cumulative unbalanced embryos , n (%)	74/250 (29.6)	24/72 (33.3)	0.644	51/195 (26.2)	47/127 (37)	0.038
Cumulative de novo aneuploid embryos, n (%)	113/261 (43.3)	57/76 (75.0)	<0.001	100/206 (48.5)	70/131 (53.4)	0.445
Embryos normal/balanced and euploid, n (%)	99/261 (37.9)	13/76 (17.1)	0.001	78/206 (37.9)	34/131 (25.9)	0.032
No of OPU cycles that reach ET, n (%)	43 (82.7)	11 (33.3)	<0.001	38 (74.5)	16 (47.1)	0.018
Healthy live birth, n (%)	23 (53.5)	7 (63.6)	0.791	20 (52.6)	10 (62.5)	0.713
INVERSION						
No of OPU cycles, n	27	13		16	24	
Female age, y (median, min-max)	31 (23-34)	38 (35-44)		34 (26-42)	33.5 (23-44)	
Cumulative unbalanced embryos , n (%)	30/129 (23.3)	0/40 (0)	<0.001	11/59 (18.6)	19/110 (17.3)	0.991
Cumulative *de novo* aneuploid embryos, n (%)	60/133 (45.1)	24/40 (60)	0.141	31/62 (50)	53/111 (47.7)	0.900
Embryos normal/balanced and euploid, n (%)	46/133 (34.6)	16/40 (40)	0.661	22/62 (35.5)	40/111 (36)	0.926
No of OPU cycles that reach ET, n (%)	22 (81.5)	9 (69.2)	0.642	13 (81.3)	18 (75.0)	0.938
Healthy live birth, n (%)	11 (50.0)	4 (44.4)	0.908	5 (38.5)	10 (55.6)	0.564

The same statistical analysis could not be performed for CCR carriers as they have small sample size. In summary, advancing female age had a negative impact on the proportion of transferable embryos and the proportion of cycles with at least one transferable embryo for RECT and ROBT carriers but not for INV carriers. On the other hand, sex had an impact on the proportion of transferable embryos only for ROBT carriers with advantage of male carriers. In addition, once ET was achieved, the clinical outcomes (CPR and LBR) of these groups and sub-groups were similar regardless of the type of the rearrangement, age and sex (Table 3.2). In order to understand the magnitude of relationship, logistic regression analysis was performed by assessing the type of rearrangement, female age and sex by using forward stepwise binary logistic regression analysis with two separate models. In the first model, the outcome was defined as segregational abnormalities, and categorized as either cumulative balanced or unbalanced (Table 3.3), the risk factors were the type of rearrangement, female age and sex. In the second model, the outcome was categorized based on the aneuploidy state as either cumulative euploid or *de novo* aneuploid respectively (Table 3.4). According to these models, sex ($p<0.05$) and the translocation type ($p<0.001$) were found to be associated with the segregational abnormalities. In detail, female carriers were 1.29x more likely to produce unbalanced embryos compared to male carriers. Female carriers was a risk factor. Furthermore, CCR were found to be the other important risk factor among other groups. When INV carriers were taken as baseline, ROBT, RECT and CCR carriers have 2.17x, 8.34x and 12.51x more likely to produce unbalanced embryos. Female age did not seem to have any relationship with segregational abnormalities ($p>0.05$) (Table 3.3).

Table 3.3 Risk factors associated with segregational abnormalities

	p-value	**O.R.**	**95% C.I.**	
SEX (Ref:Male)	0.014	1.29	1.05	1.59
Group (Ref:Inv)				
Robt	0.001	2.17	1.36	3.45
Rect	<0.001	8.3	5.5	12.6
CCR	<0.001	12.5	5.7	27.3

De novo aneuploidy was found to have only associated with female age ($p<0.001$), but not with rearrangement type and sex. This association was assessed by two models, continuous and binary. In the first model, each increase in maternal age in terms of years, the risk of aneuploidy increased by 1.09x. According to the second model with different age groups, the risk of *de novo* aneuploidy increased by 2.09x for women

who were 35 or older compared to ones who were younger (p<0.001) (Table 3.4).

Table 3.4 Risk factors associated with *de novo* aneuploidy

	p-value	O.R.	95% C.I.	
Female age (years)	<0.001	1.09	1.06	1.11

Additionally, for INV carriers, both the length and the proportion of the inverted segment were assessed if there were any associations with the segregational abnormalities and occurrence of *de novo* aneuploidy risk. According to the results the length was found to be highly associated with the segregational abnormalities such that; the longer the inverted region, the higher the chance of an embryo being unbalanced (p<0.001) (Table 3.5). In another way, the average size of the inversion was 39.1 Mb among embryos with balanced segregations, where it was 87.5 Mb among embryos with unbalanced segregations (p<0.001). Similarly, the average proportion of the inverted region was found to be associated with the risk of unbalanced segregation (p<0.001).

Table 3.5 Parameters associated with segregational abnormalities in inversion carriers

	cBalanced (n=139)	cUnbalanced (n=30)	p-value
Inverted Segment Length	39.1(24.2-180.8)	87.5(24.2-180.8)	<0.001
Inverted Segment Length/ Total Chromosome	32.9(12.4-92.8)	51.6(12.4-92.8)	<0.001
Female age	34(23-44)	30(26-34)	0.013
Sub group (Paracentric)	32(23.0)	5(16.7)	0.603

3.3 Investigation of a Possible Interchromosomal effect (ICE)

In order to test the hypothesis that an ICE was apparent from this dataset, the aneuploidy rate per embryo and the incidence of aneuploidy per each chromosome pair were compared between cases (443) and (886) controls (2 controls were enrolled per each case). The demographical comparisons of both groups (cases and controls which were matched not only per case but also per embryo) can be seen in Table 3.6. A total of 117,033 chromosomal pairs (37,458 for PGT-SR cases, 79,575 for controls) were analyzed from 5237 embryos (1777 PGT-SR cases and 3460 controls).

The chromosomes involved in the rearrangement were excluded from the analysis. There were no differences in terms of aneuploidy rate per each chromosomal pair except for chromosome 17 (4.7% in cases vs 3.1% in controls, $p<0.05$). However, the effect size measurement revealed this difference does not have any practical significance therefore negligible ($\varphi = 0.006<0.1$). The overall aneuploidy rate per chromosome was 5.3% in PGT-SR compared to 4.9% in controls. Although these statistics resulted in a p value with <0.05, the subsequent effect size measurement (φ) revealed this difference as negligible ($\varphi = 0.007$) (Table 3.7). Both in case and controls, chromosomes with the highest incidence of abnormalities were 16 and 22 (chr 16: 8.5% in cases vs 9.1% in controls, chr 22: 6.5% in cases vs 7.7% in controls) (Figure 3.1).

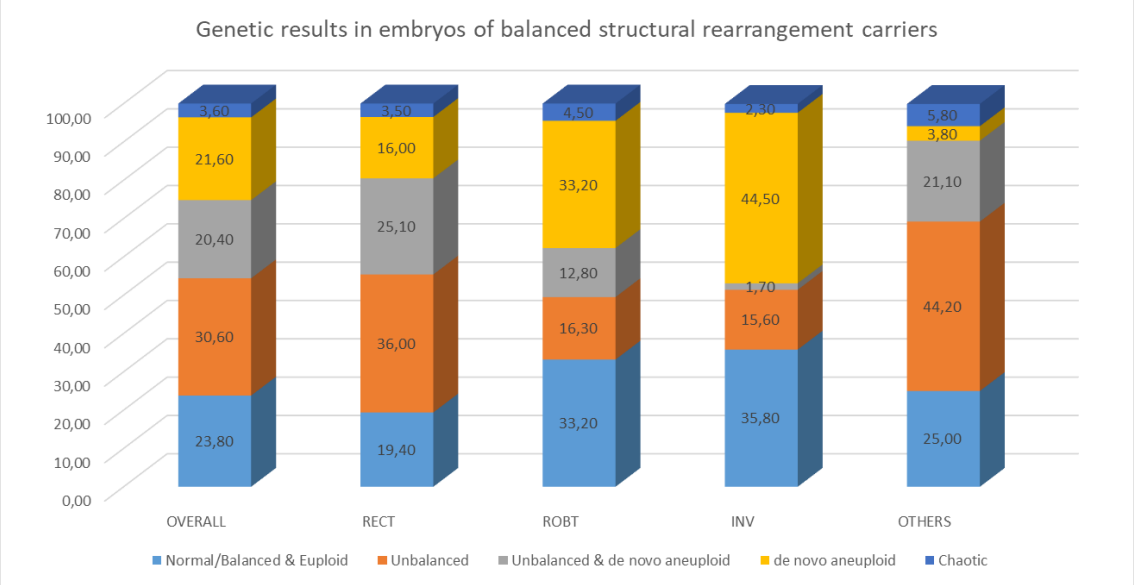


Figure 3.1 Genetic results of embryos of balanced carriers

3.3.1 ICE in sub groups

After no evidence was found in the general population of cases and control subjects, ICE was investigated also in subgroups in order to test if there is any rearrangement-specific effect. For that reason, each subgroup (RECT, ROBT, INV and CCR) was compared against their corresponding controls. The overall incidence of embryos with *de novo* abnormality was even lower in RECT (44.5%) compared to controls (53.3%) which revealed a significance but with negligible association ($p<0.001$, $\varphi<0.1$). Furthermore, the incidence of aneuploidy per chromosomal pair was statistically different in RECT (5.2%) compared to controls (4.8%) ($p<0.05$), however, the effect size measurement revealed that this association was absent or negligible ($\varphi<0.1$). In ROBT, the proportion of embryos with *de novo* abnormality did not differ between cases and controls (50.4% vs 53.0%, $p>0.5$). Moreover, the

incidence of abnormalities per chromosomal pair was higher (5.8%) in PGT-SR cases compared to controls (4.9%) ($p<0.05$). The phi effect size measurement revealed that this association was absent or negligible ($\varphi<0.1$). The other subgroups revealed little or no difference when compared between PGT-SR case and controls for both in terms of aneuploidy in embryos and in chromosomal pairs analyzed ($p>0.05$) (Table 3.8). So, in summary, although some analyses revealed apparently statistically significant differences according to p-values, more in-depth analysis of effect size only revealed either no or negligible associations. Based on this dataset, there was no evidence of an ICE.

Table 3.6 Comparison of demographical data of embryos from 443 case and 886 matched control treatment cycles

Variables	SR carriers	Controls	p-value	Cliff's Delta (δ)	%95 C. I.	Size effect Interpretation
No of cycles, n	443	886				
Total no of embryos matched by clinical parameters, n	1777	3460				
Female age (y), median (min-max)	32 (20-45)	32 (20-45)	0.100			
Total no of oocytes collected, median (min-max)	15 (1-47)	14 (1-49)	<0.001	0.1	0.067-0.134	negligible
MII, median (min-max)	13 (1-44)	12 (1-46)	<0.001	0.081	0.047-0.114	negligible
PN, median (min-max)	11 (1-35)	10 (1-41)	<0.001	0.099	0.066-0.132	negligible
Male age, median (min-max)	34 (23-59)	35 (22-61)	NA			
No of previous ART cycles, median (min-max)	3 (0-13)	3 (0-16)	<0.001	0.107	0.075-0.139	negligible
BMI (kg/m2), median (min-max)	23.9 (16.4-43.4)	23.3 (15-45.1)	<0.001	0.098	0.065-0.130	negligible
AMH (ng/ml), median (min-max)	2.92 (0.01-15.7)	2.72 (0.03-19.3)	0.032	0.039	0.003-0.076	negligible
Sperm Count (million/ml), median (min-max)	10 (0-98)	8.70 (0-88)	<0.001	0.072	0.038-0.106	negligible
Total gonadotrophin (IU), median (min-max)	1950 (400-7275)	1975 (450-8812.5)	0.285			
No of analyzed blastocysts per cycle, median (min-max)	4 (1-16)	4 (1-15)	0.764			

Table 3.7 Comparisons of individual and total chromosomal error rates between rearrangement carriers and controls based on 117,033 chromosomal pairs

CHR NO	PARAMETER	CASE	CONTROL	p-value
CHR-1	CHR pairs analyzed, n	1515	3460	.
	Abnormal, n (%)	77(5.1)	184(5.3)	0.732
CHR-2	CHR pairs analyzed, n	1609	3460	
	Abnormal, n (%)	90(5.6)	169(4.9)	0.286
CHR-3	CHR pairs analyzed, n	1687	3460	
	Abnormal, n (%)	73(4.3)	153(4.4)	0.876
CHR-4	CHR pairs analyzed, n	1599	3460	
	Abnormal, n (%)	77(4.8)	150(4.3)	0.443
CHR-5	CHR pairs analyzed, n	1669	3460	
	Abnormal, n (%)	81(4.9)	165(4.8)	0.895
CHR-6	CHR pairs analyzed, n	1628	3460	
	Abnormal, n (%)	85(5.2)	147(4.2)	0.121
CHR-7	CHR pairs analyzed, n	1646	3460	
	Abnormal, n (%)	81(4.9)	147(4.2)	0.277
CHR-8	CHR pairs analyzed, n	1679	3460	
	Abnormal, n (%)	88(5.2)	163(4.7)	0.408
CHR-9	CHR pairs analyzed, n	1634	3460	
	Abnormal, n (%)	102(6.2)	172(5)	0.060
CHR-10	CHR pairs analyzed, n	1630	3460	
	Abnormal, n (%)	91(5.6)	150(4.3)	0.051
CHR-11	CHR pairs analyzed, n	1579	3460	
	Abnormal, n (%)	75(4.7)	128(3.7)	0.079
CHR-12	CHR pairs analyzed, n	1702	3460	
	Abnormal, n (%)	65(3.8)	125(3.6)	0.711
CHR-13	CHR pairs analyzed, n	1449	3460	
	Abnormal, n (%)	82(5.7)	165(4.8)	0.193
CHR-14	CHR pairs analyzed, n	1399	3460	
CHR-14	Abnormal, n (%)	72(5.1)	138(4)	0.072
CHR-15	CHR pairs analyzed, n	1661	3460	

Table 3.7 Comparisons of chromosomal error rates (Continued)

CHR NO	PARAMETER	CASE	CONTROL	p-value
	Abnormal, n (%)	103(6.2)	192(5.5)	0.349
CHR-16	CHR pairs analyzed, n	1629	3460	
	Abnormal, n (%)	138(8.5)	315(9.1)	0.460
CHR-17	CHR pairs analyzed, n	1674	3460	
	Abnormal, n (%)	78(4.7)	108(3.1)	0.006 [1]
CHR-18	CHR pairs analyzed, n	1697	3460	
	Abnormal, n (%)	88(5.2)	176(5.1)	0.880
CHR-19	CHR pairs analyzed, n	1697	3460	
	Abnormal, n (%)	80(4.7)	178(5.1)	0.505
CHR-20	CHR pairs analyzed, n	1695	3460	
	Abnormal, n (%)	88(5.2)	166(4.8)	0.539
CHR-21	CHR pairs analyzed, n	1627	3460	
	Abnormal, n (%)	87(5.3)	205(5.9)	0.409
CHR-22	CHR pairs analyzed, n	1604	3460	
	Abnormal, n (%)	105(6.5)	267(7.7)	0.137
CHR-XY	CHR pairs analyzed, n	1749	3455	
	Abnormal, n (%)	83(4.7)	169(4.9)	0.817
TOTAL	CHR pairs analyzed, n	37458	79575	
	Abnormal, n (%)	1989(5.3)	3932(4.9)	0.007 [2]

[1]Odds ratio:1.517 (C.I. 1.127-2.042), φ 0.039
[2]Odds ratio: 1.079 (C.I. 1.021-1.140), φ 0.008

Table 3.8 Comparison of *de novo* aneuploidy and total chromosomal error rates in sub-groups of structural rearrangement carriers and controls

		CASE	CONTROL	p-value	O.R. (%95 C. I.)	Phi (φ)	Association
RECT	Embryos analyzed, n	1215	2408				
	Embryos with abnormality, n (%)	541 (44.5)	1283 (53.3)	<0.001		0.083	Negligible
	CHR pairs analyzed, n	25515	55384				
	CHR pairs with error, n (%)	1326 (5.2)	2680 (4.8)	0.030	1.078 (1.008-1.153)	0.008	Negligible
ROBT	Embryos analyzed, n	337	624				
	Embryos with abnormality, n (%)	170 (50.4)	331 (53.0)	0.482			
	CHR pairs analyzed, n	7077	14352				
	CHR pairs with error, n (%)	413 (5.8)	704 (4.9)	0.004	1.201 (1.060-1.361)	0.02	Negligible
INV	Embryos analyzed, n	173	326				
	Embryos with abnormality, n (%)	84 (48.6)	178 (54.6)	0.232			
	CHR pairs analyzed, n	3806	7498				
	CHR pairs with error, n (%)	191 (5.0)	401 (5.3)	0.484			
CCR	Embryos analyzed, n	52	102				
	Embryos with abnormality, n (%)	16 (30.8)	55 (53.9)	0.010		0.220	Moderate
	CHR pairs analyzed, n	1060	2346				
	CHR pairs with error, n (%)	59 (5.6)	147 (6.3)	0.474			
TOTAL	Embryos analyzed, n	1777	3460				
	Embryos with abnormality, n (%)	811 (45.6)	1847 (53.4)	<0.001	1.364 (1.216-1.530)	0.073	Negligible
	CHR pairs analyzed, n	37458	79575				
	CHR pairs with error, n (%)	1989 (5.3)	3932 (4.9)	0.007	1.079 (1.021-1.140)	0.008	Negligible

3.4 Performances of Prediction Models Using Machine Learning

As explained in methods section, RF and ANN models were used. The aim was to predict the embryo transfer by using pre-treatment and post treatment data.

3.4.1 Random forest model on pre- and post treatment dataset

According to this model, female age and AMH seems to be the powerful predictors in the pre-treatment dataset (Figure 3.2). When post-treatment values were added, the number of biopsied embryos was the strongest prediction parameter (Figure 3.3)

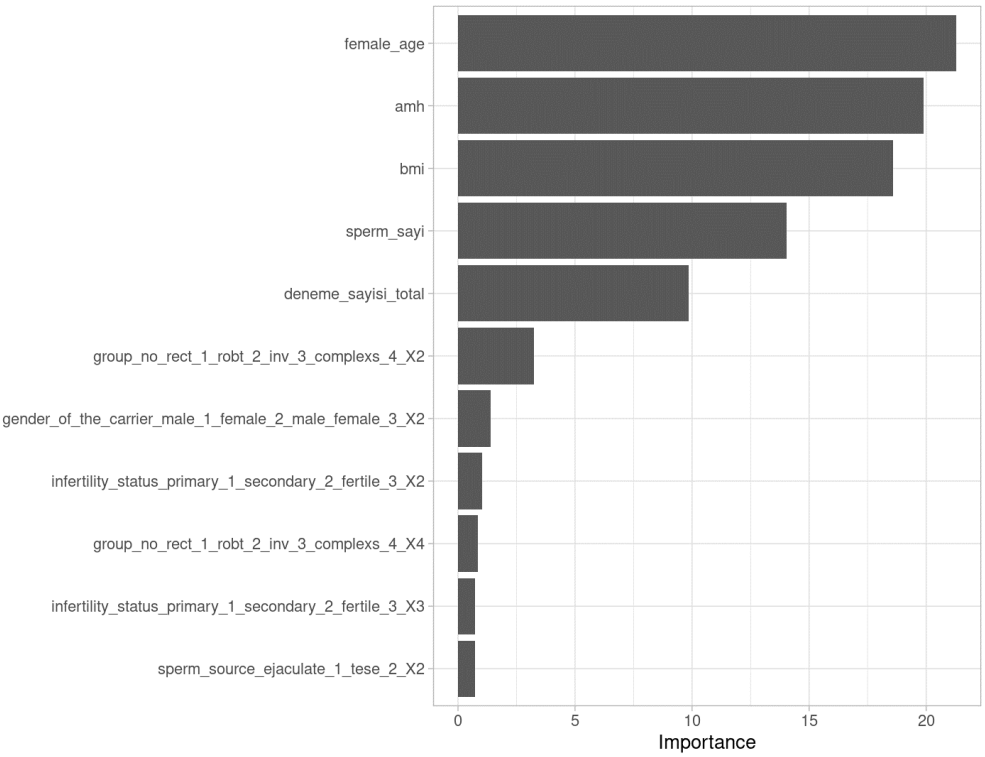


Figure 3.2 The relative importance of pre-treatment parameters in random forest model

3.4.2 Artificial neural network model on pre- and post-treatment dataset

According to this model, the type of rearrangement was the strongest predictive parameter among others (Figure 3.4) Then the same model was tested with the addition of post-treatment variables which indicated number of biopsied embryos as the strongest variable in prediction (Figure 3.5), this model resulted in a higher accuracy of prediction (Figure 3.6).

The first model performances of both random forest and neural networks were successful in the range of 59.5%-%73.6 for accuracy and %57.7-%79.4 for ROC-AUC (Table 3.9).

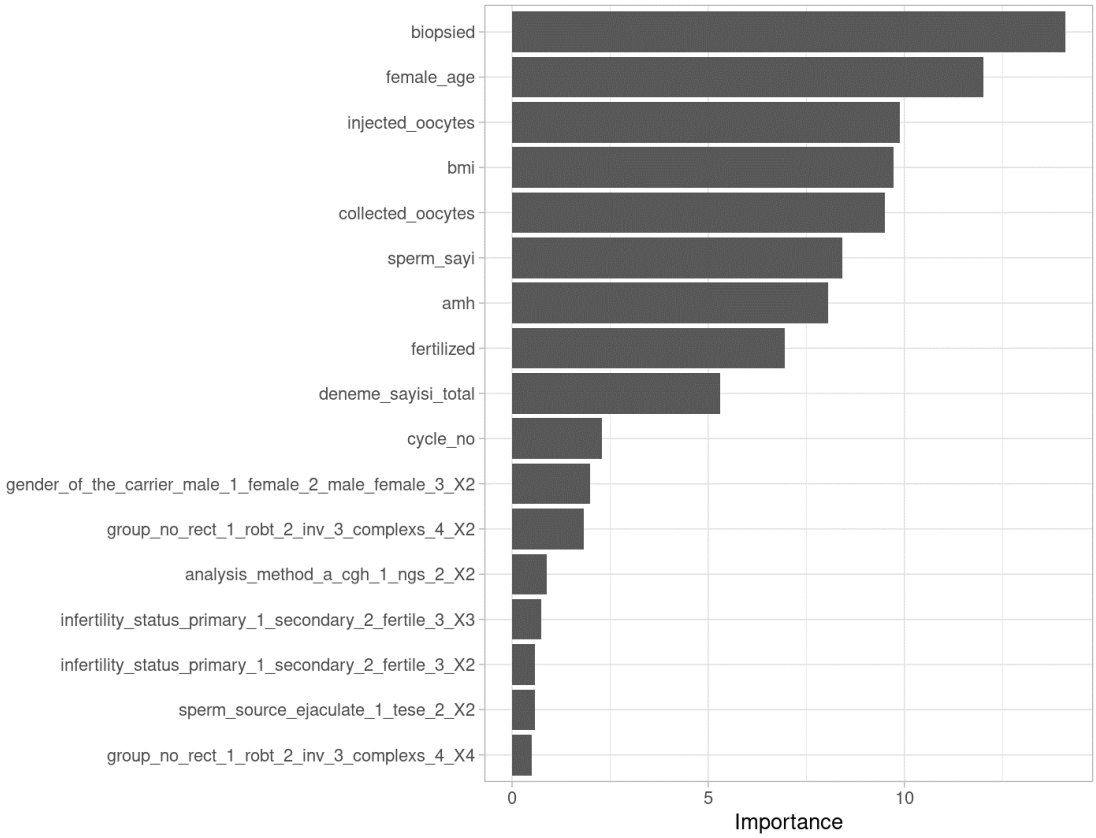


Figure 3.3 The relative importance of post-treatment parameters in random forest model

Table 3.9 Summary of performances of Random Forest and ANN using both pre and post treatment data

Treatment	Model	Accuracy	ROC_AUC
Pre	Random Forest	0.640	0.643
	Artificial Neural Network	0.595	0.577
Post	Random Forest	0.694	0.760
	Artificial Neural Network	0.736	0.794

3.4.3 Calculating variable importance with additional parameters

For an attempt to establish improved model out of the same data, the dataset was merged with another dataset which includes additional hormonal variables. Furthermore, hyperparameter tuning was applied in order to increase the prediction performance. With this facility 20 different models were checked by machine

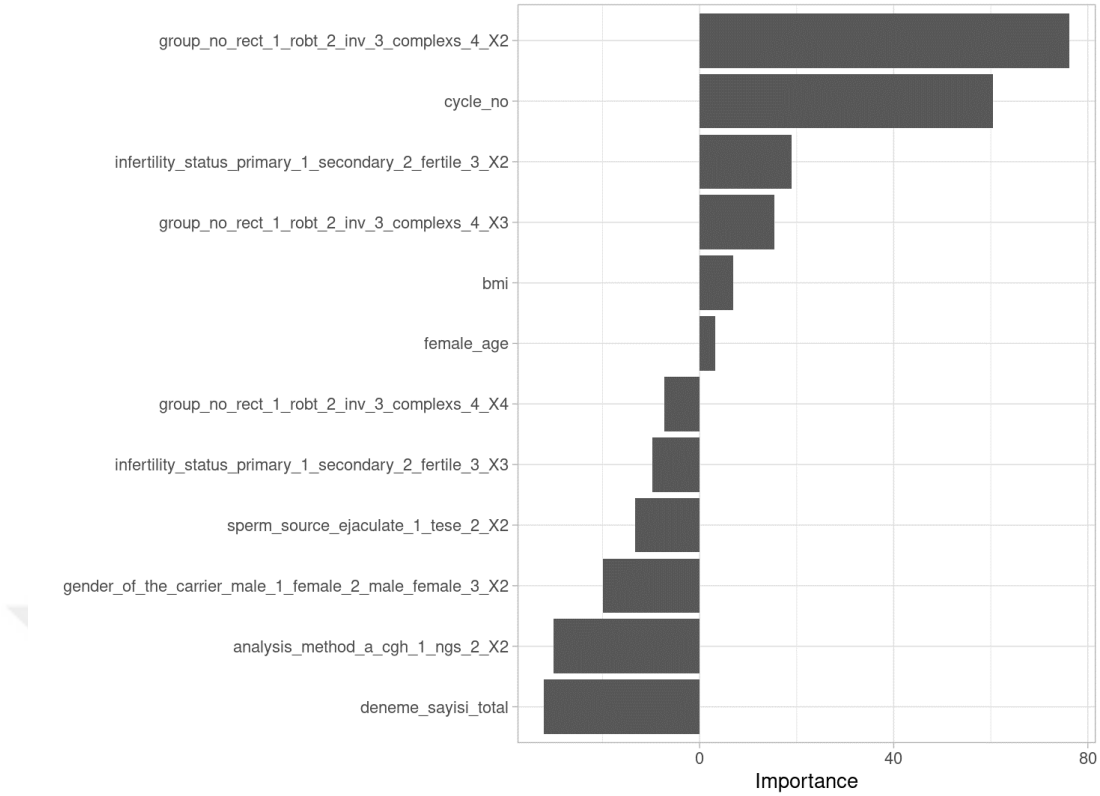


Figure 3.4 The relative importance of pre-treatment variables in neural network model

algorithms and selects the best fitting model for the data. On this merged new dataset, the same models were repeated. In this dataset, random forest model gave a better performance and it was chosen for further hyperparameter tunning.

3.4.4 Random forest with hyperparameter tunning

The 74.7% accuracy and the 78.3% ROC-AUC were achieved with random forest model after hyperparameter tunning was applied (Figure 3.7).

```
## accuracy binary  0.7473684 Preprocessor1_Model1
## roc_auc  binary  0.7836879 Preprocessor1_Model1
```

This last model gave the highest predictability score among the other models in terms of accuracy. In this model injected oocytes had the strongest predictive value (Figure 3.8, Figure 3.9).

3.5 Advantages of this study

This study is one of the largest PGT-SR study from a single center in which both genetic and clinical results were reported together. Another uniqueness of the study is its

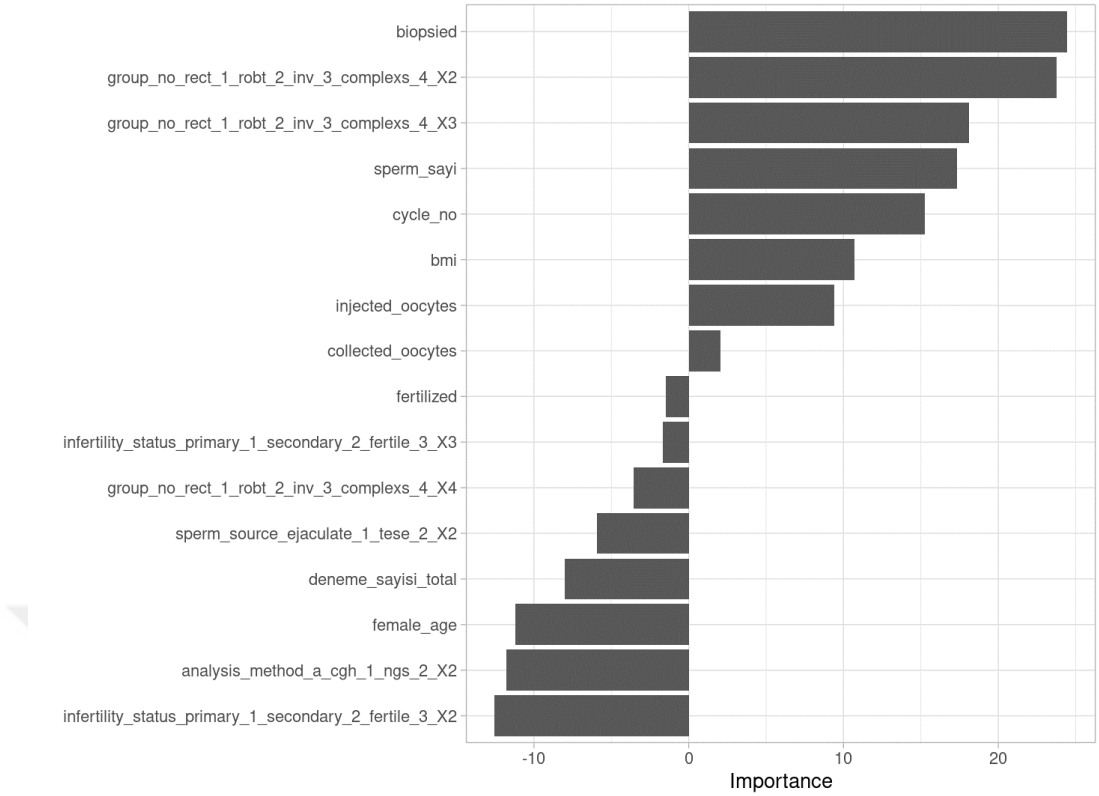


Figure 3.5 The relative importance of post-treatment variables in neural network model

design that focused on the use of analytical methods to interpret statistical results objectively through the analysis of a homogeneous group of patients, interventions, laboratory conditions and standardized data. Another advantage was that both cases and controls were acquired from a single IVF and genetic unit during the same time period. The presence of a control group, where every case has an equivalent two controls matched as closely as possible, matched by both female age, and by the number of embryos analyzed, added statistical power to the study. Matching also by embryo number helped to reduce the potential over-representation of individual high responders as well as reduce potential under-representation of low responders who may have lower number of embryos thus contribute less to the study population. Moreover, assessing each chromosomal pair in detail helped us to build more powerful statistical inferences in large sample size. In this context, the effect size measurement was used for the first time in this study which should have been provided in every study with large sample size (>1000). On the other hand, the tools available in R-program helped to standardize the data more efficiently. Different algorithms that were used in the prediction models helped to predict the probability of embryo transfer with better performance and higher accuracy further with the addition of hyperparameter tunning.

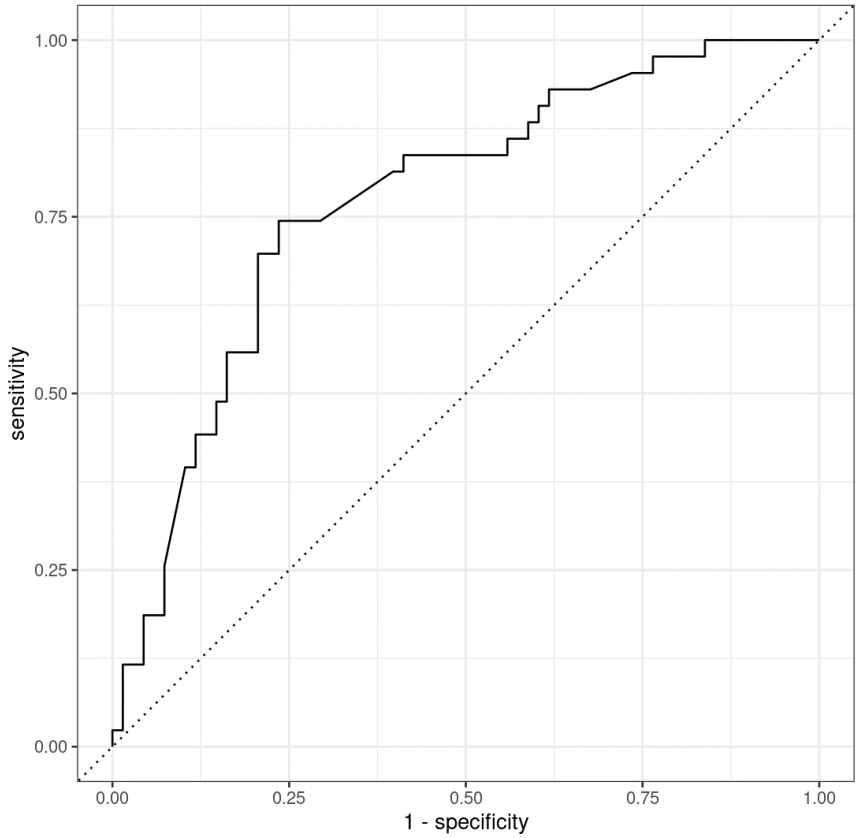


Figure 3.6 ROC Curve of neural network model with post-treatment parameters

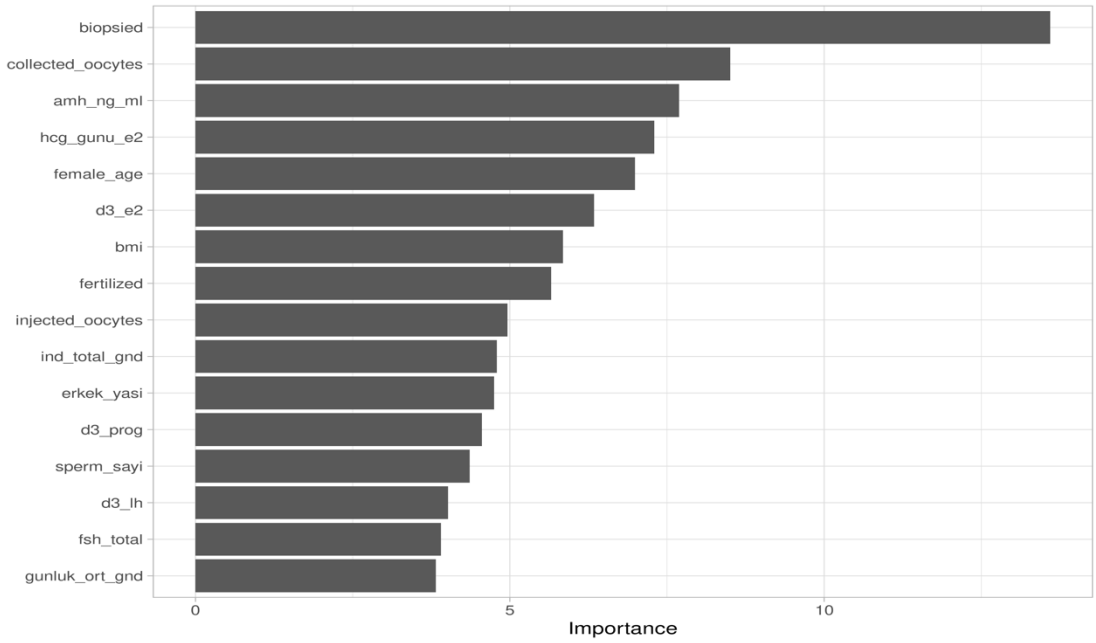


Figure 3.7 The importance of parameters in random forest model after addition of extra hormonal data

3.6 Limitations of the study

There are also some limitations in that techniques used in this study which did not allow us to differentiate embryos in terms of chromosomally normal versus those

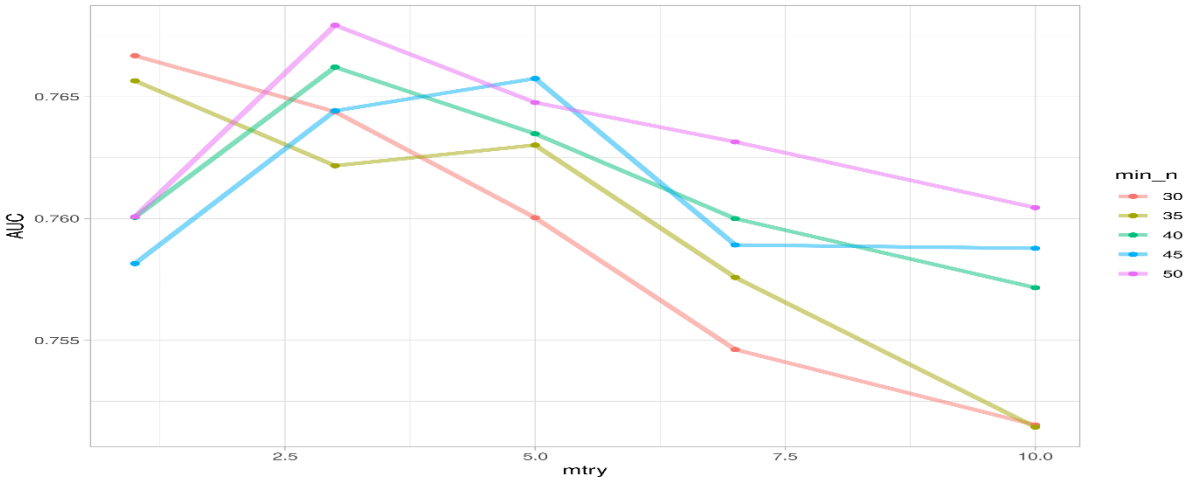


Figure 3.8 AUC graph

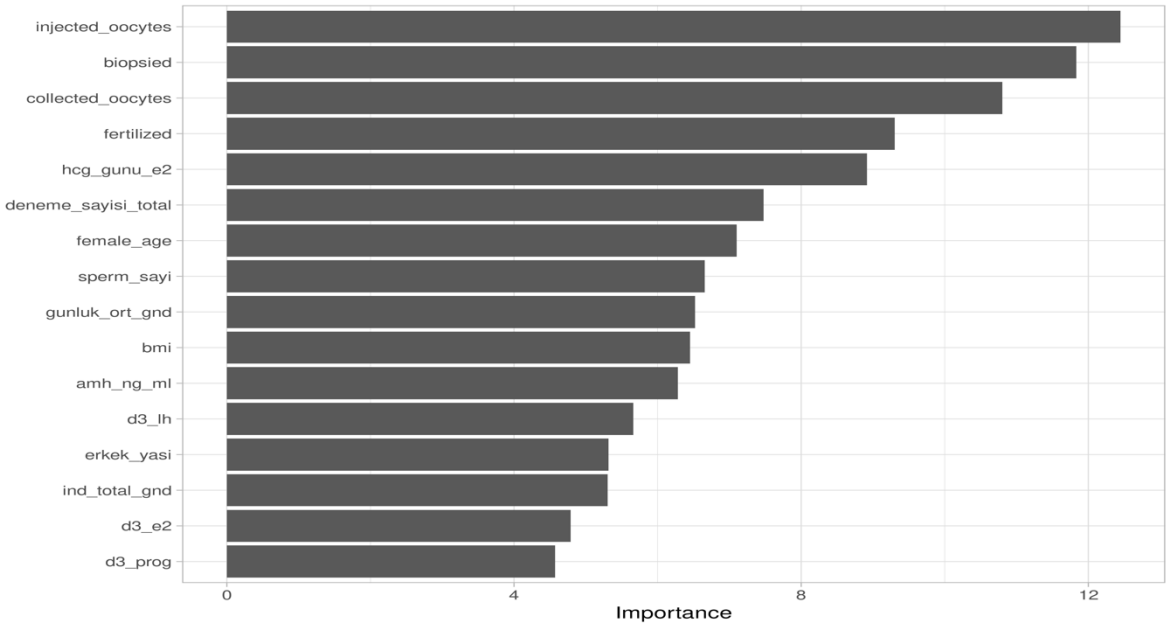


Figure 3.9 The importance of parameteres in random forest model after hyperparameter tuning technique

with balanced rearrangements. Although the idea of eliminating the transmission of balanced carriers to future generations could be beneficial, it is not an obligation of PGT-SR unless there is a gene disrupted or microdeletion is present nearby the translocation breakpoint associated with a clinical condition [39]. However this deselection further reduces the number of transferable embryos by half [70] and couples with low ovarian reserve and advanced maternal age could hardly benefit from this approach.

Although the best possible matched control group was created, there could still be selection bias due to retrospective nature of the study. However, since it is well known that; female age is the most important factor in relation with aneuploidy [71]

and these two groups were perfectly matched by female age and some more other important clinical parameters not only per case but also per embryo (Table 3.6). In most of the cases, the controls were selected among other suitable controls by the help of a web based program, so selection was randomized as much as possible.

In the past, a variety of different approaches for selection of control groups were used. Mostly, age matched couples undergoing PGT-A for infertility were chosen [72, 73] whereas some other studies used couples with monogenic disorders (PGT-M) as controls as their dataset consists mostly of secondary infertile patients. Some studies did not even use controls or the control group was not adequate in size. In such study, 356 PGT-SR case results were compared to 53 PGT-M patients which revealed evidence for ICE for young couples with female age <35, where the incidence of *de novo* aneuploidy was 24.5% in cases and 17.3 % in controls ($p<0.05$) [74]. Despite this apparent statistically significant difference, the same association failed to be demonstrated in the overall group (23.8% vs 22.0%, p=ns). As mentioned previously, in none of the previous studies, the statistically different results were tested with effect size or measuring the strength of the association found. Moreover the term ICE was misinterpreted as the spontaneous aneuploidy in embryos which affects the non-rearranged chromosomes. Most of the previous studies interpreted all *de novo* aneuploidy event as an evidence of ICE, as if there should not be any aneuploidy present in the rearrangement carrier's embryos [75]. However, it must be taken into consideration that every embryo carries a baseline risk of being aneuploid which depends on the infertility status, female age and other clinical parameters. Even considerable proportion of gametes of young donors were shown to carry aneuploidy which supports the notion that there exists a baseline for aneuploidy risk for every patient and every embryo [76] The results of the present study (with a more appropriate control group and robust statistical analysis) not only excludes the hypothesis of ICE in this dataset but also questioned the design and interpretation of the previous work published in this context.

In this study two different testing platforms (aCGH and NGS) were used due to rapid technological breakthrough which was inevitable during the time period in which this study was conducted. However these two methods are quite similar in their performance and resolution in the detection of chromosomal rearrangements. Previously, during the transition periods, some validation studies have been performed to compare these two techniques. According to these studies, the concordance rates of aCGH and NGS techniques have been found as 100% in the overall diagnosis of abnormal samples. They were also concordant for the segmental imbalances bigger than 5Mb [77, 78]. Although NGS system is more dynamic in detecting mosaic abnormalities, the use of two different techniques with similar diagnostic capabilities

should not have created important drawbacks for this study.

Some parameters such as chromosomal breakpoint positions and the involvement of acrocentric chromosomes in RECT were beyond the scope of this study. Nevertheless it was shown that although RECT carriers with acrocentric chromosome involvement had significantly lower incidence of adjacent 1 segregation patterns (29.5% vs 35.2%; $p<0.05$) and a higher incidence of 3:1 segregation patterns (11.2% vs 7.5%; $p<0.05$), the involvement of acrocentric chromosomes does not have any significant impact on the rates of alternate segregation and thus transferable embryos [79]. Embryo morphokinetics is another factor that was not included as a parameter for testing in both groups. Although there exists some evidence of a correlation between the developmental quality of embryos and the gross chromosomal abnormalities; there are contradictory results whether such a correlation exists for segmental imbalances [80, 81].

3.7 Factors affecting segregation

Results demonstrated that the type of rearrangement and (to a lesser extend) the sex of the carrier have an impact on the segregational abnormalities. Furthermore, female age was another factor although it does not have a direct effect on segregation but it decreases the number of available embryos for transfer with increasing the risk of *de novo* aneuploidy (Table 3.2 and 3.4).

3.7.1 The type of rearrangement

In the literature, although the proportion of embryos with alternate segregation (balanced) approximately ranges between 35-50% (RECT), 65-80% (ROBT) and 65-75% (INV), the percentage of transferable embryos (both normal/balanced and euploid) is roughly 10-15 % less, ranging between 20-35% (RECT), 35-50% (ROBT) and 50-60% (INV) [70, 82, 83]. The finding in this study are comparable with the previous data albeit in the lower range for similar female age groups where the proportion of normal/balanced and euploid embryos were 19.4%, 33.2%, 35.8%, 25% for carriers of RECT, ROBT, INV, CCR, respectively **(figure 1)**. In addition, although ROBT and INV carriers seem to have more *de novo* aneuploidy (33.2% and 44.5%) compared to RECT carriers (Table 3.1); when the aneuploidy was assessed in cumulative means (embryos with *de novo* aneuploidy regardless of segregation), the results did not show any difference (Table 3.2). As expected, the cumulative *de novo* aneuploidy rate was significantly higher in couples with advancing female age both in RECT (55.5% vs 40.0%, $p<0.001$) and ROBT (75% vs 43.3%, $p<0.001$) but the same

difference was not pronounced for INV (60% vs 45.1%, p=ns) carriers (Table 3.2).

On the other hand, CCR carriers are at a higher risk of infertility, spontaneous abortion and fetuses with congenital anomalies due to numerous different theoretical segregations leading to unbalanced gametes. They are extremely rare and the majority (70–75%) is *de novo* origin [8]. The chances of identifying normal/balanced blastocysts in patients with CCR are very low (<6%); and greater complexity CCRs result in fewer transplantable embryos [84]. In this study, this rearrangement group was associated with the poorest results such that it has the highest rate of segregational abnormalities (cumulative unbalanced, 65.3%). Due to the low sample size, this group could not be assessed deeply with advanced statistical tools for the impacts of maternal age and sex. However, logistic regression revealed that CCR increases the risk of an embryo being unbalanced by 12.5x compared to embryos from INV carriers (p<0.001, Table 3.3).

In the CCR group, there were 3 insertional, 1 ring chromosome and 7 complex translocation carriers. Being extremely rare, ring chromosomes are formed after fusion of the long and short arms of a chromosome, and are sometimes associated with large deletions that may occur in mosaic state. In this study, all 4 embryos were found to have segregational abnormalities and none of them (0/4) were transferable for the patient with ring chromosome (45,XX,-21[8]/46,XX,r(21)(p13q22.3)[92]). There is not much prior work performed for phenotypically normal patients who underwent PGT-SR for ring chromosomes. Nonetheless, in a case report with a woman carrier of deleted/ring 22 chromosome, among 10 embryos, none of them were normal or balanced for the ring chromosome [85].

ET was cancelled in 7 out of 12 cycles, due to the absence of available embryos for transfer (Table 1). 3 healthy live births were achieved from 4 FET for the following 3 patients with: (1) 46,XX,t(11;19)(p11.2;q12),t(17;19)(q21;p13.1); (2) 46,XX,t(5;15)(q13;q26) and 46,XY,t(5;15)(q13;q26) (where the both partners were carriers of the same translocation; and (3) 46,XY,t(5;16)(q31;p13.1),t(2;14)(q31;q32). In the remaining one cycle with transferable embryos, FET was not performed yet. Clinical results showed once a transferable embryo was found, excellent clinical outcomes were achieved in this group.

Another important finding was the outcomes of INV cases; where the risk of unbalanced segregation (15.7 %) was the lowest among all rearrangement groups. These results are in line with previous findings where only 13% of the embryos were found as unbalanced out of 140 [82]. Inversion carriers have a large range (0-54%)

of different frequencies of unbalanced spermatozoa [86]. Inversions, especially when PAI are considered, are considered as harmless possibly because the risk of formation of an inversion loop and the probability of subsequent crossing overs are very low [7, 87]. On the contrary, there are some rare cases with longer inverted segments and the risk is considerably high; which emphasizes the length of the segment as an important risk factor [88]. In 2007, Morel et al., found a significant correlation between the relative size of the inversion and the frequency of recombination (R=0.76, p=0.001). In their study they did not find any recombinant chromosomes in the cases where inverted segment is less than 30% of the chromosome length, however they claim this association was independent of the size of the inverted segment [89]. In this study, the average size of the inversion was 39.1 Mb among embryos with balanced segregations, and 87.5 Mb among embryos with unbalanced segregations ($p<0.001$). Similarly, the average proportion of the inverted region to the whole chromosome was higher among embryos with unbalanced segregation compared to the ones with balanced segregations (32.9% vs. 51.6%, $p<0.001$). Here, the present data supports the previous results in the sense of there is a similar association between segregational abnormalities and the inverted chromosome size (Table 3.5) which repeated the findings of 'the longer the inverted region the higher the risk of cross-over and segregational abnormalities'.

3.7.2 The effect of sex of the carrier

Previous studies that have assessed the effect of sex on the segregations were contradictory across and within thyself. No difference was found between male and female carriers in terms of segregations [90, 91]. Some of the studies reported higher rates of alternate segregations in embryos with paternal RECT (49.5% vs 41.7%, $p<0.001$) [79]. Likewise, in 2021, Lin et al. found that female translocation carriers had a significantly lower incidence of alternate segregation pattern than male carriers (43.26% versus 47.98%, $p=0.001$), and a higher incidence of 3:1 segregation pattern (6.70% versus 4.29%, $p<0.001$) [92]. Furthermore their data showed that this effect was more pronounced when an acrocentric chromosome was involved, which suggests these factors are affected by a combination of different factors. There are more evidences on the effect of sex in embryos from ROBT carriers. According to Zhang et al., [93] the percentage of embryos with alternate segregations for ROBT carriers are higher in male carriers (82.9% vs 55.2%, male vs female, $p<0.001$). In another recent study, Liu et al found that female carriers of both RECT and ROBT had a significantly lower (23.3 vs. 42.4%, 34.7 vs. 54.7%) percentage of normal/balanced embryos than male carriers, respectively [94]. Similarly, in the present study, sex was found to be associated with the segregational outcomes, such that; female carriers were

1.29x more likely to produce unbalanced embryos compared to male carriers ($p<0.05$) (Table 3.3). Nearly all previous evidence showed male carriers are advantageous over females due to higher proportion of alternate segregations which could be attributed both the number of gametes they can produce in a lifetime and the elimination of chromosomally unbalanced gametes by the checkpoint mechanisms in meiosis, in which it is known to be more error-prone in female counterpart [11, 95]. Sperm FISH technique is a good and a more direct approach which allows the analysis of meiotic abnormalities in sperm in the investigation of personal risks for male carriers. In a study where sperm FISH results were compared with their corresponding embryo results after PGT-SR, it was shown that sperm FISH provides a good way of prediction of the segregational abnormalities for male carriers [96]. The incidence of unbalanced segregations ranges between 19- 91%; 7- 40% and 0-54% for RECT, ROBT and INV carriers respectively [86]. Not surprisingly, different carriers of the same familial translocation from the same family show similar rates of segregational abnormalities which suggests that the risk of meiotic imbalance depends also on the nature of the chromosomes involved and the breakpoint positions [97].

3.8 Clinical outcomes

In terms of clinical outcomes, the most limiting factor was the availability of a transferable embryo where it was lowest when the proportion of cycles with cancellation of transfer were high. The cycles that did not reach for embryo transfer were 66.7% for CCR, 47.9% for RECT, 36% for ROBT and 22.5% for INV carriers ($p<0.05$). In the presence of a transferable embryo (normal/balanced and euploid), good clinical outcomes with CPR (69.5%) and LBR (55.8%) were achieved independent of female age (Table 3.2). In the literature, clinical outcomes of PGT-A cycles were highly affected by the stage of biopsy, analysis method and the patient population. Therefore CPR was highly variable and reported in a high range (34-74%) reviewed in [5]. PGT-SR has similar technical evolution as PGT-A, where the changes in the techniques such as the downward trend of FISH as a diagnostic method in parallel with the rise of comprehensive analysis methods such as aCGH, NGS and SNP-arrays and popularisation of blastocyst stage biopsy as the biopsy of choice has resulted in better outcomes in terms of diagnostics [98]. This also helped to improve clinical outcomes where CPR of SR carriers dramatically increased from 39% to 74% with the use of new approaches [99]. Even though conducted with FISH technique, in 2005, Verlinsky et al., could show a dramatic increase in take home baby rate (THBR) from 11.5 to 81.4% and decrease in miscarriage rate from 87.8% to 17.8% [100]. Similarly, Huang et al., reported another striking result of a retrospective comparison study of reproductive outcomes of 194 RECT carriers who had previously experienced

recurrent pregnancy losses or babies born with birth defects before and after PGT-SR. According to that study, healthy live birth rate increased from 2.9% (before PGT-SR) to 85.6% after PGT-SR [101]. However there are some contradictory reports as well; a meta-analysis revealed that natural conception offers similar pregnancy outcomes compared with IVF–PGD, however in this study, the carriers were selected from patients with recurrent miscarriages which might not represent the whole population of carriers [102]. Therefore, further studies with better designs, outcome measures and patient selection are needed in order to show the real benefit of PGT-SR in patients.

3.9 Rejection of the hypothesis of an inter-chromosomal effect (ICE)

ICE can be simply described as the increase in the likelihood of further chromosome abnormalities such as (segmental or whole chromosome) aneuploidy by the presence of one chromosome abnormality (such as structural abnormality in PGT-SR cases). This theory is based on the assumption that rearranged chromosomes might impact the segregation, pairing and disjunction of other chromosomes during meiosis. It is however speculated that the possibility of heterosynapsis or presence of asynaptic regions might cause an increased risk of meiotic arrest or numerical abnormalities in the resulting gametes (reviewed in [103]). First postulated in 1963 by Leujene (of Down Syndrome/Trisomy 21 which was based on the observation that there was a higher risk of having children with Down Syndrome among carriers of rearrangements of chromosomes not related to "chromosome 21" [51]. After this study, this hypothesis was tested by Warburton in 1985, with the retrospective analysis of amniocentesis data of more than 1300 pregnancies in which one of the parents were a known carrier of a RECT [104]. The evidence did not support an increased frequency of nondisjunction of other chromosomes in the presence of a translocation. With more global figures, data from a registry of individuals with trisomy 21 and of nonviable pregnancies with a trisomy did not show an increased frequency of inherited parental translocations (reviewed in [105]). The same hypothesis was also tested in sperm samples of SR carriers and in embryos from PGT-SR couples. Results are variable with conflicting data and interpretations were controversial. In general, these studies have not supported existence of an ICE, however there were some associations with increase of aneuploidy in carriers with compromised semen parameters which might indicate that poor sperm parameters can be attributed as a possible cause for the relative increase of *de novo* aneuploidies rather than the rearrangement itself [86, 106]. Similarly, in translocation carriers, who appeared fertile, the segregation patterns of ten chromosome pairs by sperm FISH from nine carriers did not show any evidence of ICE [106]. However, some studies suggest that there might be a possible ICE which

could be chromosome specific or patient specific [107–109].

Moreover, interpretation failures might also contribute to conflicting results where minute differences in the percentage values (i.e. 0.60% vs 0.39%) have been reported with statistically significant difference just because the P value was smaller than the alpha level which is generally assumed as 0.05 [107, 110]. With a sufficiently large sample, a statistical test will almost always demonstrate a significant difference, thus, reporting only the significant P value is not adequate for analytical interpretation of the results [64]. This lack of appropriate interpretation shows itself in large cohort studies. A previous large cohort study with approximately 90,000 subjects investigated the effects of taking aspirin on reducing the risk of myocardial infarction (MI) could be given as an example of how misinterpretation of the data could affect the health of millions of people in a negative way increasing the risk of adverse effects by the so-called supporting evidences for unnecessary treatments [111].

Effect size is an essential component when evaluating the strength of a statistical claim, especially while working with large datasets [64]. Therefore, both statistical significance (P value) and the substantive significance (effect size) were reported. In this study, effect size measurements were performed to reveal any practical significance of associations for variables with large sample size (i.e. >1000).

In the present dataset, analysis of over 5 thousand embryos and 117 thousand chromosomal pairs revealed apparently statistically significant difference in the form of p-values <0.05. A deeper analysis of effect size however suggest that any association is, at best negligible (Table 3.7 and 3.8). Although the realistic existence of an ICE overall is unlikely, this data does not preclude the possibility of a patient-specific ICE, where higher incidences in some particular patients and effects on particular chromosomes might be present. Another finding was that; the incidences of aneuploidy in chromosomes 16 and 22 were found to be higher than the rest (in both cases and controls). These results were similar to previous studies where a large number of embryos were analyzed by comprehensive analysis methods [13, 112].

In addition to the mentioned problems above, confusions about the terminology of "ICE" has also contributed to the misleading results in the literature. One such example is referring all *de novo* aneuploidies in embryos of rearrangement carriers as "ICE" and aneuploidies in balanced embryos as "pure ICE" without a priori comparisons with an age matched control group [75, 82]. Scriven (2021) [113] refuted the findings of Boynukalin et al., [75] who suggested significant interchromosomal effect associated with RobTs given that they observed a higher incidence of normal/balanced embryo diagnoses having aneuploidy of an unrelated chromosome as compared to RecT. That

is, reanalysis of the same data suggests the regression coefficient of the equation (line slope) actually indicates the expected incidence of unrelated chromosomes. He also performs a similar analysis on another earlier study, suggesting that level of aneuploidy for chromosomes other than those involved in the translocation was as expected, both for RECT and ROBT [113]. These incidences were falling in an expected range for a comparable age matched IVF group as reported in a recent trial [114]. Aneuploidy is common in early embryonic development [115] and may affect a significant proportion of embryos (ranging from 17.5% to 60.5%) even in young donors [76]. This suggests that there might be a baseline risk for aneuploidy for each chromosome which might be related to biological factors or other factors that have not yet been identified. As expected, studies which used better designs with control groups often failed to provide any evidence of ICE [72]. In such study, using age matched control group, they found an evidence on the opposite direction of ICE, with cumulative aneuploidy lower than controls (44.5% RECT carriers vs 53.3% in controls). This is most likely caused by over-representation of the chromosomes involved in the translocation skewing the overall percentages which we have overcome by establishing a control group by weighing the whole population by matching the cases also with their number of embryos contributed. Similar results were achieved for INV carriers such that the average error rates of all chromosome pairs of INV carriers were similar to the control group (0.0087 vs 0.0089) with no significant difference ($p>0.05$) [73], and overall incidence of aneuploidy was not significantly higher for the inversion patients compared to the controls (48.8% vs 47.2%, ns) [112] again emphasizing that the rearrangement itself does not elevate the likelihood of aneuploidy above that of the risk associated with maternal age [73, 112].

3.10 Predictive modelling via machine learning

Modern approaches in medicine state as “treat the patient not the disease” which emphasized the need to establish personalized treatment strategies which involve the tailoring of medical treatment to the individual characteristics of each patient. For that reason, patients first need to be classified into sub-populations that differ in their susceptibility to a particular disease, in the biology or prognosis of those diseases they may develop, or in their response to a specific treatment.

It is estimated that more than 80 million couples are affected by infertility [116]. IVF treatments is one of the most difficult to predict due to a high number of factors involved. Predictions are usually made by clinicians on the basis of clinical experience or gut-feeling which might not be reliable and accurate. However, the couples need to be informed well about their chances of success before undergoing the treatments

since these treatments can be stressful both emotionally and financially and may pose health risks. For the aforementioned reasons, the successful prediction of IVF outcome has long been a goal in medicine and would be a game-changer when achieved. Since the development of artificial intelligence and the use of machine learning algorithms in medicine; prediction models were established with the use of different platforms with the aim of assisting clinicians in tailoring personalized treatment of infertile patients and improve assisted reproduction outcome. There are mainly three phases of establishing a prediction model; (1) model derivation which includes the identification of predictor parameters; (2) model validation which includes testing the model performance on new patients from the same centre (internal validation) and on other datasets (external validation); lastly the impact analysis which includes testing whether the prediction model improves decisions in terms of cost-effectivity and patient care.

The previous studies mostly concentrated on the general population of infertile patients undergoing IVF treatment and used the IVF databases related with male and female medical history and clinical examinations [71, 117–119]. The model developed in this thesis is the first one designed specifically for rearrangement carriers undergoing IVF treatments coupled with PGT using R software.

The strengths of this study were as follows: the most important is the dataset belonged to one of the biggest IVF clinics in the region which was accredited by external and internal bodies with high laboratory quality standards. The data belonged to the single IVF unit with a large sample size. Usage of advanced machine learning algorithms in R-software such as random forest and k-nearest neighbour (KNN) neural network model the use of hyperparameter tuning facilities were other technical strengths of this study.

There are also limitations such as the missing data in some of the parameters due to absence of information or lost to follow up; sub-standard entry of parameters to the system by different users and the accuracy of some clinical information such as previous history of infertility was limited to the declarations of the patients which might have increased the risk of faulty training of the system during data input.

The two most important parameters for prediction are the probability of finding genetically suitable embryos for transfer -in other words, the probability of embryo transfer- and achievement of healthy live birth. This model was developed to predict the embryo transfer. With this model, we were able to establish a success rate in predicting the possibility of at least one transferable embryo with 74.7% accuracy and 78.4% ROC AUC which could be considered to be in a very good range compared

to previous studies [118, 119]. However, the model should be tested internally and externally for validation.

Conclusions

These results represent one of the largest investigations of PGT-SR practice to date and provide a reference point for future PGT-SR studies. It is a comprehensive study which combined statistical analyses with machine learning algorithms in order to predict the outcome of both genetic results per embryo and clinical outcomes per cycle. It is the first study which used advanced methods such as effect size in order to test the ICE hypothesis. In summary, for SR carriers the most important factors that affect segregations were rearrangement type and sex. Female age was the dominant impact factor in association with *de novo* aneuploidy. Although the "ICE" hypothesis was rejected, these results nonetheless illustrate the importance of comprehensive chromosomal screening of all chromosomes in parallel with rearrangement related chromosome testing. The prediction models established by using machine learning algorithms based on this dataset were highly successful in the prediction of whether a couple will achieve an embryo transfer or not. There are some more ongoing projects using the same database, such as establishing a live birth prediction model for all IVF patients including PGT-A and PGT-SR patients. We look forward to testing these models in other databases as well in order to create a more universal model and a web-based product that can be helpful for physicians which will pave the way to build a better personalized treatment strategies and can also be helpful for couples in their family planning decision making. These are the future goals of this thesis.

REFERENCES

[1] P. A. Jacobs, C. Browne, N. Gregson, C. Joyce, H. White, "Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding.," *Journal of Medical Genetics*, vol. 29, no. 2, pp. 103–108, Feb. 1992.

[2] M. De Braekeleer, T.-N. Dao, "Cytogenetic studies in male infertifity," *Human reproduction*, vol. 6, no. 2, pp. 245–250, 1991.

[3] M. J. Pettenati *et al.*, "Paracentric inversions in humans: A review of 446 paracentric inversions with presentation of 120 new cases," *American Journal of Medical Genetics*, vol. 55, no. 2, pp. 171–187, Jan. 1995.

[4] M. Braekeleer, T.-N. Dao, "Cytogenetic studies in couples experiencing repeated pregnancy losses," *Human Reproduction*, vol. 5, no. 5, pp. 519–528, Jul. 1990.

[5] C. Ogur, D. K. Griffin, "Preimplantation genetic testing for structural rearrangements," in *Preimplantation Genetic Testing*, D. Griffin, G. Harton, Eds., 1st, CRC Press, 2020, pp. 49–76.

[6] R. Gardner, G. R. Sutherland, L. G. Shaffer, *Chromosome Abnormalities and Genetic Counseling*. Oxford University Press, Nov. 2011, ISBN: 9780195375336.

[7] M. Jaarola, R. H. Martin, T. Ashley, "Direct evidence for suppression of recombination within two pericentric inversions in humans: A new sperm-FISH technique," *The American Journal of Human Genetics*, vol. 63, no. 1, pp. 218–224, Jul. 1998.

[8] F. Pellestor *et al.*, "Complex chromosomal rearrangements: Origin and meiotic behavior," *Human Reproduction Update*, vol. 17, no. 4, pp. 476–494, Apr. 2011.

[9] P. N. Scriven, S. M. Bint, A. F. Davies, C. M. Ogilvie, "Meiotic outcomes of three-way translocations ascertained in cleavage-stage embryos: Refinement of reproductive risks and implications for PGD," *European Journal of Human Genetics*, vol. 22, no. 6, pp. 748–753, Oct. 2013.

[10] A. H. Handyside, E. H. Kontogianni, K. Hardy, R. M. L. Winston, "Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification," *Nature*, vol. 344, no. 6268, pp. 768–770, Apr. 1990.

[11] T. Hassold, P. Hunt, "To err (meiotically) is human: The genesis of human aneuploidy," *Nature Reviews Genetics*, vol. 2, no. 4, pp. 280–291, Apr. 2001.

[12] M. Benkhalifa *et al.*, "Array comparative genomic hybridization profiling of first-trimester spontaneous abortions that fail to grow in vitro," *Prenatal Diagnosis*, vol. 25, no. 10, pp. 894–900, 2005.

[13] J. M. Franasiak *et al.*, "Aneuploidy across individual chromosomes at the embryonic level in trophectoderm biopsies: Changes with patient age and chromosome structure," *Journal of Assisted Reproduction and Genetics*, vol. 31, no. 11, pp. 1501–1509, Sep. 2014.

[14] T. H. Taylor, S. A. Gitlin, J. L. Patrick, J. L. Crain, J. M. Wilson, D. K. Griffin, "The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans," *Human Reproduction Update*, vol. 20, no. 4, pp. 571–581, Mar. 2014.

[15] J. Harper *et al.*, "The ESHRE PGD consortium: 10 years of data collection," *Human Reproduction Update*, vol. 18, no. 3, pp. 234–247, Feb. 2012.

[16] N. Gleicher, D. H. Barad, P. Patrizio, R. Orvieto, "We have reached a dead end for preimplantation genetic testing for aneuploidy," *Human Reproduction*, vol. 37, no. 12, pp. 2730–2734, Mar. 2022.

[17] P. N. Scriven, F. A. Flinter, Y. Khalaf, A. Lashwood, C. M. Ogilvie, "Benefits and drawbacks of preimplantation genetic diagnosis (PGD) for reciprocal translocations: Lessons from a prospective cohort study," *European Journal of Human Genetics*, vol. 21, no. 10, pp. 1035–1041, Feb. 2013.

[18] P. R. Brinsden, P. R. Brinsden, "Thirty years of IVF: The legacy of Patrick Steptoe and Robert Edwards," *Human Fertility*, vol. 12, no. 3, pp. 137–143, Jan. 2009.

[19] A. Pacchiarotti *et al.*, "Ovarian stimulation protocol in IVF: An up-to-date review of the literature," *Current Pharmaceutical Biotechnology*, vol. 17, no. 4, pp. 303–315, Feb. 2016.

[20] A. V. Steirteghem *et al.*, "Intracytoplasmic sperm injection," *Bailliére's Clinical Obstetrics and Gynaecology*, vol. 8, no. 1, pp. 85–93, Mar. 1994.

[21] You & Your Hormones. "In vitro fertilisation treatment (IVF)." (2021), [Online]. Available: `https://www.yourhormones.info/topical-issues/in-vitro-fertilisation-treatment-ivf/` (visited on 02/04/2023).

[22] Y. Verlinsky, A. Kuliev, "Preimplantation polar body diagnosis," *Biochemical and Molecular Medicine*, vol. 58, no. 1, pp. 13–17, Jun. 1996.

[23] R. T. Scott, K. M. Upham, E. J. Forman, T. Zhao, N. R. Treff, "Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: A randomized and paired clinical trial," *Fertility and Sterility*, vol. 100, no. 3, pp. 624–630, Sep. 2013.

[24] S. J. McArthur, D. Leigh, J. T. Marshall, K. A. de Boer, R. P. Jansen, "Pregnancies and live births after trophectoderm biopsy and preimplantation genetic testing of human blastocysts," *Fertility and Sterility*, vol. 84, no. 6, pp. 1628–1636, Dec. 2005.

[25] L. Gianaroli *et al.*, "Blastocentesis: A source of DNA for preimplantation genetic testing. results from a pilot study," *Fertility and Sterility*, vol. 102, no. 6, 1692–1699.e6, Dec. 2014.

[26] M. C. Magli *et al.*, "Preimplantation genetic testing: Polar bodies, blastomeres, trophectoderm cells, or blastocoelic fluid?" *Fertility and Sterility*, vol. 105, no. 3, 676–683.e5, Mar. 2016.

[27] J. Jiao *et al.*, "Minimally invasive preimplantation genetic testing using blastocyst culture medium," *Human Reproduction*, vol. 34, no. 7, pp. 1369–1379, Jun. 2019.

[28] D. K. Griffin, C. Ogur, "Chromosomal analysis in IVF: Just how useful is it?" *Reproduction*, vol. 156, no. 1, F29–F50, Jul. 2018.

[29] D. Griffin, A. Handyside, R. Penketh, R. Winston, J. Delhanty, "Fluorescent in-situ hybridization to interphase nuclei of human preimplantation embryos with X and Y chromosome specific probes," *Human Reproduction*, vol. 6, no. 1, pp. 101–105, Jan. 1991.

[30] P. Lichter, S. Joos, M. Bentz, S. Lampel, "Comparative genomic hybridization: Uses and limitations," *Seminars in Hematology*, vol. 37, no. 4, pp. 348–357, Oct. 2000.

[31] D. Wells, "Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation," *Nucleic Acids Research*, vol. 27, no. 4, pp. 1214–1218, Feb. 1999.

[32] D. Wells, T. Escudero, B. Levy, K. Hirschhorn, J. D. Delhanty, S. Munné, "First clinical application of comparative genomic hybridization and polar body testing for preimplantation genetic diagnosis of aneuploidy," *Fertility and Sterility*, vol. 78, no. 3, pp. 543–549, Sep. 2002.

[33] L. Wilton, "Preimplantation genetic diagnosis and chromosome analysis of blastomeres using comparative genomic hybridization," *Human Reproduction Update*, vol. 11, no. 1, pp. 33–41, Jan. 2005.

[34] F. Fiorentino *et al.*, "PGD for reciprocal and Robertsonian translocations using array comparative genomic hybridization," *Human Reproduction*, vol. 26, no. 7, pp. 1925–1935, Apr. 2011.

[35] T. Nakano, M. Ammae, M. Satoh, S. Mizuno, Y. Nakaoka, Y. Morimoto, "Analysis of clinical outcomes and meiotic segregation modes following preimplantation genetic testing for structural rearrangements using aCGH/NGS in couples with balanced chromosome rearrangement," *Reproductive Medicine and Biology*, vol. 21, no. 1, Jan. 2022.

[36] S. Chen *et al.*, "Comprehensive preimplantation genetic testing by massively parallel sequencing," *Human Reproduction*, Dec. 2020.

[37] N. R. Treff, L. E. Northrop, K. Kasabwala, J. Su, B. Levy, R. T. Scott, "Single nucleotide polymorphism microarray–based concurrent screening of 24-chromosome aneuploidy and unbalanced translocations in preimplantation human embryos," *Fertility and Sterility*, vol. 95, no. 5, 1606–1612.e2, Apr. 2011.

[38] A. H. Handyside *et al.*, "Karyomapping: A universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes," *Journal of Medical Genetics*, vol. 47, no. 10, pp. 651–658, Oct. 2009.

[39] N. R. Treff *et al.*, "SNP array-based analyses of unbalanced embryos as a reference to distinguish between balanced translocation carrier and normal blastocysts," *Journal of Assisted Reproduction and Genetics*, vol. 33, no. 8, pp. 1115–1119, May 2016.

[40] S. A. Natesan *et al.*, "Live birth after PGD with confirmation by a comprehensive approach (karyomapping) for simultaneous detection of monogenic and chromosomal disorders," *Reproductive BioMedicine Online*, vol. 29, no. 5, pp. 600–605, Nov. 2014.

[41] C. E. Beyer, A. Lewis, E. Willats, J. Mullen, "Preimplantation genetic testing using karyomapping for a paternally inherited reciprocal translocation: A case study," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 5, pp. 951–963, Feb. 2019.

[42] Y. Xi, H. Wang, N. Sun, "Machine learning outperforms traditional logistic regression and offers new possibilities for cardiovascular risk prediction: A study involving 143, 043 Chinese patients with hypertension," *Frontiers in Cardiovascular Medicine*, vol. 9, Nov. 2022.

[43] M. Meseguer, J. Herrero, A. Tejera, K. M. Hilligsoe, N. B. Ramsing, J. Remohi, "The use of morphokinetics as a predictor of embryo implantation," *Human Reproduction*, vol. 26, no. 10, pp. 2658–2671, Aug. 2011.

[44] A. G, B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 127–130, Jun. 2022.

[45] B. A. Akinnuwesi *et al.*, "Application of support vector machine algorithm for early differential diagnosis of prostate cancer," *Data Science and Management*, Oct. 2022.

[46] M. Langarizadeh, F. Moghbeli, "Applying naive bayesian networks to disease prediction: A systematic review," *Acta Informatica Medica*, vol. 24, no. 5, p. 364, 2016.

[47] A. Elhazmi *et al.*, "Machine learning decision tree algorithm role for predicting mortality in critically ill adult COVID-19 patients admitted to the ICU," *Journal of Infection and Public Health*, vol. 15, no. 7, pp. 826–834, Jul. 2022.

[48] R. Rabiei, "Prediction of breast cancer using machine learning approaches," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 3, Jul. 2022.

[49] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, J. Havel, "Artificial neural networks in medical diagnosis," *Journal of Applied Biomedicine*, vol. 11, no. 2, pp. 47–58, Jul. 2013.

[50] T. la Duyet. "Neural networks vs random forest." (2020), [Online]. Available: `https://blog.duyet.net/2020/05/tldr-why-rf.html` (visited on 02/23/2023).

[51] J. Lejeune, "Autosomal disorders," *Pediatrics*, vol. 32, no. 3, pp. 326–337, 1963.

[52] S. Kahraman, M. Cetinkaya, B. Yuksel, M. Yesil, C. P. Cetinkaya, "The birth of a baby with mosaicism resulting from a known mosaic embryo transfer: A case report," *Human Reproduction*, vol. 35, no. 3, pp. 727–733, Mar. 2020.

[53] Ç. Ogur *et al.*, "PGT for chromosomal structural rearrangements in 300 couples reveals specific risk factors but an interchromosomal effect is unlikely," *Reproductive BioMedicine Online*, Jul. 2022.

[54] S. Kahraman, "Biopsy techniques," in *PGDIS hands-on workshop on PGD for genetic and chromosomal disorders by blastomere and trophectoderm biopsy on 07 – 09 May 2015, Istanbul Memorial Hospital, Istanbul, TURKEY*, ser. 8th Hands-on Workshop, 2015.

[55] S. Kahraman, A. P. Cil, C. Ogur, A. Semiz, C. Yilanlioglu, "Probability of finding at least one euploid embryo and the euploidy rate according to the number of retrieved oocytes and female age using FISH and array CGH," *Journal of Reproductive Biotechnology and Fertility*, vol. 5, p. 205 891 581 665 327, Jan. 2016.

[56] Manualzz. "24sure microarray single channel reference guide." (2015), [Online]. Available: `https://manualzz.com/doc/13099519/24sure-microarray-single-channel-reference-guide` (visited on 02/04/2023).

[57] T. F. Scientific. "Ion reproseq PGS kits user guide." (2018), [Online]. Available: `https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0016712` (visited on 02/04/2023).

[58] Wikipedia. "Ion semiconductor sequencing." (2023), [Online]. Available: `https://en.wikipedia.org/wiki/Ion_semiconductor_sequencing` (visited on 02/04/2023).

[59] F. Zegers-Hochschild *et al.*, "The international glossary on infertility and fertility care, 2017," *Fertility and Sterility*, vol. 108, no. 3, pp. 393–406, Sep. 2017.

[60] Z. P. Demko, A. L. Simon, R. C. McCoy, D. A. Petrov, M. Rabinowitz, "Effects of maternal age on euploidy rates in a large cohort of embryos analyzed with 24-chromosome single-nucleotide polymorphism-based preimplantation genetic screening," *Fertility and Sterility*, vol. 105, no. 5, pp. 1307–1313, May 2016.

[61] J. McGowan-Jordan, A. Simons, M. Schmid, "ISCN 2016: An international system for human cytogenomic nomenclature (2016)," 2016.

[62] S. Lewallen, P. Courtright, "Epidemiology in practice: Case-control studies," *Community eye health*, vol. 11, no. 28, p. 57, 1998.

[63] Maple Tech. International LLC. "Random number generator." (2008), [Online]. Available: `https://www.calculator.net/random-number-generator.html` (visited on 02/04/2023).

[64] G. M. Sullivan, R. Feinn, "Using effect size or why the p-value is not enough," *Journal of Graduate Medical Education*, vol. 4, no. 3, pp. 279–282, Sep. 2012.

[65] K. Kelley, K. J. Preacher, "On effect size.," *Psychological Methods*, vol. 17, no. 2, pp. 137–152, 2012.

[66] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs," *Frontiers in Psychology*, vol. 4, 2013.

[67] D. K. Lee, "Alternatives to p value: Confidence interval and effect size," *Korean Journal of Anesthesiology*, vol. 69, no. 6, p. 555, 2016.

[68] J. Romano, J. D. Kromrey, J. Coraggio, J. Skowronek, "Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys," in *annual meeting of the Florida Association of Institutional Research*, vol. 177, 2006, p. 34.

[69] L. M. Rea, R. A. Parker, *Designing and conducting survey research: A comprehensive guide*. John Wiley & Sons, 2014.

[70] D. Cheng *et al.*, "Clinical outcomes following preimplantation genetic testing and microdissecting junction region in couples with balanced chromosome rearrangement," *Journal of Assisted Reproduction and Genetics*, vol. 38, no. 3, pp. 735–742, Jan. 2021.

[71] S. C. Esteves, J. F. Carvalho, F. C. Bento, J. Santos, "A novel predictive model to estimate the number of mature oocytes required for obtaining at least one euploid blastocyst for transfer in couples undergoing in vitro fertilization/intracytoplasmic sperm injection: The ART calculator," *Frontiers in Endocrinology*, vol. 10, Feb. 2019.

[72] Y. Xie *et al.*, "Preliminary analysis of numerical chromosome abnormalities in reciprocal and robertsonian translocation preimplantation genetic diagnosis cases with 24-chromosomal analysis with an aCGH/SNP microarray," *Journal of Assisted Reproduction and Genetics*, vol. 35, no. 1, pp. 177–186, Sep. 2017.

[73] P. Xie *et al.*, "Retrospective analysis of meiotic segregation pattern and interchromosomal effects in blastocysts from inversion preimplantation genetic testing cycles," *Fertility and Sterility*, vol. 112, no. 2, 336–342.e3, Aug. 2019.

[74] S. Zhang *et al.*, "Analysis of segregation patterns of quadrivalent structures and the effect on genome stability during meiosis in reciprocal translocation carriers," *Human Reproduction*, vol. 33, no. 4, pp. 757–767, Feb. 2018.

[75] F. K. Boynukalin *et al.*, "The impact of patient, embryo, and translocation characteristics on the ploidy status of young couples undergoing preimplantation genetic testing for structural rearrangements (PGT-SR) by next generation sequencing (NGS)," *Journal of Assisted Reproduction and Genetics*, vol. 38, no. 2, pp. 387–396, Jan. 2021.

[76] S. Munné, M. Alikani, L. Ribustello, P. Colls, P. A. Martıénez-Ortiz, D. M. and, "Euploidy rates in donor egg cycles significantly differ between fertility centers," *Human Reproduction*, vol. 32, no. 4, pp. 743–749, Mar. 2017.

[77] J. F. Chow, W. S. Yeung, V. C. Lee, E. Y. Lau, E. H. Ng, "Evaluation of preimplantation genetic testing for chromosomal structural rearrangement by a commonly used next generation sequencing workflow," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 224, pp. 66–73, May 2018.

[78] C. Cuman *et al.*, "Defining the limits of detection for chromosome rearrangements in the preimplantation embryo using next generation sequencing," *Human Reproduction*, vol. 33, no. 8, pp. 1566–1576, Jul. 2018.

[79] L. Zhang *et al.*, "Interaction of acrocentric chromosome involved in translocation and sex of the carrier influences the proportion of alternate segregation in autosomal reciprocal translocations," *Human Reproduction*, vol. 34, no. 2, pp. 380–387, Dec. 2018.

[80] S. Evsikov, J. Cieslak, Y. Verlinsky, "Effect of chromosomal translocations on the development of preimplantation human embryos in vitro," *Fertility and Sterility*, vol. 74, no. 4, pp. 672–677, Oct. 2000.

[81] G. Li *et al.*, "The influence of chromosomal polymorphism on embryo development and embryonic molecular karyotype in preimplantation genetic testing for chromosomal translocation," *Frontiers in Physiology*, vol. 11, Nov. 2020.

[82] E. Mateu-Brull *et al.*, "Interchromosomal effect in carriers of translocations and inversions assessed by preimplantation genetic testing for structural rearrangements (PGT-SR)," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 12, pp. 2547–2555, Nov. 2019.

[83] Y. Tan *et al.*, "Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing," *GigaScience*, vol. 3, no. 1, Dec. 2014.

[84] L. Hu *et al.*, "Clinical outcomes in carriers of complex chromosomal rearrangements: A retrospective analysis of comprehensive chromosome screening results in seven cases," *Fertility and Sterility*, vol. 109, no. 3, pp. 486–492, Mar. 2018.

[85] A. Mantzouratou, A. Mania, M. Apergi, S. Laver, P. Serhal, J. Delhanty, "Meiotic and mitotic behaviour of a ring/deleted chromosome 22 in human embryos determined by preimplantation genetic diagnosis for a maternal carrier," *Molecular Cytogenetics*, vol. 2, no. 1, p. 3, 2009.

[86] R. Martin, "Cytogenetic determinants of male fertility," *Human Reproduction Update*, vol. 14, no. 4, pp. 379–390, Apr. 2008.

[87] S. Bhatt *et al.*, "Breakpoint mapping and complete analysis of meiotic segregation patterns in three men heterozygous for paracentric inversions," *European Journal of Human Genetics*, vol. 17, no. 1, pp. 44–50, Aug. 2008.

[88] C. Yapan, C. Beyazyurek, C. Ekmekci, S. Kahraman, "The largest paracentric inversion, the highest rate of recombinant spermatozoa. case report: 46, XY, inv(2)(q21.2q37.3) and literature review," *Balkan Journal of Medical Genetics*, vol. 17, no. 1, pp. 55–61, Jun. 2014.

[89] F. Morel *et al.*, "Meiotic segregation analysis in spermatozoa of pericentric inversion carriers using fluorescence in-situ hybridization," *Human Reproduction*, vol. 22, no. 1, pp. 136–141, Aug. 2006.

[90] Y. Cai *et al.*, "Evaluation of preimplantation genetic testing based on next-generation sequencing for balanced reciprocal translocation carriers," *Reproductive BioMedicine Online*, vol. 38, no. 5, pp. 669–675, May 2019.

[91] J. Wang *et al.*, "Analysis of meiotic segregation modes in biopsied blastocysts from preimplantation genetic testing cycles of reciprocal translocations," *Molecular Cytogenetics*, vol. 12, no. 1, Feb. 2019.

[92] L. Lin *et al.*, "Effect of carriers' sex on meiotic segregation patterns and chromosome stability of reciprocal translocations," *Reproductive BioMedicine Online*, vol. 43, no. 6, pp. 1011–1018, Dec. 2021.

[93] L. Zhang, W. Jiang, Y. Zhu, H. Chen, J. Yan, Z.-J. Chen, "Effects of a carrier's sex and age on the segregation patterns of the trivalent of robertsonian translocations," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 9, pp. 1963–1969, Aug. 2019.

[94] H. Liu, B. Mao, X. Xu, L. Liu, X. Ma, X. Zhang, "The effectiveness of next-generation sequencing-based preimplantation genetic testing for balanced translocation couples," *Cytogenetic and Genome Research*, vol. 160, no. 11-12, pp. 625–633, 2020.

[95] A. Kolano, S. Brunet, A. D. Silk, D. W. Cleveland, M.-H. Verlhac, "Error-prone mammalian female meiosis from silencing the spindle assembly checkpoint without normal interkinetochore tension," *Proceedings of the National Academy of Sciences*, vol. 109, no. 27, May 2012.

[96] T. Escudero, I. Abdelhadi, M. Sandalinas, S. Munné, "Predictive value of sperm fluorescence in situ hybridization analysis on the outcome of preimplantation genetic diagnosis for translocations," *Fertility and Sterility*, vol. 79, pp. 1528–1534, Jun. 2003.

[97] F. Morel *et al.*, "Meiotic segregation of a t(7;8)(q11.21;cen) translocation in two carrier brothers," *Fertility and Sterility*, vol. 81, no. 3, pp. 682–685, Mar. 2004.

[98] E. Coonen *et al.*, "ESHRE PGT consortium data collection XVI–XVIII: Cycles from 2013 to 2015†," *Human Reproduction Open*, vol. 2020, no. 4, 2020.

[99] Y. Q. Tan *et al.*, "Single-nucleotide polymorphism microarray-based preimplantation genetic diagnosis is likely to improve the clinical outcome for translocation carriers," *Human Reproduction*, vol. 28, no. 9, pp. 2581–2592, Jul. 2013.

[100] Y. Verlinsky *et al.*, "Preimplantation testing for chromosomal disorders improves reproductive outcome of poor-prognosis patients," *Reproductive BioMedicine Online*, vol. 11, no. 2, pp. 219–225, Jan. 2005.

[101] C. Huang *et al.*, "Pregnancy outcomes of reciprocal translocation carriers with two or more unfavorable pregnancy histories: Before and after preimplantation genetic testing," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 11, pp. 2325–2331, Sep. 2019.

[102] M. T. M. Franssen *et al.*, "Reproductive outcome after PGD in couples with recurrent miscarriage carrying a structural chromosome abnormality: A systematic review," *Human Reproduction Update*, vol. 17, no. 4, pp. 467–475, Apr. 2011.

[103] E. Anton, F. Vidal, J. Blanco, "Interchromosomal effect analyses by sperm FISH: Incidence and distribution among reorganization carriers," *Systems Biology in Reproductive Medicine*, vol. 57, no. 6, pp. 268–278, Nov. 2011.

[104] D. Warburton, "Genetic factors influencing aneuploidy frequency," in *Aneuploidy*, Springer US, 1985, pp. 133–148.

[105] D. E. Miller, "The interchromosomal effect: Different meanings for different organisms," *Genetics*, vol. 216, no. 3, pp. 621–631, Nov. 2020.

[106] F. Pellestor, "Study of the occurrence of interchromosomal effect in spermatozoa of chromosomal rearrangement carriers by fluorescence in-situ hybridization and primed in-situ labelling techniques," *Human Reproduction*, vol. 16, no. 6, pp. 1155–1164, Jun. 2001.

[107] N. Machev, P. Gosset, S. Warter, M. Treger, M. Schillinger, S. Viville, "Fluorescence in situ hybridization sperm analysis of six translocation carriers provides evidence of an interchromosomal effect," *Fertility and Sterility*, vol. 84, no. 2, pp. 365–373, Aug. 2005.

[108] N. Douet-Guilbert *et al.*, "Interchromosomal effect in sperm of males with translocations: Report of 6 cases and review of the literature," *International Journal of Andrology*, vol. 28, no. 6, pp. 372–379, Dec. 2005.

[109] M. S. Yildirim, A. B. Arslan, A. G. Zamani, "Interchromosomal effect: Report of a father and son, bearing different translocations of the same chromosome, and a review of the current literature," *Andrologia*, vol. 53, no. 1, Aug. 2020.

[110] O. Balasar, H. Acar, "Investigation of the interchromosomal effects in male carriers with structural chromosomal abnormalities using FISH," *Turkish Journal of Urology*, vol. 46, no. 3, pp. 178–185, May 2020.

[111] A. A. Bartolucci, M. Tendera, G. Howard, "Meta-analysis of multiple primary prevention trials of cardiovascular events using aspirin," *The American Journal of Cardiology*, vol. 107, no. 12, pp. 1796–1801, Jun. 2011.

[112] D. Young, D. Klepacka, M. McGarvey, W. B. Schoolcraft, M. G. Katz-Jaffe, "Infertility patients with chromosome inversions are not susceptible to an inter-chromosomal effect," *Journal of Assisted Reproduction and Genetics*, vol. 36, no. 3, pp. 509–516, Dec. 2018.

[113] P. N. Scriven, "PGT-SR: The red-herring and the siren; interchromosomal effect and screening for unrelated aneuploidy," *Journal of Assisted Reproduction and Genetics*, vol. 38, no. 5, pp. 1015–1018, Mar. 2021.

[114] S. Munné *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*, vol. 112, no. 6, 1071–1079.e7, Dec. 2019.

[115] R. C. McCoy, "Mosaicism in preimplantation human embryos: When chromosomal abnormalities are the norm," *Trends in Genetics*, vol. 33, no. 7, pp. 448–463, Jul. 2017.

[116] S. Gurunath, Z. Pandian, R. A. Anderson, S. Bhattacharya, "Defining infertility-a systematic review of prevalence studies," *Human Reproduction Update*, vol. 17, no. 5, pp. 575–588, Apr. 2011.

[117] C. Siristatidis, P. Vogiatzi, A. Pouliakis, M. Trivella, N. Papantoniou, S. Bettocchi, "Predicting ivf outcome: A proposed web-based system using artificial intelligence," *in vivo*, vol. 30, no. 4, pp. 507–512, 2016.

[118] A. Goyal, M. Kuchana, K. P. R. Ayyagari, "Machine learning predicts live-birth occurrence before in-vitro fertilization treatment," *Scientific Reports*, vol. 10, no. 1, Dec. 2020.

[119] R. Liu *et al.*, “Multifactor prediction of embryo transfer outcomes based on a machine learning algorithm,” *Frontiers in Endocrinology*, vol. 12, Nov. 2021.

PUBLICATIONS FROM THE THESIS

Papers

1. Ç. Ogur et al., "PGT for chromosomal structural rearrangements in 300 couples reveals specific risk factors but an interchromosomal effect is unlikely," Reproductive BioMedicine Online, Jul. 2022. DOI: 10.1016/j.rbmo.2022.07.0