Faculteit Geneeskunde en Farmacie
Vrije Universiteit Brussel
Laarbeeklaan 103
B-1090 Brussel

Mitochondrial DNA variants in early development and ART

Mertens Joke

Eindwerk, ingediend voor het behalen van de graad van Master in de Biomedische Wetenschappen

22th & 23th June 2017
Academiejaar 2016 - 2017

Promotor: Prof. Dr. Claudia Spits

Begeleider: Filippo Zambelli

Reproduction and Genetics (REGE)
Vrije Universiteit Brussel
Faculteit Geneeskunde en Farmacie
Laarbeeklaan 103
1090 Brussel

Index

Abstract English

Introduction: Children born after assisted reproductive technologies (ART) have a lower birth weight and an increased risk of developing cardiometabolic disorders later in life. A possible explanation for these differences could be an increased mitochondrial DNA (mtDNA) variant load. All mtDNA copies in one individual originate from the maternal oocyte. We hypothesize that the procedure of controlled ovarian stimulation used in ART results in cellular stress in the developing oocytes, leading to mtDNA damage. The aim of this thesis is to investigate if people born after ART have a higher mtDNA variant load than people born after natural conception and to establish if pluripotent stem cell lines can be used as a representative model for the individual to explore the distribution and segregation of mtDNA variants.

Material and Methods: Blood samples of young individuals born after Intra Cytoplasmic Sperm Injection (ICSI, 57 subjects) and spontaneous conception (62 subjects) were collected in the Center for Reproductive Medicine (CRG) of the University hospital of Brussels. Four embryonic stem cell lines (VUB02, VUB03, VUB04 and VUB07) were kept in culture for more than 100 passages and DNA samples were collected at three different time points (early, mid and late passage). Additionally, 9 clonal iPSC lines from two different individuals were generated, and 9 other iPSC lines reprogrammed using different reprogramming methods were kindly provided by the University of Strasbourg.
mtDNA enrichment was done by long-range PCR using two different primer sets to cover the full mtDNA. Amplicons were subjected to massive parallel sequencing using an Illumina MiSeq and HiSeq. Alignment of the data was done using BWA-MEM and the bam files were uploaded to mtDNA server. Single nucleotide variants were detected with MuTect. Variants >1.5% frequency were further analyzed.

Results and Discussion: At first sight, the ICSI and control group gave similar results in distribution and load of mtDNA variants. When analyzing the homoplasmic private variants (i.e. variants not occurring in the haplogroup), we observed a difference in the ICSI group where more individuals showed to have more variants that have not been described yet. When we looked further into the heteroplasmies in the coding regions, we found more subjects showing a higher cumulative load in the ICSI group when compared to the control group, with a significant difference when the cumulative load is over 30% ($p<0.025$).
We also investigated if ESC lines can be used as a representative model for the individual to study the distribution and segregation of mtDNA variants, and we saw that the distribution coding vs. non-coding of mtDNA variants was similar to that of mature cell types. In addition, we explored iPSC lines as a representative model for the individual and found that they carry a significant genetic diversity, representing the cellular heterogeneity of the donor.

Conclusion: With this research, we provide evidence that people born after ICSI show a higher cumulative heteroplasmic load in the coding regions of the mtDNA when compared to a control population. We also found that ESC lines faithfully represent the distribution of the variants in early development, while iPSCs most likely reflect somatic cellular diversity.

Abstract Nederlands

Introductie: Kinderen geboren na geassisteerde reproductieve technologieën (ART) vertonen een lager geboortegewicht alsook een verhoogd risico op cardiometabole aandoeningen in hun later leven. Een mogelijke oorzaak van deze verschillen zou een toename van mitochondriale DNA (mtDNA) varianten kunnen zijn. Alle mtDNA kopies vinden hun oorsprong in de maternale oöcyt. Wij stellen een nieuwe hypothese voor die stelt dat de procedure van gecontroleerde ovariële stimulatie resulteert in cellulaire stress in de ontwikkelende oöcyten en dat dit leidt tot mtDNA schade. Het doel van deze thesis is te onderzoeken of mensen geboren na ART meer mtDNA varianten vertonen in vergelijking met mensen die geboren zijn na natuurlijke conceptie en om te bestuderen of pluripotente stamcellijnen zouden kunnen gebruikt worden als representatief model voor het individu en om de distributie en segregatie van mtDNA varianten te analyseren.

Materiaal en Methoden: Bloedstalen van jonge individuen geboren na Intracytoplasmatische Sperma-injectie (ICSI, 57 subjecten) en na spontane conceptie (62 subjecten) werden gecollecteerd in het Centrum voor Reproductieve Geneeskunde (CRG) van het universitair ziekenhuis van Brussel. Vier embryonale stamcellijnen (VUB02, VUB03, VUB04 en VUB07) werden in cultuur gehouden voor meer dan 100 passages en DNA stalen werden gecollecteerd op 3 verschillende tijdstippen (vroege, mid en late passages). Als toevoeging werden 9 klonale iPSC lijnen van twee verschillende individuen gegenereerd en bijkomend werden 9 iPSC lijnen, gereprogrammeerd door verschillende methoden, voorzien door de universiteit van Strasbourg.
mtDNA-verrijking werd verricht gebruik makend van long-range PCR met twee verschillende primer sets om het volledige mtDNA te omvatten. Deze amplicons werden onderworpen aan massive parallel sequencing met behulp van een Illumina MiSeq en HiSeq. De data werden vergeleken met het referentiegenoom door BWA-MEM, waarna de bam files werden geüpload op mtDNA server. Single nucleotide varianten werden gedetecteerd door MuTect. Varianten met een frequentie van >1.5% werden verder geanalyseerd.

Resultaten en Discussie: Op het eerste zicht geven de ICSI en controlegroep vergelijkbare resultaten wanneer men kijkt naar de distributie en de load van mtDNA varianten. Wanneer we echter de homoplasmische privé varianten bekeken (i.e. varianten die niet voorkomen in de haplogroep), zagen we dat er een verschil was in de ICSI-groep waar meer mensen varianten vertoonden die nog niet werden beschreven. Heteroplasmies werden verder geanalyseerd in de coderende regio's. Hier zagen we dat er meer subjecten een hogere cumulatieve load vertoonden in de ICSI-groep. Na statistische analyse, observeerden we een significant verschil in het aantal ICSI-subjecten wanneer we keken naar een cumulatieve load over 30% ($p<0.025$).
ESC-lijnen werden geanalyseerd om na te gaan of ze gebruikt kunnen worden als representatief model voor het individu om de distributie en segregatie van mtDNA varianten te bestuderen. We zagen dat de distributie coderend vs. niet-coderend van mtDNA varianten gelijk was aan die van meer mature celtypes. Bijkomend hebben we ook onderzoek verricht naar iPSC-lijnen als representatief model en vonden dat deze een verschillende genetische achtergrond vertoonden waarbij ze de cellulaire heterogeniteit van de donor representeren.

Conclusie: Met dit onderzoek bewijzen we dat mensen geboren na ICSI een hogere cumulatieve heteroplasmische load in de coderende regio's van het mtDNA vertonen wanneer we ze vergelijken met een controlepopulatie. We kunnen ook stellen dat ESC-lijnen betrouwbaar de distributie van de varianten in de vroege ontwikkeling representeren, terwijl iPSCs meer de somatische cellulaire diversiteit reflecteren.

1. Introduction

1.1 Assisted Reproductive Technologies

Since the development of assisted reproductive technologies (ART), many subfertile couples (i.e. one year of unprotected intercourse without resulting in a pregnancy) have the chance to fulfill their wish to become parents. Worldwide, more than 5 million children (*1*) have been born through ART procedures and their number is rapidly increasing (*2*). The most commonly used techniques are in vitro fertilization (IVF) and intra-cytoplasmic sperm injection (ICSI), which are used as a treatment for female and male infertility respectively (*3*).

Most ART procedures start with controlled ovarian stimulation. For this, exogenous hormonal stimulation is given to the woman. First, GnRH (gonadotrophin releasing hormones) analogs are administered daily, which will bind to the receptors in the hypothalamus and hypophysis to inhibit endogenous gonadotrophin secretion and to create a negative feedback. After a week, the LH (luteinizing hormone) peak will not occur and no spontaneous ovulation will take place. This is to let the developing oocytes further mature. The follicles can be monitored by ultrasound and, when they reach a sufficient stage of maturation, an injection of hCG (human chorion gonadotrophin) is given to the patient as an alternative LH peak. The follicles are then aspirated by trans-vaginal punction (*4*). Controlled ovarian stimulation is needed to allow a large cohort of oocytes mature in order to circumvent the event of developing one dominant oocyte in one menstrual cycle (*5*). After aspiration, the oocytes will be denuded and will then be inseminated in vitro. When the sperm sample is of insufficient quality, ICSI can be used to achieve fertilization by directly injecting one spermatozoon into the oocyte using a micropipette (*6*). After fertilization, the embryo is kept in vitro from two to five days and is then placed into the uterus where it will implant in the endometrium. When one of the parents carries a genetic disease, the embryos can be subjected to preimplantation genetic diagnosis (PGD) to select embryos that do not carry the disease. This is done by taking a biopsy of the embryo, polar body, blastomere or trophectoderm, to test for the presence of a specific mutation in the genotype of the embryo (*7*). This can be done in case of monogenetic diseases or chromosomal abnormalities.

1.1.1 Differences between spontaneously conceived and ART children

ART procedures are being used since the birth of Louise Brown in 1978, now for nearly 40 years (*8*). At first sight, the children born after ART did not show to be different from their naturally conceived (NC) peers. However, over time, a large number of studies have found that ART children have a higher risk of being born more prematurely, suffering from intra-uterine growth retardation, having congenital malformation, having a lower birth weight, and developing cardiometabolic syndromes (*2, 9-11*). Even though ART children show a lower birth weight than spontaneously conceived children, they gain in weight and height rapidly during late infancy (three months old to one year) (*12*).

Conversely, it is known that a lower birth weight is associated with syndromes later in life, such as renal disease (*13, 14*), diabetes (*15*), cardiovascular disease (*16*) and metabolic syndrome (*17*). In light of this, a number of follow-up studies of ART children have been focusing on cardiometabolic aspects. For instance, in 2012, a study was performed concerning the body fat distribution in IVF children. IVF boys appear to have a higher percentage of body fat and IVF girls show an increased BMI (Body Mass Index), body fat percentage, total skin folds and upper arm and waist circumferences (*18*). Another important factor is insulin resistance. Insulin resistance has been previously reported to affect ART children more than NC children. Ceelen and colleagues found that there was an increase of fasting glucose levels in IVF children compared with a control group (*19*). A comparative study was performed by Sakka et al. to look at RBP-4 (retinol-binding protein 4) and

NGAL (neutrophil gelatinase-associated lipocalin) levels in IVF children. These molecules are known to be associated with glucose intolerance and showed to be elevated in the IVF group when compared to controls (*20*). ART children also appear to be more prone to generalized endothelial dysfunction and exaggerating stiffness of their vasculature which gives a high risk of developing atherosclerosis (*21*). Also, an increase in systolic and diastolic blood pressure is seen in ART children (*19*). In addition, cardiac remodeling and function is more affected in ART children when compared with NC children (*22*).

Because of the subfertility of their parents, questions emerged about the reproductive potential of ART children. Jensen et al. performed in 2006 a large cohort study of Danish men born after fertility treatment (*23*). They found that these men had lower sperm concentrations, lower total sperm count, smaller testis, fewer motile sperm and lower serum testosterone levels. These results have been confirmed by Belva et al. in 2016 in young adults born after ICSI (*24*). Belva et al. also conducted research on the hormone levels of FSH, LH, inhibin B and testosterone in male ICSI children. They found that inhibin B and FSH concentrations were likely to be altered, lower and higher respectively, in ICSI children when compared with NC peers. However, these results did not reach statistical significance because of the small sample size (*3*). They also investigated AMH, FSH, LH and DHEAS levels in female ICSI children, but found no significant differences when compared with a control group (*25*). However, in 2008, a study of female IVF children was published showing that these children had higher DHEAS and LH levels when compared to control subjects (*26*).

The cause of these differences has been intensively investigated, and the majority of researchers have focused on the epigenome. Epigenetic modifications activate or silence genes in specific cell types without changing the sequence of the nuclear DNA, and include DNA methylation, histone modification, non-coding RNA and organization of the DNA strands (*27*). Several studies indicated an association between an increased incidence of imprinting disorders and children born after ART (*28, 29*). Some researchers reported significant changes in the methylation pattern in sperm of infertile men and in oocytes that underwent controlled ovarian stimulation (*30*). However, in 2012, Oliver et al. explored several imprinted regions and saw no difference in percentage of DNA methylation between the IVF group and their spontaneously conceived peers (*31*). Whitelaw et al. examined such imprinted regions as well and concluded that the differences between ART children and naturally conceived children were not significant (*32*). Also, Doornbos and colleagues suggested that these disorders are caused by subfertility itself rather than changes in the epigenome (*33*). Overall, the exact link between epigenetic changes and ART children remains unclear.

In this thesis, we work on a novel hypothesis that could explain why ART children show differences when compared to naturally conceived children. The core of the hypothesis is that mitochondrial DNA (mtDNA) changes, rather than epigenetic changes could explain these differences. Mitochondria play a crucial role in energy-demanding tissues. Hence, mutations in mtDNA, and in consequence alterations in the mitochondrial function, including oxidative phosphorylation (OXPHOS), result in a higher risk of dysfunction in these tissues. Variations in mtDNA could contribute to the susceptibility of several diseases in neuromuscular, musculoskeletal, cardiovascular, renal, endocrine and neural tissues. The origin of these mtDNA alterations can be caused by several factors. The two factors most associated with lower birth weight in ART children are maternal infertility and controlled ovarian stimulation in the mother (*10*). In this research, we suggest that these two factors could be linked to the mitochondrial genome.

1.2 Mitochondrial genome and its link to low birth weight and cardiometabolic abnormalities

Mitochondria are the powerhouses of the cell by producing ATP (adenosine triphosphate) through OXPHOS. Mitochondria are abundant in the cell and can control their numbers by fission and fusion according to the energy need of the cell (*34*). They contain their own DNA (mtDNA) in multiple

copies per mitochondrion. When the copies have exactly the same genetic code, this is called homoplasmy. Homoplasmic variants will be found in 100% frequency of all the mtDNA copies in one mitochondrion, cell, tissue or organism. However, when mtDNA copies with different genetic codes are found, this is called heteroplasmy. Heteroplasmic variants can be found between 1 and 99% frequency.

The mtDNA consists of 16,569 base pairs and has a double-stranded, circular structure (*35*) (figure 1). It is packaged in nucleoids which can contain multiple copies of mtDNA (*36, 37*). Two strands can be distinguished: the heavy strand and the light strand, reflecting their behavior in density gradients according to their richness in Gs and Cs respectively (*36, 38-40*). mtDNA is very compact and efficient, considering it has no introns and intergenetic sequences are reduced to a few or no base pairs at all. Moreover, some regions are overlapping for different genes. Packaging, transcription and replication are mediated by TFAM (mitochondrial transcription factor A) (*41*).

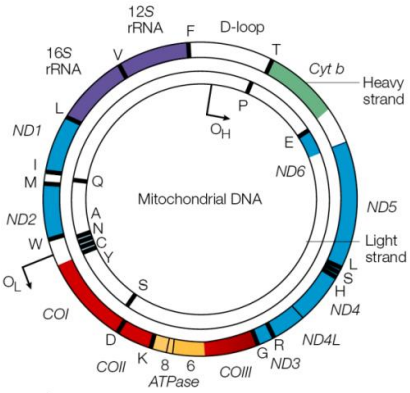


Fig.1. Structure of mtDNA.

mtDNA encodes for 13 subunits of the electron transport chain of the oxidative phosphorylation (table 1). The OXPHOS is the most important energy producing process for the cell which creates ATP from adenosine diphosphate (ADP) and inorganic phosphate (P_i). It consists of 5 complexes with almost 90 subunits, encoded by both nuclear and mtDNA, and is embedded within the inner membrane of the mitochondria (*42*). During the OXPHOS, significant amounts of reactive oxygen species (ROS) are formed. These natural byproducts can cause damage to proteins and DNA in the cell, leading to oxidative stress.

Gene	Subunit of	Locus on the mitochondrial genome
ND1	Complex I	3307-4262
ND2	Complex I	4470-5511
COI	Complex IV	5904-7445
COII	Complex IV	7586-8269
ATP8	Complex V	8366-8572
ATP6	Complex V	8527-9207
COIII	Complex IV	9207-9990
ND3	Complex I	10059-10404
ND4L	Complex I	10470-10766
ND4	Complex I	10760-12137
ND5	Complex I	12337-14148
ND6	Complex I	14149-14673
CYB	Complex III	14747-15887

Table 1. List of the 13 subunits of the electron transfer chain encoded by mtDNA and their location. ND: NADH dehydrogenase, CO: cytochrome c oxidase, ATP: ATP synthase, CYB: Cytochrome B.

The mtDNA also transcribes part of its own transcription machinery: 22 tRNAs and 2 rRNAs. Transcription of the mtDNA needs TFAM which will bind to the promoters on the heavy and the light strands (*43*). Transcription of both the H-strand and the L-strand is done to form one long polycistronic RNA-strand which will be processed by nuclease cleavage to cut the tRNAs, which flank most of the rRNAs and mRNAs, and hereby generate mature rRNAs and mRNAs (*44*).

In addition to the genes, tRNAs and rRNAs, mtDNA contains also a non-coding control region which is 1 Kbp long and is called the displacement loop (D-loop). This region has a three-stranded structure and interacts with transcription and replication factors that are encoded by the nuclear genome (*39, 45*). The D-loop contains the origin of replication on the heavy strand (O_H) and promoters for transcription on both strands (*36, 38*). It consists of 2 hypervariable regions and the control region (*35*).

1.2.1 mtDNA linked to the phenotype of ART children

1.2.1.1 Lower birth weight

During fetal life and especially during the third semester, the OXPHOS will operate at high levels to answer to the demands of the rapid growing fetus (*46, 47*). This suggests that if the foetus suffers from mtDNA mutations, its growth may be compromised, resulting in an intra-uterine growth restriction and in a lower birth weight. This particular effect has already been described using a mtDNA mutator mouse model (*48*). In human, children carrying mutations causing mitochondrial respiratory chain deficiency are frequently born with low birth weight (*49, 50*). Also, a retrospective study in 2012 by Tavares et al. investigated 44 pediatric subjects with genetic and biochemical alterations and found that the birth weight of these children was significantly lower as when compared to control children (*51*).

1.2.1.2 Cardiometabolic syndromes

mtDNA mutations are known to play a role in several aspects of the cardiometabolic syndrome. mtDNA mutations are associated with β-cell failure which is correlated with defective activity of complexes I and II of the electron transport chain (*52, 53*). Also, β-cells carrying mtDNA mutations show to have a defect in mitochondrial inner membrane potential which can induce diabetes (*53*). In skeletal muscle, it has been described that alterations in OXPHOS are associated with development of insulin resistance and type 2 diabetes (*54, 55*). Patients with mtDNA mutations commonly display diabetes as one of the clinical features. These patients represent 1-2% of the diabetic patients (*56*). Another important and energy-demanding system is the cardiovascular system which depends greatly on mitochondrial respiration. It was already suggested in the 1960s that impairment of OXPHOS plays a crucial role in the development of atherogenesis (*57*). Defects of OXPHOS leads to a decrease in ATP production which gives a reduction in the inhibition of HMG-CoA-reductase and concurrently an increase of cholesterol synthesis (*53*). Also, ischemia and hypertrophic cardiomyopathy has been associated with deterioration of OXPHOS, mtDNA deletions and an increase of ROS (*53*).

1.2.1.3 Infertility

Mitochondria play a crucial role during gametogenesis and preimplantation development, and have been repeatedly hypothesized to play an important role in reproductive ageing and female infertility (*58, 59*). It is thought, for instance, that an age-related decrease in mitochondrial function, possibly due to the accumulation of mtDNA mutations, impairs oocyte maturation, spindle activity and chromosomal segregation (*59-61*). Furthermore, pathogenic mtDNA mutations are known to be commonly present in healthy individuals at low loads, and cross-generation studies have shown that

they are heritable (*62-64*). In the light of this, it is possible that women with an above-average mtDNA mutation load (generally or germ-line restricted) are infertile, whilst being otherwise healthy. In turn, these same mutations that cause the infertility of the mother will be transmitted to the child through the procedure of ART.

1.2.2 Transmission and segregation of mtDNA in human early embryonic development

1.2.2.1 Oogenesis

mtDNA is uniparentally transmitted to the next generation and is therefore, inherited in a non-Mendelian manner. All the mtDNA copies of one individual were present in the oocyte before conception, thus coming from the mother. During oogenesis, mitochondria undergo a bottleneck (figure 2A) where the few mitochondria present in the primordial germ cell, which in turn contain very few copies of mtDNA each, will replicate to 20,000 – 593,000 mtDNA copies in the resulting mature oocyte (*65, 66*). This means that one mutant mtDNA copy in the germ cell can give rise to thousands in the oocyte and millions in the offspring. The bottleneck hypothesis is a possible model for explaining the genotypic drift of mtDNA polymorphisms within one generation (*67*).

This pool of mitochondria in the oocyte can contain exactly the same genetic material (homoplasmy), however, this is not always the case. This means that an oocyte can contain wild-type mtDNA copies as well as mutant copies (figure 2B). This heteroplasmic mixture of mtDNA copies is transmitted to the offspring. Steffann et al. found that there is a high transmission risk (84%) for mtDNA mutations when the mother is already a carrier of these mutations (*67*).

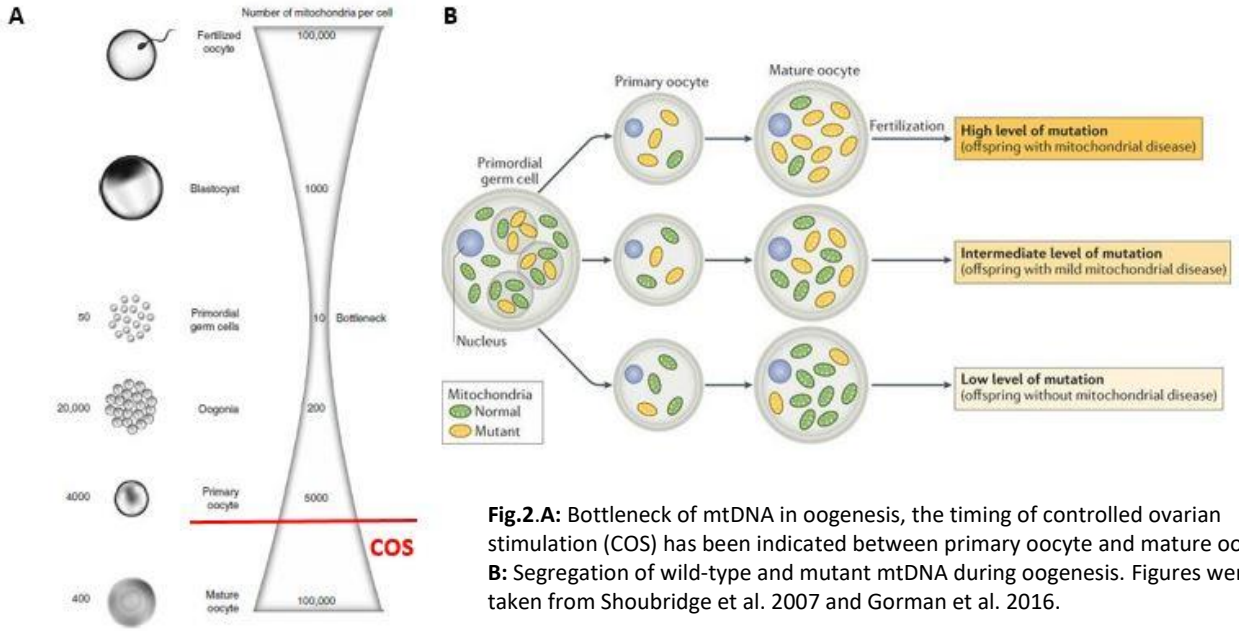


Fig.2.A: Bottleneck of mtDNA in oogenesis, the timing of controlled ovarian stimulation (COS) has been indicated between primary oocyte and mature oocyte. **B:** Segregation of wild-type and mutant mtDNA during oogenesis. Figures were taken from Shoubridge et al. 2007 and Gorman et al. 2016.

1.2.2.2 mtDNA integrity in stimulated oocytes

In ART procedures, oocytes will be stimulated through exogenous hormonal stimulation and this causes cellular stress in the developing follicles, possibly leading to an increase in mtDNA mutation load.
Most mutations in the mitochondrial genome are seen in slow, non-dividing cells (*68*). These tissues show a higher percentage in mtDNA mutations than other tissues with a rapid turnover. Oocytes are formed during the fetal development and reside in a prophase I-block in the ovary for several decades until the oocyte is recruited in a menstrual cycle. This results in an increased risk of

developing mutations in the mitochondrial genome. Due to the controlled ovarian stimulation, the oocytes that would undergo atresia in a natural cycle, are recruited and aspirated in ART procedures. It is described in literature that the mammalian ovary has developed a system to eliminate follicles harboring mtDNA deletions (*40*). However, due to the stimulated protocol in ART procedures, oocytes harboring mtDNA mutations might circumvent this system and be used to be fertilized.

Very few research has been done about the effect of controlled ovarian stimulation on mtDNA integrity in animals, and none of them in human. It has been suggested that in mice and Rhesus monkeys, hormonal stimulation leads to an acute increase of ROS levels, which in turn increases the frequency of large-scale mtDNA deletion in oocytes and mitochondrial dysfunction (*69-71*). Furthermore, Hammond et al. showed that maternal age has a significant impact on the levels of mtDNA deletions in bovine oocytes and that their frequency possibly increases due to the controlled ovarian stimulation (*72*). Overall, these results suggest a damaging effect of controlled ovarian stimulation on mtDNA integrity in oocytes.

In human oocytes, the mechanisms and segregation profile of several mutations have been described, but remain elusive. On one side, Blok and colleagues (*73*) reported the segregation pattern of the T8993G mutation in oocytes after ovarian follicular stimulation. They analyzed seven oocytes and saw that one did not show the mutation while the remaining six showed more than 95% mutation load. This data could support the hypothesis of one mtDNA copy preferentially amplified during oogenesis; on the other side, in 2001, Brown et al. (*74*) investigated the transmission of the A3243G mtDNA mutation in human primary oocytes originating from one patient, and they saw that this mutation was transmitted through random genetic drift, contradicting the data described by the group of Blok (*73*).

1.2.2.3 mtDNA in early development in vivo

After fertilization, mitochondria are located close to the pronuclei. This is to provide energy for the nuclear DNA replication and syngamy (*75*). During the first cleavage stages, mtDNA replication does not occur until the blastocyst stage (figure 3), and the number of mtDNA copies per cell decreases per cell division (*41*). Hence, all mtDNA copies present in the cleavage stages are originated from the pool that resided in the oocyte. These early cleavage cells are dependent on anaerobic respiration and lose their capacity to generate ATP from the oxidative phosphorylation (*76, 77*). When the embryo reaches the blastocyst stage, an upregulation of the mtDNA replication and transcription factors is observed (*78*), mitochondria change their structure and start to differentiate towards a more mature phenotype (*79*). This is particularly observed in the differentiated trophectodermal cells. The inner cell mass is still composed of pluripotent cells, with mitochondria that reside in a more undifferentiated phenotype, and these cells do not express mtDNA replication factors (*80*). At this stage, there are thousands of mtDNA copies per cell, but this number further decreases until 30-45 copies per cell (*81*) (figure 3). After implantation, mtDNA continues to replicate and mtDNA copies will increase in neurons, liver, muscle and heart (*67*) and will further expand during fetal development (*67, 82*).

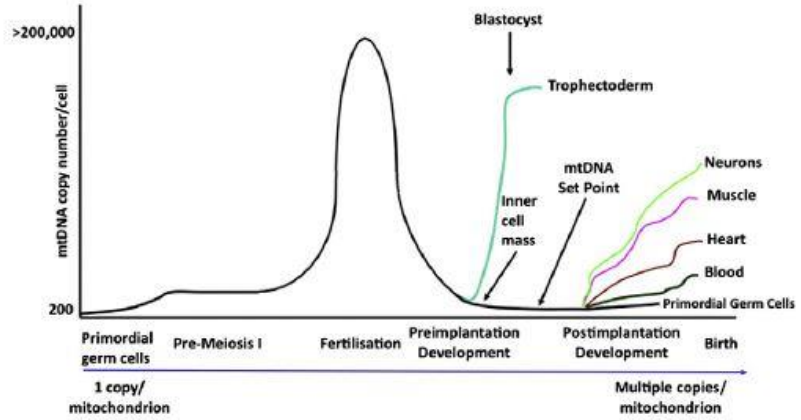


Fig.3. mtDNA copy number per cell during embryonic and fetal development. (St. John et al. 2010)

1.2.2.4 Human embryonic stem cells (hESCs) in vitro as a proxy to early human development

hESCs can be used as a research model to study early human embryonic development, since they are an in vitro culture of early human embryonic cells. hESCs are most commonly derived from embryos donated by couples undergoing ART (*83*). This makes this model particularly interesting in this study, as the cells, at least at very early passage after derivation, can be seen as representative of what is found in ART embryos.

hESCs are also an interesting model to investigate the impact of mtDNA variants on cells in culture. When a certain variant arises during culture (or was already present in the embryo, but at a low heteroplasmy) and confers a growth advantage due to alterations in the cell's metabolism, this variant can become homoplasmic in the culture. This situation can mimic the circumstances in early embryonic development in vivo as mtDNA copies are distributed in the daughter cells of the rapid growing embryo and variants with a growth advantage can quickly become homoplasmic in the embryo. The mutated mtDNA copies will migrate to the newly formed cells which eventually will become different cell types. Here, replication of mtDNA will start simultaneously with cell differentiation and the cell will become more in need of an optimal mitochondrial respiratory function.

Mitochondria play an important role in the maintenance of the stemness in pluripotent stem cells (PSCs). Their immature state is one of the most crucial features to maintain the stemness in pluripotent cells which rely on the glycolytic pathway for energy production much more than the OXPHOS (*84, 85*). It seems that mtDNA is not actively transcribed in stem cells, however, it is known that mtDNA mutations can impair self-renewal of stem cells and induce an abnormal differentiation (*86*). Several papers also described that impairment of OXPHOS using respiratory inhibitors improves gene expression of pluripotency genes and inhibits differentiation (*87, 88*).

Another pluripotent stem cell type are induced pluripotent stem cells (iPSCs) (*89, 90*). These cells are reprogrammed from mature somatic cells, and display the characteristics of pluripotent stem cells, including not relying on OXPHOS, which reflects in their low number of immature mitochondria (*41, 91*). While the cells are reprogrammed from a somatic metabolism to a stem cell metabolism, the mitochondria undergo several changes. As a consequent of epigenetic changes, there will be an upregulation of the glycolytic genes and a downregulation of the OXPHOS genes which initiates dedifferentiation. This results in a reduction of mtDNA copy number and alterations in the structure

and morphology of the mitochondria to establish a more immature state (*84, 92, 93*). This switch is essential to induce and maintain the pluripotency of the cell. It has been recently described that pathogenic mtDNA mutations over a certain threshold can impair differentiation capacity, proliferation rate, in case of iPSCs, reprogramming efficiency (*94-96*).

Overall, mtDNA in PSCs is not operating at high levels, however, it has an important function in maintaining the pluripotency of the cells and in differentiation towards mature cell types.

1.2.3 mtDNA mutagenesis

mtDNA is 10-20 times more susceptible to mutations as compared to the nuclear DNA, as mitochondria lack efficient DNA repair mechanisms (*40*). mtDNA contains no histones and is located near the OXPHOS site which produces high amounts of ROS that create a mutagenic environment (*53, 97*). The frequency in which mtDNA diseases appear is 1/5000 (*98*). These mutations can be divided into three groups: recent mutations, ancient regional variants and somatic mutations (*40*). The difference between these mutations is the time point in which these mutations have been established. Recent mutations are those that were transmitted through the maternal germline, ancient mutations were established during evolution to adapt to the environment and somatic mutations are mutations that accumulated in somatic tissues, during the life of an individual.

The effects of mtDNA mutations depend on the severity of the mutation, the threshold (i.e. heteroplasmy level) and distribution of the variant in specific tissues and the age and predisposition of the individual (*99*). Therefore, mtDNA mutations show a variety of clinical features.

1.2.3.1 Classical inherited diseases (recent mutations)

Many mitochondrial DNA mutations and their association to diseases, have been reported in literature. The most severe syndromes described are chronic progressive external ophtalmoplegia (CPEO) and Kearns-Sayre syndrome (KSS). Another severe syndrome, resulting from mtDNA mutations, is the Pearson's syndrome. These patients suffer from pancytopenia early in life. If they survive this stage, their disease progresses in the KSS (*100*). Several diseases like Leber's hereditary optic neuroretinopathy (LHON) (*101*), mitochondrial myopathy (*102*), maternally inherited Leigh's syndrome (MILS) (*103*) and neuropathy, ataxia and retinitis pigmentosa (NARP) syndrome (*104*) are caused by point mutations in genes. Mutations in tRNAs and rRNAs affect the transcription of these genes and can cause diseases like mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS) (*105*) and myoclonic epilepsy associated with ragged-red fibers (MERF) (*106*). The most common large deletion in mtDNA spans the 4977 base pairs located between nucleotides 8482 and 13460 (*107*).

1.2.3.2 Ancient regional variants

Many polymorphisms within our mtDNA have accumulated during evolution, allowing us to migrate to different climates and adapt to them. Through mtDNA analysis, researchers have been able to group these polymorphisms into different haplogroups and have calculated the time when they originated during evolution. It is remarkable that highly conserved amino acids have been changed during evolution within the same species. An example of adaption are the small alterations in membrane potential and Ca2+ metabolism induced by the variants in haplogroup N. These variants are responsible for an increase in the number of calories burned to gain more heat and increase body-heat production in severe circumstances (*108*). Polymorphisms in haplogroups are not always non-pathogenic, considering that some haplogroups are more prone to certain diseases than others (*109-111*). These polymorphisms accumulated throughout the history of our species considering they are transmitted in an uniparental way and therefore recombination does not take place.

1.2.3.3 Somatic mutations

Another category of mtDNA mutations are somatic mutations. As cells undergo mitosis, they also must replicate their mtDNA, during which there is a risk of mutagenesis. Also, these mutations correlate to tissue turn-over. Tissues with a slow turn-over, like neurons or muscle cells, have an increased risk of suffering from mtDNA mutations. This is possibly because mtDNA mutations accumulate with age and cells with a slow turn-over displaying such mutations are sustained throughout a large period, unlike tissues with a rapid turn-over like blood or epithelial cells (*112, 113*).

Somatic variations can play a role in ageing, cancer, neurodegenerative and other diseases (*114-118*). These mutations can be tissue-specific and have been found recurrently in the same tissues of different individuals and can be the cause of heterogeneity of heteroplasmies in different tissues (*116, 119*). Some researchers even hypothesized that some heteroplasmies are positively selected over time (*119*). These mutations are mostly located in the region that controls replication (*118*).

Finally, somatic mutations will appear earlier when mtDNA has already a higher mutation load coming from the mother. This was shown in mice that inherited a higher mtDNA mutation load at birth and suffered from a higher rate of somatic mutations later in life, causing an acceleration of the ageing process (*48*).

1.2.4 Sequencing of mtDNA

The first problem that is encountered when sequencing and analyzing mtDNA is that, even though mitochondria are abundant in the cell, mtDNA is still very small in comparison with the nuclear genome. Hence, mtDNA needs to be enriched before sequencing. This can be done by isolating mtDNA through ultra-centrifugation in CsCl density gradients. However, this method is very time-consuming and has a low-throughput (*120*). More high-throughput methods for purifying mtDNA are commercial available kits, microarray hybridization and PCR-based enrichment (*121, 122*).

A second problem to face is that not all variants are present in a 0 or 100% load, but in a range in between. Several methods have been proposed to quantify heteroplasmies, but all of them have their limitations, for instance, like the proposed use of PCR-RFLP (PCR-restriction fragment length polymorphism) analysis to quantify heteroplasmies that cannot go lower than a detection limit of 10% (*123, 124*). Previously, two methods were frequently used to sequence mtDNA: Sanger sequencing and mitochondrial DNA re-sequencing by Affymetrix's MitoChip (*120*). However, these methods are costly, labor intensive and did not provide quantitative information. A novel technology arose to circumvent these weaknesses: Massive parallel sequencing or Next generation sequencing (NGS).

NGS has a deep coverage or depth of sequencing (i.e. the amount of reads of one base pair), determines quantitative sequence data and can detect mutations at very low frequencies, unlike Sanger sequencing which can only identify heteroplasmy levels above 15% (*125, 126*).

There are different approaches to NGS. In all cases, before sequencing, a library needs to be prepared. DNA fragments are sheared into shorter fragments and platform specific adapter sequences are added to the 3' and 5' ends (paired-end) (*127*). The adapter sequences flank the insertion sequences to perform alignment with both the forward and the reverse oligonucleotides. In several platforms, local clonal amplification of the initial template will be performed to create clusters (*128*). This is to increase the signal-to-noise ratio. NGS can be performed on different platforms, such as, Illumina, Life Technologies, Semiconductor sequencing and Pacific Bioscience. The Illumina technology (used in this project) is performed in a flow cell. This flow cell can be divided

into different lanes: 1 lane (MiSeq), 2 lanes (HiSeq2500) or 8 lanes (HiSeq2000) (*127*). In these lanes, the forward and reverse oligos are attached to the flow cell and the adapter sequences, which are introduced during the library preparation, will bind. The DNA fragments will be denatured in order to hybridize to the oligos and to form a copy of the initial template. This initial template will be removed and the new strand will bend over to bind to the reverse oligos to form a bridge. This is called bridge amplification. The final step is to denature the double-stranded copies of the initial template and to perform bridge amplification in multiple cycles. During the formation of the strands fluorescently labeled dNTPs are incorporated. All four dNTPs are differently labeled and emit a different signal when incorporated. This signal is detected by optical sensors, hence, rebuilding the sequence of the forward and reverse strands of the initial template. This is called sequence-by-synthesis.

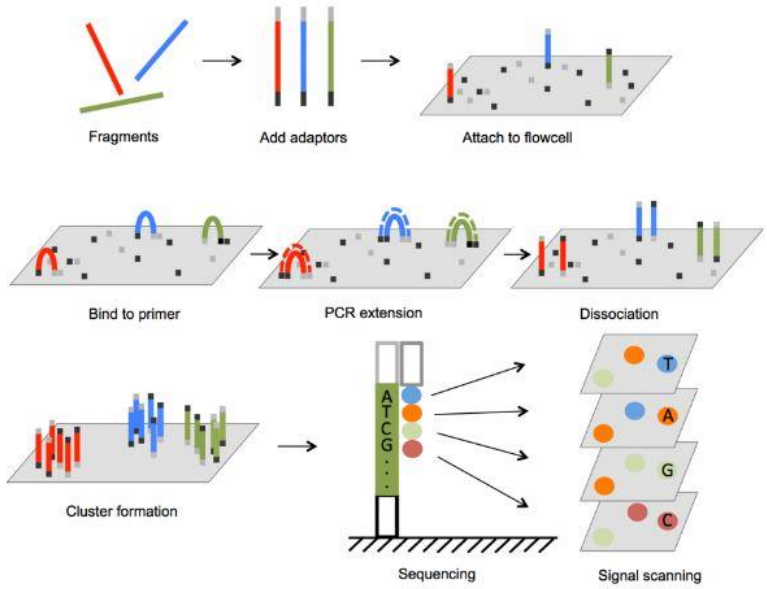


Fig.4. Sequential steps in Illumina sequence technology. Firstly, the template will bind to the oligonucleotides attached to the platform. Secondly, PCR amplification is performed and bridges will be formed to produce forward and reverse strands. Thirdly, the double-stranded DNA fragments will denature in order to perform multiple amplification steps and to form clusters. Sequencing is done by synthesis of the forward and reverse strands. This done by emitting specific fluorescent signals by the different dNTPs. This signal is detected by sensors and sequencing of the initial template is performed.

The extensive amount of data generated with NGS needs to be aligned before analyzing. These alignments can be made through several software programs available (BWA, Eland, RMAP, MAQ, ZOOM, SeqMap, Cloudburst, SHRiMP etc.) (*129*). These programs allow to align short fragments against a large reference sequence. Burrows-Wheeler Alignment tool (BWA) is a new package that efficiently reads short fragments (*129*). It supports base space reads generated from Illumina machines and is 10-20 times faster than MAQ (*129*). A component of BWA is BWA-MEM which supports paired-end reads, is robust to sequencing errors and is applicable to a wide range of sequence lengths from 70bp to several megabases (*130*).

Zang et al. conducted a comparison between Sanger sequencing and NGS and found that several heteroplasmies were not detected with the Sanger sequencing method, whilst NGS did reveal them (*126*). They show a detection limit of 1.33% frequency with 99.9% confidence. Other studies showed a detection limit of 5% (*131*) and even of 1% (*132*). It is, however, difficult to distinguish variants with very low frequencies from possible sequencing errors, such as PCR errors or errors from the sequencing platform (*120*).

2. Aims and research set up

The aim of this study is to investigate if ART children have a higher mutation load in their mitochondrial DNA when compared to children that were naturally conceived. We will investigate this by analyzing blood samples of young individuals born after ICSI and spontaneous conception (control group). In this setting, we are only investigating individuals born from fertile women, as this ICSI cohort only includes couples with a male infertility factor. A second part of this study is to establish if PSCs can be used as a representative model for the individual and to examine if the distribution of heteroplasmies depicted in somatic tissues is similar in PSCs. Also, segregation of heteroplasmies during long-term culture is explored. As an addition, iPSCs are also analyzed.

We begin with the sample collection and DNA extraction from blood and buccal swabs. After this, mtDNA is enriched through long range PCR using two different primer sets to cover the full mtDNA. The generated amplicons are sequenced through massive parallel sequencing using the Illumina MiSeq or HiSeq platform of the BrightCore facility (http://www.brightcore.be/).

The generated fasta files are then aligned with BWA-MEM (single nucleotide variant (SNV) analysis) and BLAST (large deletion pipeline). Subsequently, the SNVs are detected with MuTect and bam files (i.e. a binary format to store sequence data) are uploaded to mtDNA server. This is an online server which analyses the homoplasmic variants and correlates them to a specific haplogroup, and the heteroplasmic variants and reports their relative frequencies. mtDNA server also gives information about possible contaminations or acquired pathogenic variants. Heteroplasmic variants, present in more than 1.5% of the bulk, are thoroughly investigated. The variants that induce a non-synonymous change in the genes and therefore, a frame shift and an amino acid change in the protein are uploaded to MutPred. This is a tool that predicts the pathogenicity of the variant by reporting the molecular changes in the translated protein.

The analysis of the pathogenicity of these variants will allow us to evaluate not only the relationship between the frequency of the variant and its phenotypic effect, but also to observe if non-pathogenic and pathogenic variants are segregated differently in development, as described in literature (*67*).

This project is integrated in a greater project, named MitoART, which is part of the research conducted by Prof. Claudia Spits. MitoART studies mtDNA mutations in ART children and aims to evaluate a possible association between hormonal stimulation in ART procedures and an increased mutation load in mtDNA. It focusses on mtDNA in different stages in life and in different tissues to have a deeper insight in how mtDNA mutations can influence the metabolism of an individual. The tissues that are investigated are oocytes, embryos, stem cells, placenta and adult tissues like blood, urine and buccal swabs. This will also give a better understanding of the segregation dynamics of mtDNA throughout human development.

This project has been funded by Fonds voor Wetenschappelijk Onderzoek (FWO), the Methusalem grant of Prof. Sermon (Research Council of the VUB) and by the Scientific Research Fonds Willy Gepts of Universitair Ziekenhuis Brussel.

The research team of Prof. Spits is part of the research group Reproduction and Genetics (REGE, http://emgerege.vub.ac.be/). The laboratories of REGE are situated at campus Jette of the Vrije Universiteit Brussel (VUB), and the group is part of the Reproduction, Genetics and Regenerative Medicine (RGRG) Cluster. Overall, the research at REGE focusses on the different aspects of human genetics, reproduction, early development and stem cells. The results of these projects are discussed in meetings within the laboratory and are validated through publications in international, peer-reviewed journals.

3. Materials and methods

3.1 Recruitment and subject material

119 Young adults were recruited by the Center for Reproductive Medicine (CRG) of the University Hospital Brussels of which 57 individuals were conceived via ICSI and 62 were conceived spontaneously. Blood samples from all the volunteers and buccal swabs from a subset of individuals were taken by study nurses following a standard protocol. All volunteers were informed about the research purposes and signed an informed consent. This study was approved by the ethical committee of the UZ Brussel.

3.2 Stem cell culture and generation of iPSCs

Four human embryonic stem cell lines were kept in culture: VUB02, VUB03, VUB04 and VUB07. All ESC lines were derived from embryos donated by couples undergoing ART (*133, 134*) and kept in NutriStem® hPSC XF Medium (Biological Industries) in 10 µg/mL laminin (LN-521, BioLamina) coated dishes. The medium was renewed daily. DNA samples were taken at different passages as summarized in the table below.

Stem Cell line	Early passage	Mid passage	Late passage
VUB02	P22	P57	P209
VUB03	P19	P60	P127
VUB04	P27	P66	P125
VUB07	P22	P57	P195

In-house derived iPSCs were generated by lentiviral transduction. Fibroblasts were transduced using a mix of lentiviral vectors provided by the Laboratory of Molecular and Cellular Therapy (LMCT) of the VUB, 88% Dulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with 10% fetal calf serum (FCS), 1% non-essential amino acids (Gibco) and 1% L-glutamine, and 10 µg/mL protamine sulfate. Twenty-four hours after transduction, the cells were washed with phosphate buffer saline (PBS, Gibco). The medium was renewed daily until day 7. On day 8, the cells were passaged in a 1/1 ratio and kept in NutriStem on 10 µg/ml laminin coated dishes. After 20 days in culture, iPSC colonies were identified by live staining for TRA-1-60. Briefly, the cells were incubated in a 1/100 dilution of the mouse anti-human anti-TRA-1-60 primary antibody (MAB4360, Millipore) for 60 minutes at 37°C, followed by the secondary antibody AlexaFluor 488 goat anti-mouse IgM (A21042, Invitrogen) for 20 minutes at 37°C. The staining was visualized using an EVOS FL (Thermo Fisher). The colonies positive for TRA-1-60 were cut from the dishes and transferred in separate wells of a 24 well-plate, resulting in clonal iPSC lines. The cells were kept in NutriStem on 10 µg/mL laminin coated dishes.

In this project, we also used DNA isolated from iPSC lines derived at the laboratory of Prof. Stephane Viville from the University of Strasbourg. In this case, 9 lines were derived from the same source, using different reprogramming methods (see table below).

Table 2. Cell lines and samples studied in this project

Line name, passage (p)	Cell type, reprogramming
RV ONSL1, p7	RV ONSL
RV ONSL2, p7	RV ONSL
CD13, p52	ONSL, linear DNA insertion
32B4, p53	ONSL, linear DNA insertion
LV ONSL2, p8	LV ONSL
RV OSKM2, p9	RV OSKM
RV ONSL + OSKM1, p18	RV ONSL+ RV OSKM
RV ONSL + OSKM2, p15	RV ONSL+ RV OSKM
RV OSKM1, p27	RV OSKM
CPRE2, p9	Adult dermal fibroblasts

RV: retroviral vector, LV: lentiviral vector. All constructs are polycistronic. O: OCT4, N: NANOG, S: SOX2, L: LIN28, K: KLF4, M: C-MYC.

3.3 DNA extraction and mtDNA enrichment

DNA from blood samples was extracted using a Chemagic DNA blood kit special on a chemagic Prepito®-D instrument, according to the manufacturer's instructions. DNA from bulk of stem cells and iPSCs was isolated by proteinase K – SDS lysis and extracted by phenol/chloroform gradient followed by ethanol precipitation.

For mtDNA enrichment, long-range PCR was performed using two different primer sets covering the full mtDNA sequence. Set 1 generated an amplicon of 12.96 Kb (5041 forward-1424 reverse) and set 2 generated an amplicon of 5.3 Kb (550 forward-5895 reverse). The sequences of primer set 1 are: 5'-AGCAGTTCTACCGTACAAC-3' (forward) and 5'-ATCCACCTTCGACCCTTAAG-3' (reverse), and of primer set 2 are: 5'- TTGGTTTGGGGTTTCTGTGGG-3' (forward) and 5'- GTGAGTCGGTAAAATGGAGTG-3' (reverse).

The amplification was done by using a proof-reading polymerase for long-range PCR specifically designed to efficiently amplify long DNA fragments. The reaction mix per sample contained 10 µL of 5x LongAmp Buffer, 7.5 µM of each dNTP, 2 µL of each primer (at a 10 µM solution), 2 µL Taq polymerase (5 units) and 50 ng DNA in a total of 50 µL (LongAmp Taq DNA Polymerase kit, New England Biolabs). The PCR protocol for primer set 1 is described in the table below. For primer set 2, the elongation step was reduced to 5 minutes, as it generated a smaller amplicon. All PCRs were done on the Veriti 96 Well Thermal Cycler (Applied BioSystems). Electrophoresis on agarose gel (1%) was performed to confirm successful amplification.

Step 1	**Step 2**	**Step 3**	**Step 4**
Initiation	*8 cycles (touchdown)*	*22 cycles*	*Final*
30 sec at 94°C	15 sec at 94°C	15 sec at 94°C	11 min at 65°C
	30 sec at 64°C – 0.4°C per cycle	30 sec at 61°C	
	11 min at 65°C	11 min at 65°C	

3.4 Real-time PCR

Quantification of the mtDNA copy number was done using quantitative real-time PCR (qPCR). TaqMan assays were designed and probes binding to specific regions of the mtDNA (3319, 9755 and 11848) were used and normalized to the reference gene *B2M* (Beta 2 microglobulin). Reactions were performed in MicroAmp® optical 96-well reaction plate (0.1 mL, Thermo Fisher) on an Applied Biosystems ViiaTM 7 (Thermo Fisher) in a total volume of 20 µL. The master mix for the customized probes 3319, 9755 and *B2M* was: 10 µL mastermix (Eurogentec), 0.9 µL forward primer, 0.9 µL reverse primer, 0.625 µL probe, 1 ng of DNA input. For probe 11848 (universal probe 14, Roche): 10 µL mastermix (Eurogentec), 0.72 µL forward primer, 0.72 µL reverse primer, 0.4 µL probe and 1 ng of DNA input. Sequences of the primers and the probes can be found in the table below. All reactions were done in technical triplicates.

Locus	Sequence probe	Sequence forward primer	Sequence reverse primer
3319	TTACCGGGCTCTGCCATC	CACCCAAGAACACGGTTTGT	TGGCCATGGGTATGTTGTAA
9755	CTCAGAGTACTTCGAGTC	CGTATTACTCGCATCAGGAGT	TAGATGCCGTCGGAAATGGT
11848	CTGGGAGA	CCTCGCTAACCTCGCCTTA	GGAGAACGTGGTTACTAGCACA
B2M	TTGCTCCACAGGTAGCTCTAGGAGG	TGCTGTCTCCATGTTTGATGT	TCTCTGCTCCCCACCTCTAAG

3.5 Massive Parallel Sequencing (Next Generation Sequencing)

A library was prepared by end repair, adenylation and paired-end adapters ligation of the fragments using the TruSeq DNA PCR-free Library Preparation kit (*Illumina*). After the library preparation, the long (primer set 1, 12.96 Kbp) and short (primer set 2, 5.3 Kbp) amplicons were pooled in 0.35 ratio (short amplicon/long amplicon) to balance the number of molecules. Before sequencing, the generated PCR amplicons (≥1 µg) were sheared using a Covartis M220 sonicator (*Life Technologies*), following instrument specification, to retrieve fragments of ± 350 base pairs. Sequencing was done on a MiSeq or a HiSeq using a MiSeq or HiSeq Reagent Micro kit, v2 (*Illumina*).

3.6 Data analysis and bioinformatic processing

The sequence data were aligned to the Cambridge reference sequence (NC_012920.1) with Burrows-Wheeler Aligner (BWA). The generated files were then uploaded to mtDNA server to detect the homoplasmies (>97% heteroplasmic load) and to determine the haplogroups. The same files were also analyzed with MuTect to detect heteroplasmies (<97% heteroplasmic load). The heteroplasmic variants were also confirmed by mtDNA server and CLC Bio Genomic Workbench.

Following exclusion criteria were used: <0.1 Forward/Reverse reads balance, <20 Phred quality score, <1.5% heteroplasmic load, <1000 coverage.

3.7 Statistics

Statistical analysis was done using the Fisher's exact test with a 95% confidence interval.

4. Results

4.1 Optimization of the method

First, we optimized our method to analyze mtDNA variants using mtDNA server and MuTect. mtDNA server is an online tool that analyzes homoplasmic and heteroplasmic variants in the mtDNA and defines their relative frequencies and pathogenic potential, along with haplogroup determination and possible cross-sample contaminations. To strengthen the consistency of the single nucleotide variants (SNV) calling and to analyze small insertions and deletions, we used MuTect. This software was developed for the detection of rare mutations in a mixed population of cells (*135*).

Some nucleotide positions were excluded from the analysis for specific reasons, as indicated in the table below.

Nucleotide position(s)	*Sequence*	*Variant*	*Reason*
303	aaccccccct	Insertion C	Recurrent*
310-316	aacccccccctccccgcttctggcca	-	Recurrent*
513-515	atcctacccagcacacacac	-	Recurrent*
550-800		-	First 250 bp of forward primer
955	gttttgatcaccccc	C>A	Recurrent*
1174-1424		-	First 250 bp of reverse primer
5041-5291		-	First 250 bp of forward primer
5192	aacaccttaat	C>A	Recurrent*
5535-5785			First 250 bp of reverse primer
12148	aatatagtttaa	Deletion A	Recurrent*
15407	cgataaaatc	-	Recurrent*
16180-16199		-	Recurrent*

*Recurrently present across samples, it is in a homopolymeric stretch, prone to PCR/sequencing artifacts

Since our experimental set up employed the pooling of 2 different amplicons of different lengths, we first evaluated the coverage depth along the full mtDNA sequence resulting from our pipeline (described in Zambelli et al., under revision), to prove that all the regions were equally represented. As seen in figure 5, the mtDNA is highly covered (about 50000x) and the two amplicons are proportionally represented. We observed an extremely high coverage in the regions where the primer sets are binding, but this is consistently observed in different amplicon based strategies and with different primer sets. It is probably due to the generation of incomplete and short DNA fragments during the amplification that cause over-representation of the regions close to the primers. Hence, we decided to not consider any variant located in the initial 250 base pairs of every amplicon. Furthermore, we decided to reduce the number in reads to get an average coverage of 10000x, which is more than sufficient for our analysis and results in a faster bioinformatics processing.

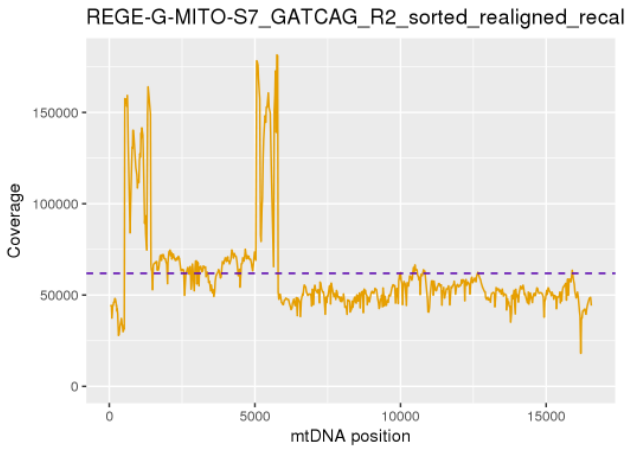


Fig.5. Example of a coverage plot by mtDNA server. Primer binding sites show a higher number of reads.

Next, we checked the mtDNA copy number of different tissues. We optimized our protocol by using three different probes binding to three different regions of the mtDNA (3319, 9755 and 11848). For this purpose, we used tissues of which the mtDNA copy number was already described in literature. Energy-demanding tissues, like liver and muscle, are known to possess a high copy number, while blood cells and fibroblast have a lower mtDNA content. We saw that all three probes provided similar results (figure 6). In muscle, 11848 gave a lower mtDNA copy number when compared to the other two probes. This was expected as the sample was known to have a deletion in that region of the mtDNA.

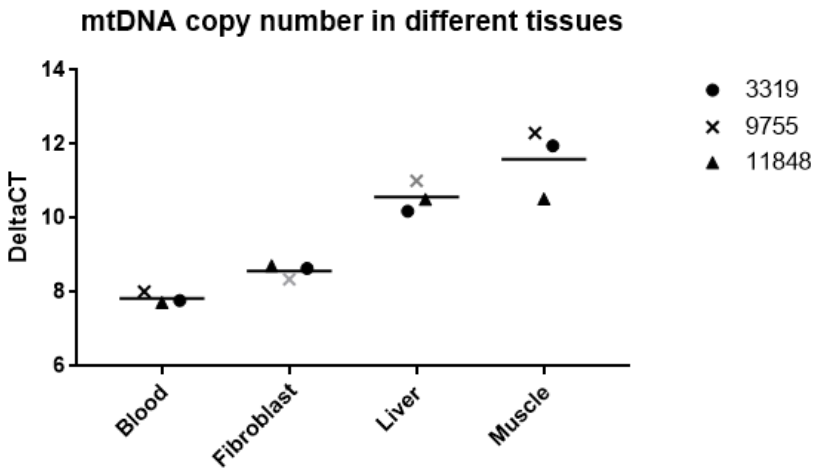


Fig.6. qPCR results of mtDNA copy number in different tissues. Each point represents the different probes, dot: 3319, cross: 9755 and triangle: 11848. Results are shown in deltaCT and are normalized to B2M as a nuclear control. All tests were done in technical triplicates.

4.2 Deep sequencing of the mtDNA in blood samples: ICSI vs. spontaneously conceived individuals

The second part of this study investigated if young adults born after ICSI show a higher mutation load in their mtDNA when compared to a control population. We examined the DNA of 119 volunteers of which 57 were conceived via ICSI and 62 were spontaneously conceived.

4.2.1 Tissue selection

Firstly, we explored which tissue would be more suitable to investigate mtDNA. Buccal swabs and blood samples of a subset of volunteers (n=12) were taken and homo- and heteroplasmies were studied.

Our analysis showed that both tissues yielded corresponding results (table 3). Four ICSI subjects displayed the same variants in both tissues in similar frequencies, except for one volunteer who had one variants with a difference of 16% (61% in blood vs. 77% in buccal swab). Also, in our control group, we observed correspondence between the two tissues, except for one volunteer who showed a difference of 14% in heteroplasmy level (40% in blood vs. 54% in buccal swab). Another three ICSI subjects and one control volunteer had no heteroplasmies but had identical homoplasmic variants. Heteroplasmic variants found are summed up in table 3. All comparison results can be found in the supplementary data.

Blood samples were preferred for analysis because we already had these samples obtained from more volunteers, since this tissue is more accessible and analyzing only one tissue reduced the cost of the study.

	ICSI_1	ICSI_2	ICSI_3			ICSI_4
Variant	236 A>G	204 T>C	1119 T>G	10026 T>C	12618 G>A	16234 C>T
Blood	61%	0%	5%	9%	16%	6%
Buccal swab	77%	3%	7%	9%	12%	3%
	Control_1	Control_2	Control_3		Control_4	
Variant	204 T>C	8557 G>A	309 C>T	6755 G>A	215 A>G	
Blood	3%	40%	2%	2%	0%	
Buccal swab	2%	54%	2%	3%	3%	

Table 3. Frequencies of heteroplasmic variants detected in blood and buccal swabs in the same individual.

4.2.2 mtDNA copy number in blood

Secondly, we checked the mtDNA copy number in 16 blood samples of our volunteers (8 ICSI and 8 controls). This was done to confirm that all samples were comparable for analysis (figure 7). For this purpose, we performed a quantitative real-time PCR (qPCR) using three different probes, as described to have a higher variance. We also checked if there was a difference between female and male volunteers because it is described in literature that sex-specific hormones can influence mtDNA copy number (*136*). Of the 16 volunteers we examined, half of them were female and showed to have no increased copy number.

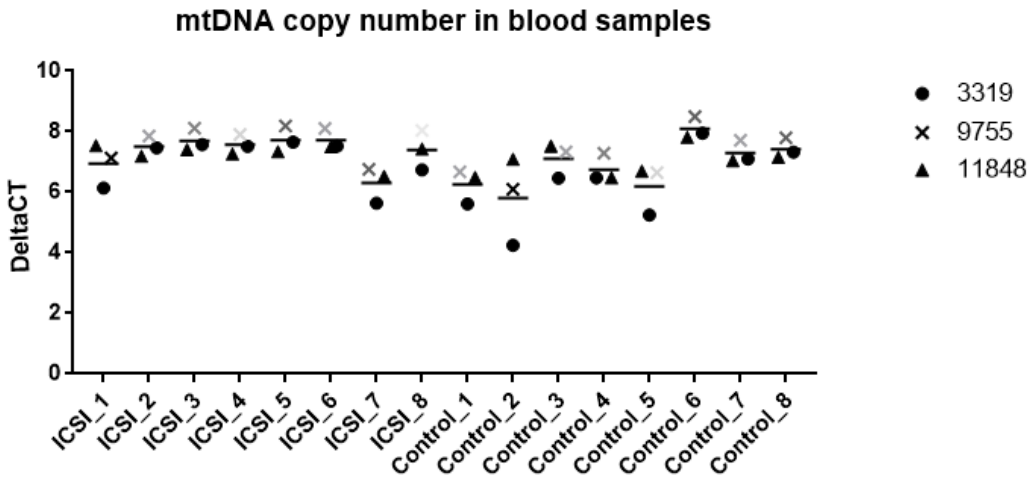


Fig.7. qPCR results of mtDNA copy number in ICSI subjects compared to control subjects. Each point represents the different probes, dot: 3319, cross: 9755 and triangle: 11848. Results are shown in deltaCT and are normalized to B2M as a nuclear control. All tests were done in technical triplicates.

4.2.3. Haplogroup determination

Thirdly, we evaluated the incidence of different haplogroups in both groups. Aside from the main prevalence of haplogroup H in both groups, we observed that haplogroup J was three times more abundant in the ICSI group (16% vs. 5%) and that haplogroup U was almost twice as much represented in the control group (10% vs. 18%). Other minor haplogroups were equally distributed in both groups (table 4).

Haplogroups ICSI	Frequency	N° individuals	Haplogroups Control	Frequency	N° individuals
H	42%	24	H	42%	26
I	2%	1	I	2%	1
J	16%	9	J	5%	3
K	14%	8	K	10%	6
T	9%	5	T	12%	8
U	10%	6	U	18%	11
V	2%	1	V	3%	2
W	0%	0	W	3%	2
X	5%	3	X	5%	3
Total	100%	57	Total	100%	62

Table 4. Distribution of haplogroups amongst the volunteers in the ICSI and control group.

4.2.4 Homoplasmies

The homoplasmic variants (i.e. variants occurring at >97% frequency in the samples) that were not included in the haplogroup of the individual were examined into further detail.

An initial categorization of the variants was made into local and global private variants. Local variants are variants described in PhyloTree (http://www.phylotree.org/), while global variants are variants that have not been yet reported. The distribution (located in coding or non-coding region) and the type (inducing synonymous or non-synonymous changes) of the variants were also analyzed. Non-synonymous variants were subdivided in likely to induce a pathogenic change (MutPred score >0.5) and likely non-pathogenic ones (MutPred score <0.5).

We found that there are no differences in the number of individuals showing no homoplasmies, homoplasmies in the coding region, in the non-coding region or in both (table 5). In addition, we observed the number of people carrying local, global or both variants. We saw that the results were similar, however, we observe a slight increase in number of individuals carrying global variants in the ICSI group (table 5). Also, the pathogenicity of the variants was investigated.

Homoplasmic variants	ICSI	Control
No homoplasmies	21% (12)	26% (16)
Coding regions	35% (20)	29% (18)
Non-coding regions	9% (5)	13% (8)
Coding + non-coding regions	35% (20)	32% (20)
Local	44% (25)	55% (34)
Global	12% (7)	8% (5)
Local + Global	23% (13)	11% (7)
Pathogenic	17% (10)	18% (11)
Non-pathogenic	7% (4)	14% (9)

Table 5. Percentages of individuals showing no homoplasmies outside the haplogroup, only in the coding region, only in the non-coding region or in both. Also, the distribution of people carrying local and/or global variants was analyzed as well as the pathogenicity.

4.2.5 Heteroplasmies

After the analysis of homoplasmic sites, we moved to the analysis of the heteroplasmic ones (i.e. variants occurring between 0% and 97%). We first looked at the distribution of the variants in the mtDNA. Again, we obtained similar results in both groups, with 86% of the variants being SNVs in the ICSI compared to 91% in the control group. The numbers of insertions and deletions were too small to compare the two groups (6 insertions and 3 deletions in the ICSI group and 5 insertions and no deletions in the control group). In addition to the previously performed analysis, we could also asses the frequency of the single variants and the cumulative frequency of heteroplasmy retrieved in single individuals, obtained by adding the frequencies of the single nucleotide variants together.

When the total amount of variants was considered, we did not detect a significantly higher cumulative frequency of variants in one of the groups (figure 8).

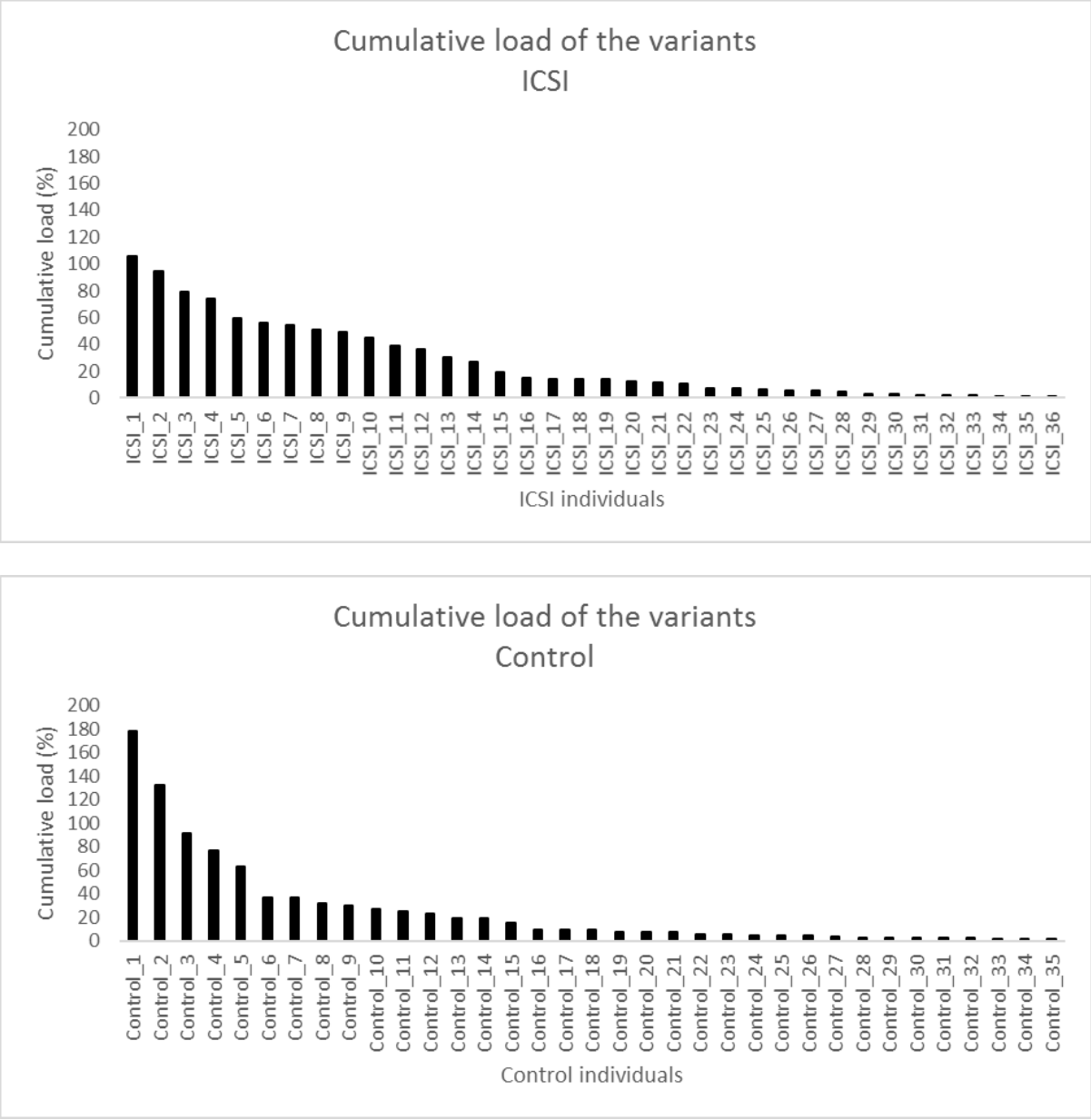


Fig.8. Cumulative variant load in percentage in ICSI and control individuals.

On the other hand, when we restricted the analysis to the variants located in coding regions and excluded the non-coding region (and particularly the hypervariable region), we observed a significantly higher presence of variants in the ICSI group. Figure 9 shows the cumulative frequencies of all the individuals with heteroplasmies in the coding region as well as the distribution in both groups.

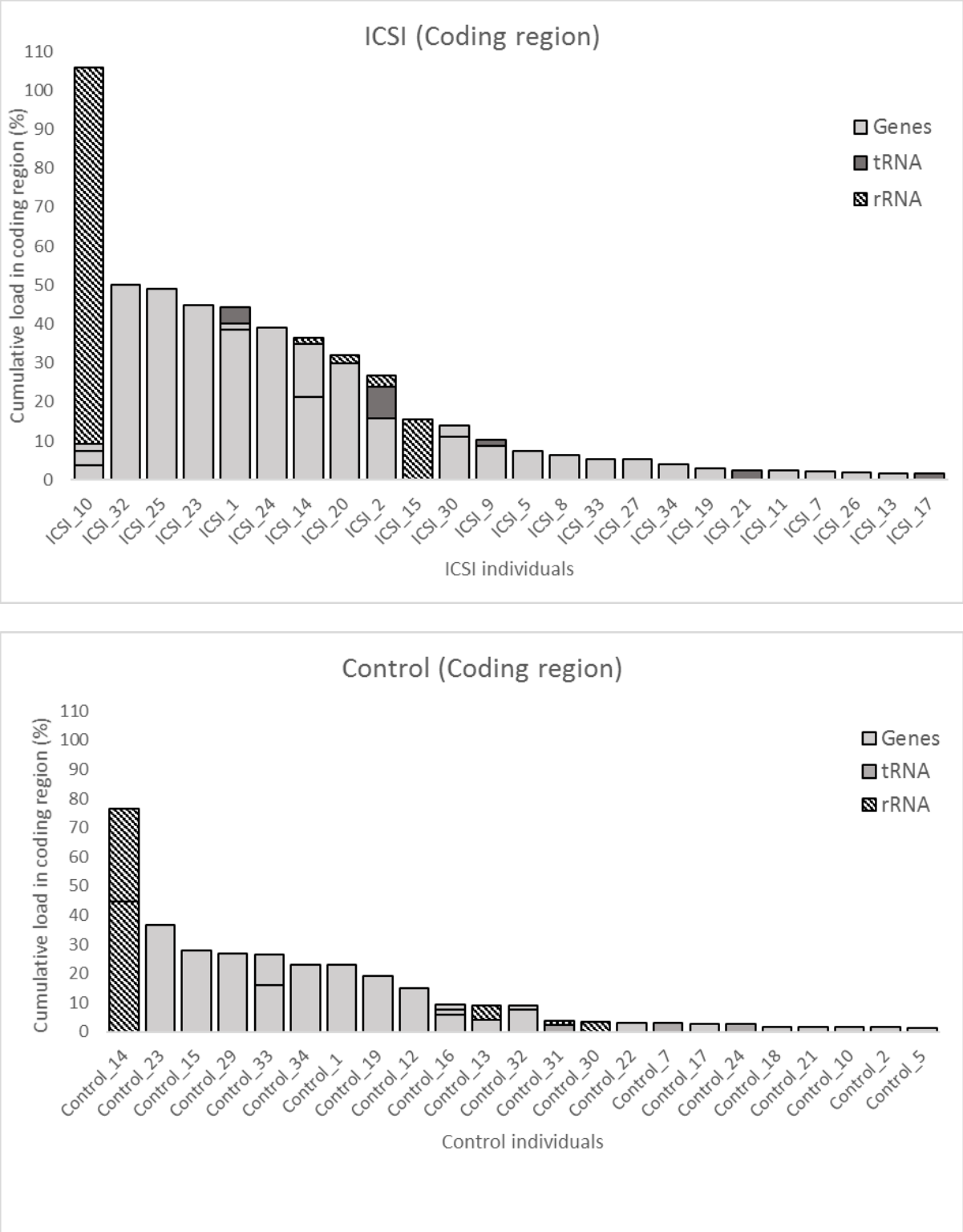


Fig.9. Cumulative variant load in the coding regions in percentage in ICSI and control individuals.
The height of each box represents the frequency of the single variant.

We performed statistical analysis using the Fischer's test to investigate which cut off could give a significant difference in cumulative load in the coding regions (figure 10) and found that the number of individuals showed to be significantly higher when the load is over 30% ($p<0.025$).

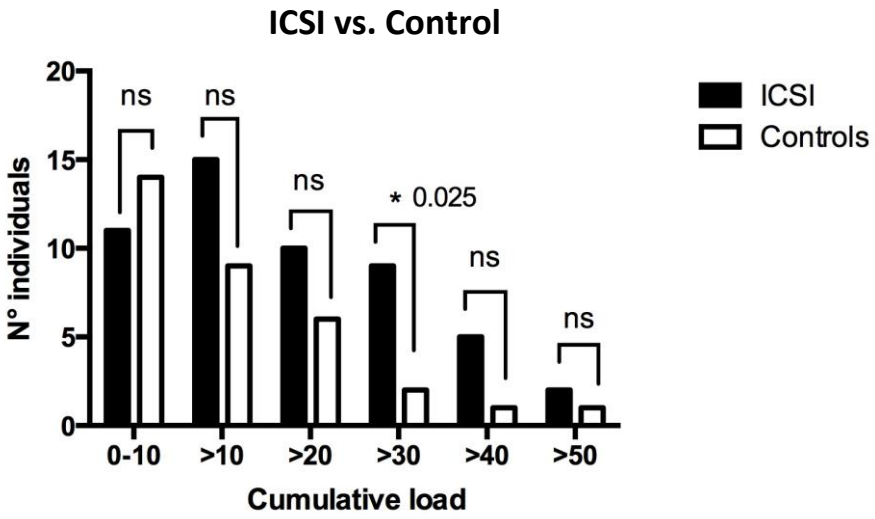


Fig.10. Number of individuals showing a cumulative load in the coding region between 0 and 10%, over 10%, 20%, 30%, 40% or 50%. The number of individuals with a cumulative load over 30% are significantly higher in the ICSI group with a p-value of 0.025. ns: not significant.

We investigated how many variants were found in the coding regions as opposed to the non-coding region and found that the distribution of variants found in the coding region was similar in both groups (60% vs. 53%). In addition, we examined the number of variants in the different genes, rRNAs and tRNAs (figure 11). The numbers are too small to be conclusive (ranging from 0 to 5 variants found per region) but it does not seem that specific regions are more affected by changes than others in both groups. We also didn't find any recurrent variant in the individuals of any group.

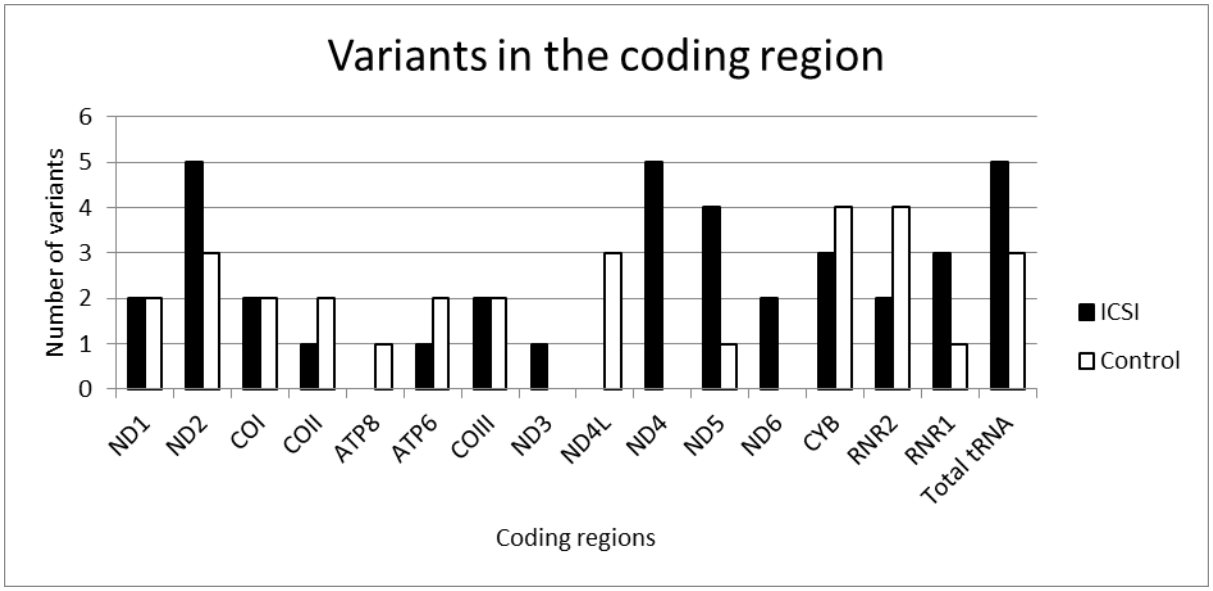


Fig.11. Number of the variants in the different coding regions.

4.3 Pluripotent stem cells: ESCs and iPSCs

The second aim of this thesis was to explore mtDNA integrity and segregation in pluripotent cells, and evaluate if this cellular model can faithfully recapitulate the earlier stages of embryo development. For this, we analyzed both ESCs and iPSCs. For the latter, we investigated four lines at three different passages (early, mid and late) to evaluate also the effect of prolonged time in culture on mtDNA stability. For the iPSCs, we analyzed nine clonal lines coming from two individuals reprogrammed with lentiviral vectors along with nine more lines derived from a third individual but reprogrammed with a wide variety of reprogramming methods. This to evaluate if different reprogramming methods could also influence the mutation load of mtDNA.

4.3.1 mtDNA copy number in pluripotent stem cells

We first compared mtDNA copy number in different cell populations, and compared them with different tissues such as blood, fibroblast, liver and muscle cells to see if the range of copy numbers in ESCs and iPSCs represent the average of copy numbers seen in these tissues. Here, we saw that ESCs and iPSCs have a similar amount of mtDNA copies and show a higher copy number than blood cells and fibroblasts, but have less mtDNA content than energy-demanding tissues like liver and muscle (figure 12).

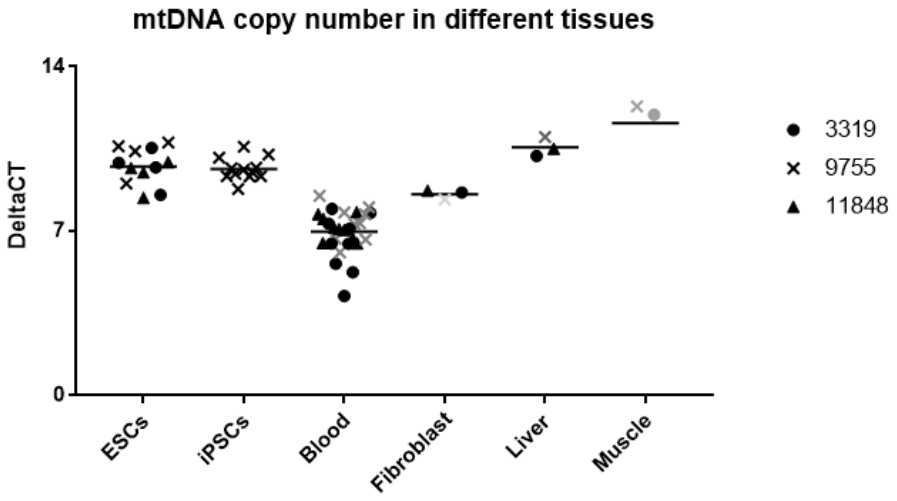


Fig.12. qPCR results of mtDNA copy number in different tissues compared to ESCs and iPSCs. Each point represents the different probes, dot: 3319, cross: 9755 and triangle: 11848. Results are shown in deltaCT and are normalized to B2M as a nuclear control. All tests were done in technical triplicates.

In addition, we examined the copy number of mtDNA in iPSCs and compared them to their source cell population, i.e. fibroblasts. Here, we saw that the number of mtDNA copies present in the iPSCs is higher than the copy number in fibroblasts (figure 13). This finding is remarkable as it is assumed in literature that the mtDNA copy number reduces after dedifferentiation towards iPSCs.

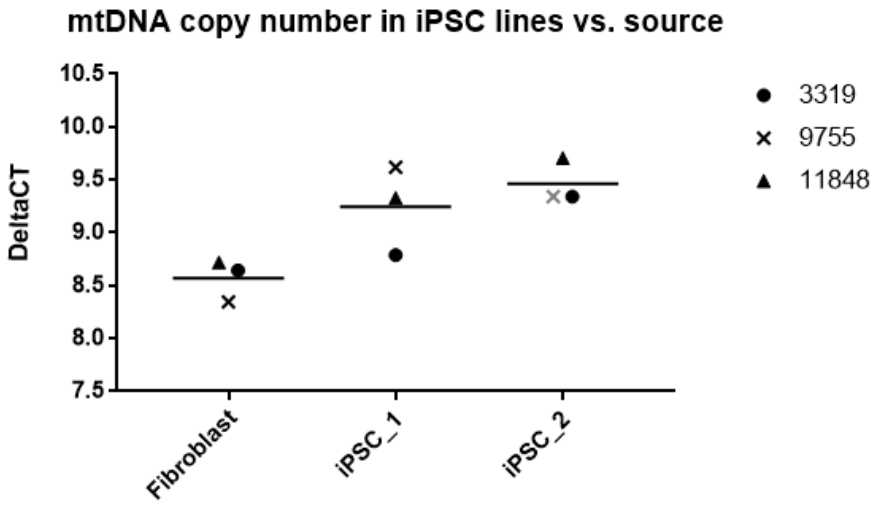


Fig.13. mtDNA copy number in two iPSC clonal lines compared to their source material. iPSCs show to have a higher copy number. Each point represents the different probes, dot: 3319, cross: 9755 and triangle: 11848. Data is shown in deltaCT and was normalized to B2M as a nuclear control. All tests were done in technical triplicates.

4.3.2 Segregation and distribution of heteroplasmies in ESCs

We compared four different ESC lines in three different stages (early, mid and late passages) to investigate the segregation and distribution of variants in an in vitro culture and to see if ESCs after prolonged proliferation can still maintain the distribution of variants in a similar way as the somatic tissues in an individual.

The total number of variants stayed constant after approximately 60 passages and increased significantly after 100 passages for VUB04 and VUB07, while it stayed the same for VUB02 and VUB03 (figure 14). We also investigated the cumulative frequency of the heteroplasmies in the different time passages and found that all ESC lines, except for VUB02, showed a higher cumulative frequency in the late passage when compared to earlier passages (figure 14).

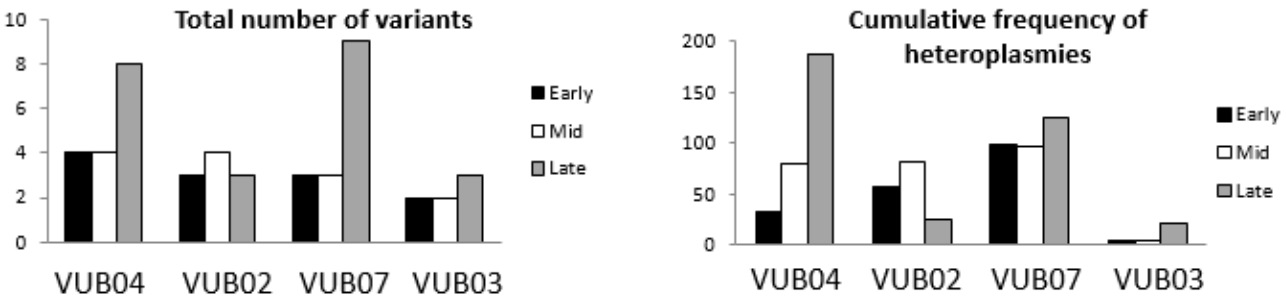


Fig.14. Analysis of heteroplasmic variants in four different ESC lines (VUB04, VUB02, VUB07 and VUB03) at three different time points (early, mid and late passage). Graphs showing the total number of variants and the cumulative frequencies of the variants

Next, we looked at the distribution of the variants over time in the different time points and saw that there was a shift towards an increase of variants in the coding regions (genes, rRNA and tRNA) in late passages, while the number of variants in the non-coding region stayed the same after 60 passages (figure 15). Furthermore, the cumulative frequency of all the variants in the different ESC lines showed a significant increase in the later passages.

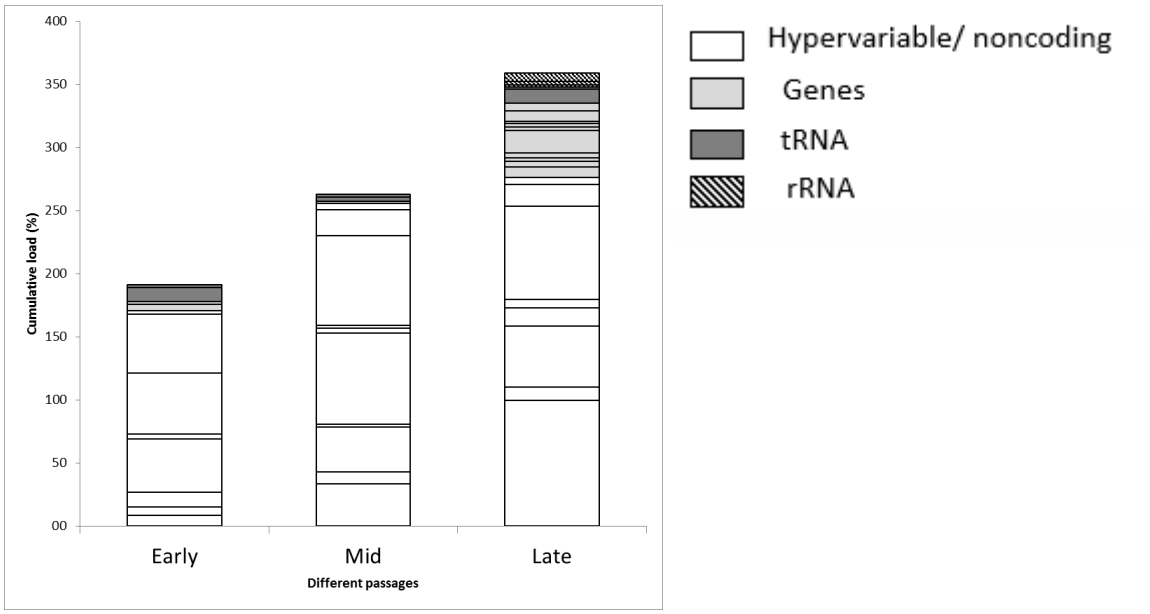
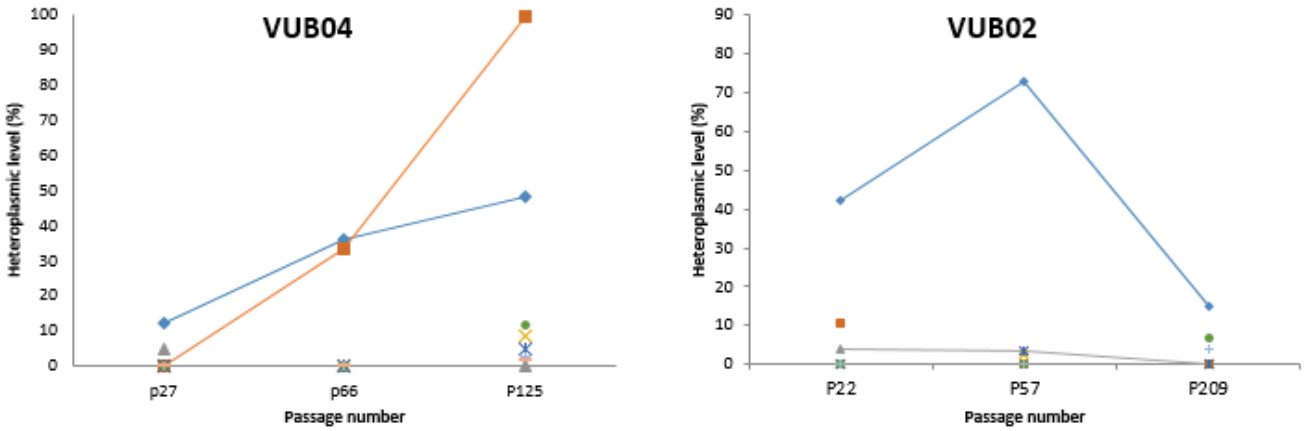


Fig.15. Graph showing the distribution of the variants in the different regions of the mtDNA. All variants of the four different ESC lines are pooled together. Every box accounts for one heteroplasmic variant. White: non-coding region, light gray: genes, dark grey: tRNA, and stripes: rRNA. The height of each box represents the frequency of the single variant.

Subsequently, we investigated the behavior of the heteroplasmic variants throughout passaging. Here, we observed that the variants did not show a specific pattern during their time in culture, some variants acquire a higher frequency while others decline in frequency (figure 16). For ESC line VUB03, we saw that the heteroplasmies exhibited very low frequencies and were not recurrent over different passages.



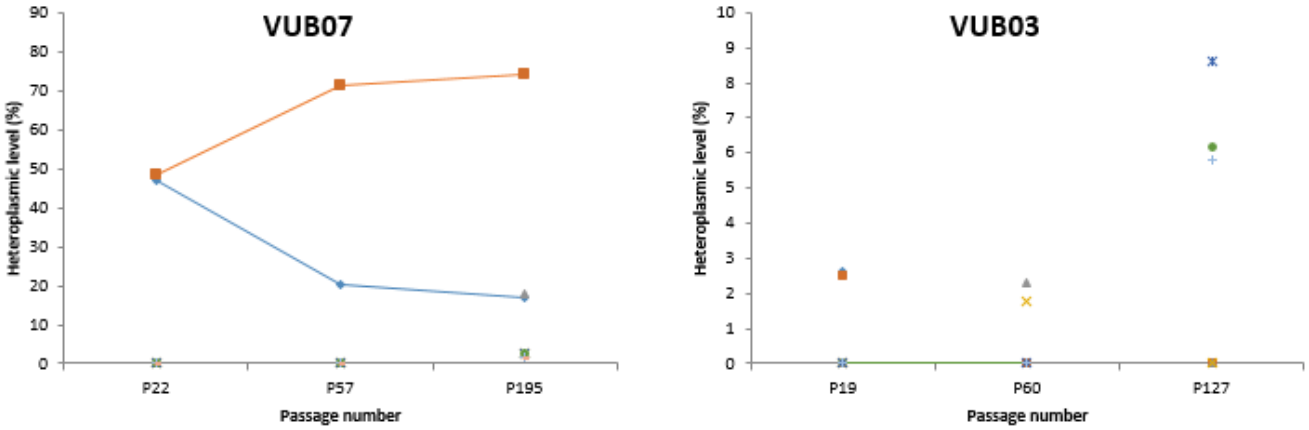


Fig.16. Graphs showing the segregation of several variants over time in culture for the four different ESC lines. Every dot represents a variants and variants present in different passages are connected with a line.

4.2.4 iPSCs as a representative for the individual?

Lastly, we studied iPSCs to examine the distribution of mtDNA variants and to investigate potential differences when comparing them to embryonic stem cells and cells derived from somatic tissues. We used four hESC lines and nine iPSCs derived from the same individual. Analysis showed that the number of mtDNA variants is similar between ESCs and iPSCs.

Interestingly, we found that the variant distribution was peculiar as 66% of the variants were falling in the coding region in ESCs, and this was comparable with somatic tissues such as blood (57%) and buccal swabs (58%), while for iPSCs the variants in the coding region represent the vast majority of the total 95% (figure 17). When looking at the distribution and frequencies of the variants in different iPSC lines derived from the same individual, this was significantly different while it stayed relatively constant for the ESC lines (figure 18) Furthermore, when analyzing potentially pathogenic changes, there seemed to be a 2.5-fold increase in the iPSCs when compared to the hESCs.

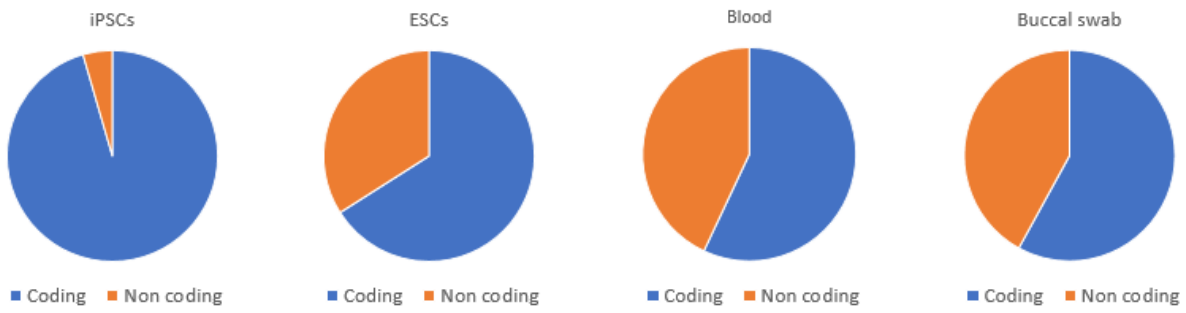


Fig.17. Distribution of the variants in the coding versus the non-coding regions in iPSCs compared to ESCs, blood tissue and buccal swab.

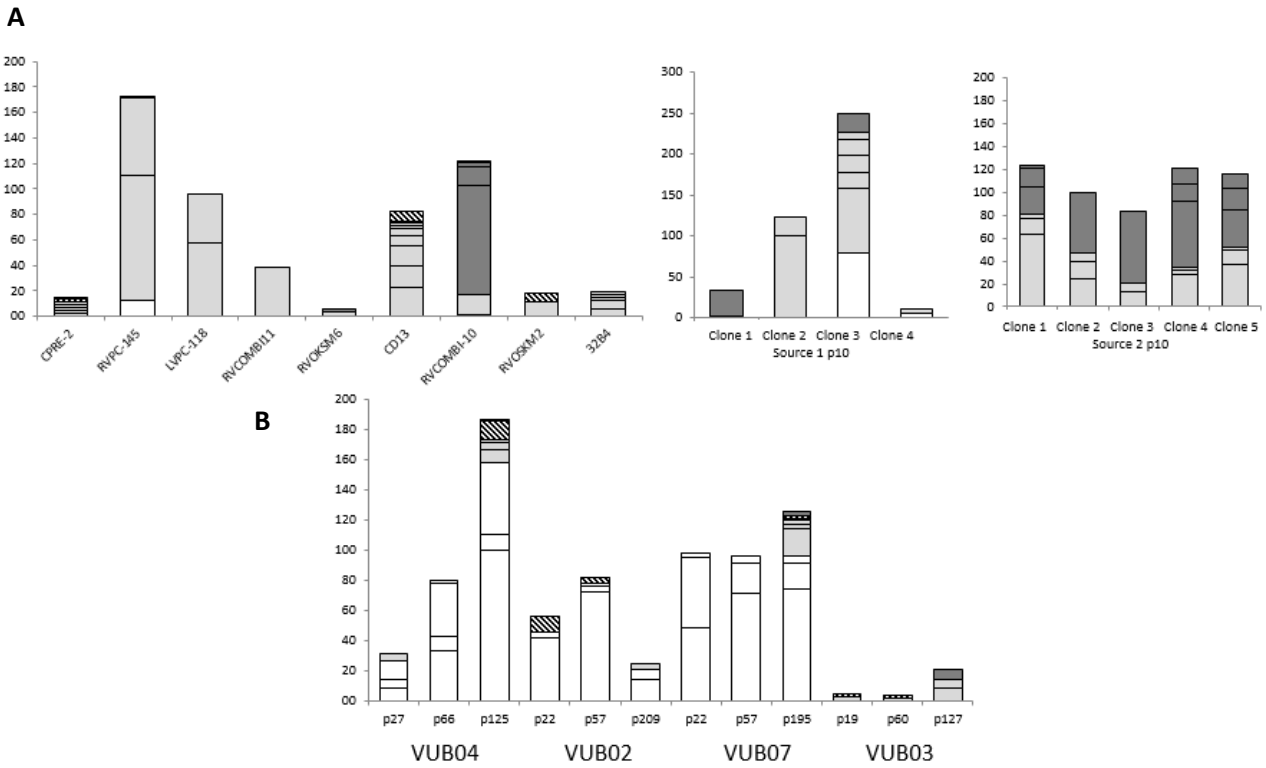


Fig. 18. Graphs showing the distribution of the variants in the different regions in iPSCs (**A**) and ESCs (**B**). Every box accounts for one heteroplasmic variant. White: non-coding region, light gray: genes, dark grey: tRNA, and stripes: rRNA. The height of each box represents the frequency of the single variant.

5. Discussion

The first part of this research explored the hypothesis that children conceived through ART have a higher variant load in their mitochondrial genome than their spontaneously conceived peers.

Blood samples were compared with buccal swabs in a subset of volunteers at the start of our investigation, to evaluate if the variants detected were comparable in term of number and frequency. The different samples have a different embryonic origin (mesodermal lineage for blood and meso-ectodermal for buccal swabs), making it possible to distinguish between inherited variants, or variants present from very early stages of development on, and de novo somatic mutagenesis. We found that both tissues yielded similar results, apart from two heteroplasmies that showed to have a difference of more than 10% in absolute frequency. Nevertheless, this difference has very limited biological relevance. The matching of this preliminary set of individuals made us also consider that the vast majority of the variants in the individuals were already present at very early stages of development, possibly inherited with very limited asymmetric distribution amongst tissues, and not acquired de novo. This equivalence made us decide to continue analyzing blood samples only.

We then compared the mtDNA copy number in the blood samples between the different groups and saw that there was no significant difference. This was done to confirm that we could compare the heteroplasmic levels of all volunteers. From our limited dataset, it appeared that the control group had a higher variance in mtDNA copy number. Knez et al. studied the mtDNA copy number in blood in the general population (*136*). They concluded that mtDNA copy number increases until the fifth decade of life and declines afterwards. This could be because of an accumulation of somatic

mutations and the cellular response of autophagy. They also showed that mtDNA copy number is influenced by estradiol and hereby confirmed the study of Lopez et al. in 2012 (*137*). In our analysis, all our volunteers were of the same age (18 years old) and we compared the mtDNA copy number of 8 female with 8 male volunteers and saw that our population showed no great differences in mtDNA content.

Next, we looked at the haplogroup distribution of our subjects in both groups. Haplogroup H accounted for the majority of both groups (42%) which reflects the prevalence of the haplotype in Europe (40-50%) (*138*). We observed a higher presence of haplogroup J in the ICSI group (16% vs. 5%). In the general European population, haplogroup J has an incidence of approximately 9% (*139*). Haplogroup J is associated with lower efficiency of electron transport chain, less mtDNA and mitochondrial RNA which results in a lower concentration of polypeptides (*140, 141*). In our control group, we saw that 18% of the people exhibited haplogroup U as opposed to 10% in the ICSI group. We must hold into account that the people born via ICSI are 18 years old and were born in the early years after the ICSI-technique was published (*6*). Our control population is of the same age, but may represent a different population as the composition of the population changed in the past years.

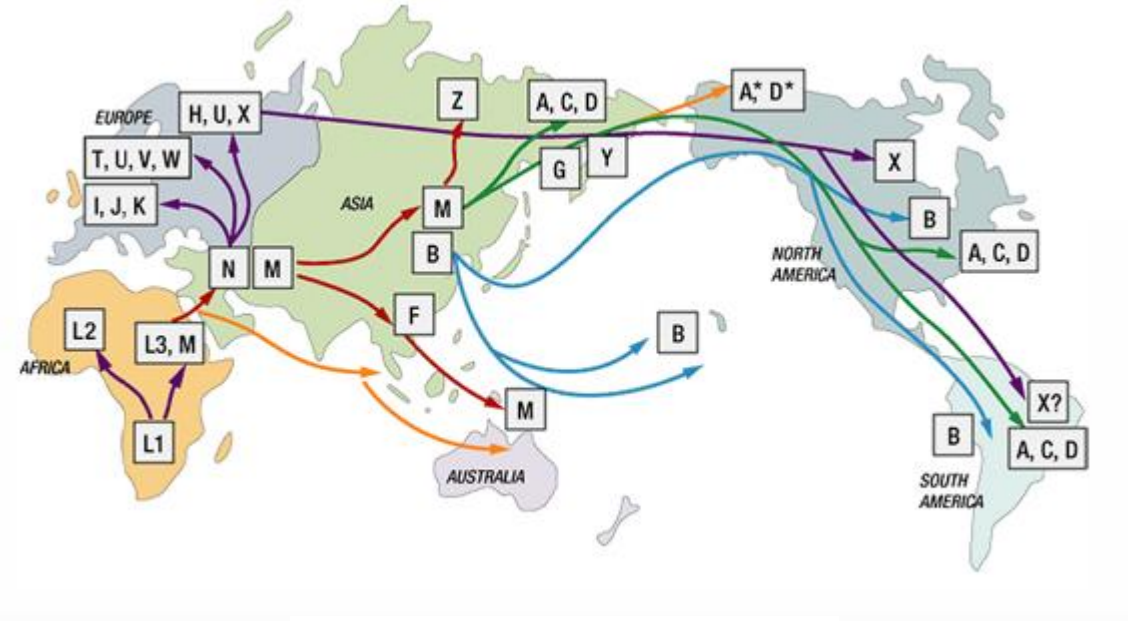


Fig.19. Haplogroup distribution around the globe.

After haplogroup determination, we examined the homoplasmic variants that were not present in the haplogroup. From our data, we saw that there were more individuals showing global variants in the ICSI group when compared to the control group. This higher presence of homoplasmic global private variants might underlie a correlation with maternal infertility, as these variants have not yet been described. However, a more accurate study of single variants and genes involved is needed.

The analysis of the heteroplasmies showed that the total number of variants between the different groups is not significantly different, meaning that ART does not seem to influence the appearance of de novo variants as such. Conversely, when we analyzed the frequency of the variants, we found a shift towards a higher SNV load in the ICSI individuals, with a statistically significant difference in number of individuals with a cumulative coding SNV load above 30%. This might suggest that ART procedures in general, and possibly controlled ovarian hormonal stimulation, have an effect in increasing the frequency of pre-existing variants, as suggested for somatic mutagenesis of ageing cells. This finding proves our hypothesis that oocytes that underwent controlled ovarian stimulation carry a higher SNV load in their mtDNA and that these variants are maintained in the offspring. We

speculate that hormonal stimulation causes an increase in the oxidative stress that mimic the same mutagenic environment encountered in the ageing cells, and therefore results in a higher mutation load in the oocytes and later the higher frequencies that are already detected in young individuals. In order to further investigate the origin of this variant distribution, we are currently sequencing placentas from mothers that underwent ART to see if this variant distribution is already present in extra-embryonic tissues in early development and, if possible, we would like to investigate the mtDNA variants of the mothers as well.

In the second part of this study, we investigated the potential of ESC and iPSCs as a representative model for early development regarding the segregation and distribution of mtDNA variants.

First, we investigated the mtDNA copy number of ESCs and found that this is similar to that of iPSCs. From the literature, we know that fibroblasts have copy numbers ranging between 2000 and 3000 molecules per cell, and this number correspond to the number found in ESCs (*142*). As an addition, we compared the mtDNA copy number of two different iPSC lines with their source material in order to confirm the hypothesis that mtDNA copy number will reduce during reprogramming, as it assumed in literature. However, in our study, we saw remarkably that the mtDNA copy number does not decrease but is slightly increased, after reprogramming, and is similar to the copy number in ESCs.

We then studied the effects on mtDNA integrity in a prolonged in vitro culture. Here, we studied the segregation and distribution of variants in ESCs, after long-term passaging. We found that the number of variants stays constant after 60 passages. Despite that, when the culture was prolonged until over 100 passages, the total number of variants increased and the distribution changed, with an increase in variants in the coding region. Overall, the cumulative load increased with passage number. Again, this points towards a more “aged” phenotype of the cells, with a higher frequency of variants and a higher incidence of polymorphisms in the coding regions. Although ESCs are dependent on the glycolytic pathway, because they derive from cells that in vivo reside in a hypoxic environment, they are kept in culture in a normoxic environment. This can induce mtDNA mutations, in analogy to the somatic mutations in vivo.

Regarding iPSCs, we found that multiple clonal lines derived from the same individual exhibited remarkable variations in the type, frequency and distribution of the variants when compared to ESCs, blood and buccal epithelia. This could reflect cellular mosaicism. iPSCs generate from one cell and when a clonal population is derived, the genotype of the founder cell will be observed, therefore the pattern of mutations that are retrieved can greatly deviate from the one observed in the source tissue as a whole. To strengthen this hypothesis, we performed a comparison between fibroblast bulk cultures and single fibroblasts. The variation found in the single cells was comparable to the variation found in the iPSCs clones (Zambelli et al., under revision). The higher mutation level in single cells (especially in the coding regions) are probably arising from somatic mutagenesis, which is in line with the work of Kang and colleagues (*143*) that showed how iPSCs derived from older patients carried a higher level of mutations than the one derived from younger individuals, regardless of the tissue used for the reprogramming.

In our results, these variants were 95% in the coding region. We hypothesize that the majority of the coding variants are reflecting somatic mutagenesis in the donor individual, rather than being inherited from their mothers. Noteworthy, we investigated iPSC lines from different reprogramming methods and still observe this shift in distribution. Therefore, we can exclude possible variants induced by reprogramming methods.

For the follow-up experiments, we have collected 120 single cells from three different ESC lines after 20 and 80 passages to explore if the distribution of mtDNA variants is different than the bulk material. The amplification is done and we are currently waiting for the sequencing results.

6. Conclusion

With this investigation, we provide evidence that people born after ICSI have a higher cumulative variant load in their coding region than naturally conceived children. We also observed a slight increase in individuals showing homoplasmic global variants in the ICSI group. These variants are not described yet in literature and can play a role in maternal subfertility.

We explored the use of ESC lines as a model to investigate mtDNA variant distribution and segregation and found that this model is representative for the individual as it showed similar distribution of the variants in the early passages, reflecting the developmental processes (childhood to adulthood), and in the later passages, reflecting the ageing process in the individual. In addition, we examined if iPSC lines can be used a model as well to represent the individual and find that they reflect the somatic variance of the individual.

7. Acknowledgements

I would like to thank my promoter Prof. Dr. Claudia Spits for giving me the chance of fulfilling my internship at the REGE lab and for guiding me through the projects. I enjoyed our little meetings and I learned a lot from them. I'm looking forward to continuing our collaboration. I would like to thank my supervisor Filippo Zambelli for helping me on a day-to-day basis. Thanks for answering all my questions and for keeping me sane every time I had a stress attack. You taught me what it means to be a good scientist and I will always be grateful for that. Thank you both for helping me to accomplish this thesis.

I would also like to thank my friends, especially Delphine and Fien, for all the little talks we had during our time in the lab and for always making me laugh. I am really lucky to call you my friends. Thank you to my parents who always supported me and encouraged me to do my best to take my chances. Lastly, I would like to thank my boyfriend Stijn, for bearing with me and to support me when I had a hard time this year. I wouldn't know what to do without you.

8. Supplementary data

Tables showing the comparison between blood and buccal swab tissue in the ICSI group (n=7) and in the control group (n=5).

ICSI_1			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a	H2a
Homoplasmies			
9575	G>A	✓	✓
9932	G>A	✓	✓
13885	C>T	✓	✓
16399	A>G	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
263	A>G	61%	77%
16063	T>C	0%	1,6%

ICSI_2			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a	H2a
Homoplasmies			
73	A>G	✓	✓
14798	T>C	✓	✓
16162	A>G	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
None			

ICSI_3			
Locus	Variant	Blood	Buccal swab
Haplogroup		U5a1c2a	U5a1c2a
Homoplasmies			
5153	A>G	✓	✓
16286	C>T	✓	✓
Heteroplasmies			
204	T>C	0,0%	2,6%
12651	G>A	2,0%	0,0%

ICSI_4			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a	H2a
Homoplasmies			
152	T>C	✓	✓
7805	G>A	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
1119	T>G	5,2%	7,4%
10026	T>C	9,0%	9,3%
12618	G>A	15,9%	11,6%

ICSI_5			
Locus	Variant	Blood	Buccal swab
Haplogroup		X2+225	X2+225
Homoplasmies			
200	A>G	✓	✓
6150	G>A	✓	✓
9449	C>T	✓	✓
15308	A>G	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
None			

ICSI_6			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a2b2	H2a2b2
Homoplasmies			
6261	G>A	✓	✓
Heteroplasmies			
16234	C>T	6,2%	2,6%

ICSI_7			
Locus	Variant	Blood	Buccal swab
Haplogroup		H1c1	H1c1
Homoplasmies			
16519	T>C	✓	✓
Heteroplasmies			
None			

Control_1			
Locus	Variant	Blood	Buccal swab
Haplogroup		K1a+195	K1a+195
Homoplasmies			
7364	A>G	✓	✓
9316	T>A	✓	✓
12651	G>A	✓	✓
16166	A>G	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
204	T>C	2,7%	2,2%

Control_2			
Locus	Variant	Blood	Buccal swab
Haplogroup		K1a4a1g	K1a4a1g
Homoplasmies			
16243	T>C	✓	✓
6722	G>A	✓	✓
9377	A>G	✓	✓
16261	C>T	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
8557	G>A	40,6%	54%

Control_3			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a	H2a
Homoplasmies			
6776	T>C	✓	✓
10837	C>T	✓	✓
13768	T>C	✓	✓
16148	C>T	✓	✓
16325	T>C	✓	✓
Heteroplasmies			
309	C>T	1,6%	2,1%
6755	G>A	1,6%	2,6%

Control_4			
Locus	Variant	Blood	Buccal swab
Haplogroup		U5b3b	U5b3b
Homoplasmies			
16526	G>A	✓	✓
152	T>C	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
None			

Control_5			
Locus	Variant	Blood	Buccal swab
Haplogroup		H2a	H2a
Homoplasmies			
14344	A>G	✓	✓
16519	T>C	✓	✓
Heteroplasmies			
215	A>G	0%	3,5%

9. Abbreviations

(h)ESC	(human) Embryonic Stem Cell
ADP	Adenosine Diphosphate
AMH	Anti-Mullerian Hormone
ART	Assisted Reproductive Technologies
ATP	Adenosine Triphosphate
ATP6/ATP8	ATP Synthase 6/8
B2M	Beta 2 Microglobulin
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
BWA	Burrows-Wheeler Alignment
Cl	Chloride
CO	Cytochrome Oxidase
COS	Controlled Ovarian Stimulation
CPEO	Chronic Progressive External Ophtalmoplegia
CRG	Center for Reproductive Medicine / Centrum voor Reproductieve Geneeskunde
Cs	Cesium
CYB	Cytochrome B
DHEAS	Dehydroepiandrosterone Sulfate
D-loop	Displacement-loop
DMEM	Dulbecco's Modified Eagle's Medium
FCS	Fetal Calf Serum
FSH	Follicle-Stimulating Hormone
FWO	Fonds voor Wetenschappelijk Onderzoek
GnRH	Gonadotrophin Releasing Hormone
hCG	human Chorion Gonadotrophin
HMG-CoA	3-Hydroxy-3-methyl-glutaryl-coenzyme-A
H-strand	Heavy strand
HV	Hypervariable
ICSI	Intra-Cytoplasmic Sperm Injection
IgM	Immunoglobulin M
iPSC	induced Pluripotent Stem Cell
IVF	In Vitro Fertilization
KSS	Kearns-Sayre Syndrome
LH	Luteinizing Hormone
LHON	Leber's Hereditary Optic Neuroretinopathy
LMCT	Laboratory of Molecular and Cellular Therapy
L-strand	light strand
MAQ	Mapping and Assembly with Quality
MEF	Mouse Embryonic Fibroblasts
MELAS	Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like Episodes

MERF	Myoclonic Epilepsy associated with Ragged-red Fibers
MILS	Maternally Inherited Leigh's Syndrome
MSP	Massive Parallel Sequencing
mtDNA	Mitochondrial DNA
NARP	Neuropathy, Ataxia and Retinitis Pigmentosa
NC	Naturally Conceived
ND	NADH Dehydrogenase
NGAL	neutrophil gelatinase-associated lipocalin
NGS	Next Generation Sequencing
NonCod	Non-Coding
O_H	Origin of Heavy strand
OXPHOS	Oxidative Phosphorylation
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PGD	Preimplantation Genetic Diagnosis
P_i	Inorganic Phosphate
PSC	Pluripotent Stem Cell
qPCR	Quantitative PCR
RBP-4	retinol-binding protein 4
REGE	Reproduction and Genetics
RFLP	Restriction Fragment Length Polymorphism
RGRG	Reproduction, Genetics and Regenerative Medicine
RMAP	Read Mapping
ROS	Reactive Oxygen Species
rRNA	ribosomal RNA
SHRiMP	Short Read Mapping Package
SNV	Single Nucleotide Variant
Taq	Taq Polymerase
TFAM	Mitochondrial Transcription Factor A
tRNA	transfer RNA

10. Bibliography

1. G. D. Adamson, M. Tabangin, M. Macaluso, J. de Mouzon, The number of babies born globally after treatment with the assisted reproductive technologies (ART). *Fertility and Sterility* **100**, S42-S42 (2013).
2. R. Hart, R. J. Norman, The longer-term health outcomes for children born as a result of IVF treatment: Part IGeneral health outcomes. *Human Reproduction Update* **19**, 232-243 (2013).
3. F. Belva *et al.*, Reproductive hormones of ICSI-conceived young adult men: the first results. *Hum Reprod* **32**, 439-446 (2017).
4. M. Ceelen, M. M. van Weissenbruch, J. P. W. Vermeiden, F. E. van Leeuwen, H. A. Delemarre-van de Waal, Growth and development of children born after in vitro fertilization. *Fertility and Sterility* **90**, 1662-1673 (2008).
5. C. Siristatidis *et al.*, Mild Versus Conventional Ovarian Stimulation for Poor Responders Undergoing IVF/ICSI. *In Vivo* **31**, 231-237 (2017).
6. G. Palermo, H. Joris, P. Devroey, A. C. Vansteirteghem, Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyteP. *Lancet* **340**, 17-18 (1992).
7. J. L. Simpson, S. Rechitsky, Preimplantation diagnosis and other modern methods for prenatal diagnosis. *Journal of Steroid Biochemistry and Molecular Biology* **165**, 124-130 (2017).
8. P. C. Steptoe, R. G. Edwards, Birth after reimplantation of a human embryo. *Lancet* **2**, 366-366 (1978).
9. M. Hansen, C. Bower, E. Milne, N. de Klerk, J. J. Kurinczuk, Assisted reproductive technologies and the risk of birth defects - a systematic review. *Human Reproduction* **20**, 328-338 (2005).
10. A. Pinborg, A. K. A. Henningsen, S. S. Malchau, A. Loft, Congenital anomalies after assisted reproductive technology. *Fertility and Sterility* **99**, 327-332 (2013).
11. B. Fauser *et al.*, Health outcomes of children born after IVF/ICSI: a review of current expert opinion and literature. *Reproductive Biomedicine Online* **28**, 162-182 (2014).
12. M. Ceelen *et al.*, Growth during infancy and early childhood in relation to blood pressure and body fat measures at age 8-18 years of IVF children and spontaneously conceived controls born to subfertile parents. *Human Reproduction* **24**, 2788-2795 (2009).
13. R. C. Painter *et al.*, Microalbuminuria in adults after prenatal exposure to the Dutch famine. *Journal of the American Society of Nephrology* **16**, 189-194 (2005).
14. C. L. Abitbol, M. M. Rodriguez, The long-term renal and cardiovascular consequences of prematurity. *Nature Reviews Nephrology* **8**, 265-274 (2012).
15. J. G. Eriksson, C. Osmond, E. Kajantie, T. J. Forsen, D. J. P. Barker, Patterns of growth among children who later develop type 2 diabetes or its risk factors. *Diabetologia* **49**, 2853-2858 (2006).
16. D. J. P. Barker, C. Osmond, J. Golding, D. Kuh, M. E. J. Wadsworth, Growth in utero, blood-pressure in childhood and adult life, and mortality from cardiovascular-disease. *British Medical Journal* **298**, 564-567 (1989).
17. R. W. J. Leunissen, G. F. Kerkhof, T. Stijnen, A. Hokken-Koelega, Timing and Tempo of First-Year Rapid Growth in Relation to Cardiovascular and Metabolic Risk Profile in Early Adulthood. *Jama-Journal of the American Medical Association* **301**, 2234-2242 (2009).

18. F. Belva *et al.*, Are ICSI adolescents at risk for increased adiposity? *Human Reproduction* **27**, 257-264 (2012).

19. M. Ceelen, M. M. van Weissenbruch, J. P. W. Vermeiden, F. E. van Leeuwen, H. de Waal, Cardiometabolic differences in children born after in vitro fertilization: Follow-up study. *Journal of Clinical Endocrinology & Metabolism* **93**, 1682-1688 (2008).

20. S. D. Sakka *et al.*, Gender dimorphic increase in RBP-4 and NGAL in children born after IVF: an epigenetic phenomenon? *European Journal of Clinical Investigation* **43**, 439-448 (2013).

21. U. Scherrer *et al.*, Systemic and Pulmonary Vascular Dysfunction in Children Conceived by Assisted Reproductive Technologies. *Circulation* **125**, 1890-1896 (2012).

22. B. Valenzuela-Alcaraz *et al.*, Assisted Reproductive Technologies Are Associated With Cardiovascular Remodeling In Utero That Persists Postnatally. *Circulation* **128**, 1442-1450 (2013).

23. T. K. Jensen *et al.*, Fertility treatment and reproductive health of male offspring: A study of 1,925 young men from the general population. *American Journal of Epidemiology* **165**, 583-590 (2007).

24. F. Belva *et al.*, Semen quality of young adult ICSI offspring: the first results. *Hum Reprod* **31**, 2811-2820 (2016).

25. F. Belva *et al.*, Serum reproductive hormone levels and ultrasound findings in female offspring after intracytoplasmic sperm injection: first results. *Fertil Steril*, (2017).

26. M. Ceelen, M. M. van Weissenbruch, J. P. W. Vermeiden, F. E. van Leeuwen, H. A. Delemarre-van de Waal, Pubertal development in children and adolescents born after IVF and spontaneous conception. *Human Reproduction* **23**, 2791-2798 (2008).

27. H. Sasaki, Y. Matsui, Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Reviews Genetics* **9**, 129-140 (2008).

28. M. R. DeBaun, E. L. Niemitz, A. P. Feinberg, Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *American Journal of Human Genetics* **72**, 156-160 (2003).

29. A. C. Moll *et al.*, Incidence of retinoblastoma in children born after in-vitro fertilisation. *Lancet* **361**, 309-310 (2003).

30. G. Lazaraviciute, M. Kauser, S. Bhattacharya, P. Haggarty, A systematic review and meta-analysis of DNA methylation levels and imprinting disorders in children conceived by IVF/ICSI compared with children conceived spontaneously. *Human Reproduction Update* **20**, 840-852 (2014).

31. V. F. Oliver *et al.*, Defects in imprinting and genome-wide DNA methylation are not common in the in vitro fertilization population. *Fertility and Sterility* **97**, 147-U219 (2012).

32. N. Whitelaw *et al.*, Epigenetic status in the offspring of spontaneous and assisted conception. *Human Reproduction* **29**, 1452-1458 (2014).

33. M. E. Doornbos, S. M. Maas, J. McDonnell, J. P. W. Vermeiden, R. C. M. Hennekam, Infertility, assisted reproduction technologies and imprinting disturbances: a Dutch study. *Human Reproduction* **22**, 2476-2480 (2007).

34. C. D. Moyes, B. J. Battersby, S. C. Leary, Regulation of muscle mitochondrial design. *Journal of Experimental Biology* **201**, 299-307 (1998).

35. S. Anderson *et al.*, Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465 (1981).

36. E. A. Shoubridge, T. Wai, Mitochondrial DNA and the mammalian oocyte. *Mitochondrion in the Germline and Early Development* **77**, 87-111 (2007).

37. X. J. Chen, R. A. Butow, The organization and inheritance of the mitochondrial genome. *Nature Reviews Genetics* **6**, 815-825 (2005).

38. J. W. Taanman, The mitochondrial genome: structure, transcription, translation and replication. *Biochimica Et Biophysica Acta-Bioenergetics* **1410**, 103-123 (1999).

39. H. Kasamatsu, J. Vinograd, Replication of circular DNA in eukaryotic cells. *Annual Review of Biochemistry* **43**, 695-719 (1974).

40. D. C. Wallace, Mitochondrial DNA Mutations in Disease and Aging. *Environmental and Molecular Mutagenesis* **51**, 440-450 (2010).

41. J. C. St John, J. Facucho-Oliveira, Y. Jiang, R. Kelly, R. Salah, Mitochondrial DNA transmission, replication and inheritance: a journey from the gamete through the embryo and into offspring and embryonic stem cells. *Human Reproduction Update* **16**, 488-509 (2010).

42. A. B. C. Otten, H. J. M. Smeets, Evolutionary defined role of the mitochondrial DNA in fertility, disease and ageing. *Human Reproduction Update* **21**, 671-689 (2015).

43. M. Falkenberg *et al.*, Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nature Genetics* **31**, 289-294 (2002).

44. N. D. Bonawitz, D. A. Clayton, G. S. Shadel, Initiation and beyond: Multiple functions of the human mitochondril transcription machinery. *Molecular Cell* **24**, 813-825 (2006).

45. J. St John, The control of mtDNA replication during differentiation and development. *Biochimica Et Biophysica Acta-General Subjects* **1840**, 1345-1354 (2014).

46. C. Yanicostas, N. Soussi-Yanicostas, R. El-Khoury, P. Benit, P. Rustin, Developmental aspects of respiratory chain from fetus to infancy. *Seminars in Fetal & Neonatal Medicine* **16**, 175-180 (2011).

47. S. DiMauro, C. Garone, Metabolic disorders of fetal life: Glycogenoses and mitochondrial defects of the mitochondrial respiratory chain. *Seminars in Fetal & Neonatal Medicine* **16**, 181-189 (2011).

48. J. M. Ross *et al.*, Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. *Nature* **501**, 412-+ (2013).

49. J. C. von Kleist-Retzow *et al.*, Antenatal manifestations of mitochondrial respiratory chain deficiency. *Journal of Pediatrics* **143**, 208-212 (2003).

50. K. Gibson *et al.*, Mitochondrial Oxidative Phosphorylation Disorders Presenting in Neonates: Clinical Manifestations and Enzymatic and Molecular Diagnoses. *Pediatrics* **122**, 1003-1008 (2008).

51. M. V. Tavares *et al.*, Antenatal manifestations of mitochondrial disorders. *Journal of Inherited Metabolic Disease* **36**, 805-811 (2013).

52. K. G. Bensch, W. deGraaf, P. A. Hansen, H. P. Zassenhaus, J. A. Corbett, A transgenic model to study the pathogenesis of somatic mtDNA mutation accumulation in beta-cells. *Diabetes Obesity & Metabolism* **9**, 74-80 (2007).

53. E. Fosslien, Review: Mitochondrial medicine - Molecular pathology of defective oxidative phosphorylation. *Annals of Clinical and Laboratory Science* **31**, 25-67 (2001).

54. V. K. Mootha *et al.*, Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629-640 (2003).

55. M. E. Patti *et al.*, Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8466-8471 (2003).

56. J. A. Maassen *et al.*, Mitochondrial diabetes - Molecular mechanisms and clinical presentation. *Diabetes* **53**, S103-S109 (2004).

57. A. F. Whereat, Oxygen consumption of normal and atherosclerotic initima. *Circulation Research* **9**, 571-& (1961).

58. J. Van Blerkom, Mitochondrial function in the human oocyte and embryo and their role in developmental competence. *Mitochondrion* **11**, 797-813 (2011).

59. Y. Bentov, T. Yavorska, N. Esfandiari, A. Jurisicova, R. F. Casper, The contribution of mitochondrial function to reproductive aging. *Journal of Assisted Reproduction and Genetics* **28**, 773-783 (2011).

60. A. K. Bartmann, G. S. Romao, E. D. Ramos, R. A. Ferriani, Why do older women have poor implantation rates? A possible role of the mitochondria. *Journal of Assisted Reproduction and Genetics* **21**, 79-83 (2004).

61. U. Eichenlaub-Ritter, M. Wieczorek, S. Luke, T. Seidel, Age related changes in mitochondrial function and new approaches to study redox regulation in mammalian oocytes in response to age or maturation conditions. *Mitochondrion* **11**, 783-796 (2011).

62. B. A. I. Payne *et al.*, Universal heteroplasmy of human mitochondrial DNA. *Human Molecular Genetics* **22**, 384-390 (2013).

63. K. X. Ye, J. Lu, F. Ma, A. Keinan, Z. L. Gu, Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10654-10659 (2014).

64. Y. Guo *et al.*, Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans. *Journal of Genetics and Genomics* **40**, 607-615 (2013).

65. P. Reynier *et al.*, Mitochondrial DNA content affects the fertilizability of human oocytes. *Molecular Human Reproduction* **7**, 425-429 (2001).

66. T. Wai *et al.*, The Role of Mitochondrial DNA Copy Number in Mammalian Fertility. *Biology of Reproduction* **83**, 52-62 (2010).

67. J. Steffann, S. Monnot, J. P. Bonnefont, mtDNA mutations variously impact mtDNA maintenance throughout the human embryofetal development. *Clinical Genetics* **88**, 416-424 (2015).

68. J. A. Barritt, C. A. Brenner, J. Cohen, D. W. Matt, Mitochondrial DNA rearrangements in human oocytes and embryos. *Molecular Human Reproduction* **5**, 927-933 (1999).

69. H. T. Chao *et al.*, Repeated ovarian stimulations induce oxidative damage and mitochondrial DNA mutations in mouse ovaries. *Role of the Mitochondria in Human Aging and Disease: from Genes to Cell Signaling* **1042**, 148-156 (2005).

70. T. C. Gibson, H. M. Kubisch, C. A. Brenner, Mitochondrial DNA deletions in rhesus macaque oocytes and embryos. *Molecular Human Reproduction* **11**, 785-789 (2005).

71. H. S. Ge *et al.*, Impaired mitochondrial function in murine oocytes is associated with controlled ovarian hyperstimulation and in vitro maturation. *Reproduction Fertility and Development* **24**, 945-952 (2012).

72. E. R. Hammond *et al.*, Oocyte mitochondrial deletions and heteroplasmy in a bovine model of ageing and ovarian stimulation. *Molecular Human Reproduction* **22**, 261-271 (2016).

73. R. B. Blok, D. A. Gook, D. R. Thorburn, H. H. M. Dahl, Skewed segregation of the mtDNA nt 8993 (T->G) mutation in human oocytes. *American Journal of Human Genetics* **60**, 1495-1501 (1997).

74. D. T. Brown, D. C. Samuels, E. M. Michael, D. M. Turnbull, P. F. Chinnery, Random genetic drift determines the level of mutant mtDNA in human primary oocytes. *American Journal of Human Genetics* **68**, 533-536 (2001).

75. R. P. S. Jansen, K. de Boer, The bottleneck: mitochondrial imperatives in oogenesis and ovarian follicular fate. *Molecular and Cellular Endocrinology* **145**, 81-88 (1998).

76. H. J. Leese, A. M. Barton, Pyruvate and glucose-uptake by mouse ova and preimplantation embryos. *Journal of Reproduction and Fertility* **72**, 9-13 (1984).

77. J. Van Blerkom, P. Davis, V. Mathwig, S. Alexander, Domains of high-polarized and low-polarized mitochondria may occur in mouse and human oocytes and early embryos. *Human Reproduction* **17**, 393-406 (2002).

78. L. Piko, K. D. Taylor, Amounts of mitochondrial-DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. *Developmental Biology* **123**, 364-374 (1987).

79. L. Piko, L. Matsumoto, Number of mitochondria and some properties of mitochondrial-DNA in mouse egg. *Developmental Biology* **49**, 1-10 (1976).

80. E. C. Spikings, J. Alderson, J. C. S. John, Regulated mitochondrial DNA replication during oocyte maturation is essential for successful porcine embryonic development. *Biology of Reproduction* **76**, 327-335 (2007).

81. L. Q. Cao *et al.*, The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nature Genetics* **39**, 386-390 (2007).

82. J. Van Blerkom, P. Davis, S. Alexander, Differential mitochondrial distribution in human pronuclear embryos leads to disproportionate inheritance between blastomeres: relationship to microtubular organization, ATP content and competence. *Human Reproduction* **15**, 2621-2633 (2000).

83. J. A. Thomson *et al.*, Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).

84. C. D. L. Folmes, P. P. Dzeja, T. J. Nelson, A. Terzic, Metabolic Plasticity in Stem Cell Homeostasis and Differentiation. *Cell Stem Cell* **11**, 596-606 (2012).

85. X. L. Xu *et al.*, Mitochondrial Regulation in Pluripotent Stem Cells. *Cell Metabolism* **18**, 325-332 (2013).

86. K. J. Ahlqvist *et al.*, Somatic Progenitor Cell Vulnerability to Mitochondrial DNA Mutagenesis Underlies Progeroid Phenotypes in Polg Mutator Mice. *Cell Metabolism* **15**, 100-109 (2012).

87. S. Chung *et al.*, Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. *Nat Clin Pract Cardiovasc Med* **4 Suppl 1**, S60-67 (2007).

88. S. Mandal, A. G. Lindgren, A. S. Srivastava, A. T. Clark, U. Banerjee, Mitochondrial Function Controls Proliferation and Early Differentiation Potential of Embryonic Stem Cells. *Stem Cells* **29**, 486-495 (2011).

89. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).

90. K. Takahashi *et al.*, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).

91. J. M. Facucho-Oliveira, J. C. St John, The relationship between pluripotency and mitochondrial DNA proliferation during early embryo development and embryonic stem cell differentiation. *Stem Cell Rev* **5**, 140-158 (2009).

92. A. Prigione *et al.*, Human Induced Pluripotent Stem Cells Harbor Homoplasmic and Heteroplasmic Mitochondrial DNA Mutations While Maintaining Human Embryonic Stem Cell-like Metabolic Reprogramming. *Stem Cells* **29**, 1338-1348 (2011).

93. S. Varum *et al.*, Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* **6**, e20914 (2011).

94. M. Wahlestedt *et al.*, Somatic Cells with a Heavy Mitochondrial DNA Mutational Load Render Induced Pluripotent Stem Cells with Distinct Differentiation Defects. *Stem Cells* **32**, 1173-1182 (2014).

95. M. Yokota, H. Hatakeyama, S. Okabe, Y. Ono, Y. I. Goto, Mitochondrial respiratory dysfunction caused by a heteroplasmic mitochondrial DNA mutation blocks cellular reprogramming. *Human Molecular Genetics* **24**, 4698-4709 (2015).

96. M. Yokota, H. Hatakeyama, Y. Ono, M. Kanazawa, Y. Goto, Mitochondrial respiratory dysfunction disturbs neuronal and cardiac lineage commitment of human iPSCs. *Cell Death & Disease* **8**, 11 (2017).

97. A. Trifunovic *et al.*, Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417-423 (2004).

98. A. M. Schaefer, R. W. Taylor, D. M. Turnbull, P. F. Chinnery, The epidemiology of mitochondrial disorders - past, present and future. *Biochimica Et Biophysica Acta-Bioenergetics* **1659**, 115-120 (2004).

99. S. W. Ballinger, Mitochondrial dysfunction in cardiovascular disease. *Free Radical Biology and Medicine* **38**, 1278-1295 (2005).

100. A. Rotig, T. Bourgeron, D. Chretien, P. Rustin, A. Munnich, Spectrum of mitochondrial-DNA rearrangements in the Pearson marrow pancreas syndrome. *American Journal of Human Genetics* **57**, 1047-1047 (1995).

101. I. J. Holt, A. E. Harding, R. K. H. Petty, J. A. Morganhughes, A new mitochondrial disease associated with mitochondrial-DNA heteroplasmy. *American Journal of Human Genetics* **46**, 428-433 (1990).

102. A. L. Andreu *et al.*, Missense mutation in the mtDNA cytochrome b gene in a patient with myopathy. *Neurology* **51**, 1444-1447 (1998).

103. A. Rojo *et al.*, NARP-MILS syndrome caused by 8993 T > G mitochondrial DNA mutation: a clinical, genetic and neuropathological study. *Acta Neuropathologica* **111**, 610-616 (2006).

104. F. G. Debray, M. Lambert, A. Lortie, M. Vanasse, G. A. Mitchell, Long-term outcome of Leigh syndrome caused by the NARP-T8993C mtDNA mutation. *American Journal of Medical Genetics Part A* **143A**, 2046-2051 (2007).

105. Y. Goto, I. Nonaka, S. Horai, A mutation in the transfer RNALEU(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348**, 651-653 (1990).

106. C. Enter *et al.*, A specific point mutation in the mitochondrial genome of caucasians with MELAS. *Human Genetics* **88**, 233-236 (1991).

107. L. Z. Diba, S. M. M. Ardebili, J. Gharesouran, M. Houshmand, Age-related decrease in mtDNA content as a consequence of mtDNA 4977 bp deletion. *Mitochondrial DNA Part A* **27**, 3008-3012 (2016).

108. A. A. Kazuno *et al.*, Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. *Plos Genetics* **2**, 1167-1177 (2006).

109. M. A. Frye *et al.*, Mitochondrial DNA sequence data reveals association of haplogroup U with psychosis in bipolar disorder. *Journal of Psychiatric Research* **84**, 221-226 (2017).

110. S. Farha *et al.*, Mitochondrial Haplogroups and Risk of Pulmonary Arterial Hypertension. *Plos One* **11**, 13 (2016).

111. H. W. Wang, Y. Xu, Y. L. Miao, H. Y. Luo, K. H. Wang, Mitochondrial DNA Haplogroup A may confer a genetic susceptibility to AIDS group from Southwest China. *Mitochondrial DNA Part A* **27**, 2221-2224 (2016).

112. G. A. Cortopassi, N. Arnheim, Detection of a specific mitochondrial-DNA deletion in tissues of older humans. *Nucleic Acids Research* **18**, 6927-6933 (1990).

113. M. Debrinski, M. T. Lott, J. M. Shoffner, D. C. Wallace, Accumulation of mitochondrial-DNA damage in chronic heart and brain disease. *American Journal of Human Genetics* **49**, 132-132 (1991).

114. L. C. Greaves, A. K. Reeve, R. W. Taylor, D. M. Turnbull, Mitochondrial DNA and disease. *Journal of Pathology* **226**, 274-286 (2012).

115. D. C. Wallace, D. Chalkia, Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harbor Perspectives in Biology* **5**, 47 (2013).

116. D. C. Samuels *et al.*, Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. *Plos Genetics* **9**, 12 (2013).

117. E. A. Schon, S. DiMauro, M. Hirano, Human mitochondrial DNA: roles of inherited and somatic mutations. *Nature Reviews Genetics* **13**, 878-890 (2012).

118. A. Chomyn, G. Attardi, MtDNA mutations in aging and apoptosis. *Biochemical and Biophysical Research Communications* **304**, 519-529 (2003).

119. M. K. Li, R. Schroder, S. Y. Ni, B. Madea, M. Stoneking, Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2491-2496 (2015).

120. F. Ye, D. C. Samuels, T. Clark, Y. Guo, High-throughput sequencing in mitochondrial DNA research. *Mitochondrion* **17**, 157-163 (2014).

121. V. Vasta, S. B. Ng, E. H. Turner, J. Shendure, S. H. Hahn, Targeted gene analysis of mitochondrial disorders by next generation sequencing. *Molecular Genetics and Metabolism* **98**, 6-6 (2009).

122. M. X. Sosa *et al.*, Next-Generation Sequencing of Human Mitochondrial Reference Genomes Uncovers High Heteroplasmy Frequency. *Plos Computational Biology* **8**, 11 (2012).

123. R. K. Bai, L. J. C. Wong, Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: A single-step approach. *Clinical Chemistry* **50**, 996-1001 (2004).

124. M. H. Liang, L. J. C. Wong, Yield of mtDNA mutation analysis in 2,000 patients. *American Journal of Medical Genetics* **77**, 395-400 (1998).

125. A. Rohlin *et al.*, Parallel Sequencing Used in Detection of Mosaic Mutations: Comparison With Four Diagnostic DNA Screening Techniques. *Human Mutation* **30**, 1012-1020 (2009).

126. W. Zhang, H. Cui, L. J. C. Wong, Comprehensive One-Step Molecular Analyses of Mitochondrial Genome by Massively Parallel Sequencing. *Clinical Chemistry* **58**, 1322-1331 (2012).

127. H. P. J. Buermans, J. T. den Dunnen, Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta-Molecular Basis of Disease* **1842**, 1932-1941 (2014).

128. R. D. Mitra, G. M. Church, In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research* **27**, 6 (1999).

129. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

130. L. Heng, C. C. Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA, Ed. (Oxford University Press 2013, 2013).

131. M. K. Li *et al.*, Detecting Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial DNA Genomes. *American Journal of Human Genetics* **87**, 237-249 (2010).

132. Y. Guo *et al.*, The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutation Research-Genetic Toxicology and Environmental Mutagenesis* **744**, 154-160 (2012).

133. I. Mateizel *et al.*, Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. *Human Reproduction* **21**, 503-511 (2006).

134. I. Mateizel, C. Spits, M. De Rycke, I. Liebaers, K. Sermon, Derivation, culture, and characterization of VUB hESC lines. *In Vitro Cellular & Developmental Biology-Animal* **46**, 300-308 (2010).

135. K. Cibulskis *et al.*, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* **31**, 213-219 (2013).

136. J. Knez *et al.*, Correlates of Peripheral Blood Mitochondrial DNA Content in a General Population. *American Journal of Epidemiology* **183**, 138-146 (2016).

137. S. Lopez *et al.*, Sex-Specific Regulation of Mitochondrial DNA Levels: Genome-Wide Linkage Analysis to Identify Quantitative Trait Loci. *Plos One* **7**, 11 (2012).

138. U. Roostalu *et al.*, Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The near eastern and Caucasian perspective. *Molecular Biology and Evolution* **24**, 436-448 (2007).

139. M. Pala *et al.*, Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. *American Journal of Human Genetics* **90**, 915-924 (2012).

140. A. Marcuello *et al.*, Human mitochondrial variants influence on oxygen consumption. *Mitochondrion* **9**, 27-30 (2009).

141. A. Gomez-Duran *et al.*, Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy. *Biochimica Et Biophysica Acta-Molecular Basis of Disease* **1822**, 1216-1222 (2012).

142. E. J. Bowles *et al.*, Contrasting effects of in vitro fertilization and nuclear transfer on the expression of mtDNA replication factors. *Genetics* **176**, 1511-1526 (2007).

143. E. J. Kang *et al.*, Age-Related Accumulation of Somatic Mitochondrial DNA Mutations in Adult-Derived Human iPSCs. *Cell Stem Cell* **18**, 625-636 (2016).