UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY
CHAIR OF BIOTECHNOLOGY

Katre Teearu

Genetic abnormalities in premature ovarian failure patients

Master's Thesis
40 EAP
Biomedicine

Supervisors: Prof. Ants Kurg, Ph.D
Dr. Masoud Zamani Esteki, PhD
Post-doctoral scientist at the Department of
Human Genetics, KU Leuven, Belgium

Tartu 2016

ABSTRACT

Somatic mosaicism, defined as the presence of different cell populations with distinct genotypes within one individual, caused by post-zygotic errors, has long been considered as a source for human genetic variation within and between individuals. It is also plausible that the presence of large structural mosaic events could have important implications for human diseases. In this study, we provide a genome-wide survey of genetic variation in premature ovarian failure (POF) patients by analyzing SNP array data with a novel algorithmic method that deciphers mosaic structural alterations. We found mosaic aberrations in 8.2% of samples, including 23 mosaic copy number variation (CNV) regions, one mosaic X monosomy and 24 (larger than 1 Mb) mosaic uniparental disomy (UPD) events. In addition, we were able to investigate 23 novel CNVs among patients.

CERCS B220 Genetics, cytogenetics
Key words: CNV, mosaicism, SNP microarray, premature ovarian failure, POF

ABSTRAKT

Somaatiline mosaiiksus on nähtus, kus viljastamisjärgselt esineb üksikindiviidi erinevates rakupopulatsioonides erineva genotüübiga rakke. Mosaiiksust on pikka aega peetud inimestevahelise ning ka -sisese geneetilise varieeruvuse põhjuseks. Suured mosaiiksed struktuursed ümberkorraldused genoomis võivad avaldada mõju indiviidi fenotüübile ning olla seeläbi erinevate haiguste põhjuseks. Antud töös teostati kogu genoomi hõlmav uuring, avastamaks mosaiikseid ümberkorraldusi enneaegse ovariaalpuudulikkusega patsientidel. Kasutades genotüpiseerimiskiipide andmeid ja rakendades uut algoritmilist meetodit, tuvastati 8,2%-l valimist mosaiikseid aberratsioone, nende hulgas 23 koopiaarvu varianti, üks X kromosoomi monosoomia ja 24 (>1 Mb) uniparentaalse disoomia juhtu. Lisaks avastati patsientide hulgas 23 uudset koopiaarvu variatsioonide regiooni.

CERCS B220 Geneetika, tsütogeneetika
Märksõnad: CNV, mosaiiksus, SNP genotüpiseerimiskiip, enneaegne ovariaalpuudulikkus, POF

TABLE OF CONTENTS

ABBREVIATIONS

aCGH	array comparative genomic hybridization
aUPD	acquired uniparental disomy
BAF	B-allele frequency
CN-LOH	copy-neutral loss of heterozygosity
CNV	copy number variation
CNVR	copy number variation region
CNA	copy number alteration
DGV	Database of Genomic Variants
DSB	double strand break
EGCUT	Estonian Genome Center of the University of Tartu
FISH	fluorescence *in situ* hybridization
FoSTeS	fork stalling and template switching
GWAS	genome-wide association study
LOH	loss of heterozygosity
LCR	low copy repeat
LRR	Log R ratio
MMBIR	microhomology-mediated break-induced repair
MMEJ	microhomology-mediated end joining
NAHR	non-allelic homologous recombination
NHEJ	non-homologous end joining
NGS	next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
POF	premature ovarian failure
SCNA	somatic chromosomal and copy number alterations
SHOX	short stature homeobox
SNP	single nucleotide polymorphism
STR	short tandem repeats
SV	structural variation
TS	Turner syndrome
UPD	uniparental disomy
VNTR	variable number tandem repeats
WGS	whole-genome sequencing

INTRODUCTION

All human individuals are about 99.9 percent identical in their DNA sequence. Understanding how that 0.1 percent of human genetic variation influences health and disease is one of medical science's highest priorities. However, an adult human organism is composed of ca 10^{14} cells and it is not surprising that errors can occur at any stages of DNA replication and cell division throughout life, producing population of cells, each with its own "personal" genome.

Part of human variability is explained by mosaicism, which describes an individual who has developed from a single fertilized egg and has two or more populations of cells with distinct genotypes (Strachan and Read, 2011). Mosaic cases are often divided into somatic (affecting only non-reproductive cells) and gonadal (affecting germ cells, therefore with the potential of being passed on to any offspring) (Biesecker and Spinner 2013). Although the majority of changes on somatic DNA level have no obvious consequences, a post-zygotic mutation depending on the specific DNA region or the stage of development or type of cells involved, can occasionally have an influence on the phenotype of an individual.

Three main types of structural genetic changes like deletions, gains and copy number neutral loss of heterozygosity are known to occur post-zygotically (Forsberg et al. 2013). Genome-wide single nucleotide polymorphism (SNP) microarray data collaborated with computational algorithmic methods provides an opportunity to detect chromosome variations and aberrations as well as somatic mosaic changes within and between individuals.

Premature ovarian failure (POF) is defined as the cessation of ovarian function before the age of 40 years and is therefore one of the causes for female infertility and poor reproductive outcome of infertility treatments. As a multifactorial disease its cause remains largely unknown, but various data indicate that POF has a strong genetic component.

As a practical outcome of this thesis is to investigate copy number and copy-neutral changes and mosaic chromosomal abnormalities across the genome in premature ovarian failure patients. Further aim is to find novel CNV regions and candidate genes related to POF phenotype. In addition, we attempt to describe possible mechanisms in which large-scale mosaic events might occurred.

1. LITERATURE OVERVIEW

Cell is the structural and functional unit of life and the building block of all the living organisms. Each human cell contains DNA or the genetic material packaged into 23 pairs of chromosomes, of which 22 pairs are autosomes and the other pair is called sex chromosomes, determining the gender of humans. DNA carries almost all the genetic information required to build and maintain the human body. The gene is the unit of inheritance, determining (a) genetic trait(s) and can be passed on during reproduction. The cells adherence to the base-pairing rules establishes that the new strand is an exact copy of the old one. Although replication of the genome is a very precise process and its proofreading capacity minimizes the frequency of errors (i.e. mutations), it is not always perfectly accurate and can result in errors that may extremely affect the living organism (somatic mutations) or its offspring (germ-line mutations).

1.1 Genetic variation

Genetic variation stems from DNA mutations and can describe differences between the two individuals. In addition to the genetic variation that we inherit from our parents and that is present in all our cells, DNA changes may occur in the DNA structure of our cells throughout life (i.e. *de novo* mutations). When the frequency of a genetic variant at a specific locus is equal or over 1% of subjects in the pool of alleles within a population, it is called polymorphism. The variants with frequency of less than 1%, are rather classified as rare variants (Frazer et al. 2009). Genetic variations in the human genome can take many forms (Table 1, see below), including single nucleotide changes or substitutions; tandem repeats; insertions and deletions (known as indels); copy number variations (CNVs); other chromosomal rearrangements such as inversions and translocations; and copy-neutral loss of heterozygosity (LOH) or homozygosity. These genetic variations are having a wide spectrum of sizes, from a single nucleotide (bp) to millions of basepairs, known as megabase (Mb) size aberrations. These changes in the human genome can be deemed as the footprints of mistakes that occur in DNA replication during cell division, even though external agents, such as viruses and chemical mutagens, can also induce changes in the DNA sequence. Regardless of the molecular mechanisms or processes that generate the genetic variations, they can be generally classified as either somatic or germline variations depending on whether they arose during mitosis or meiosis of cell division, respectively (see below) (Ku et al. 2010).

The most prevalent form of genomic variation is single nucleotide polymorphisms (SNPs) with a frequency of equal or more than 1% in the population. Over 84 million SNPs are present in each human genome and most of them have no influence on the phenotype (Auton et al. 2015). However, they are useful as genetic markers for genome research to perform linkage and association studies due to their high density and the relatively easy methods of genotyping. In general, SNPs can occur anywhere on a genome, but much more less frequently in coding regions than in noncoding regions. The phenotypic consequence of a SNP is related to the location where it occurs, as well as the nature of the variant. However, many complex disorders have been associated to common SNPs (Consortium 2004).

Table 1. Types of genetic variants and their approximate sizes. Size ranges are indicative only of the scale of each type of rearrangement and are not definitive. Adapted and modified from (Sharp et al. 2006).

Variation	Rearrangement type	Size range
Single-pair changes	SNPs, point mutations	1 bp
Indels	binary insertion/deletion events of short sequences (majority <10 bp in size)	1bp -50 bp
Tandem repeats	STRs, VNTRs, microsatellites, minisatellites and other simple repeats	1bp -500 bp
Submicroscopic structural variation	deletions, duplications, tandem repeats, inversions	500 bp – 50 kb
Large-scale structural variation	deletions, duplications, large tandem repeats, inversions	50 kb – 5Mb
Chromosomal variation	large cytogenetically visible deletions, duplications, translocations, inversions and aneuploidy	~5Mb – entire chromosomes

Tandem repeats can be broadly divided into two classes: short tandem repeats (STRs) usually refer to tandem repeats in which the sequence length is eight nucleotides or less, and longer tandem repeats labeled as variable number tandem repeats (VNTRs). These are also known as microsatellites (1-6 bp) and minisatellites (10-100 bp), respectively (Denoeud et al. 2003; Ku et al. 2010; Sawaya et al. 2013).

However, apart from single nucleotide changes, such as SNPs, all the genetic variations can be broadly grouped under the umbrella of structural variations (SV). The most

wide-spread copy number variations (CNVs) can affect chromosomal segments larger than 1 kilobase (kb) to many megabases (Mb), including entire chromosomes (aneuploidy), which are variable in copy number when compared to the reference genome. Copy number variations include deletions, duplications and insertions (Feuk et al. 2006). CNVs are widespread among humans and account for a nearly ten-fold greater proportion of variation in the genome compared to SNPs (Pang et al. 2010). CNVs larger than 40 kb in size account for about 18% of the genetic variation in gene expression (Stranger et al. 2007). Copy number changes affect approximately seven times as many base pairs as single nucleotide variants and are major contributors to interindividual differences (Sudmant et al. 2015).

The copy number variation map of the human genome, based on CNVs of healthy individuals developed from data in the Database of Genomic Variants (DGV) revealed that up to 9.5% of the human genome seems to be involved with DNA copy number variation region (CNVR) losses or gains (7.5% and 3.9%, respectively). CNVRs were found to be irregularly distributed among chromosomes. The average size of these regions are 300-3,000 bp, with losses slightly smaller than gains. For gains, chromosome 22, the Y chromosome and chromosomes 16, 9 and 15 demonstrated the highest proportion of variability, in contrast to chromosomes 3 and 18, which showed the lowest proportion. For losses, chromosomes 19, 22 and the Y chromosome displays the highest proportion of variable sequence, contrary to chromosomes 5 and 8 with the lowest proportion (Zarrei et al. 2015). Common CNVs most likely represent genomic regions that can vary without negative consequences. Over the past few years, CNVs have attracted much attention due to the fact that they can have dramatic phenotypic consequences, including the capacity to alter gene dosage, disrupt coding segments, and regulate functional genes. More than 50% of the effects of known CNVs are caused by gene disruption or disruption of the regulatory units associated with genes, rather than by gene dosage change (Lee and Scherer 2010).

Genetic variations are widely used as genetic markers in disease gene mapping, for example family linkage and genetic association studies to describe the suscepility loci or genes for monogenic and complex diseases (Ku et al. 2010).

1.1.1 Mechanisms of copy number variation formation

Different mechanisms for formation of CNVs have been described, with different types of mechanisms responsible for recurrent and rare CNVs. These include (a) recombination errors, like non-allelic homologous recombination (NAHR), (b) mistakes generated in DNA break repair, as in non-homologous end joining (NHEJ) and microhomology-mediated end joining

(MMEJ) or (c) errors in replication, such as fork stalling and template switching (FoSTes) or microhomology-mediated break-induced replication (MMBIR).

One of the most common mechanism is thought to be NAHR, which is caused during mitosis or meiosis by misalignment and cross-over of non-allelic homologous DNA segments that can be repetitive sequences such as segmental duplications (SD) or low copy repeats (LCRs). Depending on the location and orientation of the homology regions, NAHR can cause intra- or inter-chromosomal aberrations (Figure 1). Intra-chromosomal aberrations happen within the same chromosome, including deletions and duplications, when they are in direct orientation, and inversions, when they are in opposite orientation on the same chromosome. However, inter-chromosomal aberrations happen between different chromosome, such as translocations induced by LCRs (Inoue and Lupski 2002; Robberecht et al. 2013; Sharp et al. 2006). This mechanism of CNV formation is more common in recurrent events (Conrad et al. 2010). Somatic NAHR between nonsister chromatids can result in the formation of an isochromosome (Barbouti et al. 2004).

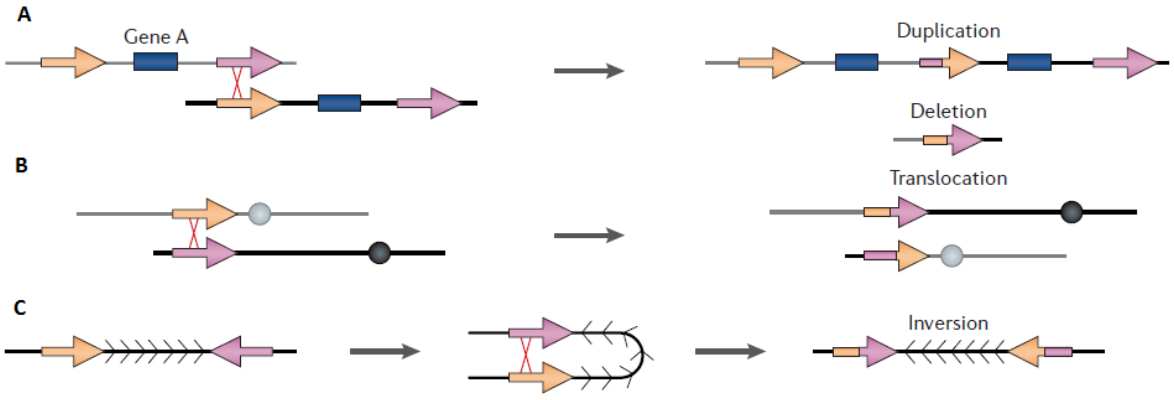


Figure 1. Schematic representation of non-allelic homologous recombination (NAHR). Repeats oriented in the same direction (A) mediate formation of duplications and deletions of the interval between the repeats, while repeats oriented in opposite directions (C) on the same chromosome mediate formation of inversions. Repeats on different chromosomes (B) mediate formation of translocations. Adapted from (Bailey and Eichler 2006).

The majority of nonrecurrent events are caused by a diversity of mechanisms. Non-homologous end joining (NHEJ) is one of the predominantly explained mechanism (Figure 2, see below), which is a very rapid process that fuses the ends of a double-strand DNA break with little or no sequence homology (<4 bp), generating blunt joins or short insertions or deletions at the breakpoint junction. Microhomology-mediated end joining (MMEJ) is a form of alternative NHEJ. It usually arises when NHEJ and homologous replication repair mechanisms are repressed and it is more error-prone than NHEJ. It can lead to deletions and

translocations, therefore it is considered as a major source of genomic instability (McVey and Lee 2008).

Second mechanism is called FoSTeS, later included in the generalized mechanism MMBIR, has been described as a mechanism for the formation of complex rearrangements (Figure 2), as well as inversions, tandem duplications or translocations. The process involves stalling of the replication at the fork and shifting of the polymerase by microhomology to any nearby single stranded DNA (Hastings et al. 2009a). The involved forks may be adjacent on the same chromosome or in a close proximity in three-dimensional space. FoSTeS may cause the joining of different sequences from discrete genomic position that are far apart from each other. Depending on the direction of the fork progression, the incorrectly joined fragment can be in direct or inverted orientation. Furthermore, depending on whether the new fork is located upstream or downstream of the original fork, the template switching results in either a deletion or a duplication (Zhang et al. 2009).

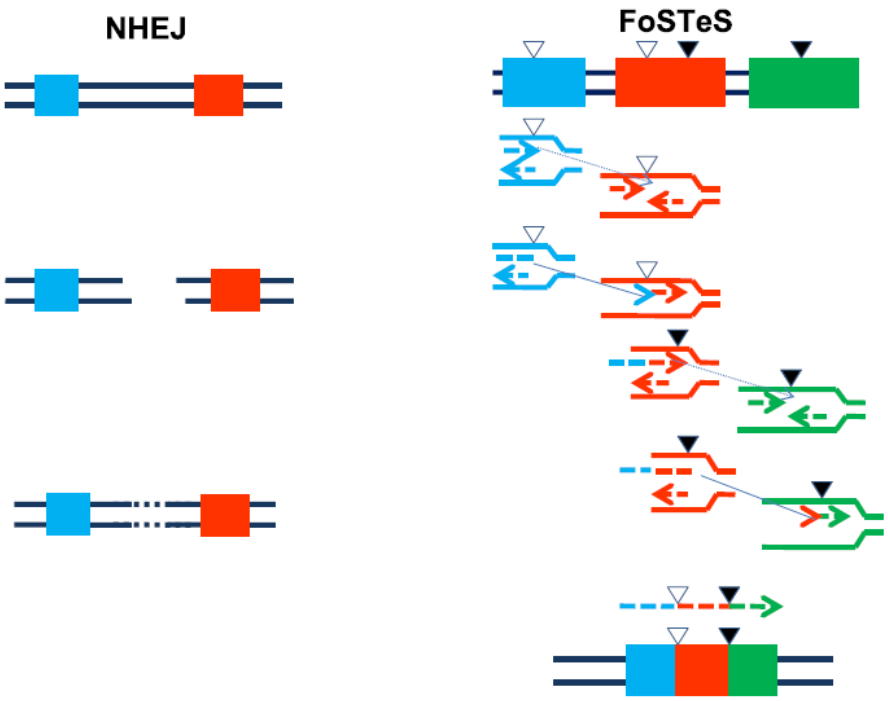


Figure 2. Genomic rearrangement mechanisms by NHEJ and FoSTeS resulting in duplication or deletion. Double-strand breaks (DSBs) are created between the two sequences represented as a blue and a red rectangle with no homology between each other. The NHEJ mechanism modifies and rejoins the two ends, resulting in the deletion of the segment between the two DSBs. FoSTeS mechanism is causing a complex deletion involving two fragments. No extensive homology is required between the substrate sequences depicted by a blue, a red and a green rectangle. However, the small open triangle heading downwards depicts a site bearing microhomology (2-5 bp) between the blue and the red sequences, and the small filled triangle heading downwards depicts another site bearing microhomology between the red and the green sequences. Adapted from (Gu et al. 2008).

Different CNV formation mechanisms are effective in different environments. For example, NAHR and NHEJ are effective during meiosis and mitosis, but MMBIR during DNA replication. Also, there must be consideration, when studying the CNV breakpoint, that NAHR and NHEJ usually correct double-stranded breaks, whereas MMBIR/FoSTeS correct single-stranded breaks. These mechanisms may cause deletions and duplications, but MMBIR is also able causing inversions, translocations, triplications and loss of heterozygosity (LOH) (Conrad et al. 2010). Nonrecurrent CNVs have breakpoints that are not clustered in particular genomic regions and are usually formed by NHEJ or microhomology-mediated DNA-replication errors (FoSTeS or MMBIR) during cell divisions, whereas recurrent CNVs refer to genomic-disorder-associated CNVs with clustered breakpoints, usually located in flanking LCR regions and mediated by NAHR (Hastings et al. 2009b). The ones that occur at mitosis and depending on the developmental stage and cell type can either be restricted to one tissue type or present in several or all tissues. For example, somatic mosaicism is the result of structural changes during post-zygotic mitosis (Piotrowski et al. 2008).

1.2 Meiosis and mitosis

The growth and reproduction of all organisms depend on two types of cell divisions, namely mitosis and meiosis. Mitosis renders to the majority of the cells via replication and segregation of the genomic DNA. The series of events occur in mitosis, called cell cycle, yielding two daughter cells. Multicellular organisms are therefore created through many cell divisions during the course of the development right after fertilization to a complete organism with multiple different tissues and organs.

In general, cells can be present in two different stages: proliferating (dividing) or quiescent (non-dividing). The typical cell cycle is divided into four phases: gap 1 (G1), synthesis (S), gap 2 (G2), and mitosis (M) phase, of which the first three phases are collectively called interphase. The G1 phase, which begins immediately after the cell division, is the primary growth phase where proteins are synthesized and new organelles are formed. Once committed to the cell cycle, the next major step is the Synthesis (S) phase, where nuclear DNA is replicated (Figure 3A, see below). In the succeeding G2 phase, the cell usually continues growing, the machinery for cell division is assembled, and the accuracy of the DNA synthesis is tightly controlled. Finally, in the M phase, sister chromatids are separated and distributed to the two daughter cells via a process called cytokinesis. The length of the cell cycle depends on several factors, including the cell and organ type, developmental stage, and physiological conditions. Some cell types continue to divide throughout the life of

an organism, while some other cell types remain in a so called non-dividing phase, G0, until they receive a signal from the environment to instigate the cycle (Alberts et al. 2014).

Numerous checkpoints have evolved to ensure that mitosis only proceeds when growth conditions are ideal and chromosomes are efficiently replicated and free of damage. This level of quality control takes time and for example, proliferating mammalian somatic cells require 12–30 hours to properly prepare for division. By contrast, mitosis itself is relatively rapid, typically lasting only 20–60 minutes, depending on chromosome number and the efficiency of spindle assembly (Yang et al. 2008). However, several mitotic errors may occur at any stage of the development leading to large spectrum of genetic mosaicism (see below).

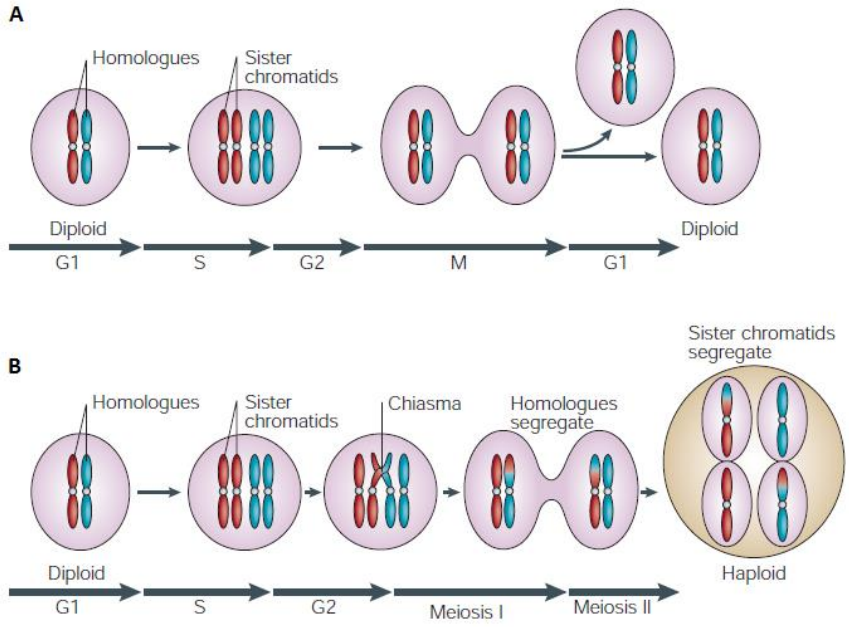


Figure 3. The mitotic and meiotic cell cycles. In mitosis (A), diploid cells replicate chromosomes during S phase and segregate sister chromatids during M phase. In meiosis (B), two chromosome-segregation phases, meiosis I and meiosis II, follow a single round of DNA replication during pre-meiotic S phase. During meiosis I, homologous chromosomes (shown in red and blue) are segregated to opposite poles. Sister chromatids segregate to opposite poles during meiosis II, which results in the formation of non-identical haploid gametes. The lengths of the cell-cycle stages are not drawn to scale. Adapted from (Marston and Amon 2004).

Meiosis is a specialized cell division, characterized by two consecutive rounds of chromosome segregation, termed meiosis I (MI) and meiosis II (MII) without an additional round of DNA replication (Figure 3B). Pre-meiotic S phase generates two sister chromatids per homolog, which are tightly connected along their entire lengths. In meiosis I, the homologous chromosomes first pair with one another and normally undergo recombination, and then segregate to different daughter cells maintaining the sister chromatids together. Due

to the fact that chromosome number is reduced to half this division is also known as reductional division. Meiosis I is followed by meiosis II that resembles mitosis, as the sister chromatids separate and segregate to opposite spindle poles giving rise to four different daughter cells. A crucial player in correct segregation of the homologous chromosomes is homologous recombination which happens in MI and instigates genetic diversity in the progeny. In most organisms, crossing over in meiosis is required to ensure accurate segregation of homologous chromosomes at the first meiotic division. Errors in this process can result in random disjunction and aneuploidy, being the leading cause of miscarriages and congenital birth defects in humans. Most gametic aneuploidy originates during oogenesis, particularly during the first meiotic division, and the frequency of such errors increases with female age (Handel and Schimenti 2010; Hassold and Hunt 2001).

1.3 Mosaicism, chimaerism and uniparental disomy

Part of human variability is explained by chromosomal mosaicism, which have been defined as the presence of both normal karyotypes as well as those with large structural genomic events resulting in alteration of copy number or loss of heterozygosity in distinct and detectable subpopulations of cells (Jacobs et al. 2012). There is variety of different mechanisms or processes by which the mosaicism can arise. The unequal distribution of DNA upon mitosis may lead to aneuploidy, the duplication or deletion of segments (CNVs) or whole chromosomes (aneuploidy), and reciprocal duplication and deletion events that appear as copy-neutral loss of heterozygosity (CN-LOH) or acquired uniparental disomy (aUPD) (Jacobs et al. 2012). Some types of structural mosaic changes are also schematically shown in Figure 6 (see page 20). These mutational processes can occur at any stage of development: in stem cells, differentiating cells, and in terminally differentiated somatic cells (Lupski 2013). Depending on the point at which a mutation occurs, the aberrant cell population may exist in only a specific part of the body, e.g. one organ, one tissue or even a part of that tissue (Biesecker and Spinner 2013). As mosaicism can arise through any type of post-zygotic alteration, it is right to claim that we are all genetically mosaic, as everyone has at least one cell in their body that differs from the others.

The consequences of mosaicism depend on how it modifies genetic architecture and more specifically development and cell-specific pathways (Rodríguez-Santiago et al. 2010). Somatic mosaicism may affect different tissues at variable levels. For example, depending on the timing at which the alteration occurred, as well as the affected gene and whether it is expressed in that certain tissue. In addition, the survival capacity of the mutant cells is also

important (Cohen et al. 2015). There can be wide range of phenotypic impact of mosaicism for the individual. The majority of somatic mosaic changes have been described in relation with a known phenotype, which represent mosaic aberrations with strong effect, but the mosaicism can result in either milder or unusual disease phenotype as well. If mosaic somatic changes have no visible phenotypic effect, it is suggested that the occurrence of mosaicism events is underestimated because of deficiency of common detection methods (Rodríguez-Santiago et al. 2010).

Mosaic aneuploidy is the most common type of mosaicism (Hassold and Jacobs 1984). In some inherited diseases, the mosaic form is phenotypically less severe than the constitutional form, e.g. mosaic trisomies 13, 18 and 21 have less severe phenotypic consequences when compared to full trisomic ones. Other mosaic trisomies reported include trisomy 8, 9, 14, 17 and 22 (Daber et al. 2011; Hassold and Jacobs 1984). Aneuploidies of the sex chromosomes are more common than autosomal aneuploidies and the post-zygotic gain or loss of an X or Y chromosome results in somatic mosaicism, e.g. the mosaic forms of Klinefelter syndrome or mosaic Turner syndrome (TS) (Hersmus et al. 2012; Jacobs et al. 1997). Mosaic forms of other structural variants such as translocations, inversions, ring chromosomes and supernumerary chromosomes are not as commonly reported (Biesecker and Spinner 2013).

Studies on human embryos have also shown that mosaicism for chromosome anomalies is detected in 50-80% of all human embryos generated following in vitro fertilization (IVF) (Bielanska et al. 2002; Zamani Esteki et al. 2015; Vanneste et al. 2009). The loss or gain of entire chromosomes is thought to be caused by errors in chromosomal segregation in anaphase during mitosis, while non-allelic homologous recombination (NAHR) may cause the gain or loss of large genomic regions (Liu et al. 2012). The phenotypic effects of these reorganizations vary based on the genomic region and the size of the event. Mosaicism has been stated in normal aging, also as a cause of miscarriage, birth defects, developmental delay and cancer (Conlin et al. 2010; Hassold 1982; Hsu et al. 1992; Jacobs et al. 2012; Laurie et al. 2012; Lu et al. 2008; Menten et al. 2006). Forsberg *et al.*, suggested that post-zygotic variation may explain a notable part of the cause of non-heritable diseases (Forsberg et al. 2013).

Mosaicism is different from chimaerism, which exists when an individual carries cell lineages from multiple zygotes. Examples of chimaerism include cells remaining in a recipient after an organ transplantation; chimaerism can also occur in twins when there is fusion of two different zygotes resulting in a single embryo; or after fertilization by two sperm of an oocyte and a polar body (Bluth et al. 2007; Khan et al. 2004; E. J. Yunis et al. 2007).

Chimaerism would not be detectable by standard cytogenetic technology. The use of a genome-wide SNP array makes the differentiation between mosaicism and chimaerism possible together with novel genome analysis methods, as the allelic admixtures can be identified within the B-allele frequency (BAF) profiles (Conlin et al. 2010; Destouni et al. 2016; Zamani Esteki et al. 2015).

Uniparental disomy (UPD) is a type of copy-neutral structural variation. It occurs when two homologous chromosomes, or segments of chromosomes, originate from a single parent, instead of one copy each coming from the mother and father (Engel 1980). UPD can either be constitutional (germline lesion) or acquired (aUPD) in somatic cells. Acquired uniparental disomy (aUPD), also known as copy-neutral loss of heterozygosity (CN-LOH) have reported in numerous studies as a common feature in a variety of human cancers (Dunbar et al. 2008; Walsh et al. 2008). There are at least four primary mechanisms by which UPD can occur for segmental or whole chromosomes (Figure 4): (1) trisomy rescue, for example, mitotic loss of one of the three copies of the trisomic chromosome; (2) monosomy rescue, for example, duplication of the single copy of a chromosome pair via nondisjunction; (3) post-fertilization error by either mitotic nondisjunction followed by reduction of the single homologue or *vice versa*; (4) gamete complementation, whereby a gamete missing one chromosome pair unites with a gamete containing two copies of the same chromosome pair by chance (Eggermann et al. 2015; Engel 2006).

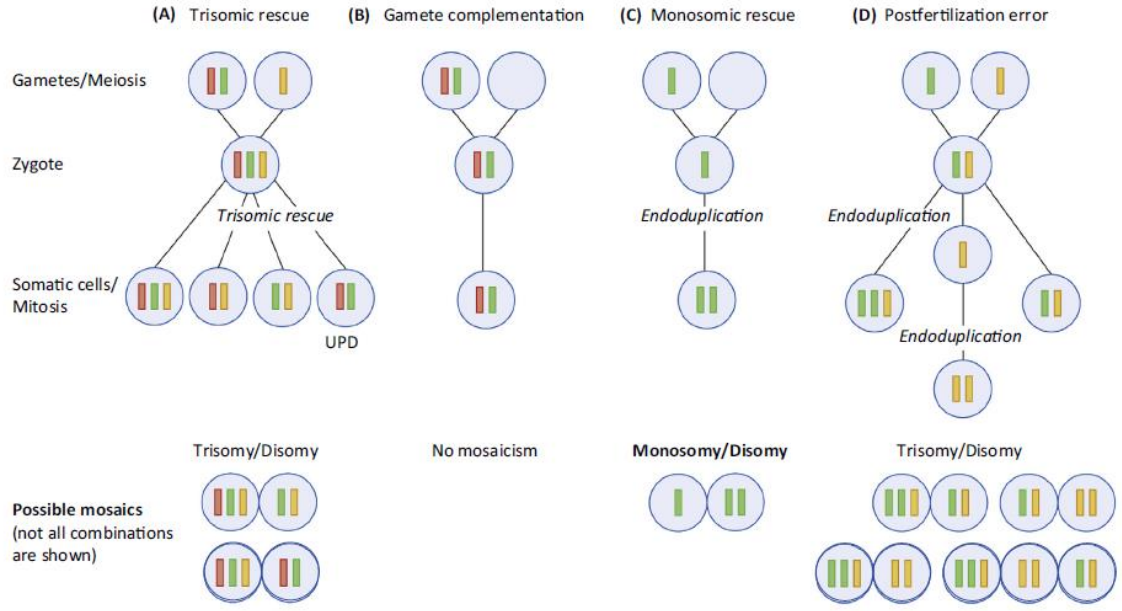


Figure 4. The schematic illustration of the mechanisms leading to UPD formation and possible mosaic constitutions. Mechanisms include trisomy rescue (A), monosomy rescue (C), gamete complementation (C) and post-fertilization mitotic error (D). Adapted from (Eggermann et al. 2015).

There are no apparent phenotypic effects from UPD for most chromosomes, but there are a few chromosomes that involve parent-specific imprinting and have clinically recognizable phenotypic effects when involved in UPD. Currently, maternally derived chromosomes 7, 14 and 15 and paternally derived chromosomes 6, 11, 14 and 15 are the only examples of definite phenotypic effect due to UPD and imprinting. Chromosomes 2, 16 and 20 are also being studied but it is unclear what are their exact phenotypic effects due to imprinting (Shaffer et al. 2001). As an example, UPD can cause disease through the biallelic silencing or biallelic expression of an imprinted gene. Segmental paternal UPD for a portion of chromosome 11p15.5 is a cause of Beckwith–Wiedemann syndrome (BWS) in 10–20% of patients. The 11p UPD is always mosaic, and it has been hypothesized that UPD for this region may be lethal early in development (Cooper et al. 2007; Romanelli et al. 2011).

CN-LOH or aUPD can occur in two different ways: loss of one chromosome followed by duplication of the remaining chromosome (with the homologous chromosomes arising either maternally or paternally) leads to whole chromosome aUPD, whereas somatic recombination leads to segmental aUPD. In both cases, aUPD has the potential to lead to loss of heterozygosity of existing aberrations such as mutation, deletion, methylation, histone-modification, or imprinted genes (Tuna et al. 2015). Because of the lack of copy number change, CN-LOH cannot be detected by conventional cytogenetics, but during the past decade, significant advances in genome-wide SNP arrays have provided strong resolution and greatly simplified assay of CN-LOH (Score and Cross 2012). Žilina and colleagues studied panels of tissue samples (11-12 tissues per individual) from four autopsy subjects using high-resolution SNP arrays to reveal the presence of possible intra-individual tissue specific CN-LOH and CNV patterns. They detected five mosaic CN-LOH regions over 5 Mb in some tissue samples in three out of four individuals. These results give further support to the idea that somatic mosaicism for CN-LOHs is a common phenomenon in phenotypically normal individuals (Žilina et al. 2015).

1.4 Strategies for detection of genomic aberrations

There are several methods (Table 2, see below) that have been used for detection of chromosomal aberrations and each of them has different throughput, coverage and resolution. The development of chromosome banding techniques, in which segments of euchromatin and heterochromatin are differentially stained, enabled the identification of large structural changes on chromosomes. The most common chromosomal banding technique, high resolution G-banding was invented in 1978 (Yunis et al. 1978) and enabled the detection of

structural changes that are at least 5-10 Mb in size and is still used nowadays for clinical diagnostic testing.

Table 2. Experimental methods to detect structural variation. Adapted and modified from (Vandeweyer and Kooy 2013)

Method	Karyotyping	FISH	aCGH	SNP array	NGS (paired-end mapping)
Type	microscope	microscope	microarray	microarray	sequencing
Translocation	>3 Mb	yes	no	no	yes
Inversion	>3 Mb	yes	no	no	yes
Deletion	>3 Mb	>200 kb	>5 kb	>5 kb	yes
Duplication	>3 Mb	>200 kb	>5 kb	>5 kb	yes
UPD	no	no	no	yes	yes
Sequence	no	no	no	yes (SNP)	yes

Traditional cytogenetics, such as karyotyping, is able to detect large balanced and unbalanced numerical and structural chromosome aberrations, but copy-neutral loss of heterozygosity (CN-LOH) is not visible. Karyotyping can also detect mosaic structural abnormalities, but this process is laborintensive, e.g. 14 cells must be examined per individual to exclude 10% mosaicism with 95% confidence (Hook 1977).

Molecular cytogenetics, however, enabled more accurate screening of the genome. The trend has started with probing a few marker at a time via fluorescent *in situ* hybridization (FISH) to hundreds of thousands markers at a time via microarray technologies and now basically every single nucleotide of the genome can be screened with the advent of next-generation sequencing (NGS). Fluorescent *in situ* hybridization offers improved resolution compared to karyotyping and interphase FISH can be performed without cultured cells. Metaphase FISH, which is culture-dependent method, enables concurrent visualization of a structural aberration and the chromosomes. Although FISH allows a better resolution in detecting smaller genomic abnormalities (approximately 80-200 kb), when compared to traditional methods such as G-banding, this approach is limited. Only a few FISH probes can be used in this assay and therefore it lacks whole-genome coverage and requires prior knowledge of certain region to identify the clinical value of the test. FISH is still used today for detection of unbalanced translocations and also for independent validation for other methods (Levsky and Singer 2003).

Comparative genomic hybridization (CGH) is a technique in which fluorophore-labeled DNA from a control and test individual are hybridized to a metaphase reference chromosome (Kallioniemi et al. 1993). The same rationale has been applied in the development of array comparative genomic hybridization (aCGH), which quantifies the frequency of thousands to millions of probes across the genome of the DNA samples, as it allows measuring the intensity of light emitted by fluorescent probes hybridized to a DNA sample when compared to the control(s).

In general, there are two formats of microarray, albeit with different chemistries: (1) two-channel platform aCGH and (2) one-channel platform, single nucleotide polymorphism (SNP) microarray. Specifically, in aCGH DNA from test and reference samples are labeled with different fluorophores and then competitively hybridized to a microarray including hundreds to millions DNA probes that are complementary to targeted genomic regions. Whereas, SNP array technology relies on hybridization of one sample to an array and the results of that hybridization is compared *in silico* to a database of standard reference DNA (Pinto et al. 2011). SNP array technology has the advantage of detecting not only copy number gains and losses but also loss of heterozygosity regions, e.g. uniparental disomy (UPD). Importantly, SNP array has the capacity of detecting mosaic allelic architectures across the genome. These technologies are cell-culture free, the mosaicism is more easily detected because of thousands of cells are assayed simultaneously (Bruno et al. 2011; Oostlander et al. 2004). Despite these facts, robust computational methods are required to detect copy number and copy-neutral anomalies and different levels of mosaic aberrations (see below).

Next generation sequencing (NGS) techniques have revolutionized sequencing possibilities. NGS includes whole-exome sequencing (WES), whole-genome sequencing (WGS), and targeted sequencing approaches. Each NGS technique starts with creating a DNA library by fragmenting genomic DNA and amplifying it. Using synthetic oligonucleotides, the fragments can be attached to a sequencing media (for example, glass) and amplified, generating many clones of fragments. Fragments are denaturated and imaging techniques capture growing strands and record strings of bases, of which each contains bases from a location in the genome. WGS can detect different types of copy number variation, CN-LOH and balanced translocations, providing very complete resolution of heterogeneous sample. The unique paired-end sequencing by lllumina usually allows detection capacity in base pair resolution for the detection of deletions, duplications, inversions and insertions (Campbell et al. 2008). On the other hand, if there is a purpose to detect mosaic mutations, the achievement of high coverage of the genome makes the technology expensive. In any next generation

sequencing approach, the analysis of the large amounts of generated data is currently the most challenging part of the process. This is also a major reason why the NGS technique has been implemented to the clinical setting slowly and with caution (Desai and Jere 2012).

1.4.1 SNP array-based detection methods for mosaic chromosomal abnormalities

It has been proven that SNP-array technology has the capacity of detecting a large spectrum of genetic disorders, including mosaic and/or chimeric genomic aberrations (Conlin et al. 2010; Destouni et al. 2016; Jacobs et al. 2012). SNP arrays provide an intermediate option between traditional cytogenetics and whole-genome sequencing (WGS). The detection of mosaic events in SNP array platform is based on assessment of allelic imbalance and copy number changes. The chromosomal abnormalities detected by Illumina SNP arrays are based on two main data tracks (Figure 5): summed allelic intensity, which is demonstrated by log R ratio (LRR) and allele balance, which is demonstrated by B-allele frequency (BAF).

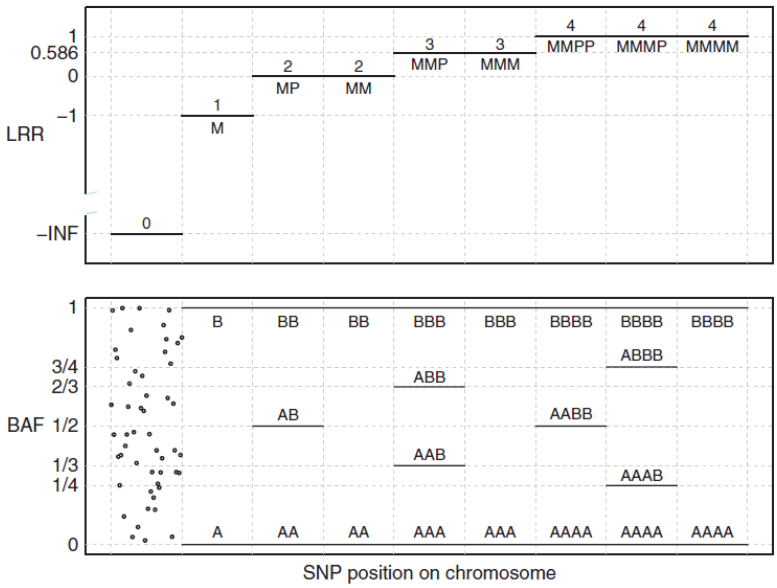


Figure 5. Expected LRR and BAF values for different copy number and copy number neutral changes. LRR plot describes the copy number state and BAF plot illustrates the SNP genotypes. M – maternal chromosome; P – paternal chromosome. The scatter of points demonstrate homozygous deletion (copy number = 0). Adapted from (Laurie et al. 2012).

For each probe on the SNP array, LRR is the normalized log2 ratio of the observed signal intensity to the expected signal intensity (Peiffer et al. 2006). If the DNA sample is normal diploid, LRR value is zero and any deviations from zero indicate copy number change. BAF is the fraction of B allele over both alleles (A allele + B allele). For normal diploid genomic regions, heterozygous AB SNP genotypes have BAF value of 0.5, while homozygous AA or

BB SNP genotypes have BAF value of 0 or 1, respectively and any deviations from these values are indicative of structural aberrations (Figure 6) (Wang et al. 2007).

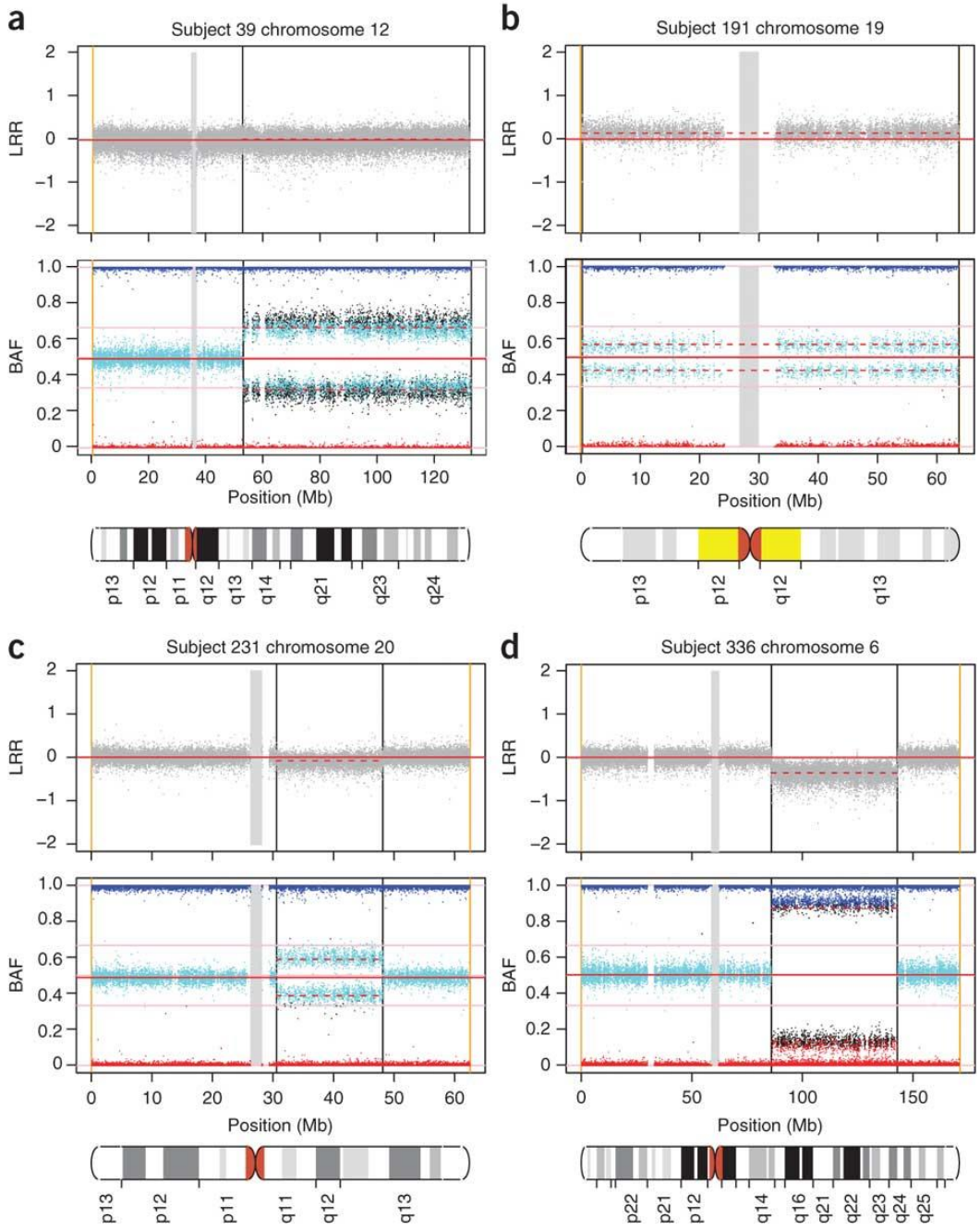


Figure 6. Post-zygotic structural genetic aberrations: mosaic aUPD/CN-LOH (a), mosaic trisomy (b) and a mosaic deletions (c, d), respectively. The results consist of two data tracks: log R ratio (LRR) and B-allele frequency (BAF), as described above. Panel A shows a mosaic aUPD for distal 12q with lack of changes in LRR value and BAF deviations from value 0.5. Panel B displays mosaic trisomy for chromosome 19, where is a narrow split in the intermediate BAF band and LRR demonstrates only a small elevation. Panel C illustrates a mosaic deletion at 20q with a small decrease in LRR and narrow cleavage from intermediate BAF band. Panel D demonstrates also a mosaic deletion at 6q, but with a wider split in BAF band and larger decrease in LRR compared with panel C. Adapted from (Laurie et al. 2012).

1.5 Large scale studies of chromosomal mosaicism

In 2010, Conlin *et al.* examined blood from 2,019 children with pervasive developmental delay or congenital abnormalities, identifying 12 with mosaic aneuploidy (0.6%) and eight with UPD. Of these eight UPD events, four were from trisomy rescue, two were from monosomy rescue, and two were mitotic in origin. Mosaicism was only detected in the two mitotic cases. The origin of the other six UPD events was inferred from the allele fraction patterns. Of the 12 aneuploidies, 9 were monosomies, and all of these monosomies arose from mitotic nondisjunction suggesting that early stage (inherited) monosomy is lethal, while half of the trisomies arose by meiotic nondisjunction. In addition, one of the children with a mosaic abnormality was chimeric (Conlin et al. 2010).

SNP array data were also used by several studies that directly analyze the rate of somatic segmental copy number variation and CN-LOH in the general population. Laurie *et al.* analyzed over 50 000 samples from 15 different case-control GWA studies (GWAS) collected as part of the GENEVA (Gene-Environment Association Studies) consortium. The studied phenotypes included cancers and non-cancer conditions, and the subjects had a wide range of ages, from newborns to individuals more than 80 years old. They found that the prevalence of a detectable mosaic event was low for younger age groups, but increased to 2-3% for subjects more than 80 years old. They also found a ten-fold increased risk of hematological cancer for individuals with detectable blood mosaicism compared to subjects without detectable blood mosaicism, suggesting that mosaic mutations may be a biomarker for cancer risk (Laurie et al. 2012). Jacobs *et al.* also found that the prevalence of mosaicism increased to about 2% in cancer-free individuals older than 75 years, and found an even stronger relationship between blood mosaicism and incident hematological cancer (Jacobs et al. 2012).

The largest analysis of 127,000 adults, which focused on events >2 Mb, showed an overall rate of mosaic events of 0.73%, of which approximately half of the detected events were mosaic copy number neutral uniparental disomy (48%), followed by mosaic losses (34%) and mosaic gains (17%). 14% of the individuals had multiple events. The mosaic changes showed an oblique chromosomal distribution: gains were most common in chromosomes 8, 12 and 15, losses affected mainly chromosomes 13 and 20, and the most common mosaic UPDs involved chromosomes 9 and 14. Gains and losses involved a higher proportion of cells than copy-neutral events. Mosaic events were rare in individuals under the age of 50. Their frequency was six-fold higher in individuals aged over 75, suggesting that mutant clones expand with age (Machiela et al. 2015).

Recently, Vattathil and Scheet applied a computational method using estimated haplotypes to characterize mosaic structural mutations in 31,100 GWAS subjects (Vattathil and Scheet 2016, Supplementary). They investigated 2.9% somatic copy number alterations (SCNAs) of all samples. There were 70 gains, 202 hemizygous losses, 30 CN-LOHs and other were left unclassified. They also provided additional survey with previously identified somatic mutations by Laurie *et al*., but they used hapLOH haplotype-based approach and identified larger number of SCNAs, 794 of them were unique (1093 vs 379). The study also confirmed that SCNA prevalence is positively associated with age, in their results over 80-years old individuals it is approximately 80% (Vattathil and Scheet 2016). Examples of microarray studies of mosaicism are summarized in Supplement 1.

1.6 Female reproductive system, ovarian reserve and fertility outcome

Healthy ovarian function is essential for the general health of a woman and for the production of sex steroids, which are needed for the development of the genital tract and for bone density (De Vos et al. 2010). In the embryo, primordial germ cells (PGCs) migrate from the yolk sac to the urogenital ridge at around 5 and 6 weeks (Mamsen et al. 2012), which in the female embryo become the ovaries containing a non-renewable reserve of germ cells. According to traditional standpoint the human ovary holds a decreasing reserve of oocytes from fetal life until the woman enters menopause. The oocyte quantity reaches its peak already before birth, with approximately seven million follicles (Figure 7). At birth the number decreases to around 1-2 million and at the onset of puberty only 300,000-400,000 follicles are left.

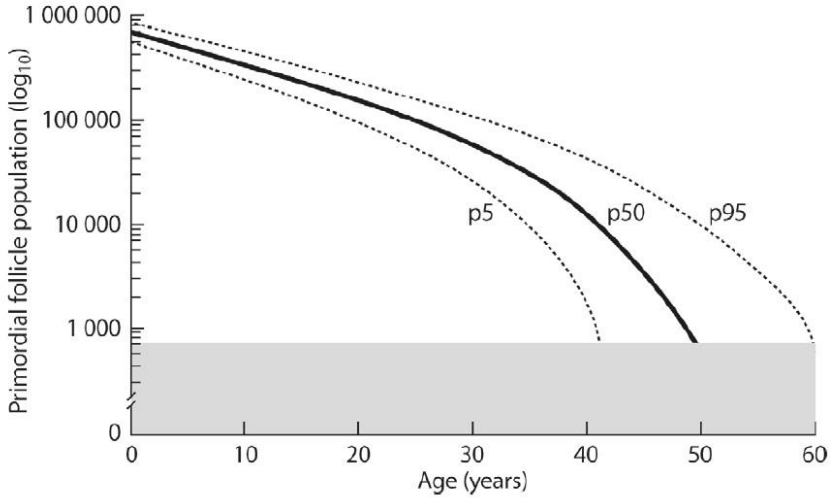


Figure 7. Reduction of ovarian follicular pool during a woman's life. Shaded area indicates level of follicle population at which ovarian failure or menopause occurs. p5=5th percentile, p50=median, p95=95th percentile. Adapted from (De Vos et al. 2010).

When the woman enters menopause at an average age of 51, approximately 1000 follicles remain in the ovary (Faddy and Gosden 1996). During the reproductive life of a woman, approximately 400 of the original follicles will ovulate and at the vast majority will undergo atresia by apoptosis (Vaskivuo and Tapanainen 2003). Besides the decrease in follicle number, also the quality of follicles and oocytes decline with age, exhibiting structural damage and aneuploidy with an increasing frequency (de Bruin et al. 2004).

As social and economic development has increased during the last century, there has been a substantial decline in fertility rates (Myrskylä et al. 2009). Ovarian ageing and its associated effects on fertility has received greater attention as an increasing number of women in modern society choose to postpone the age at which they bear children. Age is a crucial factor affecting female's fertility and reproductive outcome. Older women have to face the natural limits of their own reproductive system, due to the fact that fecundity is reduced with increasing female age (Broekmans et al. 2007). Over the past 20 years, the median age of women at their first birth in Estonia increased from 22.8 to 26.5 (Eurostat database)[1]. A similar trend has been noted all over the Europe, where data shows an average median age at first birth almost 29 years of age. Up to 10% of women in the general population are estimated to become menopausal by the age of 45 and 1% before the age of 40 (Nikolaou and Templeton 2004).

1.7 Premature ovarian failure

Premature ovarian failure (POF), also known as primary ovarian insufficiency (POI), refers to development of amenorrhea due to cessation of ovarian function before the age of 40 years with a prevalence about 1% in women under the age of 40, and 0.1% in women under the age of 30 (Coulam et al. 1986). From a clinical point of view, POF has been defined as 4–6 months of amenorrhea, a rise in serum follicle-stimulating hormone (FSH) levels to greater than 40 mIU/L, and hypoestrogenism (Goswami and Conway 2005). POF is one of the causes for female infertility and has a poor reproductive outcome of infertility treatments (de Boer et al. 2003). As a consequence of hypoestrogenism, POF is associated with a greater risk of osteoporosis, osteoarthritis and cardiovascular disease (Shuster et al. 2010).

Despite the presence of numerous studies, the pathophysiological development of POF remains unknown in most cases. The disorder is heterogeneous: possible causes include chromosomal abnormalities, gene mutations, autoimmunity, metabolic disorders, infections and iatrogenic treatments like ovarian surgery, radiotherapy or chemotherapy that somehow leads to follicle dysfunction or depletion (Nelson et al. 2009). POF usually appears

sporadically, but there can be also a family history in 4-30% of cases (Conway et al. 1996; Vegetti et al. 1998).

Chromosomal abnormalities include numerical defects like monosomy X or X chromosomal mosaicism, X deletions, X-autosome translocations, X-isochromosomes and other rearrangements. The prevalence frequency of these anomalies based on different studies is between 10-13% (Qin et al. 2015). Both, X chromosome and autosomes, has a wide range of candidate genes causing non-syndromic premature ovarian failure (see Supplementary Table 1 and 2 from Qin et al. 2015). Distinct from non-syndromic POF, in some patients POF appears to be associated with another syndrome (syndromic POF) that may manifest the disorder as part of their phenotypic spectrum. The most common single genetic explanation for premature ovarian failure is represented by Fragile X syndrome. Around 21% of cases of familiar POF are associated with the premutation of *FMR1* gene (MIM*309550)[3], located in the X chromosome (Sullivan et al. 2005). There are also remarkable numbers of genome-wide approaches to locate susceptible loci or genes causing POF, but there are notable imbalances among different populations (see Supplementary Table 5 from Qin et al. 2015).

Our lab performed genetic association study to investigate the potential associations between copy number variations and the onset of premature ovarian failure using Estonian population-based biobank samples. In addition to the three regions investigated before, in this study, 15 novel rare microdeletion and microduplication regions were found that may contribute to spontaneous POF. Among the genes found in the novel regions are for example *FMN2* (1q43), *SGOL2* (2q33.1), *TBP* (6q27), *SCARB1* (12q24.31), *BNC1* (15q25) and *ARFGAP3* (22q13.2). These genes are essential for meiotic progression or in follicular growth and oocyte maturation, respectively. This study confirmed that CNVs have possible role in pathogenesis of POF (Tšuiko et al. 2016).

2. EXPERIMENTAL PART

2.1 Objectives

The present study was carried out with the following primary objectives:

1. Detection of copy number alterations in POF patients:

The detection of DNA copy number changes, including duplications and deletions across the genome of POF patients following whole-genome genotyping by SNP-array technology.

2. Detection of copy-neutral alterations in POF patients

The prevalence of DNA copy-neutral loss of heterozygosity (CN-LOH) across the genome in POF patients versus the controls.

3. Detections of mosaic abnormalities in POF patients

The assessment of somatic mosaic changes, including mosaic deletions, duplications and UPDs, and their degree of mosaicism in POF patients and in the controls.

2.2 Materials and Methods

2.2.1 Study Objects

Participants of the study were selected from the biobank of EGCUT (Estonian Genome Center, University of Tartu) on the basis of the phenotype information. In total, we analysed 587 women with spontaneous premature ovarian failure, of which 345 were classified as cases, who has secondary amenorrhea occurred before the age of 40 and the remaining 242 women are with iatrogenic manipulation, especially gynecological operation causing amenorrhea. The latter cohort are classified as controls in this particular study. In addition, we also used Estonian general population samples genotyped with Illumina HumanOmniExpress BeadChip arrays (N=5132) from EGCUT to exclude benign population-specific CNVs.

2.2.2 SNP genotyping by the use of SNP-array technology

HumanCoreExome and HumanOmniExpress BeadChip arrays (Illumina Inc.) were used for genotyping. The HumanOmniExpress BeadChip contains >715,000 SNP markers with median marker spacing of 2.1 kb, while HumanCoreExome BeadChip has 547,644 markers, with 265,919 of them being focused in exome regions, and a median marker spacing of 1.9

kb. Sample processing and assay preparation were performed according to manufacturer protocols in EGCUT Genotyping Core Facility.

2.2.3 Algorithmic methods used for detection of genomic abnormalities

We have initially analyzed the SNP-array data by Illumina's GenomeStudio software Genotyping Module v.3.1. A call rate of >98% was accepted as the primary quality control for each sample. Log R ratio (LRR) and B-allele frequency (BAF) at each SNP loci are two important types of values for CNV detection. LRR and BAF information was analyzed with an algorithmic method which is conceptually similar to haplarithmisis (Zamani Esteki et al. 2015). The algorithm detects allelic imbalances by decifering both LRR and BAF values, with the capacity of revealing (mosaic) copy number and copy-neutral abnormalities across the genomes.

2.3 Results

2.3.1 Copy number alterations in POF patients

Total of 345 samples of premature ovarian failure patients cohort were analysed for detection of genome-wide copy number changes. LRR information indicated four patients with monosomy X (Turner syndrome) and one control with trisomy X (Triple X syndrome), which were exluded from the following analyses. Subsequently, we compared all the detected CNV regions of the patients with the controls. All overlapping regions were left out and 122 non-overlapping copy number variation regions (CNVRs) were remained. Additional quality filters were applied after obtaining remaining regions. CNVs that were less than 100 kb were excluded from the following data analyse against population-based CNVs from EGCUT (N=5132). Confidence interval for this analyse was CI=30. Elevated values provide better insight and reliability as well maximise accuracy for further analyse. Results from comparison of 40 CNVRs (>100 kb) against benign CNVs (N=5132) revealed 23 potentially novel CNV regions (Table 3, see below), 17 of these overlapped with population-based CNVs.

In detail, in the evaluation of 345 POF patients, this analysis indicated a total of 23 new CNV regions and included 16 copy number losses (70%) and 7 copy number gains (30%) ranging from 100 kb to 7,5 Mb in length (mean=672 kb, median=191 kb). In addition, our data shows that losses are more frequent than gains.

In previous research with spontaneuos POF patients, total of 55 microduplications or microdeletions were found (Tšuiko et al. 2016, Supplementary) and 15 of them were found to be relevant to POF condition. The second important matter is that nine of these 15 regions were probabaly excluded by our patient-control comparison or analysis against population-based CNVs. Tšuiko *et al.* considered CNVs as benign if these were present >1% of the population, but we decided to excluded all overlapping regions. In our analysis, however, we were able to detect 70,6% of the same regions, of which 20% were mosaic. Improtantly, we detected 11 novel (Table 3, blue rows) CNV regions (findings not validated) from the same cohort of patients and these novel aberrant regions were studied along with other detected regions using the UCSC Genome Browser database[2] to evaluate the potential clinical relevance to POF of a particular CNV.

Table 3. Overview of CNV regions (N=23) identified in this study. Novel variants (N=11) from the same cohort of POF patients examined before are presented in blue rows.

Chr	Band	Start	Stop	Length (bp)	Change
2	p25.3	1116648	1584345	467697	Deletion
2	p22.3	35597319	35709163	111844	Deletion
2	p11.2	89918400	90109261	190861	Deletion
2	q13	110863908	110983320	119412	Deletion
3	q29	197581268	197833758	252490	Duplication
4	p16.3	60055	171534	111479	Duplication
4	q13.3	69434042	69537475	103433	Deletion
4	q23	99363155	99510279	147124	Deletion
4	q24	102839161	103023963	184802	Deletion
5	p15.32	6021846	6136125	114279	Deletion
7	p22.3	2159817	2310974	151157	Deletion
7	q36.3	158456893	159119220	662327	Duplication
8	p23.3	163226	1178457	1015231	Duplication
9	p22.1	19074820	19281501	206681	Duplication
9	p12	39004140	39151736	147596	Deletion
13	p12	114507886	115091330	583444	Duplication
14	q12	27071318	27393217	321899	Deletion
15	q25.1	80500456	80642236	141780	Duplication
16	p13.3	95254	2305539	2210285	Deletion
16	p13.3	84529382	84630053	100671	Deletion
18	q22.2-q23	70467981	78014582	7546601	Deletion
19	p13.2	8969269	9298364	329095	Duplication
21	q21.2	24166144	24400039	233895	Deletion

2.3.2 Copy number neutral alterations in POF patients

Copy-neutral LOHs were assessed using our algorithmic method. Visual inspection of the profiles deduced by the algorithm revealed that the approximate copy-neutral events across the genomes of both patients and controls are relatively high. We counted the CN-LOHs of every chromosome in both group and divided it with the size of the each cohort (Figure 8). Following this analysis, we could not find a significant difference between these cohorts. Interestingly, we observed relatively large CN-LOH regions in chromosomes 2, 3, 7, 17 and 22 in one POF patient. Demonstrations of these events with approximate sizes are shown in Supplement 4.

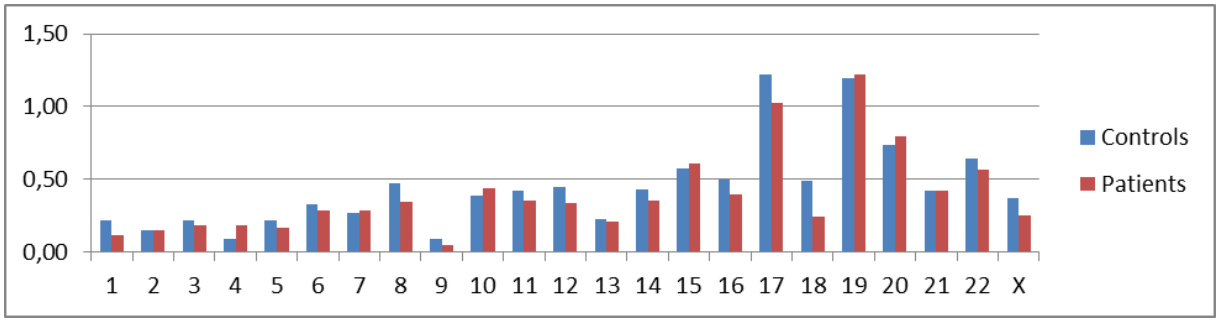


Figure 8. Proportions of CN-LOH events by chromosome from patients and controls. X axis – chromosome number, Y axis – total of CN-LOH events seen by chromosome divided with number of samples from patients or controls, respectively.

2.3.3 Mosaic abnormalities in POF patients

We also identified mosaic copy number gains and losses, as well as mosaic copy-neutral UPD events. For mosaic CNAs, each sample was explored for BAF values [BAF_G (greater), BAF_L (lower)] together with LogR in the same chromosomal region. Based on the values seen in the data, we were able to determine degree of mosaicism and the type of alteration (gain or loss). Degree of mosaicism was determined as described by Conlin *et al.* (Supplement 2). First, all mosaic findings were compared between patients and controls. All non-overlapping events (comparison of the LogR values) that were larger than 100 kb were analysed against population-based cohort (N=5132) to exclude benign copy number changes to distinguish correct mosaic changes.

Results from this analysis left only 24 mosaic events (Table 4, see below) for further investigation. In patients, 13 mosaic duplication (54,2%) and 2 mosaic deletions (8,3%) were found. One of these deletions, however, is the most significant finding of mosaic monosomy

for chromosome X (Figure 9, see below). In control group, these numbers were 8 (33,3%) and 1 (4,2%), respectively.

Table 4. Results of mosaic copy number abnormalities (N=24) detected in this study. White rows are representing the results from patients and blue rows from controls group. Mosaic findings that were the same with the CNV regions investigated in previous POF cohort study are bolded.

Chr	Band	Start	Stop	Length (bp)	Mosaic %	Change	Subject
2	p22.2-p22.1	36775797	38580233	1804436	40	deletion	patient
2	**q21.2**	**133225803**	**133902333**	**676530**	**45**	**duplication**	**patient**
2	**q33.1**	**200250898**	**201845999**	**1595101**	**40**	**duplication**	**patient**
3	p26.3	1577232	2667189	1089957	40	duplication	control
3	q29	195479250	195877603	398353	45	duplication	patient
4	p16.1	10188326	11112475	924149	45	duplication	patient
4	q28.1	127041554	127649306	607752	40	duplication	patient
4	q34.1-q34.2	174961164	175658175	697011	45	duplication	control
5	**p13.2**	**33955326**	**34753340**	**798014**	**40**	**duplication**	**patient**
7	p11.2	55798357	56504381	706024	45	duplication	control
10	q26.3	132547184	133057358	510174	40	duplication	patient
11	p11.12-p11.11	50070332	51372036	1301704	40	duplication	patient
12	p13.33	1017830	1517350	499520	45	duplication	control
12	q24.33	133448936	133810935	361999	40	duplication	patient
16	q23.1	76986417	78635193	1648776	40	duplication	control
17	p13.1	10119922	10423933	304011	40	duplication	control
17	**q12**	**34815551**	**36245768**	**1430217**	**45**	**duplication**	**patient**
19	q13.41-q13.42	52495380	54803022	2307642	50	deletion	control
20	p13	2308942	2725678	416736	40	duplication	patient
20	p13	3183919	3641868	457949	40	duplication	patient
21	**q21.3**	**27484463**	**27852668**	**368205**	**40**	**duplication**	**patient**
22	q11.21	18222118	18606947	384829	45	duplication	control
22	q11.21-q11.22	20740778	21806401	1065623	45	duplication	control
X	p22.33-q28	2700157	154780283	152080126	15	deletion	patient

Our data displays that mosaic duplications (87,5%) are more frequent than mosaic deletions (12,5%). The most frequent chromosome of event observed was chromosome 2 (12,5%) for patient cohort and chromosome 22 (8,3%) for control cohort. Average lengths for mosaic duplications in patients (754 kb) and in controls (799 kb) were the same. The abnormal cell

proportions were between 40-45% for mosaic duplications and 15-50% for mosaic deletions. Five of the mosaic duplication events in patients (see Table 4, bolded) are the same with the CNV regions described in previous study (with a range of 40-45% degree of mosaicism). Detected mosaic CNVs were examined for genes relevant to POF with using UCSC Genome Browser database[2].

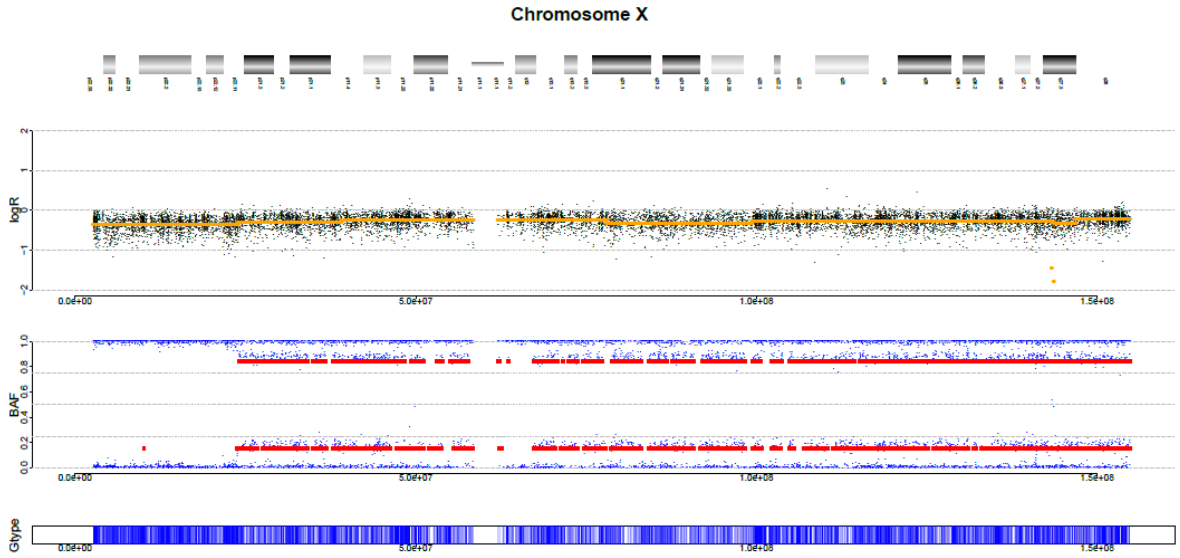


Figure 9. Illustration of mosaic monosomy of chromosome X (BAF values for p22.11-q28) from patients group. Mosaic monosomy is detected when LRR (top) shows a decrease from normal copy number state. In addition, the BAF (bottom) shows a split from the intermediate value of 0.5 and indicates the percent of abnormal population of cells. Loss of and A allele, results in a shift of the frequency towards 0%, while loss of a B allele results in a shift towards 100% and the percent of mosaicism can be calculated from the relative shifting of the BAF values.

Mosaic copy-neutral events can also be captured by the detection of BAF deviations from the expected value of 0.5 and the LRR value is not represented. The degree of UPD was determined as described previously by Conlin *et al*. (Supplement 3). We were able to detect total of 241 mosaic UPD cases, ranging in length 14,6 kb to 63 Mb, but only the changes larger than 1 Mb (N=24) were considered for further analysis. Half of these events (N=13), were located around the centromeric or telomeric regions of the chromosome (data not shown). Out of remaining 11 mosaic UPD events for further investigation, six occured in patients and five in controls (Table 5, see below). The abnormal cell proportions for UPD events ranged between 10-60%. Full chromosome mosaic UPD occured in chromosome 13 in one of the patients (Figure 10, see below) with 30% abnormal cells.

Table 5. Summary of mosaic uniparental disomy (UPD) findings larger than 1 Mb in patients and controls. White rows are representing the UPD regions from patients and blue ones from controls. The values for start and end point are from BAF values.

Chr	Band	Start	Stop	Length (bp)	Mosaic UPD % of abnormal cells	Subject
2	p11.2	86406556	88315993	1909437	15%	control
3	p21.31-p21.1	49701298	53457291	3755993	10%	control
6	p22.2-p21.33	26405835	31427139	5021304	50%	control
6	p22.1-p21.33	29230577	31000042	1769465	60%	patient
6	p22.1-p21.32	29788770	32595060	2806290	40%	patient
6	p21.33-p21.32	31000042	32680122	1680080	55%	patient
10	q11.21-q11.22	45478092	47703869	2225777	30%	patient
13	q12.12-q34	23162926	114878438	91715512	30%	patient
14	q32.33	104581072	107287663	2706591	10%	patient
16	p13.11	15129970	16291971	1162001	30%	control
19	q13.2-q13.31	42463050	43864419	1401369	30%	control

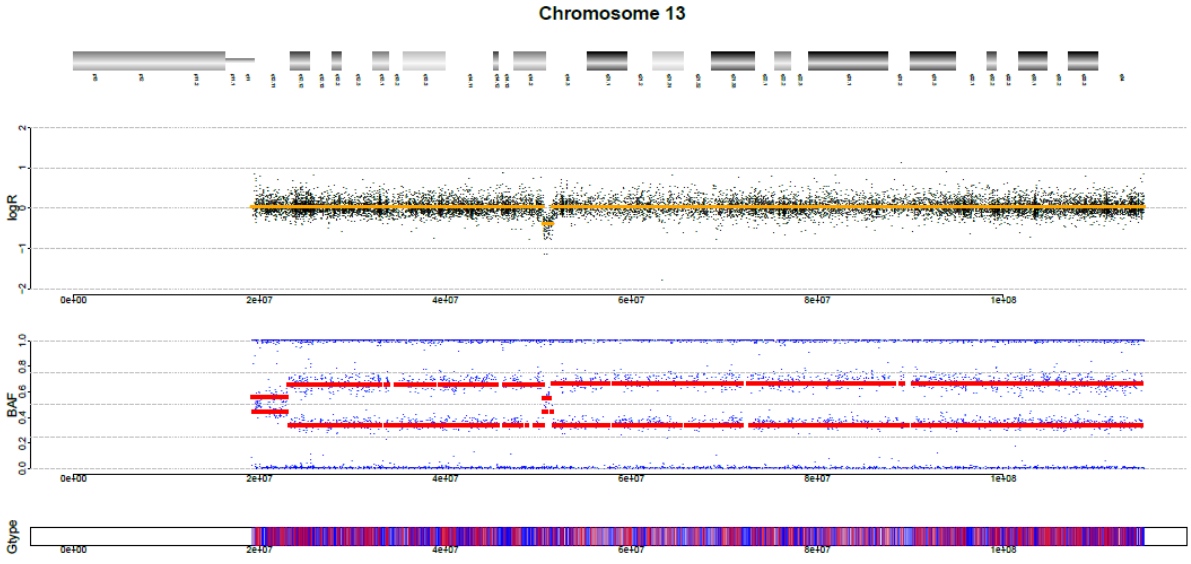


Figure 10. Output of mosaic UPD event in chromosome 13. The LRR (top) is consistent with normal copy number. The BAF plot (bottom) demonstrates a shift in genotype frequencies and the percent of mosaicism for the abnormal cell line could be estimated at 30%. In 13q14.3 is also detected a deletion.

2.4 Discussion

Numerical and structural chromosome alterations, including deletions, duplications, uniparental disomy and translocations, may cause a wide variety of genetic disorders with different severity. However, some of these variations may also be benign within a population (Feuk et al. 2006). Identifying DNA variants that contribute to a disease or syndrome is a key objective in human genetics. State-of-the-art genome technologies such as SNP microarrays and next generation sequencing have drastically enhanced our understanding about the underlying mechanisms of genetic anomalies as well as genetic heterogeneity within and between individuals. Along with technological advances, new algorithmic methods are crucial in identifing genetic variations across large datasets. Although many algorithmic methods have been designed to detect CNVs, but they are not accurate enough to detect CN-LOH and somatic mosaic aberrations, in particular aberrations with low degree of mosaicism. Within a population of cells originating from a single zygote, any post-zygotic somatic change may results in mosaicism, in which a subset of cells harbors a different genetic variant.

Mosaicism is a biological phenomenon that indicates the presence of two or more chromosomally different cell lines in an individual arising from a single zygote (Strachan and Read, 2011). The underlying mechanism of mosaicism formation involves a nondisjunction error during a mitotic cell division or during meiosis followed by a post-zygotic correction of aneuploidy. The implication of genetic mosaicism for disease phenotypes is difficult to predict in many cases, since it is dependent upon timing of the mutation event, cell type or types affected and distribution of genetically aberrant cells. However, mosaic forms of disease are generally associated with milder phenotypes compared to non-mosaic aberrations.

SNP microarray data enables quantitative measurement of aberrant cell population via the B-allele frequency (BAF) values. Mosaic events can be captured by analysing SNP array data, especially from assessing multiple clusters of heterozygous alleles showing BAF and LRR values different from those expected for regular heterozygous deletions, duplications or loss of heterozygosity (LOH) events. Interestingly, a number of studies demonstrate that structural variants occurring in mosaic forms are more frequent than expected, and thus they may play a relevant role in human diversity and disease susceptibility (Conlin et al. 2010; Jacobs et al. 2012; Laurie et al. 2012; Razzaghian et al. 2010; Vattathil and Scheet 2016).

In this study, DNA of 345 individuals with spontaneus POF and 242 individuals with iatrogenic POF have been used to investigate for evidence of copy number variations and somatic mosaicism using Illumina genome-wide SNP array data and a novel computational approach enabling an accurate detection of (mosaic) genetic alterations. The presence of CNVs in phenotypically selected premature ovarian failure patients from Estonian population-

based biobank has been investigated in a previous study (Tšuiko et al. 2016), using two independent Hidden Markov Model-based algorithms, QuantiSNP (Colella et al. 2007) and PennCNV (Wang et al. 2007). Here, we applied a novel computational algorithm, inspired from haplarithmisis concept (Zamani Esteki et al. 2015) and compared the two studies.

Copy number alterations in human genomic DNA are most likely the cause of different rare or complex disorders (Vulto-van Silfhout et al. 2013). Premature ovarian failure is highly heterogeneous disease with strong genetic component. We identified four patients with monosomy X and it is known that Turner syndrome (45,X) is the most common congenital cause of POF (Schlessinger et al. 2002). X monosomy without mosaicism is more typically found in primary amenorrhea and according to phenotype information, three of our patients had absence of menstruations, but in one of these Turner cases occurred amenorrhea at the age of 30.

From copy number variation analysis, performed with our method, we investigated a total of 23 non-overlapping CNVRs from patient cohort. In comparison with previous study, which found 55 rare microdeletion or microduplication regions (see Tšuiko et al. 2016, Supplementary), our method was able to detect 70,6% of the same regions and eleven novel regions among the same population of patients. However, 29,4% of the regions remain undetected. We did not find any susceptibility genes among these 23 regions relevant to POF pathogenesis. According to the point that CNVs from spontaneous premature ovarian failure patients were already examined before, we have to take into consideration many aspects between these studies that may lead to different results: (1) the algorithmic methods to detect genetic alterations were different; (2) the interpretation of CNVs were performed with different strategies and (3) the controls cohorts were not accordance. Despite of these annotations, the comparison results were quite expected and the main goal of this study was primarily directed for detection of mosaic abnormalities.

We identified 24 non-overlapping mosaic copy number variation regions (>100 kb), 15 of them among patients and nine among controls. It was interesting that mosaic duplications (87,5%) were much more presented than mosaic deletions (12,5%). Among mosaic alterations, the most significant finding of this study was mosaic monosomy for X chromosome, where the terminal part of a short arm is not mosaic (see Figure 9). According to phenotype information, the patient had relatively short stature (156 cm) and menopause occurred at the age of 30. With regard to short stature, it has been suggested that the *SHOX* gene (MIM *312865)[3] in pseudoautosomal region (PAR1) is the major player and that haploinsufficiency of this gene leads to the growth failure seen in Turner syndrome (Rao et al. 1997). We hypothesize that this particular monosomy may have originated from normal XX

cells with CN-LOH in the terminal part of short arm, where the *SHOX* gene is located, which probably initiated X monosomy in a subset of cells. The proportion of abnormal cells in this particular case was 15%. The low percentage of cells affected may indicate that this individual with mosaic monosomy X will have a milder phenotype than those with complete monosomy X. Mosaic TS patients are more likely to experience normal pubertal development, regular menstrual cycles and to conceive spontaneously compared to those with 45,X monosomy (El-Mansoury et al. 2007). Our patient had normal menstruations between 19-30 years of age and no other clinical features for Turner stigmata was not observed. Mosaic monosomy could either arise by mitotic non-disjunction in a diploid embryo leading to monosomy in a subset of cells, or monosomy rescue in some cells of a monosomic zygote early in development (Conlin et al. 2010).

Other mosaic changes were relatively small (approximately 885 kb in length). Five of the mosaic alterations were in the same regions as described in previous study. Three of these regions involved potential candidate genes for POF, two of them (*SGOL2* in 2q33.1 and *LHX1* in 17q12) are consistent with the findings of Tšuiko *et al*. In our opinion, these five regions detected as mosaic CNVs are more precise because the methods used in previous study can not detect mosaic aberrations. This discovery leads us to argue about the proportion of mosaic events among detected CNV results. Extrapolating our results, the difference would be as high as 9.1%. These observations are concordant with Rodriguez-Santiago *et al.*, who suggested in their study that the occurrence of mosaic events is underestimated because of deficiency of common detection methods. They also brought out, that is plausible that some of the CNVs registered in the Database of Genomic Variants might correspond to mosaic rearrangements. Thus, optimization of the analysis of data obtained with SNP arrays, as well as the development of novel algorithms for the analysis of low-degree mosaicism is required to improve the accurate detection of these events in human populations (Rodríguez-Santiago et al. 2010).

In the mosaic deletion in chromosome 2 (chr2:36,775,797-38,580,233) is located the *CYP1B1* (Cytochrome P450 1B1) gene (MIM *601771)[3], which is involved in estrogen hydroxylation (Nishida et al. 2013) and with *CYP1A1* (Cytochrome P450 1A1) it is the primary gene involved in estrogen metabolism (Tsuchiya et al. 2005). Affecting this gene, higher levels of estrogens throughout the reproductive life are believed to exist (Hanna et al. 2000), but the way in which this would affect the ovarian follicular pool remains to be elucidated. However, previous investigations showed that *CYP1B1* polymorphisms are unlikely to influence the age at menopause (Hefler et al. 2005; Long et al. 2006).

Uniparental disomy (UPD) is the presence of a chromosome pair derived only from one parent present in a disomic cell line (Liehr 2010). In this study, we identified 24 mosaic UPD events larger than 1 Mb and for further analyse was considered eleven of these regions. Remaining 13 mosaic UPD events were located near centromere (11) or telomere (2). For example, copy-neutral telomeric events are theoretically expected if mitotic recombination represents DNA repair occurring secondary to a double strand break, which is the generally favored mechanism. Previous studies confirm the strong correlation of segmental UPD with telomeric regions (Kotzot 2008). Breakpoint analyses of regions surrounding mosaic events might aid in understanding mechanisms responsible for event initiation, but the current resolution of event boundaries in SNP microarrays is limited as a result of insufficient probe density (Machiela et al. 2015). Of the eleven mosaic UPD cases, six occurred in patients and five in controls. The abnormal cell proportions for these UPD events ranged between 10-60%. If we take into account all the detected mosaic UPD events (N=241), these findings are also consistent with other studies that implicate mosaic copy-neutral UPDs as the majority of detected mosaic events (Machiela et al. 2015; Rodríguez-Santiago et al. 2010). Mosaicism for segmental UPD most probably arises through mitotic recombination or alternatively by chromosome breakage and repair, in which the DNA sequence from one allele is copied across to replace the other allele to generate homozygosity (Biesecker and Spinner 2013). SNP arrays has the advantage of precise mapping of segmental UPD regions, but this technology can only identify isodisomic cases (isodisomic regions) (Conlin et al. 2010). In this study, both patients and controls were detected the same amount of segmental mosaic UPD events.

Full chromosome mosaic UPD (isodisomy) occured in chromosome 13 in one of the patients (see Figure 10) and using the shift in allele frequencies the percent of mosaicism for the abnormal cell line was estimated at 30%. The origin of UPD often entails meiotic nondisjunction followed by a mitotic rescue event, but the possibility of crossing-over of homologs and missegregation of translocated chromosomes and other complex events are also possible (Kotzot 2008). The possible mechanism by which this particular chromosome 13 mosaic UPD occurred is probably due to post-fertilization error, occurrence of mitotic nondisjunction and the subsequent duplication of the remaining chromosome.

Mitotic recombination and gene conversion can result in acquired uniparental disomy (aUPD), where loss of heterozygosity has occurred without loss of genomic material and both alleles of a gene are derived from one parent. This is also called copy-neutral loss of heterozygosity (CN-LOH) and leads to the generation of daughter cells with reciprocal chromosomal products (Fitzgibbon et al. 2005). CN-LOH does not alter the copy number of

affected chromosomal segment and are detectable with BAF values of allelic imbalances. In this study, CN-LOH events were explored by visual inspection of the profiles deduced by the algorithm. Both patients and controls showed relatively high amount of copy-neutral events across the genomes, but we could not find a significant difference between these cohorts. Our results shows diversity between events among chromosomes, for example in our study in chromosome 9 we observed the least CN-LOH events, but in Laurie *et al*. chromosome 9 has the highest amount of occurrence (see Laurie et al. 2012, Supplementary Figure 9). Even though, the other chromosomes showed approximately the same rate of events.

In addition, we observed relatively large CN-LOH regions in chromosomes 2, 3, 7, 17 and 22 in one of the patients. CN-LOH is not expected to cause abnormalities unless genes on affected chromosome loci are subject to genomic imprinting, which is the monoallelic expression that is dependent on the allele's parental origin. When aUPD occurs in a region containing an imprinted domain, cells will inherit either two active or two repressed copies, leading to an abnormal dosage of these imprinted gene products (Lapunzina and Monk 2011). Mostly, CN-LOH has been shown to be involved in the development of cancers and developmental disorders. The list of conditions that have been associated with CN-LOH is continuously growing (Makishima and Maciejewski 2011; Tuna et al. 2009). In addition, this particular patient, has relatively large (~62,5 Mb) CN-LOH region in chromosome 7 (q21.2-q36.2), encompassing two strong candidate genes for POF (OMIM nomenclature for POF5 (*NOBOX*) and POF8 (*STAG3*)[3]. We might assume that this LOH is possible cause of POF condition as both genes are located in this region, but additional molecular validation is needed for confirmation.

In summary, it is highly possible that the mosaic occurrence of genomic variants may have been missed previously with the standard methods applied to CNV detection with SNP arrays. In the light of recent findings about the presence of detectable somatic mosaicism, our method is certainly advantagous in detecting low-grade mosaicism and resolve the mechanistic origin of constitutional genetic anomalies detected in multi-cell DNA samples. If we consider for calculation all the detected mosaic events larger than 2 Mb like many large-scale studies have performed, then the presence of detectable clonal mosaicism in our study is 3.2%.

As mosaicism is becoming an increasingly frequent observation, there are many issues that need to be aware of: (1) first and foremost, mosaicism has been underestimated within the population. Mutations accumulate in individuals over their lifetime that may not have a phenotypic effect and we surely should expect to find mosaicism in phenotypically normal individuals; (2) similarly, loss of the X chromosome might exist in mosaic form, which means

that having some 45,X cells does not mean the patient has Turner syndrome. It is therefore important to understand that finding mosaicism does not make it causal of the phenotype under investigation; (3) if the mutant cells do not have growth advantage over the normal cells, then it is related to reduced percentage of detectable mosaicism associated with a more severe phenotype. Mosaicism may have a greater impact on the phenotype when expressed in one tissue over another and it is more likely occurring earlier during development.

SUMMARY

Genetic variation between individuals has been extensively investigated, but differences between tissues within individual are far less understood. It is commonly assumed that all healthy cells that arise from the same zygote possess the same genomic content. However, somatic mosaicism arised from post-zygotic mutations for copy number and copy-neutral chromosomal rearrangements has been recently identified as a relatively common source of genetic variation between normal individuals, but its prevalence in population is poorly defined or underestimated. Despite numerous techniques available to be applied, low-grade mosaicism remains a difficult issue in both clinical diagnostics and research. At this point, it is worth to add that there is a growing concern in the field about whether some CNVs categorized as germline-acquired mutations in biobanks are actually somatically acquired alterations.

Aims of this thesis was to review the basic forms of mosaic abnormalities (deletions, duplications, UPDs and CN-LOHs) with their possible mechanisms of post-zygotic formation during mitosis. We introduced the common detection methods and demonstrated the results from recent genome-wide array approaches to detect mosaic alterations. In practical part of the thesis, we analysed genotype data of premature ovarian failure (POF) patients and controls with novel algorithmic method for detection of mosaic structural rearrangements. We found mosaic aberrations in 8.2% of samples, including 23 mosaic copy number variation (CNV) regions, one mosaic X monosomy and 24 (larger than 1 Mb) mosaic uniparental disomy (UPD) events. In addition, we were able to investigate 23 novel CNVs among patients. Reanalysis of the previously assessed SNP array analysis with POF patients demonstrated that there are hidden genomic mosaic events that cannot be detected by using standard detection methods. Based on the information provided from this study, it is apparent that some cases of previously detected CNVs may contribute mosaic form of alteration.

It is still largely unknown exactly what types of environmental or genetic factors are responsible for the development of detectable human clonal mosaicism and therefore, in this genetic area is plenty of more to discover. We emphasize that somatic mosaicism should be considered as a form of inter- and intraindividual genetic variation.

Geneetilised aberratsioonid enneaegse ovariaalpuudulikkusega patsientidel

Katre Teearu

KOKKUVÕTE

Tänapäeval arvatakse, et genoom on kõigil inimestel 99,9 protsenti identne ja ainult 0,1 protsenti erinevust teeb meid üksteisest niivõrd erinevaks. Just see väga väike osa DNA-st on meditsiiniteaduse peamisi prioriteete, et uurida, kuidas geneetiline varieeruvus avaldub ning seeläbi põhjustab erinevaid haiguseid. Inimorganism koosneb umbes 10^{14} rakust ja seega ei ole üllatav, et DNA replikatsioonil või rakujagunemiste käigus inimese eluea jooksul tekib vigu, mis genereerivad „personaalse“ genoomiga rakke. Üks osa inimkonna varieeruvusest on seletatud mosaiiksusega, mis on bioloogiline fenomen, kus viljastamisjärgselt esineb üksikindiviidi erinevates rakupopulatsioonides erineva genotüübiga rakke. Mosaiiksust on pikka aega peetud inimestevahelise ning ka -sisese geneetilise varieeruvuse põhjuseks. Mosaiiksed struktuursed ümberkorraldused genoomis võivad avaldada mõju indiviidi fenotüübile ning olla seeläbi erinevate haiguste põhjuseks, kuid see sõltub spetsiifilisest DNA regioonist, arenguetapist ja mõjutatud rakkudest.

On teada, et kolm peamist tüüpi struktuursed geneetilised ümberkorraldused, nagu deletsioonid, duplikatsioonid ja koopiaarvu neutraalsed heterosügootsuse kaotanud alad võivad tekkida ka postsügootiliselt. Ülegenoomsete genotüpiseerimiskiipide andmete analüüs koos arvutuslike meetoditega võimaldavad detekteerida kromosomaalseid aberratsioone kui ka somaatilisi mosaiikseid muutuseid. Enneaegne ovariaalpuudulikkus (POF) on multifaktoriaalne haigus, mis viib viljatuseni reproduktiivses eas naisi aastaid enne loomulikku menopausi. Olenemata haiguse heterogeensest etioloogiast, mitmed uuringud kinnitavad, et POFi esilekutsuvatest mehhanismidest on geneetikal tugev roll.

Käesoleva magistritöö eesmärgiks oli anda kirjandusel põhinev ülevaade peamistest mosaiiksuse vormidest erinevate struktuursete ümberkorralduste puhul ning tuua välja nende võimalikud tekkemehhanismid postsügootilistes protsessides. Samuti tutvustada, kuidas erinevate meetoditega on võimalik mosaiiksust detekteerida ning milliseid suuremahulisi uuringuid selles valdkonnas on läbi viidud. Töö praktilises osas teostati kogu genoomi hõlmav uuring, avastamaks mosaiikseid ümberkorraldusi enneaegse ovariaalpuudulikkusega patsientidel kui ka kontrollindiviididel. Kasutades genotüpiseerimiskiipide andmeid ja rakendades uut algoritmilist meetodit, tuvastati 8,2%-l valimist mosaiikseid aberratsioone, nende hulgas 23 koopiaarvu varianti, üks X kromosoomi monosoomia ja 24 (>1 Mb) uniparentaalse disoomia juhtu. Lisaks avastati patsientide hulgas 23 uudset koopiaarvu

variatsioonide regiooni. Korduvanalüüs enneagse ovariaalpuudulikkusega patsientidel näitas, et leidub peidetud genoomseid mosaiikseid aberratsioone, mida ei ole võimalik avastada standardmeetodeid kasutades. Uuringu tulemustest saab järeldada, et eelnevalt tuvastatud koopiaarvu variatsioonide hulgas võib mõningatel juhtudel ilmneda mosaiikseid vorme.

Siiani on täpselt teadmata, millised geneetilised- või ka keskkonnafaktorid vastutavad inimese mosaiiksete rakkude arengu eest ja seetõttu on antud geneetilisel areenil veel küllalt avastamisruumi.

ACKNOWLEDGEMENTS

I would like to express my warmest gratitude to several people:

First and foremost, I would like to express my deepest gratitude to my supervisor professor Ants Kurg for all his constant support and guidance throughout my thesis.

I am sincerely grateful to my co-supervisor dr. Masoud Zamani Esteki for the excellent guidance, data preparation, patience and corrections throughout the study. All of your support and advice was the basis to the progress of this work and I feel extremely grateful that I was given the opportunity to work with you during these months.

This thesis would not have been possible without the help and support from the staff at the EGCUT. I am thankful to Reedik Mägi for giving me good advice and helping me with computing assistance. I would also like to thank Margit Nõukas for processing data of population-based analysis.

Finally, I am grateful to my parents for believing in me. I can never thank them enough for whatever they have done for me.

REFERENCES

Alberts, B., et al. (2014), 'Essential Cell Biology', Fourth edition, *Garland Science, Taylor & Francis Group*, 603-629.

Auton, A., et al. (2015), 'A global reference for human genetic variation', *Nature,* 526 (7571), 68-74.

Bailey, J. A. and Eichler, E. E. (2006), 'Primate segmental duplications: crucibles of evolution, diversity and disease', *Nat Rev Genet,* 7 (7), 552-64.

Barbouti, A., et al. (2004), 'The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a complex genomic architecture with large, palindromic, low-copy repeats', *Am J Hum Genet,* 74 (1), 1-10.

Bielanska, M., Tan, S. L., and Ao, A. (2002), 'Chromosomal mosaicism throughout human preimplantation development in vitro: incidence, type, and relevance to embryo outcome', *Hum Reprod,* 17 (2), 413-9.

Biesecker, L. G. and Spinner, N. B. (2013), 'A genomic view of mosaicism and human disease', *Nat Rev Genet,* 14 (5), 307-20.

Bluth, M. H., Reid, M. E., and Manny, N. (2007), 'Chimerism in the immunohematology laboratory in the molecular biology era', *Transfus Med Rev,* 21 (2), 134-46.

Broekmans, F. J., et al. (2007), 'Female reproductive ageing: current knowledge and future trends', *Trends Endocrinol Metab,* 18 (2), 58-65.

Bruno, D. L., et al. (2011), 'Pathogenic aberrations revealed exclusively by single nucleotide polymorphism (SNP) genotyping data in 5000 samples tested by molecular karyotyping', *J Med Genet,* 48 (12), 831-9.

Campbell, P. J., et al. (2008), 'Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing', *Nat Genet,* 40 (6), 722-9.

Cohen, A. S., et al. (2015), 'Detecting somatic mosaicism: considerations and clinical implications', *Clin Genet,* 87 (6), 554-62.

Colella, S., et al. (2007), 'QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data', *Nucleic Acids Res,* 35 (6), 2013-25.

Conlin, L. K., et al. (2010), 'Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis', *Hum Mol Genet,* 19 (7), 1263-75.

Conrad, D. F., et al. (2010), 'Origins and functional impact of copy number variation in the human genome', *Nature,* 464 (7289), 704-12.

Consortium, International Human Genome Sequencing (2004), 'Finishing the euchromatic sequence of the human genome', *Nature,* 431 (7011), 931-45.

Conway, G. S., et al. (1996), 'Characterization of idiopathic premature ovarian failure', *Fertil Steril,* 65 (2), 337-41.

Cooper, W. N., et al. (2007), 'Mitotic recombination and uniparental disomy in Beckwith-Wiedemann syndrome', *Genomics,* 89 (5), 613-7.

Coulam, C. B., Adamson, S. C., and Annegers, J. F. (1986), 'Incidence of premature ovarian failure', *Obstet Gynecol,* 67 (4), 604-6.

Daber, R., et al. (2011), 'Mosaic trisomy 17: variable clinical and cytogenetic presentation', *Am J Med Genet A,* 155A (10), 2489-95.

de Boer, E. J., et al. (2003), 'Increased risk of early menopausal transition and natural menopause after poor response at first IVF treatment', *Hum Reprod,* 18 (7), 1544-52.

de Bruin, J. P., et al. (2004), 'Age-related changes in the ultrastructure of the resting follicle pool in human ovaries', *Biol Reprod,* 70 (2), 419-24.

De Vos, M., Devroey, P., and Fauser, B. C. (2010), 'Primary ovarian insufficiency', *Lancet,* 376 (9744), 911-21.
Denoeud, F., Vergnaud, G., and Benson, G. (2003), 'Predicting human minisatellite polymorphism', *Genome Res,* 13 (5), 856-67.
Desai, A. N. and Jere, A. (2012), 'Next-generation sequencing: ready for the clinics?', *Clin Genet,* 81 (6), 503-10.
Destouni, A., et al. (2016), 'Single-cell high resolution melting analysis: A novel, generic, pre-implantation genetic diagnosis (PGD) method applied to cystic fibrosis (HRMA CF-PGD)', *J Cyst Fibros,* 15 (2), 163-70.
Dunbar, A. J., et al. (2008), '250K single nucleotide polymorphism array karyotyping identifies acquired uniparental disomy and homozygous mutations, including novel missense substitutions of c-Cbl, in myeloid malignancies', *Cancer Res,* 68 (24), 10349-57.
Eggermann, T., et al. (2015), 'Mosaicism and uniparental disomy in prenatal diagnosis', *Trends Mol Med,* 21 (2), 77-87.
El-Mansoury, M., et al. (2007), 'Chromosomal mosaicism mitigates stigmata and cardiovascular risk factors in Turner syndrome', *Clin Endocrinol (Oxf),* 66 (5), 744-51.
Engel, E. (1980), 'A new genetic concept: uniparental disomy and its potential effect, isodisomy', *Am J Med Genet,* 6 (2), 137-43.
Engel, E. (2006), 'A fascination with chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements', *Eur J Hum Genet,* 14 (11), 1158-69.
Faddy, M. J. and Gosden, R. G. (1996), 'A model conforming the decline in follicle numbers to the age of menopause in women', *Hum Reprod,* 11 (7), 1484-6.
Feuk, L., Carson, A. R., and Scherer, S. W. (2006), 'Structural variation in the human genome', *Nat Rev Genet,* 7 (2), 85-97.
Fitzgibbon, J., et al. (2005), 'Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias', *Cancer Res,* 65 (20), 9152-4.
Forsberg, L. A., Absher, D., and Dumanski, J. P. (2013), 'Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime', *J Med Genet,* 50 (1), 1-10.
Frazer, K. A., et al. (2009), 'Human genetic variation and its contribution to complex traits', *Nat Rev Genet,* 10 (4), 241-51.
Goswami, D. and Conway, G. S. (2005), 'Premature ovarian failure', *Hum Reprod Update,* 11 (4), 391-410.
Gu, W., Zhang, F., and Lupski, J. R. (2008), 'Mechanisms for human genomic rearrangements', *Pathogenetics,* 1 (1), 4.
Handel, M. A. and Schimenti, J. C. (2010), 'Genetics of mammalian meiosis: regulation, dynamics and impact on fertility', *Nat Rev Genet,* 11 (2), 124-36.
Hanna, I. H., et al. (2000), 'Cytochrome P450 1B1 (CYP1B1) pharmacogenetics: association of polymorphisms with functional differences in estrogen hydroxylation activity', *Cancer Res,* 60 (13), 3440-4.
Hassold, T. (1982), 'Mosaic trisomies in human spontaneous abortions', *Hum Genet,* 61 (1), 31-5.
Hassold, T. and Hunt, P. (2001), 'To err (meiotically) is human: the genesis of human aneuploidy', *Nat Rev Genet,* 2 (4), 280-91.
Hassold, T. J. and Jacobs, P. A. (1984), 'Trisomy in man', *Annu Rev Genet,* 18, 69-97.
Hastings, P. J., Ira, G., and Lupski, J. R. (2009a), 'A microhomology-mediated break-induced replication model for the origin of human copy number variation', *PLoS Genet,* 5 (1), e1000327.
Hastings, P. J., et al. (2009b), 'Mechanisms of change in gene copy number', *Nat Rev Genet,* 10 (8), 551-64.

Hefler, L. A., et al. (2005), 'Estrogen-metabolizing gene polymorphisms and age at natural menopause in Caucasian women', *Hum Reprod,* 20 (5), 1422-7.
Hersmus, R., et al. (2012), 'SRY mutation analysis by next generation (deep) sequencing in a cohort of chromosomal Disorders of Sex Development (DSD) patients with a mosaic karyotype', *BMC Med Genet,* 13, 108.
Hook, E. B. (1977), 'Exclusion of chromosomal mosaicism: tables of 90%, 95% and 99% confidence limits and comments on use', *Am J Hum Genet,* 29 (1), 94-7.
Hsu, L. Y., et al. (1992), 'Proposed guidelines for diagnosis of chromosome mosaicism in amniocytes based on data derived from chromosome mosaicism and pseudomosaicism studies', *Prenat Diagn,* 12 (7), 555-73.
Inoue, K. and Lupski, J. R. (2002), 'Molecular mechanisms for genomic disorders', *Annu Rev Genomics Hum Genet,* 3, 199-242.
Jacobs, K. B., et al. (2012), 'Detectable clonal mosaicism and its relationship to aging and cancer', *Nat Genet,* 44 (6), 651-8.
Jacobs, P., et al. (1997), 'Turner syndrome: a cytogenetic and molecular study', *Ann Hum Genet,* 61 (Pt 6), 471-83.
Kallioniemi, O. P., et al. (1993), 'Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors', *Semin Cancer Biol,* 4 (1), 41-6.
Khan, F., Agarwal, A., and Agrawal, S. (2004), 'Significance of chimerism in hematopoietic stem cell transplantation: new variations on an old theme', *Bone Marrow Transplant,* 34 (1), 1-12.
Kotzot, D. (2008), 'Complex and segmental uniparental disomy updated', *J Med Genet,* 45 (9), 545-56.
Ku, C. S., et al. (2010), 'The discovery of human genetic variations and their use as disease markers: past, present and future', *J Hum Genet,* 55 (7), 403-15.
Lapunzina, P. and Monk, D. (2011), 'The consequences of uniparental disomy and copy number neutral loss-of-heterozygosity during human development and cancer', *Biol Cell,* 103 (7), 303-17.
Laurie, C. C., et al. (2012), 'Detectable clonal mosaicism from birth to old age and its relationship to cancer', *Nat Genet,* 44 (6), 642-50.
Lee, C. and Scherer, S. W. (2010), 'The clinical context of copy number variation in the human genome', *Expert Rev Mol Med,* 12, e8.
Levsky, J. M. and Singer, R. H. (2003), 'Fluorescence in situ hybridization: past, present and future', *J Cell Sci,* 116 (Pt 14), 2833-8.
Liehr, T. (2010), 'Cytogenetic contribution to uniparental disomy (UPD)', *Mol Cytogenet,* 3, 8.
Liu, P., et al. (2012), 'Mechanisms for recurrent and complex human genomic rearrangements', *Curr Opin Genet Dev,* 22 (3), 211-20.
Long, J. R., et al. (2006), 'Polymorphisms of the CYP1B1 gene may be associated with the onset of natural menopause in Chinese women', *Maturitas,* 55 (3), 238-46.
Lu, X. Y., et al. (2008), 'Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis', *Pediatrics,* 122 (6), 1310-8.
Lupski, J. R. (2013), 'Genetics. Genome mosaicism--one human, multiple genomes', *Science,* 341 (6144), 358-9.
Machiela, M. J., et al. (2015), 'Characterization of large structural genetic mosaicism in human autosomes', *Am J Hum Genet,* 96 (3), 487-97.
Makishima, H. and Maciejewski, J. P. (2011), 'Pathogenesis and consequences of uniparental disomy in cancer', *Clin Cancer Res,* 17 (12), 3913-23.
Mamsen, L. S., et al. (2012), 'The migration and loss of human primordial germ stem cells from the hind gut epithelium towards the gonadal ridge', *Int J Dev Biol,* 56 (10-12), 771-8.
Marston, A. L. and Amon, A. (2004), 'Meiosis: cell-cycle controls shuffle and deal', *Nat Rev Mol Cell Biol,* 5 (12), 983-97.

McVey, M. and Lee, S. E. (2008), 'MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings', *Trends Genet,* 24 (11), 529-38.
Menten, B., et al. (2006), 'Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports', *J Med Genet,* 43 (8), 625-33.
Myrskylä, M., Kohler, H. P., and Billari, F. C. (2009), 'Advances in development reverse fertility declines', *Nature,* 460 (7256), 741-3.
Nelson, S. M., et al. (2009), 'Anti-Müllerian hormone-based approach to controlled ovarian stimulation for assisted conception', *Hum Reprod,* 24 (4), 867-75.
Nikolaou, D. and Templeton, A. (2004), 'Early ovarian ageing', *Eur J Obstet Gynecol Reprod Biol,* 113 (2), 126-33.
Nishida, C. R., Everett, S., and Ortiz de Montellano, P. R. (2013), 'Specificity determinants of CYP1B1 estradiol hydroxylation', *Mol Pharmacol,* 84 (3), 451-8.
Oostlander, A. E., Meijer, G. A., and Ylstra, B. (2004), 'Microarray-based comparative genomic hybridization and its applications in human genetics', *Clin Genet,* 66 (6), 488-95.
Pang, A. W., et al. (2010), 'Towards a comprehensive structural variation map of an individual human genome', *Genome Biol,* 11 (5), R52.
Peiffer, D. A., et al. (2006), 'High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping', *Genome Res,* 16 (9), 1136-48.
Pham, J., et al. (2014), 'Somatic mosaicism detected by exon-targeted, high-resolution aCGH in 10,362 consecutive cases', *Eur J Hum Genet,* 22 (8), 969-78.
Pinto, D., et al. (2011), 'Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants', *Nat Biotechnol,* 29 (6), 512-20.
Piotrowski, A., et al. (2008), 'Somatic mosaicism for copy number variation in differentiated human tissues', *Hum Mutat,* 29 (9), 1118-24.
Qin, Y., et al. (2015), 'Genetics of primary ovarian insufficiency: new developments and opportunities', *Hum Reprod Update,* 21 (6), 787-808.
Rao, E., et al. (1997), 'Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome', *Nat Genet,* 16 (1), 54-63.
Razzaghian, H. R., et al. (2010), 'Somatic mosaicism for chromosome X and Y aneuploidies in monozygotic twins heterozygous for sickle cell disease mutation', *Am J Med Genet A,* 152A (10), 2595-8.
Robberecht, C., et al. (2013), 'Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations', *Genome Res,* 23 (3), 411-8.
Rodríguez-Santiago, B., et al. (2010), 'Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome', *Am J Hum Genet,* 87 (1), 129-38.
Romanelli, V., et al. (2011), 'Beckwith-Wiedemann syndrome and uniparental disomy 11p: fine mapping of the recombination breakpoints and evaluation of several techniques', *Eur J Hum Genet,* 19 (4), 416-21.
Sawaya, S., et al. (2013), 'Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements', *PLoS One,* 8 (2), e54710.
Schlessinger, D., et al. (2002), 'Genes and translocations involved in POF', *Am J Med Genet,* 111 (3), 328-33.
Score, J. and Cross, N. C. (2012), 'Acquired uniparental disomy in myeloproliferative neoplasms', *Hematol Oncol Clin North Am,* 26 (5), 981-91.
Shaffer, L. G., et al. (2001), 'American College of Medical Genetics statement of diagnostic testing for uniparental disomy', *Genet Med,* 3 (3), 206-11.
Sharp, A. J., Cheng, Z., and Eichler, E. E. (2006), 'Structural variation of the human genome', *Annu Rev Genomics Hum Genet,* 7, 407-42.

Shuster, L. T., et al. (2010), 'Premature menopause or early menopause: long-term health consequences', *Maturitas,* 65 (2), 161-6.
Strachan, T. and Read, A. P. (2011), '*Human Molecular Genetics*', Fourth edition, *Garland Science, Taylor & Francis Group*
Stranger, B. E., et al. (2007), 'Relative impact of nucleotide and copy number variation on gene expression phenotypes', *Science,* 315 (5813), 848-53.
Sudmant, P. H., et al. (2015), 'Global diversity, population stratification, and selection of human copy-number variation', *Science,* 349 (6253), aab3761.
Sullivan, A. K., et al. (2005), 'Association of FMR1 repeat size with ovarian dysfunction', *Hum Reprod,* 20 (2), 402-12.
Zamani Esteki, M., et al. (2015), 'Concurrent whole-genome haplotyping and copy-number profiling of single cells', *Am J Hum Genet,* 96 (6), 894-912.
Zarrei, M., et al. (2015), 'A copy number variation map of the human genome', *Nat Rev Genet,* 16 (3), 172-83.
Zhang, F., et al. (2009), 'Copy number variation in human health, disease, and evolution', *Annu Rev Genomics Hum Genet,* 10, 451-81.
Žilina, O., et al. (2015), 'Somatic mosaicism for copy-neutral loss of heterozygosity and DNA copy number variations in the human genome', *BMC Genomics,* 16 (1), 703.
Tsuchiya, Y., Nakajima, M., and Yokoi, T. (2005), 'Cytochrome P450-mediated metabolism of estrogens and its regulation in human', *Cancer Lett,* 227 (2), 115-24.
Tšuiko, O., et al. (2016), 'Copy number variation analysis detects novel candidate genes involved in follicular growth and oocyte maturation in a cohort of premature ovarian failure cases', *Hum Reprod,* accepted for publication
Tuna, M., Knuutila, S., and Mills, G. B. (2009), 'Uniparental disomy in cancer', *Trends Mol Med,* 15 (3), 120-8.
Tuna, M., et al. (2015), 'Prognostic relevance of acquired uniparental disomy in serous ovarian cancer', *Mol Cancer,* 14, 29.
Walsh, C. S., et al. (2008), 'Genome-wide loss of heterozygosity and uniparental disomy in BRCA1/2-associated ovarian carcinomas', *Clin Cancer Res,* 14 (23), 7645-51.
Vandeweyer, G. and Kooy, R. F. (2013), 'Detection and interpretation of genomic structural variation in health and disease', *Expert Rev Mol Diagn,* 13 (1), 61-82.
Wang, K., et al. (2007), 'PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data', *Genome Res,* 17 (11), 1665-74.
Vanneste, E., et al. (2009), 'Chromosome instability is common in human cleavage-stage embryos', *Nat Med,* 15 (5), 577-83.
Vaskivuo, T. E. and Tapanainen, J. S. (2003), 'Apoptosis in the human ovary', *Reprod Biomed Online,* 6 (1), 24-35.
Vattathil, S. and Scheet, P. (2016), 'Extensive Hidden Genomic Mosaicism Revealed in Normal Tissue', *Am J Hum Genet,* 98 (3), 571-8.
Vegetti, W., et al. (1998), 'Inheritance in idiopathic premature ovarian failure: analysis of 71 cases', *Hum Reprod,* 13 (7), 1796-800.
Vulto-van Silfhout, A. T., et al. (2013), 'Clinical significance of de novo and inherited copy-number variation', *Hum Mutat,* 34 (12), 1679-87.
Yang, Z., et al. (2008), 'Extra centrosomes and/or chromosomes prolong mitosis in human cells', *Nat Cell Biol,* 10 (6), 748-51.
Yunis, E. J., Zuniga, J., and Romero, V. (2007), 'Chimerism and tetragametic chimerism in humans: implications in autoimmunity, allorecognition and tolerance', *Immunol Res,* 38 (1-3), 213-36.
Yunis, J. J., Sawyer, J. R., and Ball, D. W. (1978), 'The characterization of high-resolution G-banded chromosomes of man', *Chromosoma,* 67 (4), 293-307.

WEB PAGES LIST

[1] Eurostat database: http://ec.europa.eu/eurostat/

[2] UCSC Genome Browser database: http://genome.ucsc.edu/

[3] OMIM (Online Mendelian Inheritance in Man®): http://www.omim.org/

SUPPLEMENTS

Supplement 1

Overview of microarray studies investigating detectable clonal mosaicism.

Reference	Platform	Variation type	Tissue	Number of samples	Number of mosaic anomalies	Frequency (%)	Subjects
(Conlin et al. 2010)	SNP microarray	CNV, LOH	blood, fibroblasts	2019	23 (+1 chimera)	1.1	childrens with developmental delay
(Rodríguez-Santiago et al. 2010)	SNP microarray	CNV, UPD	blood, buccal	1991	34 (+8 mosaic trisomy)	1.7 (autosomes)	bladder cancer cases
(Jacobs et al. 2012)	SNP microarray	CNV, CN-LOH	blood, buccal	57 853	517	0.89	31,717 cancer cases (GWAS)
(Laurie et al. 2012)	SNP microarray	CNV, CN-LOH	blood, buccal	50 222	514	<0.5% (0-50 y); 2-3% (>50 y)	GWAS subjects
(Pham et al. 2014)	aCGH	CNV	blood	10 362	57	0.55	Patients with cognitive impairment
(Machiela et al. 2015)	SNP microarray	CNV, UPD	blood, buccal	127 179	1315	0.73 (autosomes)	GWAS subjects
(Vattathil and Scheet 2016)	SNP microarray	CNV, CN-LOH	blood, buccal	31 100	1087	3.5	GWAS subjects

Supplement 2

Expected B-allele frequencies and respective degree of mosaicism for deletion/duplication due to mitotic nondisjunction.

				Percentage of abnormal cell line										
Cell line 1	BAF	Cell line 2	BAF	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	
a	0	aaa	0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
a	0	aab	0,333	0,000	0,024	0,048	0,070	0,091	0,111	0,130	0,149	0,167	0,184	
a	0	abb	0,333	0,000	0,024	0,048	0,070	0,091	0,111	0,130	0,149	0,167	0,184	
b	1	baa	0,333	1,000	0,951	0,905	0,860	0,818	0,778	0,739	0,702	0,667	0,633	
b	1	bba	0,333	1,000	0,951	0,905	0,860	0,818	0,778	0,739	0,702	0,667	0,633	
b	1	bbb	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	
Cell line 1	BAF	Cell line 2	BAF	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
a	0	aaa	0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
a	0	aab	0,333	0,200	0,216	0,231	0,245	0,259	0,273	0,286	0,298	0,310	0,322	0,333
a	0	abb	0,333	0,200	0,216	0,231	0,245	0,259	0,273	0,286	0,298	0,310	0,322	0,333
b	1	baa	0,333	0,600	0,569	0,538	0,509	0,481	0,455	0,429	0,404	0,379	0,356	0,333
b	1	bba	0,333	0,600	0,569	0,538	0,509	0,481	0,455	0,429	0,404	0,379	0,356	0,333
b	1	bbb	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

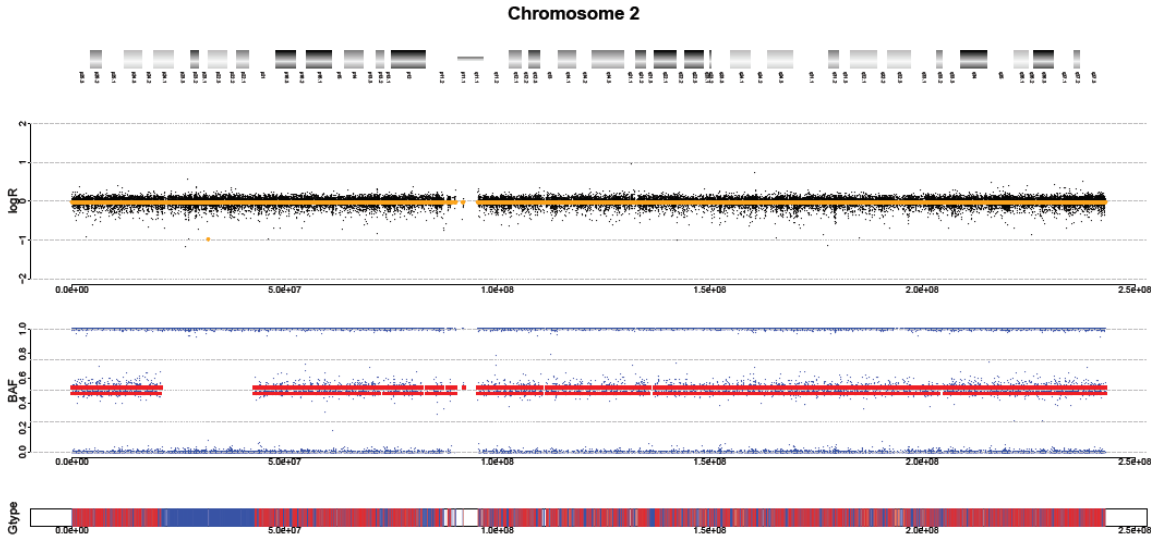
Supplement 3

Expected B-allele frequencies and respective degree of mosaicism for segmental UPD.

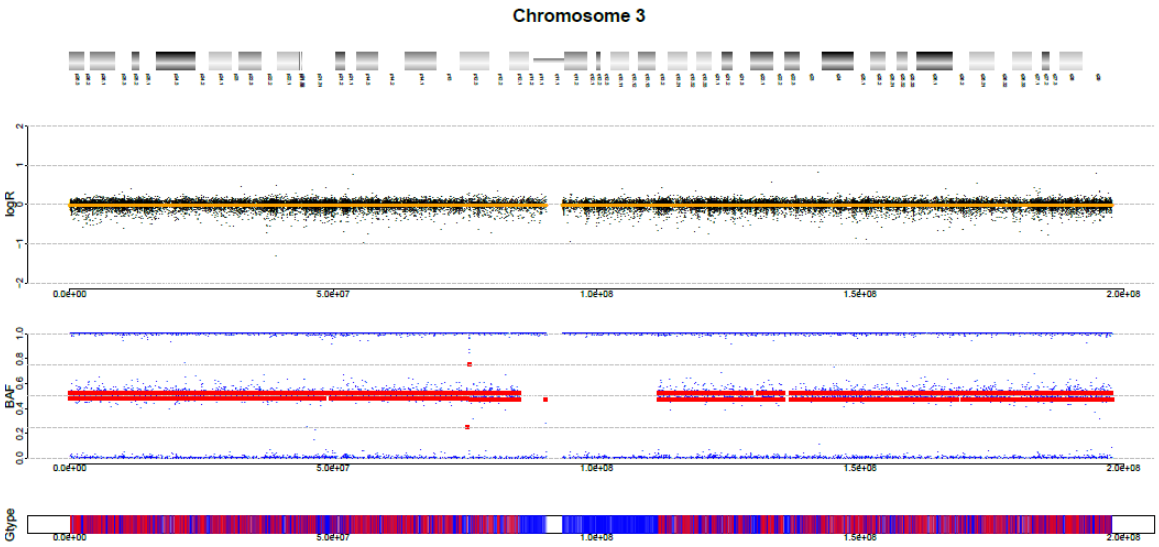
				Percentage of abnormal cell line										
Cell line 1	BAF	Cell line 2	BAF	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	
aa	0	aa	0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
ab	0,5	aa	0	0,500	0,475	0,450	0,425	0,400	0,375	0,350	0,325	0,300	0,275	
ab	0,5	bb	1	0,500	0,525	0,550	0,575	0,600	0,625	0,650	0,675	0,700	0,725	
bb	1	bb	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	
Cell line 1	BAF	Cell line 2	BAF	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
aa	0	aa	0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
ab	0,5	aa	0	0,250	0,225	0,200	0,175	0,150	0,125	0,100	0,075	0,050	0,025	0,000
ab	0,5	bb	1	0,750	0,775	0,800	0,825	0,850	0,875	0,900	0,925	0,950	0,975	1,000
bb	1	bb	1	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Supplement 4

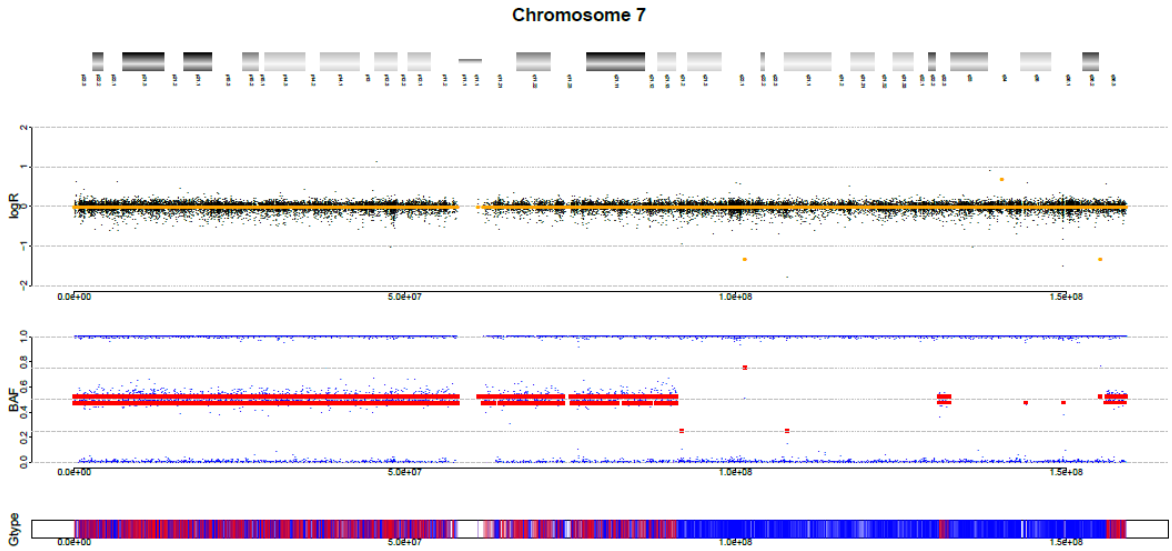
Figures of CN-LOH regions with approximate sizes from different chromosomes (chr 2, 3, 7, 17 and 22) in one of the patients.



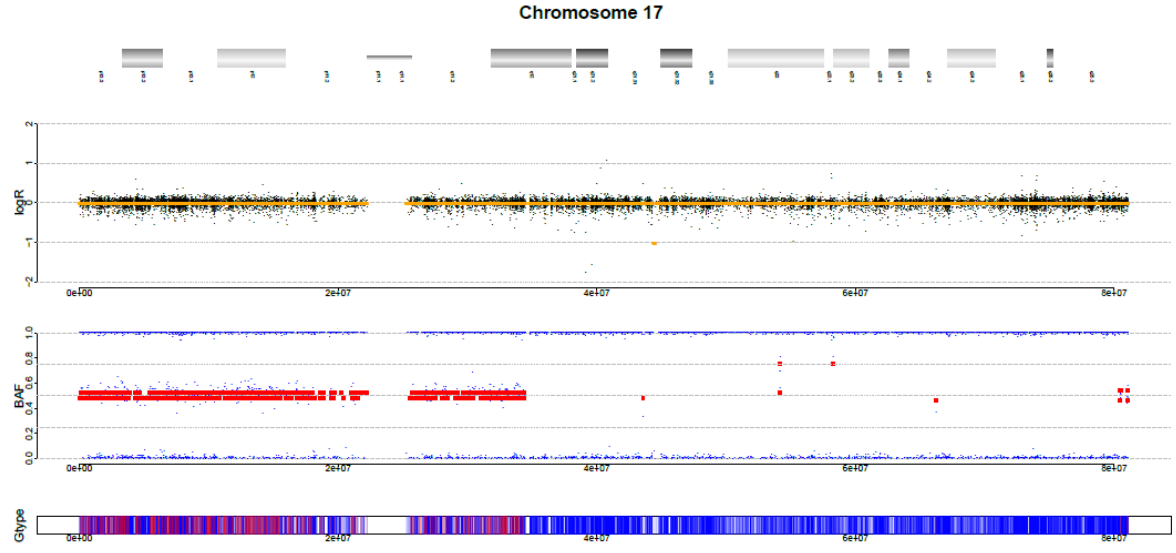
CN-LOH in 2p24.1-p21 (approximate size 20.5 Mb).



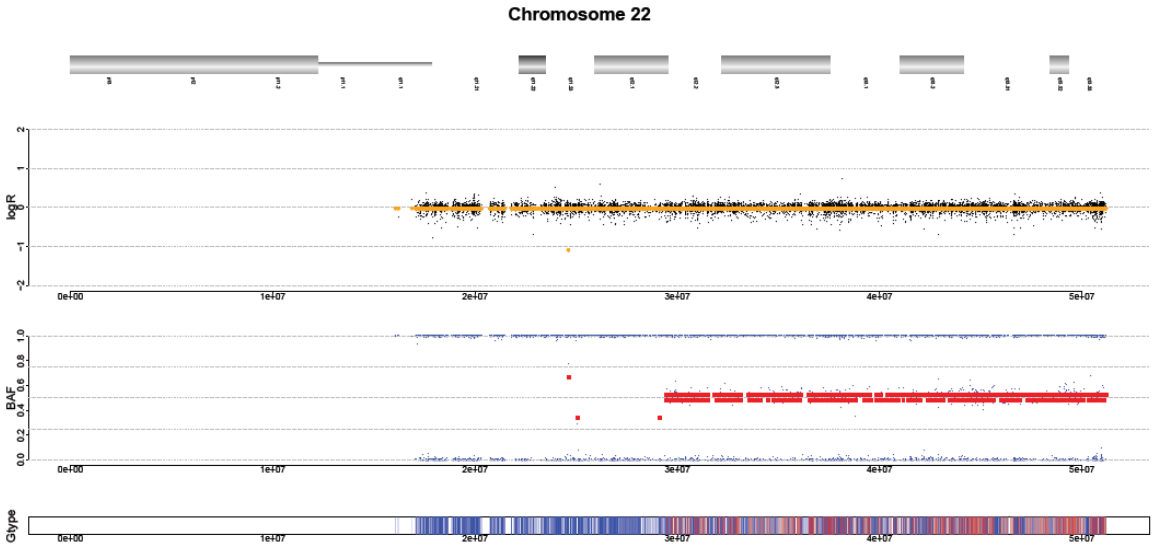
CN-LOH in 3p12.1-q13.2 (approximate size 28.5 Mb).



CN-LOH in 7q21.2-q36.2 (approximate size 62.5 Mb).



CN-LOH on 17q12-q25.3 (approximate size 46.7 Mb).



CN-LOH in 22q11.21-q12.1 (approximate size 10.8 Mb).

Non-exclusive licence to reproduce thesis and make thesis public

I, Katre Teearu

(date of birth: 21.11.1989),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the university's web environment, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Genetic abnormalities in premature ovarian failure patients,

supervised by Prof. Ants Kurg, PhD and Dr. Masoud Zamani Esteki, PhD

2. I am aware of the fact that the author retains these rights.

3. This is to certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 27.05.2016