LIINA NAGIRNAJA

Global and fine-scale genetic determinants of recurrent pregnancy loss

Tartu 2014

ISSN 1024–6479
LIINA NAGIRNAJA

Global and fine-scale genetic determinants of recurrent pregnancy loss
Institute of Molecular and Cell Biology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in molecular and cell biology) on March 19, 2014 by the Council of the Institute of Molecular and Cell Biology, University of Tartu.

Supervisor: Prof. Maris Laan, PhD
Institute of Molecular and Cell Biology
University of Tartu
Estonia

Opponent: Dr. Daniel Vaiman, PhD
Cochin Institute
INSERM
Paris Descartes University
France

Commencement: Room No 105, 23B Riia St, Tartu, on May 13th, 2014, at 10.00

The publication of this thesis is granted by the Institute of Molecular and Cell Biology, University of Tartu.

ISSN 1024–6479

Copyright: Liina Nagirnaja, 2014

University of Tartu Press
www.tyk.ee
# TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS ................................................................. 7  
LIST OF ABBREVIATIONS ............................................................................. 9  
INTRODUCTION .......................................................................................... 11  

1. REVIEW OF LITERATURE........................................................................ 13  
   1.1. Dynamics and challenges of human pregnancy............................... 13  
      1.1.1. Prerequisites of early pregnancy success ................................. 13  
      1.1.2. Recurrent early pregnancy loss ............................................. 15  
         1.1.2.1. The heterogeneity of RM etiology ............................... 16  
         1.1.2.2. Defining the genetic component of RM (also reviewed in Ref. I) ......................................................... 17  
         1.1.2.3. Hypothesis-free ‘omics’ approach in screening for novel RM risk factors ........................................ 19  
   1.2. Human chorionic gonadotropin, the ‘pregnancy hormone’ ............ 21  
      1.2.1. Structural characteristics of hCG ............................................ 21  
      1.2.2. Genomic context and evolution of hCGβ genes (also reviewed in Ref. II) ......................................................... 22  
      1.2.3. hCGβ genes as candidate risk loci in pregnancy failure .......... 24  
   1.3. DNA copy number variation ............................................................. 24  
      1.3.1. CNVs in common diseases ...................................................... 25  
         1.3.1.1. Hypothesis-free global genomic profiling of CNVs ... 25  
         1.3.1.2. Targeted analysis of copy number variable genes ....... 26  
      1.3.2. CNVs in pregnancy complications .......................................... 27  

2. AIMS OF THE PRESENT STUDY ............................................................... 30  

3. RESULTS ................................................................................................. 31  
   3.1. Genome dynamics of human LHB/CGB gene cluster (Ref. II, III) .... 31  
      3.1.1. Genomic structure of the LHB/CGB gene cluster ............... 31  
      3.1.2. Polymorphism density and gene conversion activity .......... 32  
      3.1.3. Recombination rate and structural analysis of potential recombination ‘hot spot’ ......................................................... 34  
   3.2. DNA single nucleotide variants of CGB5 and CGB8 in the context of recurrent miscarriage .................................................. 35  
      3.2.1. SNP profile of the CGB5 and CGB8 genes in recurrent miscarriage (Ref. IV, V) ................................................ 35  
         3.2.1.1. Screening for variants in Northern Europe.................... 35  
         3.2.1.2. Disease-related variants and haplotypes in CGB8 ....... 35  
         3.2.1.3. Disease-related variants and haplotypes in CGB5 ..... 38  
         3.2.1.4. Rare non-synonymous mutations in CGB5 and CGB8 ................................................................. 39
3.2.2. Structural and functional impact of non-synonymous mutations in CGB5 and CGB8 (Ref. V) .................................. 39


3.2.2.2. Structural features and assembly of recombinant hCGβ isoforms .................................................. 41

3.2.2.3. Bioactivity of recombinant hCGβ isoforms ............ 43

3.3. DNA copy number variants in recurrent miscarriage (Ref. VII) .... 44

3.3.1. Genome-wide profile of CNVs ........................................ 44

3.3.2. Functional enrichment of genes disrupted by CNVs .......... 47

3.3.3. Identification of novel common CNV regions conferring risk to recurrent miscarriage ................. 47

3.3.3.1. Prioritized CNVRs affecting RM in Estonia and Denmark ......................................................... 47

3.3.3.2. Genomic context and fine-mapping of the PDZD2:GOLPH3 duplication .................................. 48

3.3.3.3. Expression profile of PDZD2 and GOLPH3 .......... 49

4. DISCUSSION ........................................................................................... 51

4.1. Nature and impact of fine-scale genetic variation of the LHB/CGB region in RM ........................................ 51

4.2. Genome-wide effect of CNVs in RM .................................. 53

4.2.1. Genomic CNV burden as risk factor for RM ................. 53

4.2.2. Genetic variability of immunoregulatory pathways as risk factor for RM ........................................ 53

4.3. Global genomic analysis as the source of novel biomarkers .... 54

4.4. Status quo and future perspectives in assessing genetic determinants of RM .................................................. 55

SUMMARY AND CONCLUSIONS ............................................................. 57

REFERENCES .......................................................................................... 59

SUMMARY IN ESTONIAN ................................................................. 73

ACKNOWLEDGEMENTS ........................................................................ 77

PUBLICATIONS .................................................................................... 79

CURRICULUM VITAE ............................................................................. 193
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles referred to in the text by their Roman numerals:


Articles are reprinted with the permission of copyright owners.
Author’s contributions:

Ref. I, II – contributed to the preparation of the review article

Ref. III, IV – participated in experimental design, experimental data collection, analysis and interpretation and contributed to manuscript preparation

Ref. V – contributed to in silico data analysis, interpretation and manuscript preparation

Ref. VI – contributed to experimental design, conducted the experiments, participated in data analysis and interpretation and wrote the first draft of the manuscript

Ref. VII – contributed to experimental design, participated in data analysis, experimental data collection and interpretation and wrote the first draft of the manuscript
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp</td>
<td>basepair</td>
</tr>
<tr>
<td>C4A</td>
<td>complement component 4A gene</td>
</tr>
<tr>
<td>C4B</td>
<td>complement component 4B gene</td>
</tr>
<tr>
<td>cAMP</td>
<td>cyclic adenosine monophosphate</td>
</tr>
<tr>
<td>CGβ</td>
<td>chorionic gonadotropin β-subunit</td>
</tr>
<tr>
<td>CGB</td>
<td>chorionic gonadotropin, beta polypeptide gene</td>
</tr>
<tr>
<td>CHO</td>
<td>chinese hamster ovary cell line</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variant</td>
</tr>
<tr>
<td>CNVR</td>
<td>copy number variable region</td>
</tr>
<tr>
<td>CRE</td>
<td>cAMP response element</td>
</tr>
<tr>
<td>CTNNA3</td>
<td>catenin (cadherin-associated protein), alpha 3 gene</td>
</tr>
<tr>
<td>DGV</td>
<td>Database of Genomic Variants</td>
</tr>
<tr>
<td>DKK2</td>
<td>dickkopf 2 homolog gene</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>EGCUT</td>
<td>Estonian population cohort from Estonian Biobank, Estonian Genome Center, University of Tartu</td>
</tr>
<tr>
<td>FSH</td>
<td>Follicle-stimulating hormone</td>
</tr>
<tr>
<td>Golgi</td>
<td>Golgi phosphoprotein 3 gene</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>hCG</td>
<td>human chorionic gonadotropin</td>
</tr>
<tr>
<td>hCGα</td>
<td>human chorionic gonadotropin α-subunit</td>
</tr>
<tr>
<td>hCGβ</td>
<td>human chorionic gonadotropin β-subunit</td>
</tr>
<tr>
<td>hCG-h</td>
<td>hyperglycosylated human chorionic gonadotropin</td>
</tr>
<tr>
<td>HEK293</td>
<td>human embryonic kidney 293 cell line</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>hLH/CGR</td>
<td>human luteinizing hormone/chorionic gonadotropin receptor</td>
</tr>
<tr>
<td>HPRT</td>
<td>hypoxanthine phosphoribosyltransferase gene</td>
</tr>
<tr>
<td>IGF2</td>
<td>insulin-like growth factor 2 gene</td>
</tr>
<tr>
<td>IGKV</td>
<td>Immunoglobulin kappa variable gene cluster</td>
</tr>
<tr>
<td>IVF</td>
<td>in vitro fertilization</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>KIR</td>
<td>killer immunoglobulin-like receptors</td>
</tr>
<tr>
<td>KRK</td>
<td>korduv raseduse katkemine</td>
</tr>
<tr>
<td>LD</td>
<td>linkage disequilibrium</td>
</tr>
<tr>
<td>LH</td>
<td>luteinizing hormone (lutropin)</td>
</tr>
<tr>
<td>LHB</td>
<td>luteinizing hormone beta polypeptide gene</td>
</tr>
<tr>
<td>MAF</td>
<td>minor allele frequency</td>
</tr>
<tr>
<td>Mb</td>
<td>megabase</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>mTOR</td>
<td>mechanistic target of rapamycin</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>NK</td>
<td>natural killer cells</td>
</tr>
<tr>
<td>NTF5</td>
<td>neurotrophin 5 gene, currently known as NTF4</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>PDZD2</td>
<td>PDZ domain containing 2 gene</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>RM</td>
<td>recurrent miscarriage</td>
</tr>
<tr>
<td>S100A8</td>
<td>inflammatory marker calprotectin gene</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of the mean</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>TIMP2</td>
<td>TIMP metallopeptidase inhibitor 2 gene</td>
</tr>
<tr>
<td>TNF</td>
<td>tumor necrosis factor</td>
</tr>
<tr>
<td>TRAIL</td>
<td>TNF-related apoptosis-inducing ligand gene</td>
</tr>
<tr>
<td>WT</td>
<td>wild type</td>
</tr>
</tbody>
</table>
INTRODUCTION

Sporadic miscarriage is the most common pregnancy complication affecting up to 15% of all clinically recognized pregnancies (Stirrat 1990). Up to 3% of fertile couples have been diagnosed with recurrent miscarriage disease (RM) defined by three or more consecutive early pregnancy losses before gestational week 22. RM is a multifactorial disease accompanied by increased probability of other pregnancy complications such as preterm birth or small for gestational age newborns (Jivraj et al. 2001; van Oppenraaij et al. 2009). Although a spectrum of causes is known for RM, etiology behind approximately half of the cases remains unsolved and is defined as idiopathic (reviewed in (Christiansen et al. 2008)).

The genetic component in the etiology of idiopathic RM has been acknowledged and extensively addressed by numerous studies targeting almost 100 candidate genes to date (Kolte et al. 2011; Rull et al. 2012). However, no prevalent gene variants specific to the disease have been described sufficiently explaining the heritability of RM. Few hypothesis-free ‘omics’ studies, including genome-wide approach or transcriptomics of endometrial and placental samples (Lee et al. 2007; Li et al. 2010; Rull et al. 2013), have been reported identifying novel candidate genes that modify the risk of having RM and highlighting the potential of ‘omics’ research as the source of novel genetic biomarkers.

Human chorionic gonadotropin (hCG), also termed as the ‘pregnancy hormone’, is essential for pregnancy establishment and maintenance. The genes encoding the β-subunit of the heterodimeric hCG are attractive candidates in studying early pregnancy success as the aberrant expression has previously been implicated in various pregnancy disorders, including RM (Rull and Laan 2005). The four highly similar genes encoding hCGβ (CGB) are co-located in the LHB/CGB gene cluster on chromosome 19. The complex nature of the LHB/CGB region has hindered the fine-scale analysis of the polymorphisms within the hCGβ coding genes and their effect on early pregnancy success until now.

DNA copy number variants (CNVs) involving DNA rearrangements of >50 bp in size have been proposed to explain heritability of various complex diseases due to large coverage of the human genome (up to 30%) (Redon et al. 2006; Mills et al. 2011). Only few studies have addressed the contribution of CNVs in the manifestation of reproductive disorders, including a single CNV screening focusing on placental samples of RM cases (Rajcan-Separovic et al. 2010). Genome-wide profiling of CNVs among parents affected by the disease may prove as a powerful tool in mapping novel genetic biomarkers affecting the predisposition to RM independently but also in consort.

In this thesis, the review of literature provides up to date information on the current knowledge of risk factors and genetic association studies of RM performed at fine-scale and global level. A comprehensive overview of the
structural and functional characteristics of hCG, its β-subunit and hCGβ coding genes is provided and the contribution of CNVs in common diseases and pregnancy complications is reviewed.

The experimental part of the thesis aimed to map the fine-scale and global determinants of genetic susceptibility to RM setting the genetic diversity of hCGβ genes and global CNV profile as the examples. The research focused on firstly, the local genomic landscape of LHB/CGB gene cluster and evolutionary factors shaping the diversity of CGB genes in the context of RM. Fine-scale analysis of polymorphisms in two major hCGβ genes CGB5 and CGB8 as promising RM candidate genes was performed in three North European populations and the structural and functional effect of three non-synonymous mutations identified within the genes was addressed. Secondly, the impact of CNVs on the manifestation of RM was elucidated at the genome-wide scale and novel independent genetic biomarkers of RM were inferred from the global CNV profile. An improved map of genetic determinants and current status of research on the etiology of RM is discussed in the light of the findings provided in this thesis.
1. REVIEW OF LITERATURE

1.1. Dynamics and challenges of human pregnancy

Human pregnancy is a unique intricate system that involves maternal acceptance of an organism foreign by the molecular and genetic build, establishment of a transient invasive organ placenta and a complex interplay of maternal and fetal signals balancing between the maternal needs and fulfilling the fetal requirements. Well-balanced communication and timely progression of early stages of pregnancy are crucial for achieving a normal full term birth.

Hemochorial placentation defined by the direct contact of fetal chorion with maternal bloodstream is characteristic to humans, higher primates, rabbit and rodents, however the advanced invasiveness of placentation is unique to humans (Enders and Carter 2004). Following the implantation of an embryo, the extraembryonic trophoblast cells proliferate and form syncytiotrophoblast cells covering the villous trees and give rise to non-proliferative extravillous trophoblast cells that adopt an invasive phenotype and induce remodeling of maternal uterine and vascular tissues (Norwitz et al. 2001; Red-Horse et al. 2004). As the pregnancy progresses, the increasing surface of the branching placental chorionic villi is in direct contact with maternal blood providing access to nutritional, gas and waste exchange in response to the growing needs of the developing fetus. Successfully formed maternal-fetal interface is able to adapt to the changing environment and effectively resolve the challenges of pregnancy. Inadequate implantation, shallow placentation or reaction to stress factors (e.g. oxidative stress) may give rise to infertility or pregnancy complications several of which are largely human-specific, such as preeclampsia and recurrent miscarriages (Norwitz et al. 2001; Jauniaux et al. 2006). Some disorders may persist even after parturition and have an adverse impact on the long term maternal health by invoking cardiovascular and autoimmune diseases, metabolic syndrome and early mortality (Clifton et al. 2012).

1.1.1. Prerequisites of early pregnancy success

Balance and correct timing of the multitude of pathways regulating the first post-conception weeks are essential for successful establishment and subsequent course of pregnancy. Low chance of natural conception per menstrual cycle (30%) has been observed in humans with 60% of pregnancy losses occurring already prior to clinical diagnosis of pregnancy highlighting the selective nature of the implantation and placentation stages (Zinaman et al. 1996; Macklon et al. 2002).

The timely progression of blastocyst implantation (within approximately 7 – 10 days after ovulation) is strictly regulated by a bi-directional interplay between maternal and blastocyst signals (Teklenburg et al. 2010). The induction of endometrial receptivity window when the uterus supports implantation, promotion of blastocyst attachment and differentiation of endometrial cells into
decidua cells are mediated via a complex signaling network involving ovarian hormones, cytokines, growth factors, adhesion molecules and transcription factors (e.g. progesterone, estrogen, HB-EGF, KLF5, MSX1/2, LIF) (Norwitz et al. 2001; Cha et al. 2012). Failed or delayed implantation may lead to infertility or early pregnancy loss, respectively, but also predispose to late obstetric complications such as preeclampsia (Figure 1) (Wilcox et al. 1999; Leach et al. 2002; Cha et al. 2012). Interestingly, it has recently been noted that unexpectedly large proportion of women affected by repeated early miscarriage (41% versus predicted 8%) are not sub- but super-fertile and achieve pregnancy within only couple of cycles (Salker et al. 2010). The phenomenon has been attributed to impaired natural embryo selection, whereby uterine environment allows the implantation of unviable embryos (Teklenburg et al. 2010).

The final outcome of human pregnancy largely depends on the invasion of cytotrophoblast cells to an adequate depth in the uterus accompanied by remodeling of uterine spiral arteries and transformation into low-resistance vascular network (Jauniaux et al. 2006). Limited endovascular invasion has been related to suboptimal flow, subsequent hypoxia and tissue damage leading to the manifestation of preeclampsia or intrauterine growth restriction (Kaufmann et al. 2003). On the other hand, excessive placental invasion and abnormal attachment to maternal myometrium (placenta accreta/increta) that may even reach maternal organs (placenta percreta) is a serious obstetric complication with a risk of severe post-partum haemorrhage (Jauniaux and Jurkovic 2012). The advanced invasiveness characteristic to human hemochorial placentation has been correlated with the evolution of hyperglycosylated human chorionic gonadotropin (hCG-h) in primate lineage expressed by invasive extravillous trophoblasts in early pregnancy (Jauniaux et al. 2006; Cole 2009; Guibourdenche et al. 2010). Concurrently, placental insufficiency due to low expression of hCG-h has been implicated in both early and late pregnancy disorders, such as early pregnancy loss and preeclampsia (Kovalevskaya et al. 2002; Keikkala et al. 2013).

It has been shown that the largest proportion of genes upregulated in the human decidua cells in response to the paracrine signals of an implanting trophoblast are related to immune response at the feto-maternal interface (Hess et al. 2007). Maternal immune tolerance to the semi-allogeneic embryo expressing paternal antigens has been most extensively studied in normal pregnancy but also in relation to pregnancy complications (Guleria and Sayegh 2007; Redman and Sargent 2010). In this feto-maternal immune system communication, critical role has been attributed to uterine natural killer (NK) cells that constitute 40% of cells in the decidua (CD56brightCD16− NK subset) and promote trophoblast invasion and angiogenesis (Moffett-King 2002; Hanna et al. 2006). Uterine NK cells express an array of killer immunoglobulin-like receptors (KIR) that mediate the trophoblast recognition via binding MHC class I molecules (HLA-E, HLA-C and HLA-G) expressed on extravillous trophoblasts. Distinct KIR receptor-ligand combinations have been associated
with increased risk of early pregnancy loss and preeclampsia (Hiby et al. 2004; Hiby et al. 2008; Faridi and Agrawal 2011).

Upon successfully passing the selection windows at implantation and meeting all the prerequisites and challenges of a pregnancy, 30% of conceptions are estimated to reach live birth (Macklon et al. 2002).

![Figure 1](image.png)

Figure 1. Cause-and-effect scenarios in human pregnancy disorders as proposed by (Cha et al. 2012). Deviations in the timing and signaling of early pregnancy stages may lead to in-/subfertility, early pregnancy loss or give rise to late pregnancy disorders due to subsequent placental insufficiency or shallow invasion. IUGR, intrauterine growth restriction.

### 1.1.2. Recurrent early pregnancy loss

A sporadic miscarriage is the most common pregnancy complication affecting approximately quarter of all women at least once in their lifetime and up to 15% of all clinically recognized pregnancies (Wilcox et al. 1988; Stirrat 1990). In 3% of fertile couples, three or more consecutive early pregnancy losses occur before gestational week 22, defined as recurrent miscarriage disease (RM) (Christiansen et al. 2008). In patients diagnosed with RM, each subsequent
pregnancy is accompanied with an increased risk of not only miscarriage but also other obstetric complications such as preterm birth or small for gestational age newborns (Ogasawara et al. 2000; Jivraj et al. 2001; van Oppenraaij et al. 2009). It is a distressing disease for the affected couples with approximately half of the cases remaining unsolved due to the complex etiology involving a spectrum of known but also as yet unidentified risk factors (reviewed in (Christiansen et al. 2008)).

1.1.2.1. The heterogeneity of RM etiology

The major cause of sporadic early miscarriages (in up to 75% of cases) is fetal chromosome abnormalities associated with continuously increasing age of women postponing childbearing to late 30s and early 40s in western countries (Philipp et al. 2003; Group 2010). Although high maternal age is also a risk factor in RM, other causes predominantly drive this disease as the chance of having an early pregnancy loss due to large chromosomal alterations is decreasing with an increasing number of miscarriages in a couple (Figure 2) (Ogasawara et al. 2000). The etiology of RM is heterogeneous and involves interplay between maternal, paternal and cumulative (placental/fetal) risk factors in pathways related to pregnancy establishment and maintenance.

![Figure 2](image)

**Figure 2.** Rate of pregnancy success in women experiencing spontaneous miscarriages. (A) Pregnancy outcome depending on the maternal age in women with 2 or more miscarriages (adapted from (Matthiesen et al. 2012)). (B) Pregnancy success and the chance of observing fetal normal karyotype in women with 2 or more miscarriages and with average age of 31 years (based on (Ogasawara et al. 2000)).
The main factors studied in association with RM susceptibility include (reviewed in detail in (Pandey et al. 2005; Christiansen et al. 2006; Christiansen et al. 2008; Larsen et al. 2013))

- thrombophilic disturbances that contribute to thrombosis of placental vessels leading to placental infarctions
- excessive inflammatory processes causing increased apoptosis and pregnancy failure
- reproductive tract infections (e.g. HIV, malaria) triggering the inflammation of the uterine lining that disrupts embryo implantation and growth
- autoimmune (e.g. antiphospholipid syndrome) and alloimmune (couple’s immunologic mismatch) factors leading to maternal rejection of the fetus
- anatomical disorders (e.g. uterine malformations)
- endocrine dysfunction (e.g. thyroid dysfunction)
- sperm DNA fragmentation introducing deleterious DNA damage to the developing embryo
- impaired embryo selection by maternal decidua
- oxidative stress damaging feto-maternal tissues

For many of the RM risk factors studied, the association with the disease is weak and/or the underlying mechanism has remained unclear (reviewed in (Christiansen et al. 2008; Larsen et al. 2013)) highlighting the need for further research to map and describe all the risk factors of this multifactorial disease

### 1.1.2.2. Defining the genetic component of RM
(also reviewed in Ref. I)

Contribution of genetic susceptibility to the etiology of RM has been acknowledged and up to 2-fold higher occurrence of the disease has been reported within the affected families compared to control population (Christiansen et al. 1990; Kolte et al. 2011). Due to lower prevalence of large chromosomal abnormalities in miscarried embryos of RM patients compared to sporadic miscarriages (Ogasawara et al. 2000), other types of genetic variation (e.g. submicroscopic changes) likely contribute to the manifestation of the disease. Most studies addressing the genetic susceptibility to RM have used a candidate gene based approach by screening polymorphisms and mutations in genes with known functional effect on RM and/or pathways related to pregnancy establishment and maintenance. Currently, approximately 100 candidate genes have been targeted with the largest attention given to factors leading to thrombophilic, inflammatory and immunologic disturbances at the feto-maternal interface (Table 1 in Ref I).

Like other thrombophilic mutations, the most common variants G1691A in factor V (Leiden mutation) and G20210A in prothrombin, may contribute to the excessive blood coagulation and impaired placental blood circulation leading to
increased risk of early pregnancy loss in the mutation carriers (Dizon-Townson et al. 1997). Although both mutations are routinely tested in the clinical setting and have been positively associated with RM in most studies, the odds ratio (OR) ranges considerably (from 0.5 to 18) due to the heterogeneous nature of the phenotype and instead, a stronger association has been observed for late pregnancy loss (Rey et al. 2003; Kovalevsky et al. 2004; Robertson et al. 2006). Similarly, the genes regulating the balance between the action of pro-inflammatory (e.g. TNFα, IFNγ) and anti-inflammatory factors (e.g. IL6, IL10) at the feto-maternal interface have been studied extensively but have gained inconclusive results (Daher et al. 2003; Bombell and McGuire 2008). Due to weak or no impact of the studied polymorphisms, it has been proposed that instead of individual mutations, a combination and accumulation of thrombophilic and inflammation related genetic variants may shape the cumulative risk of RM (Rey et al. 2003; Jivraj et al. 2006).

As the fetal cells are expressing antigens of paternal origin, the maternal rejection of the ‘foreign’ fetus is potentially one of the major causes leading to miscarriage event (Wilczynski 2006). This condition of allograft rejection is bidirectional defined by maternal recognition and fetal presentation of paternal antigens and thus the genetic contribution of both counterparts has been addressed. For example, the allelic composition and presence of 14 bp insertion in the human leukocyte antigen HLA-G gene, the fetal genetic determinant expressed by extravillous cytotrophoblasts at the feto-maternal interface (Kovats et al. 1990), has been associated with the risk of RM, however the results are inconsistent (Aldrich et al. 2001; Hviid et al. 2004; Aruna et al. 2010; Wang et al. 2013). The cumulative effect of maternal recognition and presented fetal antigens has also been suggested as promising genetic factors describing susceptibility to RM, for example increased HLA-sharing between couples/mother-fetus or specific interaction of killer immunoglobulin-like receptors (KIR) on maternal uterine NK cells and fetal HLA-C on the trophoblast cells. However, no clear consensus has also been reached for these findings (Witt et al. 2004; Beydoun and Saftlas 2005; Hiby et al. 2010; Kolte et al. 2010; Moghraby et al. 2010).

The inconclusive or controversial results of the numerous candidate gene-based SNP studies highlight the complexity of the studied phenotype and potentially low impact of single nucleotide variants in the manifestation of RM. It should be noted that many of the studies were also underpowered due to narrow study setup (only affected mothers) and insufficient number of study subjects to detect genetic variants of small effect size and/or with low population-specific prevalence (Ref. I). Thus, increasing the sample size, addressing both the maternal and paternal contribution to the disease and/or other types of variation in association with RM would prove beneficial.
1.1.2.3. Hypothesis-free ‘omics’ approach in screening for novel RM risk factors

Unlike candidate gene based studies, the powerful tools of genome/transcriptome/proteome-wide analyses would enable the identification of novel genes and pathways essential in early pregnancy and modulating the risk to RM. In recent years, a number of studies have addressed various levels of variation shedding light on this heterogeneous phenotype.

Genomics

Two genome-wide studies have been reported searching for novel RM-associated genomic loci among the cases with unexplained RM. In a pilot study performed in unrelated Han Chinese patients and fertile controls, three microsatellite markers at 6q27, 9q33.1 and Xp22.11 were significantly associated with RM (Li et al. 2010). At 9q33.1, two potential RM candidate genes were proposed, the TNFSF8 and TNFSF15 belonging to the tumor necrosis factor (TNF) ligand family implicated in early pregnancy loss in mice (Erlebacher et al. 2004). Four additional genomic regions (3p14.2, 6q16.3, 9p22.1 and 11q13.4) that did not overlap with the results by Li et al. (2010) were revealed in a linkage study undertaken in Danish sibling pairs affected by RM (Kolte et al. 2011). No SNP-based genome-wide association studies (GWAS) extensively used for identification of the genetic component of complex diseases have been performed for RM phenotype. The informative value of GWAS in the multifactorial RM remains unclear due to expected small effect sizes of individual genetic variants as also demonstrated for other reproductive traits (Montgomery et al. 2013).

Increasing evidence suggests that genomic imprinting mediating the expression of parent-specific alleles regulates the fetus- and placenta-specific genes and pathways (Nelissen et al. 2011; Novakovic and Saffery 2012). Aberrant gene methylation profiles have been linked to various obstetric disorders such as small for gestational age, intrauterine growth restriction, preeclampsia and gestational trophoblastic disease (Xue et al. 2004; Guo et al. 2008; Diplas et al. 2009; Yuen et al. 2009) but also recurrent miscarriage. In addition to one candidate-gene based study identifying aberrant methylation of hCG β-subunit coding CGB5 gene in three cases with RM (Uuskula et al. 2010), a genome-scale methylation analysis of chromosomally normal miscarried chorionic villi from RM women has been reported (Hanna et al. 2013). Alternative global methylation patterns were observed among the RM cases when compared to normal chorionic villi from elective abortions and altered methylation levels were subsequently confirmed for distinct loci, such as AXL (receptor tyrosine kinase) and DEFB1 (defensin β 1).
Transcriptomics

A whole transcriptome analysis of maternal decidua or placental/fetal tissues may provide novel insights on pathways and specific genes with critical functions in implantation and maintenance of pregnancy. Although the expression of pre-selected genes in chorionic villi (Baek et al. 2002; Choi et al. 2003) and endometrium (Lee et al. 2007) of RM cases has been addressed, only two studies have targeted the global transcriptome profile of these tissues. Differential gene expression analysis of the decidual tissue of women suffering from RM and undergoing a miscarriage confirmed the major contribution of pathways and genes related to immune response (23% of dysregulated genes) but also cell signaling (18%) and cell invasion (17.1%) in the reoccurrence of early pregnancy loss (Krieg et al. 2012). In the miscarried placental tissue of RM cases, significant overexpression of TNF-related apoptosis-inducing ligand (TRAIL) and inflammatory marker calprotectin (S100A8) were identified (Rull et al. 2013). The increased level of TRAIL protein has also been reported for maternal serum (Agostinis et al. 2012; Rull et al. 2013) and S100A8 mRNA in the maternal decidua (Nair et al. 2013) of women affected by recurrent early pregnancy loss thus potentially representing novel biomarkers of RM disease.

Additional level of regulation in placental gene expression may be provided by short non-coding RNAs implicated in the post-transcriptional control of pathways related to male and female reproductive traits (Hawkins et al. 2011). Although not directly studied in recurrent miscarriages, differential microRNA (miRNA; ~22 nucleotides) profile has been reported for the cases of repeated implantation failure undergoing in vitro fertilization (IVF) procedures (Revel et al. 2011), indicating the potential contribution of miRNAs in the manifestation of early pregnancy loss.

Proteomics

In order to determine novel or confirm known RM risk factors at the proteome level, follicular fluid of three RM cases compared to three multiparous fertile controls was addressed with a combination of two-dimensional gel electrophoresis and mass spectrometry by (Kim et al. 2006). Aberrant expression of five proteins was observed in the follicular fluid of which the angiotensinogen, complement component C3c chain E and coagulation factors fibrinogen γ and antithrombin were subsequently also confirmed to be downregulated in the chorionic villi tissue samples thus potentially affecting embryo development or placental function. The strength of proteome analysis was further underlined when similar approach was applied for the maternal serum of RM women identifying an additional potential biomarker, the acute-phase inflammation related ITI-H4 (Kim et al. 2011).
1.2. Human chorionic gonadotropin, the ‘pregnancy hormone’

One of the key factors essential in implantation and early pregnancy maintenance is the placental human chorionic gonadotropin also termed as the ‘pregnancy hormone’. hCG is produced already by an 8-cell blastocyst prior to implantation (Lopata and Hay 1989) and upon reaching mother’s circulation it is used as an early pregnancy biomarker in conventional pregnancy tests. The concentration of hCG doubles every two days until peaking at gestational weeks 9–11 and large inter-individual variation in the levels of hCG during pregnancy has been documented (Fig. 6B in Ref. II) (Hay 1988). Nevertheless, critically low amounts in maternal circulation have been related to adverse pregnancy outcome such as early spontaneous miscarriage or ectopic pregnancy (Korhonen et al. 1994; Rull and Laan 2005).

A variety of functions have been attributed to the pleiotropic hormone hCG since early pregnancy that include supporting progesterone production by corpus luteum until the luteo-placental shift and independent expression of progesterone by the placenta but also promoting angiogenesis, trophoblast invasiveness and decidualization of endometrium, stimulating fetal testicular testosterone production and regulating maternal immunotolerance (Huhtaniemi et al. 1977; Zygmunt et al. 2002; Kayisli et al. 2003; Guibourdenche et al. 2010; Kajihara et al. 2010; Tsampalas et al. 2010; Schumacher et al. 2013). Alternative functions and sites of expression at lower amounts have been reported for assembled hCG and/or free hCG subunits in normal non-trophoblastic tissues such as pituitary, seminal fluid or secretory endometrium (Hoermann et al. 1995; Berger et al. 2007; Zimmermann et al. 2012). Also, increased level of circulatory hCG/hCGβ production in non-pregnant organism is the marker of invasive tumour progression, including bladder and gastrointestinal cancers (reviewed in (Stenman et al. 2004)). The parallel effect of hCG in implantation/placentation and cancerogenesis is attributed to its ability to modulate pathways essential in both processes, e.g. promoting cell invasion, angiogenesis and escape of immune surveillance (reviewed in (Holtan et al. 2009)).

1.2.1. Structural characteristics of hCG

Human chorionic gonadotropin belongs to the family of heterodimeric glycosylated gonadotropins together with luteinizing hormone (LH) and follicle stimulating hormone (FSH), all formed by non-covalent association of common α-subunit and unique β-subunit that defines their functional properties and binding to specific receptor (Morgan et al. 1975; Pierce and Parsons 1981). Although LH and hCG act via the same ubiquitously expressed human luteinizing hormone/chorionic gonadotropin receptor (hLH/CGR) (reviewed in (Ascoli et al. 2002)), hCG is discerned by mainly placenta-specific expression (syncytiotrophoblast cells) and increased biopotency in induction of cAMP
signaling pathway (Casarini et al. 2012). Recently, stimulatory effects independent of hLH/CGR have also been proposed for hCG isoforms in angiogenesis and trophoblast invasion (Berndt et al. 2013; Lee et al. 2013).

The hCG α- (length 116 amino acids) and β-subunits (145 amino acids) form a similar tertiary structure determined by five cystine bonds in hCGα and six in hCGβ (Lapthorn et al. 1994). Both subunits are comprised of three hairpin loops held together by three disulfide bonds that form a characteristic cystine knot motif which is a highly conserved structural feature also found among growth factors such as transforming growth factor-β2, nerve growth factor and platelet-derived growth factor-BB (Murray-Rust et al. 1993; Lapthorn et al. 1994).

Disruption of the cystine knot forming disulfide bonds in hCGβ (Cys38-Cys90, Cys34-Cys88 and Cys9-Cys57) in vitro has a detrimental effect on folding, assembly and subsequently function of the protein (Bedows et al. 1993; Mishra et al. 2003). Concordantly, no in vivo mutations affecting the production of hCG hormone have been characterized likely due to resulting compromised fertility and embryo’s viability.

The stability of hCG is determined by two N-linked and four O-linked oligosaccharide chains specific to the hCGβ protein that prolong the hormone’s half-life to >24 hours compared to 3–4 hours for FSH and only <1 h for LH (Table 1 in Ref. II) (Morgan et al. 1975; Lapthorn et al. 1994). A hyper-glycosylated form of hCG (hCG-h) with large complex glycan moieties has been described as the major form of hCG expressed by invasive extravillous cytotrophoblast during the first weeks of normal pregnancy and promoting early placentation process (Guibourdenche et al. 2010). Increased amounts of hCG-h have been detected in cases of testicular cancer, hydatidiform mole, choriocarcinomas and trisomy 21 (Elliott et al. 1997; Cole et al. 1999; Lempiainen et al. 2012), whereas reduced levels have been observed in early pregnancy loss and preeclampsia (Kovalevskaya et al. 2002; Keikkala et al. 2013).

### 1.2.2. Genomic context and evolution of hCGβ genes (also reviewed in Ref. II)

In humans, the hormone-specific β-subunit of hCG is encoded by four CGB genes (CGB, CGB5, CGB7 and CGB8) with 97–99% DNA sequence identity that give rise to an almost identical hCGβ protein (98 – 100%) in the placenta (Bo and Boime 1992; Hollenberg et al. 1994). The hCGβ genes are co-located with the LHβ gene (LHB) and two putative pseudogenes (CGB1 and CGB2) in tandem and inverted orientations within a common gene cluster (approximately 50 kb) at 19q13.32 (Figure 3) (Policastro et al., 1986).

The CGB genes have likely evolved by the expansion of the ancestral LHB gene duplicons in the primate lineage as consistently increasing number of CGB genes is evident among the primates today reaching five gene copies in chimpanzee and six in the human genome (Figure 3; Ref II) (Maston and Ruvolo 2002; Hallast et al. 2008). The duplicate CGB gene acquired a
frameshift mutation in the last exon elongating the protein by 24 amino acids (termed as the C-terminal extension) that introduced the attachment sites for novel 0-linked oligosaccharides and subsequently increased the half-life of CG (Maston and Ruvolo 2002). Nevertheless, the homology between human LHβ and CGβ proteins has remained at >80% (Pierce and Parsons 1981). Alterations in the promoter region transferred the expression of the novel gonadotropin from pituitary to placenta and from non-pregnant to pregnant organism (Hollenberg et al. 1994).

Figure 3. Genomic context of the LHB/CGB gene cluster in primates. The figure was drawn based on Ensembl database (http://www.ensembl.org/; Release 54). Boxes denote the genes and triangles above or below them point to the direction of transcription. The black boxes indicate CGB, white LHB and grey neighbouring genes. The CGB genes of rhesus macaque (Macaca mulatta) and common marmoset (Callithrix jacchus) are indicated as CGB A-C, since their ancestral status relative to the human CGB genes is unknown (reviewed in Henke and Gromoll 2008; Hallast et al. 2009) (Ref. II).

The CGB1 and CGB2 genes further diverged from the CGβ genes due to replacement of classical CGB 5’ upstream region with a novel DNA sequence changing the open reading frame and promoting speculations on the existence of a novel protein that bears no resemblance to hCGβ (Bo and Boime 1992; Hallast et al. 2007). Although the expression of CGB1/2 splice variants has been detected in placenta, testis and pituitary (Dirnhofer et al. 1996; Rull and Laan 2005; Parrott et al. 2011) and at increased levels in ovarian and breast cancer tissues (Giovangrandi et al. 2001; Kubiczak et al. 2013), only the classical hCGβ protein has been identified in testis possibly induced by the active transcription of snaR-G genes located in the novel inserted sequence upstream of CGB1/2 (Parrott et al. 2011). The functional relevance of the
genes and the existence of the hypothetical CGB1/2 protein are currently under debate.

The appearance of CGB genes has been correlated with the evolution of increasingly invasive hemochorial placenta in primates as CG mediates the trophoblast invasion and placental anchoring into maternal tissues (Maston and Ruvolo 2002; Cole 2009). Concordantly, inadequate placental invasion and hCG production may lead to obstetric complications (e.g. pre-eclampsia, spontaneous miscarriage) largely unique to humans as the representatives of most advanced hemochorial placentation (Jauniaux et al. 2006).

1.2.3. hCGβ genes as candidate risk loci in pregnancy failure

The human CGβ genes exhibit a highly variable transcriptional activity in the placenta with CGB5 and CGB8 providing the largest contribution to the pool of hCGβ transcripts (up to 82%) (Bo and Boime 1992; Miller-Lindholm et al. 1997; Rull and Laan 2005). The cumulative expression level of the hCGβ genes during the course of pregnancy is in good correlation with the total amount of the hormone as the rate-limiting step in hCG production is the formation of the hCGβ protein (Fig. 6 in Ref. II) (Huth et al. 1992). The clinical significance of the hCGβ genes is indicated by the decreased placental expression among the cases of recurrent miscarriage (Rull and Laan 2005) or excessive expression in ectopic and molar pregnancy (Rull et al. 2008). The requirement of balanced biallelic transcription of maternal and paternal alleles has been observed for early pregnancy, whereas methylation allelic polymorphism patterns in the CGB5 gene whereby paternal alleles have gained methylation has been associated with recurrent miscarriage (Uuskula et al. 2010).

Addressing the genetic variation of the CGB genes in relation with obstetric complications has been challenging due to high DNA sequence similarity between gene copies (up to 99%), imprecise map of gene-specific polymorphisms and likely compensatory effect of the multi-copy hCGβ genes. It is thus not surprising that no genetic variants with phenotypic effect have been reported in CGB genes until the studies reported in this thesis (also reviewed in Ref. II). Further research is needed to elucidate the genetic factors underlying the transcriptional variability of CGB genes seen in various pregnancy disorders but also cancer.

1.3. DNA copy number variation

Copy number variants (CNVs) (Figure 4) represent deletions, duplications or inversions of > 50 bp DNA sequence cumulatively covering up to 12–30% of human genome compared to 1% altered by SNPs (Redon et al. 2006; Zhang et al. 2009; Mills et al. 2011). Due to the large genomic coverage, CNVs introduce substantial dynamics and interindividual variation to the genome not only among humans but also other mammals, including chimpanzee, rhesus
macaque, dog, mouse and rat (Graubert et al. 2007; Kehrer-Sawatzki and Cooper 2007; Guryev et al. 2008; Lee et al. 2008; Conrad et al. 2010; Berglund et al. 2012), and in Drosophila melanogaster (Emerson et al. 2008). Recent technological advances such as next-generation sequencing have further improved fine-mapping of the CNV profiles in human genome facilitating the accurate copy number detection of even complex genomic regions (Alkan et al. 2009; Mills et al. 2011).

It has been estimated that CNVs disrupt around 13% of human RefSeq genes (McCarroll et al. 2008; Conrad et al. 2010) and approximately 18% of variability in gene expression may be attributed to these DNA variants (Stranger et al. 2007). Concordantly, a growing number of structural DNA variations have been implicated in benign phenotypic traits (e.g. human amylase gene variation) (Perry et al. 2007), rare genomic syndromes (e.g. Potocki-Lupski syndrome) (Potocki et al. 2000) and common disease susceptibility (Girirajan et al. 2011a). Both increased global burden of CNVs and distinct CNV loci have been linked to several complex disorders pointing to considerable contribution of these genetic variants in modulating individual’s disease risk.

![Figure 4](image)

**Figure 4.** Schematic representation of a DNA copy number variation (CNV). In case a CNV region involves a dosage sensitive gene/regulatory region (arrow), deletion (individual X), tandem duplication (Z) or a more complex genomic rearrangement (W) may lead to a disease among the CNV carriers.

### 1.3.1. CNVs in common diseases

#### 1.3.1.1. Hypothesis-free global genomic profiling of CNVs

Genome-wide screening for structural variations has been undertaken in an attempt to map the global ‘risk profile’ or determine particular categories of CNVs predisposing to common disease. Due to likely disease-conferring risk, rare CNVs have been preferentially addressed by numerous studies and several rare case-specific rearrangements have been implicated in the occurrence of a common disease. There is a publication bias towards the neuropsychiatric disorders likely due to the assembled consortia and access to large number of cases needed for increased statistical power of genome-wide analysis, particularly when targeting rare CNVs.
Increased burden of *de novo* and rare (<1%) CNVs among the affected individuals has been reported for several neuropsychiatric disorders, including autism (Pinto et al. 2010), schizophrenia (Consortium 2008; Walsh et al. 2008; Vrijenhoek et al. 2008; Vacic et al. 2011), bi-polar disorder (Malhotra et al. 2011) and a combined analysis of autism, intellectual disability and dyslexia (Girirajan et al. 2011a). Several novel candidate genes increasing the disease risk have been identified in these CNVs, including *VIPR2*, *NRXN1* and *SHANK2*. An enrichment of genes affecting neuronal signaling and development pathways has also been observed potentially suggesting that co-occurrence of individually rare CNVs may cumulatively increase the risk of a neurodevelopmental disorder (Walsh et al. 2008; Pinto et al. 2010). Concurrently, a two-hit model has been proposed, whereby at least two mutational hits of large (>500 kb) rare CNVs are needed for the manifestation of a neuropsychiatric disease (Girirajan et al. 2010; Girirajan et al. 2012).

In addition to neurodevelopmental disorders, genome-wide association studies involving both rare and common CNVs have been performed for complex diseases having a large impact on public health. In severe obesity cases, burden of long (>100 kb or >500 kb) rare CNVs was reported and significant enrichment of deletion CNVs affecting genes of G-protein coupled receptors was observed (Bochukova et al. 2010; Wheeler et al. 2013). A genome-wide analysis of all CNV classes in eight common diseases has been applied by The Wellcome Trust Case Control Consortium in altogether 16000 cases of bipolar disorder, breast cancer, coronary artery disease, Crohn’s disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes (Craddock et al. 2010). CNVs in or near altogether three loci were associated with Crohn’s disease (*IRGM*), type 2 diabetes (*TSPAN8*) or with Crohn’s disease, rheumatoid arthritis and type 1 diabetes (HLA region). The association of HLA region with type 1 diabetes was further confirmed by an independent study identifying 11 risk conferring CNVs, including a deletion near HLA-DQ allele (Grayson et al. 2010). Recently, two studies have reported the impact of CNVs on human longevity, whereby either burden of large (≥500 kb) common deletions or CNVs involving genes that affect RNA alternative splicing were associated with decreased lifespan (Kuningas et al. 2011; Glessner et al. 2013).

### 1.3.1.2. Targeted analysis of copy number variable genes

Targeted analysis of several distinct copy number variable loci has been performed in the context of common human diseases, majority of which are related to disturbances in the function of immune system and share the risk-associated loci.

The complex structural variation of duplicate genes encoding alpha- and beta-defensins (*DEFA* and *DEFB*) has been mapped in detail in human population (Hollox et al. 2003; Aldred et al. 2005). As the multi-functional defensin proteins have an important role in innate system, copy number
variation of these genes has been studied in complex disorders and increased or decreased copy numbers have been associated with psoriasis (Hollox et al. 2008; Stuart et al. 2012) and Crohn’s disease (Fellermann et al. 2006), respectively.

Human Fcγ receptors are IgG-binding glycoproteins that are encoded by a duplicate gene family of FCGR genes and have been implicated in autoimmune and inflammatory diseases. Copy number variation of members of this gene family increases the risk of idiopathic thrombocytopenic purpura and systemic lupus erythematosus (Breunis et al. 2008; Fanciulli et al. 2007). The systemic lupus erythematosus disease has also been associated with CNVs involving complement component C4 genes (Yang et al. 2007) and interestingly, an epistatic effect was found between FCGR3B and CCL3L1 encoding for macrophage inflammatory protein (MIP)-1α influencing both systemic lupus erythematosus and rheumatoid arthritis (Mamtani et al. 2010). Susceptibility to human immunodeficiency virus (HIV) infection has been addressed with CNV studies although with contrasting results. Contribution of the gene dosage of CCL3L1, the human immunodeficiency virus-1 (HIV-1)-suppressive chemokine, has been debated (Gonzalez et al. 2005; Shao et al. 2007), whereas a positive association was observed for the relative amounts of activating and inhibitory KIR genes that control HIV-1 infection (Pelak et al. 2011). As the KIR gene family has been associated with a variety of human diseases and extensive copy number variability has been described in the region (Jiang et al. 2012; Kusnierczyk 2013), other traits depending on KIR gene dosage are expected to be reported.

As examples of non-immunomodulatory findings, triplication of trypsinogen gene (PRSS1) in patients of hereditary pancreatitis and recently, CNVs disrupting the AUTS2 gene of unknown function among the cases of neurodevelopmental disorders have been reported (Le Marechal et al. 2006; Nagamani et al. 2013) (reviewed in (Girirajan et al. 2011b)).

### 1.3.2. CNVs in pregnancy complications

In spite of the growing evidence of CNVs modulating the risk to various complex disorders, only few studies addressing the contribution of CNVs in pregnancy complications have been reported very recently.

A combined GWAS and CNV analysis has been performed among the cases with pre-eclampsia, the pregnancy-related hypertension, and in normotensive females (Zhao et al. 2012). Although only deletion CNV regions were considered as potentially the most deleterious type of rearrangement in the original study, three CNVs with increased prevalence among white female cases compared to controls were identified (Table 1). A 15 kb deletion encompassing the PSG11 gene expressed during pregnancy by syncytiotrophoblasts was highlighted as potentially contributing to the occurrence of preeclampsia. The findings were not subsequently replicated among preeclamptic and normoten-
sive Afro-Caribbean, Hispanic and European ancestry mothers, however, a number of novel candidate rearrangements were reported (Table 1) (Zhao et al. 2013). Neither of the studies identified SNP associations of genome-wide significance likely due to the small number of analyzed cases.

Single reports for recurrent miscarriage and stillbirth have been published, both addressing a small number of cases (n = 27 and 29, respectively) and largely relying on the reference dataset of common CNVs in the Database of Genomic Variants (DGV) in order to infer the potential impact of identified rearrangements on the disease (Table 1) (Ledig et al. 2010; Rajcan-Separovic et al. 2010; Harris et al. 2011). CNVs involving the PAPPA and HLA-DPA1 genes were implicated in the occurrence of stillbirths, whereas in recurrent miscarriage, two imprinted placental genes rearranged in the placental samples, TIMP2 and CTNNA3, were highlighted as novel potential candidates increasing the risk of early pregnancy loss (Rajcan-Separovic et al. 2010; Harris et al. 2011).

Like candidate gene-based studies, CNV analyses are sensitive to sample size, ethnic background and adequate definition of phenotype of the study subjects and thus, most of the currently reported studies on pregnancy complications are likely to underestimate the number and effect of CNVs modulating the risk to disease. Also, none of the published studies have addressed the couple’s cumulative susceptibility to disease, although both mother and father contribute to the fetal/placental genome and function (Table 1). Further research is needed to elaborate on the findings and assess the contribution of CNVs in manifestation of pregnancy complications.
Table 1. CNV studies in pregnancy complications.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Study subjects</th>
<th>No of cases/controls</th>
<th>Screening platform</th>
<th>No of CNVs</th>
<th>CNV size (kb)</th>
<th>Rearranged genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preeclampsia</td>
<td>White mothers</td>
<td>177/116</td>
<td>Affymetrix Genome-Wide Human SNP Array 6.0</td>
<td>3</td>
<td>15–41</td>
<td>PDXDC1, PSG11</td>
</tr>
<tr>
<td></td>
<td>Afro-Caribbean mothers</td>
<td>21/1049</td>
<td>Illumina Human1M-Duo</td>
<td>8</td>
<td>4–89</td>
<td>GPR39, HCG26, HCG2P7, HCG4P6, HCP5, MAG12, MICA, UGT2B10</td>
</tr>
<tr>
<td></td>
<td>European ancestry mothers</td>
<td>50/1207</td>
<td>Illumina Human610-Quad</td>
<td>3</td>
<td>14–39</td>
<td>AADAC</td>
</tr>
<tr>
<td></td>
<td>Hispanic mothers</td>
<td>62/661</td>
<td>Illumina Human1M-Duo</td>
<td>11</td>
<td>11–373</td>
<td>DTX2, FRRI1, GOLPH3, HCG4P6, PDZD2, POMZP3, TXNRD2, USP37</td>
</tr>
<tr>
<td>Stillbirth</td>
<td>Placental samples</td>
<td>29/10; DGV</td>
<td>Illumina CNV370-Duo</td>
<td>24</td>
<td>9–2886</td>
<td>ADD2, AGO2, AUH, C9orf47, CADM2, CHRAC1, CKS2, CLU1, CLU1 OS, CYP2C18, CYP2C19, DRA1, GADD45G, GP6, HLA-DPA1, MIR3153, MIR290, MRPS30, NF1L3, OR4A47, OR4B1, OR4C3, OR4C45, OR4S1, OR4X1, OR4X2, OR51F2, OR51S1, OR51T1, OR52R1, PAPP-A, PTK2, PTTP1, RDH13, S1PR3, SECISBP2, SEMA4D, SHC3, SYK, TGFA, TRAPP1, TRRAP</td>
</tr>
<tr>
<td>Recurrent miscarriage</td>
<td>Placental samples</td>
<td>27/ DGV</td>
<td>Array-CGH</td>
<td>13</td>
<td>27–1593</td>
<td>C7orf36, CSS3, CTNN43, GPM6B, HDHD14, LIPI, NDUF1A2L, NPSAS3, ODF1, PARK2, PNPL4A, POU6F, PRMT3, RAB9A, RALA, STS, TIMP2, TRPPC2, VCX, 13 ZNF family genes</td>
</tr>
</tbody>
</table>

a Only deletions were considered in the study
b Database of Genomic Variants (DGV) (http://genome-euro.ucsc.edu/cgi-bin/hgGateway) was used as a reference dataset in addition to control samples if available to exclude potentially benign and common CNVs from the study.
2. AIMS OF THE PRESENT STUDY

The present thesis aimed to elucidate contribution of human genome variation at global and fine-scale level in modulating the genetic etiology of recurrent miscarriage (RM). The main research aims of the thesis were as follows:

I. Fine-scale genome dynamics of duplicated LHB/CGB genes essential in fertility and early pregnancy maintenance (Ref. III)
   • Investigation of structural features and recombination patterns shaping the genomic landscape of the LHB/CGB gene cluster

II. DNA single nucleotide variants in the hCG β-subunit coding genes CGB5 and CGB8 in the context of RM disease
   • Screening for and association study of single nucleotide variants in the CGB5 and CGB8 genes in three North European populations (Ref. IV, V)
   • Testing the impact of identified missense mutations on the structural and functional features of hCGβ and hCG in vitro (Ref. VI)

III. DNA copy number variants (CNVs) as novel genetic determinants of recurrent miscarriage (Ref. VII)
   • Investigation of genome-wide CNV profile and its effect on functional pathways in RM cases
   • Identification of novel common CNV regions conferring risk to RM
3. RESULTS

3.1. Genome dynamics of human LHB/CGB gene cluster (Ref. II, III)

Rationale of the study:
The expression of the heterodimeric human chorionic gonadotropin (hCG) hormone in the placenta is of critical importance to the maintenance of early pregnancy. Due to the extensive sequence complexity and lack of detailed information on the polymorphisms of the gene family encoding the β-subunit of the hormone (CGB genes) the region has proven difficult to target in genomic approaches. This study was undertaken to describe the nature and origin of the diversity of the LHB/CGB gene cluster in order to promote further research of this region in the context of pregnancy success.

3.1.1. Genomic structure of the LHB/CGB gene cluster

To fine-map the genomic structure of the LHB/CGB gene cluster in silico, the sequence obtained from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov; locus no NG_000019; June 26, 2002 release) was analyzed using web-based global alignment tools CLUSTALW (http://www.ebi.ac.uk/clustalw/) and BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) (Ref. III). The human LHB/CGB gene cluster includes the LHB gene (luteinizing hormone beta polypeptide; length 1.1 kb) encoding the LHβ protein expressed in pituitary, putative pseudogenes CGB1 and CGB2 (1.4 kb) and six CGB (chorionic gonadotropin, beta polypeptide) genes encoding the hCG β subunit primarily expressed in placenta – CGB, CGB5, CGB7 and CGB8 (all 1.5 kb). The region is characterized by not only extensive DNA sequence identity between the genes (97–99% among the hCGβ genes, 92–93% when compared to LHB, 85% when compared to CGB1 and CGB2) but also between the intergenic regions (up to 97%) (Figure 5). The gene cluster is closely bordered by functionally unrelated genes RuvB-like 2 (RUVBL2) and neurotrophin 5 (NTF5; now known as NTF4).

The structural features currently seen among humans have likely arisen via several steps of duplication events involving the ancestral LHB giving rise to the CGB genes and part of the NTF5 gene that has spread into the intergenic regions (Figure 5). The duplicated intergenic Alu-rich fragment originating from NTF5 included the consensus Escherichia coli χ-sequence (GCTGGTGG) that has been associated with increased recombination and gene conversion activity (Smith 1988). The presence of χ-sites together with high content of Alu repetitive sequences in the intergenic regions (currently in humans 10–56%) may have cumulatively promoted the extensive rearrangement events in the primate lineage. The concept of duplication series in the evolution of this gene cluster is supported by the identification of various number of CGB genes in
other primates, such as chimpanzee (n = 5) (Hallast et al. 2008), rhesus macaque (n = 4) and common marmoset (n = 1) (Figure 3).

3.1.2. Polymorphism density and gene conversion activity

The diversity of the LHB/CGB genes was addressed by resequencing all genes except for CGB8 (inaccessible due to sequence complexity) in three populations: Estonians (n = 47), African Mandenka (n = 23) and Chinese Han (n = 25). In total, 191 SNPs were identified with only a small fraction represented in the dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) (Supplementary Table S1 in Ref. III). Although the level of diversity varied among the populations analyzed, the highest polymorphism density was mutually observed for the genes at the edges of the gene cluster (LHB, CGB and CGB7) (Figure 2 in Ref. II; Table 1 in Ref. III). The peripheral genes LHB, CGB and CGB7 were also characterized by strong linkage disequilibrium (LD), whereas breakdown of LD was observed toward the center of the gene cluster (CGB1, CGB2, CGB5) (for Estonians, Figure 6A; Figure 3 in Ref. III). Two major gaps in LD were noted in all populations (between the CGB5 and CGB8 genes and between CGB and CGB2) potentially pointing to higher recombination activity in these regions.
Figure 6. Linkage disequilibrium (LD) and recombination activity in the human LHB/CGB region and structural features of the potential recombination 'hot spot'. LD and recombination activity of the whole gene cluster were estimated in 47 Estonian population samples (A) and of the CGB5-CGB7 region (B) in 11 Estonian samples. LD analysis was based on the \( r^2 \) statistic measuring the correlation between alleles using Haploview software (Barrett et al. 2005) and factor \( \lambda \) was calculated using PHASE 2.1 software (Li and Stephens 2003), both excluding SNPs with minor allele frequency <10%. (A, B) The regions of weak (white) and strong LD (black/dark gray) overlap with the occurrence of recombination 'warm/hot spots' and low recombination rate, respectively. (B) The recombination hotspot was identified between the CGB8 and CGB7 genes next to the recombination-associated \( \chi \)-site (denoted in red). Locations of genes are marked with orange brackets (CGB8) or arrows denoting the direction of transcription (CGB5, CGB7). (C) Prediction of single-stranded DNA secondary structure for the estimated recombination 'hot spot' (bordered by black brackets) between CGB8 and CGB7. The DNA structural analysis was performed with EMBOSS einverted software (http://emboss.sourceforge.net/) (Rice et al. 2000). The hotspot is flanked by double inverted Alu-sequences forming the stem (625 bp) of the palindrome, and its center falls within the loop (222 bp). (MIR) Mammalian-wide interspersed repeat; (FLAM_A) Alu-element-like repeat.
The high sequence similarity between duplicated regions in the \textit{LHB/CGB} cluster is the rich substrate for the phenomenon of gene conversion whereby DNA sequence information is transferred between a pair of highly identical sequences via the process of homologous recombination (Figure 4A in Ref. II). A directional transfer of gene conversion tracts was observed from the central \textit{CGB} gene copies to the distal genes with the highest number of acceptor sites identified for \textit{CGB} (\(n = 8\)) and \textit{CGB7} (\(n = 7\)) (Figure 2 in Ref. III) that also exhibited the highest polymorphism density and strongest LD. Gene conversion mechanism is likely the major force in shaping the diversity patterns of the \textit{LHB/CGB} region and spreading mutations that may modulate the function of the involved genes.

### 3.1.3. Recombination rate and structural analysis of potential recombination ‘hot spot’

The recombination rate of the \textit{LHB/CGB} gene cluster was assessed and recombination ‘hot spots’ were identified from unphased genotype data based on ‘\(\lambda\)’ that estimates by which factor the recombination rate between the loci exceeds the average background rate (Li and Stephens 2003). The value \(\lambda = 1\) corresponds to the absence and \(\lambda > 1\) to increased recombination activity. In all populations studied, potential recombination ‘warm spots’ and ‘hot spots’ were identified in the regions overlapping with the sharp LD breakdown observed in the \textit{LHB/CGB} cluster (Figure 3 in Ref. III). Among Estonians, the factor \(\lambda\) reached 2.36 for the potential recombination ‘warm spot’ between the \textit{CGB} and \textit{CGB2} genes and 57.1 for the ‘hot spot’ between \textit{CGB5} and \textit{CGB7} (Figure 6A).

In order to refine the location of the potential recombination ‘hot spot’ in the region between the \textit{CGB5} and \textit{CGB7} genes that includes \textit{CGB8} and has remained inaccessible to \textit{CGB8}-specific PCR and sequencing methods due to high sequence complexity, I conducted a pilot study by applying a combination of long-range (8.3 kb) and nested PCRs and subsequent resequencing in 11 Estonian populations samples. The analysis confirmed the weak LD in the region and fine-mapped the location of the recombination ‘hot spot’ in a \(<1\) kb region between the \textit{CGB8} and \textit{CGB7} genes, embedded within an \textit{Alu}-rich (~75\% \textit{Alu}-sequences) segment and 90–100 bp from the recombination-associated \(\chi\) – sequence (Figure 6B). The DNA structural analysis of the ‘hot spot’ identified palindromic \textit{Alu} repeats in the region that could give rise to a stem-loop secondary structure with a 625 bp stem and single-stranded loop segment (222 bp) in the middle (Figure 6C) potentially promoting double-strand breaks and subsequent recombination/gene conversion activity. Inverted repeats were also identified in the recombination ‘warm spot’ region between the \textit{CGB} and \textit{CGB2} genes, however with a much longer spacer (2788 bp) which might lead to decreased stability of the secondary structure and decreased recombination rate.
3.2. DNA single nucleotide variants of CGB5 and CGB8 in the context of recurrent miscarriage

3.2.1. SNP profile of the CGB5 and CGB8 genes in recurrent miscarriage (Ref. IV, V)

**Rationale of the study:**

Due to the irreplaceable role of hCG at the early stage of pregnancy, CGB genes encoding the hCG β-subunit are regarded as potential candidate genes for studying the genetic etiology of early pregnancy complications, such as recurrent miscarriage. The CGB5 and CGB8 genes that cumulatively provide up to 82% of hCG β-subunit transcripts in the human placenta (Miller-Lindholm et al. 1997; Rull and Laan 2005) likely have the largest effect on the quantity and quality of the hCG produced and are thus best candidates for addressing the role of this gene family in recurrent miscarriage.

3.2.1.1. Screening for variants in Northern Europe

The lack of comprehensive list of gene-specific polymorphisms for highly similar CGB5 and CGB8 (>92%, Ref. III) in the SNP databases prompted the full resequencing of these genes in RM case-control subjects from Northern Europe (Ref. IV). The 5' upstream (up to –435 bp from the start site of mRNA sequence) and genic region (down to +1082 bp relative to the mRNA start site) of CGB5 and CGB8 were resequenced in 184 female RM cases and their male partners (≥3 consecutive miscarriages before gestational week 22; Estonia n = 99, Finland n = 85) and 195 fertile women as controls (at least one live birth in Finland or three in Estonia and no previous miscarriages; Estonia n = 95, Finland n = 100). In total, 71 SNPs were identified (49 in CGB5, 22 in CGB8; Table 1 in Ref. IV) with nearly absent LD between them in both populations (Figure 2 in Ref. IV) that is largely concordant with the LD breakdown in the middle part of the LHB/CGB cluster reported in Ref. III.

3.2.1.2. Disease-related variants and haplotypes in CGB8

The SNP profile analysis of CGB8 highlighted the potential functional relevance of the 5' upstream region of the gene. The region involved only three SNPs (Table 1 in Ref. IV) compared to the 18 SNPs in the respective DNA stretch of CGB5 and was predicted to evolve under stronger functional constraints. Two neutrality tests were applied to explore observed versus expected distribution of SNPs and haplotypes – the Tajima’s D (the difference between observed (π) and expected (θ) diversity estimates) and Ewens-Watterson homozygosity estimations (tests the observed allele frequency spectrum with the expected allele frequency spectrum under the neutral model (Hardy-Weinberg Equilibrium)). Both, the Tajima’s D statistic (D^T = 2.29,
$P < 0.05$) as well as Ewens-Watterson homoygosity test ($P = 0.007$) (Table 2) indicated a possible scenario of balancing selection driving the three apparently most efficient $CGB8$ promoter variants (H1, H3, H4 in Figure 7A) to high frequency in both populations (Supplementary Table S2 and Supplementary Fig. S2 in Ref. IV).

Table 2. Neutrality tests of the $CGB5$ and $CGB8$ genes in fertile women and RM cases.

<table>
<thead>
<tr>
<th></th>
<th>CGB5</th>
<th>CGB8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full regiona</td>
<td>5’upstream regionb</td>
</tr>
<tr>
<td>No of SNPs</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>Fertile women</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tajima D$^d$</td>
<td>–1.12980</td>
<td>–1.16944</td>
</tr>
<tr>
<td>$P$-value of Ewens-Watterson $F^e$</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>RM Patients</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tajima D$^d$</td>
<td>–1.23744</td>
<td>–1.23550</td>
</tr>
<tr>
<td>$P$-value of Ewens-Watterson $F^e$</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

*aSNPs in 5’upstream and genic regions; bSNPs located in the region of –435 bp up to –1 bp from the start site of mRNA sequence; cSNPs located in mRNA sequence: +1 bp up to +1082 bp from the start site of mRNA sequence; dThe basis of the Tajima’s D statistics ($D^T$) is the difference between observed ($\pi$) and expected ($\theta$) diversity estimates: under neutral conditions $\pi = \theta$ and $D^T = 0$; eThis statistic tests the observed allele frequency spectrum with the expected allele frequency spectrum under the neutral model (Hardy-Weinberg Equilibrium). *$P = 0.0169$; ns – non-significant ($P > 0.05$)

Only one rare promoter variant H2 was identified in the 5’ upstream region defined by the minor allele A of SNP c8EF-4 (position relative to transcription start site) that was solely present in RM patients, one from Finland and two from Estonia (Cochran-Armitage trend test, $P = 0.071$; Figure 7A; Table 1 in Ref. IV). This polymorphism may potentially affect the expression of the $CGB8$ gene as it is located (i) within the AP1-like sequence overlapping the $hCG\beta$ initiator element critical for basal transcription and (ii) downstream of the Ets-2 binding site acting as a major enhancer of $hCG\beta$ gene expression (Ghosh et al. 2003).
Figure 7. Haplotype networks of predicted promoter variants in the 5' upstream region of CGB8 (A) and CGB5 (B). Promoter variants were inferred from unphased genotype data using the Bayesian statistical method in the program PHASE 2.1.1. (Stephens et al. 2001) and networks of inferred haplotypes were drawn with NEWTWORK 4.201 software using the Median-Joining network algorithm (Bandelt et al. 1999). Singleton polymorphisms were excluded from the haplotype prediction analysis because of unreliable phasing. The size of each node is proportional to the haplotype frequency in the total dataset. The relative distribution of each haplotype among the RM cases (black) and fertile controls (white) is indicated. Haplotype nomenclature is shown in Supplementary Table S2 in Ref. IV. (A) Haplotype H2 defined by the minor allele of a proximal promoter mutation c8EF-4 was exclusively identified among RM patients in both Estonia and Finland. (B) The clade with haplotypes H1, H2, H10 and H11 carry the combination of minor alleles of four SNPs (c5-155/C/c5-147del/c5-144/C/c5-142/A) that originates from the CGB8 gene and has a protective effect against RM. The locations of the four SNPs are indicated in the alignment of consensus DNA sequence of the LHβ and hCGβ genes. Positions are given according to the transcription start site of CGB5.
Interestingly, the haplotype combining the minor alleles of c8EF-287 and c8EF-186 was absent in the current dataset in spite of the relatively high minor allele frequencies (MAF) of the SNPs (25.2% and 39.7%, respectively). The discrepancy between observed (0%) and expected (10%) haplotype frequency may be explained by the localization of these SNPs within the binding sites of Sp1/AP-2 transcription factors residing in the hCGβ gene region critical for the trophoblast-specific expression as well as cAMP-responsiveness of the transcription (Albanese et al. 1991). Considering also that the CGB8 gene contributes most to the pool of hCGβ transcripts in the placenta (40%) (Rull and Laan 2005) it is thus likely that CGB8 is harboring the most optimally functioning promoter sequence of the hCGβ genes.

3.2.1.3. Disease-related variants and haplotypes in CGB5

The SNP profile of 5’ upstream region of CGB5 (in total, 18 SNPs) involved a motif of minor alleles of four polymorphisms c5-155G/C, c5-147G/del, c5-144T/C and c5-142T/A that was completely identical to the homologous region in the CGB8 gene exhibiting no genetic variation at these positions (Figure 7B). This DNA stretch (including c5-155C/c5-147del/c5-144C/c5-142A) probably originates from the CGB8 gene via a meiotic gene conversion event between the two promoter regions. The haplotypes carrying this CGB8-specific DNA sequence (H1, H2, H10, H11) formed a separate clade on the haplotype network of the CGB5 5’ upstream region (Figure 7B). It was speculated that in some pregnancies with impaired trophoblast growth (due to genetic, thrombophilic, immunological or other reasons), the placenta with the most efficient CGB5 promoter haplotype (originating from and identical to the efficient CGB8) may have a better capacity for extra hCG production that may eventually rescue the threatened fetuses.

Concordantly, the minor allele frequency of the four CGB5 SNPs was higher in fertile women (12.05–13.08%) compared to RM group (7.10–7.92%) in Estonian and Finnish populations combined and the association testing with the occurrence of RM identified a modest but significant protective effect for these SNPs (P < 0.025; OR = 0.54–0.58; Table 3 in Ref. IV). The results were successfully replicated in a meta-analysis study across the Estonian, Finnish and an independent RM case-control sample set from Denmark (RM cases with three or more consecutive miscarriages, n = 450; fertile women with at least two normal pregnancies and no miscarriages, n = 119) (P = 0.021; Table 2 in Ref. V). A strong protective effect against RM was also observed for a genic SNP c5EF1038 in CGB5 intron 2 (Cochran-Armitage trend test P < 0.007; OR = 0.53 [95% CI 0.32–0.85]) with the frequency of 14.36% in fertile women compared to 8.15% in the RM group of Estonian and Finnish combined sample (Table 3 in Ref. IV). However, the allele frequency of this polymorphism did not differ between the RM cases and controls in the Danish replication sample.
(MAF, 7.14% versus 7.42%, respectively; Cochran-Armitage trend test $P = 0.52$) and was not associated with RM (Ref. V).

### 3.2.1.4. Rare non-synonymous mutations in CGB5 and CGB8

In addition to the spectrum of SNPs described in the CGB5 and CGB8 genes, four rare mutations leading to non-synonymous amino acid changes in the hCGβ protein were identified: CGB5 p.Val56Leu in a single Finnish RM patient, p.Arg8Trp and p.Pro73Arg substitutions in the CGB8 gene of single Estonian patients, and CGB8 p.Val29Ile in one Finnish patient, two Estonian patients and seven Estonian fertile women (positions in the mature hCGβ protein; Table 1 in Ref. III). The three missense mutations (CGB5 p.Val56Leu, CGB8 p.Arg8Trp and p.Pro73Arg) exclusively present among the cases were speculated to have impact on the production of hCGβ-subunit protein in placenta and thus re-occurrence of miscarriages.

### 3.2.2. Structural and functional impact of non-synonymous mutations in CGB5 and CGB8 (Ref. V)

**Rationale of the study:**

In spite of the functional relevance of hCGβ in the establishment of pregnancy, the studies on genetic variation of the CGB genes and hCGβ protein has been challenging due to a high sequence similarity between the gene copies as well as the hCGβ proteins coded by these genes (98–100%) (Ref. II, III). Thus, only one naturally occurring variant of hCGβ (CGB5 p.Val79Met) has been functionally characterized leading to inefficient hCG assembly *in vitro* (Miller-Lindholm et al. 1999). To assess the impact of the novel missense mutations identified by our screening study exclusively among the Estonian and Finnish RM cases (p.Arg8Trp, p.Val56Leu and p.Pro73Arg; Ref. IV), I conducted series of experiments addressing the structural and functional features of recombinant hCGβ proteins carrying these amino acid changes (performed in I.T. Huhtaniemi’s lab, Imperial College London, UK; Ref. VI).


The prevalence of the rare mutations p.Arg8Trp, p.Val56Leu and p.Pro73Arg was estimated in a combined sample set of the discovery subjects from Estonia and Finland (described in Ref. IV) and replication sample from Denmark (described in Ref. V) (in total, 655 RM cases and 431 fertile female controls). Two of the studied substitutions CGB5 p.Val56Leu (rs72556325, g.1178G>C; located in exon 3, **Figure 8A**) and CGB8 p.Arg8Trp (rs72556341, g.806C>T; exon 2; **Figure 8A**) were each identified in a single heterozygous RM patient,
whereas the CGB8 p.Pro73Arg mutation (rs72556345, g.1237C>G; exon 3; Figure 8A) was identified with a similar carrier frequency (0.46%) among RM (3/655) and control (2/431) individuals in the full screened Northern-European sample set (Estonian, Finnish, Danish, n = 1086) (Table 2 in Ref. VI).

All three hCGβ missense mutations under study are located immediately next to disulfide bond forming cysteins (Cys9, Cys57 and Cys72; Figure 1A in Ref. VI). Positions Arg8 and Val56 are involved in the central cystine knot structure essential for hCGβ folding and heterodimer assembly, whereas Pro73 is incorporated in a stable turn of the protein loop 3 that does not directly associate with the hCG α-subunit or the human luteinizing hormone/chorionic gonadotropin receptor (hLH/CGR) (Figure 8B). Unlike other targeted missense mutations, Val56 was found as fully conserved among hCGβ homologs from mammals to fishes (Figure 1C in Ref. VI) and was largely buried within the hCGα/β heterodimer complex with only 3% exposed to the solvent compared to 46% in case of the unassembled β-monomer based on Solvent Accessible Surface calculations (Figure 2 in Ref. VI). Therefore out of the three mutations under study, amino acid substitutions at position Val56 were predicted to have most pronounced effect on the hCGβ protein and specifically on the formation of the hCGα/β complex.

Figure 8. Localization of identified non-synonymous mutations in the CGB5 and CGB8 genes (A) and in the structure of assembled hCG molecule (B). (A) The positions of the mutations in the CGB5 and CGB8 gene. Gray boxes represent exons with the exon number given above. Black arrows indicate the direction of gene transcription. Cen, centromere; Tel, telomere. (B) Three-dimensional (3D) structure of the assembled hCG molecule based on Protein Data Bank (PDB; http://www.pdb.org) entry 1hcn. The structure of the hCG α-subunit is depicted in blue, β-subunit in pink and the disulfide bonds in yellow. The side chains of amino acids Arg8, Val56 and Pro73 in hCGβ are shown in the space-filling representation.
3.2.2.2. Structural features and assembly of recombinant hCGβ isoforms

To study the effect of the rare missense mutations in \textit{in vitro} experimental setup, I constructed four alternative recombinant FLAG-tagged hCGβ variants – the wild-type and the hCGβ proteins carrying either p.Arg8Trp, p.Val56Leu or p.Pro73Arg mutation. Each FLAG-tagged hCGβ variant was transiently co-expressed with un-tagged α-subunit in the CHO cell line.

Forty-eight hours after transfection, the media containing the secreted hCGα and hCGβ isoforms were collected and subjected to co-immunoprecipitation using anti-FLAG antibody for specific precipitation of free FLAG-tagged hCGβ monomers and heterodimeric hCGα/β complexes. To study the structure and assembly of the expressed isoforms, the retrieved molecules were run on SDS-PAGE and detected by Western blot using either the anti-FLAG antibody which probed for both hCGβ monomers and assembled α/β heterodimers (\textbf{Figure 9A,C}), or via antiserum against hCG α-subunit which specifically visualizes assembled heterodimer hormone only (\textbf{Figure 9B}). Either non-reducing (\textbf{Figure 9A,B}) or reducing (\textbf{Figure 9C}) SDS-PAGE conditions were used based on either absence or presence of 2-mercaptoethanol and heat denaturation, respectively.

\textbf{Figure 9.} Co-immunoprecipitation and Western blot analysis of FLAG-tagged hCGβ variants co-expressed with hCGα in CHO cells. FLAG-tagged hCGβ monomers and associated complexes were immunoprecipitated from CHO cell culture media using anti-FLAG antibody-conjugated beads and separated by SDS-PAGE under non-reducing (A,B) or reducing (C) conditions. (A,C) Free and heterodimeric assembled FLAG-tagged hCGβ was detected using anti-FLAG antibody. (B) Heterodimeric hCG was specifically visualized using antiserum to hCG α-subunit. Bands corresponding to the heterodimeric hCG and unassembled hCGβ monomers are indicated by arrowheads; bands corresponding to β-subunit specific multimeric complexes that have been shown to be secreted from the cells, especially in the presence of mutations that affect the β-subunit folding pathway (Bedows et al. 1994; Feng et al. 1995; Feng et al. 1996), are indicated with a bracket. The data is drawn from the same experiment and it is representative of three independent co-immunoprecipitation experiments. Alternative α/β complex with the hCGβ conformational isoform caused by the p.Pro73Arg mutation is indicated by asterisk (*). WT, wild-type.
Under non-reducing conditions, the proportion of the p.Val56Leu \( \beta \)-subunits incorporated into \( \alpha/\beta \) heterodimers (~47 kDa) relative to the freely retained \( \beta \)-monomers (~34 kDa) was notably reduced compared to the wild-type (Figure 9A,B) indicating a decreased capability of p.Val56Leu \( \beta \)-subunits to assemble into the heterodimer and/or destabilizing the assembled hormone. The effect was confirmed by the quantitative immunoassays detecting the assembled intact hCG (hCG Human ELISA Kit; Abcam, Cambridge, MA) or total hCG\( \beta \) (hCG+\( \beta \) kit on Roche Elecsys 1010 system; measured in Tartu University Hospital, Tartu, Estonia) from the cell culture media. The efficiency of hCG assembly (measured as the ratio of assembled hCG to the total hCG\( \beta \) amount) was decreased down to 10% compared to the wild-type hCG\( \beta \) (= 100%; Student’s t test, \( P = 0.014 \)) (Figure 5 in Ref. V). The results are largely concordant with the \textit{in silico} positional context analysis predicting the hindrance of p.Val56Leu substitution upon the hCG heterodimer formation due to its location immediately next to Cys57 in the highly conserved structural feature, the cystine knot (Figure 8B). The Cys57 forms the Cys9-Cys57 disulfide bond disruption of which has previously been shown to give rise to folding and assembly deficient hCG\( \beta \) protein (Bedows et al. 1994; Mishra et al. 2003).

The hCG\( \beta \) p.Pro73Arg mutant gave rise to approximately equal amounts of two alternative hCG\( \beta \) isoforms, one corresponding to the molecular weight of the wild-type hCG\( \beta \) (~34 kDa) and an additional variant with approximately 2 kDa lower molecular weight (Figure 9A,B). Pro73 is located next to the Cys72 that forms one of the six hCG\( \beta \) disulfide bonds, Cys23-Cys72. It has been demonstrated previously that disruption of this bond affects the hCG\( \beta \) folding pathway leading to secretion of an additional isoform lacking the Cys23-Cys72 bond and exhibiting a difference of 2 kDa in size on SDS-PAGE (Bedows et al. 1993; Bedows et al. 1994). Nevertheless, both isoforms detected in this study were assembled into hCG dimer with approximately equal efficiency (Figure 9A,B).

In the case of hCG\( \beta \) p.Arg8Trp substitution, no detectable differences in the fraction of free hCG\( \beta \) or in the assembly of intact hCG compared to the wild-type variant were revealed (Figure 9). Also, no evidence of the effect of the studied three substitutions on the glycosylation pattern of the hCG\( \beta \) protein was seen based on the SDS-PAGE performed under reducing conditions that cause dissociation of hCG dimers and disruption of disulfide bonds. All tested hCG\( \beta \) variants collapsed into one major (~34 kDa) and one minor (~31 kDa) isoform (Figure 9C), previously shown to contain either two or one N-linked oligosaccharide chains, respectively (Matzuk et al. 1987).
3.2.2.3. Bioactivity of recombinant hCGβ isoforms

Bioactivity of the hCG hormone can be estimated based on its ability to induce cAMP signaling upon binding to the human LH/CG receptor. In order to achieve sufficient signaling response, a high concentration of FLAG-tagged intact hCG (termed ‘high yield’ hCG) was produced using QMCF technology (by Icosagen Cell Factory OÜ; Supplemental data, Text S1 in Ref. VI) (Silla et al., 2005). I used the series of diluted cell-conditioned media containing either the wild-type or mutant ‘high yield’ hCG adjusted for concentration of hCG heterodimer to stimulate HEK293 cell-line stably transfected with the human LH/CG receptor (HEK-hLH/CGR) and containing the cAMP-responsive (CRE) firefly luciferase reporter gene. After 5–6 h stimulation, the CRE-luciferase activity was determined and the EC\textsubscript{50} values (± standard deviation, SD) [EC\textsubscript{50} is defined as the concentration of the hormone required to produce 50% of maximal response] were estimated for each hCG isoform.

Interestingly, although the CGB5 p.Val56Leu substitution leads to inefficient assembly and/or instability of the heterodimeric hCG, the in vitro bioactivity analysis indicated an increased potency of this isoform. The cAMP response to stimulation by the p.Val56Leu heterodimer was significantly more sensitive than to wild-type hCG (Figure 10), exhibiting a half-maximal response EC\textsubscript{50} of 2.50 ± 0.81 pg/ml compared to the wild-type hCG EC\textsubscript{50} of 11.41 ± 2.32 pg/ml (t test \(P < 0.0013\)). As a consequence, the shortage of the produced heterodimeric hormone (10% compared to wild-type) may be partly or fully compensated for by its increased bioactivity.

The hCGβ variant carrying the CGB8 p.Pro73Arg mutation did not induce differences in the cAMP signaling when compared to the wild-type hCG (\(P = 0.053\); Figure 10) despite the appearance of two alternative hCGβ isoforms on SDS-PAGE (Figure 9). Importantly, the overall functional characteristics of the assembled hCG thus remained comparable to the wild-type pointing to the functional neutrality of this mutation.

The p.Arg8Trp hCGβ isoform represented a fully neutral amino acid substitution out of the three addressed in this study as no impact on structural or functional characteristics were observed compared to the wild-type hCGβ variant.
3.3. DNA copy number variants in recurrent miscarriage (Ref. VII)

**Rationale of the study:**

Previously published studies addressing the genetic etiology of RM have primarily targeted SNPs producing few confirmed disease risk variants. Although the alternative genetic variation class of copy number variants (CNVs) has been associated with various complex disorders, it has been largely understudied in pregnancy complications. My goal in this study (Ref. VII) was to elucidate the role of CNVs in modulating predisposition to RM.

**3.3.1. Genome-wide profile of CNVs**

In order to define genome-wide profile of CNVs specific to RM cases or fertile controls, screening with Illumina Human370CNV-Quad SNP array was undertaken in 70 Estonian subjects including 27 female and 16 male idiopathic RM cases with at least three miscarriages before gestational week 22 and 27 female controls with at least three live births and no miscarriages prior to recruitment. On average, 13.3 CNVs were determined per individual among the RM cases (in total, n = 43) and 12.6 CNVs/individual in fertile controls (n = 27). Identified CNVs clustered into 423 non-overlapping discrete Copy Number Variable Regions (CNVRs) (Supplemental Table S3 in Ref. VII) that were uniformly distributed across the genome (Figure 11).
Figure 11. Circos plots with frequency and distribution of deletion and duplication CNVRs across the genome among the Estonian discovery cases (n = 43) and controls (n = 27). The length of the bars (Y-axis) in the histogram represents the number of carriers of a CNV at the genomic locus, whereas the width correlates with the size of the CNVR. Single occurrence CNVs are highlighted in the yellow background. The number of carriers for outlier CNVRs of high frequency (>8 carriers in a study group) are indicated next to respective bars.

In accordance with the known heterogeneity in the etiology of RM (Rai and Regan 2006), an individual-specific increased risk of RM was inferred for a subset of patients using case-by-case CNV profile analysis. Five outlier cases (12% of RM subjects studied) were observed with (i) a 5-fold excess cumulative burden of all CNVs (up to 6.1 Mb per genome) mostly comprising of long (>100 kb) deletions (Figure 12A) or (ii) accumulation of long deletions with very high gene count (up to 189 genes per individual compared to only a median of 8.3 genes among the fertile controls) (Figure 12B). The list of rearranged genes in the outlier cases with the largest cumulative burden of all CNVs (n = 2; Figure 12A) involved several known or potential candidate genes for RM and pregnancy success, such as C4A, C4B, IGF2 or Immunoglobulin heavy chain (IGH) gene cluster that may contribute to the increased chances of miscarriages among these patients.
Figure 12. Genomic burden of all CNVs and the subset of long (≥100 kb) deletions in the Estonian discovery phase sample set. (A) Cumulative length of all deletions and duplications per individual in RM patients (n = 43) and fertile controls (n = 27). The outlier cases with increased cumulative burden of all CNVs are indicated with asterisk. (B) Cumulative length of long deletions and the number of disrupted genes per individual in the discovery phase cases and controls. The outlier cases with increased number of genes disrupted by long deletions are indicated with asterisk. Female and male patients with identical number-codes represent RM couples (e.g. RM-F45 and RM-M45). FFC, fertile female control; RM-F, female RM patient; RM-M, male RM patient.
3.3.2. Functional enrichment of genes disrupted by CNVs

To define the functional impact of the discovered CNVs and identify biological pathways significantly affected by the CNVs, functional enrichment analysis was performed for the list of all disrupted genes in either RM cases (1459 genes) or fertile female controls (553 genes) using g:Profiler software (Reimand et al. 2011). Among the cases, the results highlighted the specific impact of CNVs on the processes of immunomodulatory function at the feto-maternal interface related to maternal rejection of the semi-allogeneic fetus expressing paternally inherited alloantigens. Functional categories such as ‘Innate immunity signaling’ (REAC: 168249; 11.0% of genes disrupted by CNVs, multiple testing corrected $P = 3.57 \times 10^{-3}$), ‘Fc gamma receptors interact with antigen-bound IgG’ (REAC: 199161; 26.3% disrupted genes, $P = 9.97 \times 10^{-6}$) and ‘Complement cascade’ (REAC: 166658; 15.6% disrupted genes, $P = 1.93 \times 10^{-4}$) (Table 1 in Ref. VII) were significantly and specifically overrepresented in RM case sample. None of the immunomodulatory pathways were significantly affected by CNVs among the controls and only processes associated with general cellular function were identified (Table 1 in Ref. VII).

3.3.3. Identification of novel common CNV regions conferring risk to recurrent miscarriage

In search for novel common CNV regions that may independently modulate the predisposition to RM, I performed an experimental analysis of prioritized discrete CNV regions and conducted association study in two North European populations.

3.3.3.1. Prioritized CNVRs affecting RM in Estonia and Denmark

Nine discrete common CNV regions were selected from the whole genome SNP array genotyping data (Supplemental Table S4 in Ref. VII) for subsequent experimental testing in the Estonian discovery sample set (n = 70) using TaqMan qPCR. The selection included CNVRs that were (i) present in >1 individual, (ii) found only among RM patients or overrepresented in RM patients with OR $\geq 1.5$ and (iii) intersected with or located in the proximity (up to approximately 200 kb) of biological candidate genes with a potential impact on the course of pregnancy based on previously published literature. Three CNVRs with precise TaqMan copy number typing assays were tested in the full Estonian sample set: CNV regions of IGKV (Immunoglobulin kappa variable cluster at 2p11.2), DKK2 (Dickkopf 2 homolog, at 4q25), and PDZD2:GOLPH3 (PDZ domain containing 2; Golgi phosphoprotein 3 at 5p13.3)).

The PDZD2:GOLPH3 duplication exhibited the strongest effect (OR = 7.28) with a higher prevalence of duplication carriers among the Estonian RM cases compared to fertile controls (9/119, 7.6% versus 1/90, 1.1%, respectively)
(Supplemental Table S7 in Ref. VII). However, analysis of the Danish replication sample (in total, 439 RM patients with ≥2 consecutive miscarriages, 115 multiparous fertile female controls) identified increased prevalence of the duplication only among female patients who exhibited carrier frequency comparable to Estonian female cases (6.6% and 7.5%, respectively) (Figure 13; Supplemental Table S8 in Ref. VII). Meta-analysis combining the results of the Estonian and Danish female patient-control samples (in total, cases n = 309, controls n = 205) confirmed the association of the PDZD2:GOLPH3 CNV with an increased maternal risk of RM (OR = 4.82, P = 0.012). (Table 2 in Ref. VII). Concordantly, high risk of RM associated with the duplication was detected when the Estonian RM women were independently tested against female controls of Estonian population cohort from the Estonian Genome Center, University of Tartu (EGCUT) (n = 496; prevalence 1.0%; OR = 7.96, P = 7.9 × 10⁻⁴) (Figure 13).

The IGKV and DKK2 CNVs with small differences in carrier frequencies in the full Estonian case-control sample (Supplemental Table S7 in Ref. VII) likely represent benign common copy number variation with no major effect on the RM phenotype in our study.

![Copy number distribution and carrier frequency of the PDZD2:GOLPH3 duplication at 5p13.3 among female study subjects. CNV carriers have 3 to 4 copies of the duplication per genome in Estonia and 3 to >4 copies per genome in Denmark. Dup, duplication; EGCUT, Estonian population cohort from Estonian Biobank, Estonian Genome Center, University of Tartu.](image)

**Figure 13.** Copy number distribution and carrier frequency of the PDZD2:GOLPH3 duplication at 5p13.3 among female study subjects. CNV carriers have 3 to 4 copies of the duplication per genome in Estonia and 3 to >4 copies per genome in Denmark. Dup, duplication; EGCUT, Estonian population cohort from Estonian Biobank, Estonian Genome Center, University of Tartu.

### 3.3.3.2. Genomic context and fine-mapping of the PDZD2:GOLPH3 duplication

To define the range and nature of the PDZD2:GOLPH3 duplication (occurring in up to >4 diploid copies) at 5p13.3 predisposing to RM, I set forward to fine-map the duplication breakpoints by integrating experimental (Evagreen qPCR, PCR, sequencing) and bioinformatic approaches (including screening for
repetitive elements with RepeatMasker, http://www.repeatmasker.org/; and non-B DNA sequence motifs with Non-B DNA Motifs Search Tool, http://nonb.abcc.ncifcrf.gov/apps/nBMST/default/). The CNVR was estimated as 61.6 kb in length (positioned Chr5: 32106204 – 32167777), whereas the duplication breakpoints were located within DNA repetitive elements in the introns of PDZD2 and GOLPH3 genes that are transcribed in the opposite direction (Figure 14A,B). Application of duplication breakpoint junction-specific PCR (BP-PCR, Figure 14B) identified an identical recurrent tandem duplication event in all CNV carriers of this study. It was hypothesized that the PDZD2:GOLPH3 duplication occurred via repeat-mediated rearrangement mechanisms other than non-allelic homologous recombination. Although the PDZD2:GOLPH3 rearrangement does not involve entire coding regions of the genes, the modifications in the local genomic context may nevertheless lead to impaired function of the involved or neighboring genes as reported previously (Henrichsen et al. 2009a; Henrichsen et al. 2009b).

3.3.3. Expression profile of PDZD2 and GOLPH3

Neither PDZD2 nor GOLPH3 has been associated with pregnancy success previously and have mostly gained attention due to their role in tumorigenesis (Tam et al. 2006; Scott et al. 2009). Thus little is known on the tissue expression profile of these genes.

In order to identify the major sites of expression, I performed an expression profiling analysis using human tissue cDNA panels (Human MTC panels I and II; BD Biosciences Clontech, CA) and TaqMan qPCR assays specific to the transcripts of PDZD2 and GOLPH3. I detected the most prominent expression for both genes in the placenta that exceeded other sites of highest expression twofold for GOLPH3 (average expression relative to reference HPRT, 2.99 ± 0.06 SEM in placenta versus 1.50 ± 0.05 in prostate) or fourfold for PDZD2 (relative expression 1.33 ± 0.05 in the placenta versus 0.36 ± 0.01 in heart) (Figure 14C). Interestingly, ovary was also included in the top three sites of expression for both GOLPH3 and PDZD2 (1.13 ± 0.02 and 0.34 ± 0.0, respectively) and thus highlighting the functional relevance of these genes in reproductive organs and specifically in the placenta.
Figure 14. Fine-mapping of the PDZD2:GOLPH3 duplication CNV at 5p13.13 and expression profiling of the PDZD2 and GOLPH3 genes. (A) Genomic context of 5p13.3 involving PDZD2 and GOLPH3 genes based on UCSC database (hg19; http://genome.ucsc.edu/cgi-bin/hgGateway). The opposite transcription of the PDZD2 and GOLPH3 genes is indicated with blue and green arrows, respectively. DGV Str Var, structural variation data from the Database of Genomic Variants. (B) Schematic representation of the 5p13.3 CNV locus with or without tandem duplication. Experimentally confirmed duplication endpoints are indicated with red arrowheads and dotted lines. The breakpoint junction of the tandem duplication is marked with red arrow tail and locations of breakpoint junction-specific PCR (BP-PCR) primers are indicated with black arrows. In case of tandem duplication, a 555 bp product was amplified and subsequently confirmed by resequencing. Ex, exon. (C) Gene expression profile of the PDZD2 and GOLPH3 genes in the human cDNA tissue panels. Expression level is given relative to the reference gene HPRT and as average of three amplification reactions ± SEM. Gene expression levels in placenta and ovary are highlighted with colored bars.
4. DISCUSSION

4.1. Nature and impact of fine-scale genetic variation of the LHB/CGB region in RM

The genes encoding the hormone-specific hCGβ subunit are one of the most attractive targets in addressing fine-scale genetic etiology of early pregnancy loss as the requirement of the ‘pregnancy hormone’ hCG for the establishment and maintenance of pregnancy is absolute. The genomic landscape of the LHB/CGB gene cluster in humans described in this thesis has likely been shaped by a complex interplay between active genome dynamics and balancing act of functional constraints maintaining the effectiveness of the LHB/CGB genes. Extensive DNA sequence identity between DNA segments, density of Alu elements, formation of DNA hairpin conformation and presence of recombination warm/hot spots in the LHB/CGB region are rich substrates for increasing genome instability and gene conversion activity (Chen et al. 2007; Chuzhanova et al. 2009). As a result of the active genome dynamics, complex evolutionary rearrangements involving multiple duplication events have given rise to a varying number of CGB genes among primates with six gene copies in humans and up to even 50 copies predicted for gorilla (Dumas et al. 2007; Hallast et al. 2008). The abundant duplicated segments have been further shaped by the directional interlocus gene conversion which is a common phenomenon in the duplicated parts of the human genome and has introduced diversity at the edges of the gene cluster on one hand and on the other hand, has homogenized the duplicated genes maintaining the recombinogenic potential of the region (Dumont and Eichler 2013; Fawcett and Innan 2013).

It is noteworthy that in spite of the genomic instability and gene conversion activity spreading polymorphisms between gene copies but also introducing de novo mutations (Hicks et al. 2010; Dumont and Eichler 2013), no gene variants of substantial effect on pregnancy success have been identified in hCGβ coding genes so far. The data suggest an accompanying action of selection forces in order to maintain the functional efficacy of the genes and concurrently, balancing selection acting on the promoter region of CGB8 and driving the most effective transcription among the CGB genes was reported in this thesis (Ref. IV). Furthermore, the polymorphism screening of the two most actively transcribed hCGβ-coding genes CGB5 and CGB8 and subsequent association testing noted the lack of risk alleles associated with RM disease in Northern European populations. Only a protective effect against RM was observed for the motif of four regulatory polymorphisms in the CGB5 gene likely transferred from the respective region of CGB8 genes via non-reciprocal gene conversion event (Ref. IV and V). Similar intricate action of gene conversion and functional constraints in modifying the local genomic context has been observed for other complex genomic regions associated with pregnancy course, including growth hormone/ somatomammotropin (Sedman et al. 2008), killer
immunoglobulin-like receptor (Yawata et al. 2006; Graef et al. 2009), pregnancy-specific glycoprotein (McLellan et al. 2005; Dumont and Eichler 2013) gene families and MHC class II region (von Salome et al. 2007).

Further evidence of the functional constraints acting on hCGβ genes due to their irreplaceable role in human pregnancy is suggested by the lack of non-synonymous mutations conferring large functional effect within the genes. Out of the three missense mutations identified in the major hCGβ genes, CGB5 and CGB8, and addressed in this thesis (CGB5 p.Val56Leu, CGB8 p.Arg8Trp and p.Pro73Arg), only the p.Val56Leu substitution was initially found to hinder the assembly of the heterodimeric hormone (Ref. VI). However, the simultaneous 5-fold increase in potency partly or fully compensated for the low proportion of the assembled heterodimer (10% compared to wild-type) restoring the overall functional effect. Although inefficient hCG assembly has also been shown for the only naturally occurring variant of hCGβ (CGB5 p.Val79Met) characterized in vitro (Miller-Lindholm et al. 1999), no clinical information was available on the mutation carriers and the substitution was completely absent in a subsequent study of over 500 DNA samples from five European populations (Jiang et al. 2004). Only few other missense mutations have been identified in populations world-wide in the CGB5 (p.Arg6Gln in Han Chinese; p.Asp117Ala in African Mandekalu; Ref. III) and CGB8 genes (p.Val29Ile in Estonia and Finland; Ref. IV) with the carrier frequency of <10% in all cases but with no clinical information available for most of the mutation carriers. Furthermore, no individuals homozygous for any hCGβ mutations have been reported so far, which may indicate either an insufficient sample size in the conducted studies or that such genotypes result in complete pregnancy failure.

The lack of risk variants of substantial effect in the major hCGβ genes raises the question of missing genetic factors responsible for the large interindividual variation of hCGβ genes in gene expression during pregnancy and dysregulated expression in pregnancy disorders (Ref. II) (Miller-Lindholm et al. 1997; Rull and Laan 2005). It has been speculated whether the pathologically low hCG levels are indeed the primary cause of pregnancy loss or a secondary phenomenon resulting from inadequate placental maturation due to alternative risk factors (Larsen et al. 2013). Considering the genetic heterogeneity of the RM phenotype, it is feasible that both scenarios are relevant and the predisposing genetic profile of hCGβ genes that is likely subject to constant dynamic changes is responsible for only a subset of RM cases exhibiting decreased expression levels of hCGβ transcripts.
4.2. Genome-wide effect of CNVs in RM

4.2.1. Genomic CNV burden as risk factor for RM

The application of genome-wide CNV screening and analysis methods has enabled to compile a parental profile of CNVs predisposing to recurrent miscarriage in this thesis. A case-by-case analysis identified a subgroup of RM patients (in total, 5 out of 43; 12%) exhibiting accumulation of several long (>100 kb) deletions or long deletions with very high gene count (Figure 12; Ref. VII) potentially responsible for the manifestation of the disease in these individuals. An enrichment of large (cutoff 100 kb or larger) structural variants have been commonly associated with an increased risk of complex diseases previously, including neurodevelopmental disorders, longevity or severe obesity (Bochukova et al. 2010; Girirajan et al. 2011a; Kuningas et al. 2011). Furthermore, higher genome-wide prevalence of gene-rich long deletions has been reported for individuals affected by autism spectrum disorders and schizophrenia (Consortium 2008; Pinto et al. 2010; Griswold et al. 2012).

The deleterious nature of large rearrangements and specifically long deletions in gene-rich regions likely originates from increased chances of affecting loci relevant to pregnancy success. Gene content analysis of CNVs identified among RM cases with the largest overall genomic burden of rearrangements highlighted some potential dosage sensitive candidate genes, including $C4A$ and $C4B$ previously implicated in RM (Laitinen et al. 1991). The findings indicate that large cumulative burden of CNVs or gene-rich deletions could be regarded as an independent risk factor of RM, however the statistical significance of these findings remains to be tested further.

4.2.2. Genetic variability of immunoregulatory pathways as risk factor for RM

Adequate bi-directional communication at the feto-maternal interface and immunologic recognition of pregnancy have been proposed as of utmost importance for implantation and maintenance of early pregnancy. Fetal rejection due to either auto- or alloimmune factors is a major contributor in RM disease (reviewed in (Pandey et al. 2005; Baek et al. 2007)) and a number of candidate loci related to placental immune function have been implicated in RM independently or cumulatively, such as $KIR$ and $HLA$ gene families (Hvid et al. 2004; Hiby et al. 2008; Aruna et al. 2011). Similar to thrombophilic genetic factors, it has been suggested that only failure of several mechanisms or systematic deviations in pathways would lead to the occurrence of RM due to a likely redundancy in the immune function at the feto-maternal interface protecting against the fetal rejection (Jivraj et al. 2006; Larsen et al. 2013).

Multiple pieces of evidence indicate the contribution of structural variants in modifying the profile of genes related to immune function in pregnancy. In this thesis, significant enrichment of genes from immunomodulatory pathways were
observed within the rearranged loci of RM patients with half of the pathways specific to cases and not affected among controls, including complement cascade and innate immunity related processes (Table 3 in Ref. VII). The role of CNVs in promoting common genetic variability of immune system loci has previously been reported for human immunoglobulin heavy chain variable (IGHV) (Watson et al. 2013) and constant (IGHC) loci (Lefranc et al. 1991), beta-defensins (Hollox et al. 2008) and duplicate KIR genes (Jiang et al. 2012) indicating the potential susceptibility of these regions also to disease risk-conferring rearrangements. Furthermore, it has been shown that CNVs modulate and define the profile of presented antigens in the graft-versus-host disease, the condition of compromised immunity similar to fetal rejection in pregnancy (Guleria and Sayegh 2007; McCarroll et al. 2009). The accumulating data support the notion that CNVs may considerably affect the function of alloimmune factors and increase the risk of early miscarriages originating from feto-maternal genetic mismatch.

4.3. Global genomic analysis as the source of novel biomarkers

Hypothesis-free genome-wide approach in studying disease susceptibility is a powerful tool not only in uncovering novel pathways but also facilitating identification of novel genes and genetic variants increasing the risk of disease. Genome-wide studies targeting CNVs have been the source of numerous novel dosage sensitive candidate genes for complex diseases but also for recurrent miscarriage as reported previously by (Rajcan-Separovic et al. 2010). Aberrations of the imprinted genes TIMP2 and CTNNA3 in the placenta were implicated in increased risk of pregnancy loss and thus represent the risk factors specific to the fetus/offspring. Alternatively, screening of parental genomes may provide data on the genetic susceptibility of either or both parents to the occurrence of repetitive miscarriages.

A strong maternal risk of RM was identified for the 61.6 kb multicopy duplication of PDZD2 and GOLPH3 genes (chromosome position 5p13.3) in this thesis among women from Estonia and Denmark (meta-analysis, OR = 4.82, P = 0.012; Ref. VII). Interestingly, in addition to being transcribed in the female reproductive organ ovary, the highest expression of both of the involved genes was identified in the placenta (Figure 14C; Ref. VII). It was speculated that the increased predisposition to RM among the female duplication carriers may be attributed to the joint effect in maternal reproductive tissues and in the placental tissue carrying a maternally inherited duplication CNV.

The relevance of the PDZD2:GOLPH3 duplication as a novel biomarker in pregnancy disorders can be estimated based on two findings. Firstly, the PDZD2:GOLPH3 duplication CNV may represent a pleiotropic risk factor of pregnancy disorders as its risk-conferring effect is not restricted to RM. A
recent CNV association study identified the duplication as a potential risk factor in women affected by pre-eclampsia – a severe late pregnancy disorder attributed to placental dysfunction (Zhao et al. 2013). The overlapping causality of RM and pre-eclampsia (Li and Huang 2009; Baig et al. 2013) but also early pregnancy loss and fetal growth restriction (Ganguly et al. 2007) has been acknowledged previously.

Secondly, the PDZD2 and GOLPH3 genes affected by the 5p13.3 duplication CNV are novel candidate genes in the context of early pregnancy maintenance and the associated pathways have not been directly linked to the occurrence of early pregnancy loss previously. The function of PDZD2 is poorly defined, whereas amplification of GOLPH3, essential for Golgi trafficking and maintenance of its structure (Dippold et al. 2009; Wood et al. 2012), has been reported in various cancers and shown to induce the signaling of mechanistic target of rapamycin (mTOR) (Scott et al. 2009; Wang et al. 2012; Hu et al. 2013). The contribution of mTOR in reproductive function is however well established and alterations in mTOR signaling have been linked to multiple reproductive disorders in human and mouse, including recurrent miscarriage (Roos et al. 2007; Hirota et al. 2011; Vatin et al. 2012). The direct functional link between the GOLPH3 amplification and mTOR signaling in female reproductive tissues and placenta remains to be experimentally confirmed.

4.4. Status quo and future perspectives in assessing genetic determinants of RM

As pregnancy loss is the most common pregnancy complication, etiology of RM has been addressed by considerable number of studies for several decades establishing the multifactorial nature of the disease and the contribution of genetic susceptibility. Nevertheless, none of the reported variants have unequivocally proven as specific to the RM phenotype, neither prevalent among the cases leaving a void in the knowledge on disease heritability factors and also highlighting the multifactorial nature of the genetic component itself. Adjustments in study designs from the genetic and phenotypic perspective may prove beneficial in fine-mapping the genetic determinants of RM (also reviewed in Ref. I).

Genetic perspective

The common attractive expectation in studies of RM has been to identify single genetic variants of high effect efficiently applicable in clinical diagnostics. However, the accumulating research data indicate that very few genetic variants exist conferring high risk to RM likely due to the heterogeneity and redundancy of mechanisms in early pregnancy as also demonstrated for the polymorphisms in the hCGβ coding genes. It has been proposed that instead of single poly-
morphisms with independent prevalent effect, several genetic variants may cumulatively alter pathways essential for pregnancy maintenance (e.g. thrombophilic and pro-inflammatory mutations). Alternatively and feasibly, due to the multifactorial nature of RM, each case or family may carry a distinct genetic factor responsible for the increased risk of the disease observed among siblings (Christiansen et al. 2008; Kolte et al. 2011).

Although no SNP-based GWAS studies have been performed to identify the set of single nucleotide variants associated with RM, the genomic profiling of CNVs addressed in this thesis supports the hypothesis of cumulative risk of multiple rearrangements involving genes within similar pathways (in this case, processes related to immune function). The finding of the common PDZD2:GOLPH3 duplication independently increasing the risk of RM among women with OR = 4.82 in two populations likely reflects the lack of previous studies targeting parental CNV profile in RM. Increasing the sample size may identify additional CNVs associated with RM, however the number of common causative rearrangements is likely to remain low in accordance with and as debated for other common diseases (Craddock 2010).

In order to improve the current knowledge on the whole spectrum of genetic factors leading to RM, the focus of the association studies could be further shifted from candidate gene-based analysis to genome-wide studies and incorporating data from transcriptomics, proteomics and methylomics that have become financially, methodologically and analytically more feasible and may prove beneficial in interpretation of large-scale genomic data.

**Phenotypic perspective**

In order to increase the power of identifying novel genetic variants of lower effect size at the genome-wide level, larger case-control samples and replication samples would be needed. However, the varying phenotyping criteria of RM at different recruitment centers such as (i) the number of miscarriages and live births at recruitment of cases and controls, respectively, (ii) exclusion criteria of cases (e.g. testing for mutations in factor V (Leiden) or factor II, prothrombin) and/or (iii) selection criteria of study subjects (only women versus couples versus placenta) may hinder the pooling of sample collections or skew replication results. Also, majority of the studies addressing the genetic component of RM have focused on only maternal or fetal/placental contribution and few have included both parents (Ref. I). As mother and father contribute equally into the offspring’s genetic composition, association studies with couples may identify novel risk factors leading to early pregnancy loss. In order to uncover novel pleiotropic genetic variants not restricted to RM, parallel analysis of various pregnancy disorders may prove informative.
SUMMARY AND CONCLUSIONS

The results of this thesis can be summarized as follows:

**Fine-scale genetic determinants of RM based on LHB/CGB gene cluster**

1. The local genomic landscape of luteinizing hormone/chorionic gonadotropin β-subunit coding genes (*LHB/CGB*) in humans reflects the history of complex evolutionary processes involving expansion of repetitive elements and multiple rearrangements likely promoted by recombination ‘warm and hot spots’. A balancing interplay between gene conversion activity spreading polymorphisms between the gene copies and functional constraints maintaining the transcription efficiency of the genes shapes the region’s genomic context. It is likely that the *LHB/CGB* region is subject to further dynamic changes.

2. The most actively transcribed hCGβ genes *CGB5* and *CGB8* were characterized by the lack of genetic variants increasing the risk of RM in the case-control sample collections from Estonia, Finland and Denmark. Instead, a modest protective effect against RM was observed for a motif of four SNPs in the *CGB5* upstream region transferred from the *CGB8* gene via gene conversion event. The balancing selection acting on the upstream region of *CGB8* likely supports the highest transcription activity observed among the hCGβ genes.

3. The structural and functional analysis of three non-synonymous mutations in the *CGB5* and *CGB8* genes identified a decreased ratio of hCG heterodimer assembly for p.Val56Leu substitution. However, the low amount of formed heterodimer was compensated by its increased bioactivity upon binding to the LH/CG receptor. As only mutations of mild functional effect are tolerated and no homozygous carriers of the studied mutations or additional missense mutations with functional effect have been described in *CGB5* and *CGB8*, it was concluded that keeping the functional efficiency of the major hCGβ genes intact is essential for pregnancy maintenance.

4. Accumulating data on the fine-scale genetic variation of RM indicates that the contribution of independent variants with large effect sizes is not sufficient to explain the genetic heritability of RM and rather combinations of genetic factors or variants specific to each family lead to the manifestation of the disease.

**Global genetic determinants of RM based on CNV profiling**

1. Genome-wide analysis of copy number variants (CNVs) identified a high burden of all CNVs or long (>100 kb) gene-rich deletions in a subset of RM cases. It was postulated that increased genomic burden of CNVs and specifically long deletions may confer risk to RM due to increased chances of affecting genes relevant in early pregnancy establishment and maintenance.
2. Functional profiling of all genes affected by CNVs among the RM cases detected a significant and specific impact of rearrangements on the pathways related to immune function at the feto-maternal interface. The findings support the contribution of inadequate maternal immune response to semi-allogeneic fetus expressing paternal antigens in early pregnancy success.

3. Screening of distinct common CNV regions identified a novel genetic predisposition marker among women affected by RM in Estonia and Denmark. The multicopy duplication at 5p13.3 involved two genes PDZD2 and GOLPH3 with the highest expression in placenta but previously uncharacterized in the context of pregnancy maintenance. The PDZD2:GOLPH3 CNV may potentially represent a novel genetic risk factor not restricted to RM.

4. Global CNV profiling of parental genomes confirmed the contribution of structural variants in shaping the genetic susceptibility to RM and may prove as an informative source of novel biomarkers in other pregnancy disorders.

**Main outcome of the thesis:**
This thesis established the low impact of the fine-scale genetic variation in the etiology of RM based on the LHB/CGB gene family and highlighted the contribution of distinct common CNVs and global burden of structural variants in shaping the genetic predisposition to recurrent early pregnancy loss.
REFERENCES


Mishra AK, Mahale SD, Iyer KS. 2003. Disulfide bonds Cys(9)-Cys(57), Cys(34)-Cys(88) and Cys(38)–Cys(90) of the beta-subunit of human chorionic gonadotropin are crucial for heterodimer formation with the alpha-subunit: experimental evidence.
for the conclusions from the crystal structure of hCG. *Biochim Biophys Acta* **1645**(1): 49–55.


Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A et al. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320(5875): 539–543.


SUMMARY IN ESTONIAN

Korduva raseduse katkemise genoomsed ja geneetilised riskitegurid


Inimese koorioni gonadotropiini (human chorionic gonadotropin, hCG), mida kutsutakse ka ‘raseduse hormooniks’, toodab blastotsüst juba 8-rakulisest staadiumis ning selle olemasolu on embrüo pesastumiseks ja raseduse säilimiseks kriitilise tähtsusega. Heterodimeerse hCG spetsiifilise määrab β-aláühik, mille normist kõgem või madalam avaldamine on tuvastatud nii emakavatise raseduse, moolrasusede kui ka idiopaatilise KRK puhul. See on β-aláühikut kodeerivad geenid ja neis esinev geneetiline varieeruvus huvipakkukv kui potentsiaalne KRK riskitegur. hCGβ aláühikut kodeerivad neli duplikaat-geeni (CGB, CGB7 ning kõige kõrgema avaldumisega CGB5 ja CGB8), mis asuvad ühes geeniklasteris 19. kromosoomil koos lutropiini β-aláühikut tootva (LHB) ja kahe hCGβ aláühikut mitte-kodeeriva (CGB1 ja CGB2) geeniga. Siiani ei ole teostatud põhjalikke hCGβ geenide varieeruvuse uuringuid KRK kontekstis valdavalt metoodiiliste takistuste tõttu, mis lähtuvad LHB/CGB geenide sarnasusest DNA (97–99%) kui valgu (98–100%) tasemel.

DNA koopiaarvu varieeruvus (copy number variants, CNV), mis hõlmab DNA struktuursete ümberkorraldusi suurusega üle 50 aluspaari, on laialdaselt fenomen võimalikest 30% inimeses kogu genoomi järjestusest. Üksikud CNV põhjustasid kui ka CNV-de summas seriklastust inimese genoomis on
seostatud mitmete haiguste patogeneesiga nagu näiteks autism, skisofreenia, ülekaalulisus ja pikaealisus. Inimese reproduktiivhaiguste kontekstis on CNV-sid vähe uuritud ning idiopaatilise KRK puhul on CNV-de profiili analüüsi varasemalt teostatud ainult 24 aborteerunud platsenta proovis. Suuremahulise platsenta-põhise CNV-de analüüsi limiteerivaks faktoriks KRK puhul, mis leia aset varajase raseduse käigus, on aga kvaliteetse bioloogilise materjali raskendatud kättesaadavus. Kuna loote genoomi panustavad ema ja isa võrdselt, siis on alternatiivina võimalik hinnata vanemate geneetilist eelsoodumust KRK esinemiseks verst era-aldatud DNA põhjal.

Käesoleva doktoritöö kirjanduse ülevaates on koondatud info tänaseks teadaolevatest idiopaatilise KRK riskiteguritest ning geneetilistest assotsiatsioonijõududest, mis on teostatud nii kogu genoomi kui ka üksikute markerite tasemel. Lisaks on kirjeldatud hCG hormooni, hCGβ alaluigid, CGB geenide funktsionaalseid ja struktuurseid omadusi ning ühtlasi CNV-de rolli nii kompleks-haiguste kui ka rasedustüüsituste kujunemisel. Kuna loote genoomi panustavad ema ja isa võrdselt, siis on alternatiivina võimalik hinnata vanemate geneetilist eelsoodumust KRK esinemiseks verst era-aldatud DNA põhjal.

Doktoritöö eksperimentaalses osas on püstitatud järgmised eesmärgid:

I. **LHB/CGB geeniklaster** kaardistamine ning detailne DNA järjestuse analüüs eesmärgiga tuvastada tegureid, mis mõjutavad piirkonna struktuurset ja evolutsioonilist dünaamikut

II. DNA ühenukleotiidised variandid kõige rohkem hCG β-alaluikut tootvates geenides CGB5 ja CGB8 ning nende seos KRK esinemisega:
- DNA ühenukleotiidsete variandid tuvastamine CGB5 ja CGB8 geenides ning juht-kontrolluuringu teostamine Eesti, Soome ja Taani valimistes
- CGB5 ja CGB8 geenides esinevate mitte-sünonüümsete mutatsioonide struktuur ja funktsionaalne analüüs in vitro

III. DNA koopiaarvu varieeruvus (CNV) kui uus KRK geneetiline marker:
- CNV-de genoomse profiili analüüs ning CNV-de poolt häiritud bioloogiliste radade tuvastamine
- Uute CNV piirkondade identifitseerimine, mis suurendavad KRK kujunemise riski

Uurimistöö peamised tulemused on järgmised:

**Ühenukleotiidete geneetilise markerite roll KRK esinemisel LHB/CGB geeniklastri näitel**

1. **LHB/CGB geeniklastri** kasutamine, mis koosneb ühest LHB ja kuuest CGB geeniklastri inimesel, on kujunenud korduvate DNA ümberkorralduste tagajärjel, mida on tõenäoliselt soodustanud aktiivsed rekombinatsiooni alad piir-
konnas. LHB/CGB geenide varieeruvust on mõjutanud ühelt poolt geenikonversioon, mis levitab polümorfisme geenikoopiate vahel, ning teisalt funktsionaalsed piirangud, mis garantieerivad hCGβ geenide aktiivse avaldumise. Arvestades genoomse piirkonna dünaamikat võib eeldada, et LHB/CGB geeniklaster on ka käesolevalt struktuuri ja varieeruvuse poolest ebastabiilne.

2. Kõige rohkem hCGβ-t tootvatele geenidele CGB5 ja CGB8 on iseloomulik riskivariantide puudumine Eesti, Soome ja Taani KRK-ga paatsientidel. CGB5 promootor-allas esineb neljast polümorfismist koosnev motivi, mis omab protektiivset efekti KRK esinemise suhtes ning mis on üle kandunud vastavast CGB8 promootori piirkonnast geenikonversiooni teel. CGB8 on omakorda kõige efektiivsem hCG β-alaühikut produseeriv geenikoopia ning selle promootor-ala on balansseeriva valiku all.


Globaalsete geneetiliste markerite roll KRK esinemisel CNV-de genoomse profilili näitel

1. CNV-de genoomne kaardistamine tuvastas, et KRK paatsientidel on CNV-de summaarne ulatus ja pikkade (>100 kb) geenirikaste deleksioonide osakaal genoomi kohta suurem kui viljakatel kontrollidel. Ulatuslik CNV-de esinemine genoomis suurendab tõenäosust, et ümberkorraldused hõlmavad geene, mis on olulised varajas raseduse etapis.

2. KRK paatsientidel esinevat CNV-de poolt ümberkorraldatud geenide funktsionaalse rikastatuse analüüsil ilmnes, et struktuured ümberkorraldused mõjutavad kõige enam bioloogilisi radasid, mis on seotud immuunvastuse väljakujunemisega platsentas. Tulemused kinnitavad, et emapoolse balansseeritud immuunvastuse tõrkenete loote poolt esitletud isa antigeenidele on määrava tähistusega roll edukaks raseduse kulgemiseks.

3. Enamlevinud CNV piirkondade-põhine uuring Eesti ja Taani valimites tuvastas uue geneetilise markeri, mille esinemisega kasnevab suurem KRK risk naispartneritel. Mitme koopiana esinev duplikatsioon 5p13.3 kromosaalses piirkonnas hõlmab kahte geeni PDZD2 ja GOLPH3, mis on kõige
aktiivsemalt avaldunud platsentas ning mida pole varasemalt seostatud varajase rasedusega.

Vaatamata laialdastele uuringutele KRK geneetiliste põhjuste selgitamiseks, ei ole siiani tuvastatud haiguse-spetsiifilisi markereid. Raseduse katkemise riski oluliselt mõjutavate geenide puhul, nagu seda on hCGβ duplikaat-geenid, on ilmselt põhjustatud loote eluõulisust oluliselt härrivate geneetiliste variandite evolutsioon negatiivse valiku surve all. Üha kasvava andmistiku põhjal on pakutud, et üksikute suure efektiaga mutatsioonide ja polümorfsete markerite asemel võivad KRK kujunemist mõjutada väikese efektiiga geneetiliste variantide kombinatsioonid või esineb igal KRK perekonnal oma unikaalne riskitegur.

DNA koopiaarvu varieeruvuse analüüs kogu genoomi tasemel on võimaldanud hüpoteesi-vaba lähenemist ja uute KRK-seoseliste lookuste tuvastamist ning on kinnitanud DNA struktuursete ümberkorralduste rolli varajase raseduse katkemisel. Ühtlasi rõhutavad käsioleva doktoritöö raames esitatud tulemused KRK geneetilise etioloogia heterogeensust.
ACKNOWLEDGEMENTS

Out of clutter, find simplicity,
From discord, find harmony,
In the middle of difficulty,
lies opportunity.

Albert Einstein

On the path towards this PhD thesis, occasionally paved with clutter and discord, I owe many thanks to my supervisor, Professor Maris Laan, for shedding light on opportunities, inspiring with countless ideas and for providing a multitude of interesting research objectives.

I am grateful to all my wonderful colleagues throughout the years for always providing help and advice whenever needed. My special thanks go to the girls from the ‘Girls’ room’, Marina Grigorova and Laura Kasak, but also to former members of our research group and my dear friends Liis Uusküla-Reimand and Kärt Tomberg for sharing the PhD experience and for being there despite the distance.

The research of my thesis is predominantly focusing on human genetics with a touch of medical sciences that I cannot claim to be expert of. Thus I acknowledge the most capable multi-tasker Dr. Kristiina Rull for all the discussions and insights on the medical aspects of scientific research. And for being true to her profession by persistently encouraging us not to waste time with scientific degrees and become pregnant.

Within the framework of my PhD studies, I was presented with an opportunity to work in Professor Ilpo T. Huhtaniemi’s research group in the Institute of Reproductive and Developmental Biology, Imperial College London, in the vibrant city of London. I am grateful to Professor Huhtaniemi for this valuable experience and appreciate Hellevi Peltoketo, Kim C. Jonas and other group members for all the help and know-how.

For the last but not least, I owe everything to the pillars of my life – my family and closest friends who have always supported me in everything and anything.

This work was financially supported by personal stipends from Kristjan Jaak Stipend Program, Graduate School in Biomedicine and Biotechnology, Ernst Jaakson memorial fund, Seitsmenda Samba Fond of University of Tartu, Margot M. ja Herbert R. Linn stipend fund and Estonian Students Fund, USA.
PUBLICATIONS
CURRICULUM VITAE

Name: Liina Nagirnaja
Date of birth: 24.12.1980
Address: Institute of Molecular and Cell Biology
Riia 23, Tartu 51010
E-mail: nagir@ut.ee
Current position: Researcher, Institute of Molecular and Cell Biology

Education:
2005–2012 PhD student in gene technology, Human Molecular Genetics research group, Institute of Molecular and Cell Biology, University of Tartu
2003–2005 MSc studies in gene technology, Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu
1999–2003 BSc studies in biology (cum laude), University of Tartu
1988–1999 Miina Härma Gymnasium, Tartu

Working Experience:
Since 2011 Researcher, Institute of Molecular and Cell Biology
2009–2010 Visiting researcher, Institute of Reproductive and Developmental Biology, Imperial College London, United Kingdom
2005–2010 Instructor of course “Genetics and evolution”, The Gifted and Talented Development Centre, University of Tartu
2008 Specialist, Estonian Biocentre
2006–2009 Committee member, Estonian Junior Science Olympiad
2006–2007 Member of the Faculty of Science and Technology Council, University of Tartu
2005–2008 Committee member, Estonian Biology Olympiad
2002 Administration officer in the Reception Committee, University of Tartu

Main Fields of Research:
The role of DNA single nucleotide variants in the human luteinizing hormone and chorionic gonadotropin β-subunit (LHB/CGB) gene family and the contribution of DNA copy number variants (CNVs) in modifying the genetic predisposition to recurrent miscarriages.
**Publications:**


**Scholarships and Awards:**

2012 Kristjan Jaak stipend, Archimedes Foundation
2011 Runner-up for The Elsevier *Trophoblast Research* Award 2011 at “IFPA (International Federation of Placenta Associations) Meeting 2011”, Geilo, Norway
2011 Kristjan Jaak stipend, Archimedes Foundation
2010 Kristjan Jaak stipend, Archimedes Foundation
2009 Estonian Students Fund, USA
2009 Ernst Jaakson memorial fund scholarship, Tartu University Foundation
2007 Kristian Jaak stipend, Archimedes Foundation
2006 Stipend of Seitismenda Samba Fond, University of Tartu
2006 Stipend of Margot M. ja Herbert R. Linn, Estonian World Council
2005  II Prize for MA thesis, Estonian National Contest for Young Scientists at university level
2005  Stipend of Tartu Raefond, University of Tartu
2005  Kristjan Jaak stipend, Archimedes Foundation
2004  Harald Raudsepp’s Stipend of Estonian-Revelia Academic Fund
2003  Diploma for B.Sc. thesis, National Contest of Students’ Scientific Research

Other Scientific Activities:
Since 2001  Member of the Estonian Society of Human Genetics

Supervised dissertations:
2010–2012  Laura Kasak, MSc studies, Institute of Molecular and Cell Biology
Since 2012  Diana Nõmmemees, BSc studies, Institute of Molecular and Cell Biology

Interests:
Since 2005  Member of Mountaineering Club “Firn”
Nimi: Liina Nagirnaja
Sünniaeg: 24.12.1980
Aadress: TÜ Molekulaar- ja Rakubioloogia Instituut Riia 23, Tartu 51010
E-post: nagir@ut.ee
Praegune töökoht: Teadur, TÜ Molekulaar- ja Rakubioloogia Instituut

Haridus:
2005–2012 PhD õpingud, geenitehnoloogia eriala, Inimese Molekulaargeneetika töögrupp, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
2003–2005 MSc (teadusmagister), geenitehnoloogia eriala, Biotehnoloogia õppetool, Molekulaar- ja Rakubioloogia Instituut, Tartu Ülikool
1999–2003 BSc bioloogia eriala (cum laude), Tartu Ülikool
1988–1999 Miina Härma Gümnaasium, Tartu

Töökogemus:
Alates 2011 Teadur, Molekulaar- ja Rakubioloogia Instituut
2009–2010 Küülisteadlane, Institute of Reproductive and Developmental Biology, Imperial College London, London, Suurbritannia
2005–2010 TÜ Teaduskooli kursuse „Geneetika ja evolutsioon” koordinaator
2008 Spetsialist, Eesti Biokeskus
2007–2008 Loodusteaduste Olümpiaadi rahvusvahelise vooru Eesti delegatsiooni liige (2007 a., Taipei, Taiwan; 2008 a., Changwon, Lõuna-Korea)
2006–2009 Loodusteaduste Olümpiaadi žürii liige
2006–2007 Tartu Ülikooli Bioloogia-geograafiateaduskonna Nõukogu liige
2005–2008 Eesti Bioloogiaolümpiaadi žürii liige
2002 Üliõpilaste vastuvõtu korraldaja, Tartu Ülikool

Peamised uurimisvaldkonnad:
Inimese luteiniseeriva hormooni ja koorion gonadotropiini beeta-subühiku geeniperekonna ning DNA struktuursete ümberkorralduste potsiaalne seotus korduvate spontaanabortidega.
Publikatsioonide loetelu:

Saadud stipendiumid ja auhinnad:
2012 Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes
2011 Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes
2010 Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes
2009 Eesti Üliõpilaste Toetusfond, USA
2009 Ernst Jaaksoni mälestusfondi stipendium, TÜ Sihtasutus
2007 Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes
2006 TÜ Seitsmenda Samba Fondi stipendium
2006 Margot M. ja Herbert R. Linna stipendium, Ülemaailmne Eesti Kesk-nõukogu
2005  Il preemia, „Üliõpilaste teadustööde riiklik konkurss”, Haridus- ja Teadusministeerium/Sihtasutus Archimedes
2005  Tartu Raeфонdi stipendium, TÜ Sihtasutus
2005  Kristjan Jaagu nimeline stipendium, Sihtasutus Archimedes
2004  Harald Raudsepa nimeline stipendium, Estonian-Revelia Academic Fund
2003  Diplom, „Üliõpilaste teadustööde riiklik konkurss”, Haridus- ja Teadusministeerium

Muu teaduslik organisatsiooniline ja erialane tegevus:
Alates 2001  Eesti Inimesegeneetika Ühingu liige

Juhendatud väitekirjad:
2010–2012  Laura Kasak, MSc väitekiri, TÜ Molekulaar- ja Rakubioloogia Instituut
Alates 2012  Diana Nõmmemeses, BSc õpingud, TÜ Molekulaar- ja Rakubioloogia Instituut

Huvialad
Alates 2005  Alpiklubi „Firn” liige.


87. **Aare Abroi.** The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004. 135 p.
100. **Ilmar Tõnno.** The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N₂ fixation in some Estonian lakes. Tartu, 2004. 111 p.
113. **Riho Teras.** Regulation of transcription from the fusion promoters generated by transposition of Tn4652 into the upstream region of pheBA operon in *Pseudomonas putida*. Tartu, 2005. 106 p.
118. **Silja Kuusk.** Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006. 126 p.
120. **Lea Tummeleht.** Physiological condition and immune function in great tits (*Parus major l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006. 94 p.
<table>
<thead>
<tr>
<th>Number</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>Priit Jõers</td>
<td>Studies of the mitochondrial helicase Hmi1p in <em>Candida albicans</em> and <em>Saccharomyces cerevisia</em>. Tartu, 2006. 113 p.</td>
<td>2006</td>
<td>113</td>
</tr>
<tr>
<td>132</td>
<td>Tatjana Krama</td>
<td>Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007. 112 p.</td>
<td>2007</td>
<td>112</td>
</tr>
<tr>
<td>133</td>
<td>Signe Saumaa</td>
<td>The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in <em>Pseudomonas putida</em>. Tartu, 2007. 172 p.</td>
<td>2007</td>
<td>172</td>
</tr>
<tr>
<td>134</td>
<td>Reedik Mägi</td>
<td>The linkage disequilibrium and the selection of genetic markers for association studies in European populations. Tartu, 2007. 96 p.</td>
<td>2007</td>
<td>96</td>
</tr>
<tr>
<td>136</td>
<td>Anu Albert</td>
<td>The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007. 95 p.</td>
<td>2007</td>
<td>95</td>
</tr>
<tr>
<td>140</td>
<td>Lauri Laanisto</td>
<td>Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008. 133 p.</td>
<td>2008</td>
<td>133</td>
</tr>
<tr>
<td>141</td>
<td>Reidar Andreson</td>
<td>Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008. 105 p.</td>
<td>2008</td>
<td>105</td>
</tr>
</tbody>
</table>
154. **Marge Starast.** Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
158. **Kaja Kiiver.** Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepacviruses. Tartu, 2009, 104 p.
166. **Pille Säälik.** The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
171. **Maarja Adojaan.** Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.

207


