GENETIC STUDIES ON RECURRENT MISCARRIAGE

Milja Kaare

Folkhälsan Institute of Genetics
and
Department of Medical Genetics
University of Helsinki,
Helsinki, Finland

ACADEMIC DISSERTATION

To be publicly discussed, with the permission of
the Medical Faculty of the University of Helsinki,
in Biomedicum Helsinki, Haartmaninkatu 8, lecture hall 2,
on February 13th 2009, at 1 pm.

Helsinki 2009
Supervised by

Docent Kristiina Aittomäki, M.D., Ph.D.
Department of Clinical Genetics,
Helsinki University Central Hospital
Folkhälsan Institute of Genetics and
Department of Medical Genetics,
University of Helsinki

Reviewed by

Docent Leena Anttila, M.D., Ph.D.
Väestöliitto Fertility Clinics Ltd, Turku Clinic
Department of Obstetrics and Gynecology,
University of Turku

Docent Nina Horelli-Kuitunen, Ph.D., medical geneticist
Medix Laboratories Ltd, Espoo

Official opponent

Professor Seppo Heinonen, M.D., Ph.D.
Department of Obstetrics and Gynecology
University of Kuopio and
Kuopio University Hospital

ISBN 978-952-10-5259-0 (PDF)
http://ethesis.helsinki.fi
Yliopistopaino
Helsinki 2009
Contents

List of original publications ........................................................................................................ 6
Abbreviations .................................................................................................................................. 7
Abstract ........................................................................................................................................ 9
Introduction .................................................................................................................................. 11
Review of the literature .................................................................................................................. 12
1 Early human embryonic development ...................................................................................... 12
   1.1 Fertilization ......................................................................................................................... 12
   1.2 Implantation ......................................................................................................................... 12
   1.3 Placentation ........................................................................................................................ 14
   1.4 Gastrulation and organogenesis .......................................................................................... 14
2 Recurrent miscarriage ................................................................................................................ 15
   2.1 Identifiable causes of miscarriage ...................................................................................... 16
      2.1.1 Risk factors for miscarriage ........................................................................................... 19
   2.2 Genetic factors causing miscarriage ................................................................................... 19
      2.2.1 Chromosomal abnormalities ......................................................................................... 20
      2.2.2 Skewed X chromosome inactivation ............................................................................. 21
      2.2.3 Y chromosome microdeletions ..................................................................................... 21
      2.3.4 Gene mutations and polymorphisms ............................................................................ 22
      2.2.5 Aberrant placental gene expression .............................................................................. 24
      2.2.6 Epigenetic mechanisms ................................................................................................ 25
3 Identification of genetic causes underlying RM ..................................................................... 26
   3.1 Identification of disease genes ............................................................................................ 27
      3.1.1 The Human Genome Project ......................................................................................... 27
      3.1.2 Traditional identification of disease genes ...................................................................... 28
      3.1.3 The candidate gene approach ....................................................................................... 28
      3.1.4 Mutation screening ........................................................................................................ 29
      3.1.5 Confirming the nature of a new genetic variation .......................................................... 30
   3.2 Mouse as a model organism ............................................................................................... 30
   3.3 Candidate genes for RM ..................................................................................................... 32
      3.3.1 Amnionless ..................................................................................................................... 32
      3.3.2 Thrombomodulin and Endothelial Protein C Receptor .................................................. 35
      3.3.3 p53 .................................................................................................................................. 38
      3.3.4 The mitochondrial genome ............................................................................................. 39
Aims ............................................................................................................................................... 43
Materials and methods ............................................................................................................... 44
1 Study subjects ........................................................................................................................... 44
   1.1 Patients (I, II, III, IV, V) ...................................................................................................... 44
   1.2 Controls (I, II, III, IV, V) ...................................................................................................... 44
List of original publications

This thesis is based on the following publications, which are referred to in the text by their Roman numerals. In addition, some unpublished results are presented.


* These two authors contributed equally to this work.

Publication II also appears in the thesis of Veli-Matti Ulander (2007)

The articles are reprinted with the permission of their copyright holders.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>angiotensin I-converting enzyme</td>
</tr>
<tr>
<td>AMN</td>
<td>amnionless</td>
</tr>
<tr>
<td>APC</td>
<td>activated protein C</td>
</tr>
<tr>
<td>APP</td>
<td>amyloid beta precursor protein</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>AZF</td>
<td>azoospermia factor</td>
</tr>
<tr>
<td>BAD</td>
<td>BCL2 antagonist of cell death</td>
</tr>
<tr>
<td>BAX</td>
<td>BCL2-associated X protein</td>
</tr>
<tr>
<td>BID</td>
<td>BH3-interacting domain death agonist</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CO2</td>
<td>Cytochrome c oxidase subunit II</td>
</tr>
<tr>
<td>cpm</td>
<td>count per minute</td>
</tr>
<tr>
<td>CS</td>
<td>carnegie stage</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>cytotoxic T lymphocyte-associated 4</td>
</tr>
<tr>
<td>CUBN</td>
<td>cubilin</td>
</tr>
<tr>
<td>CytB</td>
<td>cytochrome b</td>
</tr>
<tr>
<td>dHPLC</td>
<td>denaturing high performance liquid chromatography</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>E</td>
<td>embryonic day</td>
</tr>
<tr>
<td>EAA</td>
<td>European Academy of Andrology</td>
</tr>
<tr>
<td>EMQN</td>
<td>European Molecular Genetics Quality Network</td>
</tr>
<tr>
<td>EPCR</td>
<td>endothelial protein C receptor</td>
</tr>
<tr>
<td>ESE</td>
<td>exonic splicing enhancer</td>
</tr>
<tr>
<td>FasL</td>
<td>Fas ligand</td>
</tr>
<tr>
<td>FSH-RO</td>
<td>follicle stimulating hormone-resistant ovaries</td>
</tr>
<tr>
<td>HGP</td>
<td>human genome project</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>ICM</td>
<td>inner cell mass</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>interferon-γ</td>
</tr>
<tr>
<td>IGS</td>
<td>Imerslund-Gräsbäck syndrome</td>
</tr>
<tr>
<td>IL6/10</td>
<td>interleukin 6/10</td>
</tr>
<tr>
<td>ITGB3</td>
<td>integrin beta-3</td>
</tr>
<tr>
<td>MDM2</td>
<td>mouse double minute 2</td>
</tr>
<tr>
<td>MELAS</td>
<td>mitochondrial encephalopathy with lactic acidosis and stroke-like episodes</td>
</tr>
<tr>
<td>MERRF</td>
<td>myoclonic epilepsy and ragged-red fibres</td>
</tr>
<tr>
<td>MLPA</td>
<td>multiplex ligation-dependent probe amplification</td>
</tr>
<tr>
<td>MLS</td>
<td>microphthalmia with linear skin defects syndrome</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>MTHFR</td>
<td>methylenetetrahydrofolate reductase</td>
</tr>
<tr>
<td>MUC1</td>
<td>mucin 1</td>
</tr>
<tr>
<td>ND2</td>
<td>NADH dehydrogenase subunit 2</td>
</tr>
<tr>
<td>OFD1</td>
<td>oral-facial-digital syndrome type 1</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>PAI-1</td>
<td>plasminogen activator inhibitor 1</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PGC</td>
<td>primordial germ cells</td>
</tr>
<tr>
<td>PGD</td>
<td>preimplantation genetic diagnosis</td>
</tr>
<tr>
<td>Q-PCR</td>
<td>quantitative real-time PCR</td>
</tr>
<tr>
<td>PP14</td>
<td>placental protein 14</td>
</tr>
<tr>
<td>RM</td>
<td>recurrent miscarriage</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SIFT</td>
<td>sorting intolerant from tolerant</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SR-protein</td>
<td>Ser/Arg-rich-protein</td>
</tr>
<tr>
<td>STS</td>
<td>sequence tagged sites</td>
</tr>
<tr>
<td>Th-1</td>
<td>type 1 T-helper cell</td>
</tr>
<tr>
<td>TM</td>
<td>thrombomodulin</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>VEGF</td>
<td>vascular endothelial growth factor</td>
</tr>
<tr>
<td>VTE</td>
<td>venous thromboembolism</td>
</tr>
<tr>
<td>XCI</td>
<td>X chromosome inactivation</td>
</tr>
</tbody>
</table>
Abstract

Recurrent miscarriage (RM) is defined as three or more consecutive pregnancy failures and is estimated to affect ~1% of couples trying to conceive. The cause of RM remains unknown in approximately 50% of cases although much work has been done in the past to identify the underlying mechanisms. In this study, it was hypothesized that some of the underlying factors yet to be discovered are genetic.

The aim of this study was to obtain new information on genetic causes of RM. The aim was to search for mutations in genes known to cause miscarriage in animal models and thereby find new genetic causes for unexplained miscarriages in humans. Four candidate genes – *Amnionless* (*AMN*), *Endothelial protein C receptor* (*EPCR*), *Thrombomodulin* (*TM*), and *p53* – were chosen for the study as they have all been considered to be crucial for normal embryonic development in the mouse. In addition, the mitochondrial genome was studied because mitochondria are involved in processes important in early development. Furthermore, sex chromosome characteristics suggested to underlie miscarriage were also studied.

A total of 40 couples and 8 women with unexplained RM were collected for this study. The patients were screened for mutations in the *AMN*, *TM* and *EPCR* genes using dHPLC. Six interesting exonic and potential splice site disrupting variations were detected in these genes, namely c.363G>A, c.843+11C>T, c.1339_1344dup, and c.1170-6C>T in *AMN*, c.1728+23_+40del in *TM*, and c.323-9_336dup in *EPCR*. Two of these variations, c.363G>A and c.843+11C>T were more frequent in RM women compared to controls. However, their phenotypic effects cannot be determined without further investigations. While most variations in the *AMN*, *TM* and *EPCR* genes were identified in both patients and controls and are likely to be silent polymorphisms, others may have functional significance or modifier effects in combination with mutations or polymorphisms in other genes.

An association between the C11992A polymorphism of the *p53* gene and RM was detected. The results indicate that women carrying the C/A or A/A genotype have a twofold higher risk for RM than women with a C/C genotype. This strengthens the results of previous studies reporting that *p53* sequence variations may cause miscarriage. The role
of variation C11992A in embryonic development is, however, difficult to predict without further studies. The role of p53 in RM was also investigated through expression analysis of p53 in placental tissue. The results showed normal expression of p53 in the samples studied.

When screening the mitochondrial genome a heteroplasmic mtDNA variation was found in 13 (27%) RM women and 9 (19%) control women, which was unexpected as heteroplasmic variations are reported to be rare. One novel variation within the coding regions of the mitochondrial genome was detected. In addition, 18 previously reported polymorphisms were identified. Although the detected variations are likely to be neutral polymorphisms, a role in the aetiology of miscarriage cannot be excluded as some mtDNA variations may be pathogenic only when a threshold is reached.

Recent publications have reported skewed X chromosome inactivation and Y chromosome microdeletions to be associated with RM. Therefore, these sex chromosome abnormalities in the context of RM were investigated. No association between skewed X chromosome inactivation or Y chromosome microdeletions and RM in the Finnish patients studied were detected.

In addition, data on ancestral birthplaces of the patients were collected to study any possible geographic clustering, which would indicate a common predisposing factor. The results showed clustering of the birthplaces in eastern Finland in a subset of patients. This suggests a possibility of an enriched susceptibility gene which may contribute to RM. These results encourage further studies to determine new genetic factors contributing to the aetiology of RM.
Introduction

Miscarriages are the most common complication of pregnancy, affecting approximately 15% of all clinically recognized pregnancies in the general population. The exact frequency of miscarriages is, however, unknown as miscarriages frequently occur before the woman is aware of her pregnancy. It is estimated that more pregnancies are lost spontaneously than are actually carried to term (Rai and Regan, 2006; Stephenson and Kutteh, 2007). Most of the miscarriages are sporadic and non-recurrent, and are often caused by chromosome abnormalities in the foetus (Dhont, 2003). Recurrent miscarriage (RM), defined as three consecutive pregnancy failures, is estimated to affect ~1% of all couples trying to conceive (Li et al., 2002; Rai and Regan, 2006).

There are numerous factors that may cause RM, but the underlying problem often remains undetected. Although much work has been done to identify the underlying mechanisms, the cause of miscarriage can be identified in only ~50% of cases. The known causes of RM include chromosomal and metabolic abnormalities, uterine anomalies, and immunologic factors (Li et al., 2002; Rai and Regan, 2006). Even though RM is a heterogeneous condition and the progress in identifying causative factors has been slow, the repetitive pregnancy losses in some couples and the high percentage of unexplained RM indicate that there are specific underlying causes yet to be found.

This study was conducted to gain further knowledge concerning genetic causes of RM. Identification of the underlying factors is crucial for the development of more successful treatment and improvement of the outcome of future pregnancies in women experiencing RM. Genetic factors causing RM are, however, difficult to study because the foetus is lost at an early stage of development and is therefore difficult to examine. Consequently, most of the studies conducted on RM in general, and in this thesis work, are based on studying the couples experiencing the miscarriages. We hypothesized that several genetic factors such as mutations in genes, either identified from animal models or regulating functions important in different stages of early foetal development, or abnormalities in the sex chromosomes may underlie the miscarriages in the couples studied.
Review of the literature

1 Early human embryonic development

1.1 Fertilization

The development of a human being starts from a single cell, a zygote, which results from an ovum being fertilized by a spermatozoon. Fertilization occurs in the fallopian tube within 48 hours of ovulation. After fertilization, a series of symmetrical cell divisions create a mass of 12 to 16 totipotent cells, the morula, which is still enclosed within the zona pellucida (Norwitz et al., 2001; Sariola et al., 2003). The morula enters the uterine cavity approximately three days after fertilization. The appearance of a fluid filled inner cavity within the mass of cells marks the transition from a morula to a blastocyst and is accompanied by cellular differentiation (Norwitz et al., 2001; Wang and Dey, 2006). The first differentiation event gives rise to trophoblasts, specialized epithelial cells that give rise to extraembryonic structures, including the placenta. The remaining cells segregate at one pole of the embryo to form the inner cell mass (ICM) (Figure 1). Within three days of entering the uterine cavity, the embryo hatches from the zona pellucida, thereby exposing its trophoblast cells, which enables implantation (Norwitz et al., 2001; Sariola et al., 2003; Wang and Dey, 2006).

1.2 Implantation

The newly formed zygote has a discrete time frame in which it must prepare itself for implantation. Preparation begins upon oocyte fertilization and takes approximately 6 to 7 days. During this time successive cell division occurs, and the outer trophoblast cells of the blastocyst differentiate into cytotrophoblasts which diverge into villous and invasive subtypes forming the building blocks of the placenta (Norwitz et al., 2001; Vitiello and Patrizio, 2007). At the same time, when the blastocyst prepares itself for implantation, the uterus becomes receptive. Uterine receptivity is defined as the state during the period of endometrial maturation and proliferation when the blastocyst can become implanted. Multiple signals synchronize the development of the blastocyst and the preparation of the uterus. Of the many aspects of the synchronization process, the role of steroid hormones is
the best understood. Peptide hormones, growth factors, and cytokines also have roles in the process. The cascade of signalling events that occur in both foetal and maternal tissues establishes an appropriate environment critical to the development and survival of the foetus (Norwitz et al., 2001; Wang and Dey, 2006; Vitiello and Patrizio, 2007).

Implantation occurs in three stages. The first stage is the initial adhesion of the blastocyst to the uterine wall. The next stage is characterized by increased physical interaction between the blastocyst and the uterine epithelium. Shortly thereafter invasion begins and syncytiotrophoblasts, the outer layer of the trophoblasts, penetrate the uterine epithelium (Wang and Dey, 2006; Chavatte-Palmer and Guillomot, 2007; Vitiello and Patrizio, 2007). By the 10th day after conception, the blastocyst is completely embedded in the stromal tissue of the uterus as the uterine epithelium has regrown to cover the site of implantation (Norwitz et al., 2001). The interaction between an activated blastocyst and a receptive uterus is part of a complex process that leads to implantation and the early stages of placental development. Failure to synchronize the processes involved in these interactions results in failure of implantation (Cross et al., 1994; Wang and Dey, 2006; Chavatte-Palmer and Guillomot, 2007).

Figure 1. The early stages of human development. A) An unfertilized egg cell is fertilized in the fallopian tube by a sperm cell leading to B) a fertilized cell containing both the maternal and paternal pronuclei. The cell undergoes cleavage to C) the 2-cell stage, D) the 4-cell stage and E) the 8-cell stage, which is reached two to three days after fertilization. F) The morula enters the uterine cavity approximately three days after fertilization. G) Implantation of the blastocyst, consisting of a blastocyst cavity, trophoblast cells and the inner cell mass, into the uterine epithelium is initiated around seven days after fertilization. Modified from Wang and Dey, 2006.
1.3 Placentation

The placenta is the first organ to form during mammalian embryogenesis. The placenta is a vital organ without which the embryo cannot survive in the uterine environment. The function of the placenta during early gestation is primarily to mediate implantation of the embryo into the uterus. After implantation, the major function of the placenta is to mediate, as well as to regulate, nutrient uptake from the mother to the foetus. It also establishes the interface for gas exchange between the maternal and foetal circulation. Another important function of the placenta is the regulation of the maternal immune response so that the foetal semi-allograft is tolerated during pregnancy. The placenta also acts as an important source of hormones and growth factors that are needed for the initial maternal recognition of pregnancy (Cross et al., 1994; Cross, 2006; Sadler et al., 2006).

Because the placenta is critical for survival it is very sensitive to disruption. Genetic or environmental factors that affect the development of the placenta are associated with poor pregnancy outcome. Abnormal expression of specific regulatory genes in the placenta and abnormal functions of imprinted genes may cause placental dysfunction (Coan et al., 2005). Abnormalities in placental functions can lead to a variety of problems including implantation failure, placental insufficiency, foetal growth retardation and embryonic death. Complications that become apparent relatively late in pregnancy may actually reflect errors that occurred already during placental development (Cross et al., 1994; Rossant and Cross, 2001; Cross, 2006).

1.4 Gastrulation and organogenesis

While the syncytiotrophoblasts start to penetrate into the wall of the uterus during implantation, the embryoblast/ICM also continues developing. The embryoblast forms a bilaminar embryo, composed of the epiblast and the hypoblast. The hypoblast forms the yolk sac. The epiblast undergoes gastrulation at approximately day 16 following fertilization. Gastrulation establishes the three germ layers—the endoderm, ectoderm, and mesoderm—all of which will give rise to various organ systems. The mesoderm also interacts with the trophoblast tissue to form the umbilical cord (Cross et al., 1994; Sadler et al., 2006). At this point the actual embryonic phase starts and lasts from the third week until the eighth week following fertilization. The organ systems differentiate at greatly varying rates during this phase. For example, the circulatory system is largely functional at
the end of this period, whereas the nervous system continues to undergo massive cell division and is only beginning to establish functional connections (Sadler et al., 2006). The remainder of human development, from weeks nine to thirty-eight, is called the foetal period. During this period the embryo acquires its human appearance. The foetal period is characterized by rapid growth and continued tissue and organ differentiation (Sariola et al., 2003; Sadler et al., 2006).

2 Recurrent miscarriage

Human reproduction entails a fundamental paradox: although it is critical for the survival of the species, the process is relatively inefficient. Maximal fecundity (the probability of fertilization during one menstrual cycle) is approximately 30%. In addition, although much research and many advances have been made in reproductive medicine during the past decades, miscarriage, the spontaneous loss of a pregnancy before the foetus reaches viability, remains the most common complication of pregnancy (Clark et al., 2001; Pandey et al., 2004).

There are two types of miscarriage; sporadic and recurrent. In the general population ~15% of all clinically recognized pregnancies end in miscarriage (Stray-Pederson and Stray-Perderson, 1984; Nybo Andersen et al., 2000). However, the rate of miscarriage may be as high as 50-60% if the undetected, very early pregnancy losses occurring within the first weeks following gestation are taken into account (Roberts and Lowe, 1975; Wilcox et al., 1988). Of the pregnancies that are lost, ~70-75% are caused by failures of implantation or early placentation and are therefore not clinically recognized as pregnancies. As a consequence, only about half of all conceptions advance into the second trimester of pregnancy (Figure 2) (Rai and Regan, 2006).

RM is estimated to occur in 1-2% of all couples (Tulppala et al., 1993; Katz and Kuller, 1994), but the exact prevalence of RM is dependent on its definition. RM is often defined as the occurrence of three or more consecutive, clinically detectable pregnancy failures before the 20th week of gestation (Li et al., 2002b; Quenby et al., 2002; Pandey et al., 2004). However, some studies have also included patients with two miscarriages (Laskin et al., 1997; Stephenson et al., 1998) and in some studies the miscarriages were
not necessarily consecutive. This increases the prevalence of RM up to 5% of all couples trying to conceive (Rai and Regan, 2006).

![Figure 2](image)

**Figure 2.** Number of conceptions expected to reach different stages of pregnancy. Of all fertilized oocytes only about half result in a livebirth. Most of the pregnancies are lost during early stages of development. Modified from Rai and Regan, 2006.

### 2.1 Identifiable causes of miscarriage

The repetitive miscarriages in some couples suggest that certain women are at particular risk of losing their pregnancy and that there must be an underlying explanation for this. Even though much work has been done to identify these underlying mechanisms, the aetiology of miscarriage is still unknown in many cases. Additionally, on an individual level the exact reason for a particular miscarriage is rarely defined. Despite the extensive medical testing and experimental treatments that many patients undergo, the cause often remains unclear (Plouffe et al., 1992; Clifford et al., 1994; Carrington et al., 2005). The main identifiable causes of miscarriage include the following categories (Table 1):

- **Uterine pathology:** Uterine defects, such as uterine anomalies and fibroids, can predispose to miscarriage by affecting implantation (Bulletti et al., 1996; Bajekal and Li, 2000).

- **Endocrine abnormalities:** Following implantation, the maintenance of the pregnancy is dependent on a range of endocrinological events that will eventually aid in the successful growth and development of the endometrium and the foetus (Potdar and Konje, 2005; Arredondo and Noble, 2006). Normal endometrial development and maturation is dependent on steroid hormones, particularly oestrogen and progesterone. Therefore,
miscarriage can be a result of subnormal hormone production or abnormal endometrial response to circulating steroid hormones (Li et al., 2002a; Arredondo and Noble, 2006). In addition, several other maternal endocrinological abnormalities such as uncontrolled diabetes (Mills et al., 1988), high androgen levels (Okon et al., 1998; Bussen et al., 1999), hyperprolactinaemia (Tal et al., 1991), thyroid dysfunction (Roberts and Murphy, 2000), and obesity (Wang JX et al., 2002) have been implicated as aetiological factors for RM.

**Immunological factors:** Since the conceptus displays paternal gene products and antigens, it is possible that the maternal immune system recognizes these as foreign, resulting in an immune response (Dalton et al., 1998; Hill and Choi, 2000). Survival of the semiallogenic foetus is dependent on suppression of the maternal immune response and miscarriage may be a consequence of a failure of this suppression. It is assumed that both maternal and embryonic regulating factors protect the embryo against an adverse maternal immunological reaction. Maternal adaptation to the implanting embryo is needed for successful establishment of the foetal-placental unit, and the embryo must be able to stimulate an immunological reaction from the mother in order to become protected from

| Table 1. Causes of RM and their estimated frequencies in women with RM (Stephenson and Kutteh, 2007; Warren and Silver, 2008). |
|---|---|
| **Factor** | **Frequency** |
| **Uterine defects** | ~15% |
| Anatomic deformation |  |
| Uterine fibroids |  |
| Cervical incompetence |  |
| **Endocrinologic** | ~10% |
| Abnormal hormone production |  |
| Diabetes |  |
| Thyroid dysfunction |  |
| Obesity |  |
| **Immunologic** | ~15% |
| Aniphospholipid syndrome |  |
| Maternal autoimmune factors (acquired or inherited) |  |
| **Thrombophilic** | ~10% |
| Prothrombotic state (acquired or inherited) |  |
| **Environmental** | ~5% |
| Infections |  |
| Teratogens (alcohol, drugs, tobacco) |  |
| **Parental chromosome abnormality** | ~5% |
| Balanced translocations |  |
cytotoxic molecules (Bulletti et al., 1996; Li et al., 2002b; Pandey et al., 2004). The mechanism by which the foetus escapes rejection by the maternal immune system is unknown. However, differences in the concentration of circulating immunocompetent cells, such as natural killer (NK) cells, and cytokines have been reported between RM women and controls (Wegmann et al., 1993; Reinhard et al., 1998; Ntrivalas et al., 2001; Yamada et al., 2001).

**Prothrombic state:** A prothrombotic state can be broadly defined as any condition, whether inherited or acquired, that predisposes to venous or arterial thrombosis (Hiatt and Lentz et al., 2002). It is postulated that maternal thrombophilia is a risk factor for miscarriage because a successful pregnancy is dependent on satisfactory placental development and sustained placental function is based on placental circulation (Kujovich, 2004). These processes require the establishment of an adequate foeto-maternal circulatory system, which in turn may be disturbed by a prothrombotic state. Factor V Leiden, mutations in the prothrombin gene, and protein C and protein S deficiencies are hereditary prothrombotic disorders of clinical importance and associations between increased risk of miscarriage and these thrombophilic disorders have been reported (Preston et al., 1996; Rey et al., 2003; Kujovich, 2004).

One of the major acquired prothrombotic disorders is antiphospholipid antibody syndrome, in which the body recognizes its own phospholipids as foreign and produces antibodies against them. Lupus anticoagulant and anticardiolipin antibody are the two known antiphospholipid antibodies that are associated with RM (Levine et al., 2002). It is unknown exactly how the antiphospholipid antibody syndrome adversely affects pregnancy, but one theory is that it may cause blood clots in the blood vessels of the placenta, impairing placental function (Rand et al., 1997). However, antiphospholipid syndrome is a treatable cause of RM. Various treatments, such as heparin, aspirin, and intravenous immunoglobulin, have been used in attempts to improve the pregnancy outcome of women with antiphospholipid syndrome (Empson et al., 2002; Farquharson et al., 2002). Heparin is a thromboprophylactic agent that can bind to antiphospholipid antibodies, thereby protecting the trophoblast and maternal vascular endothelium from damage in early pregnancy. Later in pregnancy, when the intervillous circulation has been established, heparin helps to decrease the risk of placental thrombosis and infarction (Girardi, 2005, Bose et al., 2005).
Environmental factors; Heavy metals, organic solvents, and ionizing radiation are confirmed teratogens, and exposure to these can contribute to pregnancy loss (Dhont, 2003). Alcohol and cocaine are also confirmed teratogens, while caffeine and smoking are suspected teratogens but their teratogenic impact is still controversial (Dominguez-Rojas et al., 1994; Parazzini et al., 1998; Cnattingius et al., 2000; Kesmodel et al., 2002).

2.1.1 Risk factors for miscarriage

In addition to identifiable causes of miscarriage, other factors increasing the risk for miscarriage have been reported. Increased maternal age is reported to be an independent risk factor for miscarriage (Risch et al., 1988; Abdalla et al., 1993; Nybo Andersen et al., 2000). The risk of miscarriage increases rapidly after the age of 35. The risk of a miscarriage is 9% in women aged 20-24 years, while in women aged over 45 years the risk is as high as 75% (Nybo Andersen et al., 2000). This is explained partly, but not totally, by the association of maternal age with increased likelihood of chromosomal abnormalities in the foetus (Munne et al., 1995; Snijders et al., 1999).

The outcome of previous pregnancies, especially of the first pregnancy, is also an independent predictor of future pregnancy outcome. For young women who have never experienced pregnancy loss the rate of miscarriage is as low as 5% (Regan et al., 1989). The risk increases to approximately 30% for women with three or more losses but with a previous live-born infant, and up to 50% for women with no live-born infants (Poland et al., 1977), indicating that the risk of miscarriage increases with the number of previous miscarriages. Additionally, a normal karyotype in a pregnancy loss is a predictor of subsequent miscarriages because an abnormal foetal karyotype is associated with sporadic foetal loss. The frequency of abnormal foetal karyotypes significantly decreases with the number of miscarriages (Ogasawara et al., 2000).

2.2 Genetic factors causing miscarriage

In addition to the previously described identifiable causes of miscarriage, genetic factors are a major cause of clinically recognized miscarriages. In the following sections, the currently reported genetic causes associated with miscarriages are described.
2.2.1 Chromosomal abnormalities

Most unbalanced chromosome aberrations result in severe phenotypes already in early pregnancy and therefore lead to miscarriage. Consequently the incidence of foetal unbalanced chromosomal abnormalities gradually decreases with duration of pregnancy to less than 1% among live-born children (Stern et al., 1996; Stephenson et al., 2002). Overall, 50-70% of all sporadic miscarriages are associated with cytogenetic abnormalities in the foetus. The most frequent of these abnormalities is trisomy, followed by polyploidy, monosomy X, and unbalanced translocation (Kalousek et al., 1993; Stephenson et al., 2002). The most common trisomy is trisomy of chromosome 16, followed by trisomy of chromosomes 21 and 22 (Stephenson and Kutteh, 2007). These chromosomal abnormalities may be inherited or may arise de novo in the germ cells due to aberrant oocyte spindle formation and meiotic division (Devine et al., 2000; Kuo, 2002).

In women with RM the rate of chromosomal abnormalities in the foetus is lower than in women with sporadic miscarriage. The miscarriages in women with RM may be due to de novo numerical chromosome abnormalities, in particular autosomal trisomies of chromosomes 13, 14, 15, 16, 21, and 22, and monosomy X (Pandey et al., 2005; Warren and Silver, 2008). In women with RM, however, the miscarriage may also be due to recurrent chromosomal anomalies in the foetus resulting from a balanced aberration in one of the parents inherited by the offspring in an unbalanced form. Cytogenetic screening of couples with RM has revealed that parental chromosomal abnormalities occur in either partner in 5-7% of couples with RM, while the rate in the normal population is approximately 0.2% (Fryns and Buggenhout, 1998). Balanced translocations in either parent, including reciprocal and Robertsonian translocations, are the most common abnormalities, detected in about 4% of couples with RM. The other chromosomal abnormalities include inversions and sex chromosome abnormalities (Tulppala et al., 1993; Clifford et al., 1994; Fryns and Buggenhout, 1998). Even though the theoretical risk for abnormal offspring for phenotypically normal carriers of balanced translocations is approximately 50%, preimplantation genetic diagnosis (PGD) analysis has shown that the risk varies greatly. For carriers of a balanced translocation the risk of conceiving a chromosomally abnormal embryo, due to abnormal segregation at meiosis, varies from 20-80%. The reproductive risk conferred by chromosome rearrangements is dependent on the size, location, and type of rearrangement and whether it is carried by the woman or her male partner (Munne et al., 2000; Otani et al., 2006; Lim C et al., 2008).
2.2.2 Skewed X chromosome inactivation

X chromosome inactivation (XCI) is defined as the process where one of the two X chromosomes in each somatic cell of a healthy female is inactivated during early embryogenesis. The process is important to achieve appropriate dosage compensation for X chromosomal genes (Lyon, 1961; Brown and Robinson, 2000). Females are therefore mosaics for two cell types; cells with the paternal X chromosome as the active X and cells with the maternal X chromosome as the active X. Normally the process occurs randomly on both the maternal and paternal X chromosome, resulting in a 50:50 distribution of the two cell types. A significant deviation from this distribution is called skewed XCI (Puck and Willard, 1998; Minks et al., 2008). Recently, several studies have shown that women experiencing RM exhibit nonrandom XCI more often than control women (Lanasa et al., 1999; Uehara et al., 2001; Kuo et al., 2008). Kuo et al. (2008) reported extremely skewed XCI in 7.7% of RM women and in 1.3% of control women (p=0.01).

Skewed XCI, defined as preferential inactivation of one of the two X chromosomes, may be the result of a chance event, or due to genetic factors involved in the X inactivation process (Brown and Robinson, 2000). In some cases a non-random inactivation pattern may be an indicator of an abnormality in the inactivated X, e.g. X chromosome aberrations such as microdeletions or mutations in X chromosomal genes (Belmont, 1996; Uehara et al., 2001). Deviation from random inactivation rates can also lead to the expression of recessive X-linked traits in heterozygous females, who are expected to be healthy carriers of the disease gene (Azofeifa et al., 1995). It has been proposed that X-linked mutations which are lethal in males cause skewed XCI in female carriers. For example, microphthalmia with linear skin defects syndrome (MLS) and oral-facial-digital syndrome type 1 (OFD1) are previously described X-linked dominant male-lethal diseases (Franco and Ballabio, 2006). Such mutations may result in loss of male conceptuses and increased frequency of miscarriage in carrier females (Lanasa et al., 1999; Uehara et al., 2001).

2.2.3 Y chromosome microdeletions

The majority of the testing for RM assesses the woman. The male partners, however, have been less thoroughly investigated. Semen analysis is not generally included in the initial assessment for RM. Yet, sperm integrity is required for sperm–egg interactions,
fertilization, and early embryonic development and sperm quality is associated with the embryo’s ability to reach the blastocyst stage and progress to implantation (Simerly et al., 1995; Van Blerkom, 1996; Moomjy et al., 1999).

Microdeletions of the long arm of the human Y chromosome are associated with variable spermatogenic failure (Vogt et al., 1996; Hopps et al., 2003). These microdeletions occur in regions called AZFa, AZFb, and AZFc (more recently named AZFa, P5/Proximal-P1, P5/Distal-P1, P4/Distal-P1, b2/b4-AZFc), containing a number of genes (Hopps et al., 2003; Ferlin et al., 2007). Although ten recurrent deletions in the AZF regions have been described in detail (Ferlin et al., 2007), surprisingly little is known about the function of the individual genes and transcription units located within these regions (Hopps et al., 2003; Foresta et al., 2005). Y chromosome microdeletions are found in approximately 7% of men with very low sperm counts (Ferlin et al., 2007). These men also often show a higher incidence of sperm chromosomal abnormalities (Foresta et al., 2005). An elevated rate of sperm chromosome abnormalities has also been found in some men of RM couples (Rubio et al., 1999; Egozcue et al., 2000). A recent study reports that the incidence of sperm chromosome abnormalities is 16.5% in male partners of women with RM, while the incidence is under 5% in controls (Al-Hassan et al., 2005). Interestingly, a pilot study conducted to determine if Y chromosome microdeletions are associated with RM, reported that a significant portion (82%) of male partners of women with RM had Y chromosomal microdeletions, which may underlie the miscarriages experienced by their partners (Dewan et al., 2006).

2.3.4 Gene mutations and polymorphisms

Mutations and polymorphisms in several genes are suggested to contribute to RM. Many of these sequence variations are located in genes involved in immunological mechanisms or blood coagulation pathways, but also genes involved in other functions have been associated with RM.

Since the conceptus displays paternal gene products and antigens, it is possible that the maternal immune system recognizes these as foreign, resulting in an immune response (Dalton et al., 1998; Hill and Choi, 2000). In recent years, several genes involved in the immune response have been associated with pregnancy loss. The CTL-A gene is a
regulator of T-cell activation and an A/G polymorphism in exon 1 of the gene has been reported to be associated with RM in the Chinese population. The frequency of the GG genotype in RM patients was significantly higher compared with controls (48.8% versus 33.3%, p = 0.011) (Wang et al., 2005). Polymorphisms in the major histocompatibility complex have also been associated with RM. A polymorphism in the HLA-G promoter region, the HLA-G *0104 and *0105N alleles (Aldrich et al., 2001; Ober et al., 2003) and the HLA-DRB1 *1505 allele have been associated with the immunopathogenesis of RM (Takakuwa et al., 2003). In addition, a number of cytokine gene polymorphisms are associated with pregnancy loss. Cytokines are critical immunoregulatory molecules which play a crucial role in early embryo development and placentation. Associations have been shown between IL-6, IL-10, and IFN-γ gene polymorphisms and RM (Daher et al., 2003; Costeas et al., 2004; Prigoshin et al., 2004; von Linsingen et al., 2005).

It is assumed that maternal thrombophilia is a risk factor for miscarriage because a successful pregnancy is dependent on an adequate foeto-maternal circulatory system, satisfactory placental development and sustained placental function. These functions may be disturbed by a prothrombotic state. Factor V and prothrombin gene mutations, protein C and protein S deficiencies, cause hereditary prothrombotic disorders (Preston et al., 1996; Kujovich, 2004). Associations between these thrombophilic disorders and an increased risk of miscarriage have been reported. Depending on the study the Factor V Leiden mutation and the prothrombin G20210A mutation are reported to increase the risk for RM between 2-fold and 5.5-fold, and between 2.5-fold and 4.9-fold, respectively (Souza et al., 1999; Foka et al., 2000; Reznikoff-Etiévan et al., 2001; Rey et al., 2003). Protein C and protein S deficiencies are much rarer and they have been reported to increase the risk of early RM ~3.5-fold and ~14.5-fold, respectively (Rey et al., 2003; Kujovich, 2004). The MTHFR gene has an important role in homocysteine metabolism and missense mutations C677T and A1298C are known to cause hyperhomocysteinemia. Hyperhomocysteinemia is a risk factor for vascular thrombophilia and has also been shown to associate with RM (Couto et al., 2005; Mtiraoui et al., 2006). In addition, polymorphisms in the ITGB3 gene, which encodes a human platelet antigen, may increase the risk of miscarriage by causing thrombophilia. The PLA2 polymorphism in the ITGB3 gene was shown to increase the risk of foetal loss 3.6-fold (Ruzzi et al., 2005).

In addition, various other genes have been associated with pregnancy loss. For example, mutations in the SCO2 gene, needed for cytochrome c oxidase assembly, were
detected in a compound heterozygous state in aborted foetuses and are suggested to cause miscarriage (Tay et al., 2004). A novel A/G polymorphism in intron 6 of the eNOS gene, which is active during implantation, was recently reported to be associated with RM (Suryanarayana et al., 2006). A polymorphism in the VEGF gene, a well characterized regulator of angiogenesis, was also recently reported as a new susceptibility factor for RM. Homozygosity of the VEGF –1154A variation was more frequent in RM women compared to controls and increased the risk of RM 2.7-fold (Coulam and Jeyendran, 2008).

2.2.5 Aberrant placental gene expression

The growth and differentiation of placental structures are regulated by many genes and the placenta is very sensitive to genetic disruption, as reflected by the increasing list of targeted mouse mutations in genes such as Mash2 and Hand1 that cause placental defects and embryonic lethality (Rossant and Cross, 2001). Genetic abnormalities that affect the development of the placenta can cause placental insufficiency, foetal growth retardation and embryonic death, both in mice and humans (Rossant and Cross, 2001; Cross, 2006).

It is assumed that normal placentation and normal placental functions are mediated by differential gene expression in the placenta. It has also been suggested that genes expressed in the chorionic villi act as critical regulators of placentation and hence of early human development (Dizon-Townson et al., 2000; Baek, 2004). Recently, over 30 genes that are expressed aberrantly in pregnancy failure have been identified. The genes showing different levels of expression between controls and patients with RM can be mainly grouped as immunity related, angiogenesis related, and apoptosis related genes (Baek et al., 2002; Choi et al., 2003). Higher expression levels of apoptosis related genes such as caspase 3, 6, 7, 8, 9, 10, 12, BAD, BAX, BID, Fas and FasL were shown in chorionic villi from patients with RM compared to controls (Choi et al, 2003). Immunity related genes with aberrant gene expression in chorionic villi include PP14, MUC1 (Baek et al., 2002), CD95 (Hoshimoto et al., 2002), and Annexin II (Aarli et al., 1997). These genes may be involved in immunological foeto-maternal reactions. Fibronectin, Integrin and VEGF are examples of angiogenesis related genes showing aberrant gene expression in RM placenta (Choi et al, 2003). These genes are involved in the formation of blood vessels that are crucial for placental and embryonic development.
2.2.6 Epigenetic mechanisms

The term epigenetics refers to processes that regulate gene activity without affecting the DNA sequence and are heritable through cell division. Epigenetic modifications of DNA such as methylation are important for genome function during development, and disturbances in the highly coordinated epigenetic processes contribute to developmental failures (Lucifero et al., 2004; Santos and Dean, 2004). In most mammalian species, including humans, there are at least two developmental periods during which methylation patterns are reprogrammed genome wide; germ cell development and preimplantation. After genome wide demethylation of the genome in developing germ cells, remethylation, which resets the parent-of-origin specific marks, occurs. The second phase of reprogramming of methylation occurs during the preimplantation stage. The paternal genome is actively demethylated within a few hours of fertilization in the human zygote, whereas the maternal genome is passively demethylated by a replication dependent mechanism after the two cell embryo stage. After the demethylation stage, de novo methylation of the genome occurs in the developing embryo. This reprogramming is likely to have a crucial role in establishing nuclear totipotency required for normal development (Reik et al., 2001; Reik and Dean, 2001; Haaf, 2006).

Genomic imprinting is a mechanism which results in differential expression of maternal and paternal genes. Only one of the inherited gene copies is active in an embryo depending on the parental origin of the gene. DNA methylation in germ cells is critical for imprinting as most imprinted genes have differentially methylated regions (Reik and Dean, 2001; Swales and Spears, 2005). Anomalous expression of imprinted genes during development is sometimes caused by imbalanced representation of maternal and paternal genetic contributions, uniparental disomy, and is associated with developmental delay or gestational loss (Mutter, 1997).

It is estimated that there are several hundred imprinted genes in the human genome. It has been argued that imprinted genes play essential roles in controlling the placental supply of maternal nutrients to the foetus, by regulating the growth of the placenta and/or the activity of transplacental transport systems. In some cases, errors in genomic imprinting are embryo lethal while in others they lead to developmental disorders and diseases (Reik and Dean, 2001; Lucifero et al., 2004; Swales and Spears, 2005).
3 Identification of genetic causes underlying RM

Despite numerous investigation the cause of miscarriage remains unknown in approximately 50% of cases (Plouffe et al., 1992; Tulppala et. al 1993, Clifford et al., 1994). A number of possible aetiologies have been proposed to explain the RM in these cases. The hypotheses, however, change over time and new findings have enriched our understanding of possible mechanisms underlying RM. For example, specific HLA-G alleles (Aldrich et al., 2001) and skewed X-inactivation (Lanasa et al., 1999; Uehara et al., 2001) are newly factors identified to be associated with RM and could, in some cases, be the cause for unexplained RM. However, additional genetic factors contributing to the aetiology of RM are likely to be detected.

There are different mechanisms by which genetic factors may contribute to miscarriage, and different modes of inheritance by which these mutations and polymorphisms are transmitted from parents to foetus. Dominant de novo mutations arising in the germ cells of the parents may result in affected foetuses and cause miscarriage. Alternatively, de novo mutations during an early developmental stage in the foetus may result in mosaicism, which also may cause a severe phenotype resulting in miscarriage (McFadden and Friedman, 1997; Los et al., 2004). De novo mutations are difficult to study because these mutations cannot be detected by studying DNA samples from parents and foetal/placental samples are rarely available.

Autosomal recessive mutations cause a number of genetic diseases, and are common especially in isolated populations (Peltonen et al., 1999; Norio, 2003a; Teebi and Teebi, 2005; Weinstein, 2007). The role of recessive mutations in specific genes in RM can be studied without foetal samples by screening the parents for heterozygous mutations, which when inherited by the foetus in a homozygous state are expected to be lethal. In isolated populations disease causing mutations are often enriched and may also show uneven geographical distribution. Evidence of such geographical clustering can be studied using genealogic data (Kere, 2001; Norio, 2003a).

Not only autosomal mutations, but also abnormalities in the sex chromosomes can underlie RM. X-linked mutations are predicted to be capable of inducing foetal loss. Some X-linked recessive mutations are expected to be lethal to all male foetuses inheriting the mutation and thereby result in miscarriage (Lanasa et al., 1999). Recently, a study
suggested an association between Y-chromosome microdeletions and RM (Dewan et al., 2006). The role of these genetic factors in the aetiology of RM can be further studied using parental DNA samples.

In addition to the nuclear genome, the mitochondria contain their own DNA, which is transmitted only from the mother to the foetus. The human mitochondrial genome is 16,569 bp in size and encodes 37 genes (DiMauro, 2001). Mutations in these genes leading to dysfunctional proteins may have a role in RM as the mitochondria are involved in processes necessary for early development (Van Blerkom, 2004).

It is also possible that mutations in imprinted genes can cause miscarriage. When the imprinting mechanism is malfunctioning loss of function or inappropriate expression can occur in intact genes and may result in miscarriage (Reik and Dean, 2001; Swales and Spears, 2005). The role of imprinted genes in RM is, however, difficult to study as imprinting still is a poorly understood phenomenon.

Recently, several genes such as CTL-A, IL6, ITGB3, and VEGF have been shown to be associated with RM (Wang et al., 2005; von Linsingen et al., 2005; Ruzzi et al., 2005; Coulam and Jeyendran, 2008) and it is likely that other genes, in which mutations and polymorphisms cause or increase the risk for miscarriage, will be identified in the near future. Therefore, further studies and screening of candidate genes are required to establish the role of genetic factors in RM.

### 3.1 Identification of disease genes

#### 3.1.1 The Human Genome Project

The human genome consists of approximately 3 billion base pairs (bp) and is estimated to contain 20 000-25 000 genes (Balakrishnan et al., 2005; Antonarakis and Beckmann, 2006). The Human Genome Project (HGP) began in 1990 and its main goals were to determine the nucleotide sequence of the human genome and to identify all human genes. A draft sequence of the human genome was published in 2001 (Lander et al., 2001; Venter et al., 2001). An updated sequence, published in 2003, covered ~99% of the euchromatic genome (International Human Genome Sequencing Consortium 2004). The first complete
genome sequence of a single human individual was published in 2007 (Levy et al. 2007). In addition to the human genome, the genoms of hundreds of other organisms have been sequenced (www.ebi.ac.uk/genomes/index.html). This genomic information has dramatically changed the process of identifying disease genes. It provides new, invaluable tools for understanding the basic human genetic make-up and how variations in the genomic sequence result in disease (Peltonen and McKusick, 2001; http://genomics.energy.gov).

3.1.2 Traditional identification of disease genes

Before the complete human genome sequence was available, disease genes were traditionally identified using functional and positional cloning methods. In functional cloning the gene is isolated based on information regarding the protein or its function. The method can be used if the amino acid sequence of a protein is at least partially known (Collins, 1992). However, most often, the protein encoded by the disease gene is not known and the positional cloning method is used to identify a gene based on the knowledge about its physical location in the genome. Using polymorphic markers and DNA samples from families with affected members the disease gene locus is localized to a genomic region by linkage analysis. Disease gene identification based on positional cloning has been widely used in the research of monogenic diseases (Collins, 1992; Collins, 1995).

3.1.3 The candidate gene approach

The candidate gene approach is an important step towards disease gene identification and relies on information about known genes. New effective methods for mutation screening together with the information offered by the HGP have made the candidate gene approach a commonly used method to search for disease genes as information on physical locations and sequences of many genes is available. The candidate gene approach can be used when the biochemical background of the defect of interest is known, when a chromosomal region has been linked to a disease or if an animal model of a disease has been established (Collins, 1992; Ballabio, 1993; Zhu and Zhao, 2007). The mouse is a widely used model organism for studying mammalian gene function and genes causing a phenotype in the mouse can be used as candidate genes in human disease studies (Strachan and Read,
Interaction partners of identified proteins defective in a disease can also be considered as candidate genes for the disease of interest. In addition, candidate genes can be selected based on results from genome wide expression profiling data showing differences in gene expression between cells or tissues from affected and control individuals (Antonarakis and Beckmann, 2006).

### 3.1.4 Mutation screening

When a candidate gene is selected, the next step in disease gene identification is searching for mutations in the gene under study. During the past years several different methods have been used for mutation screening. To date, one of the most commonly used method is direct sequencing. In addition, denaturing high-performance liquid chromatography (dHPLC) (Xiao and Oefner 2001) and multiplex ligation-dependent probe amplification (MLPA) (den Dunnen and White, 2006) are frequently used to screen DNA samples for sequence variations due to their sensitivity and specificity. A set of patients and control DNA samples are screened for mutations segregating with the disease phenotype. It is important to screen a set of healthy control individuals to be able to assess the significance of the nucleotide changes and to confirm that the variation truly is associated with the disease (Strachan and Read, 2004).

When screening candidate genes different types of sequence variations can be identified. **Nonsense mutations** are point mutations that introduce premature stop codons into the coding regions. Nonsense mutations usually lead to the degradation of the transcript or may produce truncated nonfunctional protein products. **Missense mutations** are point mutations that result in the substitution of an amino acid and are sometimes difficult to differentiate from rare neutral polymorphisms. A missense mutation is more likely to be pathogenic if it changes an amino acid that is conserved. If the change alters the chemical nature of the side chain of the amino acid this usually suggests a pathogenic role for the variation. In addition, sequence variations in parts of the gene that code for functionally important domains are more likely to be pathogenic. **Deletions, insertions, and duplications** can disrupt the reading frame, resulting in a completely different protein translation. In addition to **frameshift mutations**, deletions, insertions, and duplications can introduce premature stop codons or alter the protein by amino acid(s) deletion or insertion (Strachan and Read, 2004).
Splicing mutations affect the consensus sequences at splice donor, splice acceptor, or splice branch sites and may abolish the splicing partially or completely. In addition, cryptic splice sites can be activated or splicing enhancers and silencers altered. These types of splicing mutations may lead to complete or partial exon skipping or intron retention (Cartegni et al. 2003; Baralle and Baralle, 2005). In addition, neutral and silent mutations/polymorphisms are likely to be observed in candidate gene screening. A neutral mutation is a variation that results in the substitution of a different, but chemically similar amino acid. Silent mutations are variations that do not result in a change to the amino acid sequence. They may occur either within an exon or within the non-coding regions of the gene (Strachan and Read, 2004). Polymorphisms in the non-coding regions of the genome, including the promoter region, the 5'- and 3'-untranslated regions, and the intragenic regions are also commonly detected. Some of these variations may affect gene splicing, and transcription and translation, and thereby the expression of the gene (Wang et al., 2006).

3.1.5 Confirming the nature of a new genetic variation

Identifying a disease gene is a possible starting point to study the function of the protein encoded by the gene. A candidate disease gene has to be carefully studied to determine if there is enough evidence that the identified changes are not just neutral polymorphisms but do cause the disease under study. The candidate gene can be demonstrated to be the disease gene in different ways. Mutation screening is usually the first step in the process. Identification of mutations in several unrelated individuals strongly suggests that the candidate gene is causing the disease, but formal proof requires additional evidence. To prove that the candidate gene has a pathogenic role in the disease functional studies, such as restoration of the normal phenotype in vitro, or production of an animal model for the disease, are required (Strachan and Read, 2004).

3.2 Mouse as a model organism

Genetic approaches in model organisms provide a powerful means by which to examine the biological basis of human diseases. Over the past century, the mouse has become the most used animal model system for genetic research because of its close genetic and
physiological similarities to humans. In addition, its small size and short generation time have allowed large scale mutagenesis experiments and extensive genetic crosses. Many mouse mutations have been established as reliable models of human disease while others have provided insight into complex developmental and physiological pathways in mammals. Targeted mutagenesis approaches in the mouse have led to dramatic increases in our understanding of human disease processes (Brown, 1998; Brown and Nolan, 1998; Nolan et al., 2002). There are a number of mouse models described in the literature and databases on different mutant strains are available on the internet (http://ceolas.org/VL/mo/; Peters et al., 2007).

Sequencing of the mouse genome was initiated in 1999 and the first draft sequence was published in 2002 by Waterston and coworkers. Because of the comparatively high level of sequence conservation between human and mouse coding sequences, almost all human genes have an easily identifiable mouse homolog. Therefore, disease genes identified in mouse models are usually good candidate genes for human diseases. The mouse and human genomes both appear to contain 20 000-25 000 protein coding genes and more than 90% of the mouse genome can be aligned with a region of the human genome due to conserved synteny (Waterston et al., 2002). In addition to providing important data on mouse genes and the structure of the mouse genome, the sequence data have also been valuable in comparative genomics. One application of human-mouse comparative sequence analysis involves the identification of orthologous genes. Functionally important genomic regions are commonly homologous in human and mouse genomes and genetic information obtained from the mouse can thereby be used for human disease studies (Collins et al., 1998; Nobrega and Pennacchio, 2004).

Apart from the very early stages of preimplantation development, human embryos are inaccessible for research and studies are limited to diseased aborted foetuses. Due to the difficulties of studying implantation and foetal development in humans, much of our knowledge of the genes involved in these processes has been derived from animal models. There are several animal models used to gain information about human development including; nematode worm, fruit fly, rat, guinea pig, rabbit, pig, sheep, cow, and non-human primates (Lee and DeMayo, 2004; Dvash et al., 2006; Carter, 2007). The most commonly used model organism is the mouse, but there are, however, significant differences in the mode of implantation, placental structure, yolk sac, number and function of placental hormones, and particularly in the length of pregnancy between mouse and
human (Rossant and Cross, 2001; Carter, 2007). Despite these differences, gene ablation in the mouse has proven to be a powerful tool in elucidating gene function during embryonic development (Rossant and Cross, 2001; Dvash et al., 2006).

Knock-out mouse models have been used to identify a number of genes required for normal implantation and placentation (Rossant and Cross, 2001; Lee and DeMayo, 2004; Carter, 2007). In addition, imprinted genes required for normal embryo development have been studied in mouse models (Mutter, 1997; Lucifero et al., 2004). A number of mouse models, in which deletions of a specific gene are embryo-lethal at a specific developmental stage, have also been described (Healy et al., 1995; Sah et al., 1995; Wang et al., 1996; Gu et al., 2002). These mouse models can be used to identify candidate genes for human RM. A gene identified in mouse to be embryo lethal can be considered as a clinically important candidate gene for human RM if the mouse and human genes have a highly homologous sequence and similar normal functions in both species.

3.3 Candidate genes for RM

The following sections describe four candidate genes for human RM. They were selected based on results from studies on animal models. In addition, the genes encoded by the mitochondrial genome were considered as candidate genes because the mitochondria are involved in processes important for development but their role in early development is still unknown.

3.3.1 Amnionless

The amnionless mouse model

The amnionless mouse has a recessive prenatally lethal insertional mutation in the Amn gene (Wang et al., 1996). The gene was named Amnionless (Amn) as the first visibly striking feature of the mutant mouse was the lack of an amnion surrounding the foetus. Examination of the mutant embryos at different times of gestation indicated that the mutation causes a gastrulation stage (embryonic day (E) 6.5-E8.0) defect in the homozygous mutant mouse (Wang et al., 1996; Tomihara-Newberger et al., 1998). Although gastrulation begins normally in the Amn mutants they remain small and undifferentiated and generate no amnion. The growth and differentiation of the embryonic
ectoderm cells are severely impaired in the mutant embryos. The first signs of abnormal development in embryos homozygous for the transgene insertion can be seen between E6.5 and E7.0 and the mutant embryos die and are resorbed between E9.5 and E10.5 (Table 2) (Wang et al., 1996).

**Functions of the Amn gene**
The Amn gene encodes a type I transmembrane protein (Kalantry et al., 2001). During the early post-implantation stages mouse Amn is expressed exclusively in the visceral endoderm (Tomihara-Newberger et al., 1998; Kalantry et al., 2001). In adults, both mouse and human Amn is expressed mainly in the kidney and the intestinal epithelium (Tanner et al., 2003; Strope et al., 2004).

The exact role of the Amn protein in mouse development has not yet been determined, but the gene function is required in the visceral endoderm for normal primitive streak formation (Tomihara-Newberger et al., 1998). The visceral endoderm is an epithelial cell layer that does not contribute directly to the foetus, but is needed to absorb and digest nutrients from the maternal environment, to provide signals required for the organization of the body axes, and to secrete factors that promote correct positioning of the primitive streak (Beddington and Robertson, 1999; Martinez-Barbera and Beddington, 2001; Perea-Gomez et al., 2002). At the onset of gastrulation the middle region of the primitive streak is absent in the Amn mutant mouse (Tomihara-Newberger et al., 1998).

In wild-type mouse embryos Amn co-localizes with two other gene products, cubilin (Cubn) and megalin, at the surface of the visceral endoderm. In the Amn mutant mouse Cubn fails to localize to the surface of the visceral endoderm. This indicates that Amn is an essential component of the Cubn receptor complex, and it has been suggested that the Amn/Cubn-complex is required for endocytosis/transcytosis of one or more ligands in the visceral endoderm. (Strope et al., 2004). In humans, AMN and CUBN also form a complex. This complex is required for uptake of vitamin B$_{12}$ (Fyfe et al., 2004; Luder et al., 2008). Mutations in either of these genes cause selective intestinal malabsorption of vitamin B$_{12}$, a rare autosomal recessive disorder known as Imerslund-Gräsbeck syndrome (IGS) or megaloblastic anemia 1 (Imerslund, 1960; Gräsbeck et al., 1960; Aminoff et al., 1999; Tanner et al., 2004).
The fact that the Amn protein has a similar structure in both mice and humans suggests similar functions in both species, making Amn a candidate gene for miscarriages in humans. Amn may be required in the extra-embryonic endoderm during early embryonic development, and in the intestine for the uptake of vitamin B₁₂ in adults in both species.

**Table 2.** Comparison between mouse and human stages of early development and the timepoints at which TM, AMN, and EPCR mutant mice embryos die. The carnegie stages (CS) provide a universal system for staging and comparing the embryonic development of vertebrates. The stages are based on the external and internal morphological development of the embryo, and are not directly dependent on either age or size. The stages covers approximately the first 60 days of human development and nearly the entire intrauterine development of mouse (UNSW Embryology, http://embryology.med.unsw.edu.au/embryo.htm) The table shows the age of human and mouse embryos in days at specific stages. Implantation takes place during CS 5, gastrulation starts during CS 7, neurogenesis starts during CS 8, and heart beats can be detected at CS 10.

<table>
<thead>
<tr>
<th>Carnegie stage</th>
<th>Human Age</th>
<th>Mouse Age</th>
<th>Developmental stages at lethality of mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS 1</td>
<td>E 1</td>
<td>E 0.5</td>
<td>Implantation</td>
</tr>
<tr>
<td>CS 2</td>
<td>E 2-3</td>
<td>E 1-2</td>
<td></td>
</tr>
<tr>
<td>CS 3</td>
<td>E 4</td>
<td>E 3</td>
<td></td>
</tr>
<tr>
<td>CS 4</td>
<td>E 5-6</td>
<td>E 4</td>
<td></td>
</tr>
<tr>
<td>CS 5</td>
<td>E 7-12</td>
<td>E 5-6</td>
<td></td>
</tr>
<tr>
<td>CS 6</td>
<td>E 13-15</td>
<td>E 6.5</td>
<td></td>
</tr>
<tr>
<td>CS 7</td>
<td>E 15-17</td>
<td>E 7</td>
<td>Gastrulation</td>
</tr>
<tr>
<td>CS 8</td>
<td>E 17-19</td>
<td>E 7.5</td>
<td>Neurogenesis</td>
</tr>
<tr>
<td>CS 9</td>
<td>E 19-21</td>
<td>E 8</td>
<td></td>
</tr>
<tr>
<td>CS 10</td>
<td>E 22-23</td>
<td>E 8.5</td>
<td>Heart beats</td>
</tr>
<tr>
<td>CS 11</td>
<td>E 23-26</td>
<td>E 9</td>
<td>TM mutants die</td>
</tr>
<tr>
<td>CS 12</td>
<td>E 26-30</td>
<td>E 9.5</td>
<td></td>
</tr>
<tr>
<td>CS 13</td>
<td>E 28-32</td>
<td>E 10</td>
<td>AMN mutants die</td>
</tr>
<tr>
<td>CS 14</td>
<td>E 31-35</td>
<td>E 10.5</td>
<td>EPCR mutants die</td>
</tr>
<tr>
<td>CS 15</td>
<td>E 35-38</td>
<td>E 11</td>
<td></td>
</tr>
<tr>
<td>CS 16</td>
<td>E 37-42</td>
<td>E 11.5</td>
<td></td>
</tr>
<tr>
<td>CS 17</td>
<td>E 42-44</td>
<td>E 12</td>
<td></td>
</tr>
<tr>
<td>CS 18</td>
<td>E 44-48</td>
<td>E 13</td>
<td></td>
</tr>
<tr>
<td>CS 19</td>
<td>E 48-51</td>
<td>E 13.5</td>
<td></td>
</tr>
<tr>
<td>CS 20</td>
<td>E 51-53</td>
<td>E 14.5</td>
<td></td>
</tr>
<tr>
<td>CS 21</td>
<td>E 53-54</td>
<td>E 15</td>
<td></td>
</tr>
<tr>
<td>CS 22</td>
<td>E 54-56</td>
<td>E 15.5</td>
<td></td>
</tr>
<tr>
<td>CS 23</td>
<td>E 56-60</td>
<td>E 16</td>
<td></td>
</tr>
</tbody>
</table>
3.3.2 Thrombomodulin and Endothelial Protein C Receptor

The Thrombomodulin mouse model
In mice, absence of the blood-clotting regulator thrombomodulin (TM) causes embryonic lethality before development of a functional cardiovascular system (Healy et al., 1995). TM has been shown to have a dual role in development as TM is needed in two distinct tissues. Expression of TM in non-endothelial cells of the placenta is required for proper function of the early placenta, while the absence of TM from blood vessel endothelium causes excessive activation of the embryonic blood coagulation system (Healy et al., 1995; Isermann et al., 2001). The first indications of abnormal development are observed at E8.5 in homozygous TM-deficient (TM-/−) embryos and manifest as overall growth retardation. This initial phenotype is invariably followed by rapid and complete resorption of homozygous TM-null embryos within the next 10-12 hours (Table 2) (Healy et al., 1995; Isermann et al., 2001).

Functions of the TM gene
TM is an endothelial cell surface glycoprotein expressed on the endothelium of arteries, veins, capillaries, lymphatics, and on trophoblast cells of the placenta (Maruyama et al., 1985; Weiler-Guettler et al., 1996; Fazel et al., 1998). TM plays a key role in the protein C anticoagulant pathway, which is a major regulatory system providing an inhibitory feedback mechanism that suppresses blood coagulation by preventing fibrin formation. TM inhibits blood coagulation by forming a 1:1 complex with thrombin. Binding of thrombin to this high-affinity receptor alters its specificity toward several substrates. The complex converts protein C into the natural anticoagulant activated protein C (APC) approximately 1000 times faster than thrombin alone. APC proteolytically inhibits the coagulation cofactors Va and VIIIa, resulting in down-regulation of the activity of the coagulation system (Figure 3) (Maruyama et al., 1985; Van de Wouwer et al., 2004; Dahlbäck and Villoutreix, 2005).

A number of polymorphisms in the TM gene have been characterized and their association with thrombosis or arteriosclerosis are subjects of ongoing research. These associations have been evaluated in several studies and the results suggest that the relevance of TM mutations in these complications is small or limited to a small subgroup of individuals (Ireland et al., 1997; Doggen et al., 1998; Le Flem et al., 1999; Nakazawa et al., 2002). The most convincing evidence that reduced TM function causes thrombosis is
derived from animal studies. In the past three genetically engineered mouse lines with various degrees of TM dysfunction have been generated, showing different degrees of thrombosis (Healy et al., 1995; Healy et al., 1998; Isermann et al., 2001). It appears that mutations causing a substantial reduction of TM function are very rare in humans (Faioni et al., 2002), possibly because such loss-of-function mutations would lead to foetal loss at an early developmental stage.

**Figure 3.** The protein C anticoagulant pathway. 1) Activation of prothrombin (PT); the Xa-Va complex converts prothrombin to thrombin (T), which forms a complex with TM. 2) Protein C activation; protein C (PC) binds to EPCR, which presents PC to the T-TM complex. 3) Cofactor inactivation; activated protein C (APC) binds protein S (PS). This complex inactivates factors Va and VIII thereby preventing blood coagulation.

**The EPCR mouse model**

Deletion of the endothelial protein C receptor (*EPCR*) gene in mice leads to embryonic lethality. Genotyping of progeny obtained from EPCR^{+/−} interbreeding indicates that EPCR^{−/−} embryos die on or before E10.5 (Table 2). However, EPCR^{−/−} embryos removed from extra-embryonic membranes and tissues at E7.5 and cultured *in vitro* develop beyond E10.5, suggesting a role for EPCR in the normal function of the placenta and/or at the maternal-embryonic interface (Gu et al., 2002). EPCR is normally detected on giant trophoblast cells, which are in direct contact with the maternal circulation and its clotting factors. In the giant trophoblast cells derived from EPCR^{−/−} embryos thrombosis is observed (Li et al., 2005). These observations provide evidence that extraembryonic EPCR expression is essential for embryonic viability and plays a critical role in the control of blood coagulation at the foeto-maternal boundary (Gu et al., 2002).
Functions of the EPCR gene

EPCR is a type I transmembrane protein and is strongly expressed in the endothelial cells of arteries and veins in heart and lung, and less intensely expressed in capillaries in the lung and skin (Laszik et al., 1997; Simmonds and Lane, 1999; Esmon, 2000). It is also expressed in the placenta and the developing cardiovascular system of the foetus (Crawley et al., 2002). EPCR is seen as a complementary cofactor in protein C activation (Figure 3). It binds both protein C and APC with similar affinity and functions in the protein C pathway by binding protein C and presenting it to the TM-thrombin complex on the endothelium, thereby further enhancing the rate of protein C activation and suppressing blood coagulation (Stearns-Kurosawa et al., 1996; Van de Wouwer et al., 2004).

In human, sequence variations in the EPCR gene are associated with the risk of thrombosis. A heterozygous mutation in EPCR has been identified in humans who develop thrombosis and myocardial infarction. This mutation, c.323-9_336dup, is a 23-bp insertion that leads to the formation of a truncated receptor that lacks the extracellular and transmembrane domains and is unable to sustain protein C activation (Biguzzi et al., 2001). However, a common polymorphism in EPCR, c.717+16G>C, seems to reduce the risk of thrombosis. The CC genotype reduces the risk of venous thromboembolism 2.6-fold, probably due to the higher APC levels observed in individuals carrying this genotype (Medina et al., 2004; Medina et al., 2005).

TM and EPCR as candidate genes

Pregnancy complications in patients with prothrombotic risk factors have mainly been linked to the formation of blood clots in placental blood vessels and sinuses, eventually leading to intrauterine growth retardation and, in the most severe case, foetal loss (Kupferminc, 2003; Kujovich, 2004; Kutteh and Triplett, 2006). Therefore, mutations in the genes involved in the blood coagulation systems may increase the risk of thrombosis and thereby miscarriage. Although the importance of the TM-protein C-EPCR system for placental development and maintenance of pregnancy is firmly established in mice (Healy et al., 1995; Gu et al., 2002), the relevance of these mechanisms for pregnancy associated complications in humans remains to be established. The data from mouse models, the known sequence homology of mouse and human TM and EPCR, and the similar type of placentation in both species (Dittman and Majerus, 1989; Cross et al., 1994), however, suggests TM and EPCR as candidate genes for RM in humans.
3.3.3 p53

Functions of the p53 gene

The p53 gene encodes a transcription factor consisting of an N-terminal transactivation domain, a central sequence specific DNA binding domain, and a C-terminal oligomerization domain (Levine, 1997). p53 is expressed at low levels in most cells and the protein levels increase in response to various stress signals. p53 is expressed at high levels in the mouse embryo up to midgestation, after which the expression levels rapidly decrease (Rogel et al., 1985; Marzusch et al., 1995). Hundreds, if not thousands, of studies have been dedicated to the functions of p53 because it plays a pivotal role in tumour suppression. Indeed, more than 50% of human cancers contain somatic mutations in this gene (Levine, 1997). p53 functions as a tumour suppressor by regulating the cell cycle, protecting the genome from DNA damage and in inducing apoptosis. Moreover, p53 appears to interact with a large number of proteins involved in growth control, DNA repair and transcriptional regulation (Choi and Donehower, 1999; Cheah and Looi, 2001; Balint and Vousden, 2001). Experiments in mouse models have revealed that most of the human tumour suppressor genes are essential for embryonic development. Furthermore, inactivation of both germ line alleles of a tumour suppressor usually results in embryonic lethality either in early or midgestation (Rogel et al., 1985; Schmid et al., 1991; Jacks, 1996), making p53 a candidate gene for RM.

There is one previous study supporting the hypothesis that p53 has a potential role during early pregnancy also in humans. The results of Pietrowski et al. (2005) indicate that RM is associated with a polymorphism 12139G>C at codon 72, resulting in either a proline or arginine residue. In addition, the frequency of the Pro72 allele has been shown to be significantly higher among women experiencing recurrent implantation failure than in control women (Kay et al., 2006). The codon 72 variation of p53 have been shown to differ biochemically and biologically (Thomas et al., 1999), but the role of p53 and the codon 72 polymorphism in human development needs to be further studied.

p53 mouse models

Although p53 null mice can be viable and fertile, the absence of p53 activity can lead to fatal developmental abnormalities and a significant proportion of p53−/− mice die during embryogenesis. A substantial fraction of p53 null mice have developmental defects, including neural tube defects, and the affected embryos frequently undergo resorption.
clearly implicating p53 as an important factor in the early development (Sah et al., 1995; Armstrong et al., 1995).

p53 can inhibit DNA replication by triggering growth arrest in G1 or initiate apoptosis, and therefore its activity must be tightly regulated in order to allow normal growth. In the absence of stress inducing signals, p53 is kept at low levels via its interaction with mouse double minute 2 (MDM2) that binds to the transcriptional activation domain of p53. For cell division in the early embryo to proceed, p53 needs to be down-regulated by Mdm2 (Piette et al., 1997; Bond et al., 2005). A deletion of Mdm2 causes embryonic lethality in mice around the time of implantation (Montes de Oca Luna et al., 1995; Jones et al., 1995). In addition, p53

Mouse models have also suggested that p53 may play a key role in development through protecting embryos from teratogens and diverse environmental stresses (Nicol et al., 1995; Norimura et al., 1996). Teratogenesis experiments in mice provide strong support for the idea that p53 efficiently aborts embryonic cells with teratogen induced DNA damage. If too many cells of the embryo die through apoptosis, the entire embryo will die. Conversely, lack of p53 in the embryo presumably results in continued cell cycle progression despite DNA damage, much less apoptosis and a higher proportion of embryos with developmental anomalies (Nicol et al., 1995). Taken together, the results obtained from mouse models suggest that the relative levels of p53 in the early embryo are highly important, and the protein expression has to be very carefully controlled during developmental processes. Changes in the expression levels of p53 elevates the risk of miscarriage and developmental defects, and increases the susceptibility to chemical teratogenesis, possibly also in humans.

3.3.4 The mitochondrial genome

Mitochondria are cellular organelles that contain their own genome. It is estimated that in a somatic cell one human mitochondrion typically contains 5-10 DNA copies and, depending on the cell type, a somatic cell can contain up to 1,000 mitochondria. The human mitochondrial genome is circular, double-stranded and 16,569 bp in size. It
contains 37 genes; 13 protein coding genes, which are all components of the respiratory chain; two rRNAs; and 22 tRNAs needed for mitochondrial protein synthesis (Howell et al., 2000; DiMauro, 2001; Thorburn and Dahl, 2001). The mitochondria are involved in cellular ATP production and apoptosis (DiMauro, 2007). These processes are of great importance in the early development of embryos and may be disturbed by mutations in mitochondrial DNA (mtDNA).

**Genetics of mtDNA**

Mitochondrial genetics differs from Mendelian genetics in three major aspects:

1) *Maternal inheritance*. At fertilization, all mitochondria in the zygote derive from the oocyte. mtDNA mutations are therefore inherited in a matrilineal fashion (DiMauro, 2001; McKenzie et al., 2004).

2) *Heteroplasmy/threshold effect*. In contrast to autosomal nuclear genes, each consisting of one maternal and one paternal allele, there are hundreds or thousands of mtDNA molecules in each cell. Deleterious mutations in mtDNA usually affect some but not all of the mtDNA genomes. Therefore cells, tissues, and organs can contain two different mtDNA populations (wild-type and mutant), a state known as heteroplasmy. The situation when all mtDNAs are identical is called homoplasmy. Non-deleterious mutations/neutral polymorphisms of mtDNA are usually homoplasmic, whereas pathogenic mutations are usually heteroplasmic. For an mtDNA mutation to have a deleterious effect on tissue function it must reach a minimum threshold - a certain proportion of mutant molecules are required before a phenotype is observed (DiMauro, 2001; Thorburn and Dahl, 2001; McKenzie et al., 2004).

3) *Segregation*. As a heteroplasmic cell replicates different proportions of mutant mtDNA molecules are transmitted to the daughter cells, and when the pathogenic threshold for a given tissue is surpassed the phenotype can also change. In this way the mtDNA genotype can drift toward homoplasmy, either wild-type or mutant (DiMauro, 2001; McKenzie et al., 2004).

Today over 200 mtDNA point mutations have been reported in the Mitomap database (http://www.mitomap.org) to be associated with different human diseases. These mutations can be divided into two groups: mutations affecting mitochondrial protein synthesis, including mutations in tRNA and rRNA genes and mutations in protein encoding genes. The mutations are involved in a broad spectrum of human diseases,
including multisystem disorders as well as more tissue specific diseases (DiMauro, 2001; Taylor and Turnbull, 2005; McKenzie et al., 2004; Wong, 2007).

**The mitochondrial bottleneck**

The segregation of mtDNA alleles is unexpectedly rapid, and in some maternal lineages complete allele switching has been observed in a single generation without a heteroplasmic intermediate (Koehler et al, 1991). Rapid changes in mitochondrial allele frequencies were first observed in cattle (Olivo et al., 1983; Ashley et al., 1989) but similar changes have subsequently been described in many mammalian species, including humans. Analysis of human pedigrees segregating pathogenic mtDNA point mutations show large differences in the proportion of mutant mtDNAs among siblings and between generations (Larsson et al., 1992; Ghosh et al., 1996; Jacobi et al., 2001).

The rapid changes in mtDNA allele frequency have been explained by an mtDNA bottleneck in the germ line, responsible for the random segregation of mtDNA sequence variations between generations. A mature oocyte contains about 100,000-200,000 copies of mtDNA (Figure 4). These mtDNA copies are randomly divided, without mtDNA replication, into cells of the developing embryo. Replication of mtDNA is initiated when the primordial germ cells (PGC) migrate and differentiate to generate oocytes. It is estimated that human PGCs contain ~10-100 mitochondria, increasing to ~200-1000 in the oogonia and to several thousands in the oocyte in the primordial follicle. As the oocyte matures the number of mitochondria again increases to >100,000 (Jenuth et al., 1996; Shoubridge, 2000; Khrapko, 2008; Stewart et al., 2008). Segregation is rapid because the precursor cells contain relatively small numbers of mtDNA templates and because the replication of mtDNA is under relaxed control. This relaxed form of replication allows some templates to replicate more than once during the cell cycle, while others are not replicating at all. Because of the mitochondrial bottleneck, the mtDNA genotype in the embryo is largely determined by the genotype of the few mitochondria in the PGC in the mother (Jenuth et al., 1996; Shoubridge, 2000).

**mtDNA in development**

The mitochondria produce the energy needed for cellular processes by producing ATP through oxidative phosphorylation. Through dysfunctions of energy metabolism or regulation of apoptosis, the mitochondria are suggested to contribute to human oocyte wastage and early embryo demise (Van Blerkom, 2004). The role of mtDNA mutations in
human early development is supported by a report describing a family with an unusual homoplasmic tRNA mutation that resulted in six neonatal deaths and one surviving child with Leigh syndrome even though the mother is clinically unaffected (McFarland et al., 2002). As the proportion of mutant mtDNA copies can shift when inherited the phenotype often shows remarkable variation within a family (Ghosh et al., 1996; Blok et al., 1997; Jacobi et al., 2001). Accordingly, such a shift can occur from a mother with mild or no phenotype to an embryo with significant enrichment of mutated mtDNA causing foetal loss and possibly RM. Homoplasmy for a severe pathogenic mtDNA mutation is rarely observed, presumably because such a condition would be lethal. For example mutations causing mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS) and myoclonic epilepsy and ragged-red fibres (MERRF) are invariably heteroplasmic and homoplasmic mutations are presumed to be fatal (Jenuth et al., 1996; Howell et al., 2000). An oocyte with such a mutation in a homoplasmic state would most likely result in foetal death at an early developmental stage. It has been suggested that this phenomenon might even be beneficial, if spreading of the deleterious mtDNA mutations that slip through the bottleneck would thus be eliminated (Bergstrom and Pritchard, 1998; Howell et al., 2000).

**Figure 4.** The mitochondrial bottleneck. During oocyte maturation a limited number of mtDNA copies are transferred to each primary oocyte. As this small population of mtDNA copies is rapidly replicated it may lead to a random shift of the mutational load in the next generation. Due to the mitochondrial bottleneck the offspring of a heteroplasmic female may receive a varying number of mutant mtDNA copies at fertilization resulting in different phenotypes; A) low level of mutation (unaffected offspring), B) intermediate level of mutation, C) high level of mutation (affected offspring). Modified from Taylor and Turnbull, 2005.
Aims

The present study is based on hypotheses that some as yet unknown genetic factors may result in RM. Consequently, the main aim of this study was to gain new information about the underlying genetic causes of RM in the Finnish population. The aim was to identify candidate genes for RM and subsequently screen these genes for mutations in a series of samples with RM. In addition, we aimed to study if genetic factors previously reported to underlie RM could be the cause of RM in Finnish patients.

The following specific aims were set for the study:

1. To determine whether ancestral places of origin of the couples with RM show geographic clustering.
2. To search for mutations in four candidate genes for RM — Amnionless, Thrombomodulin, Endothelial Protein C Receptor and p53 — identified to be crucial for development in mouse models.
3. To study the expression of p53 in placental samples.
4. To search for variations associated with RM in the mitochondrial genome.
5. To study if specific sex chromosome characteristics, namely skewed X chromosome inactivation or Y chromosome microdeletions, are associated with RM in Finnish patients.
Materials and methods

1 Study subjects

1.1 Patients (I, II, III, IV, V)

The patients recruited for this study have been treated for unexplained RM at the Department of Gynaecology and Obstetrics of the Helsinki University Hospital during the years 2001-2004. The inclusion criteria for the 48 women included in the study were:

1) age 18-40 years (mean age 31 years)
2) previous history of RM, defined as three or more consecutive miscarriages
3) normal karyotype
4) no uterine abnormalities (examined by ultrasonography or hysterosonogram)

Blood samples were obtained from male partners of 40 of the women included in the study. The karyotypes of these males were checked and found to be normal. Consequently, a total of 40 couples and 8 women with unexplained RM were included in the study. The women in the study had experienced three to seven miscarriages. In 45 women all miscarriages had taken place during the first trimester (weeks 1-12 of pregnancy). Two women had experienced a second trimester (weeks 13-26 of pregnancy) miscarriage, and one woman a third trimester (from week 27 to birth) intrauterine foetal death in addition to first trimester losses.

1.2 Controls (I, II, III, IV, V)

The control group consisted of 191 healthy women who had had at least one normal pregnancy and no known history of miscarriage. They were recruited from the same hospital during the same time period as the patients. The controls were of the same age and racial origin (Finnish Caucasian) as the patients.
1.3 Placental samples (III, IV)

One or more paraffin embedded placental sample from a spontaneously aborted pregnancy were available from seven of the RM women included in the study. Altogether, placental samples were available from 19 miscarriages.

1.4 Ethical issues (I, II, III, IV, V)

Informed consent was obtained from all participants prior to enrolment in the study. The ethics committee of the Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, has approved this study.

2 Methods

2.1 Genealogic analysis (I)

Data concerning the birthplaces of the patients´ parents and grandparents were collected to determine the ancestral places of origin for the patients and to investigate if the birthplaces were clustered to specific regions of Finland. Data concerning the birthplaces were obtained through a questionnaire. Replies were received from 82 (95%) patients. To investigate the geographic distribution of ancestral birthplaces, the birthplaces were placed onto a map of Finland, including the formerly Finnish areas of Russian Karelia and compared to the population distribution of Finland.

2.2 DNA extraction (I, III)

DNA was extracted according to the manufacturer´s instructions from whole blood samples and paraffin sections using Puregene DNA isolation kits (Gentra Systems, Minneapolis, USA).
2.3 Candidate gene analysis

**Polymerase chain reaction and dHPLC sample analysis (I, II, III, IV)**
The exons and exon-intron boundaries of the candidate genes were amplified using polymerase chain reactions (PCR). Amplification of appropriately sized PCR products were confirmed by agarose gel electrophoresis before further analysis. Mutation screening of the candidate genes was mostly carried out using denaturing high performance liquid chromatography (dHPLC) with the Transgenomic WAVE Nucleic Acid Fragment Analysis System (*Transgenomic, Omaha, USA*) and the associated Navigator software. To obtain optimal resolution of homoduplex and heteroduplex DNA fragments the temperature was set for partially denaturing conditions. The melting profile of the amplicons and the optimal oven temperature was predicted for each PCR fragment using the Navigator software. The buffer gradients for the elution of the fragments were created automatically with the Navigator software.

To allow sequence analysis of low-level heteroplasmic mtDNA, fractions of small sized heteroduplex peaks were collected using an FCW-200 fraction collector integrated with the WAVE system. To increase the heteroplasmy level in the samples the DNA fractions were reamplified with the primers used to produce the original amplicon.

**Sequencing (I, II, III, IV)**
Following dHPLC analysis, samples showing heterozygous peaks were sequenced in order to determine the nature of the sequence change. PCR products were purified and sequenced using BigDye version 3.1 sequencing chemistry and an ABI 3730 DNA Analyzer (*Applied Biosystems, Foster City, USA*) according to the manufacturer's instructions.

**Genotype determination (I, II, III)**
Variations that were detected in a homozygous state by sequencing were genotyped by restriction enzymes, sequencing, allele specific PCR, or dHPLC.

**Solid-phase minisequencing (IV)**
Quantification of heteroplasmic levels were performed using solid-phase minisequencing according to a previously described protocol (Suomalainen and Syvänen, 2000). PCR
products were amplified using specific primers designed for each of the variations studied. Biotinylated PCR products were then captured in streptavidin-coated microtiter wells and single nucleotide extension of specific detection primers were performed using radioactively labelled nucleotides. The results were expressed as numeric cpm-values, where the incorporated label expressed the relative amount of sequence variation in the sample.

**Quantitative real-time PCR (IV)**
The mitochondrial cytochrome b (Cytb) and the nuclear amyloid beta precursor protein (APP) genes were simultaneously amplified by quantitative TaqMan real-time PCR assay in an ABI Prism 7000 Detection System Cycler to quantify the amount of mtDNA. Fluorescence acquisition was done at the end of each annealing step, and the relative levels of the genes were compared in the exponential phase of PCR, using the ABI Prism 7000 SDS software (Applied Biosystems). The ratio of mtDNA to nuclear DNA was used as a measure of mtDNA content (2^ΔΔCT method) (Livak and Schmittgen, 2001).

**Predicting the effects of variations (I, II, III, IV)**
The ESEfinder (http://exon.cshl.edu/ESE/) is a web based resource which scans nucleotide sequences to identify putative exonic splicing enhancers (ESEs), which act as binding sites for splicing factors. The ESEfinder was used to predict if the exonic or intronic sequence variations disrupt the sequence of known ESE-elements. The ESEfinder searches for putative ESE motifs and calculates a score for all sequences with a motif match. A sequence is considered to be potentially needed for splicing when the calculated score is greater than the threshold value defined by the program.

The variations predicted to change an amino acid were analyzed by the SIFT (sorting intolerant from tolerant) program (http://blocks.fhcrc.org/sift/SIFT.html) which predicts whether an amino acid substitution in a protein will be tolerated.

Different web based databases were used to determine if the detected variations have been previously reported; Ensemble Genome browser (http://www.ensembl.org), UCSC Genome browser (http://genome.ucsc.edu), NCBI Entrez SNP database (http://www.ncbi.nlm.nih.gov), p53 knowledgebase (http://p53.bii.a-star.edu.sg), Mitomap (http://www.mitomap.org), and mtDB-Human Mitochondrial Genome Database (http://www.genpat.uu.se/mtDB/).
2.4 Immunohistochemical analysis of p53 (III)

5 μm-thick sections were cut from paraffin blocks containing placental samples. The sections were deparaffinized in xylene and rehydrated through graded concentrations of ethanol to distilled water. Monoclonal antibody DO-7 directed against p53 protein (dilution 1:100; Dako, Glostrup, Denmark) was used as primary antibody (incubation time 55 min). The procedure was performed in a TechMate 500 automated machine (DAB detection kit; Dako ChemMate). The sections were lightly counterstained with Mayer's hematoxylin.

2.5 X chromosome inactivation analysis (V)

The XCI ratio was determined in the females by analysing the methylation status of the X-linked androgen receptor (AR). Exon 1 of the AR gene contains a polymorphic microsatellite with varying numbers of trinucleotide repeats (CAG)n which can be used to identify the different alleles of the locus. DNA was digested with a methylation sensitive restriction enzyme that digests only nonmethylated (active X) DNA segments. Methylated sites can not be digested and remain intact for amplification.

Fragment analysis of the microsatellite was performed using an ABI 3730 DNA Analyzer (Applied Biosystems). Fragment analysis data were analysed and genotyped using GeneMapper 3.0 software. The total fluorescent peak areas for both alleles in digested and undigested samples were recorded and used for calculation of the ratio of allele inactivation. Previously described equations (Lau et al., 1997; Iitsuka et al., 2001) were used to calculate the degree of inactivation. We used both 85% (mildly skewed) and 90% (extremely skewed) inactivation of a particular X chromosome as cut off points for skewed X-inactivation.

2.6 Y chromosome microdeletion analysis (V)

Males were screened for the presence of microdeletions using Y chromosome specific sequence tagged sites (STS) primer sets. Altogether, 37 STS loci spanning the whole Y chromosome were analysed in multiplex PCR reactions. The presence of the specific Y chromosome markers were shown by observing the PCR fragments on agarose gels.
2.7 Statistical analysis (I, II, III, IV, V)

$X^2$ and Fisher’s exact tests ([SISA 1997 programs] http://home.clara.net/sisa) were used for statistical analyses of the data. The odds ratios (OR) and 95% confidence intervals (CI) of the OR were calculated to assess the relative risk conferred by a particular allele or genotype. Fisher’s exact test was also used to compare the frequency of skewed XCI in the patient and control groups. The T-test was used to calculate differences in mtDNA copy number levels after real-time PCR analyses. Differences were considered as statistically significant for p-values <0.05.
Results and discussion

The fact that the observed incidence of RM is much higher than expected by chance alone suggests that RM is a clinical entity distinct from sporadic miscarriages (Regan, 1991). This is supported by the finding that RM tends to occur, unlike sporadic miscarriage, when the foetus has a normal chromosome constitution (Stirrat, 1990; Sullivan et al., 2004). Recently, the genetic knowledge about the aetiology of diseases has expanded rapidly. Progress has, however, been much slower in elucidating the genetic causes of female infertility and RM. Although many studies have been conducted to identify the underlying mechanisms, the cause of miscarriage remains unknown in approximately 50% of cases (Plouffe et al., 1992; Clifford et al., 1994). Consequently, this study was conducted to gain further information about the possible genetic causes of RM in our Finnish patients.

1 Genealogic studies (I)

Genealogical studies can be used to identify geographic enrichment of genetic disorders. Depending on the age of the founder mutation the geographical area of such an enrichment can be very restricted or include large regions. In many diseases of the Finnish disease heritage a regional clustering of cases can still be observed, reflecting the geographical origins of ancestral birthplaces (Norio, 2003a; Norio, 2003b). For example, a specific type of ovarian failure, caused by mutations in the follicle stimulating hormone receptor (FSHR), called follicle stimulating hormone-resistant ovaries (FSH-RO) is an autosomal recessive trait which shows geographical enrichment of ancestral birthplaces (Aittomäki et al., 1995; Aittomäki et al., 1996). We conducted genealogic studies to investigate if such an enrichment reflecting a common genetic cause could underlie the miscarriages in our patients. The ancestors of all couples with RM were traced by questionnaires enquiring as to the full names, dates and places of birth of the patients’ parents and grandparents. Replies were received from 82 patients. To investigate the geographic distribution of the birthplaces, the birthplaces of the parents and grandparents were placed onto a map of Finland (Figure 5) and compared to a reference map showing the distribution of the Finnish population.
The results of the genealogic studies suggest the possibility of a shared underlying genetic defect for the miscarriages in a subset of patients. Compared to the reference map of the general population distribution in Finland the geographical distribution of the birthplaces of the parents and grandparents of the patients is uneven (Figure 5). The results suggest that there may be an eastern enrichment in the province of Kuopio of the ancestral birthplaces. In addition, the Helsinki region is over-represented, but this is explained by the fact that the patients were all collected from Helsinki. The clustering around Kuopio suggests that some of the patients may, due to common ancestors, carry a mutation increasing the susceptibility to RM. This would not be exceptional in Finland due to the population history, which has increased the local incidence of rare disorders. The special assortment of rare diseases among Finns is caused by the fact that the population living in Finland today is descended from a relatively small original population. As the population has increased, some alleles increased in frequency while others disappeared due to genetic drift, population bottlenecks, and non-assortative mating (Norio, 2003a; Norio, 2003b).

Figure 5. Birthplaces of the A) parents and B) grandparents of 82 patients with RM. 34 of the parents and 24 of the grandparents were born in Helsinki. The population density of Finland at the time the parents were born is shown in the small reference map.
Candidate gene analysis (I, II, III, IV)

The results of the genealogic studies suggest a common underlying genetic cause for RM in at least a subgroup of our patients. Therefore, we screened a number of candidate genes for mutations that could possibly explain the RM experienced by the couples. The candidate genes AMN, TM, EPCR, and p53 were selected based on findings in animal models. When screening candidate genes in which homozygous mutations are expected to be lethal in early pregnancy, the affected foetuses cannot, in most cases, be studied because DNA from the spontaneously aborted foetuses is rarely available. Such mutations can be studied using placental tissue samples or by screening couples experiencing RM, in which case partners are expected to be healthy heterozygous carriers of mutations in the candidate gene of interest. In addition to candidate genes identified from animal models, the genes of the maternally inherited mitochondrial genome were chosen as candidate genes based on their known functions.

2.1 Amnionless (I)

Due to the fact that homozygous mutations in Amn are known to cause miscarriage in the mouse (Wang et al., 1996) and the human AMN and mouse Amn genes are highly homologous (Kalantry et al., 2001) AMN was considered a candidate gene for miscarriages in humans. Therefore, we examined whether there is an association between RM and sequence variations of the AMN gene in humans. To our knowledge this is the first study concerning the role of AMN in miscarriage.

2.1.1 Variations in the Amnionless gene

A total of 14 sequence variations were found by screening 85 patients (40 couples and 5 women) with a history of unexplained RM and 95 controls for mutations in the 12 exons and the untranslated regions (UTR) of the Amnionless gene. Eleven of the detected sequence variations are intronic. Of these variations, 10 are single nucleotide substitutions, and one variation is a 10 basepair (GCGTGGCGTG) duplication. Three of the sequence variations are exonic, of which two are single nucleotide substitutions and one is a duplication (Table 3). None of the exonic variations detected in AMN had been previously
reported (databases used: www.ncbi.nlm.nih.gov; www.ensembl.org). In total, 8 novel sequence variations were detected in the AMN gene (Table 3).

The sequence variation c.363G>A is a single nucleotide substitution in exon 5, changing the last nucleotide of codon 121 (GGG->GGA). It is a synonymous variation, detected only in patient samples and does not predict a change of the amino acid (glycine). Sequence variation c.829A>G is a non-synonymous SNP in exon 8, changing the first nucleotide of codon 276 (ACC->GCC). The variation is predicted to change the amino acid from threonine to alanine. According to the SIFT program this amino acid change would be tolerated even though the hydroxyl group of threonine makes it much more hydrophilic and reactive than alanine. Sequence variation c.1339_1344dup is a duplication in exon 12. It is a 6 basepair duplication (GCCGGG) of the codons 447 and 448, located 5 codons before the stop codon. This DNA sequence change is predicted to make the final protein product 2 amino acids (Ala+Gly) longer.

Table 3. The location of the variations in the AMN gene, predicted amino acid changes, and the numbers of patients and controls carrying the rarer allele in a hetero/homozygous state. Numbering of the nucleotides is relative to the adenine in the ATG start codon in the reference sequence (UCSC Genome Browser, http://genome.ucsc.edu/).

<table>
<thead>
<tr>
<th>DNA variation (predicted amino acid change)</th>
<th>Location</th>
<th>RM women+ male partners n=85</th>
<th>RM women n=45</th>
<th>controls n=95</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.-87C&gt;G</td>
<td>Us 5´UTR</td>
<td>13</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>c.-74C&gt;T</td>
<td>Us 5´UTR</td>
<td>10</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>c.-27T&gt;C</td>
<td>Us 5´UTR</td>
<td>48 (15)</td>
<td>24(9)</td>
<td>50 (19)</td>
</tr>
<tr>
<td>c.-23G&gt;C</td>
<td>Us 5´UTR</td>
<td>46 (9)</td>
<td>22(5)</td>
<td>44 (9)</td>
</tr>
<tr>
<td>c.296-75_-.66dup 1</td>
<td>intron 4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c.363G&gt;A 1</td>
<td>exon 5</td>
<td>4</td>
<td>4*</td>
<td>0</td>
</tr>
<tr>
<td>c.829A&gt;G (Thr276Ala) 1</td>
<td>exon 8</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c.843+11C&gt;T 1</td>
<td>intron 8</td>
<td>5</td>
<td>5*</td>
<td>1</td>
</tr>
<tr>
<td>c.1169+42C&gt;G 1</td>
<td>intron 10</td>
<td>43 (12)</td>
<td>27(6)</td>
<td>47 (14)</td>
</tr>
<tr>
<td>c.1170-6C&gt;T 1</td>
<td>intron 10</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c.1339_1344dup 1</td>
<td>exon 12</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>c.1362+38G&gt;C 1</td>
<td>3´UTR</td>
<td>34 (11)</td>
<td>22(3)</td>
<td>40 (14)</td>
</tr>
<tr>
<td>c.1362+518C&gt;T 1</td>
<td>Ds 3´UTR</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c.1362+523G&gt;A 1</td>
<td>Ds 3´UTR</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

1 Novel variation
Us=upstream of, Ds=downstream of
* p < 0.05
Two of the novel variations; c.363G>A in exon 5, and c.843+11C>T in intron 8 were significantly more frequent in RM women compared to control women, indicating that these variations may increase the risk of miscarriage. The c.363G>A variation was detected in 4 RM women (8.7%) and in none of the control women (p=0.010). The c.843+11C>T variation was detected in 5 RM women (10.9%) and in one of the control women (1%) (p=0.014).

2.2 Thrombomodulin and Endothelial Protein C Receptor (II, unpublished)

Mutations in the TM or EPCR genes may cause thrombophilia in the mother, and thereby constitute a risk factor for miscarriages. In addition, homozygous mutations in the foetus may cause miscarriage by other mechanisms as well (Isermann et al., 2001). An important role for the TM-protein C-EPCR system for placental development and maintenance of pregnancy has been shown in mouse models (Healy et al., 1995; Gu et al., 2002). The relevance of these mechanisms for pregnancy associated complications in humans has, however, not yet been determined.

2.2.1 Variations in the TM gene

In this study we analysed the entire coding region of TM including exon-intron boundaries and the 5´ and 3´ UTRs for mutations in 86 patients with RM and 191 controls. As a result, one exonic and one 3´UTR sequence variation in the TM gene were detected (Table 4). These variations were detected in both patients and controls. There were no significant differences in the allele or genotype frequencies between all RM patients or RM women and controls for either of the sequence variations in the TM gene.

A novel 18 bp deletion in the 3´UTR of the TM gene, 1728+23_+40del, was detected in five patients (6%) and five controls (3%). The variation is located 23 bp downstream of the exon-intron boundary. The other variation detected in TM, c.1418C>T, is a previously reported common polymorphism resulting in an amino acid substitution from an alanine to a valine at codon 455. The variation is located in the sixth epidermal growth factor like domain of the gene that is responsible for thrombin binding and thereby necessary for protein C activation (Esmon, 1989) but has, however, previously been shown to be neutral in respect to thrombophilia (van der Velden et al., 1991). The variation is likely to be
tolerated according to the SIFT program, and was commonly detected in a homozygous state in both patients and controls in our study.

Table 4. The location of the variations in the TM gene, predicted amino acid changes, and the numbers of patients and controls carrying the rarer allele in a hetero/homozygous state. Numbering of the nucleotides is relative to the adenine in the ATG start codon in the reference sequence (UCSC Genome Browser, http://genome.ucsc.edu/).

<table>
<thead>
<tr>
<th>DNA variation (predicted amino acid change)</th>
<th>Location</th>
<th>RM women+ male partners</th>
<th>RM women</th>
<th>controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.1418C&gt;T (Ala455Val)</td>
<td>exon 1</td>
<td>41 (5)</td>
<td>26(3)</td>
<td>84 (8)</td>
</tr>
<tr>
<td>c.1728+23_+40del</td>
<td>3’UTR</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

1 Novel variation

2.2.2 Variations in the EPCR gene

As a result of screening 86 patients and 191 controls for mutations in EPCR two exonic variations and two variations in the non-coding regions of the gene were detected (Table 5). The two non-coding substitutions detected in EPCR have been previously reported as common polymorphisms. Variation c.323-20T>C is located 20 bp upstream of exon 3 and variation c.717+16G>C is located in the 3’ UTR. The c.717+16G>C polymorphism has been suggested to have a modifier effect on the risk of venous thromboembolism (VTE). Individuals homozygous for the C allele have elevated levels of activated protein C and a lower risk of VTE (Medina et al., 2004). In addition, a recent study suggests that the C allele decreases the risk of RM in women with a factor V Leiden mutation (Hopmeier et al., 2008).

Variation c.655A>G is located in exon 4. It is a non-synonymous change, predicting a Ser219Gly change. The variation is predicted to be tolerated according to the SIFT program and previous studies have concluded that this variation does not increase the risk of VTE (Medina et al., 2004; Ireland et al., 2005). The 23 bp duplication, c.323-9_336dup in EPCR exon 3, duplicates the preceding 23 bases and results in a premature STOP codon downstream of the insertion point. The mutation leads to the formation of a truncated receptor lacking the extracellular and transmembrane domains. The truncated protein does not function properly because it is not localized on the cell surface, cannot be secreted from the cells, and does not bind activated protein C (Biguzzi et al., 2001). We detected
this variation in three samples; in one female patient, in one male partner, and in one control woman. All four variations in EPCR were detected in both patients and controls. There were no significant differences in the allele or genotype frequencies between all RM patients or RM women and controls for any of the EPCR sequence variations. Based on these results we cannot conclude a role for the detected EPCR variations in RM.

Table 5. The location of the variations in the EPCR gene, predicted amino acid changes, and the numbers of patients and controls carrying the rarer allele in a hetero/homozygous state. Numbering of the nucleotides is relative to the adenine in the ATG start codon in the reference sequence (UCSC Genome Browser, http://genome.ucsc.edu/).

<table>
<thead>
<tr>
<th>DNA variation (predicted amino acid change)</th>
<th>Location</th>
<th>RM women+ male partners n=86</th>
<th>RM women n=46</th>
<th>controls n=191</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.323-20T&gt;C</td>
<td>intron 2</td>
<td>41 (15)</td>
<td>18(11)</td>
<td>95 (33)</td>
</tr>
<tr>
<td>c.323-9_336dup (TyrProGlnPheLeuSTOP)</td>
<td>exon 3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c.655A&gt;G (Ser219Gly)</td>
<td>exon 4</td>
<td>22 (1)</td>
<td>12</td>
<td>49 (2)</td>
</tr>
<tr>
<td>c.717+16G&gt;C</td>
<td>3′UTR</td>
<td>41 (15)</td>
<td>19(11)</td>
<td>93 (34)</td>
</tr>
</tbody>
</table>

2.2.3 TM/EPCR variations and outcome of heparin treatment

Of the women included in this study, 42 had been treated with heparin and/or aspirin in attempts to improve the pregnancy outcome. Many of the RM women in our study had a full term pregnancy when treated with heparin, but not all (Ulander V-M, personal communication). As mutations in the TM or EPCR genes may cause disturbances of the uteroplacental vasculature, we investigated whether TM or EPCR variations could be associated with pregnancy outcome after heparin/aspirin treatment. Unfortunately, the number of patients in each group was too small to obtain statistically significant results and therefore associations between a specific TM/EPCR variation and the outcome of the heparin treatment could not be studied.
2.3 p53 (III, unpublished)

p53 functions have proven important for normal development in mice. However, its role in human embryonic development has not been studied enough to make conclusions about its role at different developmental stages. As the level of p53 expression and the ability of p53 to interact with other proteins can be disturbed by sequence variations within the gene, we screened the p53 gene for variations in our series of RM patients.

2.3.1 Variations in the p53 gene

Nine different sequence variations were found in the p53 gene by screening 40 couples and 6 women with RM and 191 controls using dHPLC (Table 6). Six variations were confirmed by sequencing to be located in the non-coding regions, and three variations were found within the coding regions of the gene. All the variations have been previously reported as common polymorphisms (p53 knowledgebase; NCBI SNP database). Two of the exonic variations; G12032A, A13399G are synonymous and do not cause a change of an amino acid, while one of the exonic variations, G12139C, is a common non-synonymous polymorphism changing the amino acid of codon 72 from an arginine to a proline. These R72P variants have been shown to differ biologically and biochemically (Thomas et al., 1999), and have differences in their ability to interact with basic elements of the transcriptional machinery, to induce apoptosis and to suppress transformed cell growth (Dumont et al., 2003; Pim and Banks, 2004). This variation has previously been studied with respect to RM. Pietrowski et al. (2005) studied 175 RM women and found an association between the R72P polymorphism and RM, while Coulam et al. (2006) studied 205 RM women and showed no significant differences in the frequencies of R72P polymorphism between RM patients and controls. In this study, we did not detect any significant association between the R72P polymorphism and RM.

We found the polymorphism C11992A, located 29 bp upstream of exon 4, to be associated with RM. The C/A and A/A genotypes of the C11992A polymorphism were shown to be significantly more frequent (p-value <0.05) in women with RM than control women. Women carrying the C/A or A/A genotype had a more than twofold increased risk for miscarriage (p=0.0414, OR 2.083, CI 1.018-4.259). Statistical calculations using the \( \chi^2 \)-test were performed following a dominant genotype model, where the C/C genotypes
were compared against the C/A and A/A genotypes. A C/A or A/A genotype was detected in 32.6% of the women with miscarriages and in 18.9% of the controls. When comparing the allele frequencies of the C11992A variation, there was no significant difference between the patient and control groups.

Table 6. The location of the variations in the p53 gene, the predicted amino acid change, and the number of patients and controls carrying the rarer allele in a hetero/homozygous state. Numbering of the nucleotides and codons is relative to the sequence presented by the p53 knowledgebase (http://p53.bii.a-star.edu.sg) 2005.

<table>
<thead>
<tr>
<th>DNA variation (predicted amino acid change)</th>
<th>Location</th>
<th>RM women+ male partners n=86</th>
<th>RM women n=46</th>
<th>controls n=191</th>
</tr>
</thead>
<tbody>
<tr>
<td>11827 C&gt;G</td>
<td>intron 2</td>
<td>23 (6)</td>
<td>12 (4)</td>
<td>42 (9)</td>
</tr>
<tr>
<td>11992 C&gt;A</td>
<td>intron 3</td>
<td>22 (2)</td>
<td>13 (2)*</td>
<td>29 (7)</td>
</tr>
<tr>
<td>12032 G&gt;A</td>
<td>exon 4</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12139 G&gt;C (Arg72Pro)</td>
<td>exon 4</td>
<td>38 (8)</td>
<td>22 (3)</td>
<td>77 (8)</td>
</tr>
<tr>
<td>13399 A&gt;G</td>
<td>exon 6</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13964 G&gt;C</td>
<td>intron 6</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>14181 C&gt;T</td>
<td>intron 7</td>
<td>5</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>14201 T&gt;G</td>
<td>intron 7</td>
<td>5</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>14766 T&gt;C</td>
<td>intron 9</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

*p < 0.05

2.3.2 p53 expression in RM placenta

Previous studies have shown that complete absence of p53 in mice can result in lethality, reduced fertility, and high susceptibility to the effects of environmental stress. In addition, deviations in expression levels of p53 can have serious detrimental consequences (Sah et al., 1995; Nicol et al., 1995; Godley et al., 1996). Furthermore, a previous study showed higher expression levels of apoptosis related genes in chorionic villi from RM patients compared to controls (Choi et al., 2003). To further study the role of p53 in human development, we studied the expression of p53 in placental tissue samples from miscarriages of RM patients.

Placental tissue samples were available from 19 miscarriages from seven women. These were subjected to immunohistological staining to study the expression of p53 in the placenta. No abnormal expression patterns were detected in these samples (Figure 6). All
samples showed heterogenic and primarily weak p53 expression. Stronger expression was mainly seen in the cytotrophoblasts and syncytiotrophoblasts.

Ten of the placental tissue samples were obtained from couples who both had a C/C genotype at position 11992, and 9 of the samples were from couples where one partner was a C/A heterozygous and the other a C/C homozygote. According to parental genotypes none of the pregnancies from which placental tissue were obtained were A/A homozygotes and hence we were not able to study whether the 11992A/A genotype would affect the expression of p53 in the placenta.

![Figure 6](image)

Figure 6. A representative p53 staining of placental tissue obtained from a miscarriage. The expression patterns were similar in all 19 tissue samples. Stronger heterogeneous p53 expression is seen in the cytotrophoblasts and syncytiotrophoblasts.

### 2.4 Mitochondrial genes (IV)

Recent studies suggest that mitochondrial functions in the oocyte may have a critical role in human embryonic development. Dysfunctions that affect the capacity of the mitochondria to produce ATP or regulate the apoptosis cascade have been suggested to cause early demise of the embryo in humans (Van Blerkom et al., 1998; Van Blerkom, 2004). Accordingly, we hypothesized that some women with RM could be healthy carriers of low-level heteroplasmic mtDNA mutations which would predispose to RM.
2.4.1 Variations in the mitochondrial genome

In this study we screened 48 RM women and 48 age matched control women for heteroplasmic mtDNA variations using dHPLC. The method is based on detection of heteroduplexes in a sample and can therefore be used to detect heteroplasmic sequence variations. However, by sequencing a number of homoplasmic deviations from the reference sequence were also detected. Because all samples were not sequenced, we could not assess the frequency of homoplasmic changes.

Table 7. The location of the variations in the mtDNA, predicted amino acid changes, and the numbers of RM women and controls carrying the variation in a heteroplasmic state. Numbering of the nucleotides are based on the reference sequence in the Mitomap database.

<table>
<thead>
<tr>
<th>DNA variation (predicted amino acid change)</th>
<th>Location</th>
<th>RM women</th>
<th>control women</th>
</tr>
</thead>
<tbody>
<tr>
<td>16362 T&gt;C</td>
<td>non-coding</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16519 T&gt;C</td>
<td>non-coding</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>73 A&gt;G</td>
<td>non-coding</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>146 T&gt;C</td>
<td>non-coding</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>152 T&gt;C</td>
<td>non-coding</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>195 T&gt;C</td>
<td>non-coding</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>573 ins +1/2 bp</td>
<td>non-coding</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4232 T&gt;C (Ile309Thr)</td>
<td>ND1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4703 T&gt;C</td>
<td>ND2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4722 A&gt;G (Thr85Ala)</td>
<td>ND2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4727 A&gt;G</td>
<td>ND2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4745 A&gt;G</td>
<td>ND2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8265 T&gt;C (Leu27Pro)</td>
<td>CO2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9758 T&gt;C</td>
<td>CO3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11404 A&gt;G</td>
<td>ND4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14653 C&gt;T 1</td>
<td>ND6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14569 G&gt;A</td>
<td>ND6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14766 T&gt;C (Ile7Thr)</td>
<td>CYB</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15257 G&gt;A (Asp171Asn)</td>
<td>CYB</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Novel variation

The 19 different heteroplasmic variations detected are listed in Table 7. A heteroplasmic mtDNA variation was detected in 13 RM women (27%) and 9 control women (19%). Of the detected 19 variations, 18 have previously been reported as polymorphic variations in the Mitomap or mtDB database, while the variation 14653C>T, a synonymous change in the ND6 gene, has not been previously reported. Altogether, 12 different variations were detected within seven mitochondrial protein coding genes. No
heteroplasmic variations were detected within the mitochondrial tRNA and mitochondrial rRNA regions. Within the non-coding control region seven heteroplasmic variations were detected. Seven of the variations located within the coding regions are synonymous changes, while five of the changes are non-synonymous, predicted to change the amino acid. A non-synonymous change was detected in four RM women and four control women. The heteroplasmy levels of the non-synonymous variations were studied using solid-phase minisequencing and the heteroplasmy levels are presented in Table 8.

To study if the 7 variations detected in the mtDNA control region affected the mtDNA copy number, the mtDNA levels in samples with a variation in the control region (n=8) were analyzed using quantitative real-time PCR (Q-PCR) and compared to the mtDNA levels in samples with no heteroplasmic mtDNA variations (n=6). We did not detect significant differences (p=0.3568) in the mtDNA copy number between samples with a control region variation and control samples with no variations (Figure 7).

Table 8. Heteroplasmy levels of mtDNA variations.

<table>
<thead>
<tr>
<th>Variation</th>
<th>Sample</th>
<th>Level of heteroplasmy</th>
</tr>
</thead>
<tbody>
<tr>
<td>195 T&gt;C</td>
<td>RM patient</td>
<td>85 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>50 %</td>
</tr>
<tr>
<td>4232 T&gt;C (Ile&gt;Thr)</td>
<td>RM patient</td>
<td>10 %</td>
</tr>
<tr>
<td>4722 A&gt;G (Thr&gt;Ala)</td>
<td>RM patient</td>
<td>30 %</td>
</tr>
<tr>
<td>8265 T&gt;C (Leu&gt;Pro)</td>
<td>control</td>
<td>75 %</td>
</tr>
<tr>
<td>11404 A&gt;G</td>
<td>RM patient</td>
<td>10 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>&lt;5 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>15 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>15 %</td>
</tr>
<tr>
<td>14653 C&gt;T</td>
<td>RM patient</td>
<td>85 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>65 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>&gt;95%</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>70 %</td>
</tr>
<tr>
<td></td>
<td>placenta</td>
<td>40 %</td>
</tr>
<tr>
<td>14766 T&gt;C (Ile&gt;Thr)</td>
<td>RM patient</td>
<td>&lt;5 %</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>&lt;5 %</td>
</tr>
<tr>
<td></td>
<td>control</td>
<td>&lt;95%</td>
</tr>
<tr>
<td>15257 G&gt;A (Asp&gt;Asn)</td>
<td>control</td>
<td>25 %</td>
</tr>
</tbody>
</table>
2.4.2 Analysis of placental samples

Analysis of human pedigrees segregating pathogenic mtDNA point mutations show large differences in the proportion of mutant mtDNAs among siblings and between generations as a heteroplasmic variation is transmitted from the mother to the offspring (Ghosh et al., 1996; Blok et al., 1997). The rapid changes in mitochondrial DNA allele frequency observed between generations have been explained by an mtDNA bottleneck in the germ line. Because of this bottleneck, the mtDNA genotype in the embryo is largely determined by the genotype of the few mitochondria in the primordial germ cell (Jenuth et al., 1996; Shoubridge, 2000). Due to the mitochondrial genetic bottleneck new mtDNA mutations are either lost during transmission or quickly reach very high levels.

Figure 7. MtDNA copy number analysis by Q-PCR in cases with mtDNA control region variations. Eight lymphocyte samples and one placental sample with at least one variation in the mtDNA control region were compared to control lymphocyte (n=6) and placental samples (n=6) without mtDNA variations.

To study the transmission of the mtDNA variations from the mother to the offspring we studied the presence of the previously mentioned variations in placental samples. Unfortunately, placental samples were available from only three women with a heteroplasmic mtDNA variation; 195T>C (1 placental sample), 11404A>G (3 placental samples), and 14653C>T (4 placental samples). By analyzing these 8 placental samples using dHPLC, we were able to show that the variations were present in the placental tissue. The heteroplasmy level of these variations was studied using solid-phase minisequencing (Table 8). In three cases the proportion of the mtDNA variation was increased in the placenta compared to the mother, and in the remaining cases the
proportion was decreased in the placenta. This result is similar to previous studies showing variable levels of heteroplasmic variations in the offspring (Ghosh et al., 1996; Blok et al., 1997).

In addition, the mtDNA level of one placental sample with a control region variation was compared to the mtDNA levels in placental samples without a control region variation (n=6) using Q-PCR (Figure 7). No significant differences in the mtDNA copy number levels were detected.

3 The impact of polymorphisms on phenotype

Single nucleotide polymorphisms (SNPs) are the most common type of human genetic variation and to date over 3 million SNPs have been reported (Balakrishnan et al., 2005). Functional SNPs can be categorized as coding SNPs that change amino acids, and regulatory SNPs that modulate expression or splicing of the gene. However, the role of a specific polymorphism in a disease or a phenotype may be difficult to assess, especially if the polymorphism is located in the non-coding regions of the genome. Most polymorphisms are neutral but a number of SNPs in both coding and non-coding regions of the genome have also been reported to have phenotypic effects (Suh and Vijg, 2005; Balakrishnan et al., 2005).

3.1 Variations in the nuclear genome (I, II, III, unpublished)

Most of the variations detected in the candidate genes AMN, TM, EPCR, and p53 were more frequent than 1% in the control population, which is one of the definitions of a genetic polymorphism. In recessive disorders it is, however, not uncommon to have carrier frequencies of known deleterious mutations >1%, particularly in certain populations. Considering the frequency of miscarriage in the general population, one would expect a fairly high carrier frequency, if there was a relatively common mutation/variation underlying miscarriages. Therefore, the frequency of a possible mutation alone is not always a good indicator of whether a variation is a polymorphism or a mutation. It is also difficult to determine the function of a SNP on the basis of the nucleotide sequence alone.

This is particularly true when the SNP neither alters an amino acid nor disrupts a well-characterized motif that affects protein function or structure (Strahan and Read, 2004).
In general, nucleotide substitutions within exons may influence the function of the protein through changes in amino acid composition. Exonic variations that do not result in a change to the amino acid sequence are called silent mutations, and their phenotypic effect is often unclear. However, variations in the non-coding regions may also disturb or alter protein function. SNPs in the promoter region of a gene may influence the transcriptional activity of the gene by modulating transcription factor binding and ultimately levels of the gene product. Polymorphisms in the noncoding sequences of introns and the 3' and 5' UTRs may also result in defects in mRNA processing or stability (Rebbeck et al., 2004; Malcolm, 2005; Suh and Vijg, 2005). Recently, a growing number of polymorphisms within the non-coding regions of genes have been shown to alter the expression or the splicing of a protein (Varley et al., 1998; Attanasio et al., 2001; Wang XB et al., 2002; Wan et al., 2005; Bourdon et al., 2005). Therefore, some of the polymorphisms detected in this study may disturb the normal functions of the proteins encoded by the genes and thereby contribute to the etiology of miscarriage. It would be interesting to further study the role of some of the variations in RM, especially the deletion in TM, the duplication in EPCR, and the novel variations c.363G>A and c.843+11C>T in AMN shown to be more frequent in RM women.

Although the majority of p53 mutations are missense mutations, generally causing a loss in the ability to bind DNA, intronic variations may also alter the function of the protein, or the binding sites for interacting proteins (Cho et al., 1994). Single base pair substitutions in intron 4 of the p53 gene have been shown to disturb the binding of unidentified transcription factors, resulting in decreased expression of mouse p53 (Beenken et al., 1991). This indicates that the p53 11992 variation may also affect protein function. In humans, functional analysis has demonstrated elevated functional activity of variation G13964C in intron 6, which further indicates the importance of intronic sequences in regulating p53 function (Lehman et al., 2000). In addition, other intronic base substitutions and duplications/deletions in the p53 gene sequence have been associated with an increased risk for cancer, indicating that these variations may alter splicing, expression levels or binding sites, and thereby the function of the protein (Voglino et al., 1997; Wang-Gohrke et al., 1999; Lacerada et al., 2005). The mechanism by which the introns regulate either transcription or mRNA stability is, however, unknown but these results suggest that also the intronic C11992A polymorphism, which we detected more frequently in patients than controls, may affect the expression level of p53.
Alternatively, the variation may affect the protein functions required for protecting the embryo from teratogens and environmental stress. Further studies are, however, necessary to define whether the intronic polymorphism has functional consequences and to determine if the association with a higher risk for miscarriage is clinically significant.

3.1.1 Possible splice site variations

Approximately 60% of the listed disease causing mutations are missense or nonsense mutations. In addition, mutations at splice sites make a significant contribution to human genetic diseases. Approximately 15% of disease causing mutations affect pre-mRNA splicing through exon skipping, activation of cryptic splice sites, creation of pseudo exons within introns, or intron retention (Krawczak et al., 1992; Nakai and Sakamoto, 1994; Baralle and Baralle, 2008). Therefore, we analysed the variations which were not previously defined as common polymorphisms and the novel sequence variations with the ESEfinder to see if any of these variations may affect splicing by altering the binding sites of Ser/Arg-rich (SR) proteins. The program predicted that three single nucleotide substitutions detected in the AMN gene and the deletion detected in the TM gene may affect splicing by deleting at least one binding site for known SR-proteins and may therefore result in exon skipping or intron retention (Table 9 and Figure 8). These variations may affect splicing of the gene and thereby its function. Conformation of the phenotypic effect of these variations, however, requires further studies.

The ESEfinder was mainly developed for analysing exonic sequences, but has also been used to analyze and identify mutations in introns that affect splicing (Gabut et al., 2005; Stogmann et al., 2006). However, the program only makes predictions about the effect a variation may have on mRNA splicing. Thus, the presence of a high score motif in a sequence does not necessarily mean that the sequence determines a splicing enhancer. In addition, the sequences with the maximum scores are not necessarily the most effective splicing enhancers. The program currently searches for ESE motifs corresponding to four SR-proteins (SF2/ASF, SC35, SRp40, and SRp55) but there are several other SR-proteins for which the ESE motifs have not yet been identified (Cartegni et al., 2003).
Table 9. Threshold (Thr) values defined by the ESEfinder and predictive values obtained for normal and variant alleles for sequence variations predicted to affect the binding sites of SR-proteins.

<table>
<thead>
<tr>
<th></th>
<th>SF2/ASF</th>
<th>SC35</th>
<th>SRp55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr = 1.956</td>
<td>Thr = 2.383</td>
<td>Thr = 2.267</td>
<td></td>
</tr>
<tr>
<td><strong>AMN c.829A&gt;G</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal allele</td>
<td>2.080 (GACACCT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant allele</td>
<td>&lt;1.956 (GACGCCT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AMN c.843+11C&gt;T</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal allele</td>
<td>2.506 (CGCGCGG)</td>
<td>2.584 (GGGCCGCG)</td>
<td></td>
</tr>
<tr>
<td>Variant allele</td>
<td>&lt;1.956 (TGCGCGG)</td>
<td>&lt;2.383 (GGGCTGCG)</td>
<td></td>
</tr>
<tr>
<td><strong>AMN c.1170-6C&gt;T</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal allele</td>
<td>2.633 (GCCCTCAG)</td>
<td>2.922 (TACGCC)</td>
<td></td>
</tr>
<tr>
<td>Variant allele</td>
<td>&lt;2.383 (GCTCTCAG)</td>
<td>&lt;2.267 (TACGCT)</td>
<td></td>
</tr>
<tr>
<td><strong>TM c.1728+23_+40del</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal allele</td>
<td>2.742 (GGAGCCTG)</td>
<td>3.831 (TCCGTC)</td>
<td></td>
</tr>
<tr>
<td>Variant allele</td>
<td>sequence deleted</td>
<td>sequence deleted</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. Diagram of the predicted values obtained by the ESEfinder for the exonic variation AMN c.829A>G. Graphics of the normal (A) and variant (B) sequence show that the binding site for SF2/ASF (black bar) present in the normal sequence (GACACCT) is abolished in the variant sequence (GACGCCT).

3.1.2 Couple analysis

The 40 couples included in this study were analysed to determine if some of the possible functional variations could be detected in both partners of a couple, thereby enabling a homozygous foetus. In two of the couples both partners had a different exonic or putative splice site affecting variation in the AMN gene. One couple had AMN variations c.363G>A and c.1170-6C>T, and another couple had AMN variations c.843+11C>T and c.1339_1344dup. In these couples there is a 25% chance for every conceptus to be a compound heterozygote for potentially deleterious AMN mutations in the two alleles. Additionally, the newly identified 1728+23_+40 deletion in TM was detected in a heterozygous state in both partners of one couple, enabling homozygosity of the deletion.
in the conceptus. This variation, located 23 bp downstream of the exon-intron boundary, may affect splicing, or alternatively may have an effect on the stability of the mRNA. Unfortunately, there were no tissue samples from miscarriages experienced by these couples to study whether these variations were inherited in a homozygous form.

3.2 Variations in the mitochondrial genome (IV)

Although mtDNA is highly polymorphic and both pathogenic and silent variations are generated in the mtDNA genome, the occurrence of heteroplasmic mtDNA sequence variations is rare (Jenuth et al., 1996; McKenzie et al., 2004). For example, MELAS and MERRF mutations are invariably heteroplasmic and homoplasmic mutations in these genes are expected to be lethal (Howell et al., 2000). As the presence of two different mtDNA populations in an individual is reported to be rare (Jenuth et al., 1996; DiMauro and Schon, 2001; McKenzie et al., 2004), the percentage of samples with a heteroplasmic variation was unexpectedly high in our study. According to our study every fifth control woman carries a heteroplasmic variation, indicating that heteroplasmy is more common than previously reported. However, it may be that the high level of homoplasmy is over-emphasized as most studies rely on techniques that cannot detect low-level heteroplasmic mtDNA variations (Grzybowski et al., 2000). The fact that we detected heteroplasmy in so many of the studied samples is most likely due to the sensitivity of the detection method used in our study. In 8/19 samples (42%) in which the level of heteroplasmy was tested the ratio of variant sequences was ≤10% or ≥90%, and could not be detected by direct sequencing.

In this study, a heteroplasmic variation, predicted to change an amino acid, was detected in four RM women. All of these non-synonymous variations are previously reported as homoplasmic polymorphisms, and are therefore likely to be non-pathogenic. It is, however, possible that these variations are neutral when present in a small percentage of cells, but pathogenic when a threshold is reached or in a homoplasmic state. In general, the proportion of mutant mtDNAs in a heteroplasmic individual is expected to be very high before a biochemical or clinical phenotype is observed (Jenuth et al., 1996). Non-synonymous changes in mitochondrial genes may affect ATP production or apoptosis, thereby leading to complications during early development. In addition, the mitochondria are involved in the placental progesterone synthesis required for the maintenance of
pregnancy (Tuckey, 2005). One of the detected coding region variations is novel, not previously reported in the MitoMap or the mtDB database. This variation, 14653C>T in the ND6 gene is, however, unlikely to be pathogenic as it does not change an amino acid. In addition, this variation was detected at a heteroplasmy level of 85% in one otherwise healthy RM woman. Consequently, this novel variation is likely to be neutral.

Analysis of the mtDNA control region revealed 7 heteroplasmic mutations among RM patients. These variations in the most polymorphic region of the mitochondrial genome have previously been reported to be polymorphisms. However, mutations in the control region may affect replication of the mitochondrial genome and such mutations are suggested to be an important factor in the outcome of a pregnancy. As mitochondrial replication begins after implantation, replication defects are not likely to compromise preimplantation embryogenesis, but depending upon mutant load, may lead to post-implantation embryonic death (Van Blerkom, 2004). It has been reported that the localization of mitochondria in the early embryo is strictly regulated and the mitochondrial distribution is thought to play a role in defining embryonic patterning and long term viability of the blastomere. Therefore, mtDNA copy number may be an important determinant of oocyte quality as too few copies would result in maldistribution in the early embryo (Van Blerkom et al., 2000; Dumollard et al., 2007). To study if the detected control region variations is affecting mtDNA replication, we analyzed the mtDNA copy number in these samples. No significant differences in the mtDNA copy number were shown.

4 Analysis of sex chromosome characteristics (V)

Recent publications have reported associations between RM and skewed XCI in females and Y chromosome microdeletions in males. Therefore, both of these sex chromosome characteristics were studied in our patients.

4.1 X chromosome inactivation analysis

In recent years several studies have shown that skewed XCI, defined as preferential inactivation of one of the two X chromosomes, is increased in women with RM (Lanasa et al., 1999; Lanasa et al., 2001; Uehara et al., 2001; Kuo et al., 2008). Normally inactivation
of the X chromosomes occurs randomly on both the maternal and paternal X chromosomes during the late blastocyst state of embryogenesis (Lyon, 1961). The suggested explanation for the association between skewed XCI and RM is an unknown X chromosomal mutation resulting in both preferential inactivation of the mutated X and lethality in male embryos carrying the mutation (Lanasa et al., 1999; Uehara et al., 2001; Bagislar et al., 2006).

To study the role of XCI in RM in our Finnish patients the XCI patterns were examined in 46 women with RM and 95 control women. The XCI ratio was determined by analysing the methylation status of the X-linked AR gene using a methylation sensitive restriction enzyme that digests only nonmethylated DNA segments (Figure 9). Samples homozygous at the AR (CAG)n locus were excluded from the X-inactivation study due to the inability to distinguish between the two alleles. 3/46 women with RM and 14/95 control women were homozygous for the AR (CAG)n polymorphism and excluded from further analysis. The ratio of inactivation between the two X chromosomes was calculated for the remaining 43 female patients and 81 controls.

![Figure 9](image-url)

**Figure 9.** Fluorescence based X-inactivation analysis using a PCR based AR methylation assay. The methylation sensitive restriction enzyme *HpaII* digests and prevents amplification of nonmethylated (active X) DNA segments. (a) An example of equivalent inactivation of both alleles. (b) An example of extremely skewed inactivation of the alleles. Undigested DNA is used as control for peak intensities of the alleles.

There are several possible reasons for the discrepancy between this study and those previously reported. Firstly, it is possible that the number of patients is too small to provide statistically reliable results if skewed XCI is the cause of RM in only a small subset of patients. However, previous studies have included approximately the same number of subjects and have been able to show an association between skewed XCI and
RM (Lanasa et al., 1999; Uehara et al., 2001). Secondly, previous studies have used different inclusion criteria, which may affect the results. Our study included patients with three or more pregnancy losses. In contrast, other studies have included patients with only two pregnancy losses (Lanasa et al., 1999; Lanasa et al., 2001; Kim et al., 2004). Finally, in our study the patients and controls were of same age, the mean age in both groups being 31 years, whereas in other studies the patients were significantly older than their controls (Lanasa et al., 1999, Lanasa et al., 2001). Because skewed XCI increases in peripheral lymphocytes with age due to depletion of the haematopoietic stem cell pool the age discrepancy in other studies could have resulted in a bias towards an increase in skewed XCI in RM cases when compared with younger controls (Sharp et al., 2000; Hogge et al., 2007).

There are, however, also studies that have not detected an association between skewed XCI and RM (Sullivan et al., 2003; Kim et al., 2004; Hogge et al., 2007; Dasoula et al., 2008). Beever and coworkers (2003) evaluated the XCI status in over 200 women with RM and did not detect an association between skewed XCI and RM. Futhermore, they detected a significant excess of boys in live births among women with RM, which is contrary to expectations if the skewed XCI is due to X-linked lethal mutations (Beever et al., 2003). In some women skewed XCI may be due to the limited number of precursor cells present during the late blastocyst stage when the X chromosome inactivation process begins. At this developmental stage, there are approximately only 10–20 cells that will subsequently form the embryo; the other cells differentiate into extra embryonic tissues. Thus, some women may show skewed XCI due to chance alone (Lyon, 1971; Belmont, 1996; Gale et al., 1997). Alternatively, it is possible that skewed XCI is a heritable genetic trait that is detected in family members in multiple generations. Genetic loci that influence the selection of the inactivated X chromosome have been detected in mice (Percec et al., 2002; Chadwick et al., 2006). Whether similar loci exist in humans is not clear, but skewed XCI has been mapped in humans to Xq28 (Pegoraro et al., 1997; Naumova et al., 1998). In addition, skewed XCI appears to be inherited in some families and in these families skewed XCI may be caused by a genetic variation that is not related to RM (Naumova et al., 1996).
4.2 Y chromosome microdeletion analysis

To date the knowledge of male contribution to RM is largely limited to karyotyping. Karyotyping gives an overview of the whole genome, but reveals only large chromosomal rearrangements. Submicroscopic duplications and deletions, such as Yq microdeletions, remain unidentified using this method. Yq microdeletions are reported to cause spermatogenetic failure and male infertility (Vogt et al., 1996; Hopps et al., 2003). There are at least three regions, AZFa, AZFb, and AZFc, on the long arm of the Y chromosome, in which deletions are associated with varying degree of spermatogenic failure. Although it is widely accepted that deletions of these regions contribute to male infertility, the function of the genes located in the AZF regions is poorly understood (Krausz and McElreavey, 1999; Noordam and Repping, 2006; Ferlin et al., 2007). It has been reported that men with Y chromosome microdeletions have low sperm counts and a higher incidence of sperm chromosomal abnormalities (Forest et al., 2005). Because sperm chromosome abnormalities have been found in some men of RM couples (Rubio et al, 1999; Al-Hassan et al., 2005) Dewan and coworkers (2006) studied if Y chromosome microdeletions are associated with RM. They reported that a significant portion of male partners (82%) of women with RM had Y chromosomal microdeletions, and suggest that the Y chromosomal genes may be important for early embryonic development in male embryos (Dewan et al., 2006).

In our study we investigated if a similar association between Y chromosome microdeletions and RM was present in our set of Finnish patients. We tested 40 male partners of women with unexplained RM for the presence of Y chromosome microdeletions. In addition, to the four STSs used in the previous study (Dewan et al., 2006) we tested our patients using 33 additional STSs to enable the analysis of all AZF regions. One STS was used to test for the presence of the short arm of the Y chromosome, the other 32 STS loci were located on the long arm of the Y chromosome. Amplification products for all of the 37 STSs used were detected in all tested patients, and thereby we did not detect any Y chromosome microdeletions in our samples.

Although our study was performed on a bigger set of samples compared to the study by Dewan et al. (2006), our study revealed no microdeletions and therefore does not support the previously reported findings. It is possible that the involvement of microdeletions in the aetiology of RM differs between populations and this may be the
cause of different results between the studies. It is known that different populations have different Y chromosome haplogroups and it has been shown that the susceptibility to Y chromosomal microdeletions varies between different haplogroups (Arredi et al., 2007; Yang et al., 2008). There may also be some technical problems affecting the results of the Dewan et al. (2006) study (Noordam et al., 2006). These technical factors could be the underlying cause of the discrepancy between the previously reported results and the results of our study. According to the NCBI UniSTS Database, the STS marker DYS262 is located on the short arm of the Y chromosome, and the markers DYS220, DYF86S1, and DYF85S1 on the long arm in the AZFc region. Dewan et al. (2006) report that they found patients with both DYS262 and an additional marker deleted. This indicates that these patients would have multiple microdeletions or alternatively a region as big as 15 Mb deleted. This is, however, unlikely in otherwise healthy males.

According to guidelines provided by the European Academy of Andrology (EAA) and the European Molecular Genetics Quality Network (EMQN) for diagnostic laboratories, the use of 6 STSs should be able to detect up to 95% of all reported Y microdeletions in the AZF regions (Simoni et al., 2004). We tested our male patients using all the STSs recommended in the guidelines, and in addition, over 30 other markers spanning the Y chromosome. Therefore, it is unlikely that we would have missed at least any of the known deletions.

Although we did not detect any microdeletions, it is important to keep in mind that a normal lymphocyte karyotype does not exclude sperm abnormalities. High rates of chromosomal abnormalities in sperm have been observed in RM patients with normal lymphocyte karyotypes because the highly repetitive structure of the Y chromosome sequence predisposes to de novo deletions (Rubio et al., 1999; Al-Hassan et al., 2005). This study was performed using DNA extracted from peripheral blood and therefore, it is possible that abnormalities leading to miscarriage have arisen in the spermatocytes during spermatogenesis through de novo mutations.
5 Limitations of the study

When considering the known causes of miscarriages, the aetiology of RM seems very heterogeneous. Therefore, it is possible that mutations in the candidate genes of this study cause miscarriage in only a small subpopulation of couples with miscarriages and one may argue that more couples would be needed to determine the exact role of the candidate genes in RM. After excluding the common variations, the mutation rate was low in the candidate genes. This may be due to the fact that mutations in these genes are rare in general or that they are a rare reason for RM. Screening more patients and controls may also have revealed a statistical difference in the frequencies of the variations between patients and controls. Additionally, a larger number of controls would be needed to calculate the actual frequency of the novel variations and to determine their significance, as when studying recessive conditions or complex diseases some of the controls are also expected to be carriers of the disease causing mutation or susceptibility alleles.

Additionally, we also may have missed some mutations in the candidate genes. Although in theory dHPLC is a highly sensitive and specific method for mutation detection, the sensitivity of mutation detection by heteroduplex analysis is unlikely to be 100%. There are some rare instances where mutations may not be detected. Some variations can only be detected at one unique temperature, and variations located within a GC-rich sequence of a fragment with otherwise normal nucleotide ratios may in some cases not be detected due to the high melt temperature of GC-rich pockets (Xiao and Oefner, 2001). In this study one variation in TM (c.1418C>T) was initially not detected using dHPLC but by sequencing a random selection of samples. Another potential problem relates to the detectability of heteroplasmic mtDNA mutations in samples with a proportion of <5% or >95% mutant mtDNA, because heteroduplex analysis becomes unreliable beyond those limits (Biggin et al., 2004; Lim K et al., 2008).

Another limitation of this study is the use of blood samples for detection of heteroplasmic mtDNA variations. It has been shown that the degree of heteroplasmy for some mitochondrial DNA mutations varies considerably among tissues. Studies have reported a decrease in the proportion of mutant mtDNA in blood by approximately 0.5% to 2% per year (Howell et al., 2000; Rajasimha et al., 2008; Lim et al., 2008). As a consequence blood may contain lower levels of the mutation compared to other tissues, making the detection of low-level mutations demanding.
Conclusions

The aim of this study was to find new genetic causes for RM, a pregnancy complication that affects around 1% of couples trying to conceive. To date there are many factors known to cause RM, but the underlying cause of a couple’s miscarriages often remains unresolved as RM is a very heterogeneous condition. The identification of the underlying causes of miscarriage is important for developing more successful treatments for women experiencing RM and thereby preventing subsequent miscarriages. Identification of underlying causes also helps to identify pregnant women that are at a risk of having a miscarriage. This study aimed to identify new genetic factors underlying RM by studying candidate genes - either identified from animal models or regulating functions important in different stages of early foetal developmental - and sex chromosome characteristics previously reported to be associated with RM.

By screening the couples for mutations in the AMN, TM and EPCR genes, two novel AMN variations were shown to be associated with RM. One variation, c.363G>A, is an exonic synonymous variation, and the other, c.843+11C>T is an intronic variation predicted to affect splicing. These variations may increase the risk of RM in carrier women. Further studies would, however, be required to confirm the phenotypic effect of these variations. In addition, four interesting exonic or potential splice site disrupting variations were detected in these genes; c.1339-1344dup and c.1170-6C>T in AMN, c.1728+23_+40del in TM, and c.323-9_336dup in EPCR. In one couple both partners were heterozygous for a deletion in the TM gene and in two couples both partners had different exonic or potential splice site disrupting AMN variations. These variations may play a role in the miscarriages of these couples. However, the phenotypic effects of the potentially deleterious variations cannot be determined without further investigations. It would be interesting to study these variations in a larger sample set or alternatively, in foetal/placental samples. However, most of the variations detected in these genes are likely to be silent polymorphisms and, taken together, the results of this study suggest that mutations in AMN, TM, and EPCR do not have a major role in RM in the patients studied, even though their contribution to RM cannot be totally excluded.
Additionally, a polymorphism in the p53 gene, C11992A, was shown to be associated with RM in Finnish patients. The results indicate that women with a C/A or A/A genotype have a twofold higher risk for RM than women with a C/C genotype. As the polymorphism was, however, detected in a large number of controls as well it is likely to be a contributing factor increasing the risk of RM in combination with other sequence variations or environmental factors, rather than the sole cause of RM. It is also possible that, instead of the 11992 locus, another genetic variation linked to this locus, is the actual susceptibility locus increasing the risk for RM. While the exact role of p53 in development is unclear, it is difficult to predict whether the variation disturbs functions needed for normal embryonic development. The results of the p53 expression analysis indicate that the miscarriages experienced by the studied couples are not due to altered p53 levels in the placenta.

When screening the mitochondrial genome heteroplasmic mtDNA variations were detected in 13 RM women, which was unexpected as heteroplasmic variations are reported to be rare. Most of these variations are previously reported as polymorphisms and the novel variation is likely to be non-pathogenic as it is synonymous. However, the role of these variations in the aetiology of miscarriage cannot be excluded as some variations may be pathogenic only when a certain threshold is reached. The transmission of the variations was studied by analysing placental samples and the result is similar to previous studies showing that a heteroplasmic variation in a mother is transmitted to her offspring in different proportions. This supports the hypothesis that a woman with no disease phenotype can carry a low-level heteroplasmic variation that may cause foetal death when transmitted to the foetus in a higher proportion. However, the variations detected in this study were judged to be polymorphisms, and not likely to contribute to the miscarriages experienced by the women studied. This, however, does not exclude the role of other heteroplasmic mtDNA variations in miscarriage.

In this study sex chromosome abnormalities, recently reported to be associated with RM, were also investigated to study if these factors could underlie the RM in Finnish patients. The female patients were studied for skewed X chromosome inactivation and the male partners were studied for Y chromosome microdeletions. Based on the results of the study it was concluded that there is no association between skewed X chromosome inactivation or Y chromosome microdeletions and RM in Finnish patients.
Although no major genetic cause explaining the miscarriages in the patients could be identified, the results of the genealogic studies indicate that a subset of patients possibly have a common genetic cause underlying their miscarriages. The results of the genealogic studies suggest that there may be an eastern enrichment of ancestral birthplaces in the province of Kuopio, which was unexpected as the patients were collected from Helsinki. It would be interesting to collect more patients from this region for further genetic studies.
Acknowledgements

This study was carried out at the Folkhälsan Institute of Genetics, and the Department of Medical Genetics, University of Helsinki during 2004-2009. I wish to thank the former and current heads of these institutes, Professors Anna-Elina Lehesjoki, Kristiina Aittomäki and Päivi Peltomäki for providing excellent working facilities in Biomedicum Helsinki.

I am grateful to the following organizations for financially supporting my work: the Folkhälsan Research Foundation, the Sigrid Juselius Foundation, the Biomedicum Helsinki Foundation, the Emil Aaltonen Foundation, the Maud Kuistila Memorial Foundation, the University of Helsinki Funds, the Oskar Öflund Foundation, and Svenska Kulturfonden.

I wish to express my gratitude to my supervisor docent Kristiina Aittomäki for giving me the opportunity to work in her research group and letting me participate in a very interesting project. I am grateful to Kristiina for giving me responsibility for my projects and for her confidence in me throughout this project. I guess you knew I would end up defending my thesis eventhough I wasn´t so convinced myself all the time. I would also like to thank Kristiina for supporting me in my aim of becoming a medical geneticist.

Docent Leena Anttila and docent Nina Horelli-Kuitunen, the official reviewers of this thesis work, are warmly thanked for critical review and for valuable suggestions that improved this thesis.

Professor Päivi Peltomäki is warmly thanked for accepting the important post of custos at the dissertation.

I do owe enormous thanks to Jodie Painter, my supervisor in the lab, for encouragement and for always being willing to help me. Jodie is thanked for help in the lab, especially in the beginning of the project, invaluable help during a number of writing processes and teaching me a scientific way of thinking. Jodie is also thanked for revising the language of this thesis work and helpful comments to improve it.

I am deeply grateful to all the co-authors and collaborators in this study for their valuable contributions to this thesis work. Veli-Matti Ulander and Risto Kaaja are thanked for collecting and treating the patients included in this study, for clinical expertise and for a fruitful and nice collaboration. Ralf Bützow is thanked for providing his expertise in immunohistochemical studies. Anu Wartiovaara (née Suomalainen) and Alexandra Götz are thanked for collaboration within the mitochondrial field and valuable contributions to my thesis work under a tight schedule. Sarah Ariansen is warmly thanked for valuable collaboration and for welcoming me to visit their lab in Oslo. Taru Ahvenainen is thanked for careful lab work with the TM gene.

I am thankful to research nurses Katja Kuosa and Eija Kortelainen for providing me with exact information about the study subjects. Aila Riikonen, Jaana Welin-Haapamäki, Stephan Keskinen, Marjatta Valkama, and Solveig Halonen are acknowledged for helping with practical matters during the years.
I would like to thank Per-Henrik Groop for encouragement and interest in my studies, thesis project, and future plans. Hannele Laivuori is thanked for fruitful collaboration within the field of reproductive medicine and a positive attitude towards my work.

During my time working in the lab and writing my thesis I have greatly enjoyed the company of my research colleagues at Folkhälsan Institute of Genetics. Maria, my fellow PhD student in our small group, your company and support have been really important for me, thank you. Vilma is thanked for friendship, interesting conversations and for sharing the frustration of working with the WAVE. I would like to thank Anne S, Eija, and Nina for friendship and helping me with all kinds of questions, especially when finishing this thesis project. Kati and Ann-Liz are thanked for their positive attitude towards most things in life, and for making me smile :) Peter and Carol, the FHRM team, thank you for amusing planning meetings during the past years. Saara, Mervi, Jaakko, Otto, Maria K, Hanna V, Elina, Reetta, and all the former and present members of the Folkhälsan team are thanked for useful advice, friendship, and for creating a nice and unique working atmosphere.

I also wish to thank Joni, Annu, Helena, Pia, and Jeanette, with whom I have been sharing the office room, for nice conversations about traveling and renovations, and for patience in listening to and sharing my frustration when things were not going the way I wished. It became quite empty in our room since you left.

I want to express my warmest thanks to my dear friends, Minttu, Ammi, and Stina for always being there when I have needed you. Minja, Jenny, Anna, Mariann, Emma, and Julia are warmly thanked for friendship, for scientific and non-scientific discussions during both lunch hours and spare-time, and for making studying biology fun! It has been a great pleasure to get to know you. I wish to sincerely thank Juhis and Toffe for friendship and for convincing me since ages ago that I should do a PhD.

My deepest appreciation goes to my family for supporting me in all my decisions and an encouraging attitude towards my studies. I own my warmest thanks to Miro, my friend, my love, for his love and support during these years and for taking my thoughts off work. Thank you for taking care of the penguin.

Finally, all the patients who have voluntarily participated in this study, thereby making it possible, are warmly thanked.

Helsinki, February 2009
References


