UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Jelisaveta Džigurski

Investigating the Genetic Basis of Obstetrical Diagnoses Related to Miscarriage and Placental Biology Using Genome-Wide Association Study Meta-Analysis

Master's Thesis (30 ECTS)

Curriculum Bioengineering

Supervisors:
Associate Professor of Genomics and Reproductive Genetics, PhD Triin Laisk
Junior Research Fellow of Reproductive Genomics, MSc Natàlia Pujol Gualdo
Professor in Bioinformatics, PhD Reedik Mägi

Tartu 2023

Investigating the Genetic Basis of Obstetrical Diagnoses Related to Miscarriage and Placental Biology Using Genome-Wide Association Study Meta-Analysis

Abstract:

The triad of genetic, environmental and lifestyle factors contribute to adverse pregnancy outcomes. While the recent work investigating maternal genetic susceptibility to miscarriage has identified several genes involved in placental biology, the specific genetic risk factors of pregnancy complications remain poorly understood. To elucidate the maternal genetic basis of seven distinct obstetrical diagnoses related to miscarriage and placental biology, we conduct a case-control genome-wide association study meta-analysis using population-based data from the Estonian Biobank and the FinnGen study. We identify novel genomic risk loci and propose the most likely associated genes for diagnoses "haemorrhage in early pregnancy" and "premature rupture of membranes". In addition, we estimate SNP-heritability, provide evidence of shared genetic background between the studied traits, and describe which other diagnoses are more common in women diagnosed with pregnancies with abortive outcomes and placenta-associated pregnancy complications.

Keywords:

Genome-wide association study, obstetrical diagnoses, miscarriage, placenta

CERCS:

B110 Bioinformatics, medical informatics, biomathematics, biometrics; B570 Obstetrics, gynaecology, andrology, reproduction, sexuality

Raseduskatkemiste ning Platsentaga Seotud Raseduskomplikatsioonide Ülegenoomne Seoseuuring

Lühikokkuvõte:

Nii geneetilised ja keskkondlikud tegurid kui ka eluviis võivad suurendada raseduskomplikatsioonide riski. Kuigi hiljutised uuringud on näidanud, et platsenta bioloogiat mõjutavad geenid võivad mõjutada raseduskatkemiste riski, ei ole raseduse katkemise geneetilised riskifaktorid siiani piisavalt põhjalikult uuritud. Käesoleva magistritöö raames viidi läbi ülegenoomsete seoseuuringute metaanalüüse kasutades Eesti Geenivaramu ja FinnGen andmeid, et selgitada välja seitsme raseduse katkemisega ja platsenta bioloogiaga seotud diagnoosi geneetilisi riskifaktoreid. Töö käigus leiti uusi riskilookuseid ja pakuti välja potentsiaalseid kandidaatgeene, mis on seotud varajase rasedusaegse verejooksuga ning lootevee enneaegse puhkemisega. Lisaks hinnati uuritavate tunnuste SNP-põhist päritavust ning kirjeldati nende jagatud geneetilist komponenti. Samuti anti ülevaade, milliste teiste haiguskoodidega korreleeruvad raseduse katkemised ja platsentaga seotud tüsistused.

Võtmesõnad:

Ülegenoomne seoseuuring, sünnitusabi diagnoosid, raseduskatkemine, platsenta

CERCS:

B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika; B570 Sünnitusabi, günekoloogia, androloogia, paljunemine, seksuaalsus

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND NOTATIONS

CADD	Combined Annotation Dependent Depletion
DEPICT	Data-driven Expression Prioritized Integration for Complex Traits
eQTL	Expression Quantitative Trait Loci
EstBB	Estonian Biobank
FDR	False Discovery Rate
FUMA	Functional Mapping and Annotation Platform
GALNT6	Polypeptide N-Acetylgalactosaminyltransferase 6
GWAS	Genome-Wide Association Study
Hi-C	High-Throughput Chromatin Conformation Capture
HIC1	Hypermethylated in Cancer 1
ICD-10	International Classification of Diseases, 10th Revision
LD	Linkage Disequilibrium
LDSC	Linkage Disequilibrium Score Regression
MAF	Minor Allele Frequency
MAGMA	Multi-marker Analysis of GenoMic Annotation
MHC	Major Histocompatibility Complex
MTAG	Multi Trait Analysis of GWAS
PC	Principal Component
phepheWAS	Phenome-Wide Phenotype Association Analysis
PPROM	Preterm Premature Rupture of Membranes
PROM	Premature Rupture of Membranes
QQ plot	Quantile-Quantile plot
SLC4A8	Solute Carrier Family 4 Member 8
SMG6	SMG6 Nonsense Mediated mRNA Decay Factor
SNP	Single Nucleotide Polymorphism
SRR	Serine Racemase

INTRODUCTION

Maintaining a healthy pregnancy and carrying a baby to term can be a challenging experience for some women, and each pregnancy trimester comes with a range of potential risks and complications. For instance, the pregnancy may end in very early stages in some cases; in others, placenta-associated complications may arise and significantly disrupt the health of both the mother and the baby.

Early pregnancy loss refers to various distinctive medical diagnoses, such as threatened abortion, missed abortion, spontaneous abortion and anembryonic gestation, all commonly regarded as miscarriage. Although different definitions are in use, depending on whether the context is clinical or scientific, according to the most recent guidelines issued by the World Health Organization and the European Society of Human Reproduction and Embryology, miscarriage is the loss of pregnancy up to 24 weeks of gestation [1,2]. It affects approximately 10–25% of pregnancies [3,4], with symptoms such as bleeding and severe abdominal pain. Influenced by genetic, environmental and lifestyle factors, miscarriage is a multifactorial condition, and the exact cause may vary between individuals.

The largest comprehensive genetic study that mapped the maternal genetic susceptibility and underlying biology of miscarriage emerged in 2020 [5]. Combining data from different ancestries for sporadic and recurrent miscarriage, Laisk *et al.*, 2020, showed that miscarriage has a clear heritable component comparable to other complex traits and that the genetic signals were linked to placental biology. Placenta is a central organ in pregnancy and, thus, is associated with many pregnancy complications [6,7]. However, the above-mentioned study combined several diagnosis codes, leaving open the question of the genetic risk factors of specific diagnoses.

In the present study, we conduct genome-wide association study meta-analysis using Estonian Biobank and FinnGen data, to further investigate the genetic landscape of seven obstetrical diagnoses related to miscarriage and placental biology, classified according to the ICD-10 system [8]. The full diagnosis names and connected ICD-10 codes are: “hydatidiform mole” (O01), “other abnormal products of conception” (O02), “spontaneous abortion” (O03), “haemorrhage in early pregnancy” (O20), “premature rupture of membranes” (O42), “placenta previa” (O44) and “premature separation of placenta (abruptio placentae)” (O45). We believe that clearer understanding of the distinct features of each diagnosis could deepen the knowledge of the biology underlying these pregnancy complications and advance diagnosis identification and patient management.

1 LITERATURE REVIEW

In this chapter, we will introduce each diagnosis and outline its prevalence, symptoms, and risk factors that increase susceptibility to developing the condition. When relevant, we will also mention interesting discoveries in genetic studies conducted in the past. Subsequently, we will familiarise the readers with the fundamentals of genome-wide association study.

Throughout this thesis, we primarily use the term 'diagnosis' in the text. Still, readers should note that it is used interchangeably with the terms 'trait', 'phenotype', 'condition' and 'disease', and refer to the same concept.

1.1 Descriptions of Diagnoses

1.1.1 Pregnancy with Abortive Outcome and Other Maternal Disorders Predominantly Related to Pregnancy

According to the statistical data of the Estonian Medical Birth Registry and the Estonian Abortion Registry, the prevalence of spontaneous and other abortions (*excl.*, legally induced abortions) is around 9% [9]. However, the actual prevalence is probably higher, as most miscarriages occur in such early stages that some women may not know they are pregnant, thus, leaving miscarriage clinically undetected.

Hydatidiform Mole (O01). This gestational trophoblastic disease, commonly called molar pregnancy, occurs when placental tissue grows abnormally due to atypical oocyte fertilisation, and chromosomal abnormalities [10]. Contrary to a healthy pregnancy with a normal placenta and a fetus, in a complete molar pregnancy a mass of cysts that lack fetal tissue is formed [11]. In a partial molar pregnancy, the fetal tissue is present, but the fetus cannot survive. The prevalence in Europe is estimated to be around 1 in 1000 pregnancies [12]. Usually asymptomatic, sometimes symptoms include bleeding, uterine enlargement, and abnormal levels of pregnancy hormones [10,11]. Despite being considered benign, the hydatidiform mole can cause abnormal vaginal bleeding, and it is associated with an increased risk of developing gestational trophoblastic neoplasia, thus, becoming invasive [13]. Risk factors include young and old maternal age (under 20 and above 35 years old) [14], previous molar pregnancy, [15] and a family history of the hydatidiform mole [16], pointing to a heritable component. Although most cases are of androgenetic origin (zygote contains only paternal chromosomes), it is suspected that the oocyte immaturity is the underlying cause [11]. Several studies have reported mutations in the *NLRP7* and *KHDC3L* genes as a cause of recurrent hydatidiform mole [17–19], implicating the maternal genetic

component and autosomal recessive inheritance [17]. However, it remains unclear if common genetic variation could also increase the risk.

Other Abnormal Products of Conception (O02). This broad category encompasses several conditions related to abnormal development of the fetus and/or placenta, such as blighted ovum and nonhydatidiform mole (*i.e.*, carneous, fleshy or unspecified intrauterine mole), missed abortion, and other specified or unspecified abnormal products of conception. The specific biology of each of these conditions may vary, but they all result in a pregnancy with an abortive outcome. Blighted ovum and nonhydatiform mole are condition in which a fertilised egg implants in the uterus but does not develop into an embryo (anembryonic pregnancy) due to chromosomal or cell division abnormalities, leaving an empty gestational sac [20]. Missed abortion, also known as a silent or delayed miscarriage, is a condition in which the fetus has died, but the maternal body has not expelled the pregnancy tissues [21]. The exact information on the prevalence of these conditions in Europe is unavailable. However, one validation study regarding miscarriage-related codes in Finnish healthcare register [22] showed that out of 643 verified miscarriage diagnosis, 43.1% were classified as spontaneous abortion, 38.4% as missed abortion, 17.4% as blighted ovum and 1.1% as other abnormal products of conception. Early symptoms are identical to those in health pregnancy, but soon after, the signs of a miscarriage appear, such as abdominal cramps, vaginal spotting, and menorrhagia (heavy menstrual bleeding) [21]. Generally, the delivery of the abnormal products of conception is spontaneous. It is thought that embryonic chromosomal abnormalities and advanced maternal age, combined with environmental factors and lifestyle, all contribute to the inadequate development of the products of conception [20,21,23].

Spontaneous Abortion (O03). Spontaneous pregnancy termination resulting in the fetus's death typically occurs in the first trimester [2] and affects roughly 10–25% of pregnancies [3,4]. Common symptoms include physical complications, such as bleeding, cramping pain in the lower abdomen, and expulsion of tissue [5]. Miscarriage can lead to anxiety and depression [24], and has been associated with long-term consequences, such as an increased risk of cardiovascular disease [25,26]. Studies conducted so far indicate that the risk of miscarriage increases with maternal age [3], while other common causes include parental and embryonic/fetal chromosomal abnormalities [27,28], infections [29], obesity [30], and underlying medical conditions, such as diabetes [31] or thyroid disorders [32].

Haemorrhage in Early Pregnancy (O20). Women experience vaginal bleeding during their first trimester of gestation in approximately 20% of pregnancies [33]. Although relatively common and considered harmless since the embryo or fetus is alive, in half of the cases bleeding and lower abdominal pain are signs of miscarriage and specified as threatened abortion [10,34]. In cases where abortion does not happen, there are increased rates of later adverse pregnancy outcomes, such as preterm birth and premature separation of placenta [35]. The exact cause of threatened abortion is rarely known, and it can occur due to a variety of factors [10,36], including subchorionic haemorrhage, implantation bleeding, infections, ectopic, and molar pregnancy, and, thus, be a symptom of other diagnoses studied in this thesis.

1.1.2 Maternal Care Related to the Fetus and Amniotic Cavity and Possible Delivery Problems

The placenta is a central organ in pregnancy built up from maternal and fetal tissues. It intermediates maternal and fetal communication by creating a unique hormonal and immunological environment, transporting oxygen and nutrients to the fetus, and removing waste products [10]. Therefore, healthy placenta is essential for a healthy pregnancy and fetal development, and dysfunctional placenta can lead to various pregnancy complications, such as miscarriage, premature rupture of membranes, and stillbirth.

Premature Rupture of Membranes (O42). Protective membranes around the embryo and later fetus, called the amnion (inner membrane) and the chorion (outer membrane), build up the amniotic sac, which is filled with amniotic fluid and connects to the allantois and placenta via an umbilical cord. Together with the yolk sac, they play a vital role in developing and protecting the embryo/fetus during pregnancy [10]. Normally, these membranes rupture during labour or in some cases within 24 hours before labour (PROM). However, if it happens before the 37th week of pregnancy, this event is diagnosed as preterm premature rupture of membranes (PPROM) [37]. Approximately 5–10% of pregnancies have a PROM diagnosis [38], while PPROM affects around 3% of pregnancies [39]. The most common symptom is a sudden leaking of amniotic fluid from the vagina alongside uterine contractions. The earlier it happens, the more serious the situation is for the mother and the baby, increasing the risk of preterm birth and other complications, such as perinatal infections [40,41]. Known risk factors are smoking [42], infections of the genital tract [43], twin pregnancy [44], and previous history of PROM or PPROM [45]. A small-scale genetic

study conducted with a subset of SNPs showed a two-fold increased risk of PPROM due to the variants in the maternal *TIMP2* gene regulating extracellular matrix [46].

Placenta Praevia (O44). During labour, the cervix dilates so that the baby can enter the vaginal birth canal. However, this physiological process can be complicated by a dysfunctional placenta. Placenta previa is diagnosed when the placenta completely, partially, or marginally covers the cervix or when it lays close to the cervical *os* (so-called low-laying placenta) and prevents safe vaginal delivery of the baby [10]. The complications associated with this condition affect both the mother and the baby. The maternal complications are bleeding before, during and after childbirth, hysterectomy (removal of the uterus) [47], sepsis [48], and death [49]. The neonate complications are delays in growth, congenital malformations [50], anaemia, premature birth, and perinatal death [51]. The prevalence in Europe is low, affecting 3.6 per 1000 pregnancies (2.6 per 1000 pregnancies in Estonia) [52,9]. The typical symptoms appear in the second or third trimester as light red vaginal bleeding with or without pain, but ultrasound is necessary for a definitive diagnosis [10]. The exact cause is unclear, but identified risk factors are young and old maternal age (under 20 and above 35 years old) [53], multiparity [54], previous Caesarean delivery [55] and substance abuse [56].

Premature Separation of Placenta (Abruptio Placentae) (O45). This term is used to describe an acute condition when the normally implanted placenta prematurely separates from the inner layer of the uterus [10]. As for placenta previa, it has been shown that abruptio placentae can cause maternal morbidity and perinatal mortality [57,58]. Its prevalence in European countries is 3–6 per 1000 pregnancies (9.8 per 1000 pregnancies in Estonia) [58,59,9]. The common symptoms, such as sudden abdominal and back pain, uterine contractions, bleeding, and fetal distress, vary based on the severity of the separation [10]. Advanced maternal age [59], smoking and cocaine use [56], chronic hypertension [60], leiomyomas [61], PPROM [62] and history of previous placental abruption [59] significantly increase the risk of premature separation of placenta. However, the etiology is yet to be explained.

1.2 Genome-Wide Association Study

From its first study conducted in 2007 [63], genome-wide association studies (GWAS) have proven to be an efficient and robust method for analysing the genetics of complex diseases and traits. Commonly occurring complex diseases, such as Alzheimer's disease [64,65], depression [66,67], type 1 diabetes [68], and type 2 diabetes [69], rarely have one decisive

cause. Instead, they result from the complex interaction between genetic, environmental, and lifestyle factors. The genetic component consists mostly of genetic variants common in the population, each with a small effect on the trait or disease risk [64,66,68].

In most GWA study designs, researchers attempt to identify genetic variants that are more common in one group of individuals that have a trait of interest and investigate variants' effect on developing that trait by comparing the genetic profiles to ancestry-similar individuals who do not have the same trait of interest [70]. As the most extensively studied type of genetic variation, millions of single-nucleotide polymorphisms (SNPs) are tested across the genome to detect statistically significant associations with the trait of interest. The p-values are visualised using a so-called Manhattan plot with the statistical significance threshold set to 5×10^{-8} to account for multiple testing.

While numerous GWA studies have identified thousands of associations across different traits and disease domains, proposing underlying genes and biological pathways contributing to the development of the disease or trait [71–73], identification of causal genes and variants still remains a challenge. This is mainly due to the non-random correlations between alleles at different loci, referred to as linkage disequilibrium (LD) [74]. Therefore, the SNP with the lowest p-value in GWAS (often named the sentinel SNP, the lead SNP or lead signal) may merely be a proxy for surrounding SNPs since they are often inherited together as a group. Despite this challenge, it has been shown over the years that GWAS data can be used to identify individual genetic disease risks and variants contributing to gene-environment interactions and drug response [75,76]. Moreover, numerous methods have been developed to facilitate moving from association signals to the underlying biology [77–80].

As described earlier by Uffelmann *et al.*, 2021, and illustrated in Figure 1, the key steps of GWAS workflow include the collection of DNA samples and phenotypic information (disease status, age, sex), genotyping using whole-genome sequencing (WGS) and genome-wide SNP arrays, quality control, imputation of missing variants using haplotype phasing and reference populations, statistical association testing, and interpretation of the results using post-GWAS analyses. Additional steps, such as meta-analysis and independent replication, are used to increase the sample size, boost the discovery of novel variants, and seek confirmation of results. Throughout this thesis, each step will be explained in detail to ensure a better comprehension of the GWAS meta-analysis study design.

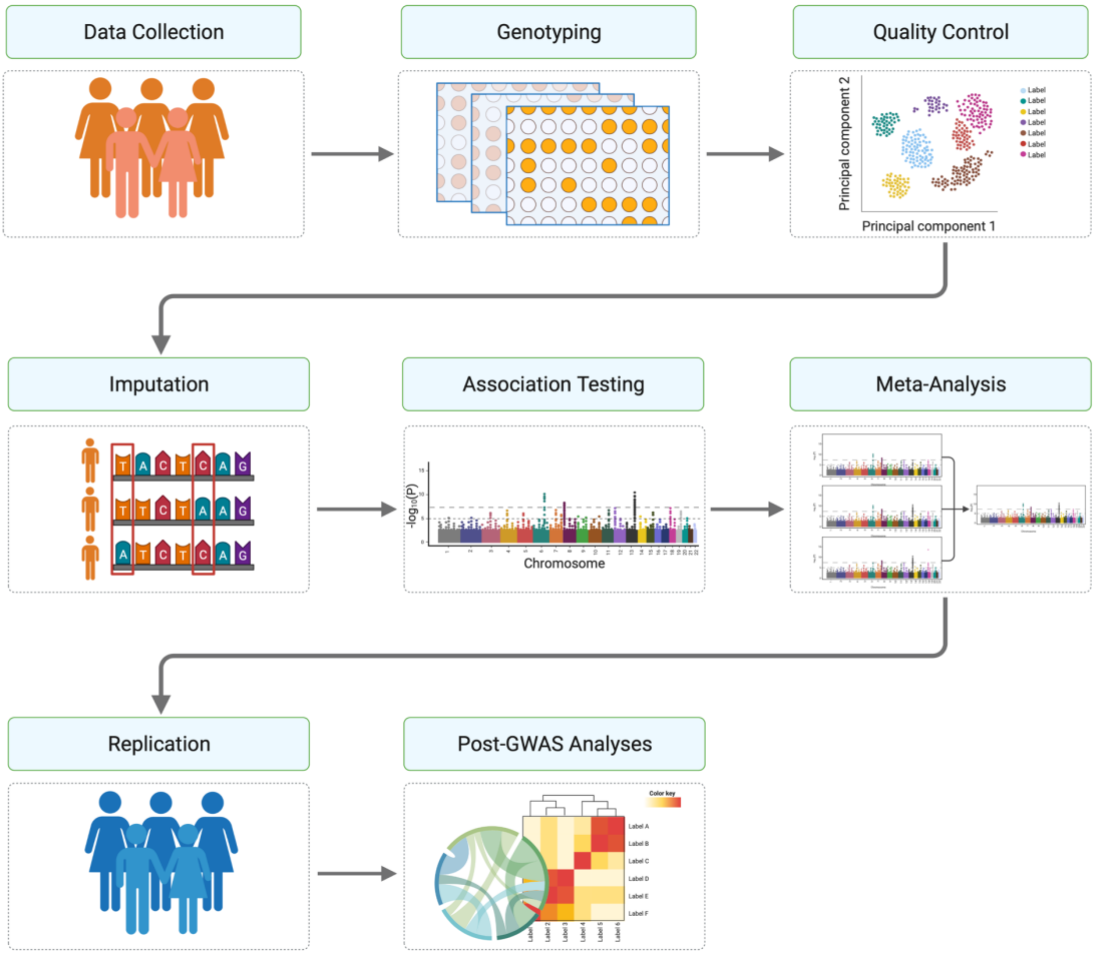


Figure 1. Overview of GWAS Workflow. The key steps of GWAS include data collection, genotyping, quality control, imputation, and association testing. The optional steps are meta-analysis, replication and post-GWAS analyses (*e.g.*, prioritisation of candidate genes, genetic correlation analysis). The illustration was created with BioRender.com and adapted from Uffelmann *et al.*, 2021.

2 THE AIMS OF THE THESIS

The central objective of this study is to investigate the underlying biology of miscarriage and placenta-associated obstetrical diagnoses by performing large-scale genetic analyses. To achieve this, we aim to:

- conduct genome-wide association study meta-analysis using Estonian Biobank and FinnGen GWAS summary statistics,
- annotate and characterise genome-wide significant signals in more detail,
- investigate the shared genetic background of the respective diagnoses between each other, with other female reproductive diagnoses and with a broad range of traits using publicly available summary statistics data.

3 EXPERIMENTAL PART

The experimental part consists of seven different approaches. An overview of the experimental framework is presented in Figure 2.

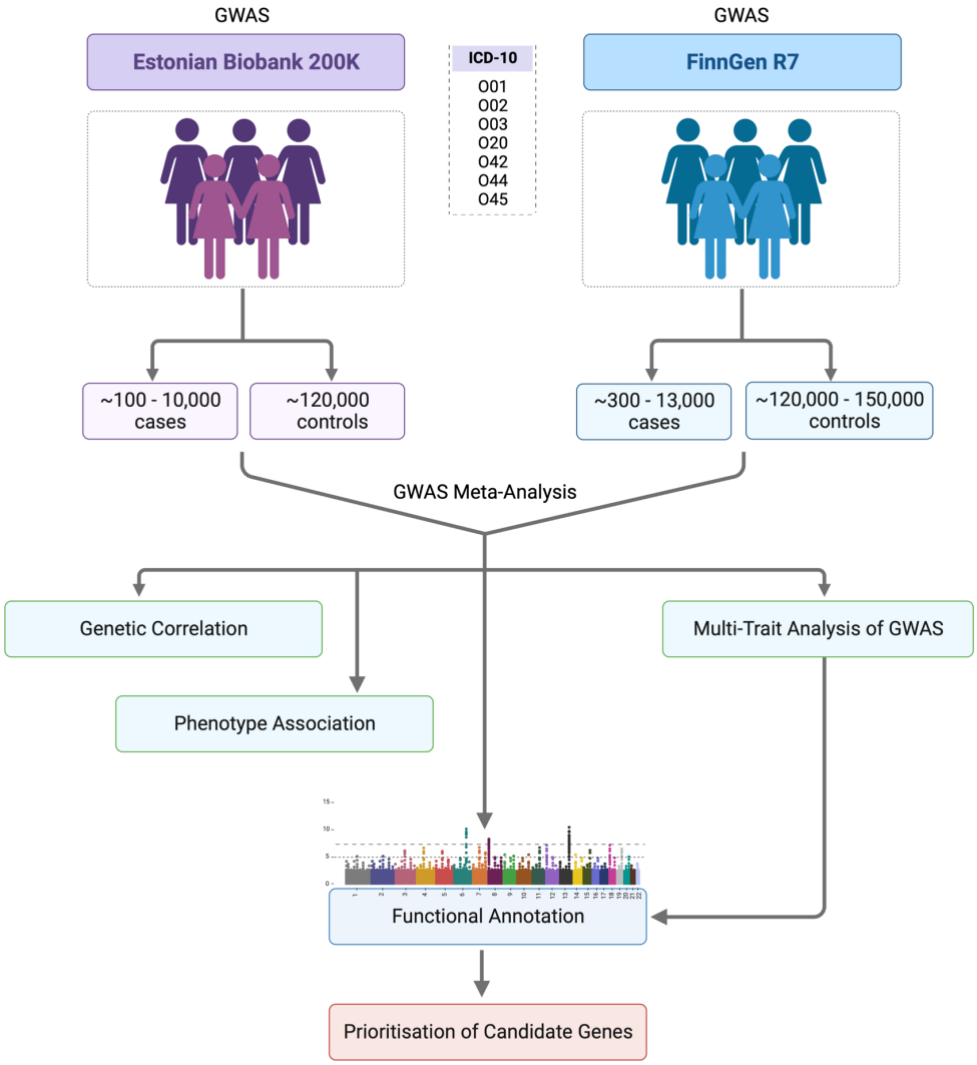


Figure 2. Overview of Experimental Workflow. Cohort-level GWAS was performed using Estonian Biobank data 200K freeze. GWAS meta-analysis was performed using Estonian Biobank and FinnGen GWAS summary statistics download from the FinnGen website release R7. The sample size in the figure reflects the approximate range for the number of cases and controls of all seven diagnoses (hydatidiform mole (O01), other abnormal products of conception (O02), spontaneous abortion (O03), haemorrhage in early pregnancy (O20), premature rupture of membranes (O42), placenta previa (O44) and premature separation of placenta (abruptio placentae) (O45)). Post-GWAS analyses included Multi-Trait Analysis of GWAS using GWAS meta-analysis summary statistics, functional annotation, and prioritisation of candidate genes. Additionally, we analysed genetic correlation and phenotype association using GWAS meta-analysis summary statistics. Figure was created with BioRender.com.

3.1 MATERIALS AND METHODS

3.1.1 Materials

Selection of Diagnoses. To select the relevant diagnoses, the International Statistical Classification of Diseases and Related Health Problems (ICD-10 Version:2019) was used [8]. Maintained by the World Health Organization, the ICD-10 system converts medical diagnoses into alphanumeric codes and provides systematic records on all epidemiological diagnostics and international statistics on mortality and morbidity. The same system is also used to store information on health outcomes at the Estonian Biobank (EstBB) [81]. This thesis focuses on several female-specific diagnoses from Chapter XV, titled Pregnancy, childbirth and the puerperium, encoded as "O" codes. These are "hydatidiform mole" (O01), "other abnormal products of conception" (O02), "spontaneous abortion" (O03), "haemorrhage in early pregnancy" (O20), "premature rupture of membranes" (O42), "placenta previa" (O44) and "premature separation of placenta (abruptio placentae)" (O45). The detailed information and classification of the selected diagnoses codes in the ICD-10 system can be seen in Supplementary Table S1.

Estonian Biobank Genotyping Details and Cohort-Level GWAS. More than 200,000 individuals (20% of the country's adult population, including over 135,000 females) are enrolled in the population-based EstBB, where each participant has signed a broad informed consent document. Ethical approval 1.1-12/624 for the study was obtained from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs). The investigations detailed in this study used the 200K data freeze, and GWAS summary statistics were used in downstream analyses. Cases were identified using the following ICD-10 codes: O01, O02, O03, O20, O42, O44 and O45. The control group for each analysed disease consisted of females who did not have respective diagnosis code. The median and range of age at agreement of cases and controls are presented in Supplementary Table S2. The initial analysis was performed with unselected controls, *i.e.*, every female in EstBB (nulliparous, primiparous, and multiparous females) without respective diagnosis. To check the robustness of the results, the analysis was also performed with selected controls (primiparous and multiparous females), which showed that the results were not biased by control selection. Since the results were quite similar, we used the results from the analysis with unselected controls in downstream analyses and present only these results. Additionally, controls without connected health registry information to the EstBB were

excluded from the analyses. The total number of cases, unselected controls, and sample size of the EstBB cohort for each diagnosis are presented in Table 1.

Table 1. Number of cases, controls, and total sample size of EstBB cohort.

ICD-10 Code	Diagnosis Name	Cases	Controls	Sample Size
O01	Hydatidiform mole	152	126,655	126,807
O02	Other abnormal products of conception	6819	119,988	126,807
O03	Spontaneous abortion	4817	121,990	126,807
O20	Haemorrhage in early pregnancy	10,276	116,531	126,807
O42	Premature rupture of membranes	8276	118,531	126,807
O44	Placenta praevia	668	126,139	126,807
O45	Premature separation of placenta (abruptio placentae)	600	126,207	126,807

As previously described by Pujol Gualdo *et al.*, 2022 [73], genotyping of the EstBB individuals was performed at the Core Genotyping Lab of the Institute of Genomics, University of Tartu, with Illumina GSAv1.0, GSAv2.0, and GSAv2.0_EST arrays. Illumina GenomeStudio v2.0.4 was used to create PLINK format files. Quality control and imputation steps were not done as a part of this study but were previously done by the Bioinformatics Core Lab of the Institute of Genomics, University of Tartu. Briefly, individuals with call-rate <95%, and mismatched sex in phenotype data and sex defined by their X chromosome heterozygosity were excluded. Only variants with call-rate >95%, Hardy-Weinberg Equilibrium (HWE) p-value > 1×10^{-4}, and minor allele frequency (MAF) > 1% were imputed, except for not using HWE filter for X chromosome. Build GRCh37 was used for position mapping of the variants, which were converted to be from the TOP strand using GSAMD-24v1-0_20011747_A1-b37.strand.RefAlt.zip files [82]. Eagle v2.3 software [83] was used for pre-phasing (number of conditioning haplotypes set to --Kpbwt=20000). Beagle v28Sep18.793 software [84] was used for imputation (Ne = 20,000), and 2297 whole genome sequencing (WGS) samples were used for the population-specific imputation reference [85].

To carry out the association testing, REGENIE v2.2.4 software [86] and its mixed logistic regression model was used, since it allows to keep relatives and adjust for covariates (year of birth and the first 10 principal components (PCs)). Among several genetic models, where each assumes different genetic effects, both additive and recessive genetic models were

evaluated for every diagnosis. Finally, variants with low imputation quality (INFO score < 0.4) and minor allele count <5 were excluded from further analyses.

FinnGen Study Cohort and Genotyping Details. The FinnGen project includes electronic health records of several registries and the Finnish biobank data on 309,154 individuals and more than 3000 disease endpoints [87]. As described by Kurki *et al.*, 2022, genotyping of the FinnGen individuals was performed with Illumina and Affymetrix arrays. Population-specific Sequencing Initiative Suomi (SISu) v3 reference panel with Beagle 4.1 v08Jun17.d8b software [88] was used for imputation. SAIGE v0.35.8.8 software [89] was used to test for association analysis with age, sex, first 10 PCs, and genotyping batch as covariates. Only variants with imputation INFO score > 0.6 or MAF > 0.01% were included.

Since FinnGen GWAS summary statistics of Chapter XV "Pregnancy, childbirth and the puerperium" diagnoses are not browsable using the ICD-10 codes, we searched for the diagnosis name (*e.g.*, hydatidiform mole). For the GWAS meta-analysis, we downloaded publicly available GWAS summary statistics of the FinnGen genetic association study from the FinnGen website release R7 [90]. In FinnGen cohort, cases were identified using the corresponding ICD-10 codes (O01, O02, O03, O20, O42, O44 and O45) and assigned FinnGen codes (Table 2). The control group for each analysed diagnosis consisted of females who did not have respective ICD-10 superclass diagnosis codes (*i.e.*, for "hydatidiform mole", "other abnormal products of conception" and "spontaneous abortion", controls were females who did not have codes from the ICD-10 superclass "Pregnancy with abortive outcome" (O00–O08)). The total number of cases, controls, and sample size for each diagnosis of the Finnish cohort are presented in Table 2.

Table 2. Number of cases, controls, and total sample size of FinnGen cohort.

FinnGen Code	Diagnosis Name	Cases	Controls	Sample Size
O15_PREG_HYDAT	Hydatidiform mole	306	124,547	124,853
O15_CONCEPT_ABNORM	Other abnormal products of conception	8933	124,547	133,480
O15_ABORT_SPONTAN	Spontaneous abortion	13,354	124,547	137,901
O15_HAEMORRH_EARLY_PREG	Haemorrhage in early pregnancy	4811	152,785	157,596
O15_MEMBR_PREMAT_RUPT	Premature rupture of membranes	5066	142,734	147,800
O15_PLAC_PRAEVIA	Placenta praevia	917	142,734	143,651
O15_PLAC_PREMAT_SEPAR	Premature separation of placenta	465	142,734	143,199

3.1.2 Estonian Biobank and FinnGen GWAS Meta-Analysis

For the GWAS meta-analysis, GWAS summary statistics data from two study cohorts (EstBB and FinnGen) was used, including up to ~285,000 females of European ancestry (the exact number of cases and controls for each diagnosis are presented in Table 3). GWAMA v2.1 open-source software [91] was used for meta-analysis using inverse variance weighted fixed-effects method with genomic control.

Table 3. Number of cases, controls, and total sample size of EstBB and FinnGen data combined.

ICD-10 Code	Diagnosis Name	Cases	Controls	Sample Size
O01	Hydatidiform mole	458	251,202	251,660
O02	Other abnormal products of conception	15,752	244,535	260,287
O03	Spontaneous abortion	18,171	246,537	264,708
O20	Haemorrhage in early pregnancy	15,087	269,316	284,403
O42	Premature rupture of membranes	13,342	261,265	274,607
O44	Placenta praevia	1585	268,873	270,458
O45	Premature separation of placenta (abruptio placentae)	1065	268,941	270,006

3.1.3 Functional Annotation of GWAS Signals

GWAS summary statistics from EstBB cohort-level data, as well as EstBB and FinnGen meta-analysis, were annotated using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA v1.3.8 and FUMA v1.4.1, respectively) platform [77]. FUMA is one of the most integrative platforms for annotating SNPs and genes, where candidate gene expression patterns and shared molecular functions can be effectively visualised and interpreted. Several implemented tools used for functional annotation are ANNOVAR [78], CADD [92], RegulomeDB [93], and 15-core chromatin state [94]. According to FUMA, any variant causally associated with a trait should have a functional effect on a gene (by altering protein structure or transcription level).

The significant threshold for lead SNPs and candidate SNPs identification was set to $p < 5 \times 10^{-8}$. The threshold for the measure of linkage disequilibrium between two loci for independent significant SNPs was set to $r^2 >= 0.6$. The reference population panel used in the analysis was 1000G Phase3 EUR. The maximum distance of LD blocks for merging into locus was set to <=250kb. For the rest of the parameters, the default settings in FUMA were used (*e.g.*, minimum CADD score >= 12.37, maximum RegulomeDB score = 7, 15-core

chromatin state maximum state <= 7). CADD score is a measure of the SNPs deleteriousness, RegulomeDB is a measure of SNPs regulatory functionality (1 being the strongest evidence), and first seven 15-core chromatin state include only open chromatin states with active and regulatory regions. Additionally, genetic variants not present in both cohorts (n_studies < 2), variants with MAF < 1%, and the highly polymorphic and gene-dense MHC region were excluded. In the results section, genomic risk loci are reported according to the lead SNP as defined by FUMA. Every identified lead SNP was further assessed if it passed a more stringent genome-wide significance threshold that accounts for the number of performed analyses (*i.e.*, $p < 5 \times 10^{-8}/14 = 3.6 \times 10^{-9}$; 7 different traits and two genetic models (additive and recessive)) and if its heterogeneity index (I^2; a measure of heterogeneity between cohorts in meta-analyses)) is of the reasonable magnitude.

3.1.4 Prioritisation of Candidate Genes

To nominate potential candidate genes at each locus and interpret broader biological context, several approaches were used, such as the nearest gene to the association signal, gene-based testing, data on expression quantitative trait loci (eQTLs), 3D chromatin interactions (Hi-C), data from mouse models databases, gene set, and tissue/cell type enrichment analyses (Figure 3). In this section, each approach will be briefly described.

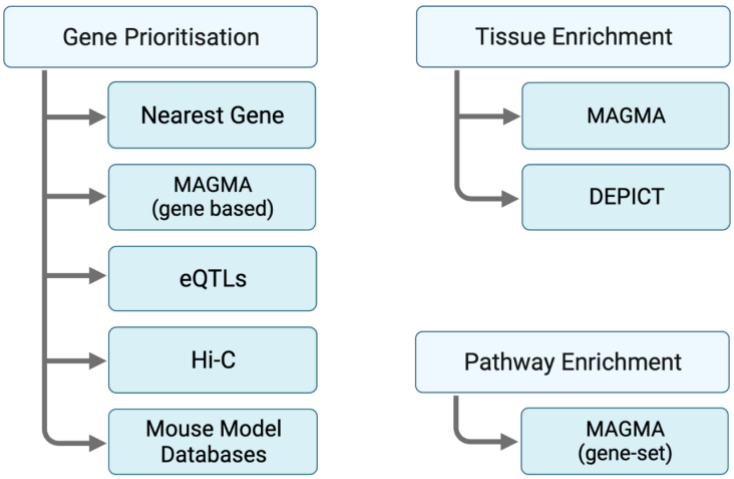


Figure 3. Strategy Used for Candidate Gene Mapping. To prioritise the most likely candidate genes at each associated locus, we have investigated the nearest gene to the association signal, output from MAGMA gene-based analysis, data on expression quantitative trait loci (eQTLs), 3D chromatin interactions (Hi-C), and from mouse models databases. To identify tissue-specific gene expression profiles, we have performed MAGMA and DEPICT tissue enrichment analysis. To search for enriched group of genes and associated biological pathway, we have performed MAGMA gene-set analysis. Figure was created with BioRender.com.

Firstly, the nearest gene to the lead association signal was investigated, as it is straightforward, yet comparably effective approach as other methods [95]. Next, using the MAGMA v1.08 tool [79] developed by De Leeuw *et al.*, 2015, implemented in FUMA, joint association effect of all SNPs within a gene was tested, taking into consideration the LD structure. Specifically, in MAGMA gene analysis based on multiple linear PCs regression model and accounting for LD (reference panel 1000G Phase3 [96]), SNPs were mapped to 18,895 protein-coding genes, while the genome-wide significant threshold was set at $p = 0.05/18895 = 2.646 \times 10^{-6}$. In the results section we will report p-value and q-value (*i.e.*, p-value adjusted for the number of performed tests).

Additionally, some SNPs in non-coding regions might have regulatory effects on genes [97–99]. For example, they can be eQTLs (genetic variants previously associated with differences in RNA levels), or they are known to physically interact with promoter regions via chromatin interactions. In both analyses, we used predefined parameters and data sources built in FUMA (*e.g.*, GTEx v6/v7/v8, BRAINEAC, eQTLGen, eQTL Catalogue, Hi-C of 21 tissue/cell types from GSE87112, Hi-C loops from Giusti-Rodriguez *et al.*, 2019 [100–104]). Although most of the tissues used in these datasets are not functionally relevant for the obstetrical and placenta-associated diagnoses, they have the potential to provide insights into the regulatory effects of genetic variants and suggest targets for future research, and, thus, are worth investigating [105]. Additionally, reported *cis*- and *trans*-eQTLs from several studies on eQTLs in human placenta [106,107] were investigated. Finally, data from experiments on mice (gene knockouts or overexpression) contributes to studying the function of genes, further disentangling the genetic determinants of complex traits. Here, we looked for mouse phenotypes that might be relevant for placental biology (phenotypes related to the reproductive tract or affecting embryonic development/viability) [108,109].

To identify tissue-specific gene expression profiles and, thus, potentially biologically relevant tissue or cell types, two complementary approaches were used, since currently no "gold standard" exists for the analytical approach. First, the GTEx (v8) 30 general tissue types dataset and the full distribution of SNP p-values were used in MAGMA gene-property analysis implemented in FUMA. Similarly, the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool [80] developed by Pers *et al.*, 2015, implemented in Complex-Traits Genetics Virtual Lab (CTG-VL 0.5-beta) [110] was used to identify tissues that show enrichment for GWAS results. Compared with MAGMA, DEPICT has a broader catalogue to infer how different genes are related to different biological

processes using the reconstituted gene sets from databases such as GO [111], KEGG [112], REACTOME [113] and Mouse Genetics Initiative [114]. In DEPICT analysis, the p-value threshold was set to 1×10^{-5}, meaning all variants with a p-value below this threshold were used in the analysis. Results from MAGMA and DEPICT are corrected for multiple testing using the Bonferroni and FDR correction (q-values), respectively.

To search for enriched group of genes and associated biological pathways, in the MAGMA gene-set analysis, SNPs were mapped to 10,678 biologically and functionally related gene sets (4761 curated gene sets 4761 and 5917 GO terms) from Molecular Signatures Database (MSigDB v7.0) [115]. Also, Bonferroni correction (q-value) for the number of performed tests was applied to reduce the risk of false positives.

3.1.5 Genetic Correlation

To better understand the shared genetic background between the obstetrical diagnoses of our interest, with other female reproductive diagnoses we are studying in our research group, and with a broad range of traits from the public genetic correlation database, their pairwise genetic correlations was examined using the LDSC v1.0.1 Python command line tool [116,117]. We chose this method as it is a widely used and efficient tool for estimating genetic correlations (r_g) and SNP-heritability (h^2_{SNP}), allowing for sample overlap. Moreover, it performs well even with underpowered GWAS summary statistics data where no significant genome-wide signals are detected. The LDSC algorithm uses a combination of test statistics and LD scores to identify true polygenic signals and correct for biases. Test statistics are measures of the statistical significance of the relationship between a genetic variant and a trait, and LD scores are measures of the correlation strength between an SNP and other SNPs in the genome. To be precise, LDSC uses only those SNPs present in the HapMap 3 reference panel (*i.e.*, common, well-imputed SNPs; *ca.* around 1.3 million SNPs).

The initial genetic correlation analysis included 47 different phenotypes from Chapters II, IV, XIV and XV of the ICD-10 system, encompassing diagnoses related to female reproductive disorders, such as “benign and malignant neoplasms”, “polycystic ovarian syndrome”, “genitourinary diseases” and “pregnancy complications”. The GWAS summary statistics datasets were available from our work group, and the sample sizes ranged from ~221,000 to ~300,000. To correct for the number of performed tests (n = 1081), the Benjamini-Hochberg method was applied. To highlight similarities and differences in genetic correlations between diagnoses, a correlation plot was created using R v4.2.1 packages *dplyr* and a modified version of *corrplot* by Kanai *et al*., 2018 [118]. Diagnoses

were also clustered by the hierarchical clustering method to identify grouping patterns based on their genetic correlation estimate (r_g).

Secondly, in the LDSC batch genetic correlation analysis implemented in Complex-Traits Genetics Virtual Lab (CTG-VL 0.5-beta) [110], each diagnosis from our list was compared with 1403 traits available in a public genetic correlation database, including phenotypes related to anthropometric traits, cardiometabolic phenotypes, education, hormonal measurements, mental health, parents' illnesses, pain, reproductive health, and smoking behaviour. Bonferroni correction ($p < 0.05/1403 = 3.5 \times 10^{-5}$) was applied to account for the number of performed tests. To visually compare the results, forest plot was created using R v4.2.1 packages *tidyverse*, *forcats* and *ggplot2*.

3.1.6 SNP-Heritability

In statistics, the proportion of variation in a trait attributed to all genetic factors is called total heritability (H^2). Part of this comes from so-called SNP-heritability (h^2_{SNP}), or the proportion of variation in the trait explained by additive effects of commonly occurring genetic variants (SNPs) from GWAS.

To calculate SNP-heritability, LDSC v1.0.1 tool [116,117] was used. The estimates on the observed scale from the LDSC were converted to the liability scale according to the work of Lee *et al.*, 2011 [119], and assuming that the population prevalence is equal to the sample prevalence. This conversion allows better interpretation and accounts for ascertainment bias when studying binary traits (*i.e.*, diseases with one of two possible states: affected/unaffected). Visual representation of results was created using R v4.2.1 packages *tidyverse*, *forcats* and *ggplot2*.

3.1.7 Multi-Trait Analysis of GWAS

In most recent studies, the Multi-Trait Analysis of GWAS (MTAG v1.0.8) [120] has been used to perform a joint GWAS analysis of the traits with high genetic correlation and SNP-heritability. This fast-computing command line tool can generate trait-specific association statistics using the generalisation of inverse-variance-weighted meta-analysis. The main idea is to boost the number of genome-wide significant loci in each trait and possibly unravel novel loci by leveraging power from its related traits.

Diagnoses selection was made according to the formed clusters in the genetic correlation plot. Therefore, diagnoses "ectopic pregnancy" (O00, n_{cases} = 7070, $n_{controls}$ = 248,810),

and “preterm labor and delivery” (O60, n_{cases} = 10,968, $n_{controls}$ = 259,753) were included due to their high genetic correlation with our diagnoses of interest. GWAS meta-analysis summary statistics files of diagnoses “ectopic pregnancy” (O00), “other abnormal products of conception” (O02), “spontaneous abortion” (O03) and “haemorrhage in early pregnancy” (O20) were formatted according to the requirements of MTAG. Only variants present in both cohorts (n_studies == 2) were included. Special option flag “--force” was added in the bash script to override the tool’s default configuration to stop the computation if the mean χ^2 statistic is lower than 1.02. The output files of individual traits were uploaded to FUMA as described in Section 3.1.3. The same steps were applied to conduct the MTAG pairwise analysis for diagnoses “premature rupture of membranes” (O42) and “preterm labor and delivery” (O60).

3.1.8 Phenotype Association

In addition to the genetic correlation, the extent of overlap between obstetrical diagnoses and other phenotypes was examined through the phenome-wide phenotype association analysis (phepheWAS) between each diagnosis from our list and the other ICD-10 diagnoses present in the EstBB 200K data freeze (~2000 diagnosis codes). The analysis was based on the framework described by Pujol Gualdo *et al.*, 2022 [121]: one diagnosis case-control status was tested against all the other ICD-10 diagnosis case-control statuses in a logistic regression framework. Diagnoses in ICD-10 system are grouped into following Chapters: I - “Certain infectious and parasitic diseases” (A00–B99), II - “Neoplasms” (C00–D48), III - “Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism” (D50–D89), IV - “Endocrine, nutritional and metabolic diseases” (E00–E90), V - “Mental and behavioral disorders” (F00–F99), VI - “Diseases of the nervous system” (G00–G99), VII - “Diseases of the eye and adnexa” (H00–H59), VIII - “Diseases of the ear and mastoid process” (H60–H95), IX - “Diseases of the circulatory system” (I00–I99), X - “Diseases of the respiratory system” (J00–J99), XI - “Diseases of the digestive system” (K00–K93), XII - “Diseases of the skin and subcutaneous tissue” (L00–L99), XIII - “Diseases of the musculoskeletal system and connective tissue” (M00–M99), XIV - “Diseases of the genitourinary system” (N00–N99), XV - “Pregnancy, childbirth and the puerperium” (O00–O99), XVI - “Certain conditions originating in the perinatal period” (P00–P96), XVII - “Congenital malformations, deformations and chromosomal abnormalities” (Q00–Q99), XVIII - “Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified” (R00–R99), XIX - “Injury, poisoning and certain other

consequences of external causes" (S00–T98), XX - "External causes of morbidity and mortality" (V01–Y98), XXI - "Factors influencing health status and contact with health services" (Z00–Z99) and XXII - "Codes for special purposes" (U00–U89).

The findings from this analysis provide insight into the likelihood of having co-occurring diagnoses, *i.e.*, whether cases of a particular diagnosis compared to the controls have some other common diagnoses. The model was adjusted for year of birth and ten genetic PCs, and Bonferroni correction was applied to account for the number of performed tests ($p < 0.05/2001 = 2.5 \times 10^{-5}$). All first- and second-degree relatives were excluded, and the analysis was conducted using unselected controls. The results were visualised using R v4.1.3 packages *stringr*, *dplyr*, *slam*, *qdap* and *PheWAS*.

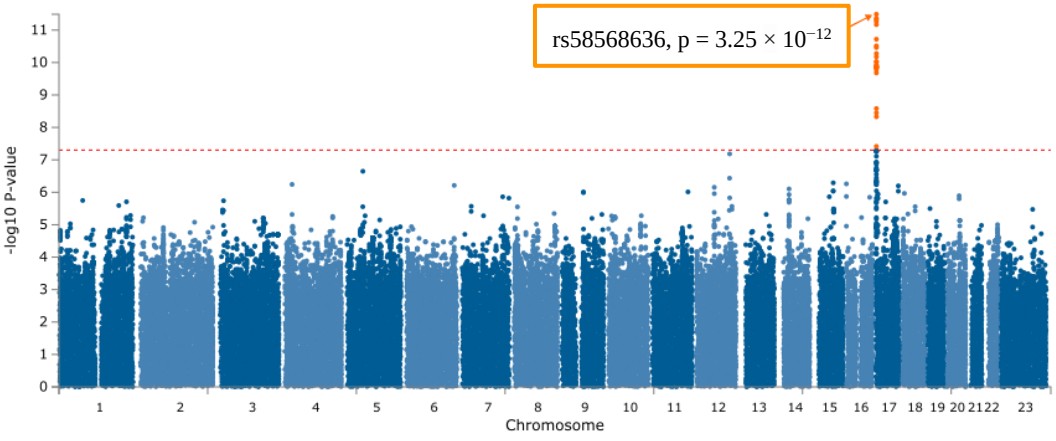
3.2 RESULTS

3.2.1 Characterisation of Genomic Loci

Estonian Biobank and FinnGen Cohort-Level GWAS Results. EstBB GWAS included a total number of 126,807 females of European ancestry (Table 1). Assuming an additive genetic model, significant locus was detected only in diagnosis "haemorrhage in early pregnancy", as shown in Figure 4a. The lead SNP on chromosome 17 (rs58568636, $p = 3.25 \times 10^{-12}$, MAF = 34%) is an intronic variant of the *SMG6* gene. Downstream of the same gene, we detected an independent significant SNP (rs2236374, $p = 3.85 \times 10^{-8}$, MAF = 40%). No evidence was found for significant genetic associations in other diagnoses.

When using a recessive genetic model, similar results were observed (Figure 4b), as we detected a significant association only for the "haemorrhage in early pregnancy" diagnosis. The significant locus on chromosome 17 is near the same *SMG6* gene, but with different lead SNP (rs216224, $p = 2.71 \times 10^{-9}$, MAF = 36%) and no additional independent significant SNPs.

a)



b)

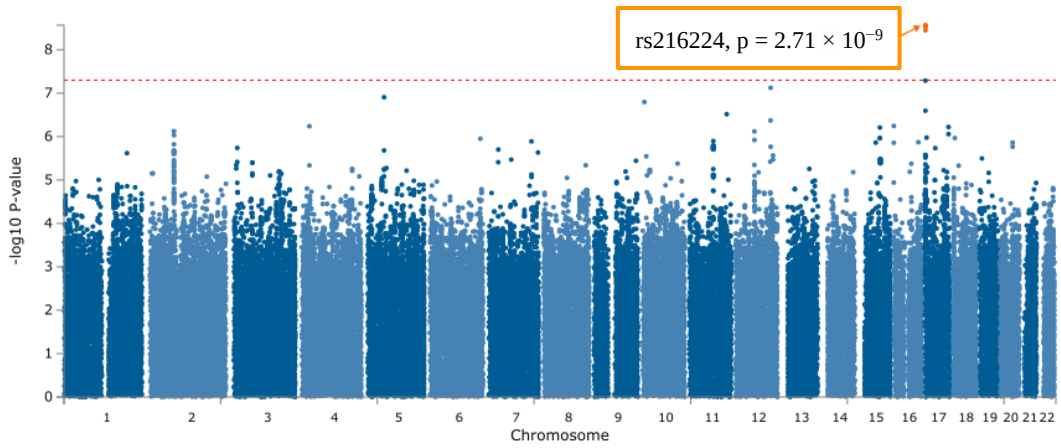


Figure 4. Manhattan Plots Made in FUMA of EstBB Cohort-Level GWAS for Diagnosis Haemorrhage in Early Pregnancy Assuming a) Additive Genetic Model and b) Recessive Genetic Model. On the x-axis are represented chromosomes and each dot is a genetic variant (SNP), while on the y-axis is the −log10 p-value from the analysis. The red dashed line marks the genome-wide significance set at 5×10^{-8}. The orange dots are statistically significant SNPs, while the annotated dot is the lead SNP. Corresponding QQ plots of GWAS summary statistics are presented in Supplementary Figure S1.

For comparison, in the FinnGen GWAS of the total sample size ranging from ~120,000–150,000 females, no significant SNPs were detected in any of the seven diagnoses.

Estonian Biobank and FinnGen GWAS Meta-Analysis Results. To conduct the GWAS meta-analysis, we combined the EstBB additive model and the FinnGen GWAS summary statistics increasing the total sample size up to ~285,000 females in the analyses (Table 3). As expected due to the increase in sample size, the meta-analyses identified novel loci in two out of seven diagnoses (Table 4).

Table 4. Identified Lead SNPs in GWAS Meta-Analysis

ICD-10 Code	Diagnosis Name	Cytogenetic Region	rs ID (EA/NEA)	p-value	EAF	OR (EA) (95% CI)
O20	Haemorrhage in early pregnancy	17p13.3	rs11657636 (C/T)	1.65×10^{-12}	0.326	0.91 (0.89–0.93)
O42	Premature rupture of membranes	12q13.13	rs4237901 (C/T)	5.2×10^{-10}	0.731	0.91 (0.86–0.94)

EA–effect allele; NEA–non-effect allele

Diagnoses “hydatidiform mole”, “other abnormal products of conception”, and “spontaneous abortion” did not have significant signals. In “haemorrhage in early pregnancy”, we detected a genome-wide significant locus on chromosome 17 (rs11657636, $p = 1.65 \times 10^{-12}$, MAF = 33%). For the placental phenotypes, we found significant signals on chromosome 12 (rs4237901, $p = 5.2 \times 10^{-10}$, MAF = 27%) for “premature rupture of membranes”, while the diagnoses “placenta previa” and “premature separation of placenta” did not have statistically significant associations. The first mentioned lead variant showed higher heterogeneity index (rs11657636, $I^2 = 61\%$) due to the higher number of cases present in EstBB and, thus, greater statistical power of the cohort data. The other lead variant showed no heterogeneity (rs4237901, $I^2 = 0\%$). Additionally, Q-Cochran p-values of 0.1 and 0.5 suggested no heterogeneity between Estonian and Finnish cohorts. The Manhattan plots from the GWAS meta-analyses are presented in Figure 5, while the corresponding QQ plots are in Supplementary Figure S1.

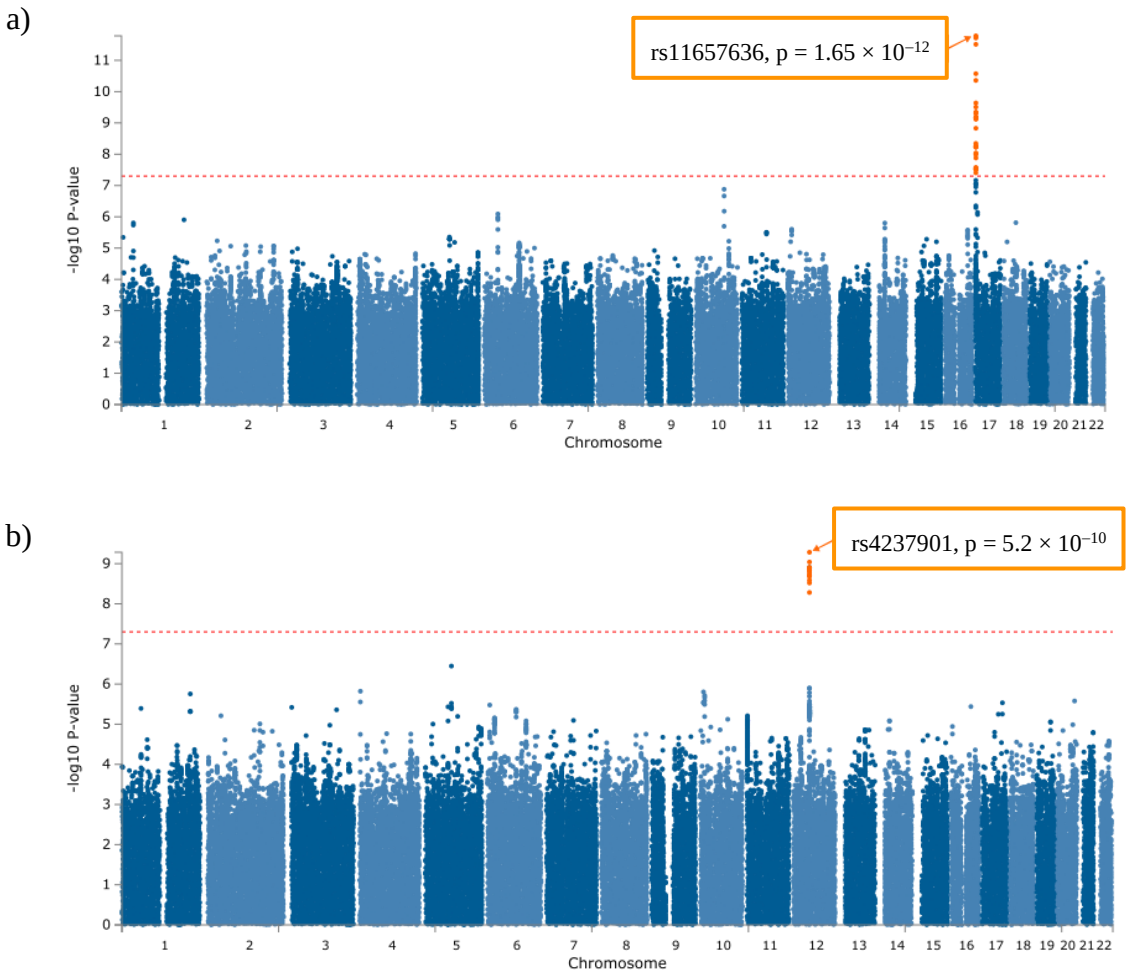


Figure 5. Manhattan Plots Made in FUMA of EstBB and FinnGen GWAS Meta-Analysis for Diagnoses a) Haemorrhage in Early Pregnancy, and b) Premature Rupture of Membranes. On

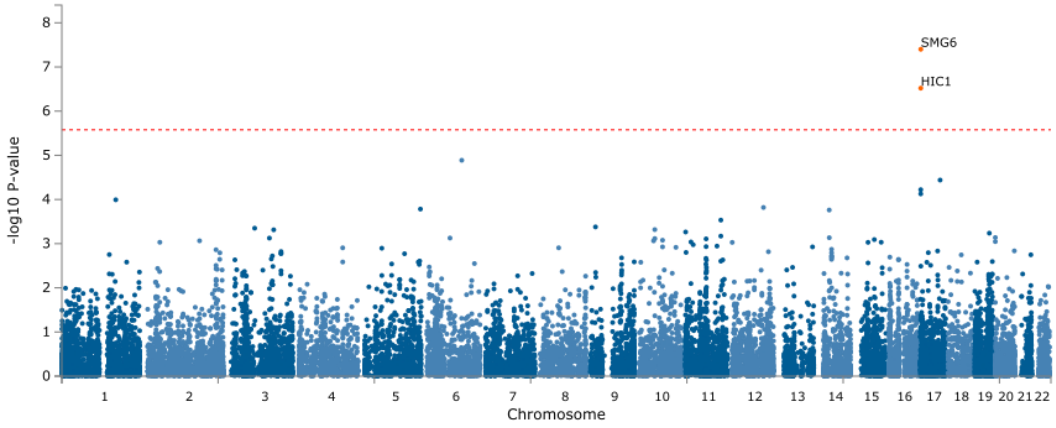
the x-axis are represented chromosomes and each dot is a genetic variant (SNP), while on the y-axis is the –log10 p-value from the meta-analysis. The red dashed line marks the genome-wide significance set at 5×10^{-8}. The orange dots are statistically significant SNPs, while the annotated dot is the lead SNP. Corresponding QQ plots of GWAS summary statistics are presented in Supplementary Figure S1.

3.2.2 Functional Annotation and Candidate Genes Prioritisation

Studying the lead SNPs associated with a trait of interest is the first, but insufficient step for understanding underlying biological mechanisms. To reliably explain observed associations and propose the most likely biologically relevant variants, we assessed outputs from databases integrated in FUMA, such as CADD scores, RegulomeDB scores, and 15 chromatin state marks for individual variants. Additionally, using an approach described in Section 3.1.4, we combined several layers of data, such as the nearest gene to the association peak, MAGMA gene-based testing, data on eQTLs, chromatin interactions (Hi-C) and from mouse models databases, to propose the most likely candidate genes at each associated locus. These findings are discussed in more detail below and summarised in Table 5.

For the diagnosis "haemorrhage in early pregnancy", the top lead SNP is in the intronic region of the protein coding gene *SMG6*, and also tags an exonic SNP (rs216195, $p = 1.32 \times 10^{-8}$, $r^2 = 0.74$) that has a high CADD score (23.1), indicating potentially detrimental functional consequences of this variation. The same *SMG6* gene was also highlighted in the gene-based analysis ($p = 3.97 \times 10^{-8}$, $q = 0.0008$), while another gene in the same region, *HIC1*, also passed the statistical significance threshold ($p = 3.02 \times 10^{-7}$, $q = 0.003$) (Figure 6a). Interestingly, in the intron of the *HIC1* gene, we found one suggestively significant SNP (rs4455005, $p = 3.5 \times 10^{-6}$, $r^2 = 0.71$) with a high CADD score (22.6), which 15-core chromatin state revealed overlap with transcription start sites in fetal lung, ESC derived cells and placenta. Regarding diagnosis "premature rupture of membranes", the lead SNP is in the intronic region of the protein-coding gene *SLC4A8*, which is next to the *GALNT6* gene. Gene-based analysis revealed two statistically significant protein-coding genes *SLC4A8* ($p = 4.38 \times 10^{-8}$, $q = 0.0008$) and *TSPAN32* ($p = 2.17 \times 10^{-6}$, $q = 0.02$) (Figure 6b).

a)



b)

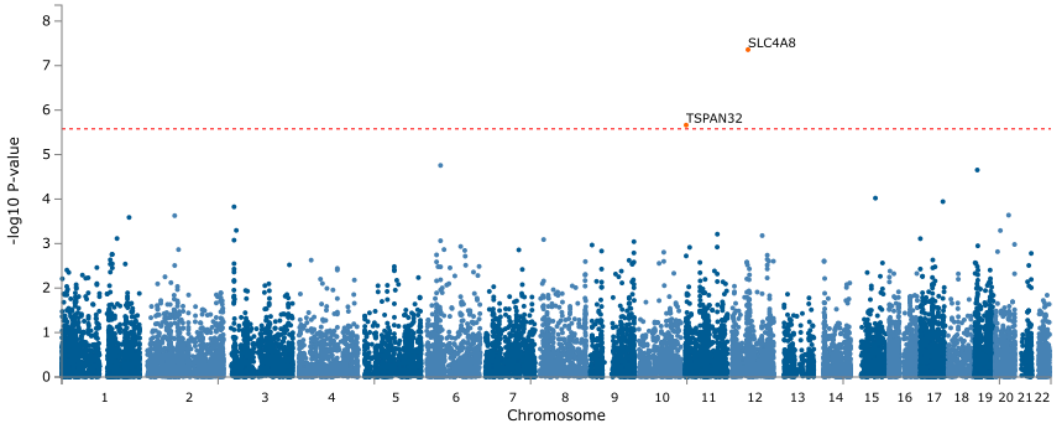


Figure 6. Manhattan Plots Made in FUMA of MAGMA Gene-Based Testing for Diagnoses a) Haemorrhage in Early Pregnancy and b) Premature Rupture of Membranes Using EstBB and FinnGen Data. In the gene-based analysis, SNPs were mapped to 18,895 protein-coding genes. On the x-axis are represented chromosomes and each dot is a gene, while on the y-axis is the –log10 p-value from the meta-analysis. The red dashed line marks the genome-wide significance set at p = 0.05/18895 = 2.646×10^{-6}. The orange dots are statistically significant highlighted genes. Corresponding QQ plots of MAGMA gene-based test are presented in Supplementary Figure S2.

Next, according to eQTL and chromatin interaction data (Supplementary Figure S3) implemented in FUMA, for "haemorrhage in early pregnancy" the following 11 genes on the chromosome 17 were proposed as candidate genes for this locus: *PAFAH1B1*, *SGSM2*, *AC006435.1*, *MNT*, *SMG6*, *SRR*, *RTN4RL1*, *DPH1*, *HIC1*, *OVCA2*, *TSR1*. The genes linked by only eQTLs were *ERBB2* and *RPA1*. Chromatin interactions individually identified interaction with another 11 different genes: *INPP5K*, *SERPINF1*, *PRPF8*, *WDR81*, *TLCD2*, *SERPINF2*, *PFN1*, *ENO3*, *SLC25A11*, *RNF167*, *SPAG7*. Concerning "premature rupture of membranes", both analyses associated lead SNP with the two additional genes on

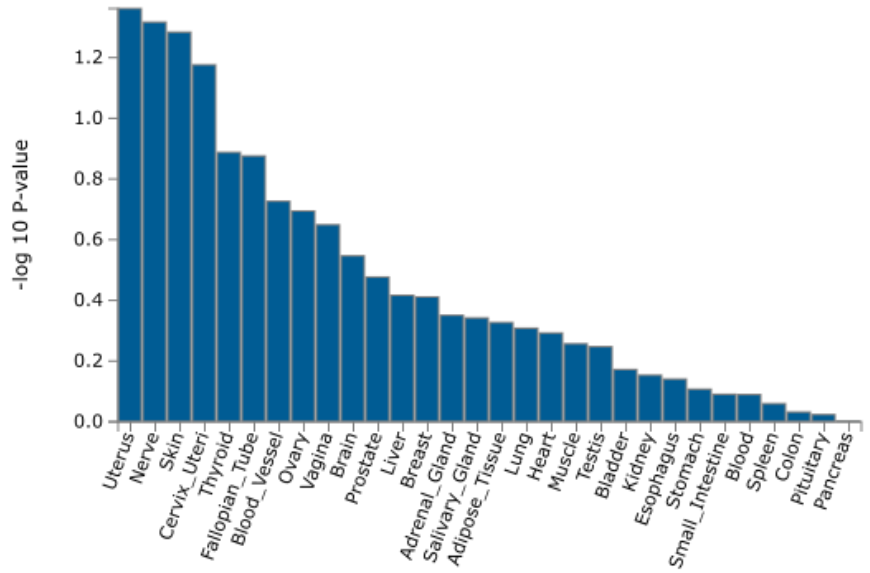
chromosome 12: *CSRNP2* and *AC068987.1*. Chromatin interactions alone identified interaction with additional 10 different genes: *ANKRD33*, *ACVR1B*, *CELA1*, *BIN2*, *SMAGP*, *DAZAP2*, *LETMD1*, *SLC11A2*, *TMPRSS12*, *ATF1*, while the genes linked by only eQTLs were: *GALNT6*, *POU6F1*, *KRT75* and *METTL7A*. Previously reported placental eQTLs were not in LD with any of our candidate SNPs ($r^2 >= 0.6$).

To obtain more insights into the potential functional roles of these genes, we used data from a mouse genotype database [108,109]. Experiments on mice showed that transgene insertion into one of the introns of the *SMG6* gene causes the death of the embryo. Moreover, homozygous knock-out mutants of the same gene died before birth, while those homozygous for the conditional allele resulted in defective gene functionality. Regarding *HIC1* gene, homozygous mutants with null mutation died in the later pregnancy stages due to the various abnormalities, while the heterozygotes developed malignant neoplasms in several organ systems. In diagnosis "premature rupture of membranes", knock-out mutants of the *SLC4A8* gene resulted in abnormal sodium and chloride ion excretion. Additionally, homozygous null mice of the genes noted in the Table 5 for diagnoses "haemorrhage in early pregnancy" (*PAFAH1B1*, *INPP5K*, *SERPINF1*, *PFN1*, *SLC25A11* and *RPA1*) and "premature rupture of membranes" (*AC068987.1*, *ACVR1B*, *SLC11A2* and *TMPRSS12*) all led to embryonic lethality, indicating their potential relevance for the embryonic development.

Based on all these diverse data sources, we propose as the most likely candidate genes *SMG6*, *HIC1* and *SRR* for "haemorrhage in early pregnancy" at the associated risk locus on chromosome 17, and *SLC4A8* and *GALNT6* genes for "premature rupture of membranes" at the associated risk locus on chromosome 12.

Following Bonferroni multiple test corrections, no significant enrichment was observed in any of the tissues examined for both diagnoses. However, as shown in Figure 7, the top nominally associated tissue was the uterus ($p = 0.04$, $q = 0.49$) for "haemorrhage in early pregnancy", and the stomach ($p = 0.03$, $q = 0.44$) for "premature rupture of membranes". Using a different tissue enrichment testing approach (the DEPICT software), we also observed a nominally significant enrichment ($p = 0.0014$, $q = 0.3$) in embryoid bodies for "haemorrhage in early pregnancy". Embryoid bodies (EB) are spherical embryonic stem cell aggregates often used as a model system of gene regulation and embryonic development [122], as the embryonic stem cells in EB differentiate and specialise into three germ layers – endoderm, ectoderm, and mesoderm.

a)



b)

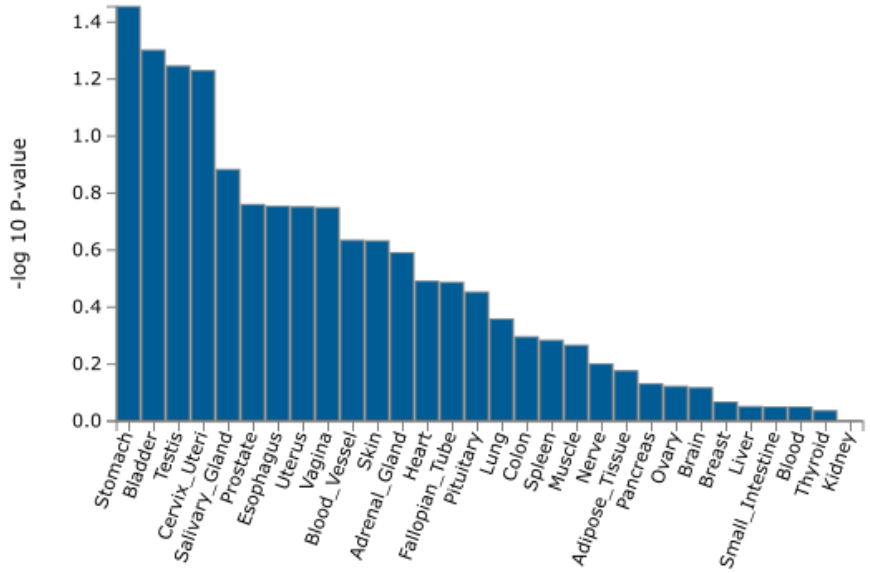


Figure 7. The Bar Plots Made in FUMA of MAGMA Tissue Enrichment Analysis of GTEx v8 (Including 30 General Tissue Types) for Diagnoses a) Haemorrhage in Early Pregnancy and b) Premature Rupture of Membranes. The p-values were transformed to log scale for clearer results.

Finally, gene-set analysis revealed nominally significant enrichment in global genome nucleotide excision repair pathway involved in DNA repair for "haemorrhage in early pregnancy" ($p = 7.2895 \times 10^{-7}$, $q = 0.011$) and the negative regulation of dendrite morphogenesis ($p = 1.9714 \times 10^{-5}$, $q = 0.31$) for "premature rupture of membranes".

Table 5. Summary Table of Candidate Gene and Enriched Tissues Mapping.

The most plausible candidate genes we propose are marked in bold.

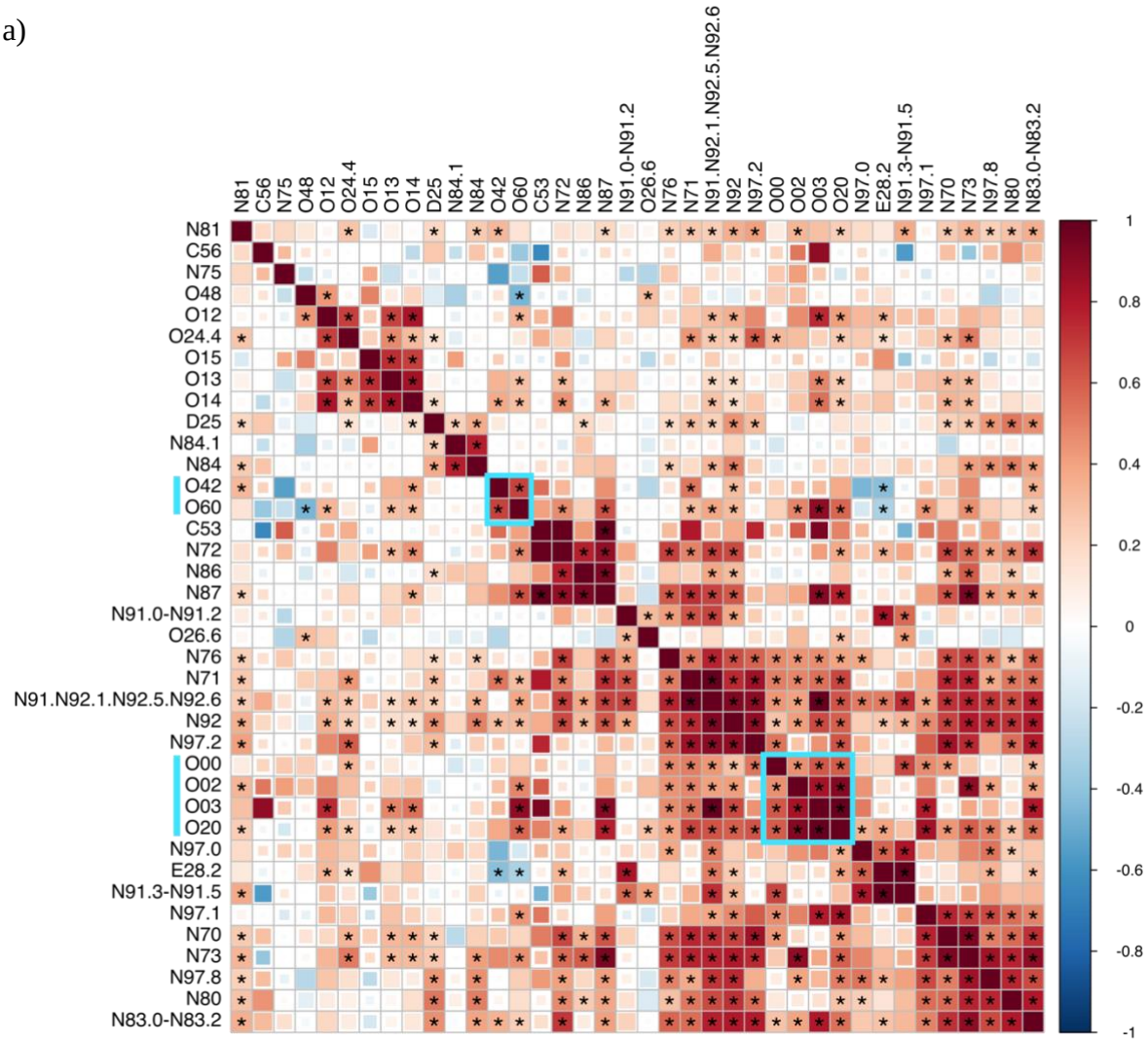
Diagnosis Name	Nearest Gene	MAGMA (gene-based)	eQTLs	Hi-C	Mouse Models	MAGMA (tissue enrichment)	DEPICT (tissue enrichment)	MAGMA (gene-set)
Haemorrhage in early pregnancy	*SMG6*	• ***SMG6*** • ***HIC1***	• *PAFAH1B1* • *SGSM2* • *AC006435.1* • *MNT* • ***SMG6*** • ***SRR*** • *RTN4RL1* • *DPH1* • ***HIC1*** • *OVCA2* • *TSR1* • *ERBB2* • *RPA1*	• *PAFAH1B1* • *SGSM2* • *AC006435.1* • *MNT* • ***SMG6*** • ***SRR*** • *RTN4RL1* • *DPH1* • ***HIC1*** • *OVCA2* • *TSR1* • *INPP5K* • *SERPINF1* • *PRPF8* • *WDR81* • *TLCD2* • *SERPINF2* • *PFN1* • *ENO3* • *SLC25A11* • *RNF167* • *SPAG7*	• *PAFAH1B1* • ***SMG6*** • ***HIC1*** • *INPP5K* • *SERPINF1* • *PFN1* • *SLC25A11* • *RPA1*	Uterus	Embryoid Bodies	Global Genome Nucleotide Excision Repair
Premature rupture of membranes	*SLC4A8*	• ***SLC4A8*** • *TSPAN32*	• *CSRNP2* • *AC068987.1* • ***GALNT6*** • *POU6F1* • *KRT75* • *METTL7A*	• *CSRNP2* • *AC068987.1* • *ANKRD33* • *ACVR1B* • *CELA1* • *BIN2* • *SMAGP* • *DAZAP2* • *LETMD1* • *SLC11A2* • *TMPRSS12* • *ATF1*	• *AC068987.1* • *ACVR1B* • *SLC11A2* • *TMPRSS12*	Stomach	/	Negative Regulation of Dendrite Morphogenesis

3.2.3 Genetic Correlation

The genetic correlation between obstetrical diagnoses related to miscarriage and placenta, with female reproductive diagnoses and other traits from publicly available genetic correlation database was tested using the LDSC v1.0.1 tool [116,117].

Firstly, the pairwise correlation analysis amongst female reproductive diagnoses revealed that certain diagnoses tend to cluster in distinct groups. Specifically, the diagnoses with abortive outcomes formed a separate cluster from the placenta-associated pregnancy complications. In particular, diagnoses "other abnormal products of conception" (O02), "spontaneous abortion" (O03), and "haemorrhage in early pregnancy" (O20) were more strongly correlated with one another (lowest: r_g = 0.82, p = 4.8×10^{-3}, highest: r_g = 0.91, p = 5.75×10^{-5}) and with the diagnosis "ectopic pregnancy" (O00) (r_g = 0.45–0.62, p < 2×10^{-2}), thus, creating a smaller cluster nested within the larger cluster, which included diagnoses "female infertility of uterine origin" (N97.2), "excessive, frequent and irregular menstruation" (N92), umbrella phenotype of menstrual disorders (N91.N92.1.N92.5.N92.6), "inflammatory disease of uterus" (N71) and "other inflammation of vagina and vulva" (N76). As for placenta-associated diagnosis "premature rupture of membranes" (O42), it had the strongest positive correlation with the diagnosis "preterm labour and delivery" (O60) (r_g = 0.68, p = 2×10^{-4}), and negative correlation with the "diseases of Bartholin gland" (N75) (r_g = –0.54, p = 6.7×10^{-2}), "liver disorders in pregnancy, childbirth and puerperium" (O26.6) (r_g = –0.28, p = 9.2×10^{-2}), "female infertility associated with anovulation" (N97.0) (r_g = –0.46, p = 9.5×10^{-2}) and "polycystic ovarian syndrome" (E28.2) (r_g = –0.41, p = 3×10^{-2}). The pairwise correlation analysis results of 38 female reproductive diagnoses and their clustering based on the genetic correlation estimate (r_g) are presented in Figure 8a, while obstetrical diagnoses of our interest and their exact r_g values are presented in Figure 8b. Due to the insufficient sample size (n_{cases} < 5000) and unreliable estimates, diagnosis "hydatidiform mole" (O01), "placenta previa" (O44) and "premature separation of placenta (abruptio placentae)" (O45) had to be excluded.

a)



ICD-10 Code Names

N81 - Female genital prolapse
C56 - Malignant neoplasm of ovary
N75 - Diseases of Bartholin gland
O48 - Prolonged pregnancy
O12 - Gestational oedema and proteinuria without hypertension
O24.4 - Diabetes mellitus arising in pregnancy
O15 - Eclampsia
O13 - Gestational hypertension
O14 - Pre-eclampsia
D25 - Leiomyoma of uterus
N84.1 - Polyp of cervix uteri
N84 - Polyp of female genital tract
O42 - Premature rupture of membranes
O60 - Preterm labor and delivery
C53 - Malignant neoplasm of cervix uteri
N72 - Inflammatory disease of cervix uteri
N86 - Erosion and ectropion of cervix uteri
N87 - Dysplasia of cervix uteri
N91.0-N91.2 - Amenorrhea
O26.6 - Liver disorders in pregnancy
N76 - Other inflammation of vagina and vulva
N71 - Inflammatory disease of uterus, except cervix
N91.N92.1.N92.5.N92.6 - Irregular menstruations
N92 - Excessive, frequent and irregular menstr.
N97.2 - Female infertility of uterine origin
O00 - Ectopic pregnancy
O02 - Other abnormal products of conception
O03 - Sponatenous abortion
O20 - Hemorrhage in early pregnancy
N97.0 - Female infertility associated with anovulation
E28.2 - Polycystic ovarian syndrome
N91.3-N91.5 - Oligomenorrhea
N97.1 - Female infertility of tubal origin
N70 - Salpingitis and oophoritis
N73 - Other female pelvic inflammatory diseases
N97.8 - Female infertility of other origin
N80 - Endometriosis
N83.0-N83.2 - Ovarian cysts

b)

Figure 8. Results of Pairwise Genetic Correlation Analysis of a) 38 Female Reproductive Diagnoses and b) Obstetrical Diagnoses Related to Miscarriage and Placenta. Column and row names are diagnosis ICD-10 codes. Positive correlations are shown in red colour, and negative ones are in blue. Larger squares correspond to lower p-values. Asterisks indicate significant correlations after Benjamini-Hochberg correction for multiple testing (n = 1081). Light blue colour highlights clusters containing diagnoses of our interest ("other abnormal products of conception" (O02), "spontaneous abortion" (O03), "haemorrhage in early pregnancy" (O20), and "premature rupture of membranes" (O42)). The numbers in subfigure b) are exact r_g values.

Secondly, from the LDSC batch genetic correlation analysis, where we compared each diagnosis of our interest with 1403 independent traits available in a public genetic correlation database from the Complex Traits Genetics Virtual Lab [110], we found that after Bonferroni correction for multiple testing ($p < 0.05/1403 = 3.5 \times 10^{-5}$), only diagnosis "haemorrhage in early pregnancy" had significant genetic correlations with any other traits (99 different traits). Figure 9 illustrates genetic correlations of "haemorrhage in early pregnancy" with selected relevant traits from several categories spanning smoking, reproductive, cardiometabolic diseases, and other diseases. We observed the largest number of traits from cardiometabolic disorders category, and the highest genetic correlations with maternal heart disease (r_g = 0.48, p = 1.08×10^{-5}), diseases of genitourinary system (r_g = 0.46, p = 4.6×10^{-6}), and maternal age at first live birth (r_g = –0.52, p = 2.5×10^{-11}).

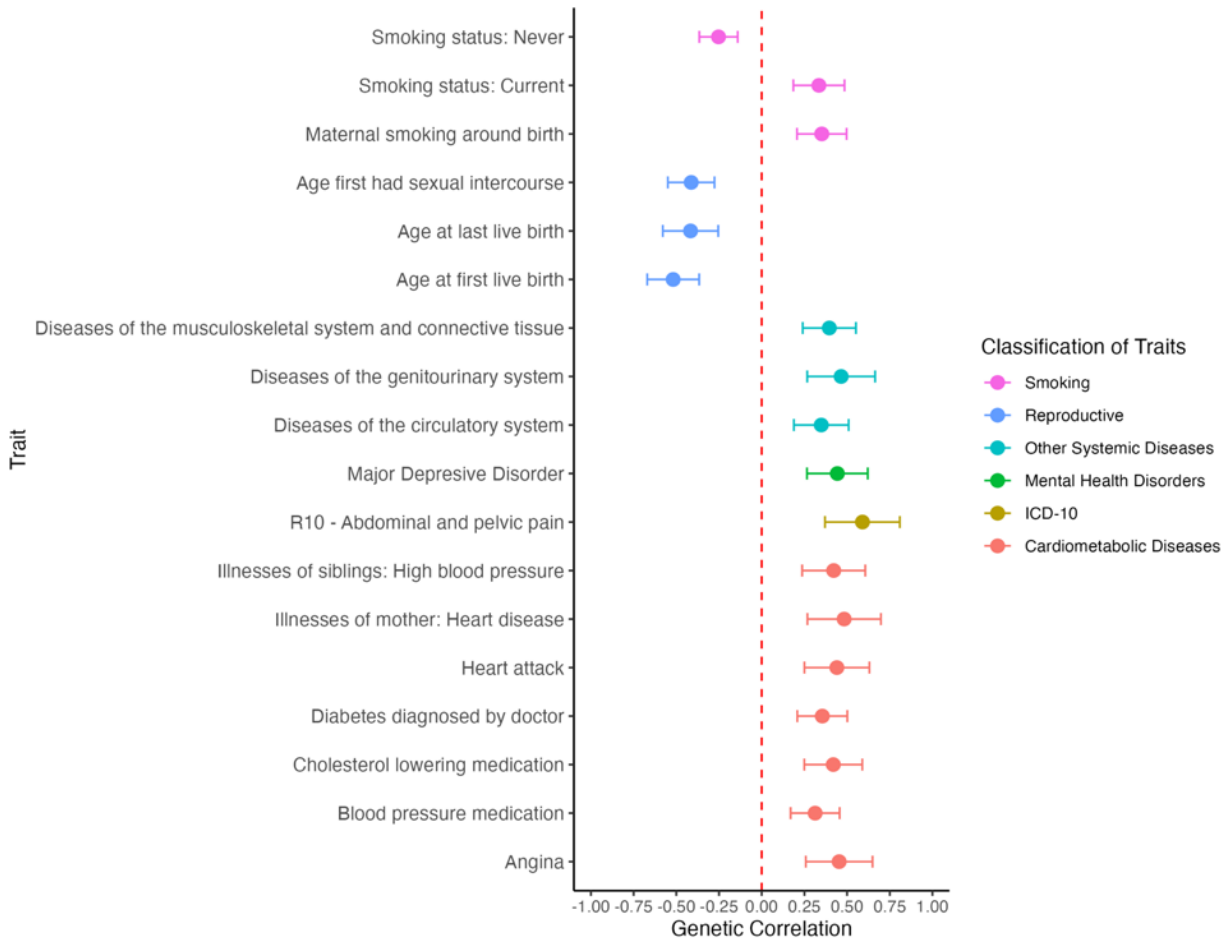


Figure 9. LDSC Batch Genetic Correlation Analysis of Diagnosis Haemorrhage in Early Pregnancy with Selected Relevant Traits. Exact r_g estimates are marked with the dot, while the error bars are 95% confidence interval. The red dashed line represents zero genetic correlation.

3.2.4 SNP-Heritability

To calculate SNP-heritability (h^2_{SNP}), we have transformed observed SNP-heritability estimates outputted by LDSC v1.0.1 tool [116,117] to the liability scale, assuming that the population prevalence is equal to the sample prevalence. The highest SNP-heritability of 4% was estimated for the diagnosis "haemorrhage in early pregnancy", while for the other three diagnoses was as follows: "premature rupture of membranes" - 2%, "other abnormal products of conception" - 2% and "spontaneous abortion" - 1%. Due to the insufficient sample size, SNP-heritability estimates were unreliable for three diagnoses ("hydatidiform mole", "placenta previa", and "premature separation of placenta (abruptio placentae)") and, thus, excluded. Detailed results are presented in Figure 10 and Supplementary Table S3.

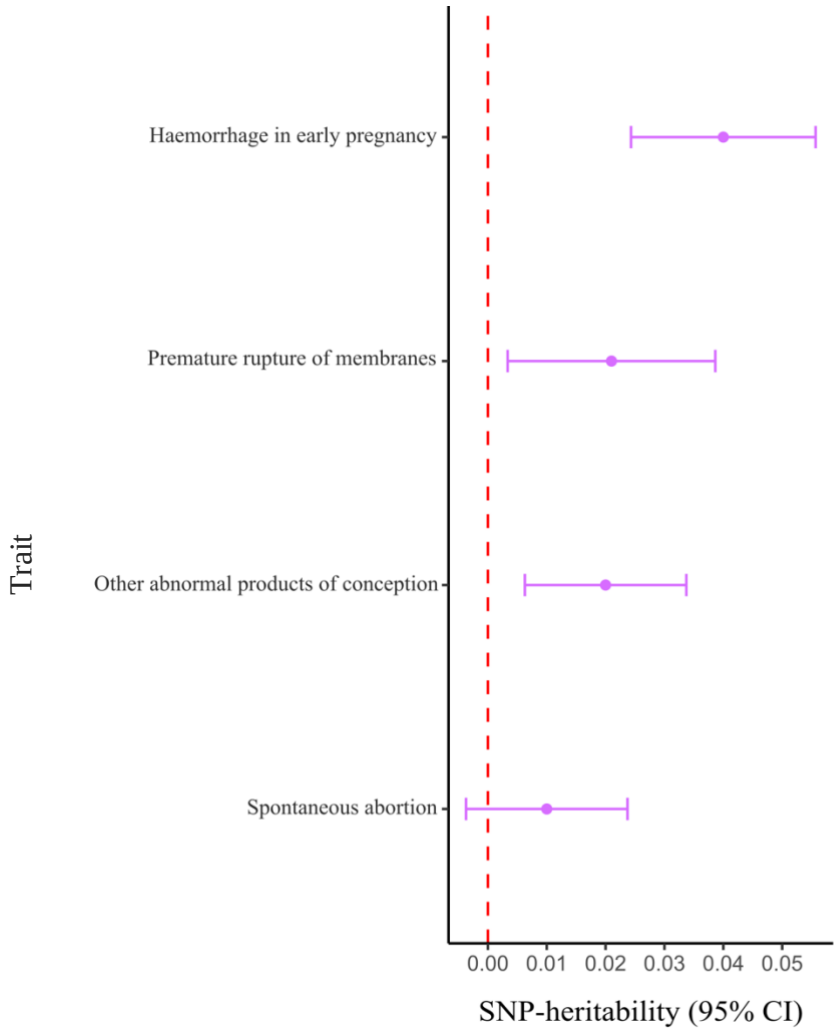


Figure 10. SNP-Heritability (h^2_{SNP}) Estimates for Four Diagnoses of Interest. Exact $h^2_{SNP}liability$ estimates are marked with the dot, while the error bars are 95% confidence interval. The red dashed line represents zero $h^2_{SNP}liability$ estimates.

3.2.5 MTAG Results

To boost the power of our GWAS meta-analyses and possibly detect novel significantly associated loci, we performed MTAG analysis. After combining diagnoses "ectopic pregnancy" (O00), "other abnormal products of conception" (O02), "spontaneous abortion" (O03) and "haemorrhage in early pregnancy" (O20), a significant novel locus was detected only in diagnosis "haemorrhage in early pregnancy" on chromosome 10 (rs192543502, p = 2.45×10^{-8}, MAF = 1.2%). The p-value of the same SNP in the original meta-analysis was 1.32×10^{-7}. However, the mean χ^2 statistic (*i.e.*, a measure of the overall statistical power of GWAS) for the individual diagnoses was as follows: O00 - 1.005, O02 - 0.996, O03 - 0.995, and O20 - 1.001. Similarly, in the pairwise analysis of the diagnoses "premature rupture of

membranes" (O42) and "preterm labor and delivery" (O60), the mean χ^2 statistic for O42 was 0.988 and for O60 was 1.007.

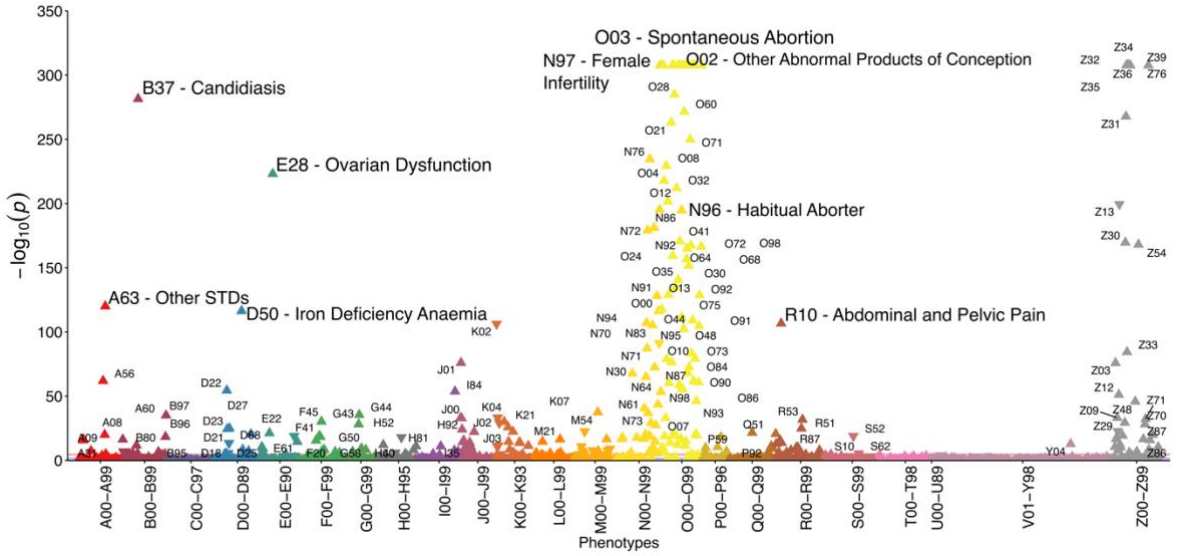
Authors of the method have specified that the robustness of calculations was not thoroughly tested when the mean value of the χ^2 statistic is lower than 1.02. Considering the low heritability of the individual diagnosis, the results may have incorrect estimates, which can inflate the type I (false-positive) error rate. Therefore, MTAG analysis did not yield reliable novel results.

3.2.6 Associated Phenotypes

To further explore four diagnoses with significant genetic correlation results ("other abnormal products of conception", "spontaneous abortion", "haemorrhage in early pregnancy" and "premature rupture of membranes") and to gain a more comprehensive understanding of their relationship with different diagnoses, we also performed a phepheWAS analysis using in-house R scripts. Specifically, we used a logistic regression framework adjusted for age and ten genetic PCs to explore which other ICD-10 diagnoses in the EstBB 200K database are more common among cases of respective obstetrical diagnoses.

According to the results, all diagnoses analysed in this analysis were associated with an increased occurrence of "diseases of the genitourinary system" (N00–N99), diagnoses related to "pregnancy, childbirth and the puerperium" (O00–O99) and "health services in circumstances related to reproduction" (Z30–Z39). Additionally, "other predominantly sexually transmitted diseases" (A63), "candidiasis" (B37), "iron deficiency anaemia" (D50), "ovarian dysfunction" (E28), "abdominal and pelvic pain" (R10) emerged as significant associations from their distinctive categories in cases diagnosed with "other abnormal products of conception", "spontaneous abortion", and "haemorrhage in early pregnancy". Interestingly, we found that "haemorrhage in early pregnancy" has exceptionally strong association with "other abnormal products of conception" (O02), "spontaneous abortion" (O03), and "female infertility" (N97), while "premature rupture of membranes" had with "preterm labour and delivery" (O60), "infections of genitourinary tract in pregnancy" (O23) and "haemorrhage in early pregnancy" (O20). Figure 11 illustrates the statistically significant associations of other diagnoses in cases diagnosed with "haemorrhage in early pregnancy" and "premature rupture of membranes". The results for the other two diagnoses are presented in Supplementary Figure S4.

a)



b)

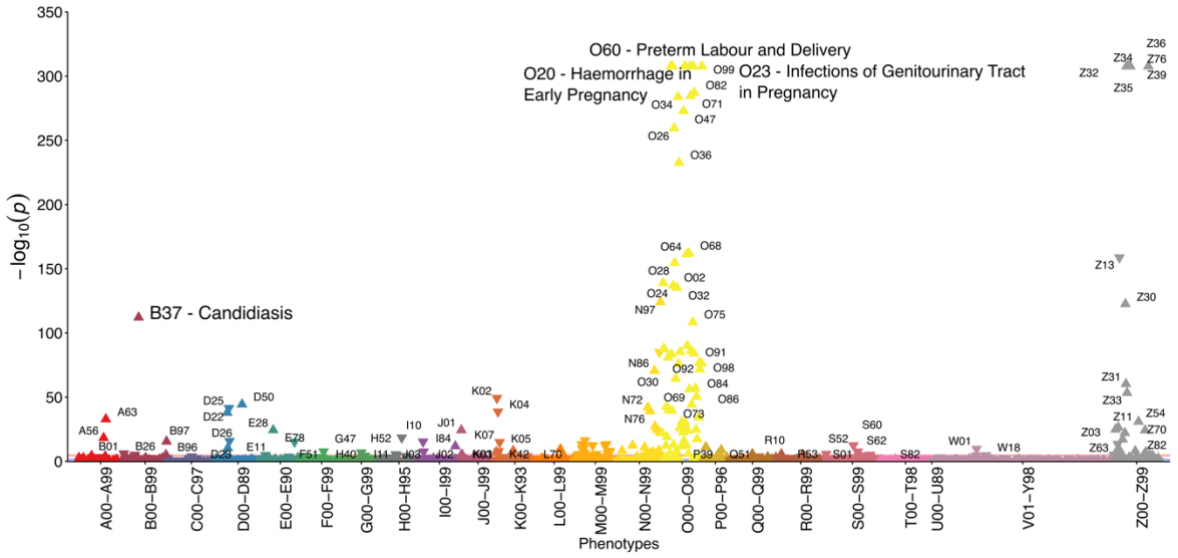


Figure 11. Associated Phenotypes in Cases Diagnosed with a) Haemorrhage in Early Pregnancy (O20) and b) Premature Rupture of Membranes (O42). The pheWAS plot illustrates the p-values of logistic regression phenotype association testing of the other diagnoses (*e.g.*, the triangles' orientation (upwards/downwards) indicates whether the cases diagnosed with "haemorrhage in early pregnancy" have an increased or decreased odds of having another diagnosis). On the x-axis are represented phenotypes/diagnoses categories from the ICD-10 system color-coded by the chapters, while on the y-axis is the –log10 p-value level of statistical significance. Each triangle is one diagnosis code. The red line marks the Bonferroni correction set at 2.5×10^{-5}. The most significantly associated phenotypes are labelled.

3.3 DISCUSSION

To elucidate causal mechanisms of adverse pregnancy outcomes, we conducted the first exploratory large-scale genetic analysis of four unique diagnoses related to miscarriage ("hydatidiform mole", "other abnormal products of conception", "spontaneous abortion", "haemorrhage in early pregnancy") and three placenta-associated obstetrical diagnoses ("premature rupture of membranes", "placenta previa" and "premature separation of placenta (abruptio placentae)"). After including up to 18,171 cases and 269,316 controls in GWAS meta-analyses, we detected two novel susceptibility loci and proposed the most likely candidate genes in each locus: 17p13.3 risk locus with *SMG6*, *HIC1* and *SRR* genes associated with the diagnosis "haemorrhage in early pregnancy" and 12q13.13 risk locus with *SLC4A8* and *GALNT6* genes associated with the diagnosis "premature rupture of membranes". We further presented evidence of shared genetic background between the studied traits of our interest, and their phenotypic association with "diseases of the genitourinary system", diagnoses related to "pregnancy, childbirth and the puerperium", and "health services related to reproduction".

We characterised SNP-heritability for four diagnoses, and found it to be low, ranging from 1 to 4% (lowest in "spontaneous abortion" and highest in "haemorrhage in early pregnancy"), whereas for other three diagnoses ("hydatidiform mole", "placenta praevia" and "premature separation of placenta") the sample sizes were too low to estimate SNP-heritability. This is also mirrored in our variant level analysis, where we could not find genetic associations for traits with lower heritability. However, this low heritability makes sense from an evolutionary point of view, since a "genetic profile" that results in embryonic lethality cannot be passed onto next generations, and, thus, we would expect very small effects from common variants, which are the focus of this study.

Next, our entirely data-driven approach confirmed genetic similarities between pregnancies with abortive outcomes and placenta-associated pregnancy complications. From the genetic correlation analysis across a wide variety of publicly available trait domains, we observed statistically significant genetic correlations only in diagnosis "haemorrhage in early pregnancy", which aligns with expectations, given that this diagnosis had the highest estimated SNP-heritability. Specifically, we observed positive correlations and shared genetic basis with diseases of genitourinary, and circulatory system. Interestingly, in the diagnosis "haemorrhage in early pregnancy", we found positive correlations with maternal and siblings cardiometabolic diseases implicating a potential multigenerational or long-term

effect of these complex diseases on adverse pregnancy outcomes. However, this result, consistent with some studies [123,124], but not universally supported [125], must be interpreted cautiously, since it may also reflect other confounders, such as socioeconomic status. Similarly, revealed negative correlation with maternal age at first/last live birth may point towards shared genetics, but could also stem from the fact that younger maternal age at first birth means more pregnancies, and, thus, a statistically higher chance of having pregnancy complications, as suggested in previous work [3,53,54]. Additionally, we observed that direction of the effect of environmental factors and lifestyle choices modifying genetic susceptibility to developing the condition is in line with discoveries of epidemiological studies [31,126].

The phenotype-level analysis showed a strong association between diagnoses "other abnormal products of conception", "spontaneous abortion", and "haemorrhage in early pregnancy", which confirms that "haemorrhage in early pregnancy" often leads to miscarriage. Associations with "other predominantly sexually transmitted diseases" and "candidiasis" require cautious interpretation. Although it may indicate that infectious diseases are a risk factor, it can be confounding due to the fact that pregnant women are screened more often for these diseases. Regarding diagnosis "premature rupture of membranes", while it exhibited the strongest association with "preterm labour and delivery" and "haemorrhage in early pregnancy", considering the results of genetic correlations, there is more likely to be shared molecular mechanism only with "preterm labour and delivery". However, this provides supporting evidence to the study of Saraswat *et al.*, 2010 [35], that "haemorrhage in early pregnancy" can also be a predictor of third-trimester pregnancy complications.

We discovered a novel susceptibility locus at 17p13.3 associated with diagnosis "haemorrhage in early pregnancy", which has been previously reported in several studies on malignancies [127–129]. The lead SNP (rs11657636) in the intronic region of the protein-coding gene *SMG6* was in high LD with an exonic variant (rs216195) of the same gene. *SMG6* gene is involved in two important cellular processes: it encodes the part of a telomerase ribonucleoprotein complex that ensures chromosome end maintenance and meiotic progression [130], and its protein also participate in initiation of the nonsense-mediated mRNA decay (NMD) pathway needed for the degradation of aberrant mRNAs [131]. Several studies have shown that NMD components are essential for normal embryonic development. For example, when genes encoding key NMD factors, such as *SMG1* [132],

UPF1 [133], *SMG6* [134], and *SMG9* [135] were knocked out in mice, the researchers observed embryonic death. Specifically, *SMG6* is involved in regulating the differentiation and reprogramming of embryonic stem cells. Another gene implicated by gene-based test was *HIC1*, a nuclear modulator of the Wnt/β-catenin signalling pathway and tumour suppressor [136]. Wnt/β-catenin pathway is involved in cell proliferation, polarisation, and fate determination during embryogenesis [137]. Also, previous studies reported a connection between *SMG6* and *HIC1* gene. Namely, the expression of *HIC1* is directly regulated by interaction of retinoic acid (RA) and RA receptors found within an intron of the *SMG6* gene [138,139]. Another important gene to mention in this locus is the *SRR* gene, located right next to the *SMG6* gene, and mapped by both eQTLs and Hi-C. *SRR* gene product (serine racemase) catalyses the synthesis of D-serine, an amino acid important in fetal brain development [140,141]. Moreover, one study provides evidence that placenta could be directly involved in supplying D-serine to the fetal circulation, [142] as they have found mRNAs of serine racemase in it. However, our results must be interpreted cautiously because "haemorrhage in early pregnancy" is a heterogenous ICD-10 code and a common symptom of various diagnoses (cervical problems, placental complications, miscarriage). To validate our findings, future studies should filter for severe cases or per specific cause of bleeding. Furthermore, accurately measuring the quantity and severity of bleeding, and assessing for potential comorbidities would be preferable. This is difficult to do in a population-based biobank setting, therefore large cohorts with relevant clinical data would be needed.

We also described a novel susceptibility locus at 12q13.13 associated with diagnosis "premature rupture of membranes". The lead variant (rs4237901) is located in the intronic region of protein-coding gene *SLC4A8*, which is next to the *GALNT6* gene. *SLC4A8* encodes NDCBE1 membrane protein, which transports sodium and bicarbonate ions across the cell membrane, thereby regulating the pH balance of the intracellular and extracellular environments, especially in neurons [143]. Study on mouse models suggests that the *SLC4A8* gene also contributes to renal fluid homeostasis and electrolyte balance [144]. Neighbouring *GALNT6* gene, also mapped by eQTLs, is mainly involved in the glycosylation of proteins [145] and highly expressed in many types of cancer [146–148]. Moreover, *GALNT6* gene has been seen as the potential candidate for the synthesis of oncofetal fibronectin [149], which has been evaluated as a predictor of premature birth in a low sample size study [150]. Also, elevated levels of vaginal fetal fibronectin have been found in women who experience PPROM [151]. All this warrants for further exploration of the *GALNT6* gene, since the PPROM is a common cause of premature birth.

Although GWAS is an efficient tool for studying disease biology, we acknowledge that it has limitations. We could not detect any genetic associations in five diagnoses ("hydatidiform mole", "other abnormal products of conception", "spontaneous abortion", "placenta previa", and "premature separation of placenta"), indicating we most likely lack the statistical power with the current sample sizes. In general, investigation of common variants with small effect sizes requires large sample sizes. The estimation of the required number of cases and total sample size is difficult, as it may vary widely, depending on the heritability, the underlying genetic architecture of the trait, and the population characteristics [152]. We also could not replicate significantly associated SNPs and loci reported in Laisk *et al.*, 2020 study. Our study's lack of supporting evidence may be due to the different phenotype definition, sample size or population ancestry.

In the underpowered GWA studies that yield few genome-wide significant signals, pathway and tissue analyses are often ineffective in providing significant results. Although we could not find statistically significant tissue enrichments for our GWAS signals in diagnoses "haemorrhage in early pregnancy" and "premature rupture of membranes", nominal associations with tissues, such as uterus and blood vessels, may still provide some insight into the pathogenesis of these conditions. It is important to emphasise that it is a commonly recognised limitation that the datasets used for these types of enrichment analyses (such as the GTEx database) lack sufficient information from tissues/timepoints relevant for female reproductive health and pregnancy.

Finally, studies of early pregnancy complications are limited partly due to ethical concerns that arise when studying pregnancies and developing fetuses, especially pregnancies that result in miscarriage [153]. With the current EstBB and FinnGen population-based biobank setting, we were able to explore only maternal genetic variation contributing to the development of obstetrical diagnoses related to miscarriage and placental biology. Additional large-scale genetic studies of the paternal and fetal genomes are needed to develop a complete picture of the underlying biology of adverse pregnancy outcomes. Moreover, combination with human placental gene expression and placental histology data could significantly improve our understanding of the placentation and how its alterations may contribute to miscarriage and adverse pregnancy outcomes.

SUMMARY

In this thesis, we set out to characterise the maternal genetic basis of seven specific obstetrical diagnoses related to miscarriage and placental biology: “hydatidiform mole”, “other abnormal products of conception”, “spontaneous abortion”, “haemorrhage in early pregnancy”, “premature rupture of membranes”, “placenta previa” and “premature separation of placenta (abruptio placentae)”. Additionally, we attempted to comprehend the extent of the shared genetic and phenotypic background of the respective diagnoses between each other, with other female reproductive diagnoses and with a broad range of traits.

Using the Estonian and FinnGen Biobank GWAS summary statistic data in a robust GWAS meta-analysis, and a set of established bioinformatics tools for large-scale genetic analyses, we discovered genomic risk loci associated with diagnoses “haemorrhage in early pregnancy” and “premature rupture of membranes”. For “haemorrhage in early pregnancy”, we dissected the 17p13.3 risk locus in detail and proposed several potentially involved genes, among which we highlight *SMG6*, *HIC1*, and *SRR* genes for follow-up studies. Similarly, for diagnosis “premature rupture of membranes”, we propose *SLC4A8* and *GALNT6* genes as potential candidate genes at 12q13.13 genomic locus. For other diagnoses, we could not draw clear conclusions about involved genetic factors due to the low estimated SNP-heritability and the limited statistical power of the data. Nonetheless, our genetic correlation and phenotype-level analyses provide supporting evidence of shared biology and risk factors. However, these results require careful interpretation, as they might reflect socioeconomic status and potential comorbidities.

In conclusion, the rich genomic and health-related data of Estonian and Finnish population-based biobanks enabled us to obtain novel insights into understudied obstetrical complications that lack suitable animal models, and relevant gene expression and tissue enrichment information. Further study in an independent cohort is needed to validate our results. Given the substantial genetic and phenotypic overlap we characterised between investigated diagnoses, additional efforts are needed to distinguish between these closely related conditions. We suggest continuing the research on obstetrical diagnoses with larger and more diverse datasets and approaching the topic from several perspectives (large-scale genetic analyses of maternal, paternal, fetal, and placental genomes) to move the field of female reproductive health and neonatal care forward.

Acknowledgements

First and foremost, I express my biggest gratitude to my supervisors, Triin Laisk, Natàlia Pujol Gualdo and Reedik Mägi, for accepting me in their research group, their endless patience, constant support in every manner, and for guidance in the fascinating world of genomic epidemiology.

Next, I express my gratitude to the participants and researchers of the Estonian Biobank and the FinnGen study for their generous contribution to this work.

I am grateful to my programme director Ilona Faustova for accepting me to the Bioengineering master's programme, the Education and Youth Board of Estonia, and the University of Tartu for the financial support with the Dora Plus and Achievement scholarships. In addition, I wish to thank all of my professors, who have been both exceptional educators and thoughtful and kind individuals.

Finally, I wish to thank my life partner for being the strongest wind in my sails, my mom, my aunt and all my friends for their support through my master's journey.

REFERENCES

1. World Health Organization. (2022). *Abortion care guideline*. Geneva: WHO.
2. ESHRE Guideline Group on RPL, Bender Atik, R., Christiansen, O. B., Elson, J., Kolte, A. M., Lewis, S., ... and Goddijn, M. (2023). ESHRE guideline: recurrent pregnancy loss: an update in 2022. *Human reproduction open*, *2023*(1), hoad002.
3. Nybo Andersen, A. M., Wohlfahrt, J., Christens, P., Olsen, J., and Melbye, M. (2000). Maternal age and fetal loss: population based register linkage study. *BMJ (Clinical research ed.)*, *320*(7251), 1708–1712.
4. Wang, X., Chen, C., Wang, L., Chen, D., Guang, W., and French, J. (2003). Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertility and sterility*, *79*(3), 577–584.
5. Laisk, T., Soares, A. L. G., Ferreira, T., Painter, J. N., Censin, J. C., Laber, S., Bacelis, J., Chen, C. Y., Lepamets, M., Lin, K., Liu, S., Millwood, I. Y., Ramu, A., Southcombe, J., Andersen, M. S., Yang, L., Becker, C. M., Børglum, A. D., Gordon, S. D., Bybjerg-Grauholm, J., … Lindgren, C. M. (2020). The genetic architecture of sporadic and multiple consecutive miscarriage. *Nature communications*, *11*(1), 5980.
6. Jauniaux, E., Jurkovic, D., and Campbell, S. (1991). In vivo investigations of the anatomy and the physiology of early human placental circulations. *Ultrasound in obstetrics and gynecology: the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, *1*(6), 435–445.
7. Norton, C., Clarke, D., Holmstrom, J., Stirland, I., Reynolds, P. R., Jenkins, T. G., and Arroyo, J. A. (2023). Altered Epigenetic Profiles in the Placenta of Preeclamptic and Intrauterine Growth Restriction Patients. *Cells*, *12*(8), 1130.
8. World Health Organization. (2019). *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. WHO.
9. Tervise Arengu Instituut. (2020). *Statistical data of EMBR and EAR*. Available at: https://en.tai.ee/en/r-and-d/registers/estonian-medical-birth-registry-and-estonian-abortion-registry/statistical-data-of-embr-and-ear (10.05.2023)
10. Cunningham, F. G., Leveno, K. J., Dashe, J. S., Hoffman, B. L., Spong, C. Y., and Casey, B. M. (2022). *Williams Obstetrics 26e*. McGraw Hill Professional.
11. Candelier J. J. (2016). The hydatidiform mole. *Cell adhesion and migration*, *10*(1-2), 226–235.

12. Lund, H., Vyberg, M., Eriksen, H. H., Grove, A., Jensen, A. Ø., and Sunde, L. (2020). Decreasing incidence of registered hydatidiform moles in Denmark 1999–2014. *Scientific reports*, *10*(1), 17041.
13. Hammond, C. B., Weed Jr, J. C., Barnard, D. E., and Tyrey, L. (1981). Gestational trophoblastic neoplasia. *CA: A Cancer Journal for Clinicians*, *31*(6), 322–332.
14. Savage, P. M., Sita-Lumsden, A., Dickson, S., Iyer, R., Everard, J., Coleman, R., ... and Seckl, M. J. (2013). The relationship of maternal age to molar pregnancy incidence, risks for chemotherapy and subsequent pregnancy outcome. *Journal of Obstetrics and Gynaecology*, *33*(4), 406–411.
15. Eagles, N., Sebire, N. J., Short, D., Savage, P. M., Seckl, M. J., and Fisher, R. A. (2015). Risk of recurrent molar pregnancies following complete and partial hydatidiform moles. *Human reproduction*, *30*(9), 2055–2063.
16. Moglabey, Y. B., Kircheisen, R., Seoud, M., El Mogharbel, N., Van den Veyver, I., and Slim, R. (1999). Genetic mapping of a maternal locus responsible for familial hydatidiform moles. *Human molecular genetics*, *8*(4), 667–667.
17. Murdoch, S., Djuric, U., Mazhar, B., Seoud, M., Khan, R., Kuick, R., Bagga, R., Kircheisen, R., Ao, A., Ratti, B., Hanash, S., Rouleau, G. A., and Slim, R. (2006). Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans. *Nature genetics*, *38*(3), 300–302.
18. Fallahian, M., Sebire, N. J., Savage, P. M., Seckl, M. J., and Fisher, R. A. (2013). Mutations in NLRP7 and KHDC3L confer a complete hydatidiform mole phenotype on digynic triploid conceptions. *Human mutation*, *34*(2), 301–308.
19. Mahadevan, S., Wen, S., Wan, Y. W., Peng, H. H., Otta, S., Liu, Z., Iacovino, M., Mahen, E. M., Kyba, M., Sadikovic, B., and Van den Veyver, I. B. (2014). NLRP7 affects trophoblast lineage differentiation, binds to overexpressed YY1 and alters CpG methylation. *Human molecular genetics*, *23*(3), 706–716.
20. Moshtaghi, A., Vaziri, H., Sariri, R., and Shaigan, H. (2017). Polymorphism of *MnSOD (Val16Ala)* gene in pregnancies with blighted ovum: A case-control study. *International journal of reproductive biomedicine*, *15*(8), 503–508.
21. Jiang, W. Z., Yang, X. L., and Luo, J. R. (2022). Risk factors for missed abortion: retrospective analysis of a single institution's experience. *Reproductive biology and endocrinology: RBandE*, *20*(1), 115.

22. Helle, N., Niinimäki, M., Linnakaari, R., But, A., Gissler, M., Heikinheimo, O., and Mentula, M. (2022). National register data are of value in studies on miscarriage-Validation of the healthcare register data in Finland. *Acta obstetricia et gynecologica Scandinavica*, *101*(11), 1245–1252.

23. Pylyp, L. Y., Spynenko, L. O., Verhoglyad, N. V., Mishenko, A. O., Mykytenko, D. O., and Zukin, V. D. (2018). Chromosomal abnormalities in products of conception of first-trimester miscarriages detected by conventional cytogenetic analysis: a review of 1000 cases. *Journal of assisted reproduction and genetics*, *35*(2), 265–271.

24. Cumming, G. P., Klein, S., Bolsover, D., Lee, A. J., Alexander, D. A., Maclean, M., and Jurgens, J. D. (2007). The emotional burden of miscarriage for women and their partners: trajectories of anxiety and depression over 13 months. *BJOG: an international journal of obstetrics and gynaecology*, *114*(9), 1138–1145.

25. Wagner, M. M., Bhattacharya, S., Visser, J., Hannaford, P. C., and Bloemenkamp, K. W. (2015). Association between miscarriage and cardiovascular disease in a Scottish cohort. *Heart (British Cardiac Society)*, *101*(24), 1954–1960.

26. Maino, A., Siegerink, B., Algra, A., Martinelli, I., Peyvandi, F., & Rosendaal, F. R. (2016). Pregnancy loss and risk of ischaemic stroke and myocardial infarction. *British journal of haematology*, *174*(2), 302–309.

27. Nagaishi, M., Yamamoto, T., Iinuma, K., Shimomura, K., Berend, S. A., and Knops, J. (2004). Chromosome abnormalities identified in 347 spontaneous abortions collected in Japan. *The journal of obstetrics and gynaecology research*, *30*(3), 237–241.

28. Chen, S., Liu, D., Zhang, J., Li, S., Zhang, L., Fan, J., Luo, Y., Qian, Y., Huang, H., Liu, C., Zhu, H., Jiang, Z., and Xu, C. (2017). A copy number variation genotyping method for aneuploidy detection in spontaneous abortion specimens. *Prenatal diagnosis*, *37*(2), 176–183.

29. Giakoumelou, S., Wheelhouse, N., Cuschieri, K., Entrican, G., Howie, S. E., and Horne, A. W. (2016). The role of infection in miscarriage. *Human reproduction update*, *22*(1), 116–133.

30. Bellver, J., Rossal, L. P., Bosch, E., Zúñiga, A., Corona, J. T., Meléndez, F., Gómez, E., Simón, C., Remohí, J., and Pellicer, A. (2003). Obesity and the risk of spontaneous abortion after oocyte donation. *Fertility and sterility*, *79*(5), 1136–1140.

31. Wright, A. D., Nicholson, H. O., Pollock, A., Taylor, K. G., and Betts, S. (1983). Spontaneous abortion and diabetes mellitus. *Postgraduate Medical Journal*, *59*(691), 295–298.
32. Wakim, A. N., Polizotto, S. L., Buffo, M. J., Marrero, M. A., and Burholt, D. R. (1993). Thyroid hormones in human follicular fluid and thyroid hormone receptors in human granulosa cells. *Fertility and sterility*, *59*(6), 1187–1190.
33. Everett C. (1997). Incidence and outcome of bleeding before the 20th week of pregnancy: prospective study from general practice. *BMJ (Clinical research ed.)*, *315*(7099), 32–34.
34. Hasan, R., Baird, D. D., Herring, A. H., Olshan, A. F., Jonsson Funk, M. L., and Hartmann, K. E. (2009). Association between first-trimester vaginal bleeding and miscarriage. *Obstetrics and gynecology*, *114*(4), 860–867.
35. Saraswat, L., Bhattacharya, S., Maheshwari, A., and Bhattacharya, S. (2010). Maternal and perinatal outcome in women with threatened miscarriage in the first trimester: a systematic review. *BJOG : an international journal of obstetrics and gynaecology*, *117*(3), 245–257.
36. Breeze C. (2016). Early pregnancy bleeding. *Australian family physician*, *45*(5), 283–286.
37. Duff P. (1996). Premature rupture of the membranes in term patients. *Seminars in perinatology*, *20*(5), 401–408.
38. Menon, R., and Richardson, L. S. (2017). Preterm prelabor rupture of the membranes: A disease of the fetal membranes. *Seminars in perinatology*, *41*(7), 409–419.
39. Sari, I. M., Adisasmita, A. C., Prasetyo, S., Amelia, D., and Purnamasari, R. (2020). Effect of premature rupture of membranes on preterm labor: a case-control study in Cilegon, Indonesia. *Epidemiology and health*, *42*, e2020025.
40. Riegel, K., Söhne, B., Fischer, P., Ort, B., Wolke, D., and Osterlund, K. (1999). Der vorzeitige Blasensprung, das Infektionsrisiko und die kindliche Prognose--Zwei Regionen im Vergleich [Premature rupture of fetal membranes, risk of infection and infant prognosis--a comparison of 2 regions]. *Zeitschrift fur Geburtshilfe und Neonatologie*, *203*(4), 152–160.
41. Caughey, A. B., Robinson, J. N., and Norwitz, E. R. (2008). Contemporary diagnosis and management of preterm premature rupture of membranes. *Reviews in obstetrics and gynecology*, *1*(1), 11–22.

42. Harger, J. H., Hsing, A. W., Tuomala, R. E., Gibbs, R. S., Mead, P. B., Eschenbach, D. A., ... and Polk, B. F. (1990). Risk factors for preterm premature rupture of fetal membranes: a multicenter case-control study. *American journal of obstetrics and gynecology*, *163*(1), 130–137.

43. Nakubulwa, S., Kaye, D. K., Bwanga, F., Tumwesigye, N. M., and Mirembe, F. M. (2015). Genital infections and risk of premature rupture of membranes in Mulago Hospital, Uganda: a case control study. *BMC research notes*, *8*(1), 1–9.

44. Vaas, P., and Rull, K. (2009). Twin pregnancy: maternal risks, pregnancy complications, obstetric and perinatal outcome at the Women's Clinic of Tartu University Hospitals (2003–2007). *Eesti Arst*, *88*(3), 164–173.

45. Assefa, N. E., Berhe, H., Girma, F., Berhe, K., Berhe, Y. Z., Gebreheat, G., ... and Welu, G. (2018). Risk factors of premature rupture of membranes in public hospitals at Mekele city, Tigray, a case control study. *BMC pregnancy and childbirth*, *18*(1), 1–7.

46. Romero, R., Friel, L. A., Edwards, D. R. V., Kusanovic, J. P., Hassan, S. S., Mazaki-Tovi, S., ... and Menon, R. (2010). A genetic association study of maternal and fetal candidate genes that predispose to preterm prelabor rupture of membranes (PROM). *American journal of obstetrics and gynecology*, *203*(4), 361-e1.

47. Knight, M., and UKOSS (2007). Peripartum hysterectomy in the UK: management and outcomes of the associated haemorrhage. *BJOG: an international journal of obstetrics and gynaecology*, *114*(11), 1380–1387.

48. Wang, Y., and Huang, X. (2018). Sepsis after uterine artery embolization-assisted termination of pregnancy with complete placenta previa: a case report. *Journal of International Medical Research*, *46*(1), 546–550.

49. Kainer, F., and Hasbargen, U. (2008). Emergencies associated with pregnancy and delivery: peripartum hemorrhage. *Deutsches Arzteblatt international*, *105*(37), 629–638.

50. Crane, J. M., van den Hof, M. C., Dodds, L., Armson, B. A., and Liston, R. (1999). Neonatal outcomes with placenta previa. *Obstetrics and gynecology*, *93*(4), 541–544.

51. Nørgaard, L. N., Pinborg, A., Lidegaard, Ø., and Bergholt, T. (2012). A Danish national cohort study on neonatal outcome in singleton pregnancies with placenta previa. *Acta obstetricia et gynecologica Scandinavica*, *91*(5), 546–551.

52. Cresswell, J. A., Ronsmans, C., Calvert, C., and Filippi, V. (2013). Prevalence of placenta praevia by world region: a systematic review and meta-analysis. *Tropical medicine and international health: TM and IH*, *18*(6), 712–724.

53. FASTER Consortium. (2005). Impact of maternal age on obstetric outcome. *Obstetrics and Gynecology*, *105*(5 Pt 1), 983–990.

54. Babinszki, A., Kerenyi, T., Torok, O., Grazi, V., Lapinski, R. H., and Berkowitz, R. L. (1999). Perinatal outcome in grand and great-grand multiparity: effects of parity on obstetric risk factors. *American journal of obstetrics and gynecology*, *181*(3), 669–674.

55. Silver, R. M., Landon, M. B., Rouse, D. J., Leveno, K. J., Spong, C. Y., Thom, E. A., Moawad, A. H., Caritis, S. N., Harper, M., Wapner, R. J., Sorokin, Y., Miodovnik, M., Carpenter, M., Peaceman, A. M., O'Sullivan, M. J., Sibai, B., Langer, O., Thorp, J. M., Ramin, S. M., Mercer, B. M., … National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network (2006). Maternal morbidity associated with multiple repeat cesarean deliveries. *Obstetrics and gynecology*, *107*(6), 1226–1232.

56. Kistin, N., Handler, A., Davis, F., and Ferre, C. (1996). Cocaine and cigarettes: a comparison of risks. *Paediatric and perinatal epidemiology*, *10*(3), 269–278.

57. Rasmussen, S., Irgens, L. M., and Dalaker, K. (2000). Outcome of pregnancies subsequent to placental abruption: a risk assessment. *Acta obstetricia et gynecologica Scandinavica*, *79*(6), 496–501.

58. Kåregård, M., and Gennser, G. (1986). Incidence and recurrence rate of abruptio placentae in Sweden. *Obstetrics and gynecology*, *67*(4), 523–528.

59. Tikkanen, M., Riihimäki, O., Gissler, M., Luukkaala, T., Metsäranta, M., Andersson, S., Ritvanen, A., Paavonen, J., and Nuutila, M. (2012). Decreasing incidence of placental abruption in Finland during 1980-2005. *Acta obstetricia et gynecologica Scandinavica*, *91*(9), 1046–1052.

60. Zetterström, K., Lindeberg, S. N., Haglund, B., and Hanson, U. (2005). Maternal complications in women with chronic hypertension: a population-based cohort study. *Acta obstetricia et gynecologica Scandinavica*, *84*(5), 419–424.

61. Rice, J. P., Kay, H. H., and Mahony, B. S. (1989). The clinical significance of uterine leiomyomas in pregnancy. *American journal of obstetrics and gynecology*, *160*(5 Pt 1), 1212–1216.

62. Major, C. A., de Veciana, M., Lewis, D. F., and Morgan, M. A. (1995). Preterm premature rupture of membranes and abruptio placentae: is there an association between these pregnancy complications?. *American journal of obstetrics and gynecology*, *172*(2 Pt 1), 672–676.

63. The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678.

64. Bellenguez, C., Küçükali, F., Jansen, I. E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A. C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., Holmans, P. A., Boland, A., Damotte, V., van der Lee, S. J., Costa, M. R., Kuulasmaa, T., Yang, Q., de Rojas, I., Bis, J. C., Yaqub, A., … Lambert, J. C. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature genetics*, *54*(4), 412–436.

65. Norton, S., Matthews, F. E., Barnes, D. E., Yaffe, K., and Brayne, C. (2014). Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. *The Lancet. Neurology*, *13*(8), 788–794.

66. Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., Coleman, J. R. I., Hagenaars, S. P., Ward, J., Wigmore, E. M., Alloza, C., Shen, X., Barbu, M. C., Xu, E. Y., Whalley, H. C., Marioni, R. E., Porteous, D. J., Davies, G., Deary, I. J., Hemani, G., … McIntosh, A. M. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*, *22*(3), 343–352.

67. Zhang, L., and Wu, L. (2021). Effects of environmental quality perception on depression: Subjective social class as a mediator. *International journal of environmental research and public health*, *18*(11), 6130.

68. Barrett, J. C., Clayton, D. G., Concannon, P., Akolkar, B., Cooper, J. D., Erlich, H. A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D. J., Stevens, H., Todd, J. A., Walker, N. M., Rich, S. S., and Type 1 Diabetes Genetics Consortium (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nature genetics*, *41*(6), 703–707.

69. Mahajan, A., Spracklen, C. N., Zhang, W., Ng, M. C. Y., Petty, L. E., Kitajima, H., Yu, G. Z., Rüeger, S., Speidel, L., Kim, Y. J., Horikoshi, M., Mercader, J. M., Taliun, D., Moon, S., Kwak, S. H., Robertson, N. R., Rayner, N. W., Loh, M., Kim, B. J., Chiou, J., … Morris, A. P. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nature genetics*, *54*(5), 560–572.

70. Uffelmann, E., Huang, Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1).

71. Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A. M., Ness, A. R., Ebrahim, S., Lawlor, D. A., … McCarthy, M. I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, N.Y.)*, *316*(5826), 889–894.

72. Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J. P., Russell, R. K., Sleiman, P. M., Imielinski, M., Glessner, J., Hou, C., Wilson, D. C., Walters, T., Kim, C., Frackelton, E. C., Lionetti, P., Barabino, A., Van Limbergen, J., Guthery, S., Denson, L., Piccoli, D., … Hakonarson, H. (2009). Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *American journal of human genetics*, *84*(3), 399–405.

73. Pujol-Gualdo, N., Läll, K., Lepamets, M., Estonian Biobank Research Team, Rossi, H. R., Arffman, R. K., Piltonen, T. T., Mägi, R., and Laisk, T. (2022). Advancing our understanding of genetic risk factors and potential personalized strategies for pelvic organ prolapse. *Nature communications*, *13*(1), 3584.

74. Slatkin M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature reviews. Genetics*, *9*(6), 477–485.

75. Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., and Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, *50*(9), 1219–1224.

76. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M., Pappas, D. A., … Plenge, R. M. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, *506*(7488), 376–381.

77. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature communications*, *8*(1), 1826.

78. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, *38*(16), e164.

79. de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*, *11*(4), e1004219.

80. Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H. J., Wood, A. R., Yang, J., Lui, J. C., Vedantam, S., Gustafsson, S., Esko, T., Frayling, T., Speliotes, E. K., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Boehnke, M., Raychaudhuri, S., Fehrmann, R. S., Hirschhorn, J. N., and Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications*, *6*, 5890.

81. Leitsalu, L., Haller, T., Esko, T., Tammesoo, M. L., Alavere, H., Snieder, H., ... and Metspalu, A. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *International journal of epidemiology*, *44*(4), 1137–1147.

82. *Strand Home*. Available at: https://www.well.ox.ac.uk/~wrayner/strand/ (20.05.2023)

83. Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., and L Price, A. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics*, *48*(11), 1443–1448.

84. Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics*, *81*(5), 1084–1097.

85. Mitt, M., Kals, M., Pärn, K., Gabriel, S. B., Lander, E. S., Palotie, A., Ripatti, S., Morris, A. P., Metspalu, A., Esko, T., Mägi, R., and Palta, P. (2017). Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *European journal of human genetics: EJHG*, *25*(7), 869–876.

86. Mbatchou, J., Barnard, L., Backman, J. D., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J. S., Abecasis, G. R., Maxwell, E., and Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, *53*(7), 1097–1103.

87. Kurki, M. I., Karjalainen, J., Palta, P., Sipilä, T. P., Kristiansson, K., Donner, K. M., Reeve, M. P., Laivuori, H., Aavikko, M., Kaunisto, M. A., Loukola, A., Lahtela, E., Mattsson, H., Laiho, P., Della Briotta Parolo, P., Lehisto, A. A., Kanai, M., Mars, N., Rämö, J., Kiiskinen, T., … Palotie, A. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, *613*(7944), 508–518.

88. *Beagle 4.1*. Available at: https://faculty.washington.edu/browning/beagle/b4_1.html (20.05.2023)

89. Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Gabrielsen, M. E., Wolford, B. N., LeFaive, J., VandeHaar, P., Gagliano, S. A., Gifford, A., Bastarache, L. A., Wei, W. Q., Denny, J. C., Lin, M., Hveem, K., Kang, H. M., Abecasis, G. R., Willer, C. J., and Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics*, *50*(9), 1335–1341.

90. *FinnGen results*. Available at: http://r7.finngen.fi/ (20.05.2023)

91. Mägi, R., and Morris, A. P. (2010). GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, *11*(1).

92. Rentzsch, P., Schubach, M., Shendure, J., and Kircher, M. (2021). CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome medicine*, *13*(1), 31.

93. Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., Park, J., Hitz, B. C., Weng, S., Cherry, J. M., and Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, *22*(9), 1790–1797.

94. *Roadmap Epigenomics*. Available at: https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html (20.05.2023)
95. Zhou, W., Kanai, M., Wu, K. H., Rasheed, H., Tsuo, K., Hirbo, J. B., Wang, Y., Bhattacharya, A., Zhao, H., Namba, S., Surakka, I., Wolford, B. N., Lo Faro, V., Lopera-Maya, E. A., Läll, K., Favé, M. J., Partanen, J. J., Chapman, S. B., Karjalainen, J., Kurki, M., … Neale, B. M. (2022). Global Biobank Meta-Analysis Initiative: Powering genetic discovery across human disease. *Cell genomics*, *2*(10), 100192.
96. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74.
97. Edwards, S. L., Beesley, J., French, J. D., and Dunning, A. M. (2013). Beyond GWASs: illuminating the dark road from association to function. *American journal of human genetics*, *93*(5), 779–797.
98. Hrdlickova, B., de Almeida, R. C., Borek, Z., and Withoff, S. (2014). Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et biophysica acta*, *1842*(10), 1910–1922.
99. Mirza, A. H., Kaur, S., Brorsson, C. A., and Pociot, F. (2014). Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PloS one*, *9*(8), e105723.
100. *GTEx Portal*. Available at: https://www.gtexportal.org/home/ (20.05.2023)
101. *BRAINEAC*. Available at: http://www.braineac.org/ (20.05.2023)
102. Võsa, U., Claringbould, A., Westra, H. J., Bonder, M. J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., Brugge, H., Oelen, R., de Vries, D. H., van der Wijst, M. G. P., Kasela, S., Pervjakova, N., Alves, I., Favé, M. J., Agbessi, M., Christiansen, M. W., … Franke, L. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature genetics*, *53*(9), 1300–1310.
103. Embl-Ebi. *Project < eQTL Catalogue < EMBL-EBI*. Available at: https://www.ebi.ac.uk/eqtl/ (20.05.2023)
104. *GEO Accession viewer*. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87112 (20.05.2023)

105. GTEx Consortium, Laboratory, Data Analysis andCoordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration andVisualization—EBI, Genome Browser Data Integration andVisualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis andCoordinating Center (LDACC):, NIH program management:, … Montgomery, S. B. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550*(7675), 204–213.

106. Peng, S., Deyssenroth, M. A., Di Narzo, A. F., Lambertini, L., Marsit, C. J., Chen, J., and Hao, K. (2017). Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. *Human molecular genetics*, *26*(17), 3432–3441.

107. Delahaye, F., Do, C., Kong, Y., Ashkar, R., Salas, M., Tycko, B., Wapner, R., and Hughes, F. (2018). Genetic variants influence on the placenta regulatory landscape. *PLoS genetics*, *14*(11), e1007785.

108. *MGI-Mouse Genome Informatics-The international database resource for the laboratory mouse*. Available at: https://www.informatics.jax.org/ (20.05.2023)

109. IMPC, International Mouse Phenotyping Consortium. (2023, May 9). *Home | IMPC | International Mouse Phenotyping Consortium*. Available at: https://www.mousephenotype.org/ (20.05.2023)

110. *CTG-VL Complex Traits Genetics Virtual Lab*. Available at: https://vl.genoma.io/ (20.05.2023)

111. *Gene Ontology Resource*. Gene Ontology Resource. Available at: http://geneontology.org/ (20.05.2023)

112. *KEGG PATHWAY Database*. Available at: https://www.genome.jp/kegg/pathway.html (20.05.2023)

113. Croft, D. B., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M. S., Garapati, P. V., Gopinath, G. R., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S. S., May, B. L., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C. K., Birney, E., ... Stein, L. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research*, *39*(Database), D691–D697.

114. Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A., and Richardson, J. E. (2014). The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Research*, *42*(D1), D810–D817.

115. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A. G., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550.

116. Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, ... and Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, *47*(3), 291–295.

117. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... and Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, *47*(11), 1236–1241.

118. Kanai, M., Akiyama, M., Takahashi, A., Matoba, N., Momozawa, Y., Ikeda, M., ... and Kamatani, Y. (2018). Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature genetics*, *50*(3), 390–400.

119. Lee, S. H., Wray, N. R., Goddard, M. E., and Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *American journal of human genetics*, *88*(3), 294–305.

120. Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., Nguyen-Viet, T. A., Wedow, R., Zacher, M., Furlotte, N. A., 23andMe Research Team, Social Science Genetic Association Consortium, Magnusson, P., Oskarsson, S., Johannesson, M., Visscher, P. M., Laibson, D., Cesarini, D., Neale, B. M., and Benjamin, D. J. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics*, *50*(2), 229–237.

121. Gualdo, N. P., Mägi, R., and Laisk, T. (2022). Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy. *medRxiv (Cold Spring Harbor Laboratory)*.

122. Rhodes, K., Barr, K. A., Popp, J. M., Strober, B. J., Battle, A., and Gilad, Y. (2022). Human embryoid bodies as a novel system for genomic studies of functionally diverse cell types. *Elife*, *11*, e71361.

123. Smith, G., Wood, A. W., Pell, J. P., and Hattie, J. (2011). Recurrent miscarriage is associated with a family history of ischaemic heart disease: a retrospective cohort study. *Bjog: An International Journal of Obstetrics and Gynaecology*, *118*(5), 557–563.

124. Ranthe, M. F., Diaz, L. J., Behrens, I., Bundgaard, H., Frisch, M., Melbye, M., and Boyd, H. A. (2016). Association between pregnancy losses in women and risk of atherosclerotic disease in their relatives: a nationwide cohort study. *European Heart Journal*, *37*(11), 900–907.

125. Yeung, E., Park, H., Nobles, C. J., Kannan, K., Silver, R. M., and Schisterman, E. F. (2019). Cardiovascular disease family history and risk of pregnancy loss. *Annals of Epidemiology*, *34*, 40–44.

126. Tong, V. T., Jones, J. R., Dietz, P. M., D'Angelo, D., Bombard, J. M., and Centers for Disease Control and Prevention (CDC) (2009). Trends in smoking before, during, and after pregnancy - Pregnancy Risk Assessment Monitoring System (PRAMS), United States, 31 sites, 2000–2005. *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C.: 2002)*, *58*(4), 1–29.

127. Konishi, H., Takahashi, T., Kozaki, K., Yatabe, Y., Mitsudomi, T., Fujii, Y., Sugiura, T., Matsuda, H., Takahashi, T., and Takahashi, T. (1998). Detailed deletion mapping suggests the involvement of a tumor suppressor gene at 17p13.3, distal to p53, in the pathogenesis of lung cancers. *Oncogene*, *17*(16), 2095–2100.

128. Phillips, N., Ziegler, M., and Deaven, L. L. (1996). A cDNA from the ovarian cancer critical region of deletion on chromosome 17p13.3. *Cancer Letters*, *102*(1–2), 85–90.

129. Phelan, C. M., Borg, A., Cuny, M., Crichton, D. N., Baldersson, T., Andersen, T. I., Caligo, M. A., Lidereau, R., Lindblom, A., Seitz, S., Kelsell, D., Hamann, U., Rio, P., Thorlacius, S., Papp, J., Olah, E., Ponder, B., Bignon, Y. J., Scherneck, S., Barkardottir, R., … Larsson, C. (1998). Consortium study on 1280 breast carcinomas: allelic loss on chromosome 17 targets subregions associated with family history and clinical parameters. *Cancer research*, *58*(5), 1004–1012.

130. Chawla, R., & Azzalin, C. M. (2008). The telomeric transcriptome and SMG proteins at the crossroads. *Cytogenetic and genome research*, *122*(3–4), 194–201.

131. Boehm, V., Kueckelmann, S., Gerbracht, J. V., Kallabis, S., Britto-Borges, T., Altmüller, J., Krüger, M., Dieterich, C., and Gehring, N. H. (2021). SMG5–SMG7 authorize nonsense-mediated mRNA decay by enabling SMG6 endonucleolytic activity. *Nature Communications*, *12*(1).

132. McIlwain, D. R., Pan, Q., Reilly, P. T., Elia, A. J., McCracken, S., Wakeham, A. C., Itie-Youten, A., Blencowe, B. J., and Mak, T. W. (2010). Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12186–12191.

133. Medghalchi, S. M., Frischmeyer, P. A., Mendell, J. T., Kelly, A. G., Lawler, A. M., and Dietz, H. C. (2001). Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Human molecular genetics*, *10*(2), 99–105.

134. Li, T., Shi, Y., Wang, P., Guachalla, L. M., Sun, B., Joerss, T., Chen, Y. S., Groth, M., Krueger, A., Platzer, M., Yang, Y. G., Rudolph, K. L., and Wang, Z. Q. (2015). Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *The EMBO journal*, *34*(12), 1630–1647.

135. Shaheen, R., Anazi, S., Ben-Omran, T., Seidahmed, M. Z., Caddle, L. B., Palmer, K., Ali, R., Alshidi, T., Hagos, S., Goodwin, L., Hashem, M., Wakil, S. M., Abouelhoda, M., Colak, D., Murray, S. A., and Alkuraya, F. S. (2016). Mutations in SMG9, Encoding an Essential Component of Nonsense-Mediated Decay Machinery, Cause a Multiple Congenital Anomaly Syndrome in Humans and Mice. *American journal of human genetics*, *98*(4), 643–652.

136. Valenta, T., Lukas, J., Doubravska, L., Fafilek, B., and Korinek, V. (2006). HIC1 attenuates Wnt signaling by recruitment of TCF-4 and beta-catenin to the nuclear bodies. *The EMBO journal*, *25*(11), 2326–2337.

137. MacDonald, B. T., Tamai, K., and He, X. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Developmental cell*, *17*(1), 9–26.

138. Hassan, H. M., Kolendowski, B., Isovic, M., Bose, K., Dranse, H. J., Sampaio, A. V., Underhill, T. M., and Torchia, J. (2017). Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex. *Cell reports*, *19*(8), 1685–1697.

139. Ubaid Ullah, Andrabi, S. B. A., Tripathi, S. K., Dirasantha, O., Kanduri, K., Rautio, S., Gross, C. C., Lehtimäki, S., Bala, K., Tuomisto, J., Bhatia, U., Chakroborty, D., Elo, L. L., Lähdesmäki, H., Wiendl, H., Rasool, O., and Lahesmaa, R. (2018). Transcriptional Repressor HIC1 Contributes to Suppressive Function of Human Induced Regulatory T Cells. *Cell reports*, *22*(8), 2094–2106.

140. Hashimoto, A., Kumashiro, S., Nishikawa, T., Oka, T., Takahashi, K., Mito, T., Takashima, S., Doi, N., Mizutani, Y., and Yamazaki, T. (1993). Embryonic development and postnatal changes in free D-aspartate and D-serine in the human prefrontal cortex. *Journal of neurochemistry*, *61*(1), 348–351.

141. Van Horn, M. R., Sild, M., and Ruthazer, E. S. (2013). D-serine as a gliotransmitter and its roles in brain development and disease. *Frontiers in cellular neuroscience*, *7*, 39.

142. Chen, Z., Huang, W., Srinivas, S. R., Jones, C. R., Ganapathy, V., and Prasad, P. D. (2004). Serine racemase and D-serine transport in human placenta and evidence for a transplacental gradient for D-serine in humans. *Journal of the Society for Gynecologic Investigation*, *11*(5), 294–303.

143. Grichtchenko, I. I., Choi, I., Zhong, X., Bray-Ward, P., Russell, J. M., and Boron, W. F. (2001). Cloning, Characterization, and Chromosomal Mapping of a Human Electroneutral Na^+-driven Cl-HCO_3 Exchanger. *Journal of Biological Chemistry*, *276*(11), 8358–8363.

144. Leviel, F., Hübner, C. A., Houillier, P., Morla, L., El Moghrabi, S., Brideau, G., Hassan, H., Parker, M. D., Kurth, I., Kougioumtzes, A., Sinning, A., Pech, V., Riemondy, K. A., Miller, R. L., Hummler, E., Shull, G. E., Aronson, P. S., Doucet, A., Wall, S. M., Chambrey, R., … Eladari, D. (2010). The Na+-dependent chloride-bicarbonate exchanger SLC4A8 mediates an electroneutral Na+ reabsorption process in the renal cortical collecting ducts of mice. *The Journal of clinical investigation*, *120*(5), 1627–1635.

145. Park, J. H., Nishidate, T., Kijima, K., Ohashi, T., Takegawa, K., Fujikane, T., Hirata, K., Nakamura, Y., and Katagiri, T. (2010). Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer research*, *70*(7), 2759–2769.

146. Berois, N., Mazal, D., Ubillos, L., Trajtenberg, F., Nicolas, A., Sastre-Garau, X., Magdelenat, H., and Osinaga, E. (2006). UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. *The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society*, *54*(3), 317–328.

147. Gomes, J., Marcos, N. T., Berois, N., Osinaga, E., Magalhães, A., Pinto-de-Sousa, J., Almeida, R., Gärtner, F., and Reis, C. A. (2009). Expression of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 in gastric mucosa, intestinal metaplasia, and gastric carcinoma. *The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society*, *57*(1), 79–86.

148. Kitada, S., Yamada, S., Kuma, A., Ouchi, S., Tasaki, T., Nabeshima, A., Noguchi, H., Wang, K. Y., Shimajiri, S., Nakano, R., Izumi, H., Kohno, K., Matsumoto, T., and Sasaguri, Y. (2013). Polypeptide N-acetylgalactosaminyl transferase 3 independently predicts high-grade tumours and poor prognosis in patients with renal cell carcinomas. *British journal of cancer*, *109*(2), 472–481.

149. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999). Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *The Journal of biological chemistry*, *274*(36), 25362–25370.

150. Morrison, J. C., Allbert, J. R., McLaughlin, B. N., Whitworth, N. S., Roberts, W. E., and Martin, R. W. (1993). Oncofetal fibronectin in patients with false labor as a predictor of preterm delivery. *American journal of obstetrics and gynecology*, *168*(2), 538–542.

151. Tarrant, J. M., Groom, J., Metcalf, D., Li, R., Borobokas, B., Wright, M. D., Tarlinton, D., and Robb, L. (2002). The absence of Tssc6, a member of the tetraspanin superfamily, does not affect lymphoid development but enhances in vitro T-cell proliferative responses. *Molecular and cellular biology*, *22*(14), 5006–5018.

152. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., van der Sluis, S., Andreassen, O. A., Neale, B. M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics*, *51*(9), 1339–1348.

153. Institute of Medicine (US) Committee on Ethical and Legal Issues Relating to the Inclusion of Women in Clinical Studies, Mastroianni, A. C., Faden, R., and Federman, D. (Eds.). (1994). *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies*. National Academies Press (US).

Appendix

I. Supplementary Materials

Table S1. Selected Diagnoses and Corresponding ICD-10 Codes. The exact diagnosis codes used in analyses are marked in bold.

ICD-10 Code	Category and Diagnosis Name
	Pregnancy with abortive outcome
O01	**Hydatidiform mole**
O01.0	Classical hydatidiform mole
O01.1	Incomplete and partial hydatidiform mole
O01.9	Hydatidiform mole, unspecified
O02	**Other abnormal products of conception**
O02.0	Blighted ovum and nonhydatidiform mole
O02.1	Missed abortion
O02.8	Other specified abnormal products of conception
O02.9	Abnormal product of conception, unspecified
O03	**Spontaneous abortion**
	Other maternal disorders predominantly related to pregnancy
O20	**Haemorrhage in early pregnancy**
O20.0	Threatened abortion
O20.8	Other haemorrhage in early pregnancy
O20.9	Haemorrhage in early pregnancy, unspecified

	Maternal care related to the fetus and amniotic cavity and possible delivery problems
O42	**Premature rupture of membranes**
O42.0	Premature rupture of membranes, onset of labour within 24 hours
O42.1	Premature rupture of membranes, onset of labour after 24 hours
O42.2	Premature rupture of membranes, labour delayed by therapy
O42.9	Premature rupture of membranes, unspecified
O44	**Placenta praevia**
O44.0	Placenta praevia specified as without haemorrhage
O44.1	Placenta praevia with haemorrhage
O45	**Premature separation of placenta (abruptio placentae)**
O45.0	Premature separation of placenta with coagulation defect
O45.8	Other premature separation of placenta
O45.9	Premature separation of placenta, unspecified

Table S2. Median and Range of Age at Agreement Group of Cases and Controls Used in EstBB Cohort-Level GWAS.

ICD-10 Code	Cases (age at agreement)		Controls (age at agreement)	
	median	range	median	range
O01	(36)	(19–58)	(44)	(18–105)
O02	(36)	(18–64)	(45)	(18–105)

O03	(37)	(18–79)	(44)	(18–105)
O20	(35)	(18–75)	(45)	(18–105)
O42	(35)	(18–67)	(45)	(18–105)
O44	(36)	(18–82)	(44)	(18–105)
O45	(36)	(18–69)	(44)	(18–105)

Table S3. SNP-Heritability (h^2_{SNP}) Estimates for Diagnoses Other Abnormal Products of Conception (O02), Spontaneous Abortion (O03), Haemorrhage in Early Pregnancy (O20) and Premature Rupture of Membranes (O42). Table contains SNP-Heritability on the observed scale and standard errors outputted by LDSC v1.0.1 tool [116,117], calculated proportion of cases in meta-analysis (n_{cases} / (n_{cases} + $n_{controls}$)), population and sample prevalence used in calculation for SNP-heritability conversion by Lee *et al*., 2011 [119] and converted SNP-heritability on the liability scale and standard errors. Estimates for three diagnoses ("hydatidiform mole" (O01), "placenta previa" (O44) and "premature separation of placenta (abruptio placentae)" (O45)) were excluded due to the insufficient sample size and unreliable estimates.

ICD-10 Code	h^2_{SNP} (observed)	Standard Error (observed)	Proportion of Cases in Meta-Analysis	Population/Sample Prevalence	h^2_{SNP} (liability)	Standard Error (liability)
O02	0.0043	0.0018	0.06	0.06	0.02	0.007
O03	0.0038	0.0018	0.07	0.07	0.01	0.007
O20	0.0089	0.0019	0.05	0.05	0.04	0.008
O42	0.0046	0.0019	0.05	0.05	0.02	0.009

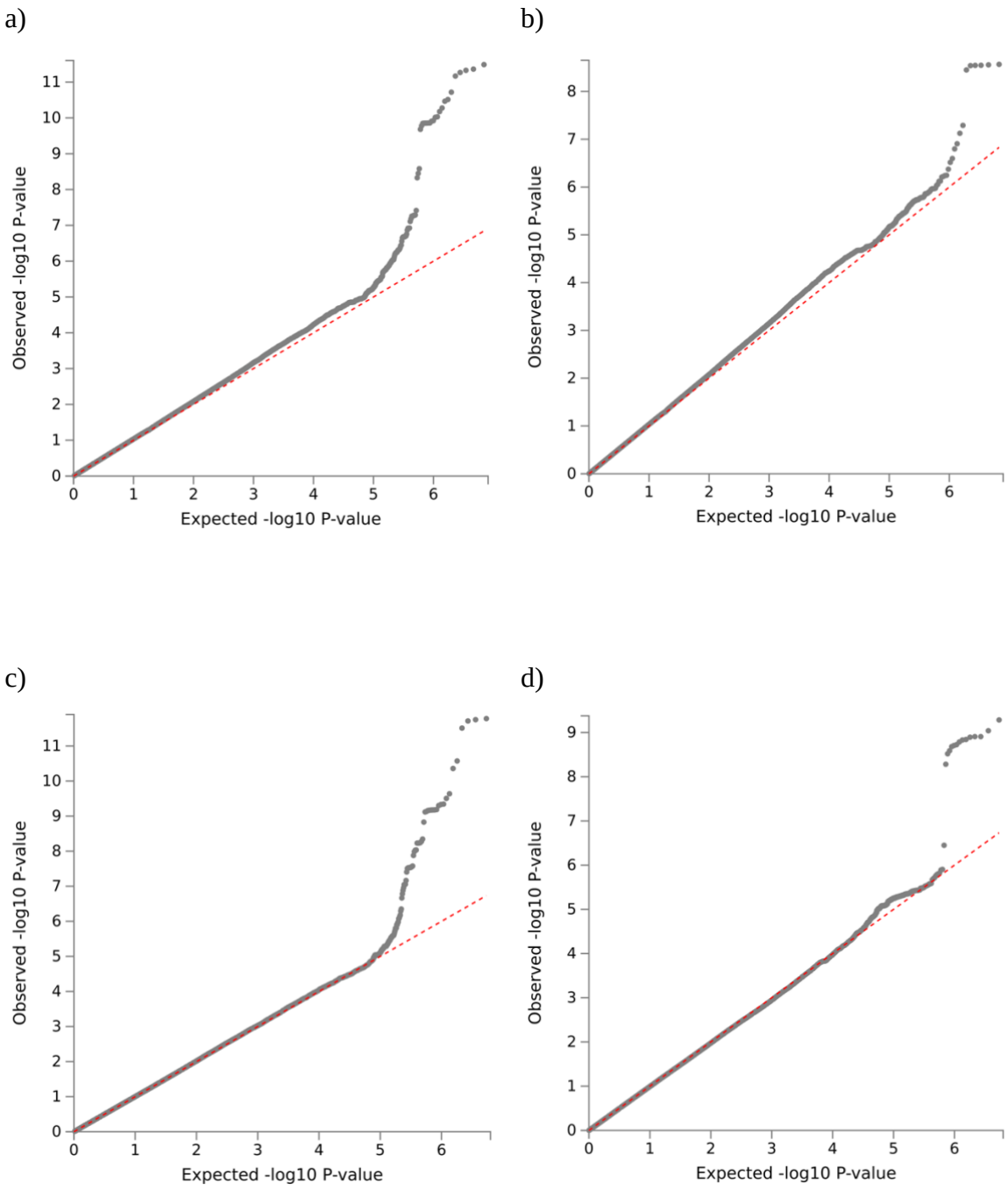


Figure S1. QQ Plots of GWAS Summary Statistics Made in FUMA of EstBB Cohort-Level GWAS for Diagnosis Haemorrhage in Early Pregnancy Assuming a) an Additive Genetic Model and b) a Recessive Genetic Model, and of EstBB and FinnGen GWAS Meta-Analysis for Diagnoses c) Haemorrhage in Early Pregnancy, and d) Premature Rupture of Membranes. The plots depict distribution of –log10 transformed observed p-values versus –log10 transformed expected p-values under a theoretical null model (no significant association). The red dashed line

follows the expected null distribution, while the deviation from the line indicates true genetic associations.

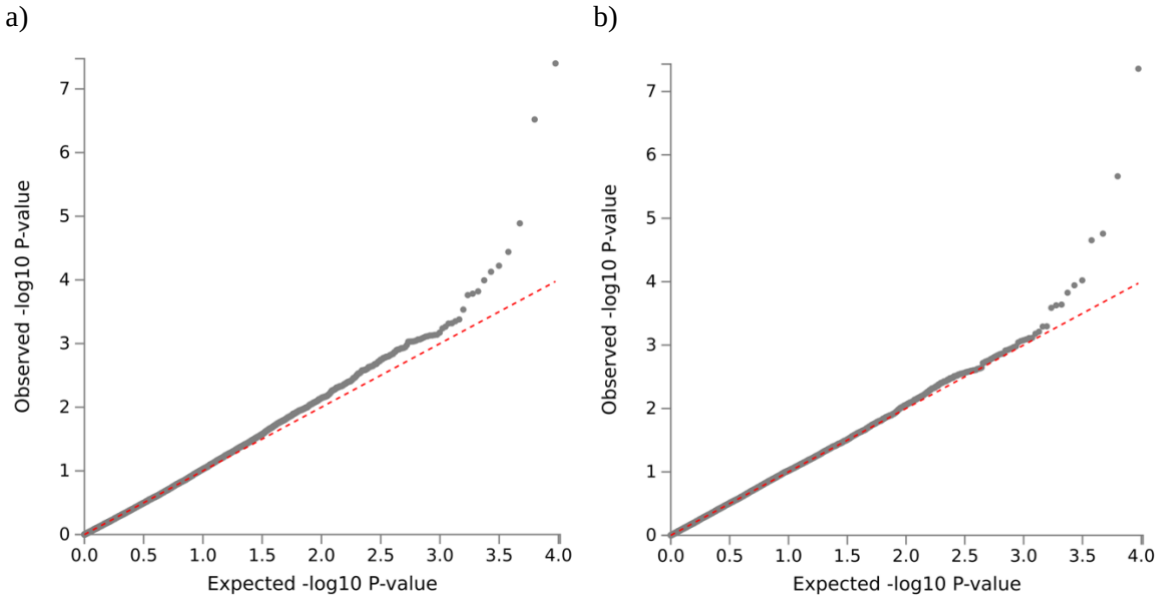
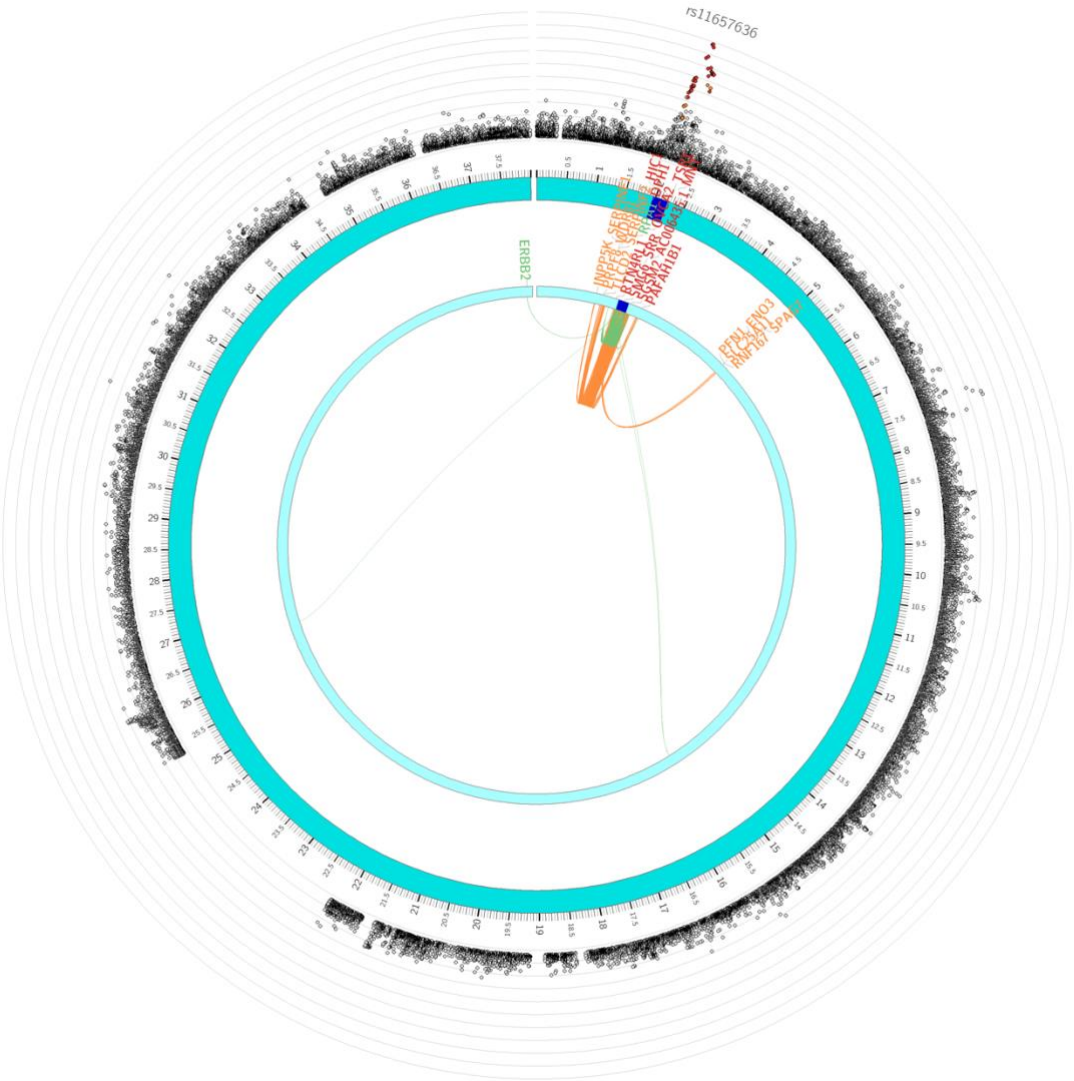
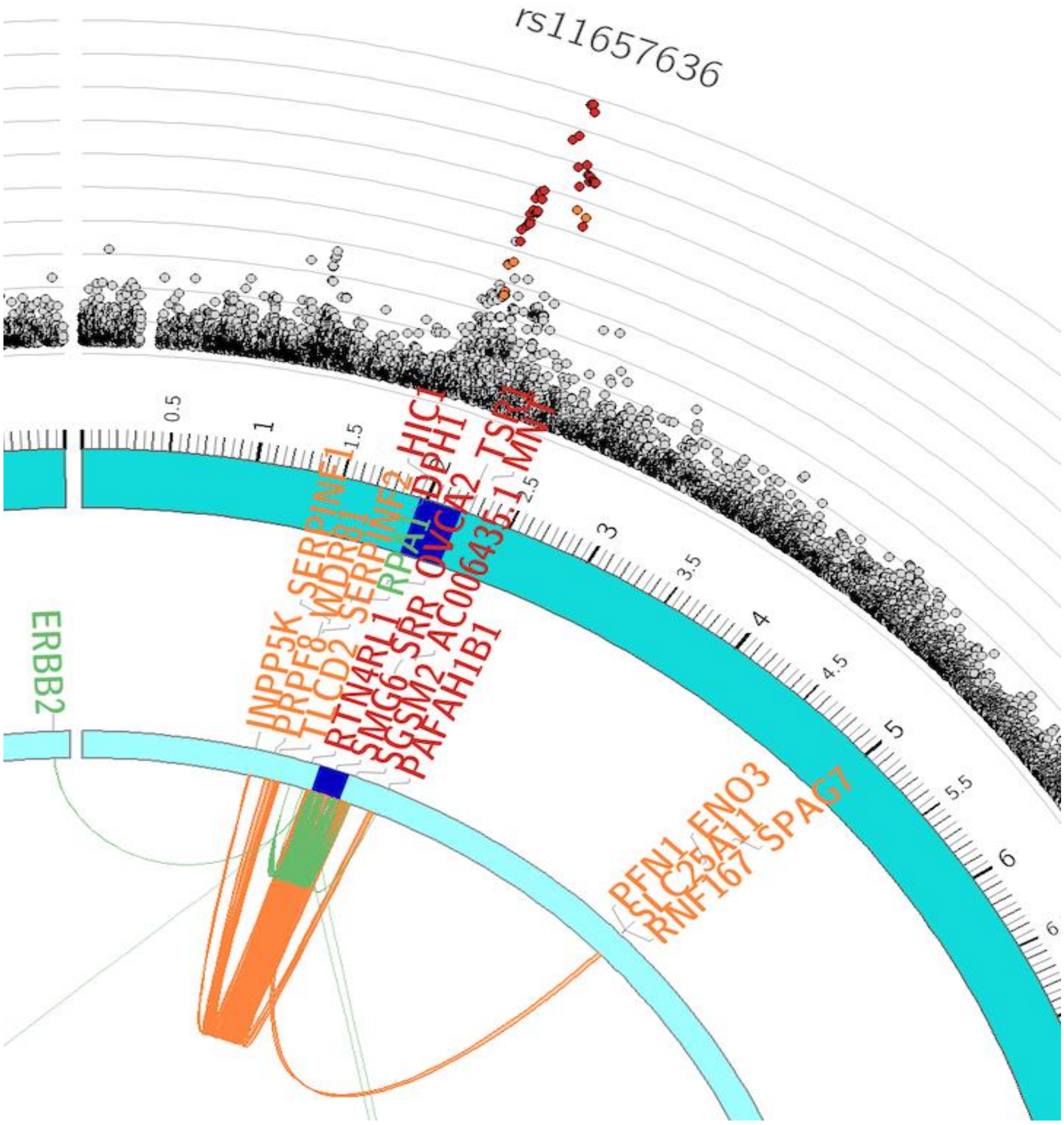


Figure S2. QQ Plots of MAGMA Gene-Based Test Made in FUMA of EstBB and FinnGen GWAS Meta-Analysis for Diagnoses a) Haemorrhage in Early Pregnancy, and b) Premature Rupture of Membranes. The plots depict distribution of –log10 transformed observed p-values versus –log10 transformed expected p-values under a theoretical null model (no significant association). The red dashed line follows the expected null distribution, while the deviation from the line indicates true genetic associations.

a)




rs11657636
0.5
1
1.5
2.5
3
3.5
4
4.5
5
5.5
6
ERBB2
INPP5K
SERPINF1
PRPF8
WDR81
TLCD2
SERPINF2
RPA1
RTN4RL1
SMG6
SRR
SGSM2
PAFAH1B1
HIC1
DPH1
OVCA2
TSR1
MNT
PFN1
ENO3
SLC25A11
RNF167
SPAG7

b)

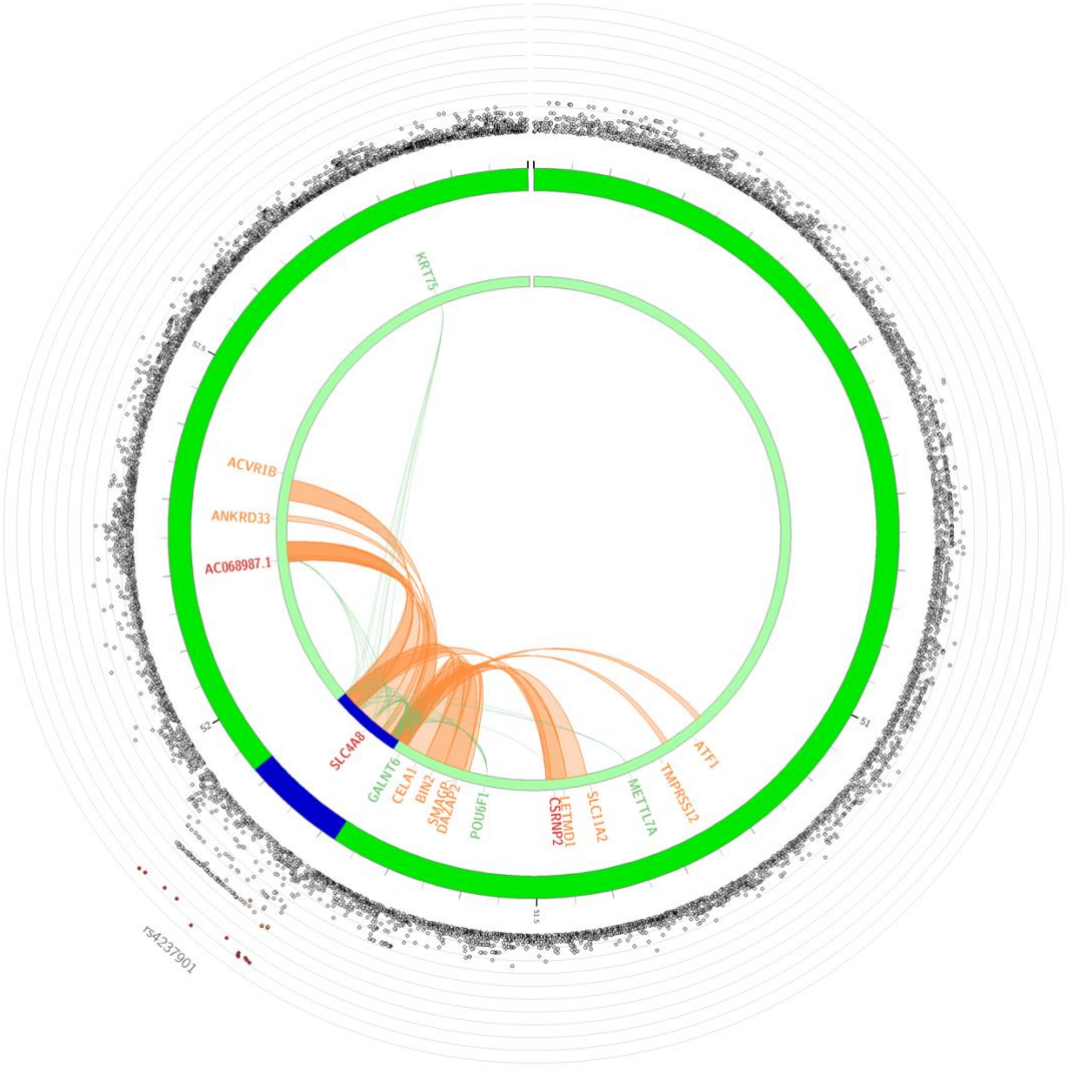
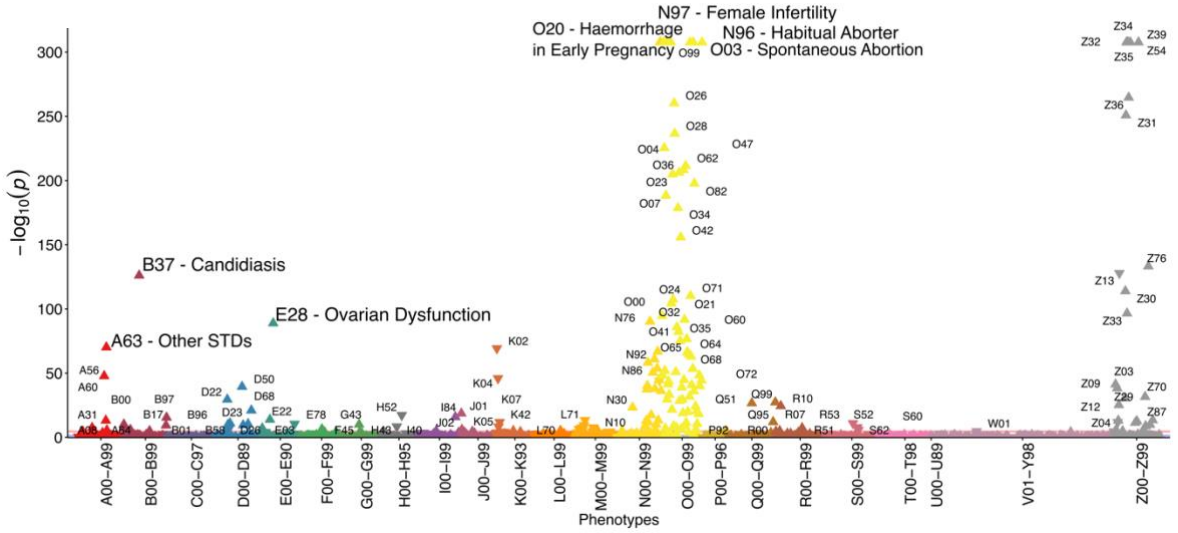


Figure S3. Circos Plots Made in FUMA of Mapped Genes by eQTLs and Chromatin Interactions at Risk Locus in Diagnosis a) Haemorrhage in Early Pregnancy and b) Premature Rupture of Membranes. The most outer ring of the plot is the Manhattan plot with labelled rsID of the lead SNP, while the other SNPs (p-value < 0.05) are as dots coloured by their maximum r^2 value to the one of the independent significant SNPs (red dot – $r^2 > 0.8$, orange dot – $r^2 > 0.6$, green dot – $r^2 > 0.4$ and blue dot – $r^2 > 0.2$; grey dot – $r^2 <= 0.2$). Middle and inner ring are representation of chromosome, with only difference of middle ring having chromosome coordinates. The area marked in dark blue is genomic risk loci. Genes mapped by eQTLs, chromatin interactions and both are coloured green, orange, and red, respectively.

a)



b)

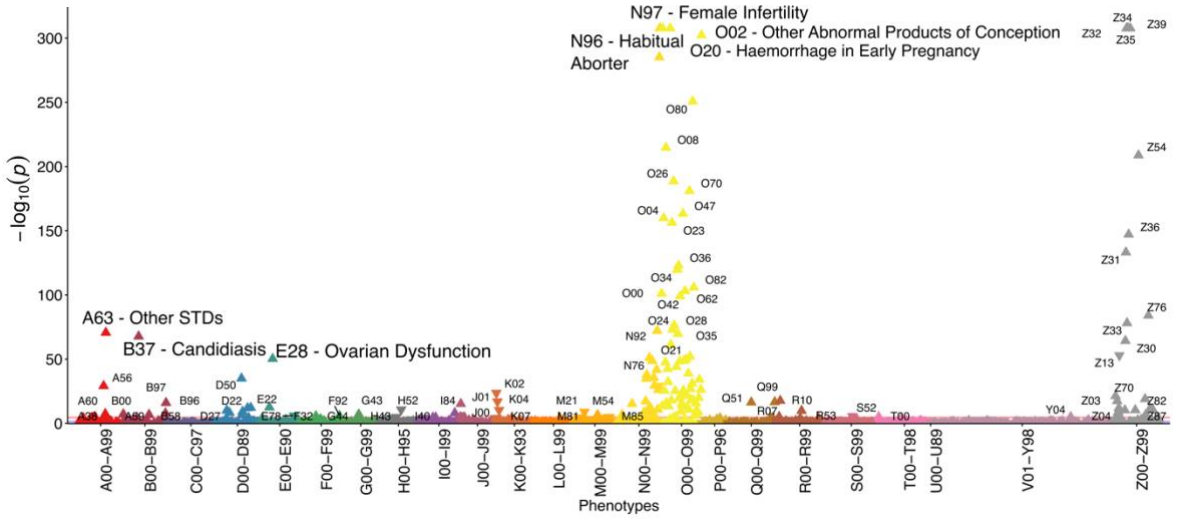


Figure S4. Associated Phenotypes in Cases Diagnosed with a) Other Abnormal Products of Conception (O02) and b) Spontaneous Abortion (O03). The pheWAS plot illustrates the p-values of logistic regression phenotype association testing of the other diagnoses (*e.g.*, the triangles' orientation (upwards/downwards) indicates whether the cases diagnosed with "other abnormal products of conception" have an increased or decreased odds of having another diagnosis). On the x-axis are represented phenotypes/diagnoses categories from the ICD-10 system color-coded by the chapters, while on the y-axis is the –log10 p-value level of statistical significance. Each triangle is one diagnosis code. The red line marks the Bonferroni correction set at 2.5×10^{-5}. The most significantly associated phenotypes are labelled.

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I, **Jelisaveta Džigurski**,

(*author's name*)

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

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

Investigating the Genetic Basis of Obstetrical Diagnoses Related to Miscarriage and Placental Biology Using Genome-Wide Association Study Meta-Analysis,

(*title of thesis*)

supervised by Triin Laisk, Natàlia Pujol Gualdo and Reedik Mägi.

(*supervisor's name*)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Jelisaveta Džigurski

22/05/2023