DISSERTATIONES
BIOLOGICAE
UNIVERSITATIS
TARTUENSIS
431

NATÀLIA PUJOL GUALDO

Decoding genetic associations of female reproductive health traits

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
431

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS
431

NATÀLIA PUJOL GUALDO

Decoding genetic associations of female reproductive health traits

UNIVERSITY OF TARTU
Press
1632

Institute of Molecular and Cell Biology / Institute of Genomics, University of Tartu, Estonia

This dissertation is accepted for the commencement of the degree of Doctor of Philosophy in Gene Technology on May 10, 2024 by the Council of the Institute of Molecular Cell Biology, University of Tartu

Supervisors: Assoc. Professor Triin Laisk, PhD. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia.

Prof. Reedik Mägi, PhD. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia.

Prof. Terhi Piltonen, MD, PhD. Dpt. of Obstetrics and Gynecology, Research Unit of Clinical Medicine, University of Oulu, Finland

Riikka Arffman, PhD. Dpt. of Obstetrics and Gynecology, Research Unit of Clinical Medicine, University of Oulu, Finland

Reviewer: Prof. Kristiina Tambets, PhD. Institute of Genomics, University of Tartu, Tartu, Estonia.

Opponent: Ilkka Kalliala, MD, PhD, Dpt. of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Commencement: Room No. 105, 23B Riia St., Tartu, on June 5 2024, at 1:15 pm.

The publication of this dissertation is granted by the Institute of Genomics and the Institute of Molecular and Cell Biology at the University of Tartu.

This research was funded by MATER Marie Skłodowska-Curie which received funding from the EU Horizon 2020 research and innovation program under grant agreement No. 813707. GENTRANSMED, Oulu university scholarship foundation and Paulon Säätiö Foundation. Data analyses were carried out in the High-Performance Computing Center of the University of Tartu.

ISSN 1024-6479 (print)
ISBN 978-9916-27-535-1 (print)
ISSN 2806-2140 (pdf)
ISBN 978-9916-27-536-8 (pdf)

University of Tartu Press
www.tyk.ee

Als meus avis
To my grandparents

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

Publications included in the thesis

This thesis is based on the following original publications, referred to in the text by Roman numerals (Ref. I to Ref. IV):

I **Pujol-Gualdo N.**, Läll K., Lepamets M., Arffman R., Rossi H., Piltonen T., Mägi R.*, Laisk T.* Advancing our understanding of genetic risk factors and personalized strategies for pelvic organ prolapse (2022) **Nature Communications**

II Tyrmi JK*, Arffman R*, **Pujol-Gualdo N*.**,…, Piltonen T., Laisk T., Kettunen J.*, Laivuori H.* Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome (2021) **Human Reproduction** https://doi.org/10.1093/humrep/deab250

III **Pujol-Gualdo N.**, Karjalainen M., Võsa U., Arffman R., Mägi R., Ronkainen J., Laisk T.*, Piltonen T.* Circulating anti-Müllerian hormone levels in pre-menopausal women: novel genetic insights from a GWAS meta-analysis **(accepted** for publication in **Human Reproduction** (2024) preprint: https://doi.org/10.1101/2023.09.07.23295182 **)**

IV Zhu J., **Pujol-Gualdo N.**, Wittemans L., Lindgren M. Cecilia, Laisk T., Hirschhorn J.N., Chan Y. Evidence from Men for Ovary-Independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome (2021) **JCEM** https://doi.org/10.1210/clinem/dgab838

* These authors contributed equally. The publications listed above have been reprinted with the permission of the copyright owners.

My contributions to the listed publications were as follows:

Ref. I Co-designed study, performed all analyses except for colocalisation analysis, wrote the manuscript, made all figures, edited the manuscript

Ref. II Co-designed the study, Replicated the analysis on the Estonian Biobank data, carried out meta-analysis and main postGWAS analysis, co-made figures, co-wrote the manuscript, co-edited the manuscript

Ref. III Designed the study, performed analysis, wrote the manuscript, made all figures, edited the manuscript

Ref. IV Replicated the analysis on the Estonian Biobank data, participated in planning the replication analysis, co-made figures, commented and edited the manuscript

Publications not included in the thesis

V **Pujol-Gualdo N**; Estonian Biobank Research Team; Mägi R, Laisk T. Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy. Hum Reprod. 2023 Dec 4; 38(12):2516–2525. doi: 10.1093/humrep/dead217. PMID: 37877466

VI Pathare A.*, **Pujol-Gualdo N***, Rukins V., Džigurski J., Peters M., Estonian Biobank Research Team, Mägi R., Salumets A., Saare M., Laisk T. Large-scale genome-wide association study to determine the genetic underpinnings of female genital tract polyps (**preprint:** https://doi.org/10.1101/2024.01.29.24301773)

VII Koel M, Võsa U, Jõeloo M, Läll K, **Pujol-Gualdo N**,…, Mägi R, Laisk T. GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. Hum Mol Genet. 2023 Jun 5; 32(12):2103–2116. doi: 10.1093/hmg/ddad043. PMID: 36929174

VIII Bourgault J, Abner E, Manikpurage HD, **Pujol-Gualdo N**, Laisk T … , Arsenault BJ. Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis. Gastroenterology. 2023 May; 164(6):953–965.e3. doi: 10.1053/j.gastro.2023.01.028. Epub 2023 Feb 1. PMID: 36736436.

IX **Pujol-Gualdo N**, Sánchez-Mora C, Ramos-Quiroga JA, Ribasés M, Soler Artigas M. Integrating genomics and transcriptomics: Towards deciphering ADHD. Eur Neuropsychopharmacol. 2021 Mar; 44:1–13. doi: 10.1016/j.euroneuro.2021.01.002. Epub 2021 Jan 23. PMID: 33495110.

X Nieuwenhuis D, **Pujol-Gualdo N**, Arnoldussen IAC, Kiliaan AJ. Adipokines: A gear shift in puberty. Obes Rev. 2020 Jun; 21(6):e13005. doi: 10.1111/obr.13005. Epub 2020 Jan 30. PMID: 32003144

XI Arnoldussen IAC, Morrison MC, ..., **Pujol-Gualdo N**, van der Logt L, Gross G, Kleemann R, Kiliaan AJ. Milk fat globule membrane attenuates high fat diet-induced neuropathological changes in obese Ldlr-/-.Leiden mice. Int J Obes (Lond). 2022 Feb; 46(2):342–349. doi: 10.1038/s41366-021-00998-w. Epub 2021 Oct 29. PMID: 34716425.

XII Custers E, Vreeken D, Kaufmann LK, **Pujol-Gualdo N**, …, Kiliaan AJ. Cognitive Control and Weight Loss After Bariatric Surgery: the BARICO Study. Obes Surg. 2023 Sep; 33(9):2799–2807. doi: 10.1007/s11695-023-06744-7. Epub 2023 Jul 21. PMID: 37477832.

LIST OF ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
AMH	Anti-Müllerian Hormone
BMI	Body Mass Index
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTG-VL	Complex Traits Genetics Virtual Lab
EstBB	Estonian Biobank
EHRs	Electronic Health Records
eQTL	Expression Quantitative Trait Locus
GTEx	Genotype-Tissue Expression project
GWAS	Genome-Wide Association Studies
HPG	Hypothalamic-Pituitary-Gonadal axis
ICD	International Classification of Diseases
ICDA	International Common Disease Alliance
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
NFBC	Northern Finland Birth Cohorts
OR	Odds Ratio
PCs	Principal Components
PCOS	Polycystic Ovary Syndrome
POP	Pelvic Organ Prolapse
PRS	Polygenic Risk Scores
scRNA	Single-cell RNA
SHBG	Sex hormone-binding globulin
SNPs	Single-Nucleotide Polymorphisms
T2D	Type 2 Diabetes
UKB	UK Biobank
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Note: Despite using woman/women and female and man/men and male terms interchangeably, we are always referring to biological sex, not gender identity, and we acknowledge that the presented biological sex (based on chromosomal make-up) does not always align with self-identified gender.

INTRODUCTION

Genetic variation, particularly single-nucleotide polymorphisms (SNPs), has shown to influence health and disease susceptibility for multiple complex diseases. This has been supported by many genome-wide association studies (GWAS) which have unravelled thousands of genetic variants in association with health traits. However, studies of genetic variation underlying female reproductive health traits remain limited, with only a small proportion of all GWAS focusing on this area. Currently, the availability of population-based biobanks, such as the Estonian Biobank, and population-based birth cohorts such as the Northern Finland Birth Cohort 1966, provide a valuable framework for studies in this field. Additionally, GWAS set the ground to move from genetic variations to potentially affected genes, proteins, biological pathways and tissues, serving as a foundation for forming hypotheses that can be validated through functional experiments. Understanding the genetic variation that is associated with a trait is important as it can provide insights into disease aetiology, prediction, and potential treatments. Another notable outcome of GWAS is the construction of polygenic risk scores (PRS), which provide a summary of an individual's genetic predisposition for a certain trait. PRS has the potential to predict disease susceptibility and serve as a tool to further explore disease biology, for example, relationships with a certain trait's comorbidities. PRS have attracted massive attention in the pursuit of the so-called personalised medicine. This thesis aims to decode the genetic underpinnings of selected female reproductive health traits through GWAS and explores PRS as a tool for both risk stratification and for informing a trait's biology. In conclusion, this research, placed at the intersection of genomics and female reproductive health, is poised to address a knowledge gap in both national and international research systems through the availability of large genomic datasets coupled with electronic health records (EHRs) and biological measurements.

In the first part of the thesis, I will touch on the motivation for studying women's health, introduce the three female reproductive health traits studied and the technological framework of the studies. Then, I will review the main genetic epidemiology tools we have used to study the genetic associations underlying different female reproductive health traits (GWAS and PRS), and introduce the state-of-the-science in the field of female reproductive genomics. In the second part, firstly I will describe the novel genetic findings and their characterisation for traits such as pelvic organ prolapse (POP) and anti-Müllerian hormone (AMH) levels, illustrating how larger sample sizes can advance our understanding of the biology underlying those traits. Secondly, I will highlight the value of population-specific initiatives with a unique demographic history for detecting population-enriched alleles in association with polycystic ovary syndrome (PCOS) and AMH levels. Thirdly, I will explore the use of PRS as a tool to both stratify and predict disease risk in POP and as a tool to inspect PCOS biology and its comorbidities.

1. REVIEW OF THE LITERATURE

1.1. Studying women's health: why does it matter?

In today's rapidly changing healthcare landscape, it is crucial to recognize the importance of the study on women's health and address some of the challenges that have led to its undervaluation and underrepresentation. Despite women representing half of humanity, we are often being invisibilized or treated as an atypical population by the medical and research community. Numerous factors have contributed to the neglect of research on women's health over the years, including a lack of acknowledgement and insufficient funding. Consequently, this leads us to a field which has been understudied and poorly understood.

Women face the challenge of existing within a healthcare system that was largely designed by and for men. Historically, the majority of studies, whether on humans or animals, primarily focused on males. A gender bias is evident across the full spectrum of research, spanning from basic science, translational studies, to clinical trials, where for instance women were not included until the end of the 20th century in the United States (Beery & Zucker, 2011; Merkatz, 1998; Nielsen et al., 2017; "Putting Gender on the Agenda.," 2010).

This broad overlook of women in research has led to a dangerous precedent where the medical and research community assumed they could apply male-focused views and results universally. This leads us to a world where, for example, women's heart attack symptoms are largely overlooked (Mehta et al., 2016), or drugs that are present in the market have disproportionately adverse side effects for women (Watson et al., 2019). However, various national initiatives, such as the White House Initiative on Women's Health Research and the Women's Health Interest Group from the European Institute of Women's Health, highlight the importance of understanding the underlying biology of women's health.

A similar situation applies when looking at conditions affecting the female reproductive system, falling into an area (reproductive health) which has not been typically the main dish on lunchtime conversations. The societal community's taboo and also the medical reluctance towards sexual and reproductive health issues in the last centuries has further extended this problem, discouraging their study and acknowledgement, and perpetuating a culture of silence, stigma and misunderstanding. This situation creates significant barriers for seeking trustworthy information, support, and access to necessary healthcare services.

Another significant obstacle to the study of overall reproductive health is limited research funding. For instance when looking at grant databases maintained by the Canadian Institutes of Health Research and the National Institutes of Health in the US, these show that the number of grants for research on non-reproductive organs is 6–7 times higher than the number for reproductive organs (Mercuri & Cox, 2022). This problem extends beyond the reproductive domain, since a gender-biased funding disparity has also been reported when looking at funding directed to the study of conditions that affect more women than men,

which receive significantly less funding compared to the reverse situation (*Women's Health Research Lacks Funding – These Charts Show How*, n.d.).

These research gaps and disparities in funding are likely to result in more significant and severe health issues. For instance, when women present with symptoms of disease, delays in diagnosis are the rule rather than an exception, for instance for endometriosis, where a long diagnostic delay after symptom onset is common (Horne & Missmer, 2022; *One in Three Women with Female Health Conditions Forced to Wait Three Years for Diagnosis | The Independent*, n.d.), or POP, where a lack of awareness seems to be perpetuated by the stigma and embarrassment around the symptoms, which may prevent seeking medical help. At the same time, early diagnosis opportunities and referral to specialist services are often missed by the general practitioners according to women's experiences when receiving care for POP (Abhyankar et al., 2019). Until the study of women's health and reproductive health is made a priority, their health outcomes will continue to be poorly understood and mistreated, imposing a substantial health burden. With women constituting two thirds of the worldwide health and social care workforce (*Value Gender and Equity in the Global Health Workforce*, n.d.), their well-being directly impacts the prosperity of communities at large.

In this thesis, I hope to illustrate how progress can be made when women's reproductive health traits are brought to the forefront, having a special focus on the role that genetics might have in shaping susceptibility to different reproductive traits in women. Genetics offers a valuable tool to discover the causal biological mechanisms underlying these conditions. The advancements in genetic technologies and genetic epidemiology methods from the last decades enable a good opportunity for the study of genetics underlying women's reproductive health traits, facilitating the discovery and characterisation of affected variants, genes, mechanisms and drug targets. However, I acknowledge genetics is only an incomplete part of the picture and there is a long way and many questions still ahead to cover. It is important to note that this work represents just a partial and limited view of the broader spectrum of women's health concerns.

The path to change those lines however is to start – even with small actions – recognising, talking about, and researching those.

1.2. Female reproductive health: beyond fertility

According to the World Health Organisation, reproductive health is '*a state of complete physical, mental and social well-being in all matters relating to the reproductive system. It implies that people are able to have a satisfying and safe sex life, the capability to reproduce and the freedom to decide if, when, and how often to do so*' (*Reproductive Health*, n.d.).

From that perspective and definition, the primary concerns related to reproductive health are those affecting an individual's sexual life, reproductive capacity, and freedom and autonomy to make choices in this regard, yet these are not the only problems stemming from reproductive health dysfunction. In the case

of women, there is a wide-range of female reproductive health outcomes across the lifespan and matters affecting the reproductive system extend beyond their impact on fertility and the events constraining the reproductive lifespan, with an extended association with various health domains such as metabolic traits (Nichols et al., 2024) and mental health (Zaks et al., 2023). Several conditions originating in the reproductive tract significantly affect women's quality of life and can develop at various stages throughout their lives independently of the pregnancy status and/or choices in this regard.

Traditionally, 'reproductive lifespan' is defined from the time of onset of puberty (known as the start of reproductive maturity with onset of menstruations, or menarche, as a main hallmark) until the menopause (known as the time when the pool of oocytes is depleted and menstrual cycles cease, around the age of 50–52 years on average ("Variations in Reproductive Events across Life: A Pooled Analysis of Data from 505 147 Women across 10 Countries.," 2019)). It is estimated that natural fertility in women ceases around 10 years before menopause (Lambalk et al., 2009). Beyond the context of oocyte ageing and fertility issues, other reproductive health issues may emerge during the so-called reproductive lifespan, such as PCOS (a common and multifaceted endocrinopathy), menstrual irregularities or menstrual dysfunctions such as heavy menstrual bleeding, endometriosis (a chronic painful inflammatory condition where the endometrial-like tissue grows outside the uterus), uterine fibroids (benign growth of the uterus), sexual dysfunctions, sexually transmitted infections, etc. Also, significant problems may arise around menopause, including pelvic floor dysfunctions such as POP, a descent of the pelvic organ into the vaginal cavity, urinary and anal incontinence, and pelvic pain. Furthermore, women may be susceptible to various reproductive cancers like cervical, ovarian or endometrial cancers, and other hormone-dependent cancers, such as breast cancer. For many of these conditions, hormones and hormonal fluctuations play a central role (Fauser et al., 2011; Moolhuijsen & Visser, 2020), which also underlines the importance of hormonal regulation in women's reproductive health.

Women's reproductive health issues affect not only women themselves. During pregnancy, issues such as ectopic pregnancy (a premature implantation of the embryo in the fallopian tube), can also arise at early stages, or in later stages, including preeclampsia (characterised by high blood pressure and the presence of proteins in urine during pregnancy), gestational diabetes (high blood sugar disorder developing during pregnancy), or even during childbirth, including, but not limited to, preterm birth. There is also an increased recognition that pregnancy complications are associated with chronic disease later in life both for the baby and the mother, such as an increased risk for cardiovascular disease or mental health disorders (Andraweera & Lassi, 2019; Lahti-Pulkkinen et al., 2020; McNestry et al., 2023; Riise et al., 2019; Täufer Cederlöf et al., 2022; Zhao & Xia, 2022).

This diverse list of traits, while not exhaustive of all female reproductive conditions, illuminates the broad spectrum of traits and varied stages throughout a

woman's lifetime where reproductive health issues can cause an impact. Consequently, female reproductive health constitutes a multifaceted and complex domain, influenced by a combination of environmental, lifestyle, and genetic risk factors. This multifactorial nature forms the foundational basis for the current thesis. Here, we study three common heritable traits that significantly influence women's reproductive health in different stages of life as examples to illustrate how genomic epidemiology can be used to study underlying biology and develop potential risk stratification tools (Figure 1).

1.2.1. Female reproductive conditions used as examples

The three traits studied include two prevalent conditions, POP and PCOS, alongside with the exploration of genetic factors influencing variations AMH levels in pre-menopausal women (Figure 1). The subsequent section will introduce the main clinical and epidemiological characteristics of these specific phenotypes.

1.2.1.1. Pelvic Organ Prolapse (POP)

POP is characterised by a descent of pelvic organs into the vaginal cavity (M. D. Barber, 2016). The prolapse of the anterior vaginal wall, or cystocele, is the most common form of POP, followed by rectocele (posterior vaginal prolapse) and uterine prolapse (M. D. Barber, 2016). Overall, POP affects around 40% of women after menopause (Hendrix et al., 2002; Kirby et al., 2013; Nygaard et al., 2004) and it is estimated that around 3–6% of those present a descent of the organs beyond the vaginal opening, which is considered the most severe form of POP (M. D. Barber, 2016). Notably, prolapse is the most common indication for hysterectomy (a surgical procedure where the uterus is removed) in postmenopausal women. The lifetime risk of gynaecological surgery for POP is up to 19% in the general female population (Nygaard et al., 2004).

The main symptoms of prolapse include a bothersome sense of vaginal bulge, urinary, bowel, and/or sexual dysfunction, which substantially affect a woman's quality of life (Ellerkmann et al., 2001; Jelovsek & Barber, 2006). Common risk factors are age, number of children, operative vaginal delivery, and body mass index (BMI) (Blomquist et al., 2018; Giri et al., 2017; Vergeldt et al., 2015). However, despite its health and economic impact, the aetiology of this complex disorder remains poorly understood, although lifestyle, environmental and genetic risk factors likely play a role. Genetic factors have been estimated to explain 43% of the variation in POP risk (Altman et al., 2008), based on comparing mono- and dizygotic female twins. Only recently has there been a large-scale assessment of genetic susceptibility to POP (Olafsdottir et al., 2020), pointing towards connective tissue metabolism and estrogen exposure in its aetiology, and marking the beginning of the genome-wide association study era for this trait. This study estimated the SNP heritability (SNP-h^2, that is the proportion of the phenotypic variance explained by SNPs) in the meta-analysis to be 12.4% (95% CI 9.9–

14.8%). While the first study into the genetic underpinnings of POP was encouraging, it was still only the first attempt with limited sample size, leaving the possibility open that many more genetic associations are yet to be revealed.

Despite there being evidence POP symptomatology aggravation could in some cases be prevented with timely interventions, such as implementing lifestyle changes and pelvic floor muscle training (Abhyankar et al., 2019; Hagen et al., 2017), currently, the identification of women with increased risk of POP cannot be accurately achieved by any clinical, imaging and/or manual exploration tests performed in the gynaecological practice. Additionally, these tests are recording changes in structural anatomy that have already taken place, disabling the long-term prediction and potential prevention of POP. Thus, there is lack of evidence and means for early detection of women who are at risk of developing POP, potentially useful information for preventing symptomatology aggravation and reducing the need for surgical treatment.

1.2.1.2. Polycystic Ovary Syndrome (PCOS)

PCOS is one of the most common endocrinopathies affecting reproductive aged women, with impacts across the lifespan from adolescence to post menopause (Teede et al., 2023). PCOS is defined by a combination of signs and symptoms of androgen excess and ovarian dysfunction. The latest international evidence-based guideline estimates PCOS prevalence to be around 10–13% (Teede et al., 2023) and recommends using the Rotterdam criteria for PCOS diagnosis, requiring the presence of at least two of the following symptoms: oligo- or anovulation (irregular or absent menstruation), clinical or biochemical hyperandrogenism, and/or polycystic ovaries either seen in ultrasound or inferred by AMH measurements, after exclusion of related disorders (Teede et al., 2023).

PCOS is the most common cause for anovulatory infertility, caused by disrupted follicle development owing to dysregulation of the Hypothalamic–Pituitary-Gonadal (HPG) axis. This results in follicle arrest and an increase in the number of antral follicles in the ovaries, as well as a 2- to 3-fold increase in levels of AMH (Silva & Giacobini, 2021).

While ovulatory dysfunction in women with PCOS often subsides with age, they might still display higher AMH and later onset of menopause (de Ziegler et al., 2018; Forslund et al., 2019; Minooee et al., 2018; Piltonen et al., 2005; Tyrmi et al., 2022). In addition to the reproductive features, PCOS is also characterised by metabolic disturbances such as obesity, insulin resistance and dyslipidemia (T. M. Barber & Franks, 2021; S. S. Lim et al., 2019; Ollila et al., 2016), as well as a higher risk for Type 2 Diabetes (T2D) (Rubin et al., 2017) and depression (Hollinrake et al., 2007). Women with PCOS also have an increased risk for endometrial cancer and a recent study suggested an increased risk of ovarian cancer among postmenopausal women with PCOS (Frandsen et al., 2023), although the majority of studies do not indicate a higher susceptibility to other types of cancer (Barry et al., 2014; Ding et al., 2018; Dumesic & Lobo, 2013; Gottschau et al., 2015).

Despite the high prevalence of the syndrome, the aetiology of this syndrome remains largely unknown, a challenging study considering the high heterogeneity of the condition. Similarly, a clear view of which factors and how those factors influence AMH levels remains to be determined. Considering the complex nature of PCOS, it is likely that both genetic and environmental factors contribute to its development (Abbott et al., 2019; Koivuaho et al., 2019; Moghetti & Tosi, 2021). Notably, the heritability of PCOS is estimated to be around 70% (Risal et al., 2019; Vink et al., 2006). In 2018, the first large genetic study for PCOS was published (F. Day et al., 2018), comparing genome-wide profiles from around 10,000 women with PCOS to around 100,000 female controls without the diagnosis. Three main findings arose from that study: firstly, it clearly supported a shared genetic architecture across diverse diagnostic criteria for PCOS; secondly, it pointed to the identification of genetic links between PCOS and various metabolic traits, as well as male-pattern balding and thirdly, the authors detected three novel genetic signals near the genes *PLGRKT*, *ZBTB16*, and *MAPRE1* beyond other 11 previously reported genetic regions identified in smaller-scale genetic studies (Chen et al., 2011; F. R. Day, Hinds, et al., 2015; Hayes et al., 2015; Y. Shi et al., 2012). However, given the high heritable component of the condition, there is likely room for more genetic risk factors to be identified through analysing larger sample sizes, implementing newer methods or technologies, or focusing on datasets available for specific populations with unique demographic history. Those avenues may lead to the detection of more genetic risk factors of the condition and potentially uncover affected underlying biology, a step to deepen our understanding towards multifaceted health conditions such as PCOS.

1.2.1.3. Anti-Müllerian hormone (AMH)

Despite owing its name to its classical role in male sexual differentiation, in women AMH is expressed by the ovarian granulosa cells during the primary to small antral stage of follicle development (Weenen et al., 2004) (Figure 1). Starting to rise postnatally, serum AMH levels peak at mid-twenties and after this the levels decrease with age, with undetectable levels following menopause, signalling depletion of ovarian reserve (Finkelstei et al., 2020) (Figure 1). As a result, AMH is primarily known as a serum marker for ovarian reserve (Moolhuijsen & Visser, 2020), becoming widely utilised in infertility clinics (Tobler et al., 2015), despite little knowledge on which factors affect its variation. The generalised implementation of serum AMH measurement has also led to an increase in diagnostic assays, including automated assays, although direct comparison of results between assays remains problematic (Moolhuijsen & Visser, 2020). Furthermore, little is known about endogenous and exogenous factors that influence serum AMH levels, which also limits proper interpretation of AMH values in a clinical setting. Therefore, identifying factors that contribute to the variation in AMH levels is of significance.

Recent genetic studies in around 7,000 pre-menopausal women have estimated a heritable component in the variance of AMH (SNP-h^2 estimation of 15%

(SE = 7%), an estimate based on the contribution of SNPs only) (Verdiesen et al., 2022), highlighting genetic variation in regions in or near the genes *MCM8*, *AMH*, *TEX41* and *CDCA7* to associate with AMH levels. Given this heritable component, there is likely room to detect more genetic risk factors associating with AMH variation, which might offer valuable insights into the biology underlying other proxy phenotypes of ovarian reserve. Examples include antral follicle count (AFC), which is assessed through transvaginal ultrasound and folliculogenesis, the process by which follicles in the ovary develop and mature, involving several stages of growth and culminating in the release of a mature oocyte) (Baerwald et al., 2012; Gershon & Dekel, 2020) (Figure 1).

Disrupted follicle development originating from disrupted HPG axis, can lead to anovulatory infertility, which results in follicle arrest and an increase in the number of antral follicles in the ovaries, which associates with increased levels of AMH (Silva & Giacobini, 2021). Anovulatory infertility and/or irregular menstruation may be a hallmark of PCOS, and therefore higher AMH levels have been associated with PCOS as well (Homburg & Crawford, 2014). Additionally to PCOS, previous studies have indicated that variations in age-specific circulating AMH levels are linked with several health conditions, including an association between higher AMH levels and breast cancer (W. Ge et al., 2018) and also an association between low AMH levels and higher prevalence of autoimmune disease such as systemic lupus erythematosus (Luo et al., 2020) (Figure 1). The identification of genetic risk factors influencing AMH levels also give tools to assess potential genetic associations between AMH and other health conditions, such as breast cancer.

1.2.2. The value of genetics and genomics

Genetics and genomics have emerged as a key layer in modern scientific research, offering valuable insights into a wide range of phenotypes, including reproductive health conditions. The slow progression in the comprehension of these conditions' aetiology contrasts with the rapid evolution and increased accessibility of genotyping and sequencing technologies. Through the establishment of genomically profiled biobanks worldwide, there is now unprecedented access to vast repositories of biological samples and associated data linked to health traits, facilitating investigations into the genetic underpinnings of female reproductive phenotypes. These biobanks serve as dynamic resources, enabling the integration of molecular data with health records, such as EHRs, offering ways to study the complex interplay between genetic susceptibility and environmental factors in disease development. The collaborative nature of genomic studies, often involving the combination of data from multiple biobanks or cohorts, has notably expanded sample sizes, enhancing the robustness and statistical power for detecting genetic risk factors underlying complex traits. In the next section, I will introduce the main population-based initiatives that have catalysed such studies, offering opportunities to provide genetic insights into biological mechanisms governing female reproductive health and disease.

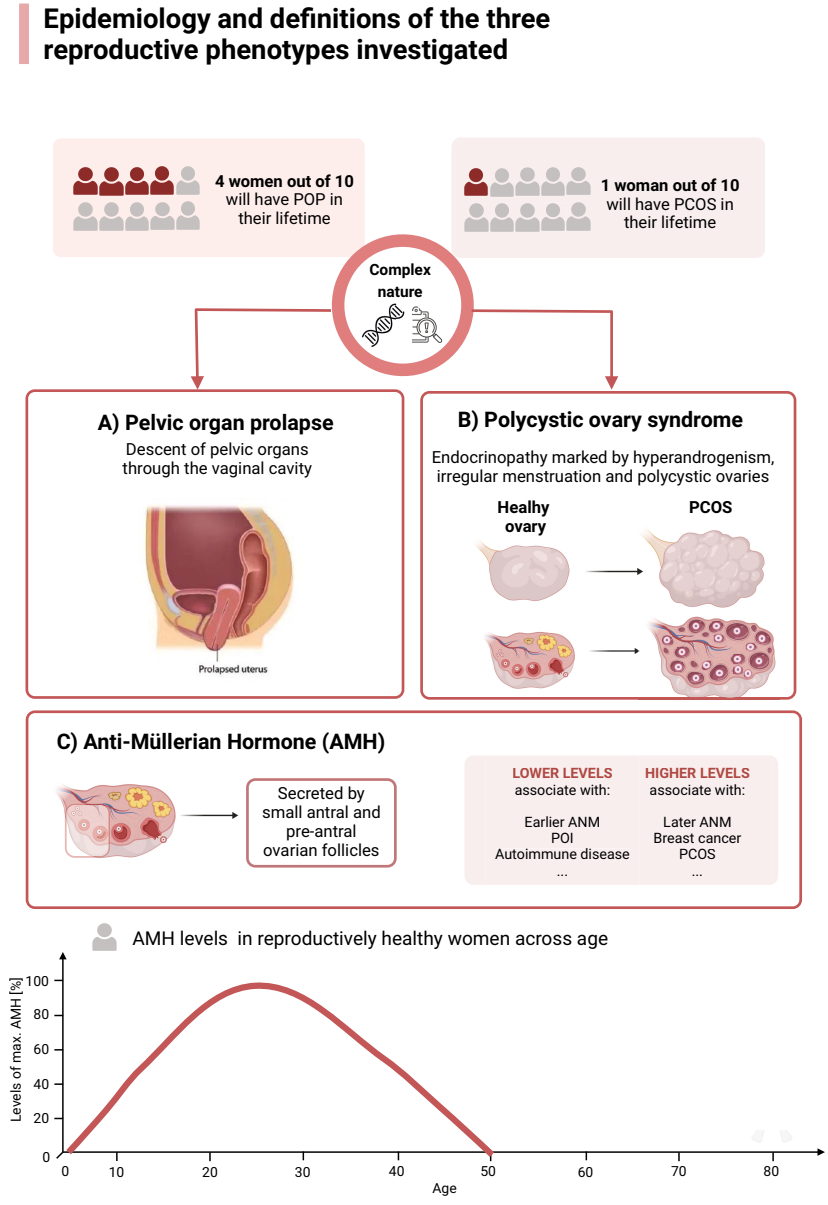

Figure 1. Clinical summary of the traits investigated (A) pelvic organ prolapse (POP) (B) polycystic ovary syndrome (PCOS), and (C) Anti-Müllerian hormone (AMH). ANM: age at natural menopause, POI: premature ovarian insufficiency. AMH levels trajectory are an approximation of average % levels in the average healthy population, note that age 50 marks the onset of menopause in this representation. Created with BioRender.com and adapted from "Human Ovary Anatomy (healthy)" and "The Effects of GnRH and Gonadotropin Secretion". Retrieved from https://app.biorender.com/biorender-templates

1.3. The value of biobanks, cohorts and electronic health records

The term 'biobank' has been used since about 1996 (Loft & Poulsen, 1996) in order to describe collections of human biological samples. In 2006, the Organisation for Economic Co-operation and Development (Sampogna, 2006) has defined a 'biobank' as 'a collection of biological material and the associated data and information stored in an organised system for a population or a large subset of a population'. Often, biobanks contain health data linked with genomic information, which provide valuable resources for the genetic research community. So far, biobanks have been central to examining the role of individual genetic susceptibility and exposures to external factors in the development of specific health conditions by combining molecular data with other associated health data. For instance, large-scale genetics analyses have been pursued leveraging data from nearly two thousand diagnoses across three large biobanks (Kurki et al., 2023).

In comparison to disease- or trait-based cohorts centred around a particular phenotype or several relevant phenotypes, biobanks enable cost-effective genetic discovery for hundreds to thousands of phenotypes, in a fast and systematic manner since these phenotypes can be usually curated from real-time longitudinal records, EHRs (Bowton et al., 2014). Other sources of information might include registry-based data (such as death or cancer registry data), epidemiological questionnaires, imaging data, laboratory test results, etc. For instance some studies presented joint analyses of genetics and imaging measurements such as neuroimaging measures (S. M. Smith et al., 2021) or optical coherence tomography measures (Currant et al., 2021, 2023), hormone measurements (Ruth et al., 2020; Venkatesh et al., 2024), circulating metabolic markers (Karjalainen et al., 2024) or drug adverse responses (Krebs et al., 2020), amongst others.

The earliest nation-wide biobank is Iceland's deCODE genetics which started in 1996, and was subsequently acquired by a company. The deCODE study has so far gathered genotypic and medical data from more than 160,000 volunteer participants, comprising well over half of the adult population in Iceland (Gulcher & Stefansson, 1999). Being instrumental in the current thesis, another of the first biobank initiatives was established in Estonia in 2001, the Estonian Biobank (EstBB). The EstBB cohort is a volunteer-based sample of the Estonian resident adult population (aged ≥18 years) hosted in the Estonian Genome Centre of the University of Tartu. The EstBB currently includes genomic and electronic health records data for over 200,000 individuals (representing around 20% of the Estonian adult population) (Leitsalu et al., 2015; Ojalo et al., 2024). Additional centres in Europe have more recently launched their biobank initiatives, subsequently becoming the leading initiatives in the field, such as the UK Biobank (UKB) (Bycroft et al., 2018) and the FinnGen study (Kurki et al., 2023). The UKB, which was initiated in 2006, captures detailed phenotype information matched to genetic data for more than 500,000 individuals (Bycroft et al., 2018). The FinnGen study is a large public-private partnership initiated in 2017 and aiming to collect and analyse genome and health data from 500,000 Finnish biobank participants,

currently analysing genetic and health information from nearly 300,000 individuals (Kurki et al., 2023). In comparison to other populations, biobanks from population isolates with a special genetic makeup such as Finland, and to a lesser extend Estonia, provide an added value, since they can provide opportunities for the detection of population-specific or population-enriched alleles, where GWAS in other populations remain largely underpowered to identify those (Kurki et al., 2023).

There have been multiple other significant initiatives worldwide such as Biobank Japan (Nagai et al., 2017), China Kadoori Biobank (Walters et al., 2023), Lifelines in the Netherlands (Sijtsma et al., 2022), the Million Veteran Program in the USA (Gaziano et al., 2016), the All of Us Research Program in the US (Denny et al., 2019), etc. While biobank initiatives have emerged in nation-wide settings, joint initiatives have also been launched, such as the Global Biobank Meta-analysis Initiative created in 2019. This initiative brings together 23 biobanks with a primary aim to work jointly to understand the genetic basis of human health and disease (Zhou et al., 2022) (Figure 2).

Since their inception, biobanks have facilitated biomedical discoveries at an unprecedented scale (Bycroft et al., 2018; Kurki et al., 2023; Leitsalu et al., 2015). Also, some biobanks returned or plan to return individual genomic research results to their participants, for instance the GeneRISK study in Finland (Widén et al., 2022) in deCODE or EstBB, which assessed hereditary breast and ovarian cancer genetic risk disclosing the risk to the participants (Leitsalu et al., 2021; Stefansdottir et al., 2020).

At the same time, biobanks pose several challenges such as obtaining informed consents and related privacy concerns, data security and governance, achieving diverse and representative participant cohorts, harmonisation of data structures and common protocols, communication and engagement, as well as building and maintaining a trust relationship with the participants (Akyüz et al., 2021).

While large-scale biobanks have gained prominence, smaller, population-specific birth cohorts with decades of follow-up analysis offer a complementary approach to the study of genetic susceptibility to disease, as well as some advantages compared to population-based biobank initiatives. For instance, these cohorts tend to allow for more detailed and in-depth data collection on participants, which enable higher granularity and deeper characterisation of phenotypes, usually including phenotypes that would be difficult to obtain from large biobank initiatives. However, it is important to note that these might have been designed with specific established research questions, contrary to biobank initiatives which tend to be agnostic and wider in scope.

One example of such population-specific birth cohort initiative is the Northern Finland Birth Cohorts program (NFBC), which was initiated in the 1960s in the two northernmost provinces of Finland to study risk factors involved in preterm birth and intrauterine growth retardation, and the consequences of these early adverse events on subsequent morbidity and mortality (Rantakallio, 1988). The uniqueness of NFBCs is that the data of the cohorts were obtained from early fetal life (including maternal health during pregnancy) to adulthood. The

NFBC1966 includes 12,058 live births to mothers in the two northernmost provinces of Finland, with different follow-up times and data acquisitions. One advantage of NFBC1966 is that it has reduced selection bias (being more representative of the whole population), with a high data coverage, including 96.3% of all births in 1966 in the region of Northern Finland (Rantakallio, 1988).

Amongst various applications in the study of health conditions, EHR-linked biobanks or population-based birth cohorts with deeper phenotyping provide opportunities to focus on less prevalent diseases that have been understudied, and thus provide a valuable framework for the study of genetic susceptibility to reproductive conditions in women. In the context of this thesis, we have focused on the study of the implementation, interpretation and potential applications around two of the main genomic epidemiology tools in the study of complex health traits: firstly, GWAS and secondly, PRS, which definitions, interpretations and applications are reviewed in the next chapters.

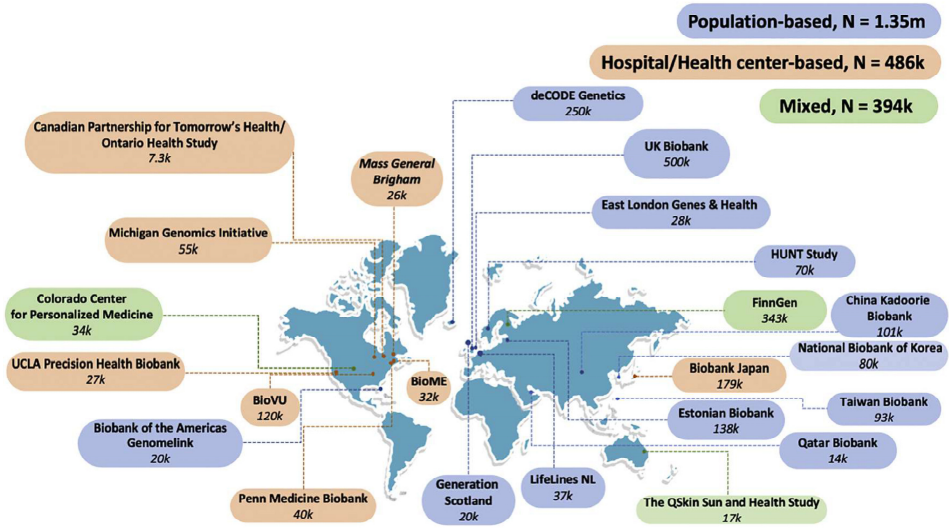

Figure 2. 23 biobanks across four continents which were included in the Global Biobank Meta-analysis Initiative as of April 2022, bringing the total number of samples with matched health data and genotypes to more than 2.2 million. Biobanks are coloured based on the sample recruiting strategies and *k* denotes thousand units. Extracted from Zhou et al. 2022 and reprinted with permission of the journal.

1.4. Genome-wide association studies (GWAS)

1.4.1. An introduction to GWAS

GWAS test for differences in the allele frequency of millions of variants between individuals who differ phenotypically and aim to identify associations of genotypes with phenotypes (Uffelmann et al., 2021). Phenotypes are commonly defined as binary traits (comparing two groups of individuals, for instance presence or absence of disease), or quantitative measurements on a whole study sample (such as hormone levels, height variation, etc.).

The most commonly studied genetic variants in GWAS are SNPs. SNPs are variations in the DNA sequence that affect a single base pair and are the most common genetic variation among people. On average, compared to a reference human genome, a person's ~6 billion-nucleotide genome sequence will have ~5 million SNPs (Carlson et al., 2003). SNPs typically have two alleles, which are two commonly occurring base-pair possibilities for a SNP location. The frequency of a SNP can be given in terms of the minor allele frequency (MAF) or the frequency of the less common allele.

Generating reliable results from GWAS requires careful quality control before and after conducting the GWAS. Some steps include removing rare or monomorphic variants, removing variants that are not in Hardy–Weinberg equilibrium, filtering SNPs that are missing from a fraction of individuals in the cohort, identifying and removing genotyping errors, and ensuring that phenotypes are well matched with genetic data (for instance in those cases where self-reported biological sex is not concordant to X and Y chromosomes), or removing individuals who are ancestry outliers (Turner et al., 2011).

GWAS are most commonly based on SNP-genotyping arrays, which are more affordable, and generally more suitable for detecting common variants. While SNP-genotyping arrays does not offer full genome-wide coverage, imputation involves the statistical inference of genotypes that have not been assayed directly using a sequenced haplotype reference panel. Some commonly used panels are the 1000 Genomes Project (Auton et al., 2015), TOPMed (Taliun et al., 2021) or population specific imputation reference panels such as in the Estonian population (Mitt et al., 2017) or the SiSu v3 in the Finnish population (Kurki et al., 2023).

Typically, in GWAS, logistic or linear regression models are used to test for associations, depending on whether the phenotype is binary or continuous, respectively. Software methods have evolved to be computationally efficient and may also allow to account for case-control imbalance and sample relatedness such as SAIGE and REGENIE (Mbatchou et al., 2021; Zhou et al., 2018). Covariates to control for confounding variables (such as age, sex), to correct for population stratification (for example using principal components (PCs) from pairwise IBD matrix), and to correct for technical variability (for example batch numbers) are included to account for stratification and avoid confounding effects (Price et al., 2010). Testing millions of associations between individual genetic variants and a

phenotype of interest requires a stringent multiple- testing threshold to avoid false positives, and usually a genome-wide significant threshold $p = 5 \times 10^{-8}$ is used (Risch & Merikangas, 1996).

Typically, after the regression model is ran, lead or sentinel variants (the variant which shows the minimal p-value in association testing) are reported from GWAS along with their genomic risk loci, which are blocks of correlated SNPs nearby (thresholds are usually set +/– 250–500kb from the lead variant). Neighbouring genetic variants are often correlated, as they tend to be inherited together due to co-segregation in meiotic recombination, what we know as linkage disequilibrium (LD) (Slatkin, 2008).

Large sample sizes are key for continued and increased variant-trait discoveries in GWAS. To this end, combining datasets through meta-analysis is a cost-effective solution, which enables a joint analysis of different GWAS studies by combining summary association statistics for each variant and phenotype of interest from different studies (Zeggini & Ioannidis, 2009). This approach increases the power to detect association signals by increasing sample size (Zeggini & Ioannidis, 2009), enabled by the development of efficient softwares such as GWAMA or METAL (Mägi & Morris, 2010; Willer et al., 2010). In response to the rapid increase in the number of published GWAS, the GWAS Catalog was initiated in 2008 aiming to provide a consistent, searchable, visualisable and freely available database of SNP-trait associations (https://www.ebi.ac.uk/gwas/home) (Sollis et al., 2023), which can be integrated with other resources. The GWAS Catalog clearly shows an exponential growth in the number of studies and sample sizes over time – of note reaching the threshold of one million participants for traits such as height (Yengo et al., 2022), smoking initiation (M. Liu et al., 2019), educational attainment (Okbay et al., 2022) and blood pressure (Evangelou et al., 2018).

1.4.2. Interpretation of GWAS

17 years far from the first well-designed GWAS by the Wellcome Trust Case Control Consortium ("Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls.," 2007), thousands of GWASs have now been published, identifying thousands of genetic loci for many human diseases and traits (Sollis et al., 2023). This framework has sometimes yielded interesting genetic associations for various phenotypes (for instance, *FTO* in association with obesity (Frayling et al., 2007; Z. Zhang et al., 2023) and valuable inputs for drug development, exemplified recently by the first Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based approved therapy for sickle cell disease & beta-thalassemia which targets the gene *BCL11A*, initially identified by GWAS (Menzel et al., 2007; Uda et al., 2008).

However, as the International Common Disease Alliance (ICDA) highlighted, the goal and huge challenge ahead now is to move from disease-associated loci to disease biology, to disease treatment (the "Maps to Mechanisms to Medicine" challenge) (*ICDA Recommendations and White Paper.Pdf – Google Drive*, n.d.).

In order to do that, there is the need to develop systematic ways to discover which are the exact causal variants affecting the trait of interest, their immediate molecular effect, the target genes on which these variants act and in which cell types and states these genes operate, along with defining which pathways are affected. At the same time there is the need to interrogate a wide spectrum of causal alleles, to investigate the potential effects of variants in other diseases and also build on effective cellular and animal models to aid study of disease processes (*ICDA Recommendations and White Paper.Pdf – Google Drive*, n.d.).

So far, these aspects have remained a challenge due to various reasons. Firstly, most traits are influenced by thousands of causal variants (Holland et al., 2020), which are correlated with causal and non-causal variants that are physically close as a result of LD (Slatkin, 2008). The genetic variants identified are often associated with many other traits (Watanabe et al., 2019), a situation known as pleiotropy (Stearns, 2010; Wagner & Zhang, 2011), which can interfere with interpretation of the variant's effect on a phenotype since the same variant can have a direct biological influence on more than one phenotypic trait in some cases (Solovieff et al., 2013). Variants may also be exhibiting gene-environment correlations, and variants identified from GWAS usually explain a relatively moderate portion of the phenotypic variance and individually confer very little risk (Abdellaoui et al., 2023). A majority of the trait-associated variants are either localised in intergenic regions or introns (Hindorff et al., 2009), which are difficult to interpret and more likely to have regulatory effects, compared to coding variants which tend to have clear functional consequences (Maurano et al., 2012; Musunuru et al., 2010). These inherent features of GWAS make direct biological, causal inferences complicated. Additionally, complex traits-associated loci often contain multiple genes, making it challenging to distinguish the affected ones, and traits' pathophysiology might involve interaction of multiple cell types.

For example, in coronary artery disease (CAD), the development of atherosclerotic plaques involves monocytes, lymphocytes, mast cells, neutrophils and smooth muscle (Insull, 2009). It is unclear in which cell types the GWAS variants act, i.e. which cell types are the true drivers of a disease. Single-cell RNA (scRNA) sequencing data allows to create cell type specific transcriptome profiles, which can then be aligned with GWAS to implicate cell type specificity of the traits. For instance, using scRNA-sequencing, a study identified 12 atherosclerosis-associated cell states and found that Vcam1+ smooth muscle cell state contributed the most to the heritability of CAD (Örd et al., 2023).

Therefore, post-GWAS analyses are of crucial importance to enhance the interpretation of GWAS and the imminent growth of several scientific driving forces in human genetics, cell biology and data science, constitute a favourable road ahead to translate GWAS findings into valuable mechanistic and medical insights (Figure 3). Many integrations of GWAS data with molecular traits such as gene expression, chromatin activity and/or other omics layers exist and here I will review a selection of methods that have been of wide use and interest. These aim to functionally annotate the associated variants, to nominate the affected gene or genes in the locus that potentially mediate the trait's association, to test which

downstream network or pathway may be affected and lead to changes in function, and which is the relevant tissue and/or cell type where the discovered genetic signals cause an effect. To this end, different *in silico* approaches have been implemented to infer the molecular effects of GWAS variants and I will introduce some of the most common tools next. While these methods can be implemented individually, also platforms have been created to enhance annotation, prioritization, visualization and interpretation of GWAS results such as FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies), an online platform for annotation of GWAS signals (Watanabe, Taskesen, Van Bochoven, et al., 2017). While certain authors have developed systematic softwares which aim to prioritise genes from GWAS signals, such as MAGMA (de Leeuw et al., 2015), DEPICT (Pers et al., 2015) or PoPs (Weeks et al., 2023), there is no gold standard way to do so, neither there is for estimating the affected regulatory pathways or tissues/cell types. Each study may rely on varied sources of information, influenced by factors such as study type, objectives, and resource availability at the time of the study. In the following sections, I will examine the current status of key steps in GWAS interpretation that I have later implemented in the studies included in the thesis.

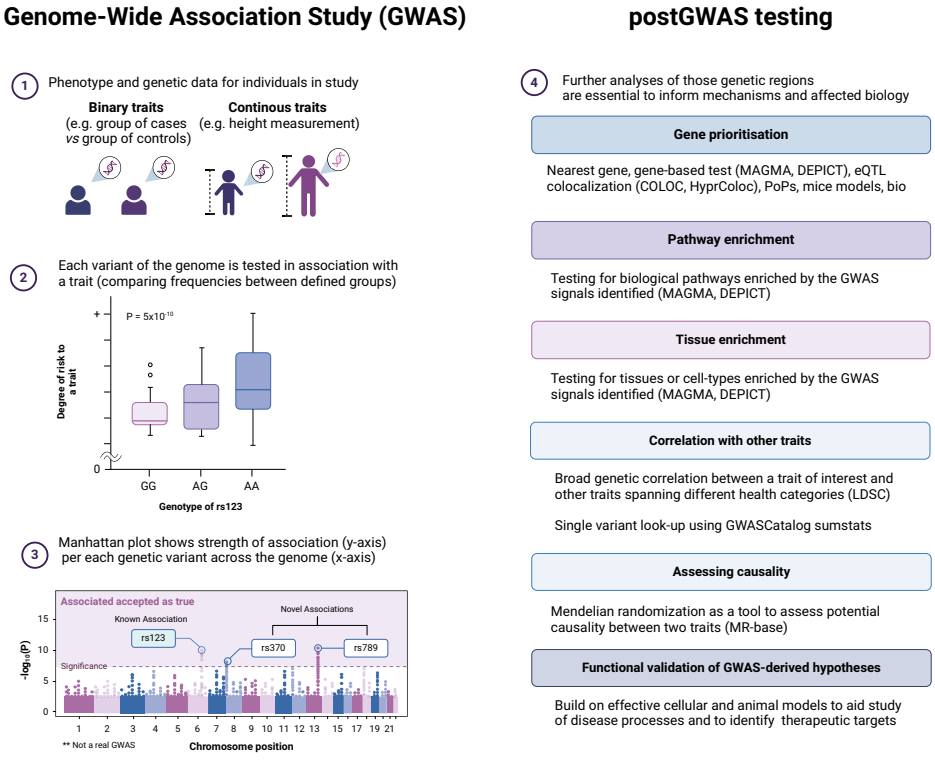

Figure 3. Overview of GWAS and post-GWAS analyses. Created from BioRender.com and adapted from "The Principle of a Genome-wide Association Study (GWAS)", by BioRender.com (2024). Retrieved from https://app.biorender.com/biorender-templates

1.4.2.1. From variants to genes

Identifying the likely affected genes in a locus remains a crucial task. While coding variants tend to be more straightforward to interpret using annotation tools like ANNOVAR (K. Wang et al., 2010) to annotate their potential effects on specific genes, most of associated SNPs are non-coding, often implicating regulatory mechanisms that influence disease associations. In this case, it has been proposed that with quite high likelihood, the gene closest to the association signal is probably the affected gene (Aragam et al., 2022; Backman et al., 2021; Barbeira et al., 2021; Nasser et al., 2021; Stacey et al., 2019). While this would be the easiest and most straightforward approach, another work suggests this is not always the case (Morris et al., 2023), and molecular evidence would still be needed to propose mechanisms of action for the genetic association. To this end, several potential avenues can be pursued. An approach for identifying regulatory target genes of genetic variants is molecular quantitative trait loci (molQTLs) analysis, which associates genetic variants with specific molecular phenotypes; for example, expression quantitative trait loci (eQTL) analysis identifies loci associated with RNA expression, which are the most common compared to other QTL-mapping traits, partly because of the robustness in RNA-sequencing technologies. By integrating this information with GWAS results, trait-associated variants can be mapped to the genes they are likely to regulate in specific tissues and infer the molecular processes mediating these associations (Y. I. Li et al., 2016; Lonsdale et al., 2013). Since annotating a variant does not guarantee the variant is causal, as it can likely show a significant association and be in high LD of the truly causal variant, eQTLs should be integrated with GWAS data using co-localization approaches, such as COLOC (Giambartolomei et al., 2014) or HyPrColoc (Foley et al., 2021), which aim to pinpoint loci where the regulatory association and disease association share the same causal variant(s) (Giambartolomei et al., 2014). This exploration into regulatory mechanisms of GWAS loci benefits from the curation of comprehensive molecular quantitative trait loci catalogues, like the eQTL catalogue (Kerimov et al., 2021, 2023). This englobes multiple comprehensive eQTL resources as for instance, the Genotype-tissue expression project (GTEx), which is the widest resource to date (Lonsdale et al., 2013).

Once a shortlist of potential affected genes has been compiled, complementary strategies, such as querying data from mouse mutant phenotypes using platforms like the Mouse Genome Database (http://www.informatics.jax.org/) might contribute to the identification of biologically most relevant genes. Integrating information from literature searches or reported associations in the GWAS Catalog aids in highlighting genes with potential biological significance, although relying purely on previous literature and biological knowledge can potentially hamper novel discoveries.

Other relevant layers of information in this step can arise from different authors' in-house pipelines, for instance Kentistou *et al.* implemented the framework 'GWAS to Genes', which integrated genomic and functional data across six sources to move from thousands of loci associated with age at menarche (AAM)

to candidate genes (Kentistou et al., 2023). They considered genes by identifying signals that co-localised with a) known enhancers and regulatory elements (Nasser et al., 2021), b) non-synonymous variants, c) eQTL specifically in tissues enriched for AAM associations, and d) circulating protein quantitative trait loci from whole blood (see Methods of Kentistou *et al*.) In addition, they integrated gene-level associations for aggregated non-synonymous common variants using MAGMA (de Leeuw et al., 2015) and gene scores from PoPs (Weeks et al., 2023), which uses bulk human and mouse data with information on scRNA, gene pathways and protein interactions to link genes to GWAS signals. Individual genes were further upweighted if they were the nearest gene to the signal.

Successful examples of candidate genes from GWAS which have been confirmed with functional studies in humans involve loci such as FTO in association with BMI and brown fat thermogenesis (Frayling et al., 2007; Z. Zhang et al., 2023) and *SORT1* in association with both plasma low-density lipoprotein cholesterol and myocardial infarction (Musunuru et al., 2010). Studies in mice have also supported and further characterised GWAS targets, for instance the female-specific locus *PAX1* in association with adolescent idiopathic scoliosis (Ushiki et al., 2024). Other works using genome-editing have also supported the regulatory effect of a non-coding variant on *ARID3A* expression, supporting the contribution of non-coding variation to primary biliary cholangitis (You Li et al., 2023).

1.4.2.2. Determining affected regulatory pathways and tissues/cell types

The signals from GWAS for any trait can potentially converge on a narrower number of biological processes, and the pathway-level effects of genetic variants can be determined and linked to cellular and physiological functions. One approach to achieve this is to use MAGMA (de Leeuw et al., 2015) which takes into account the genetic variants and the LD between them to highlight potential affected genes (a step known as gene-based test) and then by using these list of genes it tests for enriched genetic pathways/convergent functions. DEPICT (Pers et al., 2015), is also a method for gene prioritization, gene set enrichment analysis, and tissue enrichment analysis. DEPICT's primary innovation is the use of "reconstituted" gene sets, which consist of 14,462 gene sets downloaded from multiple databases that have been extended based on 77,840 publicly available expression microarrays (Fehrmann et al., 2015). The reconstituted gene sets contain z-scores for each gene in the genome for each of the 14,462 gene sets, representing how strongly each gene is predicted to be a member of each gene set. Additionally, DEPICT utilises a set of 37,427 human microarrays to identify tissue/cell types in which genes from associated loci are highly expressed. These tools evaluate sets of genes involved in specific biological pathways or associated to specific tissues, cell types, developmental stages or protein networks that are presumed to be proximal causes of the studied trait and relevant to the association

with this trait. Commonly used gene sets are, for example, extracted from the molecular signatures database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/) which is a resource of tens of thousands of annotated gene sets. The way gene sets are defined is critical and rely on the accuracy of the biological annotations made. If, for a given trait, GWAS loci are enriched (overrepresented) for genes specifically expressed in a given gene-set, tissue or cell type, these will be prioritised, accounting for multiple adjustment of tests using Bonferroni or false discovery rate correction.

Additionally, GWAS variants can be annotated with chromatin annotations such as open chromatin regions (Boyle et al., 2008; Buenrostro et al., 2013), histone modifications (Bannister & Kouzarides, 2011) or DNA methylation (Frommer et al., 1992), which may aid identifying genome elements of high levels of regulatory activity. Considering the tissue type, cell type or cell state is essential for all functional interpretation work, as genes may have pleiotropic effects across different cellular contexts. For example, whole tissue-level molecular data can include masked cell type-specific signals, further complicating interpretation or masking true signals from rare cell types. Upcoming single-cell and cell type-specific functional genomic data sets (Regev et al., 2017; van der Wijst et al., 2020) are therefore likely to advance GWAS interpretation further. This will be enhanced as well with increased statistical power gained by including more individuals or variants in those analyses, as well as increasing availability of expression and chromatin data for more cell types and states. In an in vitro setting, experimental perturbation of genes followed by cellular phenotyping is becoming increasingly scalable and informative for interpretation of GWAS loci and genes (Adamson et al., 2016; Dixit et al., 2016).

1.4.3. Applications of GWAS

Beyond detecting statistically associated variants and interpreting their role in a biological context, the results of GWAS can also be used for a range of applications, such as exploring the genetic architecture of traits, estimating its heritability, calculating genetic correlations, inferring potential causal relationships between traits, predicting disease risk, informing drug development programmes, etc. However, GWAS may face some limitations, such as ancestry transferability, phenotype resolution and population stratification, which will be further commented in the next chapters.

Understanding the genetic architecture of a trait involves estimating the number, effect sizes, and frequencies of associated variants. One of the estimates that allows us to infer genetic architecture beyond a trait is the calculation of heritability, which is the portion of trait variance explained by genetic variation. In GWAS, statistical methods and computational tools like GCTA, LDSC or SumHer, can quantify SNP-based heritability (Bulik-Sullivan et al., 2015; Finucane et al., 2015; Speed & Balding, 2019; Yang et al., 2011) from summary-level statistics, which measures the additive effects of common variants towards a trait

variance. However, SNP-based heritability might overlook the influence of rare variants and other contributing genetic factors (Manolio et al., 2009).

Another application from GWAS is to test for genetic correlations, aiming to understand the genome-wide correlation of the additive genetic effects between two traits. This is an important step, as SNPs associated with one trait might influence unrelated phenotypes. Several methods have been developed to estimate genetic correlation on a broad scale (considering only the average of the shared signal across well-referenced SNPs), such as cross-trait LDSC which uses summary statistics as input (Bulik-Sullivan et al., 2015), GNOVA (Lu et al., 2017), which introduced to estimate annotation-stratified genetic correlation, and high-definition likelihood (HDL) (Ning et al., 2020), which accounts for LD across the genome. By integrating these methods and bioinformatics website service, some tools such as LD Hub (Zheng et al., 2017) or Complex Trait Genetics Virtual Lab (CTG-VL) (https://vl.genoma.io/) have enabled to undertake genetic correlation leveraging data for many traits. Genetic correlation analyses have unravelled interesting observations, for instance a high genetic correlation (71%) between infertility of "unknown" cause and endometriosis, suggesting undiagnosed endometriosis may be responsible for many of these cases (Venkatesh et al., 2024). Genetic correlation has also been used to dissect genetic relationships between diagnoses across health domains, for instance for psychiatric disorders (Smoller et al., 2013).

Local genetic correlation has been more recently enabled by softwares such as *p*-hess (H. Shi et al., 2016), SUPERGNOVA (Yiliang Zhang et al., 2021) and LAVA (Werme et al., 2022), which in this case estimate genetic correlation locally to pre-defined genomic regions, links which are often masked or dismissed when running genetic correlation globally (H. Shi et al., 2016; Werme et al., 2022). This advance has led to interesting observations such as reporting two etiologically distinct genetic signatures with bidirectional local genetic correlations which could explain the positive yet paradoxical genetic correlation between autism spectrum disorder and cognitive performance (Yiliang Zhang et al., 2021). However, genetic correlations need cautious interpretation, as they do not imply causation and can result from various mechanisms, including pleiotropy. To this end Mendelian Randomisation (MR) enables to assess causal relations between different phenotypes using GWAS summary statistics (G. D. Smith & Ebrahim, 2003) by using genetic variants as instrumental variables as proxy measures for a randomised environmental exposure. For instance, a MR analysis identified a causal effect of genetically lowered vitamin D levels on increased susceptibility to multiple sclerosis, utilising genetic variants near genes with well-characterised effects on vitamin D synthesis, metabolism and transport (Mokry et al., 2015).

Beyond leveraging GWAS to explore molecular biology, GWAS serves as well to formulate mechanism-based therapies which can have a direct clinical impact. While two-thirds of approved drugs by the Food and Drug Administration were supported with evidence from genetics (Ochoa et al., 2022; Trajanoska et al., 2023), recent drug development such as CRISPR-editing for sickle cell anemia

or drugs repurposing examples such as ustekinumab and risankizumab to treat Crohn's disease have benefited from genetic insights and mechanistic dissection from GWAS findings (de Lange et al., 2017; Duerr et al., 2006; Menzel et al., 2007; Uda et al., 2008).

GWAS has also laid the foundation for 'personalised medicine' in different ways, for example by generating input for calculating PRS. As PRS has been of great interest in the human genetics field and this thesis includes research concerning PRS, the following section reviews its definition, calculation and introduces its main applications and implications arising from their development.

1.5. Polygenic risk scores

1.5.1. An introduction to PRS

A PRS is a single value estimate of an individual's common genetic predisposition to a phenotype. PRS nomenclature is heterogeneous, and they tend to be referred to as genetic/genomic risk scores as well. Here, we will keep with the nomenclature of "polygenic risk score (PRS)" to refer to this concept.

This single value estimate is calculated as a sum of an individual's genotypes on a genome-wide scale, weighted by corresponding risk allele effect size estimates (the effect sizes can be adjusted or non-adjusted) extracted from a particular trait's GWAS summary statistics. The trait for which the PRS will be tested may be the same as the one in the initial (base) GWAS or different; for example, testing the hypothesis that a PRS for PCOS associates with T2D case–control status. Similarly to GWAS, PRS can be constructed both for binary and continuous traits, and PRSs for hundreds of traits or diseases can be calculated from one genome-wide array or sequence.

PRS analyses can be characterised by the two input data sets that they require: firstly, the base dataset, which is the basis for the summary statistics (e.g. betas, p-values) for each SNP from a relevant GWAS which originated from individual-level data of genotypes and phenotype(s) of a base sample. Secondly, the target dataset: individual-level genotypes and phenotype(s) of a target sample (Choi et al., 2020) (Figure 4). These two samples must be independent, meaning that there must be no sample overlap between the sample from which you ran the GWAS and extract the effect sizes, and the sample in which you will calculate the individual PRS.

Careful quality control must be conducted as well on the base and target datasets beyond ensuring the standard GWAS quality control and the sample independence. For instance, it is important to ensure minimal heritability for the trait of interest (SNP-$h^2 > 0.05$) (Choi et al., 2020) and ensure the correct identification of effect alleles in GWAS data to avoid spurious results in the PRS. In the target data, sample sizes of at least 100 individuals should be guaranteed to avoid underpowered tests (Choi et al., 2020). Other quality control steps might include genome build alignment, removal of ambiguous SNPs and duplicated SNPs.

Important challenges in the construction of PRS are the number and selection of SNPs for inclusion in the score and what, if any, adjustment to apply to the GWAS effect size estimates. When parameters for generating an optimal PRS for a trait are unknown, then the target individual level data can be used for model training, allowing optimisation of model parameters. Different methods to construct and evaluate PRS have been developed over the last decade, the most commonly used currently being LDpred2 (Privé et al., 2021), PRS-CS (T. Ge et al., 2019), PRSice-2 (Choi & O'Reilly, 2019) or megaPRS (Q. Zhang et al., 2021). However, there is no universally accepted methodology and characteristics such as genotyping density and sample size (Lambert et al., 2019), or the genetic architecture of traits influence which approach better captures the genetic component of a trait (Pardiñas et al., 2023; Y. Wang et al., 2023).

The SNP-h^2 of the trait would be equal to the variance of the trait explained by the PRS in an ideal situation, where the effects of the SNPs were estimated from the GWAS without error (L. M. Evans et al., 2018). However, due to error in the effect size estimates and inevitable differences in for example LD pattern, environment and background between the base and target samples, the predictive power of PRS are typically substantially lower than SNP-h^2.

Currently over a thousand publications have focused on creating PRS for different human traits and diseases. The Polygenic Score (PGS) Catalog (https://www.pgscatalog.org/) (Lambert et al., 2021), created in 2019, is a publicly available, manually curated, open database of PRS with their relevant metadata, aiming to provide systematic evaluation and enhance reproducibility. However, the calculation of these estimates is not standardised nor straightforward, their potential applications vary and there are also known risks and limitations of PRSs, which I will introduce in the next subchapters.

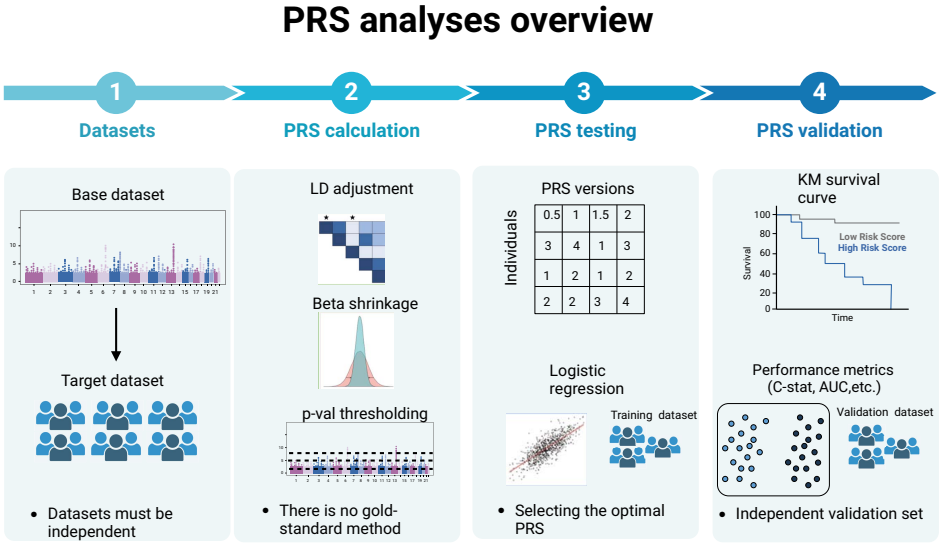

Figure 4. PRS analyses overview. Adapted from Choi *et al.* 2020 and created with BioRender.com LD: Linkage Disequilibrium, KM: Kaplan Meier, C-stat: C-statistic, AUC: Area Under the Curve.

1.5.2. Considerations for the calculation and testing of PRSs

There are several options in terms of how PRSs are calculated. Initially, PRS tended to be constructed from genome-wide significant SNPs (typically, $P < 5 \times 10^{-8}$), which for many diseases led to weakly predictive PRS as the number of genome-wide SNPs was small (D. M. Evans et al., 2009; Purcell et al., 2009). In general, more powerful PRSs tend to be constructed incorporating a larger set of SNPs.

There are different methods which consider different key factors such as the potential adjustment of GWAS estimated effect sizes, for example with applying 'shrinkage', and the task of accounting for LD. Given that SNP effects are estimated with uncertainty, and since not all SNPs influence the trait under study, the use of unadjusted effect size estimates of all SNPs could generate poorly performing PRSs with high standard error. To address this, different strategies have been adopted: firstly, the use of *P* value selection thresholds as inclusion criteria for SNPs into the score. Secondly, to perform shrinkage of the effect estimates of all SNPs (Figure 4).

In the classic PRS calculation method (Dudbridge, 2013; Euesden et al., 2015; Purcell et al., 2009) only those SNPs with a GWAS association *P* value below a certain threshold (e.g., $P < 1 \times 10^{-5}$) are included in the calculation of the PRS, while all other SNPs are excluded. This approach does not modify the effects coming from the GWAS of the included SNPs. Since the optimal *P* value threshold is unknown *a priori*, PRSs are typically calculated over a range of thresholds, association with the target trait is tested for each threshold in an independent sample, followed by selection of the optimally predicting PRS. However, there are also PRS methods that perform shrinkage of all SNPs (T. Ge et al., 2019; Mak et al., 2017; Newcombe et al., 2019; Vilhjálmsson et al., 2015) (Figure 4). Under different approaches or parameter settings, varying forms of shrinkage can be applied. Since the optimal shrinkage parameters are unknown *a priori* multiple PRS are calculated, typically by ranging possible parameter values, which in the case of LDpred, for example, includes a parameter for the theoretical fraction of causal variants (Vilhjálmsson et al., 2015).

Another aspect to consider when building a PRS is the LD structure between the variants (Figure 4). Usually association tests in GWASs are performed one SNP at a time, which, combined with the strong correlation structure across the genome, makes estimating the joint genetic effects challenging. In an ideal scenario, we would take the effect estimates from a model which includes all SNPs jointly in a GWAS. However, this is not feasible due to differences between base and target datasets, smaller sample sizes than the number of SNPs, as well as multicollinearity problems.

There are two main options for approximating the PRS that would be obtained from joint effect estimates: 1) SNPs are clumped (i.e. prioritising SNPs at the locus with the smallest GWAS *P* value) so that the retained SNPs are largely independent of each other, and, thus, their effects can be summed, assuming additivity 2) more complex methods that explicitly account for LD, such as LDpred (Vilhjálmsson et al., 2015) for all SNPs are included, accounting for the LD

between them. Both clumping and LD modelling require estimation of the LD between SNPs, which may come from the LD reference panel for the base data population or reference sample for the same ancestry (Auton et al., 2015). In the first PRS calculation approach (Dudbridge, 2013; Euesden et al., 2015; Purcell et al., 2009), this is usually combined with *P* value thresholding and called the C+T (clumping + thresholding) method (which is faster and in general more interpretable to apply), while the second option is generally favoured in methods that implement shrinkage techniques (T. Ge et al., 2019; Mak et al., 2017; Newcombe et al., 2019; Vilhjálmsson et al., 2015).

After constructing the PRS, it is essential to test its association with the trait or disease of interest. For binary traits, the effect sizes are expressed as odds ratios (OR) or hazard ratios, depending on the study design (case/control vs. prospective) and the availability of age at event (Lambert et al., 2019). The model's performance can be measured using variance explained (Nagelkerke's or pseudo-R2) or classification accuracy using area under the receiver-operating characteristic curve (AUC), the area under the precision-recall curve or Harrell's C-index (Steyerberg et al., 2010) (Figure 4). In practice, the target dataset with individual-level genotype and phenotype data can be divided into a 1) training or discovery set (which is used to optimise the PRS and select the best PRS model) and 2) separate testing or validation set where the predictive performance of the PRS can be estimated.

1.5.3. Applications of PRS

One of the main potential applications of PRS is to predict future disease risk and identify those individuals who are at the highest risk for a condition. Consequently, this information can be used to enhance prevention (e.g. target specific actions or treatments) or modify screening strategies. Prediction of disease risk is an important aim in preventative medicine, which seeks to advance personalised medicine strategies, aiming to guide clinical management and interventions.

For many clinical use cases, PRS are being evaluated around the world to determine what clinical utility they may have when included in integrated risk tools for cardiovascular disease assessment (Fuat et al., 2024), or breast and ovarian cancer (Carver et al., 2021; Esserman, 2017; Knoppers et al., 2021; Roux et al., 2022). Beyond cancer and cardiovascular risk, some trials are evaluating PRS for use in autoantibody screening of type 1 diabetes (Sims et al., 2022).

PRS have been shown to be cost-effective to optimise cancer screening (Dixon et al., 2022; Wong et al., 2021) and also for cardiovascular disease (Kiflen et al., 2022). However, more evidence is needed, and in general studies do not account for the fact that a single array/sequence could modestly improve risk stratification for multiple diseases simultaneously.

Clinical risk prediction models typically include risk factors such as age, sex, family history of disease, prevalent diseases and lifestyle factors (e.g. smoking status). In contrast, genomic risk information which can be partly captured in a

PRS is set at conception and can therefore be utilised much earlier in life. Moreover, PRS captures risk which tends to be complementary to these traditional risk factors (Inouye et al., 2018; Mars et al., 2022).

Another predictor that has been used traditionally for genetic disease risk prediction is monogenic mutations. While these mutations are often highly penetrant, their relative rarity in the population means that they only explain a small fraction of overall disease cases. Furthermore, PRS derived from GWAS summary statistics are usually missing rare variants, since those are generally not captured by genome-wide genotyping arrays and imputation. Since monogenic and polygenic risks are largely independent (Adeyemo et al., 2021; Mars et al., 2020), and with the increasing availability of whole genome sequencing (WGS) in biobank initiatives (Hawkes et al., 2023; S. Li et al., 2023), scores integrating a wide allelic spectrum (thereby combining monogenic/polygenic contribution) will likely provide the most information for individual genomic risk prediction (Dornbos et al., 2022) and the development of methods to achieve such scores is an active area of research (Chan et al., 2022; Lali et al., 2021; Z. Wang et al., 2022).

Similarly, the effects of family history and PRS have been reported to be largely independent, a pattern observed across diseases (Mars et al., 2022), so PRS is a complementary measure of information, and neither monogenic risk nor family history are interchangeable measures of information.

There is now a strong evidence base across many diseases that PRS captures disease risk information that is independent of other risk factors and improves integrated risk calculators (Xiang et al., 2024). Thus, integrating both genetic and non-genetic risk factors can provide the best estimates for complex disease risk prediction, an area that has gained interest lately. To this end, several studies show that PRS improves risk prediction when combined into integrative risk scores for conditions such as CAD (Sun et al., 2021; Weale et al., 2021), stroke (Fahed et al., 2022; Neumann et al., 2022), T2D (Hodgson et al., 2022; Neumann et al., 2022), and breast cancer (Hurson et al., 2022).

While the use of PRS in disease risk prediction has attracted a lot of attention, PRSs also hold potential utility in other areas such as improving diagnostic accuracy, for instance in e.g. discriminating type 1 diabetes from T2D (Shoaib et al., 2023), guiding treatment decisions (including pharmacological interventions, for instance defining those individuals who would mostly benefit from statin prescription (Riveros-Mckay et al., 2021; Sun et al., 2021), to increase the efficiency in clinical trials (for instance reducing trial sample size by focusing on high polygenic risk to increase the outcome rate (Fahed et al., 2022) and exploring the relationship between different traits (Richardson et al., 2019). In this latter application, exploring the relationship between a PRS for a certain trait and other traits would help identify potential risk factors for a trait of interest and the degree of comorbidity they present. In this regard, some studies have been focused on exploring associations with one specific disease, for instance, a work showing that high PRS for schizophrenia are associated with decreased cardiac volumes, stronger heart contractions, and decreased heart relaxation rates in the UKB (Pillinger et al., 2023). Other efforts have systematically evaluated different PRS

relationships across the human phenome, for instance a study that analysed more than 300 PRS across over 500 heritable traits from the UKB, creating an atlas of associations (http://mrcieu.mrsoftware.org/PRS_atlas/), uncovering novel mechanisms which contribute towards disease susceptibility (Richardson et al., 2019).

1.5.4. Limitations and challenges for the PRS field

Despite some of the demonstrated benefits of PRS and their potential in clinical utility, various technical, practical and ethical concerns need resolution before widespread clinical adoption (Figure 5). Several public and private settings have launched their own initiatives – whether for research or business purposes and consequently, both the ICDA (Adeyemo et al., 2021) and the American College of Medical Genetics and Genomics (ACMG) (Abu-El-Haija et al., 2023) have developed guidelines towards PRS use. Unfortunately, some private initiatives have even taken a controversial step forward, starting pre-implantation genetic testing using PRS of IVF embryos (such as the Orchid company (https://www.orchidhealth.com/) and more recently Gattaca Genomics (https://gattacagenomics.com/), clear examples of problematic applications which are not appropriate for clinical use and should not be offered at this time ("Direct-to-Consumer Prenatal Testing for Multigenic or Polygenic Disorders: A Position Statement of the American College of Medical Genetics and Genomics (ACMG)," 2021; Turley et al., 2021). This reality, along with the increasing interest in PRS, underscores the need for continued research and debate to enhance the awareness of their limitations and enhance the responsible use of PRS (Adeyemo et al., 2021).

First of all, the development of PRS has originated and existed in the research area, where development methods and standards are continually evolving. The risk predictions from two different PRS for a particular disease can vary significantly due to factors like the number and non-overlapping sets of SNPs, diverse effect sizes, and the characteristics of the base GWAS summary statistics. Variability is also influenced by computational methods, training samples, and adjustments for covariates like age and sex (Läll et al., 2019). This lack of consistency poses a significant concern for the PRS field and highlights the need for holding reporting standards between studies (Wand et al., 2021) in order to be reproducible and useful.

Other challenges persist, including the need to diversify genotyped cohorts (particularly among non-European ancestries) which at the moment hinder the accuracy of PRSs for diverse populations and exacerbate healthcare disparities (Martin et al., 2019). There is also a need to investigate sex-based differences in PRS performance (Jermy et al., 2023), which for instance has been proposed in cardiovascular disease events (Hajek et al., 2018) or psychiatric conditions (Docherty et al., 2020), and across different phenotype / diagnose sources since a potential limitation of the applicability of the PRS can also include differences in phenotype resolution (Burstein et al., 2023).

Additionally, the clinical benefit of PRSs for disease risk prediction depends on the availability of preventive interventions and/or medicines, the damage of the action to implement, and the cost-effectiveness of the measure to be taken. One must be cautious and aware of the risks arising from delivering 'incorrect' information. For instance, false positive results can wrongly categorise an individual as 'high risk', which could lead to mistaken clinical actions (the effect of which will depend on disease severity, the contribution of non-genetic risk factors, and the cost and consequences of any intervention) and/or unnecessary emotional harm. Communication challenges implicate conveying uncertainties associated with PRSs to the medical community and the general population, also risking misunderstanding, and the need to secure adequate resources to facilitate effective communication (Lewis et al., 2021, 2022). So far, there have been examples of both positive and no significant lifestyle changes upon receiving genetic risk estimates information, (Silarova et al., 2019; Widén et al., 2022), exemplifying the difficulty of affecting human behavioural change. Importantly, these studies did not find anxiety and depression in response to PRS information to be common.

Moreover, regulatory uncertainties globally complicate the assurance of safety, equity, and effectiveness in PRS deployment, highlighting the need for harmonization and consideration of global collaboration's impact on access and implementation. In line with GWAS, ancestral diversity is also a problem for PRS (Duncan et al., 2019; Xiang et al., 2024).

A recent statement for the ACMG (Abu-El-Haija et al., 2023), seeking to offer guidance to the health care provider, aligns with a prevailing thoughts in the PRS community, which advocates against clinical implementation of PRS testing unless the provider and patient have a clear understanding of the limitations of the testing and applicability to the specific patient, including how the results will be used to guide evidence-based clinical care (Abu-El-Haija et al., 2023). Both the ICDA (Adeyemo et al., 2021) and ACMG statements emphasise equitable use of PRS, advocating for methodological development and data collection to ensure optimal performance across all individuals regardless of genetic ancestry. They also stress the importance of avoiding unethical or harmful applications of PRS. However, there's a lack of consensus on best practices and resources when individuals present PRS results obtained from third-party providers to healthcare practitioners, which is becoming increasingly common. Overall, while PRSs offer promising insights in research, their responsible implementation calls for addressing the uncertainties, biases, communication strategies, and their impact in society.

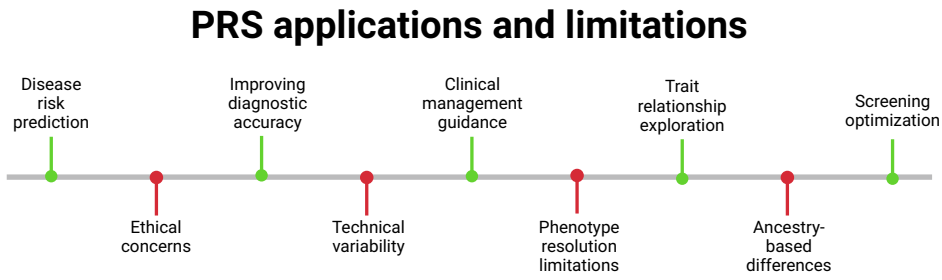

Figure 5. Main polygenic risk scores (PRS) applications (shown as green lines) and limitations (shown as red lines). Created with BioRender.com

1.6. State-of-the-science in female reproductive genomics

In recent years, there has been a transformation in the study of genetics and genomics in the field of women's reproductive health. This shift has been favoured by various circumstances. For instance, a response to the recognition of chronic underrepresentation of women's (reproductive) health in (genomics) research ("A Life-Course Approach to Women's Health.," 2024; Mercuri & Cox, 2022). Additionally, there has been an advancement in genetic technologies and an increase in the number of genomically-profiled biobanks worldwide and also an overall realisation that candidate gene studies have been quite unreliable, with a major lack of reproducibility as seen in other complex disease areas such as psychiatric genetics (Border et al., 2019; Farrell et al., 2015).

The GWAS Catalog shows an exponential growth in the number of GWAS over time and as of 11-02-2024, it contains data from 6,741 GWAS publications. Various factors make an accurate quantification of the number of GWAS completed for female reproductive traits challenging, such as the lack of sex-disaggregated information and the fact that female reproductive traits are dispersed across various disease categories in the catalog, such as "Other disease", "Lab measurements", "Other trait" and "Cancer" categories. At the same time, some of the presented GWAS publications, while including a broad number of tested traits, have been rather focused on software development or methodological/technological details instead of focusing on the biology and the characterisation of new genetic loci of a certain reproductive condition. Based on an approximate query, searching by term "female reproductive system disease" (EFO_009549), we observe that GWAS for female reproductive traits remain notably less studied compared to other health domains: the number of studies represent barely more than 5% of the total (441/6,741), compared to other health domains such as psychiatric disorders (24%, 1,620/6,741) or cardiovascular disease (22%, 1,476/6,741).

However, despite being the new kid on the block of GWAS, the female reproductive genomics field is punching above its weight. The ReproGen consortium emerged as a pioneer in characterising genetic variants and genes influencing age at menarche in 2010 (Elks et al., 2010) and age at menopause in 2012 (Stolk et al., 2012) through GWAS meta-analysis. Since then, various GWAS have shown remarkable progress in identifying the genetic landscape of common

gynecological diagnoses, such as PCOS (F. Day et al., 2018), endometriosis (Rahmioglu et al., 2023), polyps of the genital tract (Pathare et al., 2024), endometrial (O'Mara et al., 2018), ovarian (Phelan et al., 2017) and cervical cancer (Koel et al., 2023), and others. A similar trend has been observed for various obstetrical diagnoses, such as pre-eclampsia (Honigberg et al., 2023; Tyrmi et al., 2023), gestational diabetes (Pervjakova et al., 2022), hyperemesis gravidarum (Fejzo et al., 2018), ectopic pregnancy (Pujol Gualdo et al., 2023), timing of parturition (Solé-Navais et al., 2023) and others. Additionally, significant advances have been made to the study of genetic risk factors underlying sex hormones affecting both men and women, with sample sizes reaching hundreds of thousands of individuals for hormones such as serum levels of sex-hormone binding globulin (SHBG), follicle stimulating hormone, luteinizing hormone, total testosterone and estradiol (Ruth et al., 2020; Venkatesh et al., 2024), the latter study also including exome analyses from UKB (Venkatesh et al., 2024).

Towards the PRS development in the female reproductive genomics field, the PGS Catalog (https://www.pgscatalog.org/) contains 4223 PRS for 624 traits as of 26-02-2024. From those, when querying "Female reproductive system disease", we found 63 PRS (less than 2% of the total, and half of which are based on cancer or neoplasm traits), indicating that their development for gynaecological and obstetrical diagnoses has been scarce.

International collaboration has played a key role in advancing research efforts in these areas, with important initiatives created in recent years. Notably, the ReproGen consortium has established an international network of investigators, which has continued to identify genetic risk factors associated with puberty timing (F. R. Day et al., 2017) and age at menopause (Ruth et al., 2021). Furthermore, recently the consortium has also included whole exome sequencing (WES) analysis (Kentistou et al., 2023; Shekari et al., 2023). In addition to detecting and characterising novel genetic loci, this latter approach has also uncovered intriguing insights which demonstrate the value of genome-wide approaches. For instance, their analysis of WES data focused on nearly 100 genes which were historically reported monogenic causes of premature ovary insufficiency (defined as menopause before 35 years old, which affects 1% of women). This study ruled out even modest penetrance for those supposedly pathogenic variants, in other words, almost all pathogenic variants were also found in reproductively healthy women when examined on a biobank scale in the UKB (Shekari et al., 2023). This finding underscores the importance of transitioning from narrower candidate gene studies to broader genome-wide scans with rigorous statistical and study design frameworks to better understand how genetics shapes health and disease susceptibility. This shift in both research and clinical environment is crucial as it reflects the complex nature of both complex and rarer disease development, likely involving a spectrum of genetic factors (so-called polygenic or oligogenic nature) rather than simple black-and-white explanations.

In addition, a Nordic Collaboration for Women's and Reproductive Health (NCWRH) has been formed recently to promote genomics and other 'omics' studies in reproductive health. This collaboration brings together researchers from

Finland, Denmark, Norway, Sweden, UK and Estonia, and has succeeded in mapping the genetic susceptibility to common female reproductive conditions, such as postpartum haemorrhage (Westergaard et al., 2023) and have recently presented and characterised GWAS meta-analyses for female and male infertility (Venkatesh et al., 2024). In recent years, many disease-specific consortiums and collaborations have been created, for instance, focusing on the genomics study of PCOS (F. Day et al., 2018), endometriosis (Rahmioglu et al., 2023), and gestational diabetes (Pervjakova et al., 2022).

Despite the progress made in elucidating the genetic underpinnings of some female reproductive health traits, many diagnoses relevant to women's reproductive health remain inadequately and incompletely characterised at the genetic level – a main motivation to plan the current thesis. Additionally, amongst all present biobanks, this thesis focuses mostly on EstBB and FinnGen, which are valuable assets for the study of those traits since firstly, they present an age structure that captures the female reproductive traits across the lifespan compared to other similar projects such as the UKB, which better captures the post-reproductive age, and secondly, they hold a specific genetic makeup, compared to other projects based in other populations, such as the UKB. Leveraging the wealth of data from large biobanks holds promise for decoding genetic associations of female reproductive health traits, shedding light on affected biology and potentially informing personalised approaches to enhance women's reproductive health and its healthcare.

2. AIMS

The primary objective of this thesis was to investigate the genetic underpinnings of three distinct female reproductive traits (POP, PCOS and AMH levels). These traits are common, multifaceted, are to some extent heritable and represent different stages of female reproductive health across the lifespan. In addition, we aimed to assess the potential applications of PRS within the framework of population-based biobanks linked with electronic health records.

The specific objectives of this thesis were the following:

- To identify and characterise genetic variants and genes that are associated with common female reproductive traits such as POP and AMH levels through genome-wide association studies and meta-analyses including diverse European populations. Additionally, we aimed to identify enriched biological pathways and tissues in relation to the discovered genetic signals through *in silico* analyses, as well as to explore potential genetic and phenotypic associations with various traits.
- To identify and characterise genetic variants and genes that are associated with common female reproductive traits such as PCOS and AMH levels through genome-wide association studies and meta-analyses focusing on population-specific biobank resources with features of the Nordic healthcare system such as the EstBB and FinnGen, and prospective population birth cohorts such as NFBC1966.
- To assess the utility of polygenic scores for *a)* enhancing risk prediction in combination with clinical risk factors for POP and *b)* as means to investigate trait comorbidities for PCOS.

3. RESULTS AND DISCUSSION

3.1. GWAS meta-analysis of data from population-based biobanks and cohorts aids the discovery of genetic risk factors and informs biological background of female reproductive traits (Ref. I, Ref. III)

Thanks to the existing resources and that the power of discovery is getting a boost by aggregating data from different biobanks, many genome-wide loci have been detected in meta-analyses of genetic association studies across several health traits. However, as described in Chapter 1.1 in more detail, female reproductive traits have largely remained understudied and underpowered. Therefore, there is the need to join different resources to achieve larger sample sizes to enhance the identification of genetic variants associated with those traits, as well as to nominate genes, test potential pathways and tissues enriched for the GWAS signals and to explore the (genetic) relationship with other human traits. In this part of the thesis, I will present the main results for the largest GWAS meta-analysis for POP and its respective postGWAS follow-up (ref. I) and the largest GWAS meta-analysis for AMH levels in pre-menopausal women and its respective postGWAS follow-up analysis (ref. III).

In POP, for example – the only meta-analysis so far included 15,010 cases and 340,734 controls from an Icelandic cohort (3,409 cases from an Icelandic hospital-based register with respective International Classification of Diseases (ICD) ICD10 or ICD9 codes and 131,444 controls which were recruited through different genetic research projects at deCODE genetics) and the UKB (including 11,601 cases with ICD10 code N81 in hospital inpatient records and 209,228 controls) (Olafsdottir et al., 2020). They identified eight variants at seven loci that highlighted the role of connective tissue metabolism and estrogen signalling in POP, which explained just a fraction (12.4% (95% CI 9.9–14.8%) of the total heritability estimated for the condition (43%) based on a twin study (Altman et al., 2008). The potential of highlighting other loci and biological mechanisms extends further with including other biobanks allowing larger genetic studies for this condition. Larger meta-analysis including different cohorts give an additional possibility to explore risk stratification strategies by developing and testing PRS as I will discuss in Chapter 3.3.

In the case of AMH, three previous GWASs have identified a few loci associated with AMH levels in pre-menopausal women (Ruth et al., 2019; Schuh-Huerta et al., 2012; Verdiesen et al., 2022) with sample sizes ranging from 232 to 7,049 and up to four loci identified in association with AMH levels. Recently meta-analyses have also included data from younger women, which is advantageous to enhance the power of detecting additional signals since it brings more variation to the distribution of AMH-values (younger pre-menopausal women tend to have more variability in AMH-values compared to older pre-menopausal women, who tend to have lower and more similar AMH levels between them, as

exemplified in Figure 1 in Chapter 1.2) (Verdiesen et al., 2022). Therefore, an updated larger meta-analysis including a younger aged cohort is of interest, aiming to unravel potential new loci associated with AMH level variation, and to enhance the biological interpretation underlying these signals.

3.1.1. Description of methods

The following section explains the methods implemented in the first part of ref. I and in ref. III.

In ref. I, the GWAS meta-analysis for POP included 28,086 women with POP and 546,291 controls of European ancestry. The meta-analysis included three studies: firstly, results from an Icelandic and UKB GWAS meta-analysis were obtained upon request (15,010 cases, 340,734 controls) (Olafsdottir et al., 2020), secondly, we downloaded GWAS summary statistics from the FinnGen study from Release 3 (https://www.finngen.fi/en/access_results) (5,518 cases, 43,366 controls) (Kurki et al., 2023) and thirdly we performed a GWAS in the Estonian Biobank (7,896 cases, 118,865 controls) (see Figure 6). In the three studies, cases were defined as women having POP using specific ICD codes: ICD-10: N81, ICD-9: 618 and ICD-8: 623 upon availability. Controls were defined as women who did not have the respective ICD codes. In short, all EstBB participants were genotyped using Illumina GSAv1.0, GSAv2.0, and GSAv2.0_EST arrays at the Core Genotyping Lab of the Institute of Genomics, University of Tartu. Standard quality control and imputation steps were previously done by the Bioinformatics Core Lab of the Institute of Genomics, University of Tartu, and were applied as previously described (Pujol-Gualdo et al., 2022). In the EstBB, association analysis was carried out using SAIGE software to implement a mixed logistic regression model with year of birth and 10 PCs as covariates in step I (Zhou et al., 2018). SNPs with poor imputation quality (INFO score < 0.4) and minor allele count <5 were excluded.

We used GWAMA (Mägi & Morris, 2010) to run the meta-analysis, using inverse variance weighted fixed-effects method with single genomic control correction. Next, we used the FUMA platform to further annotate the GWAS meta-analysis using data from several databases (Watanabe, Taskesen, van Bochoven, et al., 2017). FUMA identified lead SNPs with $p \leq 5 \times 10^{-8}$ and merged into genomic loci within 500 kb. Gene prioritization criteria involved the inclusion of different data layers such as: a) proximity from the gene to the association peak (prioritising nearest genes), b) results from colocalization analysis implemented by COLOC R package (Giambartolomei et al., 2014) c) considering the presence of coding variants amongst the lead signals or variants in high LD (r2) with the lead variants using ANNOVAR (K. Wang et al., 2010) and d) if our candidate genes in mutant mice were displaying relevant phenotypes by querying the Mouse Genome Database (http://www.informatics.jax.org). MAGMA and DEPICT implemented in FUMA and CTG-VL (https://genoma.io/), respectively, were used to run gene-set analysis and tissue expression analysis across multiple

datasets. Summary statistics were used for genetic correlation testing using hundreds of traits available in LD-Hub portal (Zheng et al., 2017).

In ref III, we performed a GWAS meta-analysis for AMH measurements using GWAMA (Mägi & Morris, 2010) we combined 2,619 AMH measurements (at age 31 years old) from the prospective founder population cohort NFBC1966 with a previous GWAS meta-analysis that included 7,049 pre-menopausal women (spanning age range 15–48) (Verdiesen et al., 2022). NFBC1966 is a population-based prospective study in northern Finland, following individuals from birth in 1966 through ages 1, 14, 31, and 46. In NFBC1966, AMH measurements at age 31 were quantified using an automated assay (Elecsys® AMH Plus (Roche)). Meta-analysis was carried out as described in the previous paragraph. We annotated the genetic variants, combined different data layers to prioritise potential candidate genes, described significant pathways and tissues enriched by the GWAS signals, similarly as described before and mainly implementing FUMA. Colocalization analysis was run to identify plausible eQTL overlap using HyPrColoc and datasets from the eQTL catalogue (Foley et al., 2021; Kerimov et al., 2023). Publicly available summary statistics from the CTG-VL (https://vl.genoma.io/) and the GWAScatalog (Buniello et al., 2019) were used, respectively, to assess broad genetic correlations and single variant-level associations with multiple traits.

3.1.2. Discovery of novel genetic risk factors for pelvic organ prolapse (POP) (Ref. I)

In Ref. I, we conducted the largest GWAS meta-analysis for POP, which identified 26 genetic loci, including 19 novel ones (representing almost a 4-fold increase in the number of identified loci compared to the previous study) (Olafsdottir et al., 2020) (see Figure 6). Considering a prevalence of 40% for overall POP, the SNP-heritability was estimated to be 14.3%. Amongst these associations, we detected seven loci which were already reported in the previous work (near *WNT4*, *GDF7*, *EFEMP1*, *FAT4*, *IMPDH1*, *TBX5*, and *SALL1*).

Beyond those genes, we further proposed several previously unidentified candidate genes which reinforce the role of connective tissue molecular changes as a key process in the pathogenesis of POP (*LOXL1, ADAMTS5, CHRDL2*, *ACADVL*, *PLA2G6*) (Kerkhof et al., 2009; V. F. Lim et al., 2014). For instance, we were able to detect for the first time in GWAS a candidate gene previously proposed based on its association with POP in mice, *LOXL1*. Liu *et al.* described that mice lacking the protein LOXL1 (*Loxl1*$^{-/-}$) do not deposit normal elastic fibers in the uterine tract postpartum and develop POP (X. Liu et al., 2004), an association which has been furthered supported in diverse mouse and human studies (Jameson et al., 2020; Kow et al., 2016; Neupane et al., 2014). Another example which reinforces the alteration of connective tissue is the nomination of *ADAMTS5* in association with POP. *ADAMTS5* is part of the A Disintegrin And Metalloprotease with Thrombospondin Motifs (ADAMTS) family (V. F. Lim et al., 2014), which has diverse roles in tissue morphogenesis and pathophysiological

remodelling. ADAMTS members have also been associated with inguinal hernia through GWAS (Fadista et al., 2022), in line with our observation of positive genetic correlations between POP and inguinal hernia, as well as other connective tissue abnormalities such as gastroesophageal reflux, diverticular disease, osteoarthritis and hiatus hernia. Gene set enrichment results supported these observations by uncovering significant associations with various terms related to connective tissue development, such as chondrocyte differentiation (Figure 6).

Our study also underscores the urogenital development as a key process in the pathogenesis of POP highlighting previously unidentified genes such as *DVL2*, *WT1* and *HOXD13*. For instance, *DVL2* is a component of the Wnt signaling pathway (a shared pathway with the previously reported gene *WNT4*), important for epithelial tissue development and renewal, embryonic development of the sex organs, regulation of follicle maturation and controlling steroidogenesis in the postnatal ovary (Hernandez Gifford, 2015; Pitzer et al., 2021; Sharma et al., 2018; G. Zhang et al., 2017). In that way, we hypothesise the Wnt pathway to have a dual role in POP development, both by regulating organogenesis and hormonal support of tissue function, which may be protective for POP (Moalli et al., 2004). However, future studies with more granular phenotypes, such as taking into account the pre/postmenopausal status, the use of hormone replacement therapy and/or considering future genetic studies assessing sex-specific oestradiol exposure, are needed to study the effects between oestradiol lifetime exposure and POP development. Additionally, we found that gene expression of our signals identified from GWAS signals are significantly enriched in pathways such as "in utero embryonic development" and tissues such as "Cervix/ectocervix", "Uterus", "Embryoid bodies" and "Smooth muscle". The enrichment of gene expression in tissues such as "Cervix/ectocervix," "Uterus," and "Smooth muscle" supports our hypothesis since these observations align with the anatomical locations primarily affected by POP. Specifically, the cervix and uterus are crucial structures in providing support to the pelvic organs, while smooth muscle dysfunction contributes to weakened pelvic floor support, a hallmark of POP. Furthermore, the enrichment in pathways related to "in utero embryonic development" suggests a developmental component to POP susceptibility (Figure 6), indicating that genetic factors influencing early embryonic tissue formation may also predispose women to POP later in life. Highlighting tissues and pathways sheds light on the molecular mechanisms driving POP, informs selection of tissues/cell types for functional follow-up studies and also provides potential targets for therapeutic interventions which in a future may mitigate the risk of POP development.

Furthermore, *WT1*, another novel finding from this study, is a transcription factor involved in urogenital system development, which was identified to be active in fibroblasts of severe anterior vaginal prolapse in a single-cell transcriptome study (Yaqian Li et al., 2021). Moreover, *WT1* is also involved in cardiac development and disease (Duim et al., 2015, 2016; Velecela et al., 2013). In this line, a notable portion of the GWAS associations and potential candidate genes we identified also point towards a link between metabolic and cardiovascular health and POP (*KLF13*, *DUSP16*, *MAFF*, *VCL*, and *LDAH*) (Darwich et al., 2017;

Ferguson et al., 2019; Fukuda et al., 2019; Goo et al., 2014; Kimura et al., 1999; Lavallée et al., 2006; W. Li et al., 2020; Massrieh et al., 2006; Saliba et al., 2019; Von Scheidt et al., 2021). Beyond the epidemiological observation that BMI is a risk factor for POP (Vergeldt et al., 2015), to our knowledge this study is the first study to suggest a genetic link between POP and BMI, and the first to show a potential link between POP and other metabolic and cardiovascular traits, while it cannot be ruled out that these associations are being driven mainly through BMI effect.

Genetic correlation analysis supported a mild positive correlation between POP and cardiovascular and metabolic phenotypes such as angina, diabetes, triglycerides and CAD, as well as increasing BMI. Beyond these associations, genetic correlation studies mirrored well the findings of epidemiological studies, showing associations with number of births, previous hysterectomy, younger age at first birth, constipation, occupations including heavy lifting, and connective tissue disorders (Cartwright et al., 2015; Mant et al., 1997; Vergeldt et al., 2015). Additionally, a positive correlation was detected with occupations involving walking and/or standing and heavy physical work, which might have potential value in the counselling of women with higher risk to develop POP.

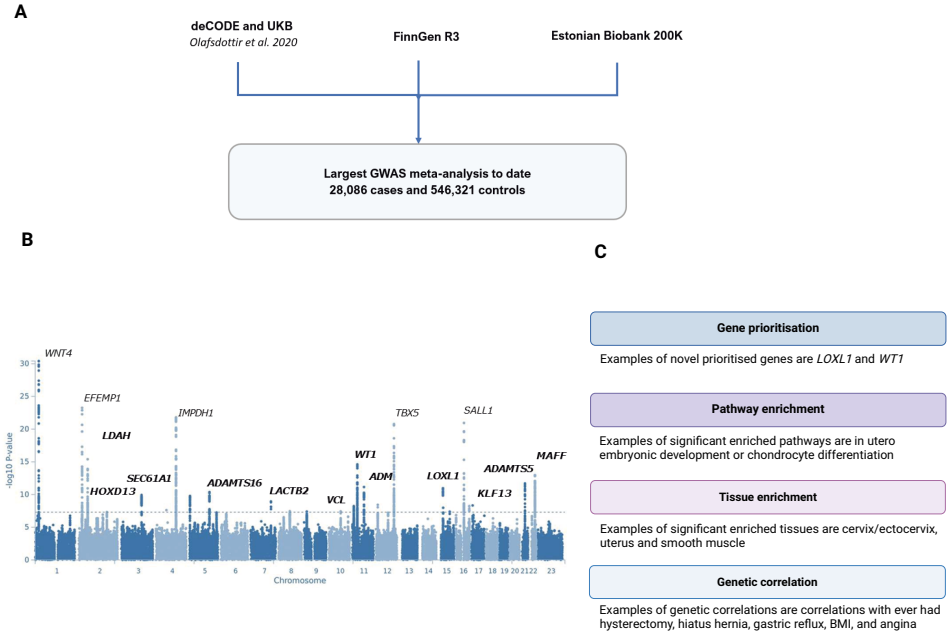

Figure 6. Summary of GWAS meta-analysis and postGWAS results for POP. A) Cohorts included in the GWAS meta-analysis for POP B) Manhattan plot of GWAS meta-analysis for POP and C) summary of postGWAS results. Created with Biorender.com and adapted from Pujol-Gualdo et al. 2022

3.1.3. Discovery of novel genetic risk factors for AMH levels (Ref. III)

In Ref. III, we identified six loci associated with AMH levels at $P < 5 \times 10^{-8}$, including the previously reported *MCM8*, *AMH* and *TEX41* loci (Verdiesen et al., 2022), and three novel signals near *CHEK2*, *BMP4* and *EIF4EBP1* (Figure 7C).

When moving from lead variants to potential associated genes, in three of the six loci the lead variants were either coding variants or variants in complete LD with coding variants, which generally facilitates the candidate gene nomination as their effects to the protein tend to be more interpretable. Therefore, we pointed to a frameshift variant in *CHEK2* (c.1100delC), in complete LD with the lead variant in that locus), a missense variant in *MCM8* (rs16991615), and a missense variant in *AMH* (rs10417628) to be the potential drivers of these respective signals. *CHEK2* is a DNA damage repair gene which encodes a checkpoint kinase that plays a role in cell cycle regulation, including in oocytes with DNA damage. I will zoom in on the biological role of that gene in the next chapter. The current GWAS meta-analysis also supported a previously identified DNA damage repair gene, *MCM8* (Park et al., 2013), in association with AMH levels (Ruth et al., 2019; Verdiesen et al., 2022). A missense variant was likely driving this signal, which has previously also been described in association with age at menopause, infertility and cancer (F. R. Day, Ruth, et al., 2015; Griffin & Trakselis, 2019; Lutzmann et al., 2019; Michailidou et al., 2017). Interpreting the association with the missense variant in *AMH* is less clear because of a potential technical issue on the presence of this variant and epitope recognition in the assays (Hoyos et al., 2020). A case report by Hoyos *et al.* suggested that the amino acid substitution corresponding to rs10417628 in *AMH* gene (same lead variant reported in the meta-analysis for AMH levels) reduces AMH detection by the picoAMH assay from Ansh Labs without influencing AMH bioactivity (Hoyos et al., 2020). Further research is needed to validate the association between rs10417628 and AMH levels using individual-level data from different assays and different groups of women with different genotypes for rs10417628 (and potentially other AMH variants).

Upon querying the AMH GWAS lead variants in the GWAS Catalog, we found shared significant signals with traits that have known epidemiological associations with AMH levels, such as breast cancer, postmenopausal status (age at menopause, estradiol levels, heel bone mineral density), and cardiovascular disease (de Kat et al., 2017; W. Ge et al., 2018; Homburg & Crawford, 2014; Moolhuijsen & Visser, 2020). Additionally, we observed novel associations with other traits in European ancestry studies, including uterine fibroids (coming likely from signal in *MCM8*) and blood cell counts (coming from signals near *CHEK2*, which has been also associated recently with clonal haematopoiesis (Kar et al., 2022).

Three other non-coding variants were identified as lead variants and we used colocalization analysis to infer plausible genes influencing AMH levels in these loci. Colocalization analysis supported regulatory effects of GWAS variants on the expression of three genes: *TEX41 (*Testis Expressed 41 gene which showed altered gene expression in the testis), *BMP4* (a known regulator of AMH)

and *EIF4EBP1* (a gene which shows association with some oncogenic processes). Therefore, colocalization analysis provided a refined characterization of possible regulatory effects of the genetic variants on different genes.

Gene set enrichment analysis highlighted renal system vasculature morphogenesis, glomerulus vasculature morphogenesis and glomerulus morphogenesis, likely reflecting the close connection between urinary and reproductive system development, originating from a common embryological precursor, the intermediate mesoderm, where *BMP4* signalling pathway is a key player during kidney development, including ureteric bud outgrowth (Grinspon & Rey, 2014; Nishinakamura & Sakaguchi, 2014; Oxburgh et al., 2014). Tissue expression analysis yielded the strongest enrichment in the pituitary gland, even though it was not statistically significant after multiple testing correction. This is in line with research from recent years showing that AMH has versatile actions in different levels of the HPG axis (Silva & Giacobini, 2021), further supporting the developmental alterations of neuroendocrine circuits regulating fertility.

From genetic correlation analysis, traits reflecting menopausal timing displayed almost complete genetic correlation, highlighting a high genetic similarity, which reinforces the use of AMH as a marker for ovarian reserve and overall reflects reproductive ageing and menopause timing. Additionally, we obtained a range of nominally significant associations such as with breast cancer. In the variant look-up, we observed that two of the lead variants in *CHEK2* and *MCM8* signals, were also identified in a previous GWAS for breast cancer (Michailidou et al., 2017), in concordance with epidemiological studies that supported an association between AMH levels and breast cancer (W. Ge et al., 2018).

3.1.4. General limitations and future directions

While GWAS is an excellent tool for exploring the genetics of complex traits, there are common limitations to consider, which also affect the studies presented. Notably, the findings are based on populations of European ancestry, potentially restricting the applicability of results to other ancestry groups with varying genetic make-up and therefore varying genetic effects. Prior family/registry studies suggest higher overall heritability for POP (43%) (Altman et al., 2008) compared to our current SNP-based estimate (14.3% (95% CI 5,9% – 22,7%), indicating that a significant portion of genetic variation remains unexplored. Unaccounted genetic variations, such as rarer variants, copy number variations, or gene-gene interactions, may explain partly this missing heritability. In the case of AMH, there is a lack of family or registry studies estimating its heritability, and in this case our GWAS estimated a SNP-heritability to be 13% (s.e. 5%), which indicates that there are likely more SNPs that contribute to the variability in AMH levels.

When moving from variants to potential candidate genes, there is no standardised or gold-standard way to do so. However, recently, methods have been developed to move from variants to genes in a more systematic way: for instance more recently the PoPS method, that uses trait-relevant gene features, such as cell-type-specific expression, to prioritise genes at GWAS loci (Weeks et al., 2023).

However, I believe combining various data sources and considering the reported literature and underlying gene biology (if known) remains the most effective strategy, especially for female reproductive conditions where cell-type specific expression datasets are scarce. Limitations in eQTL mapping also affect colocalisation analysis and involve also the lack or shortage of relevant tissues amongst reproductive tissues. While a complete catalogue of female reproductive tissues is absent, in GTEx, we found tissues similar to female reproductive tract tissues based on cellular composition and gene expression (vagina, uterus, oesophagus mucosa and gastro-oesophageal junction, sigmoid colon, skin, salivary gland and tibial nerve). Additionally it has recently been shown that there is a minimal overlap between GWAS variants and disease-relevant eQTLs, due to incomplete discovery on both sides (Mostafavi et al., 2023). Thus, while our studies identify candidate genes and pathways, further functional follow-up is also necessary to understand the regulatory functions of the uncovered loci and validate the genes nominated as potentially biologically relevant in respective POP and AMH level studies. A multi-disciplinary approach, including generation of cellular and organoid models for in vitro systems that resemble women's reproductive tract tissue, mouse models of genetic risk factors identified and CRISPR-Cas9 gene editing in human cells are potential avenues to functionally validate the given targets.

Specifically for ref. I, one limitation is that there might be potential heterogeneity in the phenotype definition, Firstly, we cannot rule out potential phenotypic misclassification (future POP cases among controls, especially in younger cohorts such as the Estonian Biobank (age range: 18–104, mean=42 and median=43) and secondly, there is potential phenotypic heterogeneity considering the sources of diagnoses, which vary across studies. Whilst in FinnGen, Icelandic data and UKB the diagnoses are obtained from in-hospital registers, in EstBB information on diagnoses was obtained by Estonian Health Insurance Fund which offers a good coverage and representation of the Estonian health care situation and includes diagnoses made by general practitioners (GPs) and specialist doctors. However, it is likely that in-patient hospital records are capturing a more severe manifestation of the condition.

With future increasing sample sizes, beyond detecting more genetic risk factors underlying overall POP, a potential added value from the use of genetics would come from stratification of the sample according to disease severity, which would enable to estimate disease progression trajectories and/or different severity stages, aiming to seek for differential genetic markers that lead to the need of surgery. Similarly, if sample sizes are big enough, future analysis could be stratified by age of onset of prolapse or even seek for any differences in genetic architecture when separating groups by prolapse type (for instance rectum prolapse, bladder o uterine prolapse), which may also expand our knowledge of the aetiology and affected biology underlying these subtypes.

Specifically for ref. III, one limitation is the lack of standardisation in the measure of AMH. The clinical importance of AMH measurement has led to the development of several AMH assays, however, direct comparison of AMH values

obtained by these assays is still problematic (Moolhuijsen & Visser, 2020). An interesting avenue for future research could be to assess AMH trajectories across age leveraging repeated measurements, as Venkatesh *et al.* have done successfully for adiposity-change in adulthood (Venkatesh et al., 2023) and would give a more accurate estimate of ovarian reserve and reproductive ageing trajectories.

While microarray genotyping, commonly used in GWAS, may not adequately capture rare genetic variants due to its design limitations, WGS provides a comprehensive view of the genome, enabling detection of coding and non-coding genetic variation, and surveying complex regions which are difficult to genotype. WGS is preferred over WES and microarray, and is expected to become the method of choice over the next years with the increasing availability of low-cost WGS technology and the achievement of landmark resources such as UKB which have started to investigate deepen the understanding of how genetics influences disease biology for the study of human biology and health (S. Li et al., 2023). Also creating consortia for collaborative efforts across studies (similar to what has been done for PCOS (F. Day et al., 2018) or gestational diabetes (Pervjakova et al., 2022), for these specific traits could amplify sample sizes, enhancing the ability to detect new genetic risk factors.

3.2. Specific population-based biobanks and cohorts: potential for the detection of population-enriched alleles (Ref. II, Ref. III)

As exemplified in the last chapter, the meta-analysis of different biobank and cohort studies has become an important source of genetic discoveries underlying complex traits. However, the implementation of GWAS within specific population-based biobanks holds an added potential; to uncover unique genetic insights by leveraging a different genetic make-up in certain populations, along with the availability of structured and digitalised national healthcare data. Owing to increased genetic drift, isolated populations with recent bottlenecks can have deleterious, disease-predisposing alleles at considerably higher frequencies than permitted by selection in larger and older outbred populations (Kurki et al., 2023). As a result, isolated populations, such as the one in Finland and to lesser extent Estonia, provide an opportunity to identify high-impact disease variants that are rare or nonexistent in other populations. Of note, these genetic discoveries would escape detection in standard GWAS meta-analysis directed to other populations, unless very extensive sample sizes would be reached or more expensive technologies such as WGS would be implemented.

Building upon these insights, in the following chapter, I aim to provide an example on how focusing on specific population-based biobanks and cohorts facilitates the identification of low-frequency deleterious alleles in *CHEK2* in association with higher risk for PCOS and higher AMH levels in women.

While our meta-analysis for AMH levels in pre-menopausal women represents the largest to date (*n*=9,668), our PCOS study builds upon the pioneering work of the first GWAS meta-analysis for PCOS within an international consortium framework (n=10,074 women with PCOS and 103,164 controls) (F. Day et al., 2018), which I will refer to from that point on as the 'PCOS consortium GWAS'. Amongst these meta-analysis results, the authors detected three novel loci near *PLGRKT*, *ZBTB16*, and *MAPRE1* and 11 previously reported loci which were detected in independent works (Chen et al., 2011; F. R. Day, Hinds, et al., 2015; Hayes et al., 2015; Y. Shi et al., 2012).

3.2.1. Description of methods

In Ref. II, we used data from the FinnGen study and the EstBB. We ran two GWASes: firstly, a GWAS in FinnGen, and secondly a GWAS in EstBB. Then, we ran a meta-analysis of both studies, aiming to identify and characterise genetic variants associated with PCOS.

FinnGen included genomic information from 141,355 women (6% of the female Finnish population) from the Release 6 (R6). In FinnGen, PCOS cases were identified using specific ICD codes: E28.2, ICD-9 code 256.4, or ICD-8 code 256.90, and controls were defined upon the absence of these codes, resulting in 797 cases and 140,558 controls. In brief, FinnGen employed Illumina and Affymetrix arrays for genotyping, followed by imputation and association analyses using SAIGE (Zhou et al., 2018). In the EstBB, we utilised the 150K data freeze for this analysis. In the EstBB, cases were identified using the ICD-10 code E28.2, with controls comprising female participants without this ICD-10 code, resulting in a total of 2812 women with PCOS and 89,230 controls. EstBB participants were genotyped and association analyses were conducted as explained in the previous Chapter 3.1.1. (Zhou et al., 2018).

An inverse-weighted meta-analysis was performed using the METAL software (Willer et al., 2010), involving 3609 cases and 229,788 controls. FUMA platform was utilised for candidate gene mapping and functional annotation, considering loci within ±1000 kb of the top variants. Gene prioritisation criteria included the identification of coding variants in high LD with the lead variants and the existence of already reported genes with a function relevant to PCOS pathophysiology. Colocalization with variants affecting gene expression was explored using the following pipeline and datasets from the eQTL Catalogue (https://github.com/eQTL-Catalogue/colocalisation) (Kerimov et al., 2021).

Conditional analyses were also conducted aiming to identify secondary association signals at the loci identified on chromosome 22, using SAIGE (Zhou et al., 2018) and a stepwise approach.

Next, an additional association analysis including BMI as a covariate was conducted in a) FinnGen including 60–65% of the original discovery sample and b) the EstBB; including approximately 75% of the validation sample. We then performed a second BMI-adjusted meta-analysis, including 2619 cases and 160,321 controls.

The study design, cohorts used and methods implemented in the GWAS meta-analysis for AMH levels in pre-menopausal women (Ref. III) have been given in the previous chapter (see point 2.3.1).

3.2.2. The discovery of low frequency alleles in association with PCOS and AMH levels (Ref. II, III)

In ref. II, the GWAS in the FinnGen study uncovered two previously reported loci (close to *ERBB4* and *DENND1A*) and one previously unreported large effect association was found on chromosome 22 (lead variant rs145598156, intronic region nearby the *CHEK2* gene) (Figure 7). Functional characterization of this locus revealed a frameshift variant, c.1100delC (rs555607708, $p = 1.68 \times 10^{-9}$, OR = 13.46 (95% CI 5.68–31.89) in *CHEK2*, with a high LD ($r^2 = 0.95$) with the lead variant (Figure 7B). Interestingly, the protein-truncating variant c.1100delC has a 3.7-fold enrichment in the Finnish population compared to non-Finnish, non-Estonian Europeans (minor allele frequencies being 0.008 for the Finnish population compared to 0.003 for the Estonian population and 0.002 for other European populations (Karczewski et al., 2020). The analysis conditioned on c.1100delC resulted in no genome-wide significant associations in this locus, with a $p = 3.29 \times 10^{-4}$ for the lead variant rs145598156, supporting that the signal was likely driven by c.1100delC.

In the validation GWAS in EstBB, this analysis also uncovered a significant association in the same region at chromosome 22. The lead variant rs182075939 showed to be tightly linked with a missense variant rs17879961 ($r^2 = 0.83$, $p = 4.23 \times 10^{-12}$), known as I157T, in *CHEK2* (Figure 7). Interestingly, the EstBB lead variant rs182075939 presents a higher MAF (0.048) in Estonian population compared to 0.029 for the Finnish population and 0.0025 for other European populations (Karczewski et al., 2020). The analysis conditioned on I157T resulted in no genome-wide significant associations in this locus, with a $p = 0.04$ for the lead variant rs182075939, supporting I157T as a likely driver of the association (Figure 7).

In the meta-analysis, beyond replicating both associations on chr22, a third novel signal (rs9312937, $p = 1.7 \times 10^{-8}$, OR = 1.16 (95% CI 1.10–1.23), MAF = 0.44) was also detected in an intronic region of the *MYO10* gene. The variant rs9312937 was previously identified in association with age at menarche (Kentistou et al., 2023), and this genetic region has been associated with T2D (Salonen et al., 2007) and metabolic syndrome traits (Yi Zhang et al., 2013). Moreover, two signals on chromosome 11 that were previously reported in the PCOS consortium GWAS (near *FSHB* and *ZBTB16*) were also replicated.

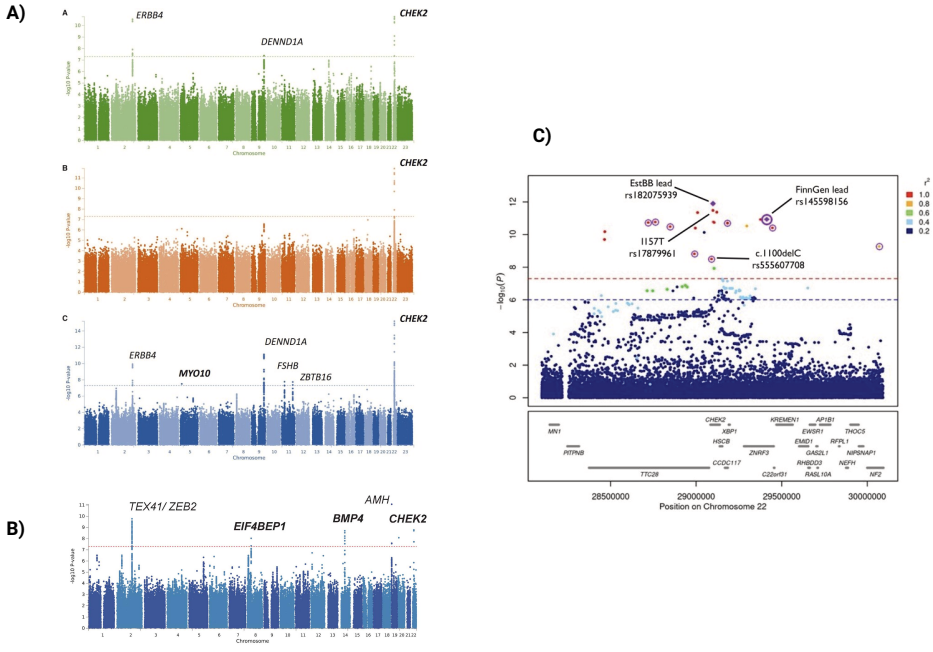

Figure 7. A) Manhattan plots of the GWAS for PCOS in FinnGen (green), EstBB (orange) and meta-analysis of both studies for PCOS (blue) B) Manhattan plot of the GWAS meta-analysis for AMH levels C) Regional plot of the *CHEK2* locus in chromosome 22. Created with Biorender.com and adapted from Tyrmi et al. 2021

When adding BMI as a covariate in each independent GWAS and meta-analysing the BMI and age-adjusted results, only the EstBB lead signal (rs182075939) passed the genome-wide significant threshold. While the FinnGen dataset had more missing data for BMI measurements (keeping only 60–65% of sample compared to EstBB where 75% of the total sample size could be kept), we cannot rule out the possibility that there might be an interaction between obesity and c.1100delC. However, an interaction between BMI and PCOS-associated variants has previously been suggested (Wojciechowski et al., 2012), and the c.1100delC variant in *CHEK2* has recently been shown to predispose particularly obese carriers to the development of breast cancer (Greville-Heygate et al., 2020). Despite those suggestions that there might be an interaction between obesity (BMI) and the c.1100delC variant on breast cancer risk, we cannot support this with our current analysis and further research would be needed to explore the potential interaction between the c.1100delC variant, obesity, and their combined impact on breast cancer risk.

In our GWAS meta-analysis for AMH levels in pre-menopausal women (n=9,968), we detected for the first time an intronic variant in the same locus (22q11). The lead variant showed to be in complete LD (r^2=1) with c.1100delC ($p = 6.60 \times 10^{-05}$ beta=0.79, se=0.12 in NFBC1966), showing to be associated with higher levels of AMH (Figure 7, Table 1).

Table 1. **Variant results for *CHEK2* variants in the main independent population GWAS studies for Ref. II and III.** EA: Effect allele, LD: Linkage Disequilibrium, OR: Odds Ratio, MAF: Minor Allele Frequency, VEP: Variant Effect Predictor, Finn: Finnish, Est: Estonian, EUR: European

Variant / EA	Pheno-type	LD with lead variant (r^2)	Main GWAS study	p-value	Effect size OR (95% CI) or BETA (SE)	MAF (FinnGen/ EstBB/ Other EUR)
c.1100delC / A	PCOS	0.95	FinnGen	1.68×10^{-9}	13.46 (5.68–31.89)	0.008/0.003/ 0.002
I157T / G	PCOS	0.83	EstBB	4.23×10^{-12}	1.53 (1.45–1.66)	0.029/0.048/ 0.0025
c.1100delC / A	AMH levels	1	NFBC1966	6.60×10^{-05}	Beta= 0.79 (s.e. 0.12)	0.008/0.003/ 0.002

3.2.3. The case of *CHEK2*: Biological functions, limitations and future directions (Ref. II, III)

In Ref. II and III, *CHEK2* stands out as a novel and interesting finding associated with both PCOS and AMH levels in pre-menopausal women from our GWAS meta-analyses. What defines *CHEK2* and what might be its biological functions in the context of PCOS and AMH?

CHEK2 (checkpoint kinase 2) encodes a tumour suppressor kinase which acts as a mediator of DNA damage signalling in response to double-stranded DNA breaks, being an important factor in the quality control of cells by causing cell cycle arrest and apoptosis damage (Ahn et al., 2004). If CHEK2 function is disturbed, DNA repair is imbalanced, which can lead to genomic instability and tumorigenesis (Mustofa et al., 2020). *CHEK2* is well-known for its association with breast cancer (Dorling et al., 2021; Wilcox et al., 2023).

Interestingly, CHEK2 also plays a crucial role in foetal oocyte attrition, a phenomenon through which 80% of the initial ovarian oocyte reserve is lost during foetal development in mammals (Tharp et al., 2020). This gene was also highlighted in the landmark paper from Ruth et al., 2021, which reported hundreds of genetic variants associated with ovarian ageing (assessed as age at natural menopause). This GWAS meta-analysis highlighted loss-of-function variants in key DNA damage repair-associated genes such as *CHEK2* to be associated with older age at natural menopause. The authors went further to test how a knockout for *CHEK2* in female mice models would influence the ovarian reserve and its rate of depletion throughout the life course. They showed that *Chk2-/-* leads to a maximised ovarian reserve at postnatal day 2 and reduced follicle atresia, a higher number of ovulated metaphase II oocytes, and higher AMH levels at 13.5 months compared to wild type female mice (Ruth et al., 2021). Our finding is in line with

this research and supports the hypothesis that loss of function in *CHEK2* might result in decreased follicular atresia and higher AMH levels also in young pre-menopausal women, and also in women with PCOS. While *CHEK2* has been suggested to potentially be useful for developing therapeutics for ovarian stimulation in the IVF setting, we exemplify potential pleiotropic effects which should be further assessed, also including many other health traits, and acknowledge that despite its potential as a drug target, a long race would be ahead to confirm this finding, considering that the median gap between a report of the original genetic observation and approval of the derived, first-in-class therapeutic agent is of 25 years (range 4–38 years) (Trajanoska et al., 2023) and drug development heavily depends on the available technologies for activation/replacement/inhibition of any given protein or target.

In humans, a clear trend was observed that *CHEK2* loss-of-function alleles are associated with later menopausal age (Ruth et al., 2021). This would be in line with women with PCOS, as they also present with an increased ovarian reserve, higher AMH levels, even at later reproductive years, and delayed menopause (de Ziegler et al., 2018; Forslund et al., 2019; Minooee et al., 2018; Piltonen et al., 2005). A specific association between menopause-delaying alleles and PCOS has also been previously demonstrated by GWAS (F. R. Day, Hinds, et al., 2015) and more recent phenome-wide association studies (PheWAS) and exome data analysis further replicated our observation reporting a connection between *CHEK2* loss of function and higher odds of PCOS (Ruth et al., 2021; Ward et al., 2022). Additionally, recent research done with exome data has also identified an association between *CHEK2* truncating variants (excluding c.1110delC) and later age at menarche (Kentistou et al., 2023).

One could wonder if the association between *CHEK2* and breast cancer translates as a higher breast cancer risk in women with PCOS, however the epidemiological studies do not show an increased risk for breast cancer, only supporting an increased risk for endometrial cancer in women with PCOS and ovarian cancer (Barry et al., 2014; Ding et al., 2018; Dumesic & Lobo, 2013; Frandsen et al., 2023; Gottschau et al., 2015). The potential hypothesis that CHEK2 is highlighting a cell-type or otherwise context-specific finding in women with PCOS and higher AMH levels needs to be further investigated.

Three studies utilising a Mendelian randomization approach (including 11, 13 or 14 independent significant SNPs reported in the PCOS consortium GWAS (F. Day et al., 2018) and data from the Breast Cancer Association Consortium, with a total of 122,977 cases and 105,974 controls (Michailidou et al., 2017) or 133,384 cases and 113,789 controls (H. Zhang et al., 2020) have suggested a modest but significant causal effect between PCOS and breast cancer (Wen et al., 2021; Wu et al., 2020; T. Zhu et al., 2021). More comprehensive MR studies might shed more light into those associations.

Whereas the association of CHEK2 c.1100delC with a moderate-risk breast cancer predisposition is well recognised (Meijers-Heijboer et al., 2002), and has been reported as well in studies in the Finnish population (Hallamies et al., 2017; Kuusisto et al., 2011; Mars et al., 2020) the pathogenic role of I157T in breast

cancer remains controversial (Kilpivaara et al., 2004; Muranen et al., 2017; Schutte et al., 2003). Overall, larger sample sizes including WES/WGS linked with EHRs in both Finland and Estonia are needed to further evaluate the pathogenic role of those variants.

In the case of AMH, variations in age-specific circulating AMH levels have been linked to breast cancer (W. Ge et al., 2018). We also found a nominally significant association between our AMH GWAS summary statistics and PCOS, pointing towards a shared genetic background between AMH levels and breast cancer. We hypothesise both *MCM8* and *CHEK2* loci may play a role in this relationship. However, the fact that the risks do not seem to translate into clinical findings when looking at women with PCOS is notable and may indicate, for example, more efficient DNA damage repair systems in women with PCOS, a feature also associated with later onset of menopause (F. Day et al., 2018; Ruth et al., 2021).

To sum up, our study highlights the value of utilising comprehensive and population-specific genetic and health datasets (FinnGen and EstBB), which aided the identification of population-enriched variants such as those in *CHEK2*. If we had not looked at the Finnish and Estonian populations specifically, we would have missed the contribution of these important variants. This highlights the need to keep and study population specific initiatives, and also is a good demonstration of the extra value the study of largely under-explored populations such as non-European populations holds, which likely might bring many more interesting genetic discoveries.

3.3. PRS: a tool to both stratify risk and explore disease biology (Ref. I, Ref. IV)

Created based on GWAS results, PRS have been researched as a tool to stratify individuals into different risk groups in complex diseases (Khera et al., 2018). However, PRS have been scarcely researched for female reproductive traits (with the exception of breast cancer) and have never been used as a tool for predicting POP development. In contrast to physical examinations, PRS could serve as a tool to identify women with higher genetic risk of POP and direct preventive strategies even decades before symptomatic POP appears.

To a lesser extent, PRS has also been used to explore trait comorbidities. For the same purpose, in ref. IV, we used an optimised PRS for PCOS and examined associations between this PRS and cardiometabolic and androgenic phenotypes in men in the absence of ovarian function using data from the UKB and EstBB. GWAS studies supported that PCOS is associated with perturbation of presumed ovarian-related factors and non-ovarian factors (F. Day et al., 2018). However, it remains to be determined which of these are the inciting events and which are the secondary consequences. Historically, PCOS has been defined as a disorder of reproductive-aged women, with ovarian dysfunction as a key feature (Escobar-Morreale, 2018). However, many have since suggested that PCOS may not

always be primarily a disorder of the female reproductive system (F. Day et al., 2018; Di Guardo et al., 2020; Kurzrock & Cohen, 2007; Sam et al., 2008; Vipin et al., 2016; Yildiz et al., 2003; Yilmaz et al., 2018). Those observations of associations between PCOS and men's health outcomes suggest that genetic variants associated with PCOS may not act through alterations in ovarian function to influence the development of the cardiometabolic and androgenic features of PCOS. To this end, we tested the degree of association between a PRS for PCOS with several cardiometabolic traits in men.

3.3.1. Description of methods

The following section explains the methods implemented for the second part of ref. I and for ref. IV.

In ref. I., a PRS for POP using meta-analysis summary statistics from Icelandic and UKB study (Olafsdottir et al., 2020), and FinnGen datasets was constructed. The EstBB dataset was used as an independent target for selecting the most optimised PRS and for validation. PRS were computed using PRSice2 (Choi & O'Reilly, 2019) and LDPred softwares (Vilhjálmsson et al., 2015).

Next, the EstBB dataset was divided into discovery (5379 prevalent cases, 21,516 randomly selected controls) and validation sets (2517 incident cases, 96,109 randomly selected controls). Logistic regression models, adjusted for first 10 PCs and age, were employed to select the best-performing PRS in the discovery set and we kept with the PRS which showed the highest degree of association to case-control status. The validation set was further filtered for minimal missing clinical data (2104 cases and 24,753 randomly selected controls), to assess the predictive ability of PRS alone or in combination with five clinical risk factors (number of children, BMI, smoking, asthma and constipation).

Standardised PRS versions were categorised into percentiles, and hazard ratios corresponding to one standard deviation were estimated. Harrell's C-statistic characterised discriminative ability using age as a time scale. Survival modelling and Cox proportional hazard models were used to assess differences in genetic risk across age strata, where age was used as a time scale for properly accounting for left-truncation and right-censoring in the data. The study also evaluated the predictive ability of PRS alone compared to the five clinical risk factors alone or in combination, and as well assessed the predictive ability of a combined genetic and clinical model. Information on clinical risk factors such as the number of children, BMI, and smoking was extracted from questionnaire data in the EstBB and the ICD10 codes J45 and K59.0 were used for asthma and constipation definitions, respectively.

In ref. IV, the study involved two population-based cohorts: the UKB with 383,212 unrelated men and the EstBB with 91,384 unrelated men, all of European ancestry. PCOS PRS were calculated using data from the consortium PCOS GWAS meta-analysis (including 10,074 PCOS cases and 103,164 controls of European ancestry) (F. Day et al., 2018). As mentioned in the list of publications included in the thesis, the main study design and PRS generation, testing and

validation using UKB data were conducted by Jia Zhu (JZ) and colleagues, and I carried all analysis linked to replication steps in EstBB. Three methods were employed by JZ to construct the PRS: PRSice-2 (Choi & O'Reilly, 2019), PRS-CS (T. Ge et al., 2019), and modified PRS-CS with probabilistic genotype dosages. To optimise PCOS PRSs, women diagnosed with PCOS in the UKB were identified (diagnoses were determined by self-report during a verbal interview with a trained nurse, primary-care clinical events, and/or ICD-9 and 10 respective codes) by JZ and colleagues. The three PRS calculation methods were applied to UKB data and the best-performing method was chosen based on phenotypic variance explained. Then, I validated the optimised PCOS PRS in EstBB women with PCOS captured via ICD-10 codes and independent from the consortium PCOS GWAS presented by F. Day et al. 2018.

The PRS method optimised in women was extended to calculate PRS for men in both the UKB and EstBB, and standardised. Cardiometabolic and androgenic phenotypes were assessed in men based on various measures as described in the paper (J. Zhu et al., 2022). Statistical analyses involved linear regression for continuous phenotypes, logistic regression for dichotomous outcomes, and comparisons between different genetic risk score categories. Adjustments were made for age, genotyping array, genetic PCs and, for UKB, also included the assessment centre. Analyses were conducted with and without BMI as a covariate.

3.3.2. Development and testing of polygenic risk score and combined risk scores for pelvic organ prolapse (Ref. I)

In ref. I, we presented the first PRS for POP, adding a new effort on the use of genomic information to stratify women who suffer from gynaecological conditions. The best-performing PRS, comprised of 3,242,959 SNPs, was constructed by LDPred, showing an OR of 1.42 (95% CI 1.37–1.47) and $p = 2.59 \times 10^{-89}$ in discriminating between cases and controls in the discovery set.

The PRS was further assessed in a validation set, showing the highest Harrell's C-statistic (C-stat) of 0.616 for the continuous distribution. Percentile analysis revealed that women in the top 5% of the PRS distribution had a hazard ratio of 1.61 (95% CI: 1.35–1.92) compared to women in other genetic risk percentiles and a hazard ratio of 1.53 (95% CI: 1.26–1.86) compared to women from the average percentiles group (40–60%). However, it must be noted that this is a Kaplan–Meier estimate and as such may overestimate incidence rates due to not accounting for competing risks such as death before developing the condition.

The study's predictive ability analyses accounted for the effect of age by taking age as the time scale, comparing only women of the same ages and avoiding prediction inflation due to age differences between cases and controls. In a validation subset, defined as part of the data which had minimal missing data in clinical covariates, the PRS alone exhibited a C-stat of 0.583, while clinical risk factors such as the number of children, ever smoked, constipation, BMI, and asthma showed varying predictive abilities on their own (range 0.51–0.56). Combining

PRS with clinical variables significantly improved predictive ability (C-stat=0.630) compared to the clinical model containing only five clinical risk factors (C-stat=0.588) (Figure 8).

The addition of the PRS to the clinical model demonstrated superior predictive ability in incident POP (C-stat=0.630) compared to analysing five clinical risk factors without PRS (C-stat=0.588) (Figure 8). This significant improvement (+4.2 percentage points) is not a common finding. For instance, another study by Inouye *et al.* presented that six traditional risk factors achieved a C-statistic of 0.670 for CAD. When combined with PRS, this increased by 2.6 percentage points to 0.696.

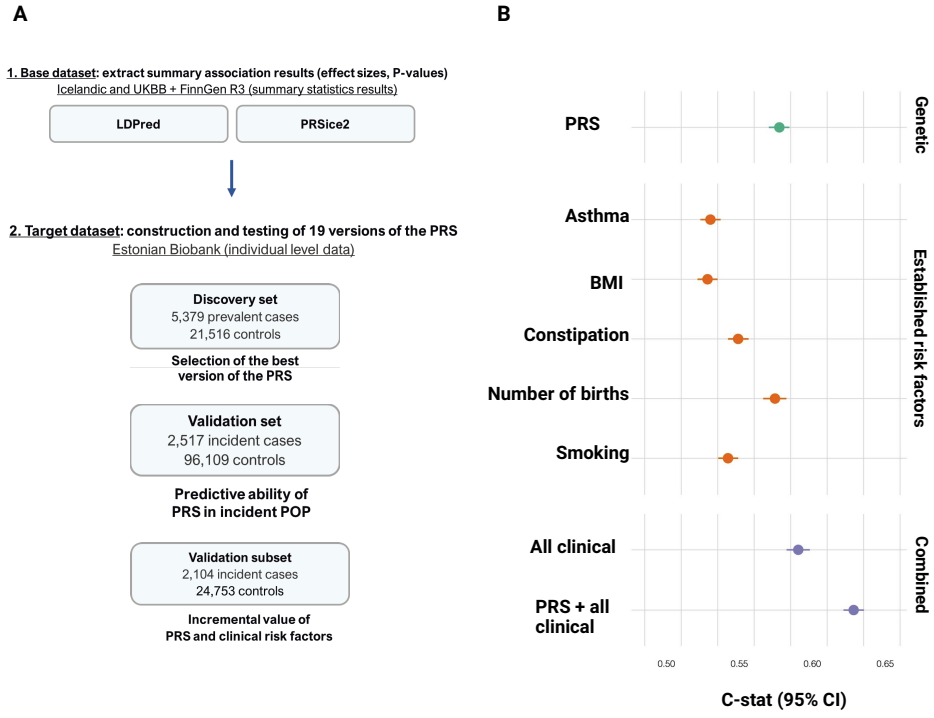

Figure 8. A) Study design for polygenic risk score (PRS) development and testing B) Predictive ability of PRS and clinical variables in incident status. Green dots represent PRS, orange dots represent five established risk factors (Asthma, Body Mass Index (BMI), Constipation, number of births and smoking) and purple dots represent genetic and/or clinical combined models C-statistic (C-stat) indexes. Data are presented as means +/− standard error of mean. C-stat adjusted by batch effects and 10 first principal components in the validation subset of Estonian Biobank (2104 cases and 24,753 controls). Created with BioRender.com and adapted from Pujol-Gualdo et al. 2022.

There have been efforts to present a screening tool for pelvic floor disorders after delivery (http://riskcalc.org/UR_CHOICE/). Unlike this tool which considers traditional risk factors, genetic risk is stable and can be evaluated throughout the lifespan. While the predictive ability shown in this study of PRS alone is limited, PRS based on larger sample sizes and larger validation processes may offer potential clinical uses in disease risk stratification and encourage earlier preventative strategies for women with higher genetic risk to develop POP. However, the clinical translation of PRS profiling for early diagnosis and targeted screening needs further examination in future cohorts with longer follow-up time and an increased number of incident POP cases. Another limitation is that common SNPs only explain a small part of total heritability, suggesting that much of heritability is yet to be discovered. Similarly, it must be noted the potential variability in stratifying risk based on different GWASs when constructing a PRS for the same trait, for instance in the evaluation Läll *et al.,* 2019 performed in breast cancer and the challenge of non-uniqueness in PRSs, especially for studies aiming to stratify individuals for prevention solely based on PRS (Läll et al., 2019). On a global scale, the influence of ancestry is important and a main obstacle nowadays, raising questions about the transferability of PRSs between different populations and on a more local scale. Additionally, potential heterogeneity coming from different diagnoses sources might also hinder PRS reproducibility.

3.3.3. A PRS for PCOS associates with cardiometabolic phenotypes in men (Ref. IV)

In ref IV, we used a PRS for PCOS and examined associations between this PRS and cardiometabolic and androgenic phenotypes in men.

For every 1 SD increase in the PCOS PRS, men in the UKB had increased odds of obesity, T2D, CAD (see association results in Figure 9) and marked androgenic alopecia (OR: 1.03, 95% CI 1.02–1.05, $p = 3 \times 10^{-5}$). BMI, haemoglobin A1c, triglycerides, and free androgen index increased as the PRS increased, whereas high-density lipoprotein cholesterol and SHBG decreased (all $p < 0.0001$).

Both bioactive androgen levels and end-organ responsiveness to androgens (how well an organ replies to the effect of androgens) contribute to the pathogenesis of PCOS. This suggests that both systemic and local bioactive androgen levels and the end-organ sensitivity to androgens play a role in the hyperandrogenism of PCOS in both women and men.

In the EstBB, I sought to replicate the associations between polygenic risk for PCOS and cardiometabolic outcomes in an independent dataset of 37,348 unrelated men. EstBB replicated the associations between a higher PRS and increased odds of obesity and T2D that we identified in the UKB. I observed a 10% increase in odds of obesity and a 6% increase in odds of T2D per 1 SD of PRS, which are comparable to the effect sizes in the UKB (Figure 9). For CAD, we did not observe any difference in the odds of disease with increases in the PRS (Figure 9). We concluded that our replication analysis of CAD in the EstBB was

underpowered to replicate this association. Additionally, the potential inclusion of milder cases in EstBB compared to UKB could also contribute to explain this observation.

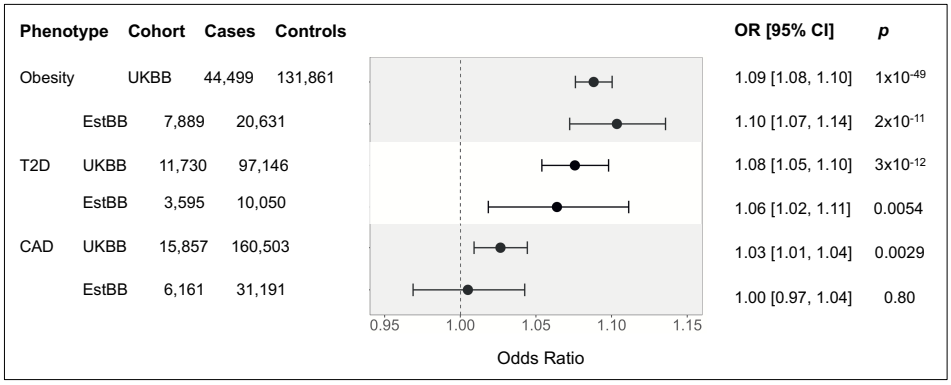

Figure 9. Association of PCOS PRS with phenotypes in the UKB and EstBB. Odds ratios of phenotypes are shown per 1 SD increase in the PCOS polygenic risk score using a logistic regression model. Error bars indicate 95% CI. All analyses were adjusted for age, age squared, genotyping array, and the first 10 principal components. In the UKB, analyses were additionally adjusted for the assessment center; in the EstBB, analyses were adjusted for batch effects. T2D: Type 2 Diabetes, CAD: Coronary artery disease, UKBB: UK Biobank, EstBB: Estonian Biobank. Extracted from Zhu et al. 2021.

The main conclusion from ref. IV is that a PRS for a condition such as PCOS which is diagnosed in women, is associated with cardiometabolic and androgenic conditions in men. This suggests that genetic risk factors for PCOS can act independently of ovarian function and demonstrates PRS as a valuable tool for untangling disease biology, of special interest when coming down to sex-specific conditions. The study's limitations include the limited predictive power of the PRS for PCOS due to the relatively low number of identifiable cases of PCOS in the UKB. The effects of polygenic risk for PCOS on men could be direct or indirect through female relatives. However, the cardiometabolic effects of polygenic risk for PCOS in men are comparable to those in women, suggesting that the association between PCOS and these outcomes is largely not mediated by ovarian function.

CONCLUSIONS

In conclusion, this work supports GWAS as a valuable tool for a wide range of genetic discoveries in different female reproductive health traits, either stemming from multiple cohorts meta-analyses or focusing on a specific population datasets. Additionally, I presented the first exploration for POP using PRS as a tool to inform risk stratification strategies for POP. This has highlighted the potential to improve predictive ability when adding genetic risk on top of clinical risk factors, despite still being far from clinical applications. Other applications of PRS have also been tested and reviewed, such as informing PCOS biology and its relationship with cardiometabolic traits.

- We presented a combination of different data layers to map potential candidate genes from the largest GWAS meta-analysis conducted for POP and AMH levels, along with pathway and tissue enrichment analyses. These analyses provided novel insights into the biology of prolapse development and AMH hormone level variation in pre-menopausal women, advancing our understanding of the genetic underpinnings for these traits and opening new avenues for further functional studies.
- The Estonian Biobank and FinnGen, large-scale biobank resources with specific features of the Nordic healthcare system, provide unique opportunities for the detection of low-frequency alleles enriched in those populations, such as our finding of *CHEK2* loss of function in association with PCOS.
- Leveraging a similar population-specific setting, prospective birth cohort datasets like NFBC1966 contain more granular information such as AMH measurements which allowed us to replicate the *CHEK2* loss of function finding in the association with higher AMH values in pre-menopausal women.
- We performed the first exploration for POP using PRS as a tool to inform risk stratification strategies, which highlighted the potential to improve predictive ability when adding genetic risk on top of some of its clinical risk factors, despite still being far from clinical applications.
- A PRS for a condition such as PCOS which is diagnosed in women, is associated with cardiometabolic and androgenic conditions in men. This suggests that genetic risk factors for PCOS can act independently of ovarian function.

SUMMARY IN ESTONIAN

Naiste reproduktiivtervisega seotud tunnuste geneetiliste seoste tõlgendamine

Geneetiline varieeruvus, eriti üksiku nukleotiidi polümorfismid (SNPid), mõjutavad erinevate komplekshaiguste eelsoodumust. Ülegenoomsed seoseuuringud (GWAS) on leidnud tuhandeid seoseid erinevate geneetiliste variantide ja haiguste vahel. Samas on teadmised naiste reproduktiivtervise tunnuste geneetilise eelsoodumuse osas piiratud, kuna vaid väike osa kõigist GWAS uuringutest keskendub sellele valdkonnale. Populatsioonipõhised biopangad, nagu Eesti geenivaramu, ning rahvastikupõhised sünnikohordid, nagu Northern Finland Birth Cohort 1966, pakuvad väärtuslikku raamistikku selle valdkonna uuringuteks. Tänu metoodika arengule saab GWAS tulemuste põhjal haigusega seotud geneetilistest variantidest liikuda edasi potentsiaalselt mõjutatud geenide, valkude, bioloogiliste radade ja kudedeni. Seega on GWAS uuringud aluseks uutele hüpoteesidele, mida saab kinnitada funktsionaalsete eksperimentide abil. Tunnuste geneetilise komponendi mõistmine on oluline, kuna see võib anda ülevaate haiguse etioloogiast, ja pakkuda uusi võimalusi haigusriskide ennustamiseks ja uuteks ravimsihtmärkideks. GWAS tulemuste põhjal saab koostada ka polügeenseid riskiskoore (PRS), mis annavad kokkuvõtliku hinnangu indiviidi geneetilisele eelsoodumusele teatud tunnuse osas. PRSi abil saab hinnata haigusriske ja lisaks saab seda kasutada vahendina haiguse bioloogia täiendavaks uurimiseks, näiteks seoses kaasuvate haigustega. PRS-id on kesksel kohal nn personaalmeditsiini poole püüdlemisel. Käesoleva doktoritöö eesmärk on dešifreerida naiste reproduktiivtervise teatud tunnuste geneetilisi aluseid ülegenoomsete seoseuuringute abil ning kasutadaa PRS-i nii riskide stratifitseerimisel kui ka tunnuste bioloogia uurimisel. Minu teadustöö ühendab genoomika ja naiste reproduktiivtervise ning loob uusi teadmisi, kasutades selleks suuri genoomikaandmestikke, mida täiendavad elektroonsed terviseandmed ja bioloogilised mõõtmised.

SUMMARY IN FINNISH

Naisten lisääntymisterveydellisten piirteiden taustalla olevien geneettisten assosiaatioiden selvittäminen

On osoitettu että geneettinen variaatio, erityisesti yksinukleotidipolymorfismi (SNP), on yhteydessä terveyteen ja sairastumisalttiuteen useissa monitekijäisissä sairauksissa. Ajatusta tukevat genominlajuiset assosiaatiotutkimukset (GWAS), joiden pohjalta on tunnistettu tuhansia terveyspiirteisiin liittyviä geneettisiä variantteja. Tästä huolimatta naisten lisääntymisterveyteen liittyvää geneettistä vaihtelua on tutkittu vähän, ja vain pieni osa kaikista GWAS-tutkimuksista on keskittynyt tähän osa-alueeseen. Väestöpohjaisten biopankkien, kuten Viron biopankin, ja väestöpohjaisten syntymäkohorttien, kuten Pohjois-Suomen syntymäkohortti 1966:n, olemassaolo tarjoaa erinomaiset puitteet alan tutkimuksille. GWAS-tutkimukset pohjustavat siirtymistä geneettisen variaation tasolta mahdollisesti muuntuneisiin geeneihin, proteiineihin, biologisiin reitteihin ja kudoksiin, ja toimivat näin perustana hypoteeseille, jotka voidaan sitten varmentaa käytännön kokeilla. Tiettyyn piirteeseen liittyvän geneettisen vaihtelun ymmärtäminen on tärkeää, koska se voi antaa uusia näkökulmia sairauden etiologiaan, ennustamiseen ja mahdollisiin hoitoihin liittyen. Yksi GWAS-tutkimusten merkittävistä tuotoksista ovat polygeeniset riskisummat (PRS), jotka tarjoavat kokonaiskuvan yksilön geneettisestä alttiudesta tietylle piirteelle tai sairaudelle. PRS on potentiaalinen työkalu ennustettaessa sairausalttiutta ja tutkittaessa sairauden biologista taustaa, esimerkiksi suhteessa piirteisiin liittyviin liitännäissairauksiin. PRS:t ovat herättäneet laajaa kiinnostusta niin kutsutun yksilöidyn lääketieteen piirissä. Tämä väitöskirja pyrkii selvittämään valikoitujen naisten lisääntymisterveyteen liittyvien piirteiden geneettiset perusteet GWAS:n avulla, ja tutkimaan PRS:n käyttöä työkaluna sekä riskien luokittelussa, että piirteiden biologisen taustan selvittämisessä. Tämä tutkimus, joka sijoittuu genomiikan ja naisten lisääntymisterveyden risteykseen, täyttää sekä kansallisissa että kansainvälisissä tutkimuksissa tunnistetun tietovajeen yhdistämällä suuria genomisia aineistoja, sähköisiä terveysrekistereitä ja biologisia määrityksiä.

REFERENCES

A life-course approach to women's health. (2024). In *Nature medicine* (Vol. 30, Issue 1, p. 1). https://doi.org/10.1038/s41591-023-02777-8

Abbott, D. H., Dumesic, D. A., & Levine, J. E. (2019). Hyperandrogenic origins of polycystic ovary syndrome – implications for pathophysiology and therapy. *Expert Review of Endocrinology & Metabolism*, *14*(2), 131–143. https://doi.org/10.1080/17446651.2019.1576522

Abdellaoui, A., Yengo, L., Verweij, K. J. H., & Visscher, P. M. (2023). 15 years of GWAS discovery: Realizing the promise. *American Journal of Human Genetics*, *110*(2), 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011

Abhyankar, P., Uny, I., Semple, K., Wane, S., Hagen, S., Wilkinson, J., Guerrero, K., Tincello, D., Duncan, E., Calveley, E., Elders, A., McClurg, D., & Maxwell, M. (2019). Women's experiences of receiving care for pelvic organ prolapse: a qualitative study. *BMC Women's Health*, *19*(1), 45. https://doi.org/10.1186/s12905-019-0741-2

Abu-El-Haija, A., Reddi, H. V, Wand, H., Rose, N. C., Mori, M., Qian, E., & Murray, M. F. (2023). The clinical application of polygenic risk scores: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *25*(5), 100803. https://doi.org/10.1016/j.gim.2023.100803

Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, J. K., Chen, Y., Villalta, J. E., Gilbert, L. A., Horlbeck, M. A., Hein, M. Y., Pak, R. A., Gray, A. N., Gross, C. A., Dixit, A., Parnas, O., Regev, A., & Weissman, J. S. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*, *167*(7), 1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048

Adeyemo, A., Balaconis, M. K., Darnes, D. R., Fatumo, S., Granados Moreno, P., Hodonsky, C. J., Inouye, M., Kanai, M., Kato, K., Knoppers, B. M., Lewis, A. C. F., Martin, A. R., McCarthy, M. I., Meyer, M. N., Okada, Y., Richards, J. B., Richter, L., Ripatti, S., Rotimi, C. N., … Zhou, A. (2021). Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nature Medicine*, *27*(11), 1876–1884. https://doi.org/10.1038/S41591-021-01549-6

Ahn, J., Urist, M., & Prives, C. (2004). The Chk2 protein kinase. *DNA Repair*, *3*(8–9), 1039–1047. https://doi.org/10.1016/j.dnarep.2004.03.033

Akyüz, K., Chassang, G., Goisauf, M., Kozera, Ł., Mezinska, S., Tzortzatou, O., & Mayrhofer, M. T. (2021). Biobanking and risk assessment: a comprehensive typology of risks for an adaptive risk governance. *Life Sciences, Society and Policy*, *17*(1), 10. https://doi.org/10.1186/s40504-021-00117-7

Altman, D., Forsman, M., Falconer, C., & Lichtenstein, P. (2008). Genetic influence on stress urinary incontinence and pelvic organ prolapse. *European Urology*, *54*(4), 918–922. https://doi.org/10.1016/j.eururo.2007.12.004

Andraweera, P. H., & Lassi, Z. S. (2019). Cardiovascular Risk Factors in Offspring of Preeclamptic Pregnancies-Systematic Review and Meta-Analysis. *The Journal of Pediatrics*, *208*, 104–113.e6. https://doi.org/10.1016/j.jpeds.2018.12.008

Aragam, K. G., Jiang, T., Goel, A., Kanoni, S., Wolford, B. N., Atri, D. S., Weeks, E. M., Wang, M., Hindy, G., Zhou, W., Grace, C., Roselli, C., Marston, N. A., Kamanu, F. K., Surakka, I., Venegas, L. M., Sherliker, P., Koyama, S., Ishigaki, K., … Butterworth, A. S. (2022). Discovery and systematic characterization of risk variants

and genes for coronary artery disease in over a million participants. *Nature Genetics*, *54*(12), 1803–1815. https://doi.org/10.1038/s41588-022-01233-6

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. In *Nature* (Vol. 526, Issue 7571, pp. 68–74). Nature Publishing Group. https://doi.org/10.1038/nature15393

Backman, J. D., Li, A. H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M. D., Benner, C., Liu, D., Locke, A. E., Balasubramanian, S., Yadav, A., Banerjee, N., Gillies, C. E., Damask, A., Liu, S., Bai, X., Hawes, A., Maxwell, E., Gurski, L., … Ferreira, M. A. R. (2021). Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*, *599*(7886), 628–634. https://doi.org/10.1038/s41586-021-04103-z

Baerwald, A. R., Adams, G. P., & Pierson, R. A. (2012). Ovarian antral folliculogenesis during the human menstrual cycle: a review. *Human Reproduction Update*, *18*(1), 73–91. https://doi.org/10.1093/humupd/dmr039

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barbeira, A. N., Bonazzola, R., Gamazon, E. R., Liang, Y., Park, Y., Kim-Hellmuth, S., Wang, G., Jiang, Z., Zhou, D., Hormozdiari, F., Liu, B., Rao, A., Hamel, A. R., Pividori, M. D., Aguet, F., Bastarache, L., Jordan, D. M., Verbanck, M., Do, R., … Im, H. K. (2021). Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biology*, *22*(1), 49. https://doi.org/10.1186/s13059-020-02252-4

Barber, M. D. (2016). Pelvic organ prolapse. *BMJ (Online)*, *354*. https://doi.org/10.1136/bmj.i3853

Barber, T. M., & Franks, S. (2021). Obesity and polycystic ovary syndrome. *Clinical Endocrinology*, *95*(4), 531–541. https://doi.org/10.1111/cen.14421

Barry, J. A., Azizia, M. M., & Hardiman, P. J. (2014). Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction Update*, *20*(5), 748–758. https://doi.org/10.1093/humupd/dmu012

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neuroscience and Biobehavioral Reviews*, *35*(3), 565–572. https://doi.org/10.1016/j.neubiorev.2010.07.002

Blomquist, J. L., Muñoz, A., Carroll, M., & Handa, V. L. (2018). Association of Delivery Mode with Pelvic Floor Disorders after Childbirth. *JAMA – Journal of the American Medical Association*, *320*(23), 2438–2447. https://doi.org/10.1001/jama.2018.18315

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *The American Journal of Psychiatry*, *176*(5), 376–387. https://doi.org/10.1176/appi.ajp.2018.18070881

Bowton, E., Field, J. R., Wang, S., Schildcrout, J. S., Van Driest, S. L., Delaney, J. T., Cowan, J., Weeke, P., Mosley, J. D., Wells, Q. S., Karnes, J. H., Shaffer, C., Peterson, J. F., Denny, J. C., Roden, D. M., & Pulley, J. M. (2014). Biobanks and electronic medical records: enabling cost-effective research. *Science Translational Medicine*, *6*(234), 234cm3. https://doi.org/10.1126/scitranslmed.3008604

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, *132*(2), 311–322. https://doi.org/10.1016/j.cell.2007.12.014

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., … Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

Burstein, D., Hoffman, G., Mathur, D., Venkatesh, S., Therrien, K., Fanous, A. H., Bigdeli, T. B., Harvey, P. D., Roussos, P., & Voloudakis, G. (2023). Detecting and Adjusting for Hidden Biases due to Phenotype Misclassification in Genome-Wide Association Studies. In *medRxiv : the preprint server for health sciences*. https://doi.org/10.1101/2023.01.17.23284670

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Carlson, C. S., Eberle, M. A., Rieder, M. J., Smith, J. D., Kruglyak, L., & Nickerson, D. A. (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nature Genetics*, *33*(4), 518–521. https://doi.org/10.1038/ng1128

Cartwright, R., Kirby, A. C., Tikkinen, K. A. O., Mangera, A., Thiagamoorthy, G., Rajan, P., Pesonen, J., Ambrose, C., Gonzalez-Maffe, J., Bennett, P., Palmer, T., Walley, A., Järvelin, M.-R., Chapple, C., & Khullar, V. (2015). Systematic review and metaanalysis of genetic association studies of urinary symptoms and prolapse in women. *American Journal of Obstetrics and Gynecology*, *212*(2), 199.e1–24. https://doi.org/10.1016/j.ajog.2014.08.005

Carver, T., Hartley, S., Lee, A., Cunningham, A. P., Archer, S., Babb de Villiers, C., Roberts, J., Ruston, R., Walter, F. M., Tischkowitz, M., Easton, D. F., & Antoniou, A. C. (2021). CanRisk Tool-A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *30*(3), 469–473. https://doi.org/10.1158/1055-9965.EPI-20-1319

Chan, A. J. S., Engchuan, W., Reuter, M. S., Wang, Z., Thiruvahindrapuram, B., Trost, B., Nalpathamkalam, T., Negrijn, C., Lamoureux, S., Pellecchia, G., Patel, R. V, Sung, W. W. L., MacDonald, J. R., Howe, J. L., Vorstman, J., Sondheimer, N., Takahashi, N., Miles, J. H., Anagnostou, E., … Scherer, S. W. (2022). Ge nome-wide

rare variant score associates with morphological subtypes of autism spectrum disorder. *Nature Communications*, *13*(1), 6463. https://doi.org/10.1038/s41467-022-34112-z

Chen, Z.-J., Zhao, H., He, L., Shi, Y., Qin, Y., Shi, Y., Li, Z., You, L., Zhao, J., Liu, J., Liang, X., Zhao, X., Zhao, J., Sun, Y., Zhang, B., Jiang, H., Zhao, D., Bian, Y., Gao, X., … Zhao, Y. (2011). Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. *Nature Genetics*, *43*(1), 55–59. https://doi.org/10.1038/ng.732

Choi, S. W., Mak, T. S. H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. In *Nature Protocols* (Vol. 15, Issue 9, pp. 2759–2772). Nature Research. https://doi.org/10.1038/s41596-020-0353-1

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, *8*(7). https://doi.org/10.1093/gigascience/giz082

Currant, H., Fitzgerald, T. W., Patel, P. J., Khawaja, A. P., Webster, A. R., Mahroo, O. A., & Birney, E. (2023). Sub-cellular level resolution of common genetic variation in the photoreceptor layer identifies continuum between rare disease and common variation. *PLoS Genetics*, *19*(2), e1010587. https://doi.org/10.1371/journal.pgen.1010587

Currant, H., Hysi, P., Fitzgerald, T. W., Gharahkhani, P., Bonnemaijer, P. W. M., Sena-bouth, A., Hewitt, A. W., Atan, D., Aung, T., Charng, J., Choquet, H., Craig, J., Khaw, P. T., Klaver, C. C. W., Kubo, M., Ong, J.-S., Pasquale, L. R., Reisman, C. A., Danis-zewski, M., … Khawaja, A. P. (2021). Genetic variation affects morphological retinal phenotypes extracted from UK Biobank optical coherence tomography images. *PLoS Genetics*, *17*(5), e1009497. https://doi.org/10.1371/journal.pgen.1009497

Darwich, R., Li, W., Yamak, A., Komati, H., Andelfinger, G., Sun, K., & Nemer, M. (2017). KLF13 is a genetic modifier of the Holt-Oram syndrome gene TBX5. *Human Molecular Genetics*, *26*(5), 942–954. https://doi.org/10.1093/hmg/ddx009

Day, F., Karaderi, T., Jones, M. R., Meun, C., He, C., Drong, A., Kraft, P., Lin, N., Huang, H., Broer, L., Magi, R., Saxena, R., Laisk, T., Urbanek, M., Hayes, M. G., Thorleifsson, G., Fernandez-Tajes, J., Mahajan, A., Mullin, B. H., … Welt, C. K. (2018). Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genetics*, *14*(12), e1007813. https://doi.org/10.1371/journal.pgen.1007813

Day, F. R., Hinds, D. A., Tung, J. Y., Stolk, L., Styrkarsdottir, U., Saxena, R., Bjonnes, A., Broer, L., Dunger, D. B., Halldorsson, B. V, Lawlor, D. A., Laval, G., Mathieson, I., McCardle, W. L., Louwers, Y., Meun, C., Ring, S., Scott, R. A., Sulem, P., … Perry, J. R. B. (2015). Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nature Communications*, *6*, 8464. https://doi.org/10.1038/ncomms9464

Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B., Esko, T., Johnson, A. D., Elks, C. E., Franceschini, N., He, C., Altmaier, E., Brody, J. A., Franke, L. L., Huff-man, J. E., … Murray, A. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics*, *47*(11), 1294–1303. https://doi.org/10.1038/ng.3412

Day, F. R., Thompson, D. J., Helgason, H., Chasman, D. I., Finucane, H., Sulem, P., Ruth, K. S., Whalen, S., Sarkar, A. K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C. M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J. J., … Perry, J. R. B. (2017). Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nature Genetics*, *49*(6), 834–841. https://doi.org/10.1038/ng.3841

de Kat, A. C., Verschuren, W. M., Eijkemans, M. J. C., Broekmans, F. J. M., & van der Schouw, Y. T. (2017). Anti-Müllerian Hormone Trajectories Are Associated With Cardiovascular Disease in Women: Results From the Doetinchem Cohort Study. *Circulation*, *135*(6), 556–565. https://doi.org/10.1161/CIRCULATIONAHA.116.025968

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., Jostins, L., Rice, D. L., Gutierrez-Achury, J., Ji, S.-G., Heap, G., Nimmo, E. R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D. C., … Barrett, J. C. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics*, *49*(2), 256–261. https://doi.org/10.1038/ng.3760

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Computational Biology*, *11*(4). https://doi.org/10.1371/journal.pcbi.1004219

de Ziegler, D., Pirtea, P., Fanchin, R., & Ayoubi, J. M. (2018). Ovarian reserve in polycystic ovary syndrome: more, but for how long? In *Fertility and sterility* (Vol. 109, Issue 3, pp. 448–449). https://doi.org/10.1016/j.fertnstert.2017.11.027

Denny, J. C., Rutter, J. L., Goldstein, D. B., Philippakis, A., Smoller, J. W., Jenkins, G., & Dishman, E. (2019). The "All of Us" Research Program. *The New England Journal of Medicine*, *381*(7), 668–676. https://doi.org/10.1056/NEJMsr1809937

Di Guardo, F., Ciotta, L., Monteleone, M., & Palumbo, M. (2020). Male Equivalent Polycystic Ovarian Syndrome: Hormonal, Metabolic, and Clinical Aspects. *International Journal of Fertility & Sterility*, *14*(2), 79–83. https://doi.org/10.22074/ijfs.2020.6092

Ding, D.-C., Chen, W., Wang, J.-H., & Lin, S.-Z. (2018). Association between polycystic ovarian syndrome and endometrial, ovarian, and breast cancer: A population-based cohort study in Taiwan. *Medicine*, *97*(39), e12608. https://doi.org/10.1097/MD.0000000000012608

Direct-to-consumer prenatal testing for multigenic or polygenic disorders: a position statement of the American College of Medical Genetics and Genomics (ACMG). (2021). *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *23*(11), 2027–2028. https://doi.org/10.1038/S41436-021-01247-1

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., Marjanovic, N. D., Dionne, D., Burks, T., Raychowdhury, R., Adamson, B., Norman, T. M., Lander, E. S., Weissman, J. S., Friedman, N., & Regev, A. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, *167*(7), 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038

Dixon, P., Keeney, E., Taylor, J. C., Wordsworth, S., & Martin, R. M. (2022). Can polygenic risk scores contribute to cost-effective cancer screening? A systematic review. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *24*(8), 1604–1617. https://doi.org/10.1016/j.gim.2022.04.020

Docherty, A. R., Shabalin, A. A., Adkins, D. E., Mann, F., Krueger, R. F., Bacanu, S.-A., Campbell, A., Hayward, C., Porteous, D. J., McIntosh, A. M., & Kendler, K. S. (2020). Molecular Genetic Risk for Psychosis Is Associated With Psychosis Risk Symptoms in a Population-Based UK Cohort: Findings From Generation Scotland. *Schizophrenia Bulletin*, *46*(5), 1045–1052. https://doi.org/10.1093/schbul/sbaa042

Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K. A., Parsons, M. T., Fortuno, C., Wang, Q., Bolla, M. K., Dennis, J., Keeman, R., Alonso, M. R., Álvarez, N., Herraez, B., Fernandez, V., Núñez-Torres, R., Osorio, A., … Easton, D. F. (2021). Breast Cancer Risk Genes – Association Analysis

in More than 113,000 Women. *The New England Journal of Medicine*, *384*(5), 428–439. https://doi.org/10.1056/NEJMoa1913948

Dornbos, P., Koesterer, R., Ruttenburg, A., Nguyen, T., Cole, J. B., Leong, A., Meigs, J. B., Florez, J. C., Rotter, J. I., Udler, M. S., & Flannick, J. (2022). A combined polygenic score of 21,293 rare and 22 common variants improves diabetes diagnosis based on hemoglobin A1C levels. *Nature Genetics*, *54*(11), 1609–1614. https://doi.org/10.1038/s41588-022-01200-1

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1003348

Duerr, R. H., Taylor, K. D., Brant, S. R., Rioux, J. D., Silverberg, M. S., Daly, M. J., Steinhart, A. H., Abraham, C., Regueiro, M., Griffiths, A., Dassopoulos, T., Bitton, A., Yang, H., Targan, S., Datta, L. W., Kistner, E. O., Schumm, L. P., Lee, A. T., Gregersen, P. K., … Cho, J. H. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science (New York, N.Y.)*, *314*(5804), 1461–1463. https://doi.org/10.1126/science.1135245

Duim, S. N., Kurakula, K., Goumans, M. J., & Kruithof, B. P. T. (2015). Cardiac endothelial cells express Wilms' tumor-1. Wt1 expression in the developing, adult and infarcted heart. *Journal of Molecular and Cellular Cardiology*, *81*, 127–135. https://doi.org/10.1016/j.yjmcc.2015.02.007

Duim, S. N., Smits, A. M., Kruithof, B. P. T., & Goumans, M. J. (2016). The roadmap of WT1 protein expression in the human fetal heart. *Journal of Molecular and Cellular Cardiology*, *90*, 139–145. https://doi.org/10.1016/j.yjmcc.2015.12.008

Dumesic, D. A., & Lobo, R. A. (2013). Cancer risk and PCOS. *Steroids*, *78*(8), 782–785. https://doi.org/10.1016/j.steroids.2013.04.004

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1), 3328. https://doi.org/10.1038/s41467-019-11112-0

Elks, C. E., Perry, J. R. B., Sulem, P., Chasman, D. I., Franceschini, N., He, C., Lunetta, K. L., Visser, J. A., Byrne, E. M., Cousminer, D. L., Gudbjartsson, D. F., Esko, T., Feenstra, B., Hottenga, J.-J., Koller, D. L., Kutalik, Z., Lin, P., Mangino, M., Marongiu, M., … Murray, A. (2010). Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nature Genetics*, *42*(12), 1077–1085. https://doi.org/10.1038/ng.714

Ellerkmann, R. M., Cundiff, G. W., Melick, C. F., Nihira, M. A., Leffler, K., & Bent, A. E. (2001). Correlation of symptoms with location and severity of pelvic organ prolapse. *American Journal of Obstetrics and Gynecology*, *185*(6), 1332–1338. https://doi.org/10.1067/MOB.2001.119078

Escobar-Morreale, H. F. (2018). Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nature Reviews. Endocrinology*, *14*(5), 270–284. https://doi.org/10.1038/nrendo.2018.24

Esserman, L. J. (2017). The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer*, *3*, 34. https://doi.org/10.1038/s41523-017-0035-5

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics (Oxford, England)*, *31*(9), 1466–1468. https://doi.org/10.1093/bioinformatics/btu848

Evangelou, E., Warren, H. R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C. P., Karaman, I., Ng, F. L., Evangelou, M., Witkowska, K., Tzanis, E., Hellwege, J. N., Giri, A., Velez Edwards, D. R., Sun, Y. V, Cho, K., … Caulfield, M. J. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature Genetics*, *50*(10), 1412–1425. https://doi.org/10.1038/s41588-018-0205-x

Evans, D. M., Visscher, P. M., & Wray, N. R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Human Molecular Genetics*, *18*(18), 3525–3531. https://doi.org/10.1093/hmg/ddp295

Evans, L. M., Tahmasbi, R., Vrieze, S. I., Abecasis, G. R., Das, S., Gazal, S., Bjelland, D. W., de Candia, T. R., Goddard, M. E., Neale, B. M., Yang, J., Visscher, P. M., & Keller, M. C. (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature Genetics*, *50*(5), 737–745. https://doi.org/10.1038/s41588-018-0108-x

Fadista, J., Skotte, L., Karjalainen, J., Abner, E., Sørensen, E., Ullum, H., Werge, T., Esko, T., Milani, L., Palotie, A., Daly, M., Melbye, M., Feenstra, B., & Geller, F. (2022). Comprehensive genome-wide association study of different forms of hernia identifies more than 80 associated loci. *Nature Communications*, *13*(1), 3200. https://doi.org/10.1038/s41467-022-30921-4

Fahed, A. C., Philippakis, A. A., & Khera, A. V. (2022). The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nature Communications*, *13*(1), 2922. https://doi.org/10.1038/s41467-022-30675-z

Farrell, M. S., Werge, T., Sklar, P., Owen, M. J., Ophoff, R. A., O'Donovan, M. C., Corvin, A., Cichon, S., & Sullivan, P. F. (2015). Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry*, *20*(5), 555–562. https://doi.org/10.1038/mp.2015.16

Fauser, B. C. J. M., Laven, J. S. E., Tarlatzis, B. C., Moley, K. H., Critchley, H. O. D., Taylor, R. N., Berga, S. L., Mermelstein, P. G., Devroey, P., Gianaroli, L., D'Hooghe, T., Vercellini, P., Hummelshoj, L., Rubin, S., Goverde, A. J., De Leo, V., & Petraglia, F. (2011). Sex steroid hormones and reproductive disorders: impact on women's health. *Reproductive Sciences (Thousand Oaks, Calif.)*, *18*(8), 702–712. https://doi.org/10.1177/1933719111405068

Fehrmann, R. S. N., Karjalainen, J. M., Krajewska, M., Westra, H.-J., Maloney, D., Simeonov, A., Pers, T. H., Hirschhorn, J. N., Jansen, R. C., Schultes, E. A., van Haagen, H. H. H. B. M., de Vries, E. G. E., te Meerman, G. J., Wijmenga, C., van Vugt, M. A. T. M., & Franke, L. (2015). Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nature Genetics*, *47*(2), 115–125. https://doi.org/10.1038/ng.3173

Fejzo, M. S., Sazonova, O. V, Sathirapongsasuti, J. F., Hallgrímsdóttir, I. B., Vacic, V., MacGibbon, K. W., Schoenberg, F. P., Mancuso, N., Slamon, D. J., & Mullin, P. M. (2018). Placenta and appetite genes GDF15 and IGFBP7 are associated with hyperemesis gravidarum. *Nature Communications*, *9*(1), 1178. https://doi.org/10.1038/s41467-018-03258-0

Ferguson, B. S., Nam, H., & Morrison, R. F. (2019). Dual-specificity phosphatases regulate mitogen-activated protein kinase signaling in adipocytes in response to inflammatory stress. *Cellular Signalling*, *53*, 234–245. https://doi.org/10.1016/j.cellsig.2018.10.011

Finkelstei, J. S., Lee, H., Karlamangla, A., Nee, R. M., Slus, P. M., Burnett-Bowie, S. A. M., Darakananda, K., Donaho, P. K., Harlo, S. D., Prizan, S. H., Joffe, H., Kumar, A., Marti, D. E., McConnell, D., Merrilat, S., Morrison, A., Pastor, L. M., Randolph, J. F., Greendal, G. A., & Santoro, N. (2020). Antimullerian Hormone and Impending Menopause in Late Reproductive Age: The Study of Women's Health Across the Nation. *The Journal of Clinical Endocrinology and Metabolism*, *105*(4), E1862–E1871. https://doi.org/10.1210/CLINEM/DGZ283

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., … Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. https://doi.org/10.1038/ng.3404

Foley, C. N., Staley, J. R., Breen, P. G., Sun, B. B., Kirk, P. D. W., Burgess, S., & Howson, J. M. M. (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nature Communications*, *12*(1). https://doi.org/10.1038/S41467-020-20885-8

Forslund, M., Landin-Wilhelmsen, K., Schmidt, J., Brännström, M., Trimpou, P., & Dahlgren, E. (2019). Higher menopausal age but no differences in parity in women with polycystic ovary syndrome compared with controls. *Acta Obstetricia et Gynecologica Scandinavica*, *98*(3), 320–326. https://doi.org/10.1111/aogs.13489

Frandsen, C. L. B., Svendsen, P. F., Nøhr, B., Viuff, J. H., Maltesen, T., Kjaer, S. K., & Jensen, A. (2023). Risk of epithelial ovarian tumors among women with polycystic ovary syndrome: A nationwide population-based cohort study. *International Journal of Cancer*, *153*(5), 958–968. https://doi.org/10.1002/ijc.34574

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R. B., 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., … 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. https://doi.org/10.1126/science.1141634

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(5), 1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Fuat, A., Adlen, E., Monane, M., Coll, R., Groves, S., Little, E., Wild, J., Kamali, F. J., Soni, Y., Haining, S., Riding, H., Riveros-Mckay, F., Peneva, I., Lachapelle, A., Giner-Delgado, C., Weale, M. E., Plagnol, V., Harrison, S., & Donnelly, P. (2024). A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. *European Journal of Preventive Cardiology*. https://doi.org/10.1093/eurjpc/zwae004

Fukuda, R., Gunawan, F., Ramadass, R., Beisaw, A., Konzer, A., Mullapudi, S. T., Gentile, A., Maischein, H. M., Graumann, J., & Stainier, D. Y. R. (2019). Mechanical Forces Regulate Cardiomyocyte Myofilament Maturation via the VCL-SSH1-CFL Axis. *Developmental Cell*, *51*(1), 62–77.e5. https://doi.org/10.1016/j.devcel.2019.08.006

Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., LaFleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., … O'Leary, T. J. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, *70*, 214–223. https://doi.org/10.1016/j.jclinepi.2015.09.016

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1), 1776. https://doi.org/10.1038/s41467-019-09718-5

Ge, W., Clendenen, T. V., Afanasyeva, Y., Koenig, K. L., Agnoli, C., Brinton, L. A., Dorgan, J. F., Eliassen, A. H., Falk, R. T., Hallmans, G., Hankinson, S. E., Hoffman-Bolton, J., Key, T. J., Krogh, V., Nichols, H. B., Sandler, D. P., Schoemaker, M. J., Sluss, P. M., Sund, M., … Zeleniuch-Jacquotte, A. (2018). Circulating anti-Müllerian hormone and breast cancer risk: A study in ten prospective cohorts. *International Journal of Cancer*, *142*(11), 2215–2226. https://doi.org/10.1002/IJC.31249

Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. (2007). *Nature*, *447*(7145), 661–678. https://doi.org/10.1038/nature05911

Gershon, E., & Dekel, N. (2020). Newly Identified Regulators of Ovarian Folliculogenesis and Ovulation. *International Journal of Molecular Sciences*, *21*(12). https://doi.org/10.3390/ijms21124565

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., & Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genetics*, *10*(5), e1004383. https://doi.org/10.1371/journal.pgen.1004383

Giri, A., Hartmann, K. E., Hellwege, J. N., Velez Edwards, D. R., & Edwards, T. L. (2017). Obesity and pelvic organ prolapse: a systematic review and meta-analysis of observational studies. *American Journal of Obstetrics and Gynecology*, *217*(1), 11–26.e3. https://doi.org/10.1016/j.ajog.2017.01.039

Goo, Y. H., Son, S. H., Kreienberg, P. B., & Paul, A. (2014). Novel lipid droplet-associated serine hydrolase regulates macrophage cholesterol mobilization. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *34*(2), 386–396. https://doi.org/10.1161/ATVBAHA.113.302448

Gottschau, M., Kjaer, S. K., Jensen, A., Munk, C., & Mellemkjaer, L. (2015). Risk of cancer among women with polycystic ovary syndrome: a Danish cohort study. *Gynecologic Oncology*, *136*(1), 99–103. https://doi.org/10.1016/j.ygyno.2014.11.012

Greville-Heygate, S. L., Maishman, T., Tapper, W. J., Cutress, R. I., Copson, E., Dunning, A. M., Haywood, L., Jones, L. J., & Eccles, D. M. (2020). Pathogenic Variants in CHEK2 Are Associated With an Adverse Prognosis in Symptomatic Early-Onset Breast Cancer. *JCO Precision Oncology*, *4*. https://doi.org/10.1200/PO.19.00178

Griffin, W. C., & Trakselis, M. A. (2019). The MCM8/9 complex: A recent recruit to the roster of helicases involved in genome maintenance. *DNA Repair*, *76*, 1–10. https://doi.org/10.1016/j.dnarep.2019.02.003

Grinspon, R. P., & Rey, R. A. (2014). When hormone defects cannot explain it: malformative disorders of sex development. *Birth Defects Research. Part C, Embryo Today : Reviews*, *102*(4), 359–373. https://doi.org/10.1002/bdrc.21086

Gulcher, J., & Stefansson, K. (1999). An Icelandic saga on a centralized healthcare database and democratic decision making. *Nature Biotechnology*, *17*(7), 620. https://doi.org/10.1038/10796

Hagen, S., Glazener, C., McClurg, D., Macarthur, C., Elders, A., Herbison, P., Wilson, D., Toozs-Hobson, P., Hemming, C., Hay-Smith, J., Collins, M., Dickson, S., & Logan, J. (2017). Pelvic floor muscle training for secondary prevention of pelvic organ prolapse (PREVPROL): a multicentre randomised controlled trial. *Lancet (London, England)*, *389*(10067), 393–402. https://doi.org/10.1016/S0140-6736(16)32109-2

Hajek, C., Guo, X., Yao, J., Hai, Y., Johnson, W. C., Frazier-Wood, A. C., Post, W. S., Psaty, B. M., Taylor, K. D., & Rotter, J. I. (2018). Coronary Heart Disease Genetic Risk Score Predicts Cardiovascular Disease Risk in Men, Not Women. In *Circulation. Genomic and precision medicine* (Vol. 11, Issue 10, p. e002324). https://doi.org/10.1161/CIRCGEN.118.002324

Hallamies, S., Pelttari, L. M., Poikonen-Saksela, P., Jekunen, A., Jukkola-Vuorinen, A., Auvinen, P., Blomqvist, C., Aittomäki, K., Mattson, J., & Nevanlinna, H. (2017). CHEK2 c.1100delC mutation is associated with an increased risk for male breast cancer in Finnish patient population. *BMC Cancer*, *17*(1), 620. https://doi.org/10.1186/s12885-017-3631-8

Hawkes, G., Beaumont, R. N., Li, Z., Mandla, R., Li, X., Albert, C. M., Arnett, D. K., Ashley-Koch, A. E., Ashrani, A. A., Barnes, K. C., Boerwinkle, E., Brody, J. A., Carson, A. P., Chami, N., Chen, Y.-D. I., Chung, M. K., Curran, J. E., Darbar, D., Ellinor, P. T., … Weedon, M. N. (2023). Whole genome association testing in 333,100 individuals across three biobanks identifies rare non-coding single variant and genomic aggregate associations with height. *BioRxiv*, 2023.11.19.566520. https://doi.org/10.1101/2023.11.19.566520

Hayes, M. G., Urbanek, M., Ehrmann, D. A., Armstrong, L. L., Lee, J. Y., Sisk, R., Karaderi, T., Barber, T. M., McCarthy, M. I., Franks, S., Lindgren, C. M., Welt, C. K., Diamanti-Kandarakis, E., Panidis, D., Goodarzi, M. O., Azziz, R., Zhang, Y., James, R. G., Olivier, M., … Dunaif, A. (2015). Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nature Communications*, *6*, 7502. https://doi.org/10.1038/ncomms8502

Hendrix, S. L., Clark, A., Nygaard, I., Aragaki, A., Barnabei, V., & McTiernan, A. (2002). Pelvic organ prolapse in the Women's Health Initiative: gravity and gravidity. *American Journal of Obstetrics and Gynecology*, *186*(6), 1160–1166. https://doi.org/10.1067/mob.2002.123819

Hernandez Gifford, J. A. (2015). The role of WNT signaling in adult ovarian folliculogenesis. *Reproduction*, *150*(4), E137–R148. https://doi.org/10.1530/REP-14-0685

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9362–9367. https://doi.org/10.1073/pnas.0903103106

Hodgson, S., Huang, Q. Q., Sallah, N., Griffiths, C. J., Newman, W. G., Trembath, R. C., Wright, J., Lumbers, R. T., Kuchenbaecker, K., van Heel, D. A., Mathur, R., Martin, H. C., & Finer, S. (2022). Integrating polygenic risk scores in the prediction of type 2 diabetes risk and subtypes in British Pakistanis and Bangladeshis: A population-based cohort study. *PLoS Medicine*, *19*(5), e1003981. https://doi.org/10.1371/journal.pmed.1003981

Holland, D., Frei, O., Desikan, R., Fan, C.-C., Shadrin, A. A., Smeland, O. B., Sundar, V. S., Thompson, P., Andreassen, O. A., & Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genetics*, *16*(5), e1008612. https://doi.org/10.1371/journal.pgen.1008612

Hollinrake, E., Abreu, A., Maifeld, M., Van Voorhis, B. J., & Dokras, A. (2007). Increased risk of depressive disorders in women with polycystic ovary syndrome. *Fertility and Sterility*, *87*(6), 1369–1376. https://doi.org/10.1016/j.fertnstert.2006.11.039

Homburg, R., & Crawford, G. (2014). The role of AMH in anovulation associated with PCOS: a hypothesis. *Human Reproduction (Oxford, England)*, *29*(6), 1117–1121. https://doi.org/10.1093/humrep/deu076

Honigberg, M. C., Truong, B., Khan, R. R., Xiao, B., Bhatta, L., Vy, H. M. T., Guerrero, R. F., Schuermans, A., Selvaraj, M. S., Patel, A. P., Koyama, S., Cho, S. M. J., Vellarikkal, S. K., Trinder, M., Urbut, S. M., Gray, K. J., Brumpton, B. M., Patil, S., Zöllner, S., … Natarajan, P. (2023). Polygenic prediction of preeclampsia and gestational hypertension. *Nature Medicine*, *29*(6), 1540–1549. https://doi.org/10.1038/s41591-023-02374-9

Horne, A. W., & Missmer, S. A. (2022). Pathophysiology, diagnosis, and management of endometriosis. *BMJ (Clinical Research Ed.)*, *379*, e070750. https://doi.org/10.1136/bmj-2022-070750

Hoyos, L. R., Visser, J. A., McLuskey, A., Chazenbalk, G. D., Grogan, T. R., & Dumesic, D. A. (2020). Loss of anti-Müllerian hormone (AMH) immunoactivity due to a homozygous AMH gene variant rs10417628 in a woman with classical polycystic ovary syndrome (PCOS). *Human Reproduction (Oxford, England)*, *35*(10), 2294–2302. https://doi.org/10.1093/humrep/deaa199

Hurson, A. N., Pal Choudhury, P., Gao, C., Hüsing, A., Eriksson, M., Shi, M., Jones, M. E., Evans, D. G. R., Milne, R. L., Gaudet, M. M., Vachon, C. M., Chasman, D. I., Easton, D. F., Schmidt, M. K., Kraft, P., Garcia-Closas, M., & Chatterjee, N. (2022). Prospective evaluation of a breast-cancer risk model integrating classical risk factors and polygenic risk in 15 cohorts from six countries. *International Journal of Epidemiology*, *50*(6), 1897–1911. https://doi.org/10.1093/ije/dyab036

ICDA Recommendations and White Paper.pdf – Google Drive. (n.d.). Retrieved April 3, 2024, from https://drive.google.com/file/d/16SVJ5lbneN9hB9E03PZMhpescAN527HO/view

Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., Lai, F. Y., Kaptoge, S., Brozynska, M., Wang, T., Ye, S., Webb, T. R., Rutter, M. K., Tzoulaki, I., Patel, R. S., Loos, R. J. F., Keavney, B., Hemingway, H., Thompson, J., … Samani, N. J. (2018). Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *Journal of the American College of Cardiology*, *72*(16), 1883–1893. https://doi.org/10.1016/j.jacc.2018.07.079

Insull, W. J. (2009). The pathology of atherosclerosis: plaque development and plaque responses to medical treatment. *The American Journal of Medicine*, *122*(1 Suppl), S3–S14. https://doi.org/10.1016/j.amjmed.2008.10.013

Jameson, S. A., Swaminathan, G., Dahal, S., Couri, B., Kuang, M., Rietsch, A., Butler, R. S., Ramamurthi, A., & Damaser, M. S. (2020). Elastin homeostasis is altered with pelvic organ prolapse in cultures of vaginal cells from a lysyl oxidase-like 1 knockout mouse model. *Physiological Reports*, *8*(11). https://doi.org/10.14814/phy2.14436

Jelovsek, J. E., & Barber, M. D. (2006). Women seeking treatment for advanced pelvic organ prolapse have decreased body image and quality of life. *American Journal of Obstetrics and Gynecology*, *194*(5), 1455–1461. https://doi.org/10.1016/j.ajog.2006.01.060

Jermy, B., Läll, K., Wolford, B., Wang, Y., Zguro, K., Cheng, Y., Kanai, M., Kanoni, S., Yang, Z., Hartonen, T., Monti, R., Wanner, J., Youssef, O., team, E. B. research, FinnGen, Lippert, C., Heel, D. van, Okada, Y., McCartney, D. L., … Ripatti, S. (2023). A unified framework for estimating country-specific cumulative incidence for 18 diseases stratified by polygenic risk. *MedRxiv*, 2023.06.12.23291186. https://doi.org/10.1101/2023.06.12.23291186

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Karjalainen, M. K., Karthikeyan, S., Oliver-Williams, C., Sliz, E., Allara, E., Fung, W. T., Surendran, P., Zhang, W., Jousilahti, P., Kristiansson, K., Salomaa, V., Goodwin, M., Hughes, D. A., Boehnke, M., Fernandes Silva, L., Yin, X., Mahajan, A., Neville, M. J., van Zuydam, N. R., … Kettunen, J. (2024). Genome-wide characterization of circulating metabolic biomarkers. *Nature*. https://doi.org/10.1038/s41586-024-07148-y

Kentistou, K. A., Kaisinger, L. R., Stankovic, S., Vaudel, M., de Oliveira, E. M., Messina, A., Walters, R. G., Liu, X., Busch, A. S., Helgason, H., Thompson, D. J., Santon, F., Petricek, K. M., Zouaghi, Y., Huang-Doran, I., Gudbjartsson, D. F., Bratland, E., Lin, K., Gardner, E. J., … Ong, K. K. (2023). Understanding the genetic complexity of puberty timing across the allele frequency spectrum. *MedRxiv*. https://doi.org/10.1101/2023.06.14.23291322

Kerimov, N., Hayhurst, J. D., Peikova, K., Manning, J. R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M. P., Kuzmin, I., Trevanion, S. J., Burdett, T., Jupp, S., Parkinson, H., Papatheodorou, I., Yates, A. D., Zerbino, D. R., & Alasoo, K. (2021). A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nature Genetics*, *53*(9), 1290–1299. https://doi.org/10.1038/s41588-021-00924-w

Kerimov, N., Tambets, R., Hayhurst, J. D., Rahu, I., Kolberg, P., Raudvere, U., Kuzmin, I., Chowdhary, A., Vija, A., Teras, H. J., Kanai, M., Ulirsch, J., Ryten, M., Hardy, J., Guelfi, S., Trabzuni, D., Kim-Hellmuth, S., Rayner, W., Finucane, H., … Alasoo, K. (2023). eQTL Catalogue 2023: New datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genetics*, *19*(9), e1010932. https://doi.org/10.1371/journal.pgen.1010932

Kerkhof, M. H., Hendriks, L., & Brölmann, H. A. M. (2009). Changes in connective tissue in patients with pelvic organ prolapse – A review of the current literature. In *International Urogynecology Journal* (Vol. 20, Issue 4, pp. 461–474). Springer London. https://doi.org/10.1007/s00192-008-0737-1

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., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. In *Nature Genetics* (Vol. 50, Issue 9, pp. 1219–1224). Nature Publishing Group. https://doi.org/10.1038/s41588-018-0183-z

Kiflen, M., Le, A., Mao, S., Lali, R., Narula, S., Xie, F., & Paré, G. (2022). Cost-Effectiveness of Polygenic Risk Scores to Guide Statin Therapy for Cardiovascular Disease Prevention. *Circulation. Genomic and Precision Medicine*, *15*(5), e003423. https://doi.org/10.1161/CIRCGEN.121.003423

Kilpivaara, O., Vahteristo, P., Falck, J., Syrjäkoski, K., Eerola, H., Easton, D., Bartkova, J., Lukas, J., Heikkilä, P., Aittomäki, K., Holli, K., Blomqvist, C., Kallioniemi, O.-P., Bartek, J., & Nevanlinna, H. (2004). CHEK2 variant I157T may be associated with increased breast cancer risk. *International Journal of Cancer*, *111*(4), 543–547. https://doi.org/10.1002/ijc.20299

Kimura, T., Ivell, R., Rust, W., Mizumoto, Y., Ogita, K., Kusui, C., Matsumura, Y., Azuma, C., & Murata, Y. (1999). Molecular cloning of a human MafF homologue, which specifically binds to the oxytocin receptor gene in term myometrium. *Biochemical and Biophysical Research Communications*, *264*(1), 86–92. https://doi.org/10.1006/bbrc.1999.1487

Kirby, A. C., Luber, K. M., & Menefee, S. A. (2013). An update on the current and future demand for care of pelvic floor disorders in the United States. *American Journal of Obstetrics and Gynecology*, *209*(6), 584.e1–584.e5. https://doi.org/10.1016/j.ajog.2013.09.011

Knoppers, B. M., Bernier, A., Granados Moreno, P., & Pashayan, N. (2021). Of Screening, Stratification, and Scores. *Journal of Personalized Medicine*, *11*(8). https://doi.org/10.3390/jpm11080736

Koel, M., Võsa, U., Jõeloo, M., Läll, K., Gualdo, N. P., Laivuori, H., Lemmelä, S., Daly, M., Palta, P., Mägi, R., & Laisk, T. (2023). GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. *Human Molecular Genetics*, *32*(12), 2103–2116. https://doi.org/10.1093/hmg/ddad043

Koivuaho, E., Laru, J., Ojaniemi, M., Puukka, K., Kettunen, J., Tapanainen, J. S., Franks, S., Järvelin, M.-R., Morin-Papunen, L., Sebert, S., & Piltonen, T. T. (2019). Age at adiposity rebound in childhood is associated with PCOS diagnosis and obesity in adulthood-longitudinal analysis of BMI data from birth to age 46 in cases of PCOS. *International Journal of Obesity (2005)*, *43*(7), 1370–1379. https://doi.org/10.1038/s41366-019-0318-z

Kow, N., Ridgeway, B., Kuang, M., Butler, R. S., & Damaser, M. S. (2016). Vaginal Expression of LOXL1 in Premenopausal and Postmenopausal Women with Pelvic Organ Prolapse. *Female Pelvic Medicine and Reconstructive Surgery*, *22*(4), 229–235. https://doi.org/10.1097/SPV.0000000000000251

Krebs, K., Bovijn, J., Zheng, N., Lepamets, M., Censin, J. C., Jürgenson, T., Särg, D., Abner, E., Laisk, T., Luo, Y., Skotte, L., Geller, F., Feenstra, B., Wang, W., Auton, A., Raychaudhuri, S., Esko, T., Metspalu, A., Laur, S., … Fadista, J. (2020). Genome-wide Study Identifies Association between HLA-B(∗)55:01 and Self-Reported Penicillin Allergy. *American Journal of Human Genetics*, *107*(4), 612–621. https://doi.org/10.1016/j.ajhg.2020.08.008

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., … Palotie, A. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, *613*(7944), 508–518. https://doi.org/10.1038/S41586-022-05473-8

Kurzrock, R., & Cohen, P. R. (2007). Polycystic ovary syndrome in men: Stein-Leventhal syndrome revisited. *Medical Hypotheses*, *68*(3), 480–483. https://doi.org/10.1016/j.mehy.2006.03.057

Kuusisto, K. M., Bebel, A., Vihinen, M., Schleutker, J., & Sallinen, S.-L. (2011). Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. *Breast Cancer Research : BCR*, *13*(1), R20. https://doi.org/10.1186/bcr2832

Lahti-Pulkkinen, M., Girchenko, P., Tuovinen, S., Sammallahti, S., Reynolds, R. M., Lahti, J., Heinonen, K., Lipsanen, J., Hämäläinen, E., Villa, P. M., Kajantie, E., Laivuori, H., & Räikkönen, K. (2020). Maternal Hypertensive Pregnancy Disorders and Mental Disorders in Children. *Hypertension (Dallas, Tex. : 1979)*, *75*(6), 1429–1438. https://doi.org/10.1161/HYPERTENSIONAHA.119.14140

Lali, R., Chong, M., Omidi, A., Mohammadi-Shemirani, P., Le, A., Cui, E., & Paré, G. (2021). Calibrated rare variant genetic risk scores for complex disease prediction using large exome sequence repositories. *Nature Communications*, *12*(1), 5852. https://doi.org/10.1038/s41467-021-26114-0

Läll, K., Lepamets, M., Palover, M., Esko, T., Metspalu, A., Tõnisson, N., Padrik, P., Mägi, R., & Fischer, K. (2019). Polygenic prediction of breast cancer: Comparison of genetic predictors and implications for risk stratification. *BMC Cancer*, *19*(1). https://doi.org/10.1186/s12885-019-5783-1

Lambalk, C. B., van Disseldorp, J., de Koning, C. H., & Broekmans, F. J. (2009). Testing ovarian reserve to predict age at menopause. *Maturitas*, *63*(4), 280–291. https://doi.org/10.1016/j.maturitas.2009.06.007

Lambert, S. A., Abraham, G., & Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Human Molecular Genetics*, *28*(R2), R133–R142. https://doi.org/10.1093/hmg/ddz187

Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., Danesh, J., MacArthur, J. A. L., & Inouye, M. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics*, *53*(4), 420–425. https://doi.org/10.1038/s41588-021-00783-5

Lavallée, G., Andelfinger, G., Nadeau, M., Lefebvre, C., Nemer, G., Horb, M. E., & Nemer, M. (2006). The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *EMBO Journal*, *25*(21), 5201–5213. https://doi.org/10.1038/sj.emboj.7601379

Leitsalu, L., Alavere, H., Tammesoo, M.-L., Leego, E., & Metspalu, A. (2015). Linking a population biobank with national health registries-the estonian experience. *Journal of Personalized Medicine*, *5*(2), 96–106. https://doi.org/10.3390/jpm5020096

Leitsalu, L., Palover, M., Sikka, T. T., Reigo, A., Kals, M., Pärn, K., Nikopensius, T., Esko, T., Metspalu, A., Padrik, P., & Tõnisson, N. (2021). Genotype-first approach to the detection of hereditary breast and ovarian cancer risk, and effects of risk disclosure to biobank participants. *European Journal of Human Genetics : EJHG*, *29*(3), 471–481. https://doi.org/10.1038/s41431-020-00760-2

Lewis, A. C. F., Green, R. C., & Vassy, J. L. (2021). Polygenic risk scores in the clinic: Translating risk into action. *HGG Advances*, *2*(4), 100047. https://doi.org/10.1016/j.xhgg.2021.100047

Lewis, A. C. F., Perez, E. F., Prince, A. E. R., Flaxman, H. R., Gomez, L., Brockman, D. G., Chandler, P. D., Kerman, B. J., Lebo, M. S., Smoller, J. W., Weiss, S. T., Blout Zawatksy, C. L., Meigs, J. B., Green, R. C., Vassy, J. L., & Karlson, E. W. (2022). Patient and provider perspectives on polygenic risk scores: implications for clinical reporting and utilization. *Genome Medicine*, *14*(1), 114. https://doi.org/10.1186/s13073-022-01117-8

Li, S., Carss, K. J., Halldorsson, B. V, Cortes, A., & Consortium, U. K. B. W.-G. S. (2023). Whole-genome sequencing of half-a-million UK Biobank participants. *MedRxiv*, 2023.12.06.23299426. https://doi.org/10.1101/2023.12.06.23299426

Li, W., Li, B., Li, T., Zhang, E., Wang, Q., Chen, S., & Sun, K. (2020). Identification and analysis of KLF13 variants in patients with congenital heart disease. *BMC Medical Genetics*, *21*(1). https://doi.org/10.1186/s12881-020-01009-x

Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., Gilad, Y., & Pritchard, J. K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science (New York, N.Y.)*, *352*(6285), 600–604. https://doi.org/10.1126/science.aad9417

Li, Yaqian, Zhang, Q.-Y., Sun, B.-F., Ma, Y., Zhang, Y., Wang, M., Ma, C., Shi, H., Sun, Z., Chen, J., Yang, Y.-G., & Zhu, L. (2021). Single-cell transcriptome profiling of the vaginal wall in women with severe anterior vaginal prolapse. *Nature Communications*, *12*(1), 87. https://doi.org/10.1038/s41467-020-20358-y

Li, You, Li, Z., Chen, R., Lian, M., Wang, H., Wei, Y., You, Z., Zhang, J., Li, B., Li, Y., Huang, B., Chen, Y., Liu, Q., Lyu, Z., Liang, X., Miao, Q., Xiao, X., Wang, Q., Fang, J., … Ma, X. (2023). A regulatory variant at 19p13.3 is associated with primary biliary cholangitis risk and ARID3A expression. *Nature Communications*, *14*(1), 1732. https://doi.org/10.1038/s41467-023-37213-5

Lim, S. S., Kakoly, N. S., Tan, J. W. J., Fitzgerald, G., Bahri Khomami, M., Joham, A. E., Cooray, S. D., Misso, M. L., Norman, R. J., Harrison, C. L., Ranasinha, S., Teede, H. J., & Moran, L. J. (2019). Metabolic syndrome in polycystic ovary syndrome: a systematic review, meta-analysis and meta-regression. *Obesity Reviews: An Official Journal of the International Association for the Study of Obesity*, *20*(2), 339–352. https://doi.org/10.1111/obr.12762

Lim, V. F., Khoo, J. K., Wong, V., & Moore, K. H. (2014). Recent studies of genetic dysfunction in pelvic organ prolapse: The role of collagen defects. In *Australian and New Zealand Journal of Obstetrics and Gynaecology* (Vol. 54, Issue 3, pp. 198–205). Blackwell Publishing. https://doi.org/10.1111/ajo.12169

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., … Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Liu, X., Zhao, Y., Gao, J., Pawlyk, B., Starcher, B., Spencer, J. A., Yanagisawa, H., Zuo, J., & Li, T. (2004). Elastic fiber homeostasis requires lysyl oxidase-like 1 protein. *Nature Genetics*, *36*(2), 178–182. https://doi.org/10.1038/ng1297

Loft, S., & Poulsen, H. E. (1996). Cancer risk and oxidative DNA damage in man. *Journal of Molecular Medicine (Berlin, Germany)*, *74*(6), 297–312. https://doi.org/10.1007/BF00207507

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. In *Nature Genetics* (Vol. 45, Issue 6, pp. 580–585). Nature Publishing Group. https://doi.org/10.1038/ng.2653

Lu, Q., Li, B., Ou, D., Erlendsdottir, M., Powles, R. L., Jiang, T., Hu, Y., Chang, D., Jin, C., Dai, W., He, Q., Liu, Z., Mukherjee, S., Crane, P. K., & Zhao, H. (2017). A Powerful Approach to Estimating Annotation-Stratified Genetic Covariance via GWAS Summary Statistics. *American Journal of Human Genetics*, *101*(6), 939–964. https://doi.org/10.1016/j.ajhg.2017.11.001

Luo, W., Mao, P., Zhang, L., Chen, X., & Yang, Z. (2020). Assessment of ovarian reserve by serum anti-Müllerian hormone in patients with systemic lupus erythematosus: a meta-analysis. *Annals of Palliative Medicine*, *9*(2), 207–215. https://doi.org/10.21037/apm.2020.02.11

Lutzmann, M., Bernex, F., da Costa de Jesus, C., Hodroj, D., Marty, C., Plo, I., Vainchenker, W., Tosolini, M., Forichon, L., Bret, C., Queille, S., Marchive, C., Hoffmann, J.-S., & Méchali, M. (2019). MCM8- and MCM9 Deficiencies Cause Lifelong Increased Hematopoietic DNA Damage Driving p53-Dependent Myeloid Tumors. *Cell Reports*, *28*(11), 2851–2865.e4. https://doi.org/10.1016/j.celrep.2019.07.095

Mägi, R., & Morris, A. P. (2010). GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, *11*, 288. https://doi.org/10.1186/1471-2105-11-288

Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X., & Sham, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genetic Epidemiology*, *41*(6), 469–480. https://doi.org/10.1002/gepi.22050

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. In *Nature* (Vol. 461, Issue 7265, pp. 747–753). NIH Public Access. https://doi.org/10.1038/nature08494

Mant, J., Painter, R., & Vessey, M. (1997). Epidemiology of genital prolapse: observations from the Oxford Family Planning Association Study. *British Journal of Obstetrics and Gynaecology*, *104*(5), 579–585. https://doi.org/10.1111/j.1471-0528.1997.tb11536.x

Mars, N., Lindbohm, J. V, Della Briotta Parolo, P., Widén, E., Kaprio, J., Palotie, A., & Ripatti, S. (2022). Systematic comparison of family history and polygenic risk across 24 common diseases. *American Journal of Human Genetics*, *109*(12), 2152–2162. https://doi.org/10.1016/j.ajhg.2022.10.009

Mars, N., Widén, E., Kerminen, S., Meretoja, T., Pirinen, M., Della Briotta Parolo, P., Palta, P., Palotie, A., Kaprio, J., Joensuu, H., Daly, M., & Ripatti, S. (2020). The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nature Communications*, *11*(1), 6383. https://doi.org/10.1038/s41467-020-19966-5

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Massrieh, W., Derjuga, A., Doualla-Bell, F., Ku, C. Y., Sanborn, B. M., & Blank, V. (2006). Regulation of the MAFF transcription factor by proinflammatory cytokines in myometrial cells. *Biology of Reproduction*, *74*(4), 699–705. https://doi.org/10.1095/biolreprod.105.045450

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., … Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)*, *337*(6099), 1190–1195. https://doi.org/10.1126/science.1222794

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E., & Marchini, J. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, *53*(7), 1097–1103. https://doi.org/10.1038/s41588-021-00870-7

McNestry, C., Killeen, S. L., Crowley, R. K., & McAuliffe, F. M. (2023). Pregnancy complications and later life women's health. *Acta Obstetricia et Gynecologica Scandinavica*, *102*(5), 523–531. https://doi.org/10.1111/aogs.14523

Mehta, L. S., Beckie, T. M., DeVon, H. A., Grines, C. L., Krumholz, H. M., Johnson, M. N., Lindley, K. J., Vaccarino, V., Wang, T. Y., Watson, K. E., & Wenger, N. K. (2016). Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association. *Circulation*, *133*(9), 916–947. https://doi.org/10.1161/CIR.0000000000000351

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., Elstrodt, F., van Duijn, C., Bartels, C., Meijers, C., Schutte, M., McGuffog, L., Thompson, D., Easton, D., Sodha, N., … Stratton, M. R. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature Genetics*, *31*(1), 55–59. https://doi.org/10.1038/ng879

Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., Best, S., Spector, T. D., Farrall, M., Lathrop, M., & Thein, S. L. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nature Genetics*, *39*(10), 1197–1199. https://doi.org/10.1038/ng2108

Mercuri, N. D., & Cox, B. J. (2022). The need for more research into reproductive health and disease. *ELife*, *11*. https://doi.org/10.7554/eLife.75061

Merkatz, R. B. (1998). Inclusion of women in clinical trials: a historical overview of scientific, ethical, and legal issues. *Journal of Obstetric, Gynecologic, and Neonatal Nursing: JOGNN*, *27*(1), 78–84. https://doi.org/10.1111/j.1552-6909.1998.tb02594.x

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., … Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284

Minooee, S., Ramezani Tehrani, F., Rahmati, M., Mansournia, M. A., & Azizi, F. (2018). Prediction of age at menopause in women with polycystic ovary syndrome. *Climacteric: The Journal of the International Menopause Society*, *21*(1), 29–34. https://doi.org/10.1080/13697137.2017.1392501

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., & 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*, *25*(7), 869–876. https://doi.org/10.1038/ejhg.2017.51

Moalli, P. A., Talarico, L. C., Sung, V. W., Klingensmith, W. L., Shand, S. H., Meyn, L. A., & Watkins, S. C. (2004). Impact of menopause on collagen subtypes in the arcus tendineous fasciae pelvis. *American Journal of Obstetrics and Gynecology*, *190*(3), 620–627. https://doi.org/10.1016/j.ajog.2003.08.040

Moghetti, P., & Tosi, F. (2021). Insulin resistance and PCOS: chicken or egg? *Journal of Endocrinological Investigation*, *44*(2), 233–244. https://doi.org/10.1007/s40618-020-01351-0

Mokry, L. E., Ross, S., Ahmad, O. S., Forgetta, V., Smith, G. D., Goltzman, D., Leong, A., Greenwood, C. M. T., Thanassoulis, G., & Richards, J. B. (2015). Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study. *PLoS Medicine*, *12*(8), e1001866. https://doi.org/10.1371/journal.pmed.1001866

Moolhuijsen, L. M. E., & Visser, J. A. (2020). Anti-Müllerian Hormone and Ovarian Reserve: Update on Assessing Ovarian Function. *The Journal of Clinical Endocrinology and Metabolism*, *105*(11), 3361–3373. https://doi.org/10.1210/clinem/dgaa513

Morris, J. A., Caragine, C., Daniloski, Z., Domingo, J., Barry, T., Lu, L., Davis, K., Ziosi, M., Glinos, D. A., Hao, S., Mimitou, E. P., Smibert, P., Roeder, K., Katsevich, E., Lappalainen, T., & Sanjana, N. E. (2023). Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science (New York, N.Y.)*, *380*(6646), eadh7699. https://doi.org/10.1126/science.adh7699

Mostafavi, H., Spence, J. P., Naqvi, S., & Pritchard, J. K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics*, *55*(11), 1866–1875. https://doi.org/10.1038/s41588-023-01529-1

Muranen, T. A., Greco, D., Blomqvist, C., Aittomäki, K., Khan, S., Hogervorst, F., Verhoef, S., Pharoah, P. D. P., Dunning, A. M., Shah, M., Luben, R., Bojesen, S. E., Nordestgaard, B. G., Schoemaker, M., Swerdlow, A., García-Closas, M., Figueroa, J., Dörk, T., Bogdanova, N. V, … Nevanlinna, H. (2017). Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *19*(5), 599–603. https://doi.org/10.1038/gim.2016.147

Mustofa, M. K., Tanoue, Y., Tateishi, C., Vaziri, C., & Tateishi, S. (2020). Roles of Chk2/CHEK2 in guarding against environmentally induced DNA damage and replication-stress. *Environmental and Molecular Mutagenesis*, *61*(7), 730–735. https://doi.org/10.1002/em.22397

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N. E., Ahfeldt, T., Sachs, K. V, Li, X., Li, H., Kuperwasser, N., Ruda, V. M., Pirruccello, J. P., Muchmore, B., Prokunina-Olsson, L., Hall, J. L., Schadt, E. E., Morales, C. R., Lund-Katz, S., Phillips, M. C., Wong, J., … Rader, D. J. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature*, *466*(7307), 714–719. https://doi.org/10.1038/nature09266

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., Murakami, Y., Yuji, K., Furukawa, Y., Zembutsu, H., Tanaka, T., Ohnishi, Y., Nakamura, Y., & Kubo, M. (2017). Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology*, *27*(3S), S2–S8. https://doi.org/10.1016/j.je.2016.12.005

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., Jones, T. R., Nguyen, T. H., Ulirsch, J. C., Lekschas, F., Mualim, K., Natri, H. M., Weeks, E. M., Munson, G., Kane, M., Kang, H. Y., Cui, A., Ray, J. P., Eisenhaure, T. M., … Engreitz, J. M. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature*, *593*(7858), 238–243. https://doi.org/10.1038/s41586-021-03446-x

Neumann, J. T., Riaz, M., Bakshi, A., Polekhina, G., Thao, L. T. P., Nelson, M. R., Woods, R. L., Abraham, G., Inouye, M., Reid, C. M., Tonkin, A. M., McNeil, J., & Lacaze, P. (2022). Prognostic Value of a Polygenic Risk Score for Coronary Heart Disease in Individuals Aged 70 Years and Older. *Circulation. Genomic and Precision Medicine*, *15*(1), e003429. https://doi.org/10.1161/CIRCGEN.121.003429

Neupane, R., Sadeghi, Z., Fu, R., Hagstrom, S. A., Moore, C. K., & Daneshgari, F. (2014). Mutation screen of LOXL1 in patients with female pelvic organ prolapse. *Female Pelvic Medicine and Reconstructive Surgery*, *20*(6), 316–321. https://doi.org/10.1097/SPV.0000000000000108

Newcombe, P. J., Nelson, C. P., Samani, N. J., & Dudbridge, F. (2019). A flexible and parallelizable approach to genome-wide polygenic risk scores. *Genetic Epidemiology*, *43*(7), 730–741. https://doi.org/10.1002/gepi.22245

Nichols, A. R., Chavarro, J. E., & Oken, E. (2024). Reproductive risk factors across the female lifecourse and later metabolic health. *Cell Metabolism*, *36*(2), 240–262. https://doi.org/10.1016/j.cmet.2024.01.002

Nielsen, M. W., Andersen, J. P., Schiebinger, L., & Schneider, J. W. (2017). One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nature Human Behaviour*, *1*(11), 791–796. https://doi.org/10.1038/s41562-017-0235-x

Ning, Z., Pawitan, Y., & Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics*, *52*(8), 859–864. https://doi.org/10.1038/s41588-020-0653-y

Nishinakamura, R., & Sakaguchi, M. (2014). BMP signaling and its modifiers in kidney development. *Pediatric Nephrology (Berlin, Germany)*, *29*(4), 681–686. https://doi.org/10.1007/s00467-013-2671-9

Nygaard, I., Bradley, C., & Brandt, D. (2004). Pelvic organ prolapse in older women: prevalence and risk factors. *Obstetrics and Gynecology*, *104*(3), 489–497. https://doi.org/10.1097/01.AOG.0000136100.10818.d8

O'Mara, T. A., Glubb, D. M., Amant, F., Annibali, D., Ashton, K., Attia, J., Auer, P. L., Beckmann, M. W., Black, A., Bolla, M. K., Brauch, H., Brenner, H., Brinton, L., Buchanan, D. D., Burwinkel, B., Chang-Claude, J., Chanock, S. J., Chen, C., Chen, M. M., … Thompson, D. J. (2018). Identification of nine new susceptibility loci for endometrial cancer. *Nature Communications*, *9*(1), 3166. https://doi.org/10.1038/s41467-018-05427-7

Ochoa, D., Karim, M., Ghoussaini, M., Hulcoop, D. G., McDonagh, E. M., & Dunham, I. (2022). Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. *Nature Reviews. Drug Discovery*, *21*(8), 551. https://doi.org/10.1038/d41573-022-00120-3

Ojalo, T., Haan, E., Kõiv, K., Kariis, H. M., Krebs, K., Uusberg, H., Sedman, T., Võsa, U., Puusepp, M., Lind, S., Hallik, I., Alavere, H., Milani, L., & Lehto, K. (2024). Cohort Profile Update: Mental Health Online Survey in the Estonian Biobank (EstBB MHoS). *International Journal of Epidemiology*, *53*(2). https://doi.org/10.1093/ije/dyae017

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., Hicks, B., Tian, C., Hinds, D. A., Ahlskog, R., Magnusson, P. K. E., Oskarsson, S., Hayward, C., Campbell, A., … Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, *54*(4), 437–449. https://doi.org/10.1038/s41588-022-01016-z

Olafsdottir, T., Thorleifsson, G., Sulem, P., Stefansson, O. A., Medek, H., Olafsson, K., Ingthorsson, O., Gudmundsson, V., Jonsdottir, I., Halldorsson, G. H., Kristjansson, R. P., Frigge, M. L., Stefansdottir, L., Sigurdsson, J. K., Oddsson, A., Sigurdsson, A., Eggertsson, H. P., Melsted, P., Halldorsson, B. V, … Stefansson, K. (2020). Genome-wide association identifies seven loci for pelvic organ prolapse in Iceland and the UK Biobank. *Communications Biology*, *3*(1), 129. https://doi.org/10.1038/s42003-020-0857-9

Ollila, M.-M. E., Piltonen, T., Puukka, K., Ruokonen, A., Järvelin, M.-R., Tapanainen, J. S., Franks, S., & Morin-Papunen, L. (2016). Weight Gain and Dyslipidemia in Early Adulthood Associate With Polycystic Ovary Syndrome: Prospective Cohort Study. *The Journal of Clinical Endocrinology and Metabolism*, *101*(2), 739–747. https://doi.org/10.1210/jc.2015-3543

One in three women with female health conditions forced to wait three years for diagnosis | The Independent. (n.d.). Retrieved April 2, 2024, from https://www.independent.co.uk/news/health/women-health-diagnosis-delay-treatment-b2280080.html

Örd, T., Lönnberg, T., Nurminen, V., Ravindran, A., Niskanen, H., Kiema, M., Õunap, K., Maria, M., Moreau, P. R., Mishra, P. P., Palani, S., Virta, J., Liljenbäck, H., Aavik, E., Roivainen, A., Ylä-Herttuala, S., Laakkonen, J. P., Lehtimäki, T., & Kaikkonen, M. U. (2023). Dissecting the polygenic basis of atherosclerosis via disease-associated cell state signatures. *American Journal of Human Genetics*, *110*(5), 722–740. https://doi.org/10.1016/j.ajhg.2023.03.013

Oxburgh, L., Brown, A. C., Muthukrishnan, S. D., & Fetting, J. L. (2014). Bone morphogenetic protein signaling in nephron progenitor cells. *Pediatric Nephrology (Berlin, Germany)*, *29*(4), 531–536. https://doi.org/10.1007/S00467-013-2589-2

Pardiñas, A. F., Kappel, D. B., Roberts, M., Tipple, F., Shitomi-Jones, L. M., King, A., Jansen, J., Helthuis, M., Owen, M. J., O'Donovan, M. C., & Walters, J. T. R. (2023). Pharmacokinetics and pharmacogenomics of clozapine in an ancestrally diverse sample: a longitudinal analysis and genome-wide association study using UK clinical monitoring data. *The Lancet. Psychiatry*, *10*(3), 209–219. https://doi.org/10.1016/S2215-0366(23)00002-0

Park, J., Long, D. T., Lee, K. Y., Abbas, T., Shibata, E., Negishi, M., Luo, Y., Schimenti, J. C., Gambus, A., Walter, J. C., & Dutta, A. (2013). The MCM8-MCM9 complex promotes RAD51 recruitment at DNA damage sites to facilitate homologous recombination. *Molecular and Cellular Biology*, *33*(8), 1632–1644. https://doi.org/10.1128/MCB.01503-12

Pathare, A. D. S., Pujol-Gualdo, N., Rukins, V., Džigurski, J., Peters, M., Team, E. B. R., Mägi, R., Salumets, A., Saare, M., & Laisk, T. (2024). Large-scale genome-wide association study to determine the genetic underpinnings of female genital tract polyps. *MedRxiv*, 2024.01.29.24301773. https://doi.org/10.1101/2024.01.29.24301773

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., Boehnke, M., Raychaudhuri, S., Fehrmann, R. S. N., Hirschhorn, J. N., & Franke, L. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*, *6*, 5890. https://doi.org/10.1038/ncomms6890

Pervjakova, N., Moen, G.-H., Borges, M.-C., Ferreira, T., Cook, J. P., Allard, C., Beaumont, R. N., Canouil, M., Hatem, G., Heiskala, A., Joensuu, A., Karhunen, V., Kwak, S. H., Lin, F. T. J., Liu, J., Rifas-Shiman, S., Tam, C. H., Tam, W. H., Thorleifsson, G., … Mägi, R. (2022). Multi-ancestry genome-wide association study of gestational diabetes mellitus highlights genetic links with type 2 diabetes. *Human Molecular Genetics*, *31*(19), 3377–3391. https://doi.org/10.1093/hmg/ddac050

Phelan, C. M., Kuchenbaecker, K. B., Tyrer, J. P., Kar, S. P., Lawrenson, K., Winham, S. J., Dennis, J., Pirie, A., Riggan, M. J., Chornokur, G., Earp, M. A., Lyra, P. C. J., Lee, J. M., Coetzee, S., Beesley, J., McGuffog, L., Soucy, P., Dicks, E., Lee, A., … Pharoah, P. D. P. (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nature Genetics*, *49*(5), 680–691. https://doi.org/10.1038/ng.3826

Pillinger, T., Osimo, E. F., de Marvao, A., Shah, M., Francis, C., Huang, J., D'Ambrosio, E., Firth, J., Nour, M. M., McCutcheon, R. A., Pardiñas, A. F., Matthews, P. M., O'Regan, D. P., & Howes, O. D. (2023). Effect of polygenic risk for schizophrenia on cardiac structure and function: a UK Biobank observational study. *The Lancet. Psychiatry*, *10*(2), 98–107. https://doi.org/10.1016/S2215-0366(22)00403-5

Piltonen, T., Morin-Papunen, L., Koivunen, R., Perheentupa, A., Ruokonen, A., & Tapanainen, J. S. (2005). Serum anti-Müllerian hormone levels remain high until late reproductive age and decrease during metformin therapy in women with polycystic ovary syndrome. *Human Reproduction (Oxford, England)*, *20*(7), 1820–1826. https://doi.org/10.1093/humrep/deh850

Pitzer, L. M., Moroney, M. R., Nokoff, N. J., & Sikora, M. J. (2021). WNT4 Balances Development vs Disease in Gynecologic Tissues and Women's Health. *Endocrinology*, *162*(7). https://doi.org/10.1210/endocr/bqab093

Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews. Genetics*, *11*(7), 459–463. https://doi.org/10.1038/nrg2813

Privé, F., Arbel, J., & Vilhjálmsson, B. J. (2021). LDpred2: better, faster, stronger. *Bioinformatics (Oxford, England)*, *36*(22–23), 5424–5431. https://doi.org/10.1093/bioinformatics/btaa1029

Pujol-Gualdo, N., Läll, K., Lepamets, M., Rossi, H.-R., Arffman, R. K., Piltonen, T. T., Mägi, R., & Laisk, T. (2022). Advancing our understanding of genetic risk factors and potential personalized strategies for pelvic organ prolapse. *Nature Communications*, *13*(1), 3584. https://doi.org/10.1038/s41467-022-31188-5

Pujol Gualdo, N., Mägi, R., & Laisk, T. (2023). Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy. *Human Reproduction (Oxford, England)*, *38*(12), 2516–2525. https://doi.org/10.1093/humrep/dead217

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–752. https://doi.org/10.1038/nature08185

Putting gender on the agenda. (2010). In *Nature* (Vol. 465, Issue 7299, p. 665). https://doi.org/10.1038/465665a

Rahmioglu, N., Mortlock, S., Ghiasi, M., Møller, P. L., Stefansdottir, L., Galarneau, G., Turman, C., Danning, R., Law, M. H., Sapkota, Y., Christofidou, P., Skarp, S., Giri, A., Banasik, K., Krassowski, M., Lepamets, M., Marciniak, B., Nõukas, M., Perro, D., … Zondervan, K. T. (2023). The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nature Genetics*, *55*(3), 423–436. https://doi.org/10.1038/s41588-023-01323-z

Rantakallio, P. (1988). The longitudinal study of the northern Finland birth cohort of 1966. *Paediatric and Perinatal Epidemiology*, *2*(1), 59–88. https://doi.org/10.1111/j.1365-3016.1988.tb00180.x

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., … Yosef, N. (2017). The Human Cell Atlas. *ELife*, *6*. https://doi.org/10.7554/eLife.27041

Reproductive health. (n.d.). Retrieved April 2, 2024, from https://www.who.int/westernpacific/health-topics/reproductive-health

Richardson, T. G., Harrison, S., Hemani, G., & Davey Smith, G. (2019). An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *ELife*, *8*. https://doi.org/10.7554/eLife.43657

Riise, H. K. R., Sulo, G., Tell, G. S., Igland, J., Egeland, G., Nygard, O., Selmer, R., Iversen, A.-C., & Daltveit, A. K. (2019). Hypertensive pregnancy disorders increase the risk of maternal cardiovascular disease after adjustment for cardiovascular risk factors. *International Journal of Cardiology*, *282*, 81–87. https://doi.org/10.1016/j.ijcard.2019.01.097

Risal, S., Pei, Y., Lu, H., Manti, M., Fornes, R., Pui, H.-P., Zhao, Z., Massart, J., Ohlsson, C., Lindgren, E., Crisosto, N., Maliqueo, M., Echiburú, B., Ladrón de Guevara, A., Sir-Petermann, T., Larsson, H., Rosenqvist, M. A., Cesta, C. E., Benrick, A., … Stener-Victorin, E. (2019). Prenatal androgen exposure and transgenerational susceptibility to polycystic ovary syndrome. *Nature Medicine*, *25*(12), 1894–1904. https://doi.org/10.1038/s41591-019-0666-1

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, *273*(5281), 1516–1517. https://doi.org/10.1126/science.273.5281.1516

Riveros-Mckay, F., Weale, M. E., Moore, R., Selzam, S., Krapohl, E., Sivley, R. M., Tarran, W. A., Sørensen, P., Lachapelle, A. S., Griffiths, J. A., Saffari, A., Deanfield, J., Spencer, C. C. A., Hippisley-Cox, J., Hunter, D. J., O'Sullivan, J. W., Ashley, E. A., Plagnol, V., & Donnelly, P. (2021). Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circulation. Genomic and Precision Medicine*, *14*(2), e003304. https://doi.org/10.1161/CIRCGEN.120.003304

Roux, A., Cholerton, R., Sicsic, J., Moumjid, N., French, D. P., Giorgi Rossi, P., Balleyguier, C., Guindy, M., Gilbert, F. J., Burrion, J.-B., Castells, X., Ritchie, D., Keatley, D., Baron, C., Delaloge, S., & de Montgolfier, S. (2022). Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the "My Personal Breast Screening" (MyPeBS) randomised clinical trial. *BMC Cancer*, *22*(1), 507. https://doi.org/10.1186/s12885-022-09484-6

Rubin, K. H., Glintborg, D., Nybo, M., Abrahamsen, B., & Andersen, M. (2017). Development and Risk Factors of Type 2 Diabetes in a Nationwide Population of Women With Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology and Metabolism*, *102*(10), 3848–3857. https://doi.org/10.1210/jc.2017-01354

Ruth, K. S., Day, F. R., Hussain, J., Martínez-Marchal, A., Aiken, C. E., Azad, A., Thompson, D. J., Knoblochova, L., Abe, H., Tarry-Adkins, J. L., Gonzalez, J. M., Fontanillas, P., Claringbould, A., Bakker, O. B., Sulem, P., Walters, R. G., Terao, C., Turon, S., Horikoshi, M., … Perry, J. R. B. (2021). Genetic insights into biological mechanisms governing human ovarian ageing. *Nature*, *596*(7872), 393–397. https://doi.org/10.1038/S41586-021-03779-7

Ruth, K. S., Day, F. R., Tyrrell, J., Thompson, D. J., Wood, A. R., Mahajan, A., Beaumont, R. N., Wittemans, L., Martin, S., Busch, A. S., Erzurumluoglu, A. M., Hollis, B., O'Mara, T. A., McCarthy, M. I., Langenberg, C., Easton, D. F., Wareham, N. J.,

Burgess, S., Murray, A., … Perry, J. R. B. (2020). Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine*, *26*(2), 252–258. https://doi.org/10.1038/s41591-020-0751-5

Ruth, K. S., Soares, A. L. G., Borges, M. C., Eliassen, A. H., Hankinson, S. E., Jones, M. E., Kraft, P., Nichols, H. B., Sandler, D. P., Schoemaker, M. J., Taylor, J. A., Zeleniuch-Jacquotte, A., Lawlor, D. A., Swerdlow, A. J., & Murray, A. (2019). Genome-wide association study of anti-Müllerian hormone levels in pre-menopausal women of late reproductive age and relationship with genetic determinants of reproductive lifespan. *Human Molecular Genetics*, *28*(8), 1392–1401. https://doi.org/10.1093/HMG/DDZ015

Saliba, J., Coutaud, B., Solovieva, V., Lu, F., & Blank, V. (2019). Regulation of CXCL1 chemokine and CSF3 cytokine levels in myometrial cells by the MAFF transcription factor. *Journal of Cellular and Molecular Medicine*, *23*(4), 2517–2525. https://doi.org/10.1111/jcmm.14136

Salonen, J. T., Uimari, P., Aalto, J.-M., Pirskanen, M., Kaikkonen, J., Todorova, B., Hyppönen, J., Korhonen, V.-P., Asikainen, J., Devine, C., Tuomainen, T.-P., Luedemann, J., Nauck, M., Kerner, W., Stephens, R. H., New, J. P., Ollier, W. E., Gibson, J. M., Payton, A., … Darvasi, A. (2007). Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *American Journal of Human Genetics*, *81*(2), 338–345. https://doi.org/10.1086/520599

Sam, S., Coviello, A. D., Sung, Y.-A., Legro, R. S., & Dunaif, A. (2008). Metabolic phenotype in the brothers of women with polycystic ovary syndrome. *Diabetes Care*, *31*(6), 1237–1241. https://doi.org/10.2337/dc07-2190

Sampogna, C. (2006). *Creation and governance of human genetic research databases*. OECD Publishing.

Schuh-Huerta, S. M., Johnson, N. A., Rosen, M. P., Sternfeld, B., Cedars, M. I., & Reijo Pera, R. A. (2012). Genetic markers of ovarian follicle number and menopause in women of multiple ethnicities. *Human Genetics*, *131*(11), 1709–1724. https://doi.org/10.1007/s00439-012-1184-0

Schutte, M., Seal, S., Barfoot, R., Meijers-Heijboer, H., Wasielewski, M., Evans, D. G., Eccles, D., Meijers, C., Lohman, F., Klijn, J., van den Ouweland, A., Futreal, P. A., Nathanson, K. L., Weber, B. L., Easton, D. F., Stratton, M. R., & Rahman, N. (2003). Variants in CHEK2 other than 1100delC do not make a major contribution to breast cancer susceptibility. *American Journal of Human Genetics*, *72*(4), 1023–1028. https://doi.org/10.1086/373965

Sharma, M., Castro-Piedras, I., Simmons, G. E., & Pruitt, K. (2018). Dishevelled: A masterful conductor of complex Wnt signals. In *Cellular Signalling* (Vol. 47, pp. 52–64). Elsevier Inc. https://doi.org/10.1016/j.cellsig.2018.03.004

Shekari, S., Stankovic, S., Gardner, E. J., Hawkes, G., Kentistou, K. A., Beaumont, R. N., Mörseburg, A., Wood, A. R., Prague, J. K., Mishra, G. D., Day, F. R., Baptista, J., Wright, C. F., Weedon, M. N., Hoffmann, E. R., Ruth, K. S., Ong, K. K., Perry, J. R. B., & Murray, A. (2023). Penetrance of pathogenic genetic variants associated with premature ovarian insufficiency. *Nature Medicine*, *29*(7), 1692–1699. https://doi.org/10.1038/s41591-023-02405-5

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *American Journal of Human Genetics*, *99*(1), 139–153. https://doi.org/10.1016/j.ajhg.2016.05.013

Shi, Y., Zhao, H., Shi, Y., Cao, Y., Yang, D., Li, Z., Zhang, B., Liang, X., Li, T., Chen, J., Shen, J., Zhao, J., You, L., Gao, X., Zhu, D., Zhao, X., Yan, Y., Qin, Y., Li, W., … Chen, Z.-J. (2012). Genome-wide association study identifies eight new risk loci for

polycystic ovary syndrome. *Nature Genetics*, *44*(9), 1020–1025. https://doi.org/10.1038/ng.2384

Shoaib, M., Ye, Q., IglayReger, H., Tan, M. H., Boehnke, M., Burant, C. F., Soleimanpour, S. A., & Gagliano Taliun, S. A. (2023). Evaluation of polygenic risk scores to differentiate between type 1 and type 2 diabetes. *Genetic Epidemiology*, *47*(4), 303–313. https://doi.org/10.1002/gepi.22521

Sijtsma, A., Rienks, J., van der Harst, P., Navis, G., Rosmalen, J. G. M., & Dotinga, A. (2022). Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *51*(5), e295–e302. https://doi.org/10.1093/ije/dyab257

Silarova, B., Sharp, S., Usher-Smith, J. A., Lucas, J., Payne, R. A., Shefer, G., Moore, C., Girling, C., Lawrence, K., Tolkien, Z., Walker, M., Butterworth, A., Di Angelantonio, E., Danesh, J., & Griffin, S. J. (2019). Effect of communicating phenotypic and genetic risk of coronary heart disease alongside web-based lifestyle advice: the INFORM Randomised Controlled Trial. *Heart (British Cardiac Society)*, *105*(13), 982–989. https://doi.org/10.1136/heartjnl-2018-314211

Silva, M. S. B., & Giacobini, P. (2021). New insights into anti-Müllerian hormone role in the hypothalamic-pituitary-gonadal axis and neuroendocrine development. *Cellular and Molecular Life Sciences: CMLS*, *78*(1), 1–16. https://doi.org/10.1007/s00018-020-03576-x

Sims, E. K., Besser, R. E. J., Dayan, C., Geno Rasmussen, C., Greenbaum, C., Griffin, K. J., Hagopian, W., Knip, M., Long, A. E., Martin, F., Mathieu, C., Rewers, M., Steck, A. K., Wentworth, J. M., Rich, S. S., Kordonouri, O., Ziegler, A.-G., & Herold, K. C. (2022). Screening for Type 1 Diabetes in the General Population: A Status Report and Perspective. *Diabetes*, *71*(4), 610–623. https://doi.org/10.2337/dbi20-0054

Slatkin, M. (2008). Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews. Genetics*, *9*(6), 477–485. https://doi.org/10.1038/nrg2361

Smith, G. D., & Ebrahim, S. (2003). "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? In *International journal of epidemiology* (Vol. 32, Issue 1, pp. 1–22). https://doi.org/10.1093/ije/dyg070

Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., & Elliott, L. T. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature Neuroscience*, *24*(5), 737–745. https://doi.org/10.1038/s41593-021-00826-4

Smoller, J. W., Kendler, K. K., Craddock, N., Lee, P. H., Neale, B. M., Nurnberger, J. N., Ripke, S., Santangelo, S., Sullivan, P. S., Neale, B. N., Purcell, S., Anney, R., Buitelaar, J., Fanous, A., Faraone, S. F., Hoogendijk, W., Lesch, K. P., Levinson, D. L., Perlis, R. P., … O'Donovan, M. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet (London, England)*, *381*(9875), 1371–1379. https://doi.org/10.1016/S0140-6736(12)62129-1

Solé-Navais, P., Flatley, C., Steinthorsdottir, V., Vaudel, M., Juodakis, J., Chen, J., Laisk, T., LaBella, A. L., Westergaard, D., Bacelis, J., Brumpton, B., Skotte, L., Borges, M. C., Helgeland, Ø., Mahajan, A., Wielscher, M., Lin, F., Briggs, C., Wang, C. A., … Consortium, D. B. D. S. G. (2023). Genetic effects on the timing of parturition and links to fetal birth weight. *Nature Genetics*, *55*(4), 559–567. https://doi.org/10.1038/s41588-023-01343-9

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., … Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(D1), D977–D985. https://doi.org/10.1093/nar/gkac1010

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews. Genetics*, *14*(7), 483–495. https://doi.org/10.1038/nrg3461

Speed, D., & Balding, D. J. (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nature Genetics*, *51*(2), 277–284. https://doi.org/10.1038/s41588-018-0279-5

Stacey, D., Fauman, E. B., Ziemek, D., Sun, B. B., Harshfield, E. L., Wood, A. M., Butterworth, A. S., Suhre, K., & Paul, D. S. (2019). ProGeM: A framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Research*, *47*(1). https://doi.org/10.1093/nar/gky837

Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics*, *186*(3), 767–773. https://doi.org/10.1534/genetics.110.122549

Stefansdottir, V., Thorolfsdottir, E., Hognason, H. B., Patch, C., van El, C., Hentze, S., Cordier, C., Mendes, Á., & Jonsson, J. J. (2020). Web-based return of BRCA2 research results: one-year genetic counselling experience in Iceland. *European Journal of Human Genetics : EJHG*, *28*(12), 1656–1661. https://doi.org/10.1038/s41431-020-0665-1

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, *21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Stolk, L., Perry, J. R. B., Chasman, D. I., He, C., Mangino, M., Sulem, P., Barbalic, M., Broer, L., Byrne, E. M., Ernst, F., Esko, T., Franceschini, N., Gudbjartsson, D. F., Hottenga, J.-J., Kraft, P., McArdle, P. F., Porcu, E., Shin, S.-Y., Smith, A. V, … Lunetta, K. L. (2012). Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nature Genetics*, *44*(3), 260–268. https://doi.org/10.1038/ng.1051

Sun, L., Pennells, L., Kaptoge, S., Nelson, C. P., Ritchie, S. C., Abraham, G., Arnold, M., Bell, S., Bolton, T., Burgess, S., Dudbridge, F., Guo, Q., Sofianopoulou, E., Stevens, D., Thompson, J. R., Butterworth, A. S., Wood, A., Danesh, J., Samani, N. J., … Di Angelantonio, E. (2021). Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. *PLoS Medicine*, *18*(1), e1003498. https://doi.org/10.1371/journal.pmed.1003498

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S.-B., Tian, X., Browning, B. L., Das, S., Emde, A.-K., Clarke, W. E., Loesch, D. P., … Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, *590*(7845), 290–299. https://doi.org/10.1038/s41586-021-03205-y

Täufer Cederlöf, E., Lundgren, M., Lindahl, B., & Christersson, C. (2022). Pregnancy Complications and Risk of Cardiovascular Disease Later in Life: A Nationwide Cohort Study. *Journal of the American Heart Association*, *11*(2), e023079. https://doi.org/10.1161/JAHA.121.023079

Teede, H. J., Tay, C. T., Laven, J. J. E., Dokras, A., Moran, L. J., Piltonen, T. T., Costello, M. F., Boivin, J., Redman, L. M., Boyle, J. A., Norman, R. J., Mousa, A., & Joham, A. E. (2023). Recommendations From the 2023 International Evidence-based Guideline for the Assessment and Management of Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology and Metabolism*, *108*(10), 2447–2469. https://doi.org/10.1210/clinem/dgad463

Tharp, M. E., Malki, S., & Bortvin, A. (2020). Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. *Nature Communications*, *11*(1), 330. https://doi.org/10.1038/s41467-019-14055-8

Tobler, K. J., Shoham, G., Christianson, M. S., Zhao, Y., Leong, M., & Shoham, Z. (2015). Use of anti-mullerian hormone for testing ovarian reserve: a survey of 796 infertility clinics worldwide. *Journal of Assisted Reproduction and Genetics*, *32*(10), 1441–1448. https://doi.org/10.1007/S10815-015-0562-7

Trajanoska, K., Bhérer, C., Taliun, D., Zhou, S., Richards, J. B., & Mooser, V. (2023). From target discovery to clinical drug development with human genetics. *Nature*, *620*(7975), 737–745. https://doi.org/10.1038/s41586-023-06388-8

Turley, P., Meyer, M. N., Wang, N., Cesarini, D., Hammonds, E., Martin, A. R., Neale, B. M., Rehm, H. L., Wilkins-Haug, L., Benjamin, D. J., Hyman, S., Laibson, D., & Visscher, P. M. (2021). Problems with Using Polygenic Scores to Select Embryos. *The New England Journal of Medicine*, *385*(1), 78–86. https://doi.org/10.1056/NEJMsr2105065

Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., de Andrade, M., Doheny, K. F., Haines, J. L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I. J., Li, R., Ling, H., Manolio, T. A., Matsumoto, M., McCarty, C. A., McDavid, A. N., … Ritchie, M. D. (2011). Quality control procedures for genome-wide association studies. *Current Protocols in Human Genetics*, *Chapter 1*, Unit1.19. https://doi.org/10.1002/0471142905.hg0119s68

Tyrmi, J. S., Arffman, R. K., Pujol-Gualdo, N., Kurra, V., Morin-Papunen, L., Sliz, E., Piltonen, T. T., Laisk, T., Kettunen, J., & Laivuori, H. (2022). Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome. *Human Reproduction (Oxford, England)*, *37*(2), 352–365. https://doi.org/10.1093/humrep/deab250

Tyrmi, J. S., Kaartokallio, T., Lokki, A. I., Jääskeläinen, T., Kortelainen, E., Ruotsalainen, S., Karjalainen, J., Ripatti, S., Kivioja, A., Laisk, T., Kettunen, J., Pouta, A., Kivinen, K., Kajantie, E., Heinonen, S., Kere, J., & Laivuori, H. (2023). Genetic Risk Factors Associated With Preeclampsia and Hypertensive Disorders of Pregnancy. *JAMA Cardiology*, *8*(7), 674–683. https://doi.org/10.1001/jamacardio.2023.1312

Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V. G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., Piras, M. G., Sestu, N., Lai, S., Dei, M., Mulas, A., Crisponi, L., Naitza, S., Asunis, I., Deiana, M., … Cao, A. (2008). Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1620–1625. https://doi.org/10.1073/pnas.0711566105

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59. https://doi.org/10.1038/s43586-021-00056-9

Ushiki, A., Sheng, R. R., Zhang, Y., Zhao, J., Nobuhara, M., Murray, E., Ruan, X., Rios, J. J., Wise, C. A., & Ahituv, N. (2024). Deletion of Pax1 scoliosis-associated regulatory elements leads to a female-biased tail abnormality. *Cell Reports*, *43*(3), 113907. https://doi.org/10.1016/j.celrep.2024.113907

Value gender and equity in the global health workforce. (n.d.). Retrieved April 2, 2024, from https://www.who.int/activities/value-gender-and-equity-in-the-global-health-workforce

van der Wijst, M., de Vries, D. H., Groot, H. E., Trynka, G., Hon, C. C., Bonder, M. J., Stegle, O., Nawijn, M. C., Idaghdour, Y., van der Harst, P., Ye, C. J., Powell, J., Theis, F. J., Mahfouz, A., Heinig, M., & Franke, L. (2020). The single-cell eQTLGen consortium. *ELife*, *9*. https://doi.org/10.7554/eLife.52155

Variations in reproductive events across life: a pooled analysis of data from 505 147 women across 10 countries. (2019). *Human Reproduction (Oxford, England)*, *34*(5), 881–893. https://doi.org/10.1093/humrep/dez015

Velecela, V., Lettice, L. A., Chau, Y. Y., Slight, J., Berry, R. L., Thornburn, A., Gunst, Q. D., Van Den Hoff, M., Reina, M., Martínez, F. O., Hastie, N. D., & Martínez-Estrada, O. M. (2013). WT1 regulates the expression of inhibitory chemokines during heart development. *Human Molecular Genetics*, *22*(25), 5083–5095. https://doi.org/10.1093/hmg/ddt358

Venkatesh, S. S., Ganjgahi, H., Palmer, D. S., Coley, K., Wittemans, L. B. L., Nellaker, C., Holmes, C., Lindgren, C. M., & Nicholson, G. (2023). The genetic architecture of changes in adiposity during adulthood. In *medRxiv : the preprint server for health sciences*. https://doi.org/10.1101/2023.01.09.23284364

Venkatesh, S. S., Wittemans, L. B. L., Palmer, D. S., Baya, N. A., Ferreira, T., Hill, B., Lassen, F. H., Parker, M. J., Reibe, S., Elhakeem, A., Banasik, K., Bruun, M. T., Erikstrup, C., Jensen, B. A., Juul, A., Mikkelsen, C., Nielsen, H. S., Ostrowski, S. R., Pedersen, O. B., … Lindgren, C. M. (2024). Genome-wide analyses identify 21 infertility loci and over 400 reproductive hormone loci across the allele frequency spectrum. *MedRxiv*. https://doi.org/10.1101/2024.03.19.24304530

Verdiesen, R. M. G., van der Schouw, Y. T., van Gils, C. H., Verschuren, W. M. M., Broekmans, F. J. M., Borges, M. C., Gonçalves Soares, A. L., Lawlor, D. A., Eliassen, A. H., Kraft, P., Sandler, D. P., Harlow, S. D., Smith, J. A., Santoro, N., Schoemaker, M. J., Swerdlow, A. J., Murray, A., Ruth, K. S., & Onland-Moret, N. C. (2022). Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women. *Human Reproduction (Oxford, England)*, *37*(5), 1069–1082. https://doi.org/10.1093/humrep/deac028

Vergeldt, T. F. M., Weemhoff, M., IntHout, J., & Kluivers, K. B. (2015). Risk factors for pelvic organ prolapse and its recurrence: a systematic review. *International Urogynecology Journal*, *26*(11), 1559–1573. https://doi.org/10.1007/s00192-015-2695-8

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P. R., Bhatia, G., Do, R., Hayeck, T., Won, H. H., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., … Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Human Genetics*, *97*(4), 576–592. https://doi.org/10.1016/j.ajhg.2015.09.001

Vink, J. M., Sadrzadeh, S., Lambalk, C. B., & Boomsma, D. I. (2006). Heritability of polycystic ovary syndrome in a Dutch twin-family study. *The Journal of Clinical Endocrinology and Metabolism*, *91*(6), 2100–2104. https://doi.org/10.1210/jc.2005-1494

Vipin, V. P., Dabadghao, P., Shukla, M., Kapoor, A., Raghuvanshi, A. S., & Ramesh, V. (2016). Cardiovascular disease risk in first-degree relatives of women with polycystic ovary syndrome. *Fertility and Sterility*, *105*(5), 1338–1344.e3. https://doi.org/10.1016/j.fertnstert.2016.01.024

Von Scheidt, M., Zhao, Y., De Aguiar Vallim, T. Q., Che, N., Wierer, M., Seldin, M. M., Franzén, O., Kurt, Z., Pang, S., Bongiovanni, D., Yamamoto, M., Edwards, P. A., Ruusalepp, A., Kovacic, J. C., Mann, M., Björkegren, J. L. M., Lusis, A. J., Yang, X., & Schunkert, H. (2021). Transcription Factor MAFF (MAF Basic Leucine Zipper Transcription Factor F) Regulates an Atherosclerosis Relevant Network Connecting Inflammation and Cholesterol Metabolism. *Circulation*, *143*(18), 1809–1823. https://doi.org/10.1161/CIRCULATIONAHA.120.050186

Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nature Reviews. Genetics*, *12*(3), 204–213. https://doi.org/10.1038/nrg2949

Walters, R. G., Millwood, I. Y., Lin, K., Schmidt Valle, D., McDonnell, P., Hacker, A., Avery, D., Edris, A., Fry, H., Cai, N., Kretzschmar, W. W., Ansari, M. A., Lyons, P. A., Collins, R., Donnelly, P., Hill, M., Peto, R., Shen, H., Jin, X., … Chen, Z. (2023). Genotyping and population characteristics of the China Kadoorie Biobank. *Cell Genomics*, *3*(8), 100361. https://doi.org/10.1016/j.xgen.2023.100361

Wand, H., Lambert, S. A., Tamburro, C., Iacocca, M. A., O'Sullivan, J. W., Sillari, C., Kullo, I. J., Rowley, R., Dron, J. S., Brockman, D., Venner, E., McCarthy, M. I., Antoniou, A. C., Easton, D. F., Hegele, R. A., Khera, A. V, Chatterjee, N., Kooperberg, C., Edwards, K., … Wojcik, G. L. (2021). Improving reporting standards for polygenic scores in risk prediction studies. *Nature*, *591*(7849), 211–219. https://doi.org/10.1038/s41586-021-03243-6

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. https://doi.org/10.1093/nar/gkq603

Wang, Y., Namba, S., Lopera, E., Kerminen, S., Tsuo, K., Läll, K., Kanai, M., Zhou, W., Wu, K.-H., Favé, M.-J., Bhatta, L., Awadalla, P., Brumpton, B., Deelen, P., Hveem, K., Lo Faro, V., Mägi, R., Murakami, Y., Sanna, S., … Hirbo, J. (2023). Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. *Cell Genomics*, *3*(1), 100241. https://doi.org/10.1016/j.xgen.2022.100241

Wang, Z., Choi, S. W., Chami, N., Boerwinkle, E., Fornage, M., Redline, S., Bis, J. C., Brody, J. A., Psaty, B. M., Kim, W., McDonald, M.-L. N., Regan, E. A., Silverman, E. K., Liu, C.-T., Vasan, R. S., Kalyani, R. R., Mathias, R. A., Yanek, L. R., Arnett, D. K., … Loos, R. J. F. (2022). The Value of Rare Genetic Variation in the Prediction of Common Obesity in European Ancestry Populations. *Frontiers in Endocrinology*, *13*, 863893. https://doi.org/10.3389/fendo.2022.863893

Ward, L. D., Parker, M. M., Deaton, A. M., Tu, H.-C., Flynn-Carroll, A. O., Hinkle, G., & Nioi, P. (2022). Rare coding variants in DNA damage repair genes associated with timing of natural menopause. *HGG Advances*, *3*(2), 100079. https://doi.org/10.1016/j.xhgg.2021.100079

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., & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, *51*(9), 1339–1348. https://doi.org/10.1038/s41588-019-0481-0

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Watanabe, K., Taskesen, E., Van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1). https://doi.org/10.1038/S41467-017-01261-5

Watson, S., Caster, O., Rochon, P. A., & den Ruijter, H. (2019). Reported adverse drug reactions in women and men: Aggregated evidence from globally collected individual case reports during half a century. *EClinicalMedicine*, *17*, 100188. https://doi.org/10.1016/j.eclinm.2019.10.001

Weale, M. E., Riveros-Mckay, F., Selzam, S., Seth, P., Moore, R., Tarran, W. A., Gradovich, E., Giner-Delgado, C., Palmer, D., Wells, D., Saffari, A., Sivley, R. M., Lachapelle, A. S., Wand, H., Clarke, S. L., Knowles, J. W., O'Sullivan, J. W., Ashley, E. A., McVean, G., … Donnelly, P. (2021). Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. *The American Journal of Cardiology*, *148*, 157–164. https://doi.org/10.1016/j.amjcard.2021.02.032

Weeks, E. M., Ulirsch, J. C., Cheng, N. Y., Trippe, B. L., Fine, R. S., Miao, J., Patwardhan, T. A., Kanai, M., Nasser, J., Fulco, C. P., Tashman, K. C., Aguet, F., Li, T., Ordovas-Montanes, J., Smillie, C. S., Biton, M., Shalek, A. K., Ananthakrishnan, A. N., Xavier, R. J., … Finucane, H. K. (2023). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nature Genetics*, *55*(8), 1267–1276. https://doi.org/10.1038/s41588-023-01443-6

Weenen, C., Laven, J. S. E., Von Bergh, A. R. M., Cranfield, M., Groome, N. P., Visser, J. A., Kramer, P., Fauser, B. C. J. M., & Themmen, A. P. N. (2004). Anti-Müllerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Molecular Human Reproduction*, *10*(2), 77–83. https://doi.org/10.1093/molehr/gah015

Wen, Y., Wu, X., Peng, H., Li, C., Jiang, Y., Su, Z., Liang, H., Liu, J., He, J., & Liang, W. (2021). Breast cancer risk in patients with polycystic ovary syndrome: a Mendelian randomization analysis. *Breast Cancer Research and Treatment*, *185*(3), 799–806. https://doi.org/10.1007/s10549-020-05973-z

Werme, J., van der Sluis, S., Posthuma, D., & de Leeuw, C. A. (2022). An integrated framework for local genetic correlation analysis. *Nature Genetics*, *54*(3), 274–282. https://doi.org/10.1038/s41588-022-01017-y

Westergaard, D., Steinthorsdottir, V., Stefansdottir, L., Rohde, P. D., Wu, X., Geller, F., Tyrmi, J., Havulinna, A. S., Navais, P. S., Flatley, C., Ostrowski, S. R., Pedersen, O. B., Erikstrup, C., Sørensen, E., Mikkelsen, C., Brun, M. T., Jensen, B. A., Brodersen, T., Ullum, H., … Nielsen, H. S. (2023). Pregnancy-Associated Bleeding and Genetics: Five Sequence Variants in the Myometrium and Progesterone Signaling Pathway are associated with postpartum hemorrhage. In *medRxiv : the preprint server for health sciences*. https://doi.org/10.1101/2023.08.10.23293932

Widén, E., Junna, N., Ruotsalainen, S., Surakka, I., Mars, N., Ripatti, P., Partanen, J. J., Aro, J., Mustonen, P., Tuomi, T., Palotie, A., Salomaa, V., Kaprio, J., Partanen, J., Hotakainen, K., Pöllänen, P., & Ripatti, S. (2022). How Communicating Polygenic and Clinical Risk for Atherosclerotic Cardiovascular Disease Impacts Health Behavior: an Observational Follow-up Study. *Circulation. Genomic and Precision Medicine*, *15*(2), e003459. https://doi.org/10.1161/CIRCGEN.121.003459

Wilcox, N., Dumont, M., González-Neira, A., Carvalho, S., Joly Beauparlant, C., Crotti, M., Luccarini, C., Soucy, P., Dubois, S., Nuñez-Torres, R., Pita, G., Gardner, E. J., Dennis, J., Alonso, M. R., Álvarez, N., Baynes, C., Collin-Deschesnes, A. C., Desjardins, S., Becher, H., … Simard, J. (2023). Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. *Nature Genetics*, *55*(9), 1435–1439. https://doi.org/10.1038/s41588-023-01466-z

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Wojciechowski, P., Lipowska, A., Rys, P., Ewens, K. G., Franks, S., Tan, S., Lerchbaum, E., Vcelak, J., Attaoua, R., Straczkowski, M., Azziz, R., Barber, T. M., Hinney, A., Obermayer-Pietsch, B., Lukasova, P., Bendlova, B., Grigorescu, F., Kowalska, I., Goodarzi, M. O., … Malecki, M. T. (2012). Impact of FTO genotypes on BMI and weight in polycystic ovary syndrome: a systematic review and meta-analysis. *Diabetologia*, *55*(10), 2636–2645. https://doi.org/10.1007/s00125-012-2638-6

Women's health research lacks funding – these charts show how. (n.d.). Retrieved April 2, 2024, from https://www.nature.com/immersive/d41586-023-01475-2/index.html

Wong, J. Z. Y., Chai, J. H., Yeoh, Y. S., Mohamed Riza, N. K., Liu, J., Teo, Y.-Y., Wee, H. L., & Hartman, M. (2021). Cost effectiveness analysis of a polygenic risk tailored breast cancer screening programme in Singapore. *BMC Health Services Research*, *21*(1), 379. https://doi.org/10.1186/s12913-021-06396-2

Wu, P.-F., Li, R.-Z., Zhang, W., Hu, H.-Y., Wang, W., & Lin, Y. (2020). Polycystic ovary syndrome is causally associated with estrogen receptor-positive instead of estrogen receptor-negative breast cancer: a Mendelian randomization study. In *American journal of obstetrics and gynecology* (Vol. 223, Issue 4, pp. 583–585). https://doi.org/10.1016/j.ajog.2020.05.016

Xiang, R., Kelemen, M., Xu, Y., Harris, L. W., Parkinson, H., Inouye, M., & Lambert, S. A. (2024). Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. *Genome Medicine*, *16*(1), 33. https://doi.org/10.1186/s13073-024-01304-9

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A. U., Jiang, Y., Raghavan, S., Miao, J., Arias, J. D., Graham, S. E., Mukamel, R. E., Spracklen, C. N., Yin, X., Chen, S.-H., Ferreira, T., Highland, H. H., … Hirschhorn, J. N. (2022). A saturated map of common genetic variants associated with human height. *Nature*, *610*(7933), 704–712. https://doi.org/10.1038/s41586-022-05275-y

Yildiz, B. O., Yarali, H., Oguz, H., & Bayraktar, M. (2003). Glucose intolerance, insulin resistance, and hyperandrogenemia in first degree relatives of women with polycystic ovary syndrome. *The Journal of Clinical Endocrinology and Metabolism*, *88*(5), 2031–2036. https://doi.org/10.1210/jc.2002-021499

Yilmaz, B., Vellanki, P., Ata, B., & Yildiz, B. O. (2018). Metabolic syndrome, hypertension, and hyperlipidemia in mothers, fathers, sisters, and brothers of women with polycystic ovary syndrome: a systematic review and meta-analysis. *Fertility and Sterility*, *109*(2), 356–364.e32. https://doi.org/10.1016/j.fertnstert.2017.10.018

Zaks, N., Batuure, A., Lin, E., Rommel, A.-S., Reichenberg, A., Grice, D., Bergink, V., Fox, N. S., Mahjani, B., & Janecka, M. (2023). Association Between Mental Health and Reproductive System Disorders in Women: A Systematic Review and Meta-analysis. *JAMA Network Open*, *6*(4), e238685. https://doi.org/10.1001/jamanetworkopen.2023.8685

Zeggini, E., & Ioannidis, J. P. A. (2009). Meta-analysis in genome-wide association studies. *Pharmacogenomics*, *10*(2), 191–201. https://doi.org/10.2217/14622416.10.2.191

Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L. M., Juodakis, J., Miller, D. E., Litterman, N., Jiang, P.-P., Russell, L., Hinds, D. A., Hu, Y., Weirauch, M. T., Chen, X., Chavan, A. R., Wagner, G. P., Pavličev, M., Nnamani, M. C., Maziarz, J., … Muglia, L. J. (2017). Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *New England Journal of Medicine*, *377*(12), 1156–1167. https://doi.org/10.1056/nejmoa1612665

Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T. A., Zhao, N., Bolla, M. K., Dunning, A. M., Dennis, J., Wang, Q., Ful, Z. A., Aittomäki, K., Andrulis, I. L., Anton-Culver, H., Arndt, V., Aronson, K. J., … García-Closas, M. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics*, *52*(6), 572–581. https://doi.org/10.1038/s41588-020-0609-2

Zhang, Q., Privé, F., Vilhjálmsson, B., & Speed, D. (2021). Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nature Communications*, *12*(1), 4192. https://doi.org/10.1038/s41467-021-24485-y

Zhang, Yi, Kent, J. W. J., Olivier, M., Ali, O., Cerjak, D., Broeckel, U., Abdou, R. M., Dyer, T. D., Comuzzie, A., Curran, J. E., Carless, M. A., Rainwater, D. L., Göring, H. H. H., Blangero, J., & Kissebah, A. H. (2013). A comprehensive analysis of adiponectin QTLs using SNP association, SNP cis-effects on peripheral blood gene expression and gene expression correlation identified novel metabolic syndrome (MetS) genes with potential role in carcinogenesis and systemic in. *BMC Medical Genomics*, *6*, 14. https://doi.org/10.1186/1755-8794-6-14

Zhang, Yiliang, Lu, Q., Ye, Y., Huang, K., Liu, W., Wu, Y., Zhong, X., Li, B., Yu, Z., Travers, B. G., Werling, D. M., Li, J. J., & Zhao, H. (2021). SUPERGNOVA: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. *Genome Biology*, *22*(1), 262. https://doi.org/10.1186/s13059-021-02478-w

Zhang, Z., Chen, N., Yin, N., Liu, R., He, Y., Li, D., Tong, M., Gao, A., Lu, P., Zhao, Y., Li, H., Zhang, J., Zhang, D., Gu, W., Hong, J., Wang, W., Qi, L., Ning, G., & Wang, J. (2023). The rs1421085 variant within FTO promotes brown fat thermogenesis. *Nature Metabolism*, *5*(8), 1337–1351. https://doi.org/10.1038/s42255-023-00847-2

Zhao, J., & Xia, L. (2022). Association between hypertensive disorders of pregnancy and risk of attention-deficit/hyperactivity disorder in the offspring: a systematic review and meta-analysis. *Hypertension in Pregnancy*, *41*(3–4), 149–158. https://doi.org/10.1080/10641955.2022.2079674

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B. S., Warrington, N. M., Finucane, H. K., Price, A. L., Bulik-Sullivan, B. K., Anttila, V., Paternoster, L., Gaunt, T. R., Evans, D. M., & Neale, B. M. (2017). LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*(2), 272–279. https://doi.org/10.1093/bioinformatics/btw613

Zhou, W., Kanai, M., Wu, K.-H. 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., … Neale, B. M. (2022). Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics*, *2*(10), 100192. https://doi.org/10.1016/j.xgen.2022.100192

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., & Lee, S. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, *50*(9), 1335–1341. https://doi.org/10.1038/s41588-018-0184-y

Zhu, J., Pujol-Gualdo, N., Wittemans, L. B. L., Lindgren, C. M., Laisk, T., Hirschhorn, J. N., & Chan, Y.-M. (2022). Evidence From Men for Ovary-independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome. *The Journal of Clinical Endocrinology and Metabolism*, *107*(4), e1577–e1587. https://doi.org/10.1210/clinem/dgab838

Zhu, T., Cui, J., & Goodarzi, M. O. (2021). Polycystic ovary syndrome and breast cancer subtypes: a Mendelian randomization study. In *American journal of obstetrics and gynecology* (Vol. 225, Issue 1, pp. 99–101). https://doi.org/10.1016/j.ajog.2021.03.020

ACKNOWLEDGEMENTS

This journey began in late August 2020 in the middle of a pandemic, with a one-way ticket and a cocktail of emotions – excitement, curiosity, determination, but also a lot of fear and self-doubt. Landing in Estonia, little did I know then about the journey ahead, that I now look back in pride and a slight touch of nostalgia. Many factors make a PhD journey a successful one, but definitely surrounding yourself with the right people is key, and here I have been fortunate.

Triin Laisk, you are one of the most remarkable and supportive mentors a PhD student could ever dream of! I am deeply grateful for that. From my very first GWAS and Manhattan plot to completing projects and presentations in places I'd never imagined: your active listening and countless scientific insights have been key. Beyond science, I admire your good sense of humor, your ability to see 'the bigger picture', and an optimal dose of pragmatism. Thank you for showing me that actions speak louder than words, for highlighting the importance of self-care, and embracing humor as a source of resilience in both science and life. It takes two (and a fair share of orange-cardamom kringel!) to push through many cool scientific stories, many laughs, and I am sincerely grateful that you've consistently stood by my side and trusted me always through this journey. Beyond being an outstanding mentor, you are a true role model – thank you, Triin!

I also want to express my gratitude to Reedik Mägi, an exceptional person and scientist (or 'bioinformatician' if you prefer it). Thanks for your help and endless patience in sharing your statistical genetics wisdom, helping through this last coding bug before the next one arrives. Talking to you has always been supportive, leaving me with the feeling that no matter what, a solution will be found (well, unless you find yourself on a badminton court!). Thank you for your care, humor, altruism and unconditional help, which shine above any methodological detail. With you on board, science and Estonia are better places to be.

Thanks to my co-supervisors professor Terhi Piltonen and Dr. Riikka Arffman for hosting me in University of Oulu as part of the MATER project and double PhD degree. Thank you for giving me the opportunity to analyze valuable cohort data, AMH measurements and more importantly thanks to both for welcoming me as part of your research family. Terhi, your energy, positivism, and enthusiasm are contagious and a backbone for many! A true example of the Finnish word '*Sisu*'. Riikka, thank you for your hospitality while I was in Finland and your support. I also want to thank all colleagues I met and worked with at the University of Oulu, and all the scientific and training opportunities received by this institution. Additionally, I would like to thank Heli Ruotsalainen and Karita Saravesi for their valuable administrative support.

And from the very north to the south of Europe, we land in lovely Bologna. I would like to thank Dr. Luca Gianaroli for the opportunity to conduct my internship in SISMeR within the MATER framework and extend my gratitude to all the colleagues in the clinic.

Indeed, this PhD journey would not have been possible without the MATER framework. To this end, I want to thank prof. Andres Salumets for putting together the MSCA grant proposal, supporting many mobility and work opportunities. I am also grateful to Kadri Raav and Gerly Põder for the administrative support. My deepest admiration goes to my PhD fellows in MATER: Nidhi, Laura, Yiqun, Lena, Seungbaek, Nina, Yan, Kate, Maria, Spyros, Meruert and Bariş, last ones with whom I shared Tartu as a base, and closely shared pains and glories from the start to the end of PhD, thank you!

To all co-authors of the publications during my PhD studies, I am grateful for your input and help through the projects. Specially I would like to thank Jia Zhu, you are such an awesome researcher and I learnt a lot from our collaboration!

It is a special feeling when brilliant colleagues become friends. Kristi Läll, I'm forever grateful and happy to have crossed paths with you! Our conversations have been always a source of inspiration, I greatly admire you in work and in life. Thanks for your trust and the lovely times in Jüri and Tartu. Kertu, you're an amazing scientist and a pillar of support for many! Thank you for your constant hospitality, countless wine-fueled conversations, best live translations, breakfasts on fire (literally), and many more great memories, with more yet to come.

In the thrilling journey of a PhD, my respect and admiration extend to every single PhD student I have met along the way, especially those students and researchers who have faced challenges in the past and have actively contributed for us (and future generations) to have improved work conditions.

A special shout-out to the loveliest master students whom I had the privilege to work with closely and who are now rockstar PhD students: Gala, Jelisaveta, and Valentina. Also I'd like to give an extended thank you to all the members of Genomic Epidemiology team, especially to Maarja, Jéssica, Fanny and Mariann for sharing many good times in Tartu and abroad.

I'd also like to give an acknowledgment to all the colleagues from the Estonian Genome Centre. Your humility and modesty haven't gone unnoticed! Here's to bigger celebrations! I am grateful to every of you who have shared your time, wisdom and encouragement during seminars, coffee and lunch breaks. A special acknowledgment to Kelli Lehto for being such a chill person and cool colleague. Additionally, I want to thank the writing retreats organized by the Institute of Genomics, a key factor in boosting my productivity writing-wise!

I would also like to thank prof. Kristiina Tambets for reviewing my thesis and suggesting helpful corrections. My gratitude also goes to prof. Andres Metspalu, for welcoming me to the Estonian Genome Centre and providing this awesome resource for doing science.

Thanks as well to all cohorts' participants for their generosity and trust, which made this work possible and allow science to advance.

Last but not least, a special shout-out to two truly inspirational leaders and exceptional women in science, prof. Elin Org and prof. Lili Milani. You exemplify not only scientific excellence but a deep commitment to advocating for students. It has been wonderful to meet you and I greatly admire you!

No puc concloure aquests agraïments sense pensar en els inicis. Marta Ribasés, María Soler i Cristina Sánchez, moltes gràcies per acollir-me a un lab de primer nivell (també humanament) durant el màster i encoratjar-me a prendre aquest camí. Cris, tenim uns croissanets i una peli a guionitzar pendent. And to my very first supervisors in my BSc studies, Ilse Arnoldussen and Prof. Amanda Kiliaan, thanks for being such inspirational scientists, Nijmegen and Dpt. of Anatomy will always be a special place.

To all the lovely bunch of international and local people in Estonia, Finland and Italy that I met during this time: Thanks to all who have made those places feel like home. If 'Ayesh and friends', Elektriteater, Sooty, 'It is raining!', Barlova, Raatuse 22, Delta, Koitelinkoski, Northern lights hunting, talks in Tuba, Khinkali, Domenica di Pasqua, Aperol o'clock, or Siena sound familiar to you... I deeply thank you for the lovely memories and so many learnings and experiences together.

Gràcies també a la meva segona família, Vedeka's, pel vostre suport i amistat. Als meus compis de la cobla, per recolzar-me sempre, i a la música en general per acostar-me a persones fantàstiques i ser una font inesgotable de salut i equilibri. A Solsona i la meva gent per ser casa i brillar amb més força. A les super mares, profes i amigues sense fronteres: gràcies Aroa i Iris. A la meva gent de la uni, gràcies per tot el suport rebut des de l'esquerra de l'Eixample, Nova York, Kuopio o Miami. Un agraïment especial a la meva partner in crime de la uni a arreu Marta Piqué i al Néstor pel seu suport.

I per acabar, voldria donar gràcies infinites a la meva família: als meus pares Lluís i Tate, i als meus germans Marc i Lluís. Tot i que no acabeu d'entendre que faig exactament, gràcies per recolzar-me i ajudar-me quan ho he necessitat. També vull agrair el suport dels meus tiets Carme i Jaume i la meva tieta Teresa. I per acabar, gràcies de tot cor als meus avis Ramona i Llàtzer, i Montserrat i Ramon, estaré sempre en deute amb vosaltres! Moltes gràcies pel vostre amor incondicional i per haver cregut sempre en mi. Us estimo!

PUBLICATIONS

CURRICULUM VITAE

Name: Natàlia Pujol Gualdo
Date of birth: January 9, 1996
Address: University of Tartu, Institute of Genomics, Riia 23b, 51010 Tartu, Estonia
E-mail: npujolgualdo@gmail.com

Education

2020–2024 MSCA PhD fellow – Double PhD degree between Institute of Genomics, University of Tartu, Estonia and University of Oulu, Finland

2019–2020 MSc, Master's Degree in Translational Biomedical Research, Authonomous University of Barcelona (UAB) and Vall d'Hebron Research Institute (VHIR), Barcelona, Catalonia, Spain

2014–2019 BSc, Bachelor in Biomedical Science, Faculty of Biology, University of Barcelona (UB)

Professional employment

2020–2024 Junior Research Fellow of Reproductive Genomics

Teaching and dissemination activities

2022–2023 Participation in the course Genetic and molecular epidemiology, Master in Epidemiology and biomedical data science, University of Oulu, Oulu, Finland

Participation in the platform exhibition and radio interview "Women in Science" "Dones de Ciència" in Saló de les Homilies d'Organyà de la Biblioteca Carles Morató, Solsona, Catalonia. Platform exhibited during 1–30 of November 2021.

Creation and presentation of two workshops due to national initiative '100tifiques' due to International Day of Women and Girls in Science (11th of February 2022): "Women in science: my personal experience and research path" directed to: first and fourth course students from high school in Escola Arrels Secundaria, Solsona, Catalonia and FEDAC Guissona (February 2022 and February 2023)

Creation of workshop directed to high school students: "GenEthics", FEDAC Guissona, Catalonia, 21st of February 2024

Scientific administrative activities

Student representative of the MATER Marie Sklodowska-Curie EU H2020 program (2021)

Oral presentations in conferences

Selected oral presentation at

Androgen Excess – PCOS annual meeting 2023, Rotterdam, The Netherlands, 5th–7th October 2023. Pujol-Gualdo N., …, Triin Laisk*, Piltonen T.* Circulating anti-Müllerian hormone levels in pre-menopausal women: novel genetic insights from a GWAS meta-analysis

European Society of Human Reproduction and Embriology (ESHRE), 28th June 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

International Congress of Human Genetics, 22nd February 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

Institute of Molecular and Cell Biology and Institute of Genomics annual conference, 20th January 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

Arctic PCOS meeting – PCOS-Metabolic community for the Northern Finland Birth Cohort projects (1st March 2022 – 4th March 2022), Levi, Finland; Immelkartano. "Novel insights in PCOS genetic risk: new genetic variants and phenotypic effects in men"

Androgen Excess – PCOS annual meeting 2021 (held virtually) 12th–14th November 2021

Pujol-Gualdo N*, Tyrmi. J., Arffman R.,… Laisk T., Kettunen J., Laivuori H. "Leveraging Northern European history: novel low frequency variants for polycystic ovary syndrome."

Nordic Society of Human Genetics and Precision Medicine (NSHG-PM) conference (held virtually), 9th November 2021. Pujol-Gualdo N., Läll K,…, Piltonen T., Mägi R., Laisk T. Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci."

European Society of Human Reproduction and Embriology (ESHRE), 2021, (Held virtually). Pujol-Gualdo N., Läll K,…, Piltonen T., Mägi R., Laisk T. Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci."

Awards and scholarships

Award to best oral presentation in Nordic Society of Human Genetic and Precision Medicine meeting (NSHG-PM): Pujol-Gualdo N., Läll K,…, Mägi R., Laisk T. Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci." (held virtually), 9th November 2021

Oulu University Foundation Scholarship "Genetic risk factors underlying Anti-Müllerian hormone variation in pre-menopausal women: largest meta-analysis to date" (May 2023)

Paulon Säätiön scholarship for dissertation writing (November 2023)

Supervision

"Investigating the Genetic Basis of Obstetrical Diagnoses Related to Miscarriage and Placental Biology Using Genome-Wide Association Study Meta-Analysis", 2023

Student: Jelisaveta Džigurski. Degree: MSc, Institute of Technology, Faculty of Science and Technology, University of Tartu. Supervisors: Triin Laisk, Reedik Mägi, Natàlia Pujol Gualdo.

Additionally, I have participated in mentoring several other students.

Publications

Pujol-Gualdo N., Läll K., Lepamets M.,Arffman R.,Rossi H., Piltonen T., Mägi R.*, Laisk T.* Advancing our understanding of genetic risk factors and personalized strategies for pelvic organ prolapse (2022) Nature Communications

Tyrmi JK*, Arffman R*, **Pujol-Gualdo N***.,…, Piltonen T., Laisk T., Kettunen J.*, Laivuori H.* Leveraging Northern European population history: novel low frequency variants for polycystic ovary syndrome (2021) Human Reproduction https:// doi.org/10.1093/humrep/deab250

Pujol-Gualdo N., Karjalainen M., Vosa U., Arffman R., Mägi R., Ronkainen J., Triin Laisk*, Piltonen T.* Circulating anti-Müllerian hormone levels in premenopausal women: novel genetic insights from a GWAS meta-analysis (accepted for publication in Human Reproduction, preprint: https://doi.org/10.1101/2023.09.07.23295182)

Zhu J., **Pujol-Gualdo N.,** Wittemans L., Lindgren M. Cecilia, Laisk T., Hirschhorn J.N., Chan Y. Evidence from Men for Ovary-Independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome (2021) JCEM https://doi.org/10.1210/clinem/dgab838

Pujol-Gualdo N; Estonian Biobank Research Team; Mägi R, Laisk T. Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy. Hum Reprod. 2023 Dec 4;38(12):2516–2525. https://doi.org/10.1093/humrep/dead217. PMID: 37877466

Pathare A.*, **Pujol-Gualdo N***, Rukins V., Džigurski J., Peters M., Estonian Biobank Research Team, Mägi R., Salumets A., Saare M., Laisk T. Large-scale genome-wide association study to determine the genetic underpinnings of female genital tract polyps (**preprint:** https://doi.org/10.1101/2024.01.29.24301773)

Koel M, Võsa U, Jõeloo M, Läll K, **Pujol-Gualdo N**,…, Mägi R, Laisk T. GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. Hum Mol Genet. 2023 Jun 5;32(12):2103–2116. https://doi.org/10.1093/hmg/ddad043. PMID: 36929174

Bourgault J, Abner E, Manikpurage HD, **Pujol-Gualdo N**, Laisk T;… ,Arsenault BJ. Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis. Gastroenterology. 2023 May;164(6):953–965.e3. https://doi.org/10.1053/j.gastro.2023.01.028. Epub 2023 Feb 1. PMID: 36736436.

Pujol-Gualdo N, Sánchez-Mora C, Ramos-Quiroga JA, Ribasés M, Soler Artigas M. Integrating genomics and transcriptomics: Towards deciphering ADHD. Eur Neuropsychopharmacol. 2021 Mar;44:1–13. https://doi.org/10.1016/j.euroneuro.2021.01.002. Epub 2021 Jan 23. PMID: 33495110.

Nieuwenhuis D, **Pujol-Gualdo N**, Arnoldussen IAC, Kiliaan AJ. Adipokines: A gear shift in puberty. Obes Rev. 2020 Jun;21(6):e13005. https://doi.org/10.1111/obr.13005. Epub 2020 Jan 30. PMID: 32003144

Arnoldussen IAC, Morrison MC, ..., **Pujol-Gualdo N**, van der Logt L, Gross G, Kleemann R, Kiliaan AJ. Milk fat globule membrane attenuates high fat diet-induced neuropathological changes in obese Ldlr-/-.Leiden mice. Int J Obes (Lond). 2022 Feb;46(2):342–349. https://doi.org/10.1038/s41366-021-00998-w. Epub 2021 Oct 29. PMID: 34716425.

Custers E, Vreeken D, Kaufmann LK, **Pujol-Gualdo N**, …, Kiliaan AJ. Cognitive Control and Weight Loss After Bariatric Surgery: the BARICO Study. Obes Surg. 2023 Sep;33(9):2799–2807. https://doi.org/10.1007/s11695-023-06744-7. Epub 2023 Jul 21. PMID: 37477832.

ELULOOKIRJELDUS

Nimi: Natàlia Pujol Gualdo
Sünniaeg: 09.01.1996
Kontakt: Tartu Ülikool, genoomika instituut, Riia 23v, 51010, Tartu, Eesti
E-post: npujolgualdo@gmail.com

Haridus

2020–2024 MSCA PhD tudeng,Tartu Ülikool, genoomika instituut ja Oulu Ülikool
2019–2020 MSc (*Translational Biomedical Research*), Barcelona Autonoomne Ülikool ja Vall d'Hebron Teadusinstituut
2014–2019 BSc (*Biomedical Science*), Barcelona Ülikool

Teenistuskäik

2020–2024 Tartu Ülikool, genoomika instituut, reproduktiivgenoomika noorem-teadur

Õpetamiskogemus

2022–2023 Osalemine Oulu Ülikooli magistriõppekursuse "Geneetiline ja molekulaarne epidemioloogia" õppetöös
2021 Osalemine "Women in Science"/"Dones de Ciència" näitusel ja raadiointervjuul (Carles Morató raamatukogu, Solsona, Kataloonia)
2022–2023 Rahvusvahelise "Naised ja tüdrukud teaduses"päeva raames riikliku algatuse "100tifiques" raames kahe töötoa loomine ja esitlus: "Naised teaduses: minu isiklik kogemus ja uurimistee", mis on suunatud keskkooliõpilastele (Solsona ja FEDAC Guissona, Kataloonia)
2024 Gümnaasiumiõpilastele suunatud töötoa "GenEthics" loomine (FEDAC Guissona, Kataloonia)

Teadusorganisatsiooniline ja- administratiivne tegevus

MATER Marie Sklodowska-Curie EU H2020 programmi üliõpilaste esindaja (2021)

Rahvusvahelised konverentsid

Suulised ettekanded

Androgen Excess – PCOS annual meeting 2023, Rotterdam, The Netherlands, 5th–7th October 2023. Pujol-Gualdo N., …, Triin Laisk*, Piltonen T.* Circulating anti-Müllerian hormone levels in pre-menopausal women: novel genetic insights from a GWAS meta-analysis

European Society of Human Reproduction and Embriology (ESHRE), 28th June 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

International Congress of Human Genetics, 22nd February 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

Institute of Molecular and Cell Biology and Institute of Genomics annual conference, 20th January 2023. Pujol-Gualdo N., Mägi R., Laisk T. "Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy".

Arctic PCOS meeting – PCOS-Metabolic community for the Northern Finland Birth Cohort projects (1st March 2022 – 4th March 2022), Levi, Finland; Immelkartano. "Novel insights in PCOS genetic risk: new genetic variants and phenotypic effects in men"

Androgen Excess – PCOS annual meeting 2021 (held virtually) 12th–14th November 2021. Pujol-Gualdo N*, Tyrmi. J., Arffman R.,… Laisk T., Kettunen J., Laivuori H. "Leveraging Northern European history: novel low frequency variants for polycystic ovary syndrome."

Nordic Society of Human Genetics and Precision Medicine (NSHG-PM) conference (held virtually), 9th November 2021. Pujol-Gualdo N., Läll K,…, Piltonen T., Mägi R., Laisk T. Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci."

ESHRE2021 (Held virtually). Pujol-Gualdo N., Läll K,…, Piltonen T., Mägi R., Laisk T. "Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci."

Stipendiumid ja auhinnad

Parim suuline ettekanne Nordic Society of Human Genetic and Precision Medicine konverentsil: Pujol-Gualdo N., Läll K,…, Mägi R., Laisk T. Advancing or understanding of genetic risk factors and personalized strategies in pelvic organ prolapse: largest GWAS to date unravels 19 novel genetic loci." (held virtually), 9th November 2021

Oulu Ülikooli Sihtasutuse stipendium "Genetic risk factors underlying Anti-Müllerian hormone variation in pre-menopausal women: largest meta-analysis to date" (May 2023)

Paulon Säätiön stipendium (November 2023)

Juhendatud väitekirjad

"Investigating the Genetic Basis of Obstetrical Diagnoses Related to Miscarriage and Placental Biology Using Genome-Wide Association Study Meta-Analysis", 2023

Juhendatav: Jelisaveta Džigurski, magistritöö. Juhendajad: Triin Laisk; Natalia Pujol Gualdo; Reedik Mägi, Asutused: Tartu Ülikool, Loodus- ja täppisteaduste valdkond, tehnoloogiainstituut

Lisaks olen osalenud teiste tudengite juhendamisel analüüside läbiviimisel.

Teaduspublikatsioonid

Pujol-Gualdo N., Läll K., Lepamets M.,Arffman R.,Rossi H., Piltonen T., Mägi R.*, Laisk T.* Advancing our understanding of genetic risk factors and personalized strategies for pelvic organ prolapse (2022) Nature Communications

Tyrmi JK*, Arffman R*, **Pujol-Gualdo N***.,…, Piltonen T., Laisk T., Kettunen J.*, Laivuori H.* Leveraging Northern European population history: novel low frequency variants for polycystic ovary syndrome (2021) Human Reproduction https://doi.org/10.1093/humrep/deab250

Pujol-Gualdo N., Karjalainen M., Vosa U., Arffman R., Mägi R., Ronkainen J., Triin Laisk*, Piltonen T.* Circulating anti-Müllerian hormone levels in premenopausal women: novel genetic insights from a GWAS meta-analysis (accepted for publication in Human Reproduction, preprint: https://doi.org/10.1101/2023.09.07.23295182)

Zhu J., **Pujol-Gualdo N.,** Wittemans L., Lindgren M. Cecilia, Laisk T., Hirschhorn J.N., Chan Y. Evidence from Men for Ovary-Independent Effects of Genetic Risk Factors for Polycystic Ovary Syndrome (2021) JCEM https://doi.org/10.1210/clinem/dgab838

Pujol-Gualdo N; Estonian Biobank Research Team; Mägi R, Laisk T. Genome-wide association study meta-analysis supports association between MUC1 and ectopic pregnancy. Hum Reprod. 2023 Dec 4;38(12):2516–2525. https://doi.org/10.1093/humrep/dead217. PMID: 37877466)

Pathare A.*, **Pujol-Gualdo N***, Rukins V., Džigurski J., Peters M., Estonian Biobank Research Team, Mägi R., Salumets A., Saare M., Laisk T. Large-scale genome-wide association study to determine the genetic underpinnings of female genital tract polyps (**preprint:** https://doi.org/10.1101/2024.01.29.24301773)

Koel M, Võsa U, Jõeloo M, Läll K, **Pujol-Gualdo N**,…, Mägi R, Laisk T. GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. Hum Mol Genet. 2023 Jun 5;32(12):2103–2116. https://doi.org/10.1093/hmg/ddad043. PMID: 36929174

Bourgault J, Abner E, Manikpurage HD, **Pujol-Gualdo N**, Laisk T;… ,Arsenault BJ. Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis. Gastroenterology. 2023 May;164(6):953–965.e3. https://doi.org/10.1053/j.gastro.2023.01.028. Epub 2023 Feb 1. PMID: 36736436.

Pujol-Gualdo N, Sánchez-Mora C, Ramos-Quiroga JA, Ribasés M, Soler Artigas M. Integrating genomics and transcriptomics: Towards deciphering ADHD. Eur Neuropsychopharmacol. 2021 Mar;44:1–13. https://doi.org/10.1016/ j.euroneuro.2021.01.002. Epub 2021 Jan 23. PMID: 33495110.

Nieuwenhuis D, **Pujol-Gualdo N**, Arnoldussen IAC, Kiliaan AJ. Adipokines: A gear shift in puberty. Obes Rev. 2020 Jun;21(6):e13005. https://doi.org/10.1111/obr.13005. Epub 2020 Jan 30. PMID: 32003144

Arnoldussen IAC, Morrison MC, ..., **Pujol-Gualdo N**, van der Logt L, Gross G, Kleemann R, Kiliaan AJ. Milk fat globule membrane attenuates high fat diet-

induced neuropathological changes in obese Ldlr-/-.Leiden mice. Int J Obes (Lond). 2022 Feb;46(2):342–349. https://doi.org/10.1038/s41366-021-00998-w. Epub 2021 Oct 29. PMID: 34716425.

Custers E, Vreeken D, Kaufmann LK, **Pujol-Gualdo N**, …, Kiliaan AJ. Cognitive Control and Weight Loss After Bariatric Surgery: the BARICO Study. Obes Surg. 2023 Sep;33(9):2799–2807. https://doi.org/10.1007/s11695-023-06744-7. Epub 2023 Jul 21. PMID: 37477832.

DISSERTATIONES BIOLOGICAE UNIVERSITATIS TARTUENSIS

1. **Toivo Maimets**. Studies of human oncoprotein p53. Tartu, 1991, 96 p.
2. **Enn K. Seppet**. Thyroid state control over energy metabolism, ion transport and contractile functions in rat heart. Tartu, 1991, 135 p.
3. **Kristjan Zobel**. Epifüütsete makrosamblike väärtus õhu saastuse indikaatoritena Hamar-Dobani boreaalsetes mägimetsades. Tartu, 1992, 131 lk.
4. **Andres Mäe**. Conjugal mobilization of catabolic plasmids by transposable elements in helper plasmids. Tartu, 1992, 91 p.
5. **Maia Kivisaar**. Studies on phenol degradation genes of *Pseudomonas* sp. strain EST 1001. Tartu, 1992, 61 p.
6. **Allan Nurk**. Nucleotide sequences of phenol degradative genes from *Pseudomonas sp.* strain EST 1001 and their transcriptional activation in *Pseudomonas putida.* Tartu, 1992, 72 p.
7. **Ülo Tamm**. The genus *Populus* L. in Estonia: variation of the species biology and introduction. Tartu, 1993, 91 p.
8. **Jaanus Remme**. Studies on the peptidyltransferase centre of the *E.coli* ribosome. Tartu, 1993, 68 p.
9. **Ülo Langel**. Galanin and galanin antagonists. Tartu, 1993, 97 p.
10. **Arvo Käärd**. The development of an automatic online dynamic fluorescense-based pH-dependent fiber optic penicillin flowthrought biosensor for the control of the benzylpenicillin hydrolysis. Tartu, 1993, 117 p.
11. **Lilian Järvekülg**. Antigenic analysis and development of sensitive immunoassay for potato viruses. Tartu, 1993, 147 p.
12. **Jaak Palumets**. Analysis of phytomass partition in Norway spruce. Tartu, 1993, 47 p.
13. **Arne Sellin**. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different enviromental conditions. Tartu, 1994, 119 p.
13. **Mati Reeben**. Regulation of light neurofilament gene expression. Tartu, 1994, 108 p.
14. **Urmas Tartes**. Respiration rhytms in insects. Tartu, 1995, 109 p.
15. **Ülo Puurand**. The complete nucleotide sequence and infections *in vitro* transcripts from cloned cDNA of a potato A potyvirus. Tartu, 1995, 96 p.
16. **Peeter Hõrak**. Pathways of selection in avian reproduction: a functional framework and its application in the population study of the great tit (*Parus major*). Tartu, 1995, 118 p.
17. **Erkki Truve**. Studies on specific and broad spectrum virus resistance in transgenic plants. Tartu, 1996, 158 p.
18. **Illar Pata**. Cloning and characterization of human and mouse ribosomal protein S6-encoding genes. Tartu, 1996, 60 p.
19. **Ülo Niinemets**. Importance of structural features of leaves and canopy in determining species shade-tolerance in temperature deciduous woody taxa. Tartu, 1996, 150 p.

20. **Ants Kurg**. Bovine leukemia virus: molecular studies on the packaging region and DNA diagnostics in cattle. Tartu, 1996, 104 p.
21. **Ene Ustav**. E2 as the modulator of the BPV1 DNA replication. Tartu, 1996, 100 p.
22. **Aksel Soosaar**. Role of helix-loop-helix and nuclear hormone receptor transcription factors in neurogenesis. Tartu, 1996, 109 p.
23. **Maido Remm**. Human papillomavirus type 18: replication, transformation and gene expression. Tartu, 1997, 117 p.
24. **Tiiu Kull**. Population dynamics in *Cypripedium calceolus* L. Tartu, 1997, 124 p.
25. **Kalle Olli**. Evolutionary life-strategies of autotrophic planktonic microorganisms in the Baltic Sea. Tartu, 1997, 180 p.
26. **Meelis Pärtel**. Species diversity and community dynamics in calcareous grassland communities in Western Estonia. Tartu, 1997, 124 p.
27. **Malle Leht**. The Genus *Potentilla* L. in Estonia, Latvia and Lithuania: distribution, morphology and taxonomy. Tartu, 1997, 186 p.
28. **Tanel Tenson**. Ribosomes, peptides and antibiotic resistance. Tartu, 1997, 80 p.
29. **Arvo Tuvikene**. Assessment of inland water pollution using biomarker responses in fish *in vivo* and *in vitro.* Tartu, 1997, 160 p.
30. **Urmas Saarma**. Tuning ribosomal elongation cycle by mutagenesis of 23S rRNA. Tartu, 1997, 134 p.
31. **Henn Ojaveer**. Composition and dynamics of fish stocks in the gulf of Riga ecosystem. Tartu, 1997, 138 p.
32. **Lembi Lõugas**. Post-glacial development of vertebrate fauna in Estonian water bodies. Tartu, 1997, 138 p.
33. **Margus Pooga**. Cell penetrating peptide, transportan, and its predecessors, galanin-based chimeric peptides. Tartu, 1998, 110 p.
34. **Andres Saag**. Evolutionary relationships in some cetrarioid genera (Lichenized Ascomycota). Tartu, 1998, 196 p.
35. **Aivar Liiv**. Ribosomal large subunit assembly *in vivo*. Tartu, 1998, 158 p.
36. **Tatjana Oja**. Isoenzyme diversity and phylogenetic affinities among the eurasian annual bromes (*Bromus* L., Poaceae). Tartu, 1998, 92 p.
37. **Mari Moora**. The influence of arbuscular mycorrhizal (AM) symbiosis on the competition and coexistence of calcareous grassland plant species. Tartu, 1998, 78 p.
38. **Olavi Kurina**. Fungus gnats in Estonia (*Diptera: Bolitophilidae, Keroplatidae, Macroceridae, Ditomyiidae, Diadocidiidae, Mycetophilidae*). Tartu, 1998, 200 p.
39. **Andrus Tasa**. Biological leaching of shales: black shale and oil shale. Tartu, 1998, 98 p.
40. **Arnold Kristjuhan**. Studies on transcriptional activator properties of tumor suppressor protein p53. Tartu, 1998, 86 p.
41. **Sulev Ingerpuu**. Characterization of some human myeloid cell surface and nuclear differentiation antigens. Tartu, 1998, 163 p.

42. **Veljo Kisand**. Responses of planktonic bacteria to the abiotic and biotic factors in the shallow lake Võrtsjärv. Tartu, 1998, 118 p.
43. **Kadri Põldmaa**. Studies in the systematics of hypomyces and allied genera (Hypocreales, Ascomycota). Tartu, 1998, 178 p.
44. **Markus Vetemaa**. Reproduction parameters of fish as indicators in environmental monitoring. Tartu, 1998, 117 p.
45. **Heli Talvik**. Prepatent periods and species composition of different *Oesophagostomum* spp. populations in Estonia and Denmark. Tartu, 1998, 104 p.
46. **Katrin Heinsoo**. Cuticular and stomatal antechamber conductance to water vapour diffusion in *Picea abies* (L.) karst. Tartu, 1999, 133 p.
47. **Tarmo Annilo**. Studies on mammalian ribosomal protein S7. Tartu, 1998, 77 p.
48. **Indrek Ots**. Health state indicies of reproducing great tits (*Parus major*): sources of variation and connections with life-history traits. Tartu, 1999, 117 p.
49. **Juan Jose Cantero**. Plant community diversity and habitat relationships in central Argentina grasslands. Tartu, 1999, 161 p.
50. **Rein Kalamees**. Seed bank, seed rain and community regeneration in Estonian calcareous grasslands. Tartu, 1999, 107 p.
51. **Sulev Kõks**. Cholecystokinin (CCK) – induced anxiety in rats: influence of environmental stimuli and involvement of endopioid mechanisms and serotonin. Tartu, 1999, 123 p.
52. **Ebe Sild**. Impact of increasing concentrations of O_3 and CO_2 on wheat, clover and pasture. Tartu, 1999, 123 p.
53. **Ljudmilla Timofejeva**. Electron microscopical analysis of the synaptonemal complex formation in cereals. Tartu, 1999, 99 p.
54. **Andres Valkna**. Interactions of galanin receptor with ligands and G-proteins: studies with synthetic peptides. Tartu, 1999, 103 p.
55. **Taavi Virro**. Life cycles of planktonic rotifers in lake Peipsi. Tartu, 1999, 101 p.
56. **Ana Rebane**. Mammalian ribosomal protein S3a genes and intron-encoded small nucleolar RNAs U73 and U82. Tartu, 1999, 85 p.
57. **Tiina Tamm**. Cocksfoot mottle virus: the genome organisation and translational strategies. Tartu, 2000, 101 p.
58. **Reet Kurg**. Structure-function relationship of the bovine papilloma virus E2 protein. Tartu, 2000, 89 p.
59. **Toomas Kivisild**. The origins of Southern and Western Eurasian populations: an mtDNA study. Tartu, 2000, 121 p.
60. **Niilo Kaldalu**. Studies of the TOL plasmid transcription factor XylS. Tartu, 2000, 88 p.
61. **Dina Lepik**. Modulation of viral DNA replication by tumor suppressor protein p53. Tartu, 2000, 106 p.
62. **Kai Vellak**. Influence of different factors on the diversity of the bryophyte vegetation in forest and wooded meadow communities. Tartu, 2000, 122 p.

63. **Jonne Kotta**. Impact of eutrophication and biological invasionas on the structure and functions of benthic macrofauna. Tartu, 2000, 160 p.
64. **Georg Martin**. Phytobenthic communities of the Gulf of Riga and the inner sea the West-Estonian archipelago. Tartu, 2000, 139 p.
65. **Silvia Sepp**. Morphological and genetical variation of *Alchemilla L.* in Estonia. Tartu, 2000. 124 p.
66. **Jaan Liira**. On the determinants of structure and diversity in herbaceous plant communities. Tartu, 2000, 96 p.
67. **Priit Zingel**. The role of planktonic ciliates in lake ecosystems. Tartu, 2001, 111 p.
68. **Tiit Teder**. Direct and indirect effects in Host-parasitoid interactions: ecological and evolutionary consequences. Tartu, 2001, 122 p.
69. **Hannes Kollist**. Leaf apoplastic ascorbate as ozone scavenger and its transport across the plasma membrane. Tartu, 2001, 80 p.
70. **Reet Marits**. Role of two-component regulator system PehR-PehS and extracellular protease PrtW in virulence of *Erwinia Carotovora* subsp. *Carotovora*. Tartu, 2001, 112 p.
71. **Vallo Tilgar**. Effect of calcium supplementation on reproductive performance of the pied flycatcher *Ficedula hypoleuca* and the great tit *Parus major,* breeding in Nothern temperate forests. Tartu, 2002, 126 p.
72. **Rita Hõrak**. Regulation of transposition of transposon Tn*4652* in *Pseudomonas putida*. Tartu, 2002, 108 p.
73. **Liina Eek-Piirsoo**. The effect of fertilization, mowing and additional illumination on the structure of a species-rich grassland community. Tartu, 2002, 74 p.
74. **Krõõt Aasamaa**. Shoot hydraulic conductance and stomatal conductance of six temperate deciduous tree species. Tartu, 2002, 110 p.
75. **Nele Ingerpuu**. Bryophyte diversity and vascular plants. Tartu, 2002, 112 p.
76. **Neeme Tõnisson**. Mutation detection by primer extension on oligonucleotide microarrays. Tartu, 2002, 124 p.
77. **Margus Pensa**. Variation in needle retention of Scots pine in relation to leaf morphology, nitrogen conservation and tree age. Tartu, 2003, 110 p.
78. **Asko Lõhmus**. Habitat preferences and quality for birds of prey: from principles to applications. Tartu, 2003, 168 p.
79. **Viljar Jaks**. p53 – a switch in cellular circuit. Tartu, 2003, 160 p.
80. **Jaana Männik**. Characterization and genetic studies of four ATP-binding cassette (ABC) transporters. Tartu, 2003, 140 p.
81. **Marek Sammul**. Competition and coexistence of clonal plants in relation to productivity. Tartu, 2003, 159 p
82. **Ivar Ilves**. Virus-cell interactions in the replication cycle of bovine papillomavirus type 1. Tartu, 2003, 89 p.
83. **Andres Männik**. Design and characterization of a novel vector system based on the stable replicator of bovine papillomavirus type 1. Tartu, 2003, 109 p.

84. **Ivika Ostonen**. Fine root structure, dynamics and proportion in net primary production of Norway spruce forest ecosystem in relation to site conditions. Tartu, 2003, 158 p.
85. **Gudrun Veldre**. Somatic status of 12–15-year-old Tartu schoolchildren. Tartu, 2003, 199 p.
86. **Ülo Väli.** The greater spotted eagle *Aquila clanga* and the lesser spotted eagle *A. pomarina*: taxonomy, phylogeography and ecology. Tartu, 2004, 159 p.
87. **Aare Abroi**. The determinants for the native activities of the bovine papillomavirus type 1 E2 protein are separable. Tartu, 2004, 135 p.
88. **Tiina Kahre**. Cystic fibrosis in Estonia. Tartu, 2004, 116 p.
89. **Helen Orav-Kotta**. Habitat choice and feeding activity of benthic suspension feeders and mesograzers in the northern Baltic Sea. Tartu, 2004, 117 p.
90. **Maarja Öpik**. Diversity of arbuscular mycorrhizal fungi in the roots of perennial plants and their effect on plant performance. Tartu, 2004, 175 p.
91. **Kadri Tali**. Species structure of *Neotinea ustulata*. Tartu, 2004, 109 p.
92. **Kristiina Tambets**. Towards the understanding of post-glacial spread of human mitochondrial DNA haplogroups in Europe and beyond: a phylogeographic approach. Tartu, 2004, 163 p.
93. **Arvi Jõers**. Regulation of p53-dependent transcription. Tartu, 2004, 103 p.
94. **Lilian Kadaja**. Studies on modulation of the activity of tumor suppressor protein p53. Tartu, 2004, 103 p.
95. **Jaak Truu**. Oil shale industry wastewater: impact on river microbial community and possibilities for bioremediation. Tartu, 2004, 128 p.
96. **Maire Peters**. Natural horizontal transfer of the *pheBA* operon. Tartu, 2004, 105 p.
97. **Ülo Maiväli**. Studies on the structure-function relationship of the bacterial ribosome. Tartu, 2004, 130 p.
98. **Merit Otsus**. Plant community regeneration and species diversity in dry calcareous grasslands. Tartu, 2004, 103 p.
99. **Mikk Heidemaa**. Systematic studies on sawflies of the genera *Dolerus, Empria,* and *Caliroa* (Hymenoptera: Tenthredinidae). Tartu, 2004, 167 p.
100. **Ilmar Tõnno**. The impact of nitrogen and phosphorus concentration and N/P ratio on cyanobacterial dominance and N_2 fixation in some Estonian lakes. Tartu, 2004, 111 p.
101. **Lauri Saks**. Immune function, parasites, and carotenoid-based ornaments in greenfinches. Tartu, 2004, 144 p.
102. **Siiri Rootsi**. Human Y-chromosomal variation in European populations. Tartu, 2004, 142 p.
103. **Eve Vedler**. Structure of the 2,4-dichloro-phenoxyacetic acid-degradative plasmid pEST4011. Tartu, 2005. 106 p.
104. **Andres Tover**. Regulation of transcription of the phenol degradation *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 126 p.
105. **Helen Udras**. Hexose kinases and glucose transport in the yeast *Hansenula polymorpha*. Tartu, 2005, 100 p.

106. **Ave Suija**. Lichens and lichenicolous fungi in Estonia: diversity, distribution patterns, taxonomy. Tartu, 2005, 162 p.
107. **Piret Lõhmus**. Forest lichens and their substrata in Estonia. Tartu, 2005, 162 p.
108. **Inga Lips**. Abiotic factors controlling the cyanobacterial bloom occurrence in the Gulf of Finland. Tartu, 2005, 156 p.
109. **Krista Kaasik**. Circadian clock genes in mammalian clockwork, metabolism and behaviour. Tartu, 2005, 121 p.
110. **Juhan Javoiš**. The effects of experience on host acceptance in ovipositing moths. Tartu, 2005, 112 p.
111. **Tiina Sedman**. Characterization of the yeast *Saccharomyces cerevisiae* mitochondrial DNA helicase Hmi1. Tartu, 2005, 103 p.
112. **Ruth Aguraiuja**. Hawaiian endemic fern lineage *Diellia* (Aspleniaceae): distribution, population structure and ecology. Tartu, 2005, 112 p.
113. **Riho Teras**. Regulation of transcription from the fusion promoters generated by transposition of Tn*4652* into the upstream region of *pheBA* operon in *Pseudomonas putida*. Tartu, 2005, 106 p.
114. **Mait Metspalu**. Through the course of prehistory in India: tracing the mtDNA trail. Tartu, 2005, 138 p.
115. **Elin Lõhmussaar**. The comparative patterns of linkage disequilibrium in European populations and its implication for genetic association studies. Tartu, 2006, 124 p.
116. **Priit Kupper**. Hydraulic and environmental limitations to leaf water relations in trees with respect to canopy position. Tartu, 2006, 126 p.
117. **Heili Ilves**. Stress-induced transposition of Tn*4652* in *Pseudomonas Putida.* Tartu, 2006, 120 p.
118. **Silja Kuusk**. Biochemical properties of Hmi1p, a DNA helicase from *Saccharomyces cerevisiae* mitochondria. Tartu, 2006, 126 p.
119. **Kersti Püssa**. Forest edges on medium resolution landsat thematic mapper satellite images. Tartu, 2006, 90 p.
120. **Lea Tummeleht**. Physiological condition and immune function in great tits (*Parus major* l.): Sources of variation and trade-offs in relation to growth. Tartu, 2006, 94 p.
121. **Toomas Esperk**. Larval instar as a key element of insect growth schedules. Tartu, 2006, 186 p.
122. **Harri Valdmann**. Lynx (*Lynx lynx*) and wolf (*Canis lupus*) in the Baltic region: Diets, helminth parasites and genetic variation. Tartu, 2006. 102 p.
123. **Priit Jõers**. Studies of the mitochondrial helicase Hmi1p in *Candida albicans* and *Saccharomyces cerevisia*. Tartu, 2006. 113 p.
124. **Kersti Lilleväli**. Gata3 and Gata2 in inner ear development. Tartu, 2007, 123 p.
125. **Kai Rünk**. Comparative ecology of three fern species: *Dryopteris carthusiana* (Vill.) H.P. Fuchs, *D. expansa* (C. Presl) Fraser-Jenkins & Jermy and *D. dilatata* (Hoffm.) A. Gray (Dryopteridaceae). Tartu, 2007, 143 p.

126. **Aveliina Helm**. Formation and persistence of dry grassland diversity: role of human history and landscape structure. Tartu, 2007, 89 p.
127. **Leho Tedersoo**. Ectomycorrhizal fungi: diversity and community structure in Estonia, Seychelles and Australia. Tartu, 2007, 233 p.
128. **Marko Mägi**. The habitat-related variation of reproductive performance of great tits in a deciduous-coniferous forest mosaic: looking for causes and consequences. Tartu, 2007, 135 p.
129. **Valeria Lulla**. Replication strategies and applications of Semliki Forest virus. Tartu, 2007, 109 p.
130. **Ülle Reier**. Estonian threatened vascular plant species: causes of rarity and conservation. Tartu, 2007, 79 p.
131. **Inga Jüriado**. Diversity of lichen species in Estonia: influence of regional and local factors. Tartu, 2007, 171 p.
132. **Tatjana Krama**. Mobbing behaviour in birds: costs and reciprocity based cooperation. Tartu, 2007, 112 p.
133. **Signe Saumaa**. The role of DNA mismatch repair and oxidative DNA damage defense systems in avoidance of stationary phase mutations in *Pseudomonas putida.* Tartu, 2007, 172 p.
134. **Reedik Mägi**. The linkage disequilibrium and the selection of genetic markers for association studies in european populations. Tartu, 2007, 96 p.
135. **Priit Kilgas**. Blood parameters as indicators of physiological condition and skeletal development in great tits (*Parus major*): natural variation and application in the reproductive ecology of birds. Tartu, 2007, 129 p.
136. **Anu Albert**. The role of water salinity in structuring eastern Baltic coastal fish communities. Tartu, 2007, 95 p.
137. **Kärt Padari**. Protein transduction mechanisms of transportans. Tartu, 2008, 128 p.
138. **Siiri-Lii Sandre**. Selective forces on larval colouration in a moth. Tartu, 2008, 125 p.
139. **Ülle Jõgar**. Conservation and restoration of semi-natural floodplain meadows and their rare plant species. Tartu, 2008, 99 p.
140. **Lauri Laanisto**. Macroecological approach in vegetation science: generality of ecological relationships at the global scale. Tartu, 2008, 133 p.
141. **Reidar Andreson**. Methods and software for predicting PCR failure rate in large genomes. Tartu, 2008, 105 p.
142. **Birgot Paavel**. Bio-optical properties of turbid lakes. Tartu, 2008, 175 p.
143. **Kaire Torn**. Distribution and ecology of charophytes in the Baltic Sea. Tartu, 2008, 98 p.
144. **Vladimir Vimberg**. Peptide mediated macrolide resistance. Tartu, 2008, 190 p.
145. **Daima Örd**. Studies on the stress-inducible pseudokinase TRB3, a novel inhibitor of transcription factor ATF4. Tartu, 2008, 108 p.
146. **Lauri Saag**. Taxonomic and ecologic problems in the genus *Lepraria* (*Stereocaulaceae*, lichenised *Ascomycota*). Tartu, 2008, 175 p.

147. **Ulvi Karu**. Antioxidant protection, carotenoids and coccidians in greenfinches – assessment of the costs of immune activation and mechanisms of parasite resistance in a passerine with carotenoid-based ornaments. Tartu, 2008, 124 p.
148. **Jaanus Remm**. Tree-cavities in forests: density, characteristics and occupancy by animals. Tartu, 2008, 128 p.
149. **Epp Moks**. Tapeworm parasites *Echinococcus multilocularis* and *E. granulosus* in Estonia: phylogenetic relationships and occurrence in wild carnivores and ungulates. Tartu, 2008, 82 p.
150. **Eve Eensalu**. Acclimation of stomatal structure and function in tree canopy: effect of light and CO_2 concentration. Tartu, 2008, 108 p.
151. **Janne Pullat**. Design, functionlization and application of an *in situ* synthesized oligonucleotide microarray. Tartu, 2008, 108 p.
152. **Marta Putrinš**. Responses of *Pseudomonas putida* to phenol-induced metabolic and stress signals. Tartu, 2008, 142 p.
153. **Marina Semtšenko**. Plant root behaviour: responses to neighbours and physical obstructions. Tartu, 2008, 106 p.
154. **Marge Starast**. Influence of cultivation techniques on productivity and fruit quality of some *Vaccinium* and *Rubus* taxa. Tartu, 2008, 154 p.
155. **Age Tats**. Sequence motifs influencing the efficiency of translation. Tartu, 2009, 104 p.
156. **Radi Tegova**. The role of specialized DNA polymerases in mutagenesis in *Pseudomonas putida.* Tartu, 2009, 124 p.
157. **Tsipe Aavik**. Plant species richness, composition and functional trait pattern in agricultural landscapes – the role of land use intensity and landscape structure. Tartu, 2009, 112 p.
158. **Kaja Kiiver**. Semliki forest virus based vectors and cell lines for studying the replication and interactions of alphaviruses and hepaciviruses. Tartu, 2009, 104 p.
159. **Meelis Kadaja**. Papillomavirus Replication Machinery Induces Genomic Instability in its Host Cell. Tartu, 2009, 126 p.
160. **Pille Hallast**. Human and chimpanzee Luteinizing hormone/Chorionic Gonadotropin beta (*LHB/CGB*) gene clusters: diversity and divergence of young duplicated genes. Tartu, 2009, 168 p.
161. **Ain Vellak**. Spatial and temporal aspects of plant species conservation. Tartu, 2009, 86 p.
162. **Triinu Remmel**. Body size evolution in insects with different colouration strategies: the role of predation risk. Tartu, 2009, 168 p.
163. **Jaana Salujõe**. Zooplankton as the indicator of ecological quality and fish predation in lake ecosystems. Tartu, 2009, 129 p.
164. **Ele Vahtmäe**. Mapping benthic habitat with remote sensing in optically complex coastal environments. Tartu, 2009, 109 p.
165. **Liisa Metsamaa**. Model-based assessment to improve the use of remote sensing in recognition and quantitative mapping of cyanobacteria. Tartu, 2009, 114 p.

166. **Pille Säälik**. The role of endocytosis in the protein transduction by cell-penetrating peptides. Tartu, 2009, 155 p.
167. **Lauri Peil**. Ribosome assembly factors in *Escherichia coli.* Tartu, 2009, 147 p.
168. **Lea Hallik**. Generality and specificity in light harvesting, carbon gain capacity and shade tolerance among plant functional groups. Tartu, 2009, 99 p.
169. **Mariliis Tark**. Mutagenic potential of DNA damage repair and tolerance mechanisms under starvation stress. Tartu, 2009, 191 p.
170. **Riinu Rannap**. Impacts of habitat loss and restoration on amphibian populations. Tartu, 2009, 117 p.
171. **Maarja Adojaan**. Molecular variation of HIV-1 and the use of this knowledge in vaccine development. Tartu, 2009, 95 p.
172. **Signe Altmäe**. Genomics and transcriptomics of human induced ovarian folliculogenesis. Tartu, 2010, 179 p.
173. **Triin Suvi**. Mycorrhizal fungi of native and introduced trees in the Seychelles Islands. Tartu, 2010, 107 p.
174. **Velda Lauringson**. Role of suspension feeding in a brackish-water coastal sea. Tartu, 2010, 123 p.
175. **Eero Talts**. Photosynthetic cyclic electron transport – measurement and variably proton-coupled mechanism. Tartu, 2010, 121 p.
176. **Mari Nelis**. Genetic structure of the Estonian population and genetic distance from other populations of European descent. Tartu, 2010, 97 p.
177. **Kaarel Krjutškov**. Arrayed Primer Extension-2 as a multiplex PCR-based method for nucleic acid variation analysis: method and applications. Tartu, 2010, 129 p.
178. **Egle Köster**. Morphological and genetical variation within species complexes: *Anthyllis vulneraria* s. l. and *Alchemilla vulgaris* (coll.). Tartu, 2010, 101 p.
179. **Erki Õunap**. Systematic studies on the subfamily Sterrhinae (Lepidoptera: Geometridae). Tartu, 2010, 111 p.
180. **Merike Jõesaar**. Diversity of key catabolic genes at degradation of phenol and *p*-cresol in pseudomonads. Tartu, 2010, 125 p.
181. **Kristjan Herkül**. Effects of physical disturbance and habitat-modifying species on sediment properties and benthic communities in the northern Baltic Sea. Tartu, 2010, 123 p.
182. **Arto Pulk**. Studies on bacterial ribosomes by chemical modification approaches. Tartu, 2010, 161 p.
183. **Maria Põllupüü**. Ecological relations of cladocerans in a brackish-water ecosystem. Tartu, 2010, 126 p.
184. **Toomas Silla**. Study of the segregation mechanism of the Bovine Papillomavirus Type 1. Tartu, 2010, 188 p.
185. **Gyaneshwer Chaubey**. The demographic history of India: A perspective based on genetic evidence. Tartu, 2010, 184 p.

186. **Katrin Kepp**. Genes involved in cardiovascular traits: detection of genetic variation in Estonian and Czech populations. Tartu, 2010, 164 p.
187. **Virve Sõber**. The role of biotic interactions in plant reproductive performance. Tartu, 2010, 92 p.
188. **Kersti Kangro**. The response of phytoplankton community to the changes in nutrient loading. Tartu, 2010, 144 p.
189. **Joachim M. Gerhold**. Replication and Recombination of mitochondrial DNA in Yeast. Tartu, 2010, 120 p.
190. **Helen Tammert**. Ecological role of physiological and phylogenetic diversity in aquatic bacterial communities. Tartu, 2010, 140 p.
191. **Elle Rajandu**. Factors determining plant and lichen species diversity and composition in Estonian *Calamagrostis* and *Hepatica* site type forests. Tartu, 2010, 123 p.
192. **Paula Ann Kivistik**. ColR-ColS signalling system and transposition of Tn*4652* in the adaptation of *Pseudomonas putida.* Tartu, 2010, 118 p.
193. **Siim Sõber**. Blood pressure genetics: from candidate genes to genome-wide association studies. Tartu, 2011, 120 p.
194. **Kalle Kipper**. Studies on the role of helix 69 of 23S rRNA in the factor-dependent stages of translation initiation, elongation, and termination. Tartu, 2011, 178 p.
195. **Triinu Siibak**. Effect of antibiotics on ribosome assembly is indirect. Tartu, 2011, 134 p.
196. **Tambet Tõnissoo**. Identification and molecular analysis of the role of guanine nucleotide exchange factor RIC-8 in mouse development and neural function. Tartu, 2011, 110 p.
197. **Helin Räägel**. Multiple faces of cell-penetrating peptides – their intracellular trafficking, stability and endosomal escape during protein transduction. Tartu, 2011, 161 p.
198. **Andres Jaanus**. Phytoplankton in Estonian coastal waters – variability, trends and response to environmental pressures. Tartu, 2011, 157 p.
199. **Tiit Nikopensius**. Genetic predisposition to nonsyndromic orofacial clefts. Tartu, 2011, 152 p.
200. **Signe Värv**. Studies on the mechanisms of RNA polymerase II-dependent transcription elongation. Tartu, 2011, 108 p.
201. **Kristjan Välk**. Gene expression profiling and genome-wide association studies of non-small cell lung cancer. Tartu, 2011, 98 p.
202. **Arno Põllumäe**. Spatio-temporal patterns of native and invasive zooplankton species under changing climate and eutrophication conditions. Tartu, 2011, 153 p.
203. **Egle Tammeleht**. Brown bear (*Ursus arctos*) population structure, demographic processes and variations in diet in northern Eurasia. Tartu, 2011, 143 p.
205. **Teele Jairus**. Species composition and host preference among ectomycorrhizal fungi in Australian and African ecosystems. Tartu, 2011, 106 p.

206. **Kessy Abarenkov**. PlutoF – cloud database and computing services supporting biological research. Tartu, 2011, 125 p.
207. **Marina Grigorova**. Fine-scale genetic variation of follicle-stimulating hormone beta-subunit coding gene (*FSHB*) and its association with reproductive health. Tartu, 2011, 184 p.
208. **Anu Tiitsaar**. The effects of predation risk and habitat history on butterfly communities. Tartu, 2011, 97 p.
209. **Elin Sild**. Oxidative defences in immunoecological context: validation and application of assays for nitric oxide production and oxidative burst in a wild passerine. Tartu, 2011, 105 p.
210. **Irja Saar**. The taxonomy and phylogeny of the genera *Cystoderma* and *Cystodermella* (Agaricales, Fungi). Tartu, 2012, 167 p.
211. **Pauli Saag**. Natural variation in plumage bacterial assemblages in two wild breeding passerines. Tartu, 2012, 113 p.
212. **Aleksei Lulla**. Alphaviral nonstructural protease and its polyprotein substrate: arrangements for the perfect marriage. Tartu, 2012, 143 p.
213. **Mari Järve**. Different genetic perspectives on human history in Europe and the Caucasus: the stories told by uniparental and autosomal markers. Tartu, 2012, 119 p.
214. **Ott Scheler**. The application of tmRNA as a marker molecule in bacterial diagnostics using microarray and biosensor technology. Tartu, 2012, 93 p.
215. **Anna Balikova**. Studies on the functions of tumor-associated mucin-like leukosialin (CD43) in human cancer cells. Tartu, 2012, 129 p.
216. **Triinu Kõressaar**. Improvement of PCR primer design for detection of prokaryotic species. Tartu, 2012, 83 p.
217. **Tuul Sepp**. Hematological health state indices of greenfinches: sources of individual variation and responses to immune system manipulation. Tartu, 2012, 117 p.
218. **Rya Ero**. Modifier view of the bacterial ribosome. Tartu, 2012, 146 p.
219. **Mohammad Bahram**. Biogeography of ectomycorrhizal fungi across different spatial scales. Tartu, 2012, 165 p.
220. **Annely Lorents**. Overcoming the plasma membrane barrier: uptake of amphipathic cell-penetrating peptides induces influx of calcium ions and downstream responses. Tartu, 2012, 113 p.
221. **Katrin Männik**. Exploring the genomics of cognitive impairment: whole-genome SNP genotyping experience in Estonian patients and general population. Tartu, 2012, 171 p.
222. **Marko Prous**. Taxonomy and phylogeny of the sawfly genus *Empria* (Hymenoptera, Tenthredinidae). Tartu, 2012, 192 p.
223. **Triinu Visnapuu**. Levansucrases encoded in the genome of *Pseudomonas syringae* pv. tomato DC3000: heterologous expression, biochemical characterization, mutational analysis and spectrum of polymerization products. Tartu, 2012, 160 p.
224. **Nele Tamberg**. Studies on Semliki Forest virus replication and pathogenesis. Tartu, 2012, 109 p.

225. **Tõnu Esko**. Novel applications of SNP array data in the analysis of the genetic structure of Europeans and in genetic association studies. Tartu, 2012, 149 p.
226. **Timo Arula**. Ecology of early life-history stages of herring *Clupea harengus membras* in the northeastern Baltic Sea. Tartu, 2012, 143 p.
227. **Inga Hiiesalu**. Belowground plant diversity and coexistence patterns in grassland ecosystems. Tartu, 2012, 130 p.
228. **Kadri Koorem**. The influence of abiotic and biotic factors on small-scale plant community patterns and regeneration in boreonemoral forest. Tartu, 2012, 114 p.
229. **Liis Andresen**. Regulation of virulence in plant-pathogenic pectobacteria. Tartu, 2012, 122 p.
230. **Kaupo Kohv**. The direct and indirect effects of management on boreal forest structure and field layer vegetation. Tartu, 2012, 124 p.
231. **Mart Jüssi**. Living on an edge: landlocked seals in changing climate. Tartu, 2012, 114 p.
232. **Riina Klais**. Phytoplankton trends in the Baltic Sea. Tartu, 2012, 136 p.
233. **Rauno Veeroja**. Effects of winter weather, population density and timing of reproduction on life-history traits and population dynamics of moose (*Alces alces*) in Estonia. Tartu, 2012, 92 p.
234. **Marju Keis**. Brown bear (*Ursus arctos*) phylogeography in northern Eurasia. Tartu, 2013, 142 p.
235. **Sergei Põlme**. Biogeography and ecology of *alnus*- associated ectomycorrhizal fungi – from regional to global scale. Tartu, 2013, 90 p.
236. **Liis Uusküla**. Placental gene expression in normal and complicated pregnancy. Tartu, 2013, 173 p.
237. **Marko Lõoke**. Studies on DNA replication initiation in *Saccharomyces cerevisiae.* Tartu, 2013, 112 p.
238. **Anne Aan**. Light- and nitrogen-use and biomass allocation along productivity gradients in multilayer plant communities. Tartu, 2013, 127 p.
239. **Heidi Tamm**. Comprehending phylogenetic diversity – case studies in three groups of ascomycetes. Tartu, 2013, 136 p.
240. **Liina Kangur**. High-Pressure Spectroscopy Study of Chromophore-Binding Hydrogen Bonds in Light-Harvesting Complexes of Photosynthetic Bacteria. Tartu, 2013, 150 p.
241. **Margus Leppik**. Substrate specificity of the multisite specific pseudouridine synthase RluD. Tartu, 2013, 111 p.
242. **Lauris Kaplinski**. The application of oligonucleotide hybridization model for PCR and microarray optimization. Tartu, 2013, 103 p.
243. **Merli Pärnoja**. Patterns of macrophyte distribution and productivity in coastal ecosystems: effect of abiotic and biotic forcing. Tartu, 2013, 155 p.
244. **Tõnu Margus**. Distribution and phylogeny of the bacterial translational GTPases and the Mqsr/YgiT regulatory system. Tartu, 2013, 126 p.
245. **Pille Mänd**. Light use capacity and carbon and nitrogen budget of plants: remote assessment and physiological determinants. Tartu, 2013, 128 p.

246. **Mario Plaas**. Animal model of Wolfram Syndrome in mice: behavioural, biochemical and psychopharmacological characterization. Tartu, 2013, 144 p.
247. **Georgi Hudjašov**. Maps of mitochondrial DNA, Y-chromosome and tyrosinase variation in Eurasian and Oceanian populations. Tartu, 2013, 115 p.
248. **Mari Lepik**. Plasticity to light in herbaceous plants and its importance for community structure and diversity. Tartu, 2013, 102 p.
249. **Ede Leppik**. Diversity of lichens in semi-natural habitats of Estonia. Tartu, 2013, 151 p.
250. **Ülle Saks**. Arbuscular mycorrhizal fungal diversity patterns in boreonemoral forest ecosystems. Tartu, 2013, 151 p.
251. **Eneli Oitmaa**. Development of arrayed primer extension microarray assays for molecular diagnostic applications. Tartu, 2013, 147 p.
252. **Jekaterina Jutkina**. The horizontal gene pool for aromatics degradation: bacterial catabolic plasmids of the Baltic Sea aquatic system. Tartu, 2013, 121 p.
253. **Helen Vellau**. Reaction norms for size and age at maturity in insects: rules and exceptions. Tartu, 2014, 132 p.
254. **Randel Kreitsberg**. Using biomarkers in assessment of environmental contamination in fish – new perspectives. Tartu, 2014, 107 p.
255. **Krista Takkis**. Changes in plant species richness and population performance in response to habitat loss and fragmentation.Tartu, 2014, 141 p.
256. **Liina Nagirnaja**. Global and fine-scale genetic determinants of recurrent pregnancy loss. Tartu, 2014, 211 p.
257. **Triin Triisberg**. Factors influencing the re-vegetation of abandoned extracted peatlands in Estonia. Tartu, 2014, 133 p.
258. **Villu Soon**. A phylogenetic revision of the *Chrysis ignita* species group (Hymenoptera: Chrysididae) with emphasis on the northern European fauna. Tartu, 2014, 211 p.
259. **Andrei Nikonov**. RNA-Dependent RNA Polymerase Activity as a Basis for the Detection of Positive-Strand RNA Viruses by Vertebrate Host Cells. Tartu, 2014, 207 p.
260. **Eele Õunapuu-Pikas**. Spatio-temporal variability of leaf hydraulic conductance in woody plants: ecophysiological consequences. Tartu, 2014, 135 p.
261. **Marju Männiste**. Physiological ecology of greenfinches: information content of feathers in relation to immune function and behavior. Tartu, 2014, 121 p.
262. **Katre Kets**. Effects of elevated concentrations of CO_2 and O_3 on leaf photosynthetic parameters in *Populus tremuloides*: diurnal, seasonal and interannual patterns. Tartu, 2014, 115 p.
263. **Külli Lokko**. Seasonal and spatial variability of zoopsammon communities in relation to environmental parameters. Tartu, 2014, 129 p.
264. **Olga Žilina**. Chromosomal microarray analysis as diagnostic tool: Estonian experience. Tartu, 2014, 152 p.

265. **Kertu Lõhmus**. Colonisation ecology of forest-dwelling vascular plants and the conservation value of rural manor parks. Tartu, 2014, 111 p.
266. **Anu Aun**. Mitochondria as integral modulators of cellular signaling. Tartu, 2014, 167 p.
267. **Chandana Basu Mallick**. Genetics of adaptive traits and gender-specific demographic processes in South Asian populations. Tartu, 2014, 160 p.
268. **Riin Tamme**. The relationship between small-scale environmental heterogeneity and plant species diversity. Tartu, 2014, 130 p.
269. **Liina Remm**. Impacts of forest drainage on biodiversity and habitat quality: implications for sustainable management and conservation. Tartu, 2015, 126 p.
270. **Tiina Talve**. Genetic diversity and taxonomy within the genus *Rhinanthus.* Tartu, 2015, 106 p.
271. **Mehis Rohtla**. Otolith sclerochronological studies on migrations, spawning habitat preferences and age of freshwater fishes inhabiting the Baltic Sea. Tartu, 2015, 137 p.
272. **Alexey Reshchikov**. The world fauna of the genus *Lathrolestes* (Hymenoptera, Ichneumonidae). Tartu, 2015, 247 p.
273. **Martin Pook**. Studies on artificial and extracellular matrix protein-rich surfaces as regulators of cell growth and differentiation. Tartu, 2015, 142 p.
274. **Mai Kukumägi**. Factors affecting soil respiration and its components in silver birch and Norway spruce stands. Tartu, 2015, 155 p.
275. **Helen Karu**. Development of ecosystems under human activity in the North-East Estonian industrial region: forests on post-mining sites and bogs. Tartu, 2015, 152 p.
276. **Hedi Peterson**. Exploiting high-throughput data for establishing relationships between genes. Tartu, 2015, 186 p.
277. **Priit Adler**. Analysis and visualisation of large scale microarray data, Tartu, 2015, 126 p.
278. **Aigar Niglas**. Effects of environmental factors on gas exchange in deciduous trees: focus on photosynthetic water-use efficiency. Tartu, 2015, 152 p.
279. **Silja Laht**. Classification and identification of conopeptides using profile hidden Markov models and position-specific scoring matrices. Tartu, 2015, 100 p.
280. **Martin Kesler**. Biological characteristics and restoration of Atlantic salmon *Salmo salar* populations in the Rivers of Northern Estonia. Tartu, 2015, 97 p.
281. **Pratyush Kumar Das**. Biochemical perspective on alphaviral nonstructural protein 2: a tale from multiple domains to enzymatic profiling. Tartu, 2015, 205 p
282. **Priit Palta**. Computational methods for DNA copy number detection. Tartu, 2015, 130 p.
283. **Julia Sidorenko**. Combating DNA damage and maintenance of genome integrity in pseudomonads. Tartu, 2015, 174 p.

284. **Anastasiia Kovtun-Kante**. Charophytes of Estonian inland and coastal waters: distribution and environmental preferences. Tartu, 2015, 97 p.
285. **Ly Lindman**. The ecology of protected butterfly species in Estonia. Tartu, 2015, 171 p.
286. **Jaanis Lodjak**. Association of Insulin-like Growth Factor I and Corticosterone with Nestling Growth and Fledging Success in Wild Passerines. Tartu, 2016, 113 p.
287. **Ann Kraut**. Conservation of Wood-Inhabiting Biodiversity – Semi-Natural Forests as an Opportunity. Tartu, 2016, 141 p.
288. **Tiit Örd**. Functions and regulation of the mammalian pseudokinase TRIB3. Tartu, 2016, 182. p.
289. **Kairi Käiro**. Biological Quality According to Macroinvertebrates in Streams of Estonia (Baltic Ecoregion of Europe): Effects of Human-induced Hydromorphological Changes. Tartu, 2016, 126 p.
290. **Leidi Laurimaa**. *Echinococcus multilocularis* and other zoonotic parasites in Estonian canids. Tartu, 2016, 144 p.
291. **Helerin Margus**. Characterization of cell-penetrating peptide/nucleic acid nanocomplexes and their cell-entry mechanisms. Tartu, 2016, 173 p.
292. **Kadri Runnel**. Fungal targets and tools for forest conservation. Tartu, 2016, 157 p.
293. **Urmo Võsa**. MicroRNAs in disease and health: aberrant regulation in lung cancer and association with genomic variation. Tartu, 2016, 163 p.
294. **Kristina Mäemets-Allas**. Studies on cell growth promoting AKT signaling pathway – a promising anti-cancer drug target. Tartu, 2016, 146 p.
295. **Janeli Viil**. Studies on cellular and molecular mechanisms that drive normal and regenerative processes in the liver and pathological processes in Dupuytren's contracture. Tartu, 2016, 175 p.
296. **Ene Kook**. Genetic diversity and evolution of *Pulmonaria angustifolia* L. and *Myosotis laxa sensu lato* (Boraginaceae). Tartu, 2016, 106 p.
297. **Kadri Peil**. RNA polymerase II-dependent transcription elongation in *Saccharomyces cerevisiae*. Tartu, 2016, 113 p.
298. **Katrin Ruisu**. The role of RIC8A in mouse development and its function in cell-matrix adhesion and actin cytoskeletal organisation. Tartu, 2016, 129 p.
299. **Janely Pae**. Translocation of cell-penetrating peptides across biological membranes and interactions with plasma membrane constituents. Tartu, 2016, 126 p.
300. **Argo Ronk**. Plant diversity patterns across Europe: observed and dark diversity. Tartu, 2016, 153 p.
301. **Kristiina Mark**. Diversification and species delimitation of lichenized fungi in selected groups of the family Parmeliaceae (Ascomycota). Tartu, 2016, 181 p.
302. **Jaak-Albert Metsoja**. Vegetation dynamics in floodplain meadows: influence of mowing and sediment application. Tartu, 2016, 140 p.

303. **Hedvig Tamman**. The GraTA toxin-antitoxin system of *Pseudomonas putida*: regulation and role in stress tolerance. Tartu, 2016, 154 p.
304. **Kadri Pärtel**. Application of ultrastructural and molecular data in the taxonomy of helotialean fungi. Tartu, 2016, 183 p.
305. **Maris Hindrikson**. Grey wolf (*Canis lupus*) populations in Estonia and Europe: genetic diversity, population structure and -processes, and hybridization between wolves and dogs. Tartu, 2016, 121 p.
306. **Polina Degtjarenko**. Impacts of alkaline dust pollution on biodiversity of plants and lichens: from communities to genetic diversity. Tartu, 2016, 126 p.
307. **Liina Pajusalu**. The effect of CO_2 enrichment on net photosynthesis of macrophytes in a brackish water environment. Tartu, 2016, 126 p.
308. **Stoyan Tankov**. Random walks in the stringent response. Tartu, 2016, 94 p.
309. **Liis Leitsalu**. Communicating genomic research results to population-based biobank participants. Tartu, 2016, 158 p.
310. **Richard Meitern**. Redox physiology of wild birds: validation and application of techniques for detecting oxidative stress. Tartu, 2016, 134 p.
311. **Kaie Lokk**. Comparative genome-wide DNA methylation studies of healthy human tissues and non-small cell lung cancer tissue. Tartu, 2016, 127 p.
312. **Mihhail Kurašin**. Processivity of cellulases and chitinases. Tartu, 2017, 132 p.
313. **Carmen Tali**. Scavenger receptors as a target for nucleic acid delivery with peptide vectors. Tartu, 2017, 155 p.
314. **Katarina Oganjan**. Distribution, feeding and habitat of benthic suspension feeders in a shallow coastal sea. Tartu, 2017, 132 p.
315. **Taavi Paal**. Immigration limitation of forest plants into wooded landscape corridors. Tartu, 2017, 145 p.
316. **Kadri Õunap**. The Williams-Beuren syndrome chromosome region protein WBSCR22 is a ribosome biogenesis factor. Tartu, 2017, 135 p.
317. **Riin Tamm**. In-depth analysis of factors affecting variability in thiopurine methyltransferase activity. Tartu, 2017, 170 p.
318. **Keiu Kask**. The role of RIC8A in the development and regulation of mouse nervous system. Tartu, 2017, 184 p.
319. **Tiia Möller**. Mapping and modelling of the spatial distribution of benthic macrovegetation in the NE Baltic Sea with a special focus on the eelgrass *Zostera marina* Linnaeus, 1753. Tartu, 2017, 162 p.
320. **Silva Kasela**. Genetic regulation of gene expression: detection of tissue- and cell type-specific effects. Tartu, 2017, 150 p.
321. **Karmen Süld**. Food habits, parasites and space use of the raccoon dog *Nyctereutes procyonoides*: the role of an alien species as a predator and vector of zoonotic diseases in Estonia. Tartu, 2017, p.
322. **Ragne Oja**. Consequences of supplementary feeding of wild boar – concern for ground-nesting birds and endoparasite infection. Tartu, 2017, 141 p.
323. **Riin Kont**. The acquisition of cellulose chain by a processive cellobiohydrolase. Tartu, 2017, 117 p.

324. **Liis Kasari**. Plant diversity of semi-natural grasslands: drivers, current status and conservation challenges. Tartu, 2017, 141 p.
325. **Sirgi Saar**. Belowground interactions: the roles of plant genetic relatedness, root exudation and soil legacies. Tartu, 2017, 113 p.
326. **Sten Anslan**. Molecular identification of Collembola and their fungal associates. Tartu, 2017, 125 p.
327. **Imre Taal**. Causes of variation in littoral fish communities of the Eastern Baltic Sea: from community structure to individual life histories. Tartu, 2017, 118 p.
328. **Jürgen Jalak**. Dissecting the Mechanism of Enzymatic Degradation of Cellulose Using Low Molecular Weight Model Substrates. Tartu, 2017, 137 p.
329. **Kairi Kiik**. Reproduction and behaviour of the endangered European mink (*Mustela lutreola*) in captivity. Tartu, 2018, 112 p.
330. **Ivan Kuprijanov**. Habitat use and trophic interactions of native and invasive predatory macroinvertebrates in the northern Baltic Sea. Tartu, 2018, 117 p.
331. **Hendrik Meister**. Evolutionary ecology of insect growth: from geographic patterns to biochemical trade-offs. Tartu, 2018, 147 p.
332. **Ilja Gaidutšik**. Irc3 is a mitochondrial branch migration enzyme in *Saccharomyces cerevisiae.* Tartu, 2018, 161 p.
333. **Lena Neuenkamp**. The dynamics of plant and arbuscular mycorrhizal fungal communities in grasslands under changing land use. Tartu, 2018, 241 p.
334. **Laura Kasak**. Genome structural variation modulating the placenta and pregnancy maintenance. Tartu, 2018, 181 p.
335. **Kersti Riibak**. Importance of dispersal limitation in determining dark diversity of plants across spatial scales. Tartu, 2018, 133 p.
336. **Liina Saar**. Dynamics of grassland plant diversity in changing landscapes. Tartu, 2018, 206 p.
337. **Hanna Ainelo**. Fis regulates *Pseudomonas putida* biofilm formation by controlling the expression of *lapA*. Tartu, 2018, 143 p.
338. **Natalia Pervjakova**. Genomic imprinting in complex traits. Tartu, 2018, 176 p.
339. **Andrio Lahesaare**. The role of global regulator Fis in regulating the expression of *lapF* and the hydrophobicity of soil bacterium *Pseudomonas putida*. Tartu, 2018, 124 p.
340. **Märt Roosaare**. *K*-mer based methods for the identification of bacteria and plasmids. Tartu, 2018, 117 p.
341. **Maria Abakumova**. The relationship between competitive behaviour and the frequency and identity of neighbours in temperate grassland plants. Tartu, 2018, 104 p.
342. **Margus Vilbas**. Biotic interactions affecting habitat use of myrmecophilous butterflies in Northern Europe. Tartu, 2018, 142 p.

343. **Liina Kinkar**. Global patterns of genetic diversity and phylogeography of *Echinococcus granulosus* sensu stricto – a tapeworm species of significant public health concern. Tartu, 2018, 147 p.
344. **Teivi Laurimäe**. Taxonomy and genetic diversity of zoonotic tapeworms in the species complex of *Echinococcus granulosus* sensu lato. Tartu, 2018, 143 p.
345. **Tatjana Jatsenko**. Role of translesion DNA polymerases in mutagenesis and DNA damage tolerance in Pseudomonads. Tartu, 2018, 216 p.
346. **Katrin Viigand**. Utilization of α-glucosidic sugars by *Ogataea* (*Hansenula*) *polymorpha.* Tartu, 2018, 148 p.
347. **Andres Ainelo**. Physiological effects of the *Pseudomonas putida* toxin grat. Tartu, 2018, 146 p.
348. **Killu Timm**. Effects of two genes (DRD4 and SERT) on great tit (*Parus major*) behaviour and reproductive traits. Tartu, 2018, 117 p.
349. **Petr Kohout**. Ecology of ericoid mycorrhizal fungi. Tartu, 2018, 184 p.
350. **Gristin Rohula-Okunev**. Effects of endogenous and environmental factors on night-time water flux in deciduous woody tree species. Tartu, 2018, 184 p.
351. **Jane Oja**. Temporal and spatial patterns of orchid mycorrhizal fungi in forest and grassland ecosystems. Tartu, 2018, 102 p.
352. **Janek Urvik**. Multidimensionality of aging in a long-lived seabird. Tartu, 2018, 135 p.
353. **Lisanna Schmidt**. Phenotypic and genetic differentiation in the hybridizing species pair *Carex flava* and *C. viridula* in geographically different regions. Tartu, 2018, 133 p.
354. **Monika Karmin**. Perspectives from human Y chromosome – phylogeny, population dynamics and founder events. Tartu, 2018, 168 p.
355. **Maris Alver**. Value of genomics for atherosclerotic cardiovascular disease risk prediction. Tartu, 2019, 148 p.
356. **Lehti Saag**. The prehistory of Estonia from a genetic perspective: new insights from ancient DNA. Tartu, 2019, 171 p.
357. **Mari-Liis Viljur**. Local and landscape effects on butterfly assemblages in managed forests. Tartu, 2019, 115 p.
358. **Ivan Kisly**. The pleiotropic functions of ribosomal proteins eL19 and eL24 in the budding yeast ribosome. Tartu, 2019, 170 p.
359. **Mikk Puustusmaa**. On the origin of papillomavirus proteins. Tartu, 2019, 152 p.
360. **Anneliis Peterson**. Benthic biodiversity in the north-eastern Baltic Sea: mapping methods, spatial patterns, and relations to environmental gradients. Tartu, 2019, 159 p.
361. **Erwan Pennarun**. Meandering along the mtDNA phylogeny; causerie and digression about what it can tell us about human migrations. Tartu, 2019, 162 p.

362. **Karin Ernits**. Levansucrase Lsc3 and endo-levanase BT1760: characterization and application for the synthesis of novel prebiotics. Tartu, 2019, 217 p.
363. **Sille Holm**. Comparative ecology of geometrid moths: in search of contrasts between a temperate and a tropical forest. Tartu, 2019, 135 p.
364. **Anne-Mai Ilumäe**. Genetic history of the Uralic-speaking peoples as seen through the paternal haplogroup N and autosomal variation of northern Eurasians. Tartu, 2019, 172 p.
365. **Anu Lepik**. Plant competitive behaviour: relationships with functional traits and soil processes. Tartu, 2019, 152 p.
366. **Kunter Tätte**. Towards an integrated view of escape decisions in birds under variable levels of predation risk. Tartu, 2020, 172 p.
367. **Kaarin Parts**. The impact of climate change on fine roots and root-associated microbial communities in birch and spruce forests. Tartu, 2020, 143 p.
368. **Viktorija Kukuškina**. Understanding the mechanisms of endometrial receptivity through integration of 'omics' data layers. Tartu, 2020, 169 p.
369. **Martti Vasar**. Developing a bioinformatics pipeline gDAT to analyse arbuscular mycorrhizal fungal communities using sequence data from different marker regions. Tartu, 2020, 193 p.
370. **Ott Kangur**. Nocturnal water relations and predawn water potential disequilibrium in temperate deciduous tree species. Tartu, 2020, 126 p.
371. **Helen Post**. Overview of the phylogeny and phylogeography of the Y-chromosomal haplogroup N in northern Eurasia and case studies of two linguistically exceptional populations of Europe – Hungarians and Kalmyks. Tartu, 2020, 143 p.
372. **Kristi Krebs**. Exploring the genetics of adverse events in pharmacotherapy using Biobanks and Electronic Health Records. Tartu, 2020, 151 p.
373. **Kärt Ukkivi**. Mutagenic effect of transcription and transcription-coupled repair factors in *Pseudomonas putida.* Tartu, 2020, 154 p.
374. **Elin Soomets**. Focal species in wetland restoration. Tartu, 2020, 137 p.
375. **Kadi Tilk**. Signals and responses of ColRS two-component system in *Pseudomonas putida.* Tartu, 2020, 133 p.
376. **Indrek Teino**. Studies on aryl hydrocarbon receptor in the mouse granulosa cell model. Tartu, 2020, 139 p.
377. **Maarja Vaikre**. The impact of forest drainage on macroinvertebrates and amphibians in small waterbodies and opportunities for cost-effective mitigation. Tartu, 2020, 132 p.
378. **Siim-Kaarel Sepp**. Soil eukaryotic community responses to land use and host identity. Tartu, 2020, 222 p.
379. **Eveli Otsing**. Tree species effects on fungal richness and community structure. Tartu, 2020, 152 p.
380. **Mari Pent**. Bacterial communities associated with fungal fruitbodies. Tartu, 2020, 144 p.

381. **Einar Kärgenberg**. Movement patterns of lithophilous migratory fish in free-flowing and fragmented rivers. Tartu, 2020, 167 p.
382. **Antti Matvere**. The studies on aryl hydrocarbon receptor in murine granulosa cells and human embryonic stem cells. Tartu, 2021, 163 p.
383. **Jhonny Capichoni Massante**. Phylogenetic structure of plant communities along environmental gradients: a macroecological and evolutionary approach. Tartu, 2021, 144 p.
384. **Ajai Kumar Pathak**. Delineating genetic ancestries of people of the Indus Valley, Parsis, Indian Jews and Tharu tribe. Tartu, 2021, 197 p.
385. **Tanel Vahter**. Arbuscular mycorrhizal fungal biodiversity for sustainable agroecosystems. Tartu, 2021, 191 p.
386. **Burak Yelmen**. Characterization of ancient Eurasian influences within modern human genomes. Tartu, 2021, 134 p.
387. **Linda Ongaro**. A genomic portrait of American populations. Tartu, 2021, 182 p.
388. **Kairi Raime**. The identification of plant DNA in metagenomic samples. Tartu, 2021, 108 p.
389. **Heli Einberg**. Non-linear and non-stationary relationships in the pelagic ecosystem of the Gulf of Riga (Baltic Sea). Tartu, 2021, 119 p.
390. **Mickaël Mathieu Pihain**. The evolutionary effect of phylogenetic neighbourhoods of trees on their resistance to herbivores and climatic stress. Tartu, 2022, 145 p.
391. **Annika Joy Meitern**. Impact of potassium ion content of xylem sap and of light conditions on the hydraulic properties of trees. Tartu, 2022, 132 p.
392. **Elise Joonas**. Evaluation of metal contaminant hazard on microalgae with environmentally relevant testing strategies. Tartu, 2022, 118 p.
393. **Kreete Lüll**. Investigating the relationships between human microbiome, host factors and female health. Tartu, 2022, 141 p.
394. **Triin Kaasiku**. A wader perspective to Boreal Baltic coastal grasslands: from habitat availability to breeding site selection and nest survival. Tartu, 2022, 141 p.
395. **Meeli Alber**. Impact of elevated atmospheric humidity on the structure of the water transport pathway in deciduous trees. Tartu, 2022, 170 p.
396. **Ludovica Molinaro**. Ancestry deconvolution of Estonian, European and Worldwide genomic layers: a human population genomics excavation. Tartu, 2022, 138 p.
397. **Tina Saupe**. The genetic history of the Mediterranean before the common era: a focus on the Italian Peninsula. Tartu, 2022, 165 p.
398. **Mari-Ann Lind**. Internal constraints on energy processing and their consequences: an integrative study of behaviour, ornaments and digestive health in greenfinches. Tartu, 2022, 137 p.
399. **Markus Valge**. Testing the predictions of life history theory on anthropometric data. Tartu, 2022, 171 p.
400. **Ants Tull**. Domesticated and wild mammals as reservoirs for zoonotic helminth parasites in Estonia. Tartu, 2022, 152 p.

401. **Saleh Rahimlouye Barabi**. Investigation of diazotrophic bacteria association with plants. Tartu, 2022, 137 p.
402. **Farzad Aslani**. Towards revealing the biogeography of belowground diversity. Tartu, 2022, 124 p.
403. **Nele Taba**. Diet, blood metabolites, and health. Tartu, 2022, 163 p.
404. **Katri Pärna**. Improving the personalized prediction of complex traits and diseases: application to type 2 diabetes. Tartu, 2022, 190 p.
405. **Silva Lilleorg**. Bacterial ribosome heterogeneity on the example of bL31 paralogs in *Escherichia coli.* Tartu, 2022, 189 p.
406. **Oliver Aasmets**. The importance of microbiome in human health. Tartu, 2022, 123 p.
407. **Henel Jürgens**. Exploring post-translational modifications of histones in RNA polymerase II-dependent transcription. Tartu, 2022, 147 p.
408. **Mari Tagel**. Finding novel factors affecting the mutation frequency: a case study of tRNA modification enzymes TruA and RluA. Tartu, 2022, 176 p.
409. **Marili Sell**. The impact of environmental change on ecophysiology of hemiboreal tree species – acclimation mechanisms in belowground. Tartu, 2022, 163 p.
410. **Kaarin Hein**. The hissing behaviour of Great Tit (*Parus major*) females reflects behavioural phenotype and breeding success in a wild population. Tartu, 2022, 96 p.
411. **Maret Gerz**. The distribution and role of mycorrhizal symbiosis in plant communities. Tartu, 2022, 206 p.
412. **Kristiina Nõomaa**. Role of invasive species in brackish benthic community structure and biomass changes. Tartu, 2023, 151 p.
413. **Anton Savchenko**. Taxonomic studies in Dacrymycetes: *Cerinomyces* and allied taxa. Tartu, 2023, 181 p.
414. **Ahto Agan**. Interactions between invasive pathogens and resident mycobiome in the foliage of trees. Tartu, 2023, 155 p.
415. **Diego Pires Ferraz Trindade**. Dark diversity dynamics linked to global change: taxonomic and functional perspective. Tartu, 2023, 134 p.
416. **Madli Jõks**. Biodiversity drivers in oceanic archipelagos and habitat fragments, explored by agent-based simulation models. Tartu, 2023, 116 p.
417. **Ciara Baines**. Adaptation to oncogenic pollution and natural cancer defences in the aquatic environment. Tartu, 2023, 164 p.
418. **Rain Inno**. Placental transcriptome and miRNome in normal and complicated pregnancies. Tartu, 2023, 145 p.
419. **Daniyal Gohar**. Diversity, genomics, and potential functions of fungus-inhabiting bacteria. Tartu, 2023, 138 p.
420. **Sirli Rosendahl**. Fitness effects of chromosomal toxin-antitoxin systems in *Pseudomonas putida.* Tartu, 2023, 154 p.
421. **Mathilde Frédérique E. André**. New Guinea, a hotspot for Human evolution: settlement history and adaptation in northern Sahul. Tartu, 2023, 202 p.

422. **Vlad-Julian Piljukov**. Biochemical characterization of Irc3 helicase. Tartu, 2023, 137 p.
423. **Gerli Albert**. Carbon use strategies of macrophyte communities in the northeastern Baltic Sea: implications for a high CO_2 environment. Tartu, 2023, 128 p.
424. **Mariann Koel**. The molecular interactions between trophoblast and endometrial cells in embryo implantation. Tartu, 2023, 171 p.
425. **Robin Gielen**. Diversity and ecological role of pathogenic fungi in insect populations. Tartu, 2023, 139 p.
426. **Kaspar Reier**. Quantity, stability and disparity of ribosomal components in *Escherichia coli* stationary phase. Tartu, 2023, 151 p.
427. **Linda Rusalepp**. The impact of environmental drivers and competition on phenolic metabolite profiles in hybrid aspen and silver birch. Tartu, 2023, 153 p.
428. **Eliisa Pass**. The effect of managed forest-wetland landscapes on forest grouse and nest predation. Tartu, 2023, 115 p.
429. **Sanni Färkkilä**. Methods for studying plant-fungal interactions – reflecting on the old, the new and the upcoming. Tartu, 2024, 147 p.
430. **Maarja Jõeloo.** Advances in microarray-based copy number variation discovery and phenotypic associations. Tartu, 2024, 209 p.