TITLE: GENOME-WIDE ANALYSIS OF 64 MALE REPRODUCTIVE PHENOTYPES REVEALS NOVEL LOCI AND SHARED BIOLOGICAL PATHWAYS

Jéssica Figuerêdo[1], Toomas Haller[1], Estonian Biobank Research Team[1], Health Informatics Research Team[2,3], Margus Punab[4,5], Triin Laisk[1]*, and Reedik Mägi[1]*

[1]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia.
[2]Institute of Computer Science, University of Tartu, Tartu, Estonia
[3]STACC, Tartu, Estonia
[4]Institute of Biomedicine and Translational Medicine, University of Tartu, Ravila tn 19, Tartu, Estonia
[5]Andrology Clinic, Tartu University Hospital, Ludvig Puusepa 8, Tartu, Estonia

*****Triin Laisk and Reedik Mägi share the senior authorship

Estonian Biobank research team collaborators: Andres Metspalu, Lili Milani, Tõnu Esko, Mait Metspalu, Mari Nelis, and Georgi Hudjashov. (EstBBresearch@ut.ee)
Health Informatics research team collaborators: Raivo Kolde, Sven Laur, Sulev Reisberg, and Jaak Vilo. (health-informatics@ut.ee)

ABSTRACT

Male reproductive disorders affect millions of men worldwide, yet their genetic basis remains poorly understood. Conditions such as infertility, erectile dysfunction, prostate disorders and hormonal abnormalities are common, interconnected and often diagnosed late, underscoring the need for broader biological insight. Using genetic data from over 530,000 men across the Estonian Biobank, FinnGen, and UK Biobank, we analysed 64 male reproductive phenotypes. We identified 143 lead variants associated with these conditions, including 47 not previously linked to male reproductive traits, and detected the first genome-wide signals for 3 understudied disorders. Twelve lead variants showed at least 1.5-fold enrichment in the Finnish and Estonian populations, indicating population-specific effects and utility of isolated cohorts for variant discovery. We also highlighted 328 candidate genes and mapped shared biological pathways across prostate, testicular, penile and hormonal traits. These findings expand knowledge of male reproductive biology and offer entry points for improved diagnosis and risk prediction.

INTRODUCTION

Male reproductive health is a key component of overall well-being, yet it remains underrepresented in genomic research on complex traits. This gap largely reflects challenges in defining and capturing many male reproductive phenotypes at population scale,

rather than a lack of clinical relevance (1). Conditions such as infertility, sexual dysfunction, prostate diseases, and hormonal imbalances, are common and often interconnected, affecting men across lifespan and contributing to substantial long-term health, psychological, and quality-of-life burden (2). Despite growing awareness of their clinical and societal importance, the genetic architecture of many male-specific reproductive conditions remains poorly understood.

The expansion of large population-based biobanks, including the Estonian Biobank (EstBB), Finnish biobank research project (FinnGen), and UK Biobank (UKBB) have become essential for mapping the genetic basis of complex traits. Their scale, combined with longitudinal follow-up and linkage to electronic health records (EHRs), enables well-powered genome-wide association studies (GWAS) across a broad range of phenotypes (3–5). The use of standardized coding systems, such as the International Classification of Diseases (ICD) for diagnoses and Logical Observation Identifiers Names and Codes (LOINC) for laboratory measurements, further supports consistent and reproducible phenotype definitions across cohorts (6,7). Together, these structures make it possible to harmonize analyses across studies and to perform meta-analyses that improve robustness and generalizability. Recent work using these resources has uncovered shared and population-specific variants for a wide range of traits, including metabolic, cardiovascular, and reproductive conditions (8–11). In particular, genetically isolated populations such as those of Finland and Estonia, offer increased power to detect risk variants that may be undetected in more heterogeneous cohorts (4).

Despite these advances, prior GWAS of male reproductive health traits have concentrated on a limited number of traits. Testosterone levels, erectile dysfunction, and prostate cancer account for much of the existing literature, with multiple genome-wide associations reported for these phenotypes (9,11,12). By contrast, many genitourinary and endocrine conditions affecting male reproductive function have received less attention in large-scale genetic studies. As a result, sex-specific disease mechanisms and pleiotropic effects across reproductive phenotypes remain poorly resolved, reflecting limited trait coverage and the lack of well-powered, male-focused datasets (1).

To address these limitations, here we present the largest and most comprehensive GWAS and meta-analysis of male reproductive phenotypes to date, leveraging data from EstBB, FinnGen, and UKBB (Figure 1). Across all analyses, the combined meta-analysis sample size reached up to 594,127 individuals, depending on phenotype availability (Supplementary table 2). 64 male reproductive phenotypes were examined, representing the full set of traits available across cohorts that met predefined inclusion and power criteria. These phenotypes encompass major domains of male reproductive health, including neoplastic, endocrine, structural, inflammatory, and functional traits, and were defined using curated ICD-10 and LOINC codes. By integrating variant-level association, colocalization, gene prioritisation, pathway enrichment, and genetic correlation analyses, we identified 143 genome-wide significant variants, including 47 novel trait-specific associations, and uncovered the first genome-wide signals for previously unexplored conditions namely abnormal spermatozoa, phimosis, and testicular dysfunction, while prioritising 328 candidate causal genes that highlight shared biological mechanisms across prostate, testicular, and penile traits.

METHODS

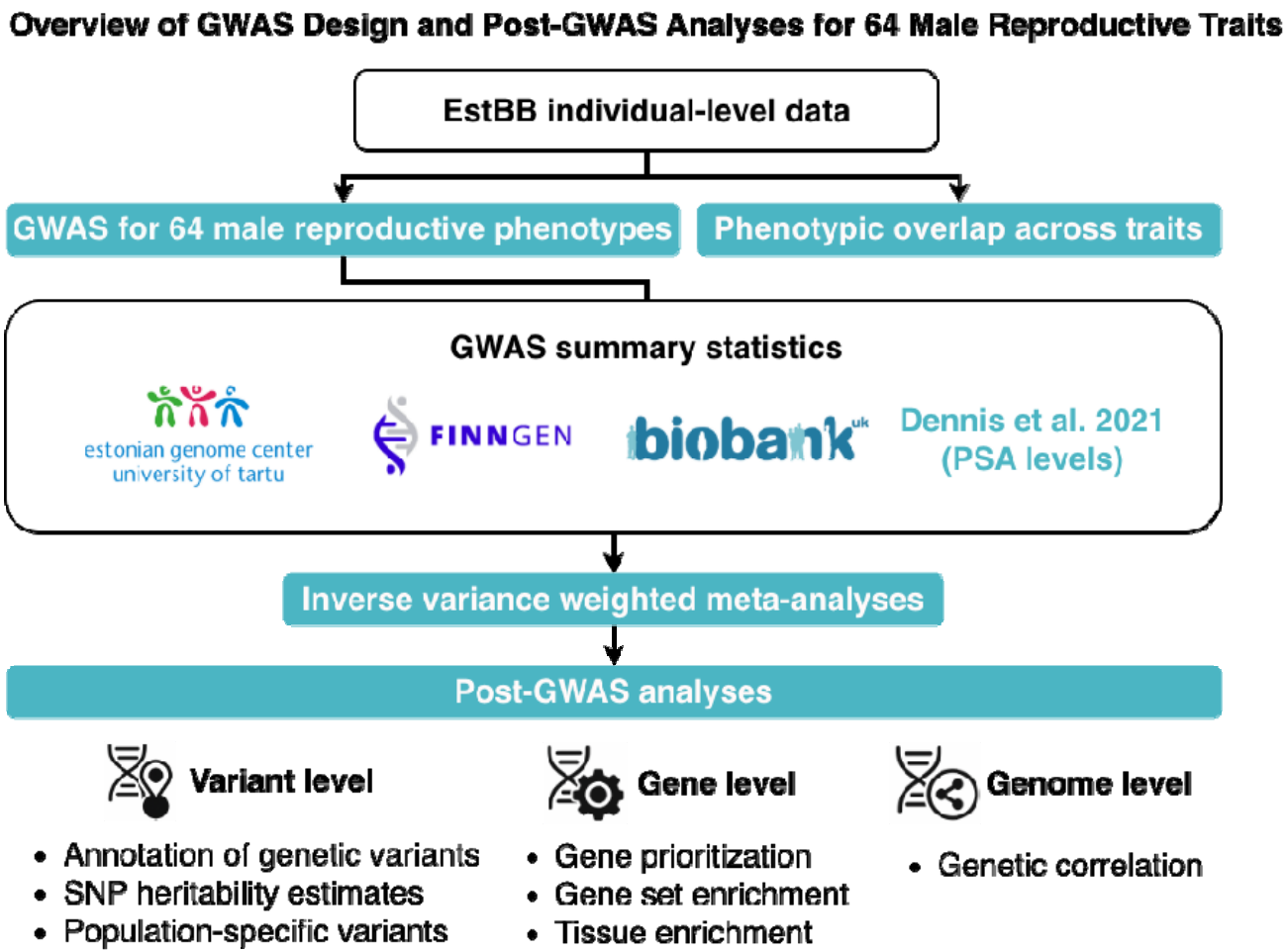

Figure 1. Flowchart of the study design and analysis workflow. The study comprised three main steps: (1) genome-wide association studies (GWAS) for 64 male reproductive phenotypes using individual-level data from the Estonian biobank (EstBB); (2) phenotypic overlap across traits using individual-level data from the EstBB; (3) GWAS meta-analyses incorporating summary statistics from EstBB, FinnGen, UK Biobank and/or Dennis et al. (2021); and (4) post-GWAS analyses performed at three levels: variant, gene and genome. PSA= prostate-specific antigen.

Estonian Biobank cohort, phenotype definition and GWAS analysis

EstBB is a population-based cohort of approximately 212,000 participants (~20% of the Estonian adult population) with extensive phenotypic and health-related information collected for each individual (3,13). At recruitment, participants provide broad consent allowing longitudinal linkage to national health registries and EHRs, including laboratory measurements. EstBB is regularly linked to the National Health Insurance Fund database (from 2004 onwards), as well as other national registries, including population, mortality, cancer, and prescription databases. For each participant, diagnostic information is available using ICD-10 codes (3).

Genotyping of DNA samples from the EstBB was done at the Core Genotyping Laboratory of the Institute of Genomics, University of Tartu using the Illumina Global Screening Array v3.0_EST. Altogether 295,810 samples were genotyped and then PLINK format files were created using Illumina GenomeStudio v2.0.4 (14). During the quality control, individuals were excluded if they had a call-rate < 95%, showed discordance between reported sex and genetically inferred sex based on X-chromosome heterozygosity, or were outliers in autosomal heterozygosity test. Variants were filtered by call-rate < 95% and Hardy-Weinberg equilibrium $P < 1 \times 10^{-4}$ (autosomal variants only) (14). Genotyped variants positions were in build 37 and were lifted over to build 38 using Picard (https://broadinstitute.github.io/picard/). Before imputation, variants with minimum allele frequency (MAF) < 1% and indels were removed. Phasing and imputation were done using Beagle v5.4 (beagle v.22Jul22.46e) and default settings (15). An Estonian population-specific imputation panel based on 2,695 whole-genome-sequenced samples was used as the reference, together with standard Beagle hg38 recombination maps (14). Based on principal components (PCs) analysis, samples that were not of European ancestry were excluded. Duplicate samples and monozygotic twins were identified using Kinship-based Inference for GWAS (KING) 2.2.7 (16), and one individual from each pair of duplicates was excluded.

Briefly, we used ICD-10 codes to identify individuals for male health phenotypes. Cases and controls were selected using ICD-10 codes for 52 male health phenotypes. Controls were individuals with no evidence of the analysed phenotypes, as defined in this study. Detailed information on the ICD codes/phenotypes used and cases and controls definitions and sample size are in the supplementary tables 2 and 3.

For the GWAS of male reproductive health-related laboratory parameters we selected 12 measurements of LOINC codes obtained from EHRs between the period of 2008-2020. More information on the codes and measurements used can be found in the supplementary table 3. During quality control, observations lying more than three standard deviations from the mean were excluded. The raw data extraction workflow can be found in the supplementary information (Text S3) of Kurvits et al. (2023). For individuals with multiple measurements, the most recent available measurement was used for GWAS analysis.

We conducted GWASs for the defined phenotypes (up to 67,744 men) using the REGENIE software (version 3.2) (18) adjusting for the first ten principal components, and year of birth.

FinnGen cohort, phenotype definition and GWAS analysis

FinnGen research project is a public-private partnership which contains 392,000 Finnish individual's genome information combined with digital health record data (4). The data release R8 (https://r8.finngen.fi/) was used for most of the analyses presented here. The data release R7 (https://r7.finngen.fi/) was used for two phenotypes ("Abnormal spermatozoa" and "Testicular dysfunction"), because they were not present in data release R8. Individuals in the FinnGen study were genotyped using Illumina and Affymetrix arrays (Illumina Inc., San Diego, CA, USA; Thermo Fisher Scientific, Santa Clara, CA, USA). Genotype imputation was performed with the Finnish population-specific SISu v3 reference panel, based on 3,775 whole-genome sequences. Disease endpoints were obtained from nation-wide health registries. GWAS analyses were conducted using a modified version of REGENIE v2.0.2, which includes dosage-based allele frequency calculation in cases and controls. The GWAS models were adjusted for age, 10 PCs and genotyping batch as covariates (4).

A total of 22 endpoints were available for the selected phenotypes and their GWAS summary statistics were used for the analyses. See Supplementary Table 4 for phenotype definitions.

UK Biobank cohort, phenotype definition and GWAS analysis

UKBB is a large population-based study that has genome and phenotype data for more than 500,000 individuals (5,19). For the meta-analysis, GWAS summary statistics were obtained via the PheWeb portal (https://pheweb.org/UKB-TOPMed/phenotypes) which contains EHRs-derived ICD codes from the White British participants of the UKBB. Genotype data had been imputed using Trans-Omics for Precision Medicine (TOPMed) and GWAS analyses were conducted using SAIGE adjusting for genetic relatedness, sex, year of birth and the first 4 PCs (20,21). A total of 19 phenotypes with available GWAS summary statistics were used in this study. See Supplementary Table 4 for phenotype definitions.

GWAS meta-analysis

GWAS meta-analyses were conducted for 27 phenotypes according to the availability of the phenotypes in the EstBB, FinnGen, and UKBB cohorts (Supplementary table 2, Figure 1). The "Prostate-specific antigen levels" (PSA levels) phenotype differed from other analyses in that it combined EstBB data with GWAS summary statistics from Dennis et al. (22), including 7,837 individuals of European ancestry (GWAS Catalog: GCST90012723). Phenotype and cohort specific sample sizes are reported in Supplementary Table 2.

Meta-analyses were done using the inverse of variance based fixed effect method implemented in Genome-Wide Association Meta-Analysis (GWAMA) (23). Genome-wide statistical significance in each analysis was defined as $P < 5 \times 10^{-8}$ and only variants that were present in at least two cohorts across all meta-analyses and passed the MAF filter of 1% were used in the follow-up analysis. Additionally, we report the number of genome-wide significant variants that met a stricter study-wide significance threshold of $P < 5 \times 10^{-8}/64 = 7.81 \times 10^{-10}$, accounting for the total number of phenotypes analysed. Heterogeneity for the variants was evaluated using I^2 index by Higgins et al. (24) implemented in GWAMA.

Annotation of GWAS signals

Annotation of GWAS signals was done using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) platform (v1.3.6a) (25). FUMA defines genomic risk loci from independent significant single nucleotide polymorphisms (SNPs) by merging linkage disequilibrium (LD) blocks, gene mapping and annotates (using ANNOVAR) all input variants.

Signals with lead variants rs184810529 and rs4907778 from "Malignant neoplasm of prostate" GWAS Meta-analysis were excluded since they were present in only one cohort. Variants located inside the Major Histocompatibility complex (MHC) region were not annotated due to the complex LD structure. Thus, the genomic locus with lead variant rs12662632 from "Redundant prepuce, phimosis and paraphimosis" was excluded from the follow-up analysis, along with the three signals (lead variants rs2523570, rs28371333, and rs9277621) in the human leukocyte antigen (HLA) region detected in the GWAS meta-analysis for the phenotype "Disorders of penis". "Urethritis and urethral syndrome" showed two genome-wide significant signals, but they were not included in the annotation, because: (1) the signals were singletons (no other SNPs have $r^2 > 0.6$); (2) there was no supporting evidence available from the biological perspective for the variants and candidate genes; (3) the analysis was done using only EstBB data and we lacked a suitable replication dataset.

Novel loci were defined as those showing no prior association with the corresponding phenotypes. Loci were delineated by the lead variant, variants in LD ($r^2 > 0.6$; FUMA-based), and SNPs within a 1Mb window around the lead variant. Evidence for prior associations was assessed using GWAS Catalog lookups implemented in FUMA, complemented by direct searches of the GWAS Catalog website and relevant published literature.

Gene prioritisation

To identify candidate causal genes at each genomic risk locus, we performed gene prioritisation based on four criteria: 1) proximity to the lead variant; 2) presence of missense variants within the gene, or in high LD ($r^2 > 0.6$) with it; 3) significant colocalization (posterior probability ≥ 0.8) between GWAS signal and gene expression (expression quantitative trait locus; eQTL) in relevant tissues; and 4) evidence of reproductive-related phenotypes in mouse models or involvement in human monogenic diseases.

For the Nearest gene parameter (proximity to the lead variant), protein-coding genes were identified using both FUMA and the Open Target Genetics platform (26,27). When results differed between these tools, priority was given to genes previously associated with reproductive health traits in the GWAS Catalog or supported by additional evidence from our prioritisation framework.

Candidate genes were ranked based on the number and relative strength of supporting criteria. Genes supported by multiple lines of evidence were prioritised over those supported by a single criterion. When only one criterion was available for a given locus, candidate genes were selected according to a predefined hierarchy of evidence strength, ordered were weighted as follows: 1) association with reproductive health phenotypes in mouse model; 2) eQTL-GWAS colocalization; 3) GWAS signal tagging a

missense variant within the gene; and 4) proximity to GWAS signal. This hierarchy reflects decreasing evidence specificity for causal involvement.

a. Correlation with mouse phenotypes

The Mouse Genome Database (MGD; https://www.informatics.jax.org/) was used to assess whether candidate genes identified in this study were associated with reproductive health-related phenotypes in mouse models or with monogenic human diseases affecting female or male reproductive traits. MGD integrates mouse phenotypes annotations with human gene-disease relationships from the National Center for Biotechnology Information (NCBI) and Online Mendelian Inheritance in Man (OMIM), as well as human disease-to-phenotype from the Human Phenotype Ontology (HPO) (28).

b. Colocalization

Colocalization analyses between GWAS summary statistics and the eQTL Catalog were performed using HyprColoc (29,30). Genotype-Tissue Expression (GTEx) v8 datasets for blood, prostate and testis tissues were used (dataset IDs: QTD000356, QTD000306, and QTD000336, respectively) (31). Lead SNPs were defined using a P threshold of 5×10^{-8}, and colocalization was tested within a 1Mb window around each lead variant.

SNP heritability

To calculate the SNP-based heritability (h^2SNP) for the phenotypes we used LD Score regression (LDSC) v1.0.1 regression implemented in Complex Trait Genetics Virtual Lab (0.4-beta) (32,33). For this analysis, we used only the phenotypes that presented effective sample size [Neff = 4/((1/N total cases) + (1/N total controls))] greater than 5,000 in total (N=14). Observed-scale heritability estimates and their standard errors were converted to the liability scale using the formula described in Lee et al. (34), implemented in R v4.2.2. Population prevalence was assumed to equal the sample prevalence (i.e., the proportion of cases in the meta-analysis). Calculated sample prevalences are shown in the Supplementary Table 8.

Genetic correlation analysis

We used LD Score regression implemented in the Complex Trait Genetics Virtual Lab (CTG-VL, version 0.4-beta) (32,33,35) to estimate genetic correlations. To account for multiple testing, we applied the Benjamini–Hochberg procedure using p.adjust function in R v4.2.3 to control the false discovery rate (FDR), considering associations with FDR-adjusted $P < 0.05$ as statistically significant.

Genetic correlations were performed for all male phenotypes with an effective sample size greater than 5,000, using two sources of GWAS summary statistics: (1) reproductive health-related hormone level GWAS datasets from the GWAS Catalog and published studies by Kachuri et al. (2023), Pott et al. (2019), Ruth et al. (2021), and Schumacher et al. (2018) (Supplementary table 16); and (2) 1,436 traits from the CTG-VL database, including eight lifestyle factors: smoking status (current, former and never), body mass index (BMI), educational attainment, and alcohol drinker status (current, former, and never). These traits were included to assess shared genetic architecture and pleiotropy between male reproductive traits and health-related lifestyle factors, which are partly modifiable.

Several phenotypes were excluded from the analysis due to their nonspecific definitions and substantial overlap with more clearly defined phenotypes included in the study. These comprised “Failure of genital response”, “Other disorders of penis”, “Impotence of organic origin”, “Disorders of prostate, unspecified”, “Hydrocele and spermatocele”, and “Orchitis, epididymitis and epididymo-orchitis without abscess”

Phenotype-level correlation analysis

To assess phenotype-level associations and evaluate the correspondence between genetic and phenotypic correlations, we examined case-level co-occurrence among binary clinical phenotypes included in the genetic correlation analysis (see above). Only individual-level data from the EstBB were used for this analysis. Case overlap was quantified using the Jaccard Index, calculated as c11 / (c10 + c01 + c11), where c11 is the number of individuals with both diagnoses, and c10 and c01 are the numbers of individuals with only one of the two diagnoses. Lifestyle (e.g., alcohol consumption, smoking status), behavioural (e.g., educational attainment), and quantitative (e.g. BMI) traits were excluded, as they are not defined as case-control variables. Results were visualized using the corrplot package in R (v0.92; https://github.com/taiyun/corrplot).

In addition, given prior evidence linking male infertility to cancer (36), we specifically evaluated the phenotypic overlap between “Male infertility” and malignant neoplasms. Cancer cases were defined using all ICD-10 C-codes from the malignant neoplasms category (C00-C97), with all remaining individuals defined as controls. Since the EstBB EHRs are available from 2004 onwards, the analysis was restricted to individuals aged 18-40 years at recruitment, corresponding to the reproductive age range. The association was tested using logistic regression, adjusting for age at recruitment and the first ten genetic principal components.

Gene-set and tissue enrichment

To investigate the biological pathways and tissue-specific expression patterns underlying male reproductive health phenotypes, we conducted gene-set and tissue enrichment analyses using FUMA, which implements Multi-marker Analysis of GenoMic Annotation (MAGMA) v1.08 for gene-property and gene-set analyses (25,37). MAGMA gene-set analysis was performed to test for enrichment of associated genes in predefined biological processes and molecular functions from the Gene Ontology (GO) database (38). MAGMA tissue expression analysis was performed using gene expression data from 53 tissue types of GTEx v8 to test for

preferential expression of phenotype-associated genes. All analyses were run using default parameters, and statistical significance was assessed after Bonferroni correction for the number of tested gene sets or tissues.

Population-specific variants

To identify population-specific variants in Finnish and Estonian populations, allele frequencies of lead variants were compared with those in broader European populations (non-Finnish Europeans). Population-specific minor allele frequencies were obtained from the Open Targets Genetics portal (https://genetics.opentargets.org/). Allele frequency enrichment was defined as the ratio of the minor allele frequency in the Finnish or Estonian populations to that in non-Finnish Europeans. Variants showing more than 1.5-fold enrichment were considered potentially population-specific (Supplementary table 7).

RESULTS

GWAS of male reproductive health traits

A total of 11 phenotypes showed genome-wide significant signals, where 143 genome-wide significant lead variants were identified (Supplementary table 5), 102 of them passing a multiple-testing threshold ($P < 7.81 \times 10^{-10}$) accounting for the number of tested phenotypes. The mean genomic inflation factor across all traits was 1.01 (range from 0.94 to 1.13) in binary traits and 0.94 (0.89 to 1.02) in quantitative traits, indicating that the test statistic inflation was well controlled.

We assessed the heterogeneity of genetic effects across studies using the Cochran's Q-test and the corresponding I^2 statistic for each lead SNP (Supplementary table 5). Overall, most genome-wide significant loci showed low heterogeneity, indicating consistent effects across cohorts (EstBB, FinnGen and UKBB). However, 17 loci exhibited high heterogeneity ($I^2 > 70\%$), possibly due differences in effect allele frequencies, phenotype definitions, or cohort-specific ascertainment and data collection procedures (Supplementary table 5; Supplementary figure 1).

Based on systematic comparison with previously reported GWAS findings, 47 trait-specific loci were classified as novel for the corresponding phenotypes. Among these, 26 variants have not been previously associated with any reproductive trait (Figure 2; Supplementary table 6). Three phenotypes were associated with significant genome-wide association signals for the first time: "Abnormal spermatozoa" (lead variant: rs17632542), "Redundant prepuce, phimosis and paraphimosis" (lead variant: rs11198372), and "Testicular dysfunction" (lead variant: rs1799941) (Table 1).

Gene prioritisation

A total of 699 genes showed at least one level of evidence of being associated with male phenotypes (Supplementary table 11). Notably, 8 genes prioritised for 4 phenotypes presented 3 levels of evidence: "Hyperplasia of prostate" - *CHEK2* (novel locus),

FOXF1 (novel locus), *POU5F1B*, *RSPO2* (novel locus), and *TFAP4* (novel locus); "Malignant neoplasm of prostate" - *AR* (novel locus), *LMTK2*, and *POU5F1B*; "Spermatocele" - *PAX8*; "Prostate-specific antigen levels" - *POU5F1B*.

All genes with three levels of evidence have been previously linked to reproductive traits and have well-characterized roles in reproduction. For example, *RSPO2* (*R Spondin 2 or roof plate-specific spondin-2*) regulates Wnt signaling and plays a key role in developmental processes and morphogenesis (39). Due to its function as an oncogenic driver, *RSPO2* has been widely implicated in various human carcinomas, including prostate cancer (12,39). Similarly, *CHEK2* (*checkpoint kinase 2*), a known tumour suppressor gene encoding a protein kinase activated in response to DNA damage, has been associated with several reproductive health traits, including androgen index free, breast cancer, female infertility, polycystic ovary syndrome, prostate carcinoma, PSA, testosterone levels, sex-hormone binding globulin, and uterine fibroids (9–12,40–43).

Based on our gene prioritisation approach, we identified 328 candidate causal genes across all the genomic risk loci (Supplementary table 11). Among these, four genes (*FGFR2*, *KLK3*, *POU5F1B*, and *TERT*) were prioritised for several prostate-related phenotypes, suggesting potential shared genetic mechanisms (Figure 3). *KLK3* and *POU5F1B* showed consistent links with prostate hyperplasia, prostate cancer, and PSA levels, in line with their established roles in prostate biology, while *TERT* was additionally associated with testicular malignancy, suggesting a broader involvement in male reproductive cancers. The recurrence of these genes across several phenotypes highlights their potential relevance to shared biological mechanisms underlying male reproductive function and disease.

Population-specific variants

Effect allele enrichment analyses in the Finnish and Estonian populations identified 12 variants with at least 1.5-fold enrichment compared to the broader European reference population, suggesting population-specific effects (Supplementary table 7). Among these, 4 variants were novel for the studied phenotypes, and 7 candidate genes are linked to phenotypes for the first time. Specifically, *MAF* and *WDR35* were newly associated with "Hyperplasia of prostate", while *MYC*, *WASHC5*, *PNPO*, *LRRC46* and *ITGB3* were linked to "Malignant neoplasm of prostate". These findings underscore the contribution of isolated populations in uncovering previously undetected disease loci.

For "Malignant neoplasm of prostate", several lead variants exhibited population-specific allele enrichment. The strongest signal was a missense variant on chromosome X (rs200737258), which displayed an 8.6-fold higher enrichment in the Finnish population and a 3.2-fold higher frequency in Estonian population compared with non-Finnish Europeans. This variant represents a novel genome-wide signal mapped to the androgen receptor (AR) locus. Additional population-enriched variants included rs77482050 on chromosome 2, which showed 6.5-fold enrichment in Finns and 1.9-fold in Estonians, and rs191767420, associated with "Hyperplasia of prostate", with 2.9-fold and 1.7-fold in Finnish and Estonian populations, respectively (Supplementary table 7).

SNP- Heritability

Heritability from common SNPs was calculated for 27 phenotypes with an effective sample size > 5,000. Of these, 16 phenotypes showed statistically significant SNP-based heritability ($P < 0.05$ from chi-square test; Figure 4; Supplementary table 8), with liability-scale estimates ranging from 2% to 13%. Among the reproductive health traits, "Spermatocele" had the highest SNP-heritability estimate (13%) followed by "Malignant neoplasm of prostate" (12%). "Other disorders of prostate" (3%), "Inflammatory diseases of prostate" (2%), and "Other disorders of male genital organs" (2%) exhibited the lowest estimates, possibly reflecting a greater contribution of non-genetic factors to these phenotypes.

Gene-set and tissue enrichment

MAGMA gene-set enrichment analysis across male reproductive health phenotypes identified several biological pathways enriched for associated genes (Supplementary tables 9 and 10). "Hyperplasia of prostate" had the largest number of significantly enriched gene sets, including pathways related to transcriptional regulation (e.g. "DNA- binding transcription factor activity" ($P_{bon} = 3.82 \times 10^{-4}$), transcriptional repressor complexes, cell cycle pathways, and fibroblast growth factor receptor signalling. These enrichments indicate that genes involved in transcriptional regulation and cell cycle- related process are overrepresented among loci associated with prostate hyperplasia. Similarly, "Malignant neoplasm of prostate" and "Phimosis" showed enrichment for gene sets involved in transmembrane transport and immune-related molecular functions, respectively.

To further explore the tissue-specific expression of genes associated with male reproductive health, we performed MAGMA tissue expression analysis using FUMA. Significant associations were identified for tissues relevant to male reproductive function. For "Malignant neoplasm of prostate", associated genes were significantly enriched for expression in the prostate ($P = 2.86 \times 10^{-6}$) and testis ($P = 2.60 \times 10^{-4}$), and for "Hyperplasia of prostate" for gene expression in the prostate ($P = 6.34 \times 10^{-3}$), consistent with the tissue origin and endocrine involvement of the diseases.

Additional enrichment for "Malignant neoplasm of prostate" was noted in stomach ($P = 3.83 \times 10^{-1}$) and *esophagus mucosa* ($P = 2.09 \times 10^{-4}$), which reflects a similar gene expression profile between these tissues (31). Surprisingly, "Premature ejaculation" was linked to expression in the fallopian tube ($P = 4.15 \times 10^{-4}$), although a female reproductive tissue, it may indicate shared gene expression patterns with male reproductive tissues or involvement in broader reproductive signalling pathways.

Genetic correlation

Pairwise genetic correlations

To investigate shared etiological architecture across male reproductive and genitourinary phenotypes, we conducted genetic correlation analyses using 29 phenotypes encompassing prostate inflammation, testicular and penile disorders, sexual dysfunction, infertility, and lifestyle traits such as BMI, smoking and alcohol use (Supplementary table 15).

At the genetic level, 28 pairwise correlations remained significant after FDR correction ($P_{FDR} < 0.05$; Figure 5). Hyperplasia of prostate exhibited strong positive genetic correlations with prostate-related conditions, including "Inflammatory disorders of prostate" ($_{rg} = 0.86$, $P_{FDR} = 1.02 \times 10^{-24}$), "Malignant neoplasm of prostate" ($_{rg} = 0.34$, $P_{FDR} = 2.39 \times 10^{-9}$), and "Prostate-specific antigen levels" ($_{rg} = 0.57$, $P_{FDR} = 0.003$). It also showed positive correlations with "Hydrocele", "Spermatocele", "Orchitis and epididymitis" and "Erectile dysfunction". Similarly, "Erectile dysfunction" was genetically correlated with "Chronic prostatitis" ($_{rg} = 0.61$, $P_{FDR} = 1.90 \times 10^{-5}$), "Hyperplasia of prostate" ($_{rg} = 0.31$, $P_{FDR} = 7.08 \times 10^{-5}$), "Malignant neoplasm of prostate" ($_{rg} = 0.30$, $P_{FDR} = 7.21 \times 10^{-4}$), "Sexual dysfunction" ($_{rg} = 0.95$, $P_{FDR} = 3.85 \times 10^{-11}$), and "Testicular dysfunction" ($_{rg} = 0.85$, $P_{FDR} = 1.21 \times 10^{-3}$)].

BMI showed significant negative genetic correlations with "Chronic prostatitis" ($_{rg} = -0.23$, $P_{FDR} = 0.0015$) and "Inflammatory disorders of prostate" ($_{rg} = -0.20$, $P_{FDR} = 0.0039$), and positive genetic correlations with "Redundant prepuce, phimosis and paraphimosis" ($_{rg} = 0.18$, $P_{FDR} = 0.00015$), and "Testicular dysfunction" ($_{rg} = 0.28$, $P_{FDR} = 0.025$). Among other lifestyle traits, only "Alcohol drinker status: Former" was genetically correlated with "Orchitis and epididymitis" ($_{rg} = 0.48$, $P_{FDR} = 0.03$).

Reproductive hormones and other reproductive traits

In addition to the genetic correlation analyses based on the GWAS summary statistics generated in this study, we also performed analyses using publicly available summary statistics from the GWAS Catalog and from studies with larger sample sizes for reproductive traits and hormone levels (Supplementary table 16). This expanded approach revealed significant genetic correlations (FDR-adjusted $P < 0.05$) for 6 reproductive traits that had not reached statistical significance in our original analysis, as well as for 14 reproductive hormones (Figure 6; Supplementary table 17).

Among the newly identified associations for reproductive phenotypes, the most notable were those involving "PSA levels", which showed strong positive genetic correlations with several prostate-related conditions, including "Malignant neoplasm of prostate", "Inflammatory disorders of prostate", "Other disorders of prostate", and "Chronic prostatitis" ($_{rg} = 0.34 - 0.69$). "PSA levels" were also positively correlated with "Hydrocele" ($_{rg} = 0.20$, $P_{FDR} = 4.79 \times 10^{-2}$). In addition, urinary tract infection exhibited a positive genetic correlation with "Hyperplasia of prostate" ($_{rg} = 0.47$, $P_{FDR} = 1.06 \times 10^{-3}$).

Regarding reproductive hormone correlations, testicular dysfunction showed eight significant negative genetic correlations with total testosterone, bioavailable testosterone, and SHBG levels ($_{rg}$ range: -0.47 to -0.36). Bioavailable testosterone levels in men were positively correlated with "Hydrocele" ($_{rg} = 0.20$, $P_{FDR} = 8.70 \times 10^{-3}$), "Hyperplasia of prostate" ($_{rg} = 0.10$, $P_{FDR} = 2.10 \times 10^{-2}$), "Malignant neoplasm of prostate" ($_{rg} = 0.11$, $P_{FDR} = 5.29 \times 10^{-3}$), and "Erectile dysfunction" ($_{rg} = 0.20$, $P_{FDR} = 4.79 \times 10^{-2}$).

Other phenotypes

Finally, we performed LDSC to estimate genetic correlations between the male reproductive phenotypes and traits from the CTG database. After FDR correction, seven phenotypes showed statistically significant genetic correlations: "Chronic prostatitis", "Hydrocele", "Hyperplasia of prostate", "Inflammatory disorders of prostate", "Malignant neoplasm of prostate", "Orchitis and

epididymitis" and "Redundant prepuce, phimosis and paraphimosis" (Supplementary table 18). The CTG traits were grouped into 4 categories: mental health, obesity/metabolic, reproductive, and other.

The mental health category included traits related to emotional symptoms (e.g. anxiety, nervousness, excessive worrying) and psychiatric diagnoses (e.g., depression, attention-deficit/hyperactivity disorder (ADHD), schizophrenia). Four male phenotypes ("Chronic prostatitis", "Hyperplasia of prostate", "Inflammatory disorders of prostate" and "Redundant prepuce, phimosis and paraphimosis") exhibited genetic correlations with this category and all the genetic correlations were positive (ranging from $_{rg}$ = 0.11 to 0.72).

In the obesity/metabolic category, genetic correlations with "Chronic prostatitis" and "Inflammatory disorders of prostate" were predominantly negative. In contrast, "Orchitis and epididymitis" and "Redundant prepuce, phimosis and paraphimosis" were mostly positively genetically correlated with traits in this category, except for "HDL cholesterol" and "Started insulin within one year diagnosis of diabetes" that showed negative genetic correlation with "Redundant prepuce, phimosis and paraphimosis".

In the reproductive category, all 7 male phenotypes demonstrated positive genetic correlations. This category was mostly composed of diagnoses related to prostate, female reproductive system, and genitourinary tract. Most of the correlations in this analysis confirmed our previous findings from the pairwise genetic correlation analysis (Supplementary table 18). In addition to these replicated results, we identified novel genetic correlations, including between "Chronic prostatitis" and "Age at first birth" (defined in women) ($_{rg}$ = 0.26, $P_{FDR} = 8.47 \times 10^{-3}$); "Female genital prolapse" and both "Hyperplasia of prostate" ($_{rg}$ = 0.19, $P_{FDR} = 4.17 \times 10^{-2}$) and "Inflammatory disorders of prostate" ($_{rg}$ = 0.35, $P_{FDR} = 2.72 \times 10^{-2}$); and "Abdominal and pelvic pain" with both "Hyperplasia of prostate" ($_{rg}$ = 0.19, $P_{FDR} = 1.34 \times 10^{-2}$) and "Inflammatory disorders of prostate" ($_{rg}$ = 0.60, $P_{FDR} = 4.96 \times 10^{-5}$).

Phenotype-level correlation

Case-level co-occurrence was accessed across 19 phenotypes using the Jaccard Index and between male infertility and malignant neoplasms (Supplementary table 19). We found that male reproductive and urogenital disorders displayed varying degrees of case overlap, with Jaccard indices ranging from 0.0009 to 0.75. The highest overlap was found between inflammation-related conditions, such as "Chronic prostatitis" and "Inflammatory disorders of prostate" (Jaccard = 0.75), followed by "Erectile dysfunction" and "Sexual dysfunction" (Jaccard = 0.43). Moderate overlap was found between "Hyperplasia of prostate" and "Inflammatory disorders of prostate" (Jaccard = 0.30), and between "Chronic prostatitis" and "Hyperplasia of prostate" (Jaccard = 0.25). "Male infertility" was associated with increased odds of "Malignant neoplasms" (OR = 1.60, $P = 4.47 \times 10^{-6}$), indicating that infertile men had nearly twice the odds of having a cancer diagnosis compared with those without infertility.

DISCUSSION

This large-scale genome-wide study of 64 male reproductive phenotypes was designed to advance understanding of the genetic architecture underlying a broad spectrum of genitourinary health in men. Genetic studies are essential for identifying biological

mechanisms that cannot be captured through clinical or environmental data alone, enabling deeper comprehensions into disease etiology and opportunities for early risk stratification. We report 143 genome-wide significant lead variants, including 47 novel trait-specific associations and the first genome-wide insights into three phenotypes: “Abnormal spermatozoa”, “Redundant prepuce, phimosis and paraphimosis”, and “Testicular dysfunction”. Many of our associations replicate previously reported genetic findings, providing validation across multiple cohorts and supporting the robustness of these genetic signals. These findings have important implications for understanding the genetics of male reproductive health and disease and offer a foundation for future translational efforts in precision urology and andrology, a domain historically underrepresented in complex trait genetics.

By integrating multiple lines of evidence, we prioritised 328 candidate causal genes across the genomic risk loci, including several with known relevance to reproductive function, such as *CHEK2*, *KLK3*, *RSPO2* and others newly implicated in prostate phenotypes (44–46). The recurrence of specific genes across distinct yet related conditions (e.g., *FGFR2*, *POU5F1B*, *TERT*) highlights potential pleiotropic mechanisms in prostate development, function, and disease (47–49). Notably, *RSPO2* and *CHEK2*, both implicated in oncogenic and DNA repair pathways, respectively, may serve as functional links between reproductive function and cancer susceptibility (39,50).

Extending these gene-level insights, our analyses demonstrate the value of genetically bottlenecked populations for identifying population-enriched risk alleles that may be rare or absent in broader European cohorts. A prominent example is the missense variant rs200737258 at the *AR* locus, identified as novel genome-wide association signal for prostate cancer and markedly enriched in Finnish and Estonian populations. Given the central role of *AR* in prostate development and androgen insensitivity syndromes (51–53), this finding underscores the biological relevance of population-specific discoveries. More broadly, the enrichment of multiple lead variants in Finland and Estonia is consistent with the founder effects and reduced admixture increasing the frequency of otherwise rare alleles, thereby enhancing power for discovery (4,42). Similar patterns have been reported in FinnGen and the Estonian Biobank, where population-specific allele frequency structures have enabled the detection of clinically relevant variants not captured in pan-European analyses (10). Together, these results highlight the importance of including diverse European populations to improve variant discovery and genetic representation in large-scale human genetic studies.

Among the newly associated phenotypes, we identified distinct signals that broaden our understanding of the genetic architecture of male reproductive health. Notably, rs17632542 in *KLK3* (kallikrein related peptidase 3), which encodes PSA, was associated with “Abnormal spermatozoa”. This variant has known links with prostate cancer, PSA levels, and chronic prostatitis (9,54), and may affect sperm quality via its role in semen liquefaction (55,56). These findings suggest that variation in *KLK3* may contribute to both oncogenic and reproductive pathways through a shared biochemical mechanism.

The variant rs11198372 was significantly associated with “Redundant prepuce, phimosis, and paraphimosis”, and our gene prioritisation approach identified three genes (*EMX2*, *KCNK18*, and *ENO4*) with known reproductive phenotypes in knockout mouse models, suggesting plausible roles in human preputial development. *EMX2* knockout mice exhibit agenesis of internal male genitalia, including absence of the vas deferens, epididymis, and seminal vesicles, as well as agonadism in both sexes (57). *ENO4* deficiency leads to severe sperm flagellar defects and male infertility (58,59), while *KCNK18* mutants show abnormal and enlarged epididymis

morphology. While none of these genes were prioritised based on coding or regulatory annotations at this locus, their shared reproductive phenotypes in mouse models support a potential developmental mechanism linking them to the studied trait. Functional validation will be essential to clarify whether these genes contribute to penile and foreskin morphology in humans.

A particularly noteworthy finding in our study is the potential interplay between immune-related genetic factors and metabolic traits influencing these conditions. Specifically, we identified a genome-wide significant association at the HLA region (rs12662632) and observed a positive genetic correlation with BMI, as well as with several other obesity/metabolic traits. The HLA region plays a pivotal role in immune system regulation and, especially the HLA Class II, has been implicated in various inflammatory and autoimmune conditions (60). In the context of phimosis, that is a chronic inflammation of the foreskin, it may be influenced by HLA-mediated immune responses, contributing to the development of fibrotic changes and foreskin constriction (61). Obesity is characterized by chronic low-grade inflammation, which can exacerbate immune responses and tissue remodelling processes (62). The positive genetic correlation between the traits suggests that higher adiposity may enhance susceptibility to foreskin pathologies through inflammatory pathways. Furthermore, studies have shown that obesity-related metabolic disturbances, such as diabetes mellitus, can impair skin integrity and healing, potentially influencing the progression of phimosis (63). This integrative understanding highlights the importance of considering both immunogenetic and metabolic factors in the etiology of foreskin pathologies.

Unexpectedly, we observed statistically significant negative genetic correlation between BMI and obesity/metabolic traits and both inflammatory diseases of prostate and chronic prostatitis. This is surprising given that prior studies have reported positive associations between obesity and prostatic inflammation (64). Obesity has been linked to systemic inflammatory states and alterations in local immune responses, which may influence prostate tissue remodelling and inflammatory profiles (65). However, an observational study reported an inverse association between BMI and prostatitis-like symptoms, although the authors attributed this finding to potential confounding by physical activity and did not pursue it further (66). While inconclusive, this evidence aligns with the direction of our genetic findings, which are less susceptible to behavioural confounding. These results suggest a more complex relationship between metabolic traits and prostate inflammation, potentially involving distinct mechanisms of genetic predisposition and environmental exposure. Further investigation is needed to clarify how obesity-related genetic factors might modulate prostate immune responses and inflammatory disease risk.

Beyond correlations among male phenotypes, we also observed positive genetic correlations between inflammatory prostate phenotypes and female traits related to reproductive timing and pelvic health. These associations may reflect shared genetic influences on biological aging and chronic inflammation, as leukocyte telomere length, a heritable marker of cellular aging, has been linked to systemic inflammation and age-related disease risk (67). In this context, age-dependent female traits such as genital prolapse and age at first birth, together with inflammatory prostate disorders in men, may partly capture common genetic susceptibility to aging-related inflammatory and connective tissue pathways rather than sex-specific pathology alone.

Consistent with previous evidence linking male infertility and cancer risk (36), we also observed a statistically significant positive association between Male infertility and malignant neoplasms, with infertile men showing 1.6-fold higher odds of cancer after

accounting for age and population structure. This finding supports the notion of shared genetic or biological mechanisms predisposing to both impaired spermatogenesis and tumorigenesis.

Our findings reveal consistent genetic and phenotypic patterns across male reproductive and urological phenotypes, highlighting a shared genetic architecture among conditions traditionally viewed as distinct. In particular, the clustering of "Hyperplasia of prostate", "Chronic prostatitis", and "Erectile dysfunction" mirrors prior reports that link prostatic and sexual disorders through inflammatory and endocrine pathways, with chronic inflammation playing a central role in prostate disease development and progression (68).

"Erectile dysfunction" also correlated with broader sexual and structural disorders, suggesting common etiologies, potentially involving androgen or vascular dysfunction. The observed association between higher BMI and both "Testicular dysfunction" and "Redundant prepuce, phimosis and paraphimosis" aligns with prior work linking obesity to male hypogonadism and compromised reproductive health (69). Similarly, the genetic correlation between former alcohol use and epididymal inflammation may reflect persistent effects of alcohol-induced mitochondrial dysfunction, even after cessation (70).

Although we replicated many previous findings, this study has some limitations. First, since we do not have supporting data for the "Premature ejaculation" GWAS using only EstBB data, the results should be interpreted with care. Second, "Prostate-specific antigen levels" from EstBB should also be interpreted with caution, as they were extracted from electronic health records and likely reflect measurements taken in the context of clinical suspicion, which may influence their levels. Third, the definition of the phenotypes was solely based on ICD-10 and LOINC codes which may not capture the clinical and biological complexity of certain conditions, potentially masking heterogeneity or overlapping etiologies across phenotypes. Finally, this study was based on European ancestry cohorts, studies using other ancestries need to be done to confirm the association of these variants for the respective phenotypes in non-European populations. Together, our integrative analysis underscores the value of combining genetic and clinical data to disentangle the complex architecture of male reproductive health.

CONCLUSIONS

This study presents the most comprehensive genome-wide analysis of male reproductive health to date, spanning 64 phenotypes and integrating data from over 500,000 individuals across EstBB, FinnGen, and UK Biobank. We identified 143 genome-wide significant variants, including 47 novel trait-specific associations and the first genetic signals for conditions such as "Abnormal spermatozoa", "Redundant prepuce, phimosis and paraphimosis" and "Testicular dysfunction". Through rigorous gene prioritisation and integrative analyses, including tissue-specific expression, functional annotation, and genetic correlation, we highlighted shared biological mechanisms underlying prostate, testicular, penile, and hormonal traits.

Our findings reveal extensive pleiotropy among male reproductive conditions, underscore the role of population-specific variants in discovery, and provide strong evidence for immune, endocrine, and developmental pathways contributing to genitourinary disease. The identification of novel loci in genes such as *CHEK2*, *RSPO2*, *KLK3*, *SHBG*, and *EMX2* further expands the catalogue of candidates for functional follow-up and potential clinical translation.

By addressing a critical knowledge gap in men's health, this work lays a foundation for improved diagnostics, risk prediction, and therapeutic targeting in male reproductive disorders. Future efforts should prioritize longitudinal clinical integration, functional validation of candidate genes, and exploration of gene-environment interactions to advance precision medicine approaches in andrology and urology.

ACKNOWLEDGEMENTS

This work was supported by the Estonian Research Council grants PRG1911, PRG1844, and PSG776; by the Ministry of Education and Research Centres of Excellence grant TK214 (Centre of Excellence in Genomics and Translational Medicine); and by the Estonian Ministry of Education and Research (project TEM-TA72). This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060011 and Project No. 2021-2027.1.01.24-0444. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

The research was conducted using the Estonian Center of Genomics/Roadmap II funded by the Estonian Research Council (project number TT17). The activities of the EstBB are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the EstBB. Individual level data analysis in the EstBB was carried out under ethical approval 1.1-12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 6-7/GI/29022 from the Estonian Biobank. Data analysis was carried out in part in the High-Performance Computing Center of University of Tartu. We want to acknowledge the participants and investigators of the EstBB, FinnGen, and UKBB studies.

The GTEX data used for the analyses described in this manuscript were obtained from eQTL Catalog on October 1, 2024. The Genotype Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

DATA AVAILABILITY

Summary statistics from UK Biobank and FinnGen used in this study are accessible through the UKBB PheWeb portal (https://pheweb.org/UKB-TOPMed/phenotypes) and FinnGen PheWeb (https://r7.finngen.fi/ and https://r8.finngen.fi/), respectively. Full GWAS summary statistics and meta-analysis will be deposited in the GWAS Catalog upon acceptance. Genome-wide association analyses were conducted using REGENIE (version 3.2), and meta-analyses were performed with GWAMA. Additional resources consulted include the Mouse Genome Database (http://www.informatics.jax.org), GTEx Portal (https://gtexportal.org/home/), eQTL Catalog (https://www.ebi.ac.uk/eqtl/), GWAS Catalog (https://www.ebi.ac.uk/gwas/), and the Complex Trait Genetics Virtual Lab (https://vl.genoma.io/).

Table 1. A summary of first time reported GWAS and GWAS meta-analyses results for male reproductive health phenotypes.

ABNORMAL SPERMATOZOA								
Chr (band)	**POS (hg19)**	**Lead variant**	**EA/ NEA**	**EAF**			**OR (95% CI)**	**P-value**
				EstBB	**FinnGen**	**UKBB**		
chr19(q13.33)	51361757	rs17632542	T/C	0.06	0.07	0.07	1.50 (1.32-1.71)	8.83×10^{-10}
REDUNDANT PREPUCE, PHIMOSIS AND PARAPHIMOSIS								
Chr (band)	**POS (hg19)**	**Lead variant**	**EA/ NEA**	**EAF**			**OR (95% CI)**	**P-value**
				EstBB	**FinnGen**	**UKBB**		
chr6(p21.32)	32578383	rs12662632	G/C	0.05	0.03	0.02	1.78 (1.66-1.92)	6.13×10^{-53}
chr10(q26.11)	120108324	rs11198372	G/A	0.59	0.63	NA	1.11 (1.07-1.15)	7.74×10^{-10}
TESTICULAR DYSFUNCTION								
Chr (band)	**POS (hg19)**	**Lead variant**	**EA/ NEA**	**EAF**			**OR (95% CI)**	**P-value**
				EstBB	**FinnGen**	**UKBB**		
chr17(p13.1)	7533423	rs1799941	G/A	0.29	0.27	0.26	1.31 (1.21-1.42)	1.45×10^{-10}

Abbreviations: Chr, chromosome; POS, genomic position (hg19); EA, effect alle; NEA, non-effect allele; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.

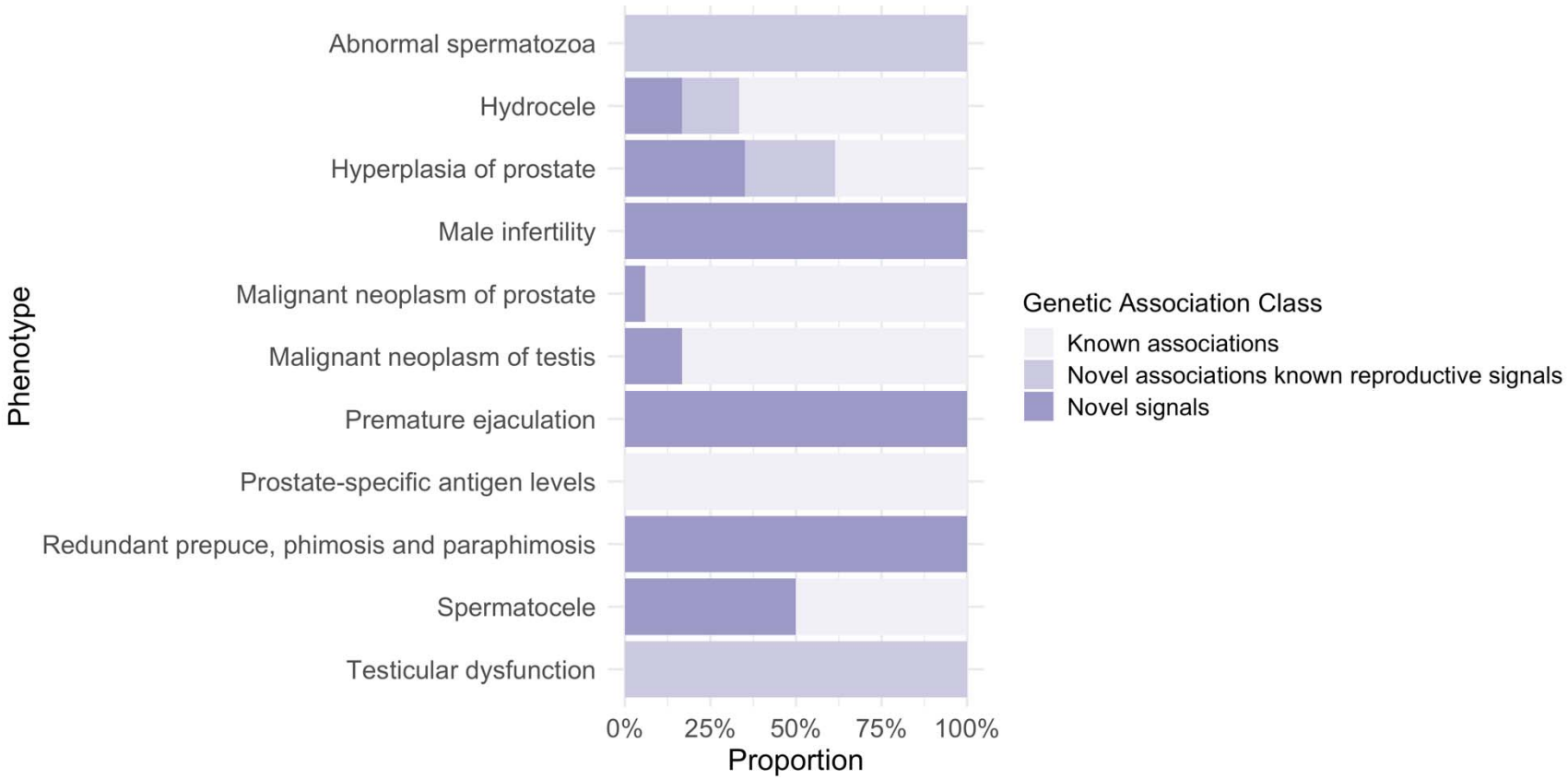

Figure 2. Overview of the known and novel genetic associations across male reproductive health phenotypes.

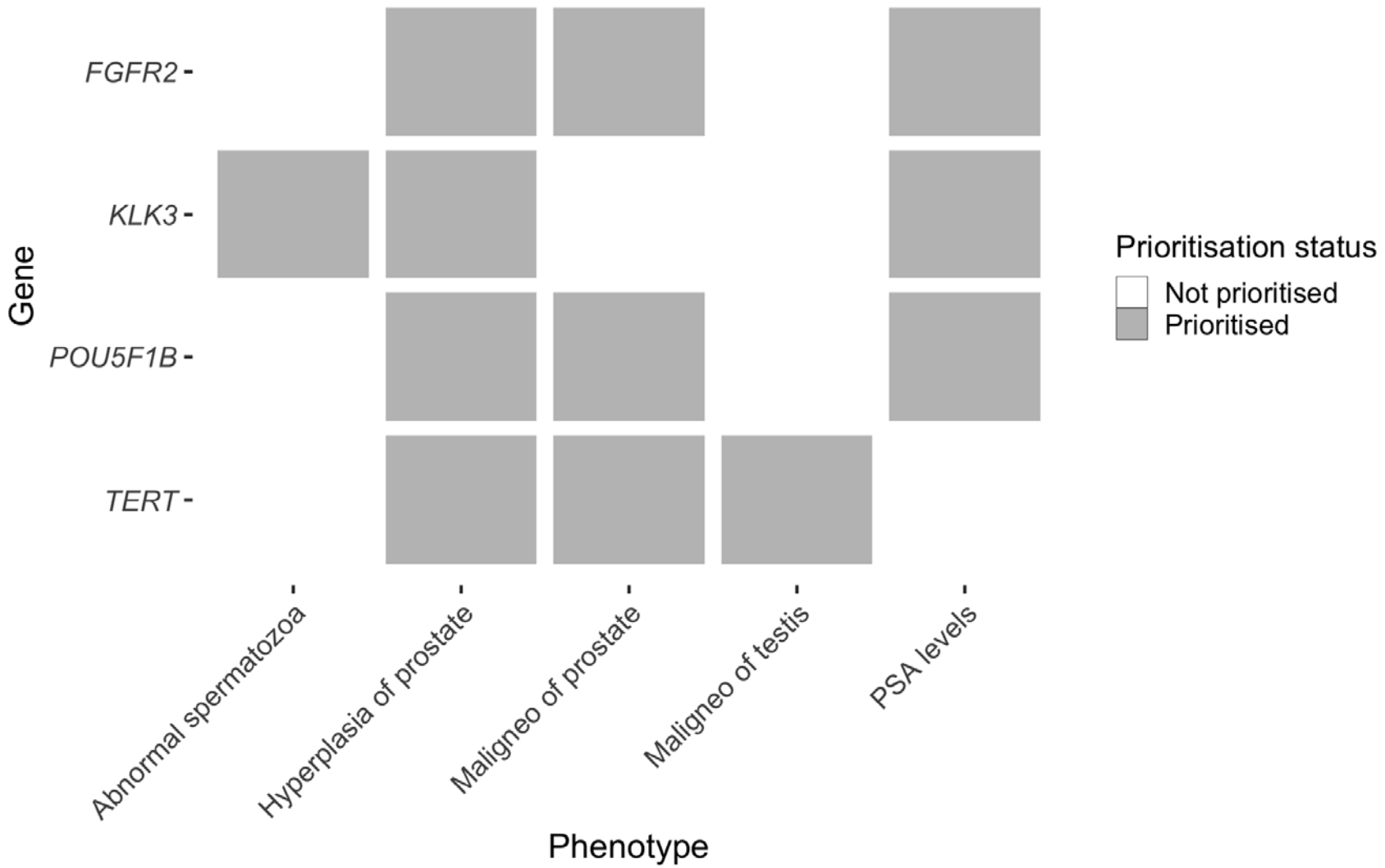

Figure 3. **Recurrence of prioritised genes across prostate-related traits.** Heatmap indicates whether a gene was prioritised as a candidate causal gene for each phenotype based on the multi-criteria framework described in Methods. Only genes associated with multiple prostate-related traits are shown. Maligneo= Malignant neoplasm; PSA= Prostate-specific antigen.

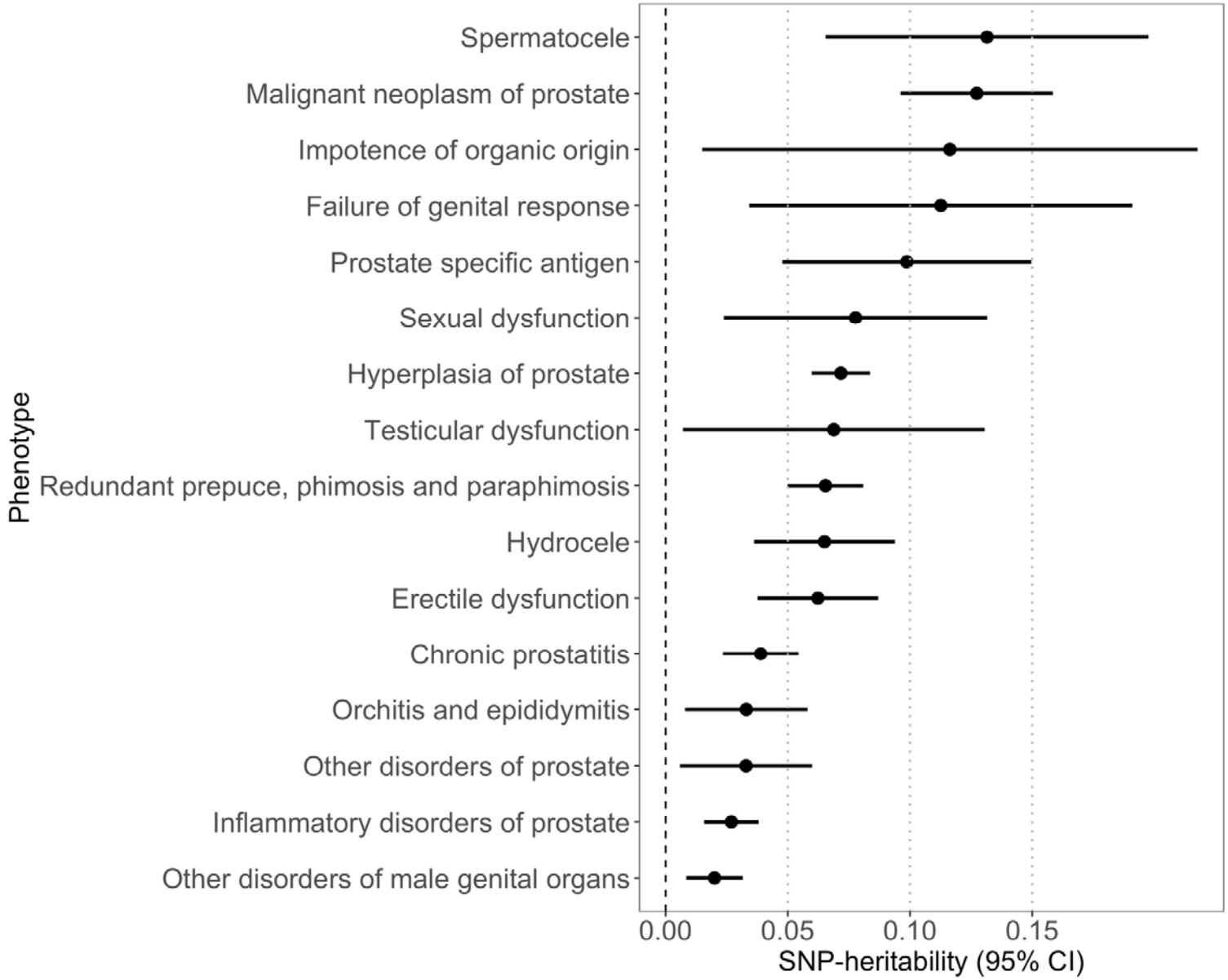

Figure 4. Statistically significant SNP heritability (P < 0.05, chi-square test) for male reproductive phenotypes with effective sample size > 5,000.

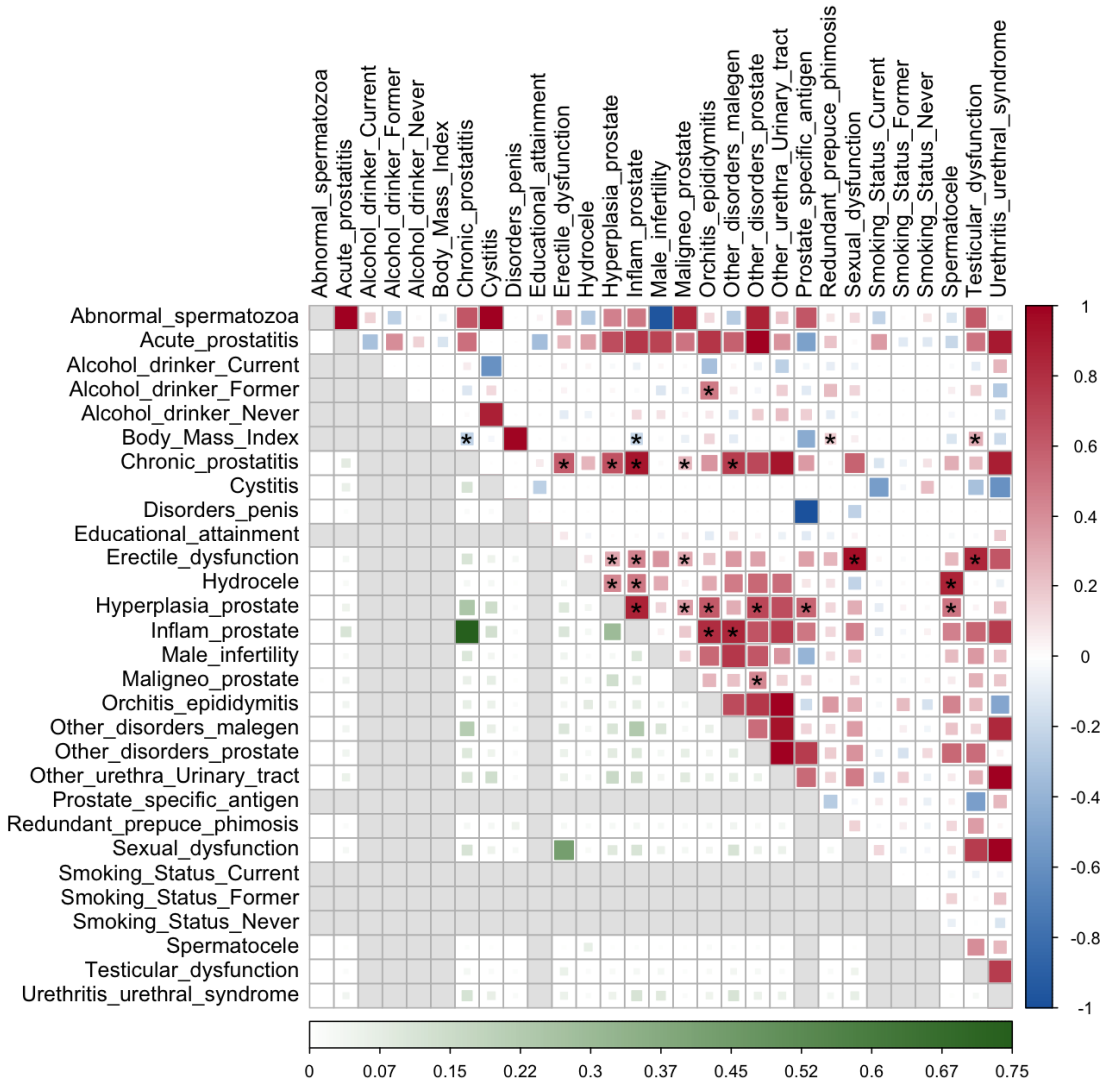

Figure 5. Results of Phenotypic overlap (bottom plot) and Genetic Correlation (upper plot) Analyses. The figure displays results for male reproductive binary diagnoses/traits (effective sample size >5,000). For the phenotypic overlap plot, darker green colours indicate a higher similarity/case overlap of the two phenotypes. Grey colour means that the phenotypic overlap was not computed for the phenotype pair. For the genetic correlation plot, the red colour indicates positive correlation, and blue indicates negative correlation. Correlation with FDR-adjusted P-values < 0.05 are denoted by asterisks in the upper plot. The size of the square corresponds to the respective estimated values. Inflam= Inflammatory disorders; Maligneo= Malignant neoplasm.

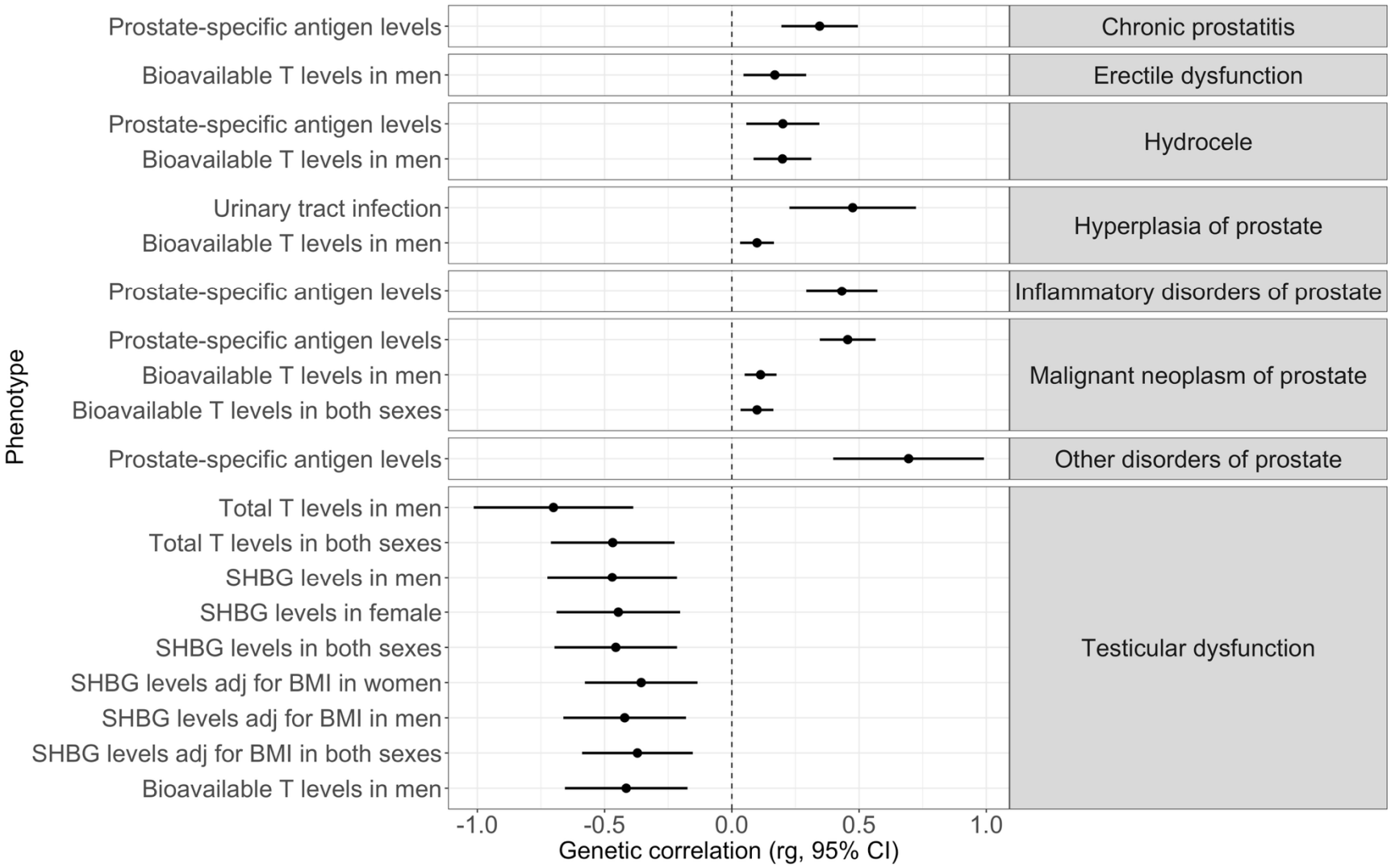

Figure 6. Genetic correlation for the male phenotypes and GWASs summary statistics for reproductive hormones/traits performed in other studies. The source of the GWASs can be found in the Supplementary table 17. Only genetic correlations that were statistically significant after FDR correction and were not reported in the pairwise genetic correlations between our GWASs are displayed in the figure. SHBG = sex-hormone binding globulin; T = testosterone; BMI = body mass index.

REFERENCES

1. Benonisdottir S, Straub VJ, Kong A, Mills MC. Genetics of female and male reproductive traits and their relationship with health, longevity and consequences for offspring. Nat Aging. 2024 Dec 13;4(12):1745–59.

2. De Jonge CJ, Barratt CLR, Aitken RJ, Anderson RA, Baker P, Chan DYL, et al. Current global status of male reproductive health. Hum Reprod Open. 2024 Mar 5;2024(2):hoae017.

3. Milani L, Alver M, Laur S, Reisberg S, Haller T, Aasmets O, et al. The Estonian Biobank's journey from biobanking to personalized medicine. Nat Commun. 2025 Apr 5;16(1):3270.

4. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023 Jan;613(7944):508–18.

5. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct 11;562(7726):203–9.

6. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update. Clin Chem. 2003 Apr 1;49(4):624–33.

7. Organization WH. International Statistical Classification of Diseases and Related Health Problems: Alphabetical index. World Health Organization; 2004. 826 p.

8. Figuerêdo J, Krebs K, Pujol-Gualdo N, Haller T, Võsa U, Volke V, et al. Uncovering the shared genetic components of thyroid disorders and reproductive health. Eur J Endocrinol. 2024 Aug 1;191(2):211–22.

9. Verma A, Huffman JE, Rodriguez A, Conery M, Liu M, Ho YL, et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. Science. 2024 July 19;385(6706):eadj1182.

10. Pujol Gualdo N, Džigurski J, Rukins V, Pajuste FD, Wolford BN, Võsa M, et al. Atlas of genetic and phenotypic associations across 42 female reproductive health diagnoses. Nat Med. 2025 May;31(5):1626–34.

11. Venkatesh SS, Wittemans LBL, Palmer DS, Baya NA, Ferreira T, Hill B, et al. Genome-wide analyses identify 25 infertility loci and relationships with reproductive traits across the allele frequency spectrum. Nat Genet. 2025 May;57(5):1107–18.

12. Wang A, Shen J, Rodriguez AA, Saunders EJ, Chen F, Janivara R, et al. Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. Nat Genet. 2023 Dec;55(12):2065–74.

13. Mägi R. Diverse landscape of genomic research within the Estonian Biobank. Hum Mol Genet. 2025 May 2;ddaf026.

14. Mitt M, Kals M, Pärn K, Gabriel SB, Lander ES, Palotie A, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. Eur J Hum Genet. 2017 July;25(7):869–76.

15. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021 Oct 7;108(10):1880–90.

16. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010 Nov 15;26(22):2867–73.

17. Kurvits S, Harro A, Reigo A, Ott A, Laur S, Särg D, et al. Common clinical blood and urine biomarkers for ischemic stroke: an Estonian Electronic Health Records database study. Eur J Med Res. 2023 Mar 25;28(1):133.

18. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021 July;53(7):1097–103.

19. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. 2015 Mar 31;12(3):e1001779.

20. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature. 2021 Feb;590(7845):290–9.

21. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet. 2018 Sept;50(9):1335–41.

22. Dennis JK, Sealock JM, Straub P, Lee YH, Hucks D, Actkins K, et al. Clinical laboratory test-wide association scan of polygenic scores identifies biomarkers of complex disease. Genome Med. 2021 Jan 13;13(1):6.

23. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010 May 28;11(1):288.

24. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003 Sept 6;327(7414):557–60.

25. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017 Nov 28;8(1):1826.

26. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. 2021 Jan 8;49(D1):D1311–20.

27. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet. 2021 Nov;53(11):1527–33.

28. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ. Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. Nucleic Acids Res. 2020 Nov 24;49(D1):D981–7.

29. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet. 2021 Sept;53(9):1290–9.

30. Foley CN, Staley JR. cnfoley/hyprcoloc: First release of software [Internet]. Zenodo; 2020 [cited 2022 Sept 5]. Available from: https://zenodo.org/record/4293559

31. THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sept 11;369(6509):1318–30.

32. Cuellar-Partida G, Lundberg M, Kho PF, D'Urso S, Gutierrez-Mondragon L, Hwang LD. Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyse [Internet]. 2019 [cited 2023 Jan 18]. Available from: https://doi.org/10.1101/518027

33. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015 Mar;47(3):291–5.

34. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating Missing Heritability for Disease from Genome-wide Association Studies. Am J Hum Genet. 2011 Mar 11;88(3):294–305.

35. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Nov;47(11):1236–41.

36. Valkna A, Juchnewitsch AG, Põlluaas L, Lillepea K, Tjagur S, Dutta A, et al. Significantly increased load of hereditary cancer-linked germline variants in infertile men. Hum Reprod Open. 2025 Mar 7;2025(2):hoaf008.

37. Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Comput Biol. 2015 de abr. de;11(4):e1004219.

38. Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 4;224(1):iyad031.

39. Srivastava A, Rikhari D, Srivastava S. RSPO2 as Wnt signaling enabler: Important roles in cancer development and therapeutic opportunities. Genes Dis. 2024 Mar;11(2):788–806.

40. Boonen RACM, Vreeswijk MPG, van Attikum H. CHEK2 variants: linking functional impact to cancer risk. Trends Cancer. 2022 Sept;8(9):759–70.

41. Leinonen JT, Mars N, Lehtonen LE, Ahola-Olli A, Ruotsalainen S, Lehtimäki T, et al. Genetic analyses implicate complex links between adult testosterone levels and health and disease. Commun Med. 2023 Jan 18;3(1):4.

42. Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, et al. Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome. Hum Reprod. 2022 Feb 1;37(2):352–65.

43. Kim J, Williams A, Noh H, Jasper EA, Jones SH, Jaworski JA, et al. Genome-wide meta-analysis identifies novel risk loci for uterine fibroids within and across multiple ancestry groups. Nat Commun. 2025 Mar 6;16(1):2273.

44. Kirchner K, Seidel C, Paulsen FO, Sievers B, Bokemeyer C, Lessel D. Further Association of Germline CHEK2 Loss-of-Function Variants with Testicular Germ Cell Tumors. J Clin Med. 2023 Jan;12(22):7065.

45. De Cian MC, Gregoire EP, Le Rolle M, Lachambre S, Mondin M, Bell S, et al. R-spondin2 signaling is required for oocyte-driven intercellular communication and follicular growth. Cell Death Differ. 2020 Oct;27(10):2856–71.

46. Gupta N, Sudhakar DVS, Gangwar PK, Sankhwar SN, Gupta NJ, Chakraborty B, et al. Mutations in the prostate specific antigen (PSA/KLK3) correlate with male infertility. Sci Rep. 2017 Sept 11;7(1):11225.

47. Breyer JP, Dorset DC, Clark TA, Bradley KM, Wahlfors TA, McReynolds KM, et al. An Expressed Retrogene of the Master Embryonic Stem Cell Gene POU5F1 Is Associated with Prostate Cancer Susceptibility. Am J Hum Genet. 2014 Mar;94(3):395–404.

48. Wang C, Liu Z, Ke Y, Wang F. Intrinsic FGFR2 and Ectopic FGFR1 Signaling in the Prostate and Prostate Cancer. Front Genet. 2019 Jan 30;10:12.

49. Dratwa M, Wysoczańska B, Łacina P, Kubik T, Bogunia-Kubik K. TERT—Regulation and Roles in Cancer Formation. Front Immunol. 2020 Nov 19;11:589929.

50. Kirchner K, Gamulin M, Kulis T, Sievers B, Kastelan Z, Lessel D. Comprehensive Clinical and Genetic Analysis of CHEK2 in Croatian Men with Prostate Cancer. Genes. 2022 Nov;13(11):1955.

51. Tan ME, Li J, Xu HE, Melcher K, Yong E leong. Androgen receptor: structure, role in prostate cancer and drug discovery. Acta Pharmacol Sin. 2015 Jan;36(1):3–23.

52. Weng H, Li S, Huang JY, He ZQ, Meng XY, Cao Y, et al. Androgen receptor gene polymorphisms and risk of prostate cancer: a meta-analysis. Sci Rep. 2017 Jan 16;7(1):40554.

53. Li C, Wang X, Wang X, Li X, Chen W, Zhao M, et al. Novel pathogenic variants in the androgen receptor gene associated with androgen insensitivity syndrome identified through exome sequencing and *in silico* analysis. Gene. 2023 Apr 15;860:147225.

54. Kachuri L, Hoffmann TJ, Jiang Y, Berndt SI, Shelley JP, Schaffer KR, et al. Genetically adjusted PSA levels for prostate cancer screening. Nat Med. 2023 June;29(6):1412–23.

55. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat Rev Cancer. 2008 Apr;8(4):268–78.

56. Srinivasan S, Kryza T, Bock N, Tse BWC, Sokolowski KA, Janaththani P, et al. A PSA SNP associates with cellular function and clinical outcome in men with prostate cancer. Nat Commun. 2024 Nov 6;15(1):9587.

57. Miyamoto N, Yoshida M, Kuratani S, Matsuo I, Aizawa S. Defects of urogenital development in mice lacking Emx2. Dev Camb Engl. 1997 May;124(9):1653–64.

58. Nawaz S, Hussain S, Bilal M, Syed N, Liaqat K, Ullah I, et al. A variant in sperm-specific glycolytic enzyme enolase 4 (ENO4) causes human male infertility. J Gene Med. 2024;26(1):e3583.

59. Nakamura N, Dai Q, Williams J, Goulding EH, Willis WD, Brown PR, et al. Disruption of a Spermatogenic Cell-Specific Mouse Enolase 4 (Eno4) Gene Causes Sperm Structural Defects and Male Infertility1. Biol Reprod. 2013 Apr 1;88(4):90, 1–12.

60. Berryman MA, Ilonen J, Triplett EW, Ludvigsson J. Important denominator between autoimmune comorbidities: a review of class II HLA, autoimmune disease, and the gut. Front Immunol. 2023 Sept 26;14:1270488.

61. Hieta NK, Haataja MAM, Tapana L. Comorbidities in Male Patients With Lichen Sclerosus: A Case-Control Study. J Low Genit Tract Dis. 2023 Oct;27(4):378.

62. Soták M, Clark M, Suur BE, Börgeson E. Inflammation and resolution in obesity. Nat Rev Endocrinol. 2025 Jan;21(1):45–61.

63. Wang CC, Wang I, Lai YH. Type 2 diabetes mellitus increases the risk of penile inflammatory disorders in men aged between 30 and 49: a 5-year follow-up study. Int J Diabetes Dev Ctries. 2023 Apr 1;43(2):267–73.

64. Wang YB, Yang L, Deng YQ, Yan SY, Luo LS, Chen P, et al. Causal relationship between obesity, lifestyle factors and risk of benign prostatic hyperplasia: a univariable and multivariable Mendelian randomization study. J Transl Med. 2022 Oct 29;20(1):495.

65. Saha A, Kolonin MG, DiGiovanni J. Obesity and prostate cancer — microenvironmental roles of adipose tissue. Nat Rev Urol. 2023 Oct;20(10):579–96.

66. Wallner LP, Clemens JQ, Sarma AV. Prevalence of and risk factors for prostatitis in African American men: The Flint Men's Health Study. The Prostate. 2009;69(1):24–32.

67. Wong JYY, De Vivo I, Lin X, Fang SC, Christiani DC. The Relationship between Inflammatory Biomarkers and Telomere Length in an Occupational Prospective Cohort Study. Rouet P, editor. PLoS ONE. 2014 Jan 27;9(1):e87348.

68. De Nunzio C, Presicce F, Tubaro A. Inflammatory mediators in the development and progression of benign prostatic hyperplasia. Nat Rev Urol. 2016 Oct;13(10):613–26.

69. Service CA, Puri D, Al Azzawi S, Hsieh TC, Patel DP. The impact of obesity and metabolic health on male fertility: a systematic review. Fertil Steril. 2023 Dec 1;120(6):1098–111.

70. Roach AN, Bhadsavle SS, Higgins SL, Derrico DD, Basel A, Thomas KN, et al. Alterations in sperm RNAs persist after alcohol cessation and correlate with epididymal mitochondrial dysfunction. Andrology. 2024 July;12(5):1012–23.