

1 **Large-scale genome-wide association study to determine the genetic underpinnings of**  
2 **female genital tract polyps**

3 **Running title:** GWAS of female genital tract polyps

4 Amruta D. S. Pathare<sup>1,3†\*</sup>, Natàlia Pujol-Gualdo<sup>2†</sup>, Valentina Rukins<sup>2</sup>, Jelisaveta Džigurski<sup>2</sup>,  
5 Maire Peters<sup>1,3</sup>, Estonian Biobank Research Team, Reedik Mägi<sup>2</sup>, Andres Salumets<sup>1,3,4</sup>, Merli  
6 Saare<sup>1,3</sup>, Triin Laisk<sup>2</sup>

7 1 Competence Centre on Health Technologies, Tartu, Estonia

8 2 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

9 3 Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of  
10 Tartu, Tartu, Estonia

11 4 Division of Obstetrics and Gynaecology, Department of Clinical Science, Intervention and  
12 Technology (CLINTEC), Karolinska Institute and Karolinska University Hospital, Stockholm,  
13 Sweden

14 **\*Corresponding author:** Amruta D. S. Pathare, [amruta.pathare@ut.ee](mailto:amruta.pathare@ut.ee) ORCID No: 0000-  
15 0002-5608-2370

16

17

---

† The authors consider that the first two authors should be regarded as joint First Authors to guide clinical practice.

18 **Abstract**

19 STUDY QUESTION

20 Can a large-scale genome-wide association study (GWAS) meta-analysis identify the genomic  
21 risk loci and associated candidate genes for female genital tract (FGT) polyps, provide insights  
22 into the mechanism underlying their development, and inform potential overlap with other  
23 traits, including endometrial cancer?

24 SUMMARY ANSWER

25 GWAS meta-analysis of FGT polyps highlighted the potentially shared mechanisms between  
26 polyp development and cancerous processes.

27 WHAT IS KNOWN ALREADY

28 Small-scale candidate gene studies have focused on biological processes such as estrogen  
29 stimulation and inflammation to clarify the biology behind FGT polyps. However, the exact  
30 mechanism for the development of polyps is still elusive. At the same time, a genome-wide  
31 approach, which has become the gold standard in complex disease genetics, has never been  
32 used to uncover the genetics of the FGT polyps.

33 STUDY DESIGN, SIZE, DURATION

34 We performed a genome wide association study (GWAS) meta-analysis including a total of  
35 25,100 women with FGT polyps (International Classification of Disease, ICD-10 diagnosis  
36 code N84) and 207,193 female controls (without N84 code) of European ancestry from the  
37 FinnGen study (11,092 cases and 94,394 controls) and the Estonian Biobank (EstBB, 14,008  
38 cases and 112,799 controls).

## 39 PARTICIPANTS/MATERIALS, SETTING, METHODS

40 A meta-analysis and functional annotation of GWAS signals were performed to identify and  
41 prioritise genes in associated loci. To determine associations with other phenotypes, we  
42 performed a look-up of associated variants across multiple traits and health conditions, a genetic  
43 correlation analysis, and a phenome-wide association study (PheWAS) with ICD10 diagnosis  
44 codes.

## 45 MAIN RESULTS AND THE ROLE OF CHANCE

46 Our GWAS meta-analysis revealed ten significant ( $P < 5 \times 10^{-8}$ ) genomic risk loci. Two signals,  
47 rs2277339 ( $P = 7.6 \times 10^{-10}$ ) and rs1265005 ( $P = 1.1 \times 10^{-9}$ ) (in linkage disequilibrium (LD) with  
48 rs805698  $r^2 = 0.75$ ), are exonic missense variants in *PRIMI*, and *COL17A1* genes, respectively.  
49 Based on the literature, these genes may play a role in cellular proliferation. Several of the  
50 identified genomic loci had previously been linked to endometrial cancer and/or uterine  
51 fibroids. Thus, highlighting the potentially shared mechanisms underlying tissue overgrowth  
52 and cancerous processes, which may be relevant to the development of polyps. Genetic  
53 correlation analysis revealed a negative correlation between sex hormone-binding globulin  
54 (SHBG) and the risk of FGT polyps ( $rg = -0,21$ ,  $se = 0.04$ ,  $P = 2.9 \times 10^{-6}$ ), and on the phenotypic  
55 level (PheWAS), the strongest associations were observed with endometriosis, leiomyoma of  
56 the uterus and excessive, frequent and irregular menstruation.

## 57 LARGE SCALE DATA

58 The complete GWAS summary statistics will be made available after publication through the  
59 GWAS Catalogue (<https://www.ebi.ac.uk/gwas/>).

## 60 LIMITATIONS, REASONS FOR CAUTION

61 In this study, we focused broadly on polyps of FGT and did not differentiate between the polyp  
62 subtypes. The prevalence of FGT polyps led us to assume that most women included in the  
63 study had endometrial polyps. Further study on the expression profile of FGT polyps could  
64 complement the GWAS study to substantiate the functional importance of the identified  
65 variants.

## 66 WIDER IMPLICATIONS OF THE FINDINGS

67 The study findings have the potential to significantly enhance our understanding of the genetic  
68 mechanisms involved, paving the way for future functional follow-up, which in turn could  
69 improve the diagnosis, risk assessment, and targeted treatment options, since surgery is the only  
70 line of treatment available for diagnosed polyps.

## 71 TRIAL REGISTRATION NUMBER:

72 Not applicable

73 Key words: Genome-wide association study, Female genital tract polyps, Endometrial polyps,  
74 Benign disorders, cell proliferation, *PRIMI*, *COL17A1*.

## 75 Introduction

76 Polyps of the female genital tract (FGT) are generally benign tissue overgrowths found  
77 in both reproductive-aged and postmenopausal women. While the prevalence of polyps can be  
78 as high as 50%, most women are asymptomatic. Hence, polyps are usually detected incidentally  
79 during routine ultrasound examinations, diagnostic hysteroscopy for other gynaecological  
80 disorders, or infertility treatment among women of reproductive age (Hinckley and Milki, 2004;  
81 Fatemi *et al.*, 2010; Karayalcin *et al.*, 2010; Bettocchi *et al.*, 2011). However, the occurrence  
82 of polyps is age-dependent, with a higher prevalence among postmenopausal women compared

83 to premenopausal women (Dreisler *et al.*, 2009). Despite the asymptomatic and benign nature  
84 of FGT polyps, for some women, they can negatively impact the daily quality of life by causing  
85 abnormal vaginal bleeding and infertility.

86 Endometrial polyps (EPs) are the most common type of FGT polyps, with a prevalence  
87 ranging from 7.8% to 50% (Dreisler *et al.*, 2009; de Azevedo *et al.*, 2016; Tanos *et al.*, 2017).  
88 In contrast, endocervical polyps occur only in 2% to 5% of cases, whereas vaginal polyps are  
89 rarer (Tanos *et al.*, 2017); therefore, limited evidence is available about their biology and  
90 clinical implications. Some known risk factors for the development of EPs are advanced age,  
91 hypertension, diabetes, obesity, hyperoestrogenism, administration of hormone replacement  
92 therapy (HRT), and tamoxifen (Vitale *et al.*, 2021; Vieira *et al.*, 2022). Nevertheless, the exact  
93 pathogenesis of FGT polyps remains unclear. Histologically, EPs are characterised by large,  
94 thickened blood vessels with fibrous stroma and irregularly shaped glandular spaces (Tanos *et*  
95 *al.*, 2017). Moreover, EPs can be associated with concomitant intrauterine pathologies like  
96 endometrial hyperplasia, adenomyosis, endometriosis and chronic endometritis (Annan *et al.*,  
97 2012; Raz *et al.*, 2021). Despite the benign nature of EPs, around 3.5% of cases progress into  
98 carcinoma (Lee *et al.*, 2010; Sasaki *et al.*, 2018; Uglietti *et al.*, 2019). At the same time, the  
99 extent of biological mechanisms and pathways shared between EPs and carcinoma, as well as  
100 unique molecular characteristics of each condition remain unclear.

101 Very little is known about the heritability and genetic background of FGT polyps.  
102 Genetic factors, including chromosomal translocations in 6p21-22, 12q13-15, or 7q22 regions,  
103 may contribute to polypoid morphology (Dal Cin *et al.*, 1995; Nijkang *et al.*, 2019; Vieira *et*  
104 *al.*, 2022), and genetic disorders like Lynch or Cowden syndrome have been reported to be  
105 accountable for the development of EPs (Vieira *et al.*, 2022). However, a recent study by Sahoo  
106 *et al.* did not confirm the presence of chromosomal rearrangements in endometrial polyps  
107 (Sahoo *et al.*, 2022).

108           Thus, the confirmed pathophysiology for the development of FGT polyps is still  
109 unknown. However, it can be assumed that the development and mechanism of polyps are  
110 multifactorial and can also depend on genetic predisposition. Thus far, the proposed genetic  
111 mechanisms are primarily based on small candidate gene studies and are inconclusive (Altaner  
112 *et al.*, 2006; Pal *et al.*, 2008; Banas *et al.*, 2018; Doria *et al.*, 2018; Takeda *et al.*, 2019). Studies  
113 in other complex diseases have shown that genome-wide association studies (GWAS) can  
114 provide valuable insight into disease biology (Claussnitzer *et al.*, 2016), but as far as we are  
115 aware, no large-scale GWAS have been published for FGT polyps. Therefore, we performed a  
116 GWAS meta-analysis to identify the genetic variants associated with FGT polyps, followed by  
117 numerous post-GWAS analyses to understand the shared and unique genetic underpinnings of  
118 FGT polyps and endometrial cancer. Ultimately, this knowledge can highlight potential clinical  
119 avenues for appropriate diagnosis and management of FGT polyps.

## 120 **Materials and methods**

### 121 **Ethical approval**

122           All Estonian Biobank (EstBB) participants have signed a broad informed consent form,  
123 and analyses were carried out under ethical approvals 1.1-12/624 and 1.1-12/2733 from the  
124 Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs)  
125 and data release application 6-7/GI/630 from the EstBB. For the FinnGen study, we used only  
126 publicly available GWAS summary statistics without individual-level data and thus, a separate  
127 ethics approval was not needed.

### 128 **Study cohorts**

129           Our analyses included a total of 25,100 women with polyps of the FGT (International  
130 Classification of Disease, ICD-10 diagnosis code N84) and 207,193 female controls (without

131 the N84 code) of European ancestry from two studies: summary level statistics from the  
132 FinnGen R7 data release (11,092 cases and 94,394 controls) and individual-level data from the  
133 EstBB (14,008 cases and 112,799 controls). In FinnGen, cases were defined using the ICD-10  
134 code N84 or corresponding ICD-9 (6210, 6227, 6237, 6246) and ICD-8 (62520) codes. Similar  
135 to EstBB, controls were defined as women who did not have the abovementioned disease codes.

136 To increase study power, we did not distinguish between different types of FGT polyp  
137 location in our phenotype definition. However, according to the FinnGen Risteys browser  
138 ([https://r7.risteys.finngen.fi/phenocode/N14\\_POLYPPFEMGEN](https://r7.risteys.finngen.fi/phenocode/N14_POLYPPFEMGEN)), 20% of FGT polyp cases  
139 were also cases for phenotype “Uterine polyps”, while in the EstBB data, 70.5 % of N84 cases  
140 had a diagnosis for uterine polyps (as defined by the presence of the ICD-10 code N84.0). This  
141 most likely reflects differences in the source of phenotype information - FinnGen phenotype  
142 definitions mostly use hospital records and thus involve more severe cases, while the EstBB  
143 also uses primary care records.

#### 144 **Cohort-level analyses**

145 The EstBB is a population-based biobank including more than 200,000 individuals  
146 (approximately 135,000 of them women) representing 20% of the Estonian adult population.  
147 Information on ICD codes is obtained via regular linking with the National Health Insurance  
148 Fund and other relevant databases. Individuals with ICD-10 code N84 (mean age at recruitment  
149 49.2 years, standard deviation 12.1) were categorised as cases having been diagnosed with  
150 polyps of the genital tract, and all female biobank participants without the diagnosis were  
151 considered as controls (mean age at recruitment 44.3 years, sd 16.6). The genotyping procedure  
152 for EstBB has been described previously (Koel *et al.*, 2023; Laisk *et al.*, 2021; Pujol-Gualdo *et*  
153 *al.*, 2022). Briefly, Illumina GSAv1.0, GSAv2.0, and GSAv2.0\_EST arrays were used for the  
154 genotyping of biobank participants at the Core Genotyping Lab of the Institute of Genomics,

155 University of Tartu. Individuals with a call-rate less than 95% and sex mismatch between  
156 phenotypic and chromosomal data were excluded from the analysis. Before imputation,  
157 genotyped variants were filtered by applying criteria of call rate  $< 95\%$ , Hardy-Weinberg  
158 equilibrium  $P < 10^{-4}$  (for autosomal variants only), and minor allele frequency  $< 1\%$ . Eagle  
159 v2.3 software was used for pre-phasing, and Beagle was used for imputation. The population-  
160 specific imputation reference of 2297 whole genome sequencing samples was used. Association  
161 analysis was performed using REGENIE v2.2.4 with year of birth and 10 principal components  
162 as covariates in step I, and variants with a minor allele count  $< 5$  were excluded by default. In  
163 downstream association analysis, poorly imputed variants with an INFO score  $< 0.4$  were  
164 excluded.

165 For FinnGen, GWAS summary statistics from the R7 data release were used, and  
166 therefore, individual-level data was not available. The summary statistics were obtained from  
167 [https://www.finnngen.fi/en/access\\_results](https://www.finnngen.fi/en/access_results), whereas the FinnGen cohort and the genotyping/data  
168 analysis details have been previously described in Kurki et al. (Kurki *et al.*, 2023). To  
169 summarise, age, 10 principal components, and genotyping batch were used in REGENIE v2.0.2  
170 analysis as covariates, and for FinnGen summary statistics, variants with a minor allele count  
171  $> 5$  and imputation INFO score  $> 0.6$  were included.

## 172 **GWAS meta-analysis**

173 A meta-analysis using fixed-effects inverse variance weighting with genomic control  
174 was performed with the GWAMA v2.1 tool. The genome-wide significance level was set to  
175  $P < 5 \times 10^{-8}$ . The genomic inflation factors (lambda) for the individual study summary statistics  
176 were 1.046 (EstBB) and 1.045 (FinnGen). Variants present in both cohorts ( $n = 12,363,169$ )  
177 were included in downstream analyses.

## 178 **Annotation of GWAS signals**



179 To annotate the GWAS signals and prioritise potential biologically relevant genes at  
180 associated loci, we adopted the following approach.

181 First, we used the Functional Mapping and Annotation of Genome-Wide Association  
182 Studies (FUMA GWAS) platform v1.5.2 to identify genetic association loci. FUMA is an online  
183 platform designed for the annotation, prioritisation, and interpretation of GWAS results that  
184 uses data from multiple databases to annotate GWAS signals (Watanabe *et al.*, 2017). In the  
185 first step of this analysis, independent significant variants, lead signals and genomic risk loci  
186 were defined. Independent significant variants (IndSigSNPs) were defined as variants that were  
187 genome-wide significant ( $P < 5 \times 10^{-8}$ ) and had a pairwise LD  $r^2 < 0.6$ , according to the 1000G  
188 EUR reference panel. From this subset, lead variants were derived. Finally, risk loci were  
189 defined from independent significant SNPs by merging LD blocks if they are less apart than  $r^2$   
190  $< 0.6$ . Thus, a genomic risk locus can contain several lead SNPs and/or independent significant  
191 SNPs, depending on the size of the locus and the LD structure. Thereafter, potential significant  
192 candidate SNPs (GWAS meta-analysis  $P > 0.05$ ) were determined to be in LD with any of the  
193 IndSigSNPs ( $r^2 \geq 0.6$ ) within a 1Mb window and had a MAF of  $\geq 1\%$ . These candidate SNPs  
194 were subjected to further annotation using multiple databases such as Annotate Variation  
195 (ANNOVAR) (Wang *et al.*, 2010), RegulomeDB scores (ranging from 1 to 7, where lower  
196 score indicates greater evidence for having regulatory function) (Boyle *et al.*, 2012), and  
197 Combined Annotation-Dependent Depletion (CADD) (a continuous score showing how  
198 deleterious the SNP is to protein structure/function; scores  $>12.37$  indicate potential  
199 pathogenicity) (Kircher *et al.*, 2014), 15 chromatin states from the Roadmap Epigenomics  
200 Project (ENCODE Project Consortium, 2012; Roadmap Epigenomics Consortium *et al.*, 2015),  
201 expression quantitative trait loci (eQTL) data (genotype-tissue expression (GTEx) v6 and v7)  
202 (GTEx Consortium, 2013) and 3D chromatin interactions from HI-C experiments of 21  
203 tissues/cell types (Schmitt *et al.*, 2016). This process provides information on the location,

204 functional impact, and potential regulatory effects of detected SNPs. For eQTL annotations, we  
205 focused on tissues similar to uterine/vaginal tissue (GTEx Consortium, 2020).

206 Using the lead signal identified by FUMA, we performed a look-up in the Open Targets  
207 genetics database (Ghoussaini *et al.*, 2021). This database combines several layers of evidence  
208 across a wide range of cell types and tissues to generate an aggregate ‘variant to gene’ (V2G)  
209 score. The V2G score provides identification and correlation of likely causal variants and genes  
210 to prioritise the potential functional genes associated with the identified variants. V2G score  
211 aggregates several parameters like distance from the canonical transcript start site, eQTLs and  
212 protein QTLs (pQTLs) datasets, datasets for chromatin interactions and conformation,  
213 molecular phenotypes, and *in silico* functional predictions (using the Variant Effect Predictor  
214 or VEP score).

215 Briefly, we prioritised genes in the identified loci by selecting three genes having the  
216 highest V2G score. Then we additionally identified those loci where the GWAS signal includes  
217 a missense variant using annotation data from FUMA. To provide additional support for  
218 prioritisation and further explore regulatory effects, we looked at eQTL associations according  
219 to FUMA annotations, and if none were reported, we looked at potential chromatin interactions  
220 as these may also indicate regulatory effects. To gain insight into endometrial-specific eQTLs,  
221 we queried the candidate SNPs defined by FUMA in the endometrial eQTL database (Fung *et*  
222 *al.*, 2018).

### 223 **Gene-based testing**

224 Analysis by Multi-marker Analysis of GenoMic Annotation (MAGMA) v1.6 (de Leeuw  
225 *et al.*, 2015), with the default settings in FUMA (Watanabe *et al.*, 2017), was used to perform  
226 gene-based association analysis to complement the single variant analyses. Gene-based analysis  
227 enables to detect the joint effect of multiple genetic variants and can thus increase the power to  
228 detect associations. Briefly, variants located in the gene body were assigned to protein-coding

229 genes (n = 18,895; Ensembl build 85), and the SNP P-values were merged into a gene test  
230 statistic using the SNP-wise mean model (de Leeuw *et al.*, 2015). The level of genome-wide  
231 significance was set at  $0.05/18,895 = 2.6 \times 10^{-6}$ , taking into account the number of tested genes.

### 232 **Genetic associations with other traits:**

233 During the FUMA functional mapping, candidate SNPs were linked with the GWAS  
234 catalogue (<https://www.ebi.ac.uk/gwas/>, GWAS Catalogue e0\_r2022-11-29, FUMA v1.5.1) to  
235 explore the association of genetic variants with previously published GWAS of different  
236 phenotypic traits.

### 237 **Genetic correlation analysis**

238 To estimate the genetic correlations between our FGT polyps meta-analysis and 1,335  
239 other traits, we utilised the LD Score Regression method implemented in the Complex Traits  
240 Genetics Virtual Lab (CTG-VL) (<https://genoma.io/>) and 3 additional endometrial cancer traits  
241 available in the GWAS catalogue (accession codes GCST006464, GCST006465 and  
242 GCST006466) (Bulik-Sullivan *et al.*, 2015). Statistical significance was determined by  
243 applying a multiple testing correction (FDR < 5%) using the p.adjust function in R v3.6.3.

### 244 **Associations with other phenotypic traits**

245 To determine the associations between ICD-10 diagnosis main codes and the N84  
246 diagnosis of polyps in the FGT, we conducted an analysis using individual level data from the  
247 EstBB. Logistic regression was used to test the associations between N84 and other ICD-10  
248 codes while controlling for age at recruitment and 10 genetic principal components to account  
249 for population stratification and avoid false associations due to ancestry/regional differences  
250 between cases and controls. Age was included as a covariate to account for incomplete  
251 electronic diagnosis data for older participants in the biobank. Our analysis was limited to the  
252 diagnosis main codes along with all the subcodes to increase the power of the data. Statistically

253 significant associations were determined by applying Bonferroni correction (2000 tested ICD  
254 main codes, corrected P threshold of  $2.5 \times 10^{-5}$ ). Odds ratios (ORs) were calculated by the  
255 logistic regression method and represented as adjusted ORs. The resulted associations were  
256 filtered to remove the diagnoses related to exogenous factors (such as injuries, poisoning,  
257 accidents, assaults, etc. in the S, T, U, V, W, X, and Y subchapters) and the PheWAS library  
258 (<https://github.com/PheWAS/PheWAS>) was used to visualise the results. All analyses were  
259 performed using R v4.1.3

260 In the FinnGen data, we did not have access to the individual level data, but using the  
261 Risteys portal ([https://r7.risteys.finnngen.fi/phenocode/N14\\_POLYPFEMGEN](https://r7.risteys.finnngen.fi/phenocode/N14_POLYPFEMGEN)), we explored  
262 the results of the survival analysis evaluating associations between FGT polyps and other  
263 selected phenotypes. A detailed description of the survival analysis can be found in the Risteys  
264 documentation (<https://r7.risteys.finnngen.fi/documentation>), but briefly, this type of analysis  
265 tests the association between an exposure endpoint and an outcome endpoint. For example, in  
266 the context of FGT polyps, what is the association between a diagnosis of FGT polyp (exposure  
267 endpoint) and endometrial carcinoma (outcome endpoint).

## 268 **Results**

### 269 **Summary of GWAS**

270 GWAS of a total of 25,100 women with polyps of the FGT and 207,193 female controls  
271 revealed ten significant ( $P < 5 \times 10^{-8}$ ) genomic risk loci and 23 independent signals (Fig. 1,  
272 Table 1). The strongest signal, rs1702136 ( $P = 2.17 \times 10^{-19}$ ), was observed on chromosome 3,  
273 which is an intronic variant of the *EEFSEC* gene. The majority of the significant signals  
274 (91.3%) were either intronic or intergenic variants, whereas two signals, rs2277339 ( $P = 7.57 \times$   
275  $10^{-10}$ ) and rs1265005 ( $P = 1.09 \times 10^{-9}$ , in LD with rs805698,  $r^2 = 0.75$ ) on chromosomes 12 and  
276 10, respectively, were exonic missense variants.

277

278 **Table 1. Summary statistics of significant genomic risk loci with lead SNP and**  
 279 **independent significant SNP.**

chr:pos:A1:A2	rsID of Lead SNP (effect allele)	P-value	OR (95% CI)	IndSigSNPs	Effect Allele frequency (Est/Fin)	Heterogeneity p-value
3:128118711:A:G	rs1702136 (G)	2.17 x 10 <sup>-19</sup>	0.90 (0.88-0.92)	rs1702136 rs3732402 rs4857866 rs7650365 rs2999051 rs13095166 rs74924715 rs1735527	0.75/0.75	0.99
8:116869477:C:T	rs800578 (T)	4.91 x 10 <sup>-11</sup>	1.08 (1.05-1.10)	rs800578 rs2736213	0.24/0.26	0.90
4:95731394:A:G	rs2865375 (A)	4.36 x 10 <sup>-10</sup>	0.93 (0.92-0.95)	rs2865375 rs10033997	0.54/0.52	0.45
12:57146069:G:T	rs2277339 (T)	7.57 x 10 <sup>-10</sup>	1.10 (1.06-1.13)	rs2277339	0.87/0.86	0.37
10:105585753:C:T	rs19309775 (C)	7.59 x 10 <sup>-10</sup>	0.83 (0.78-0.88)	rs19309775 rs17116149 rs7911816 rs11360949 rs1265005	0.97/0.95	0.92
1:61592380:A:G	rs12751005 (A)	1.09 x 10 <sup>-9</sup>	1.06 (1.04-1.09)	rs12751005	0.68/0.67	0.90
13:41869725:A:G	rs77478686 (G)	8.21 x 10 <sup>-9</sup>	1.18 (1.12-1.26)	rs77478686	0.97/0.95	0.67

5:142162633:A: G	rs7728894 (G)	$1.84 \times 10^{-8}$	0.93 (0.91- 0.96)	rs7728894	0.73/0.75	0.60
5:1285974:A:C	rs7705526 (C)	$2.33 \times 10^{-8}$	0.94 (0.92- 0.96)	rs7705526	0.68/0.68	0.45
19:8786624:A:G	rs2967684 (G)	$2.72 \times 10^{-8}$	0.93 (0.90- 0.95)	rs2967684	0.81/0.76	0.40

280 \*Alleles presented in alphabetical order

281 IndSigSNPs: Independent significant SNP

282

283

284

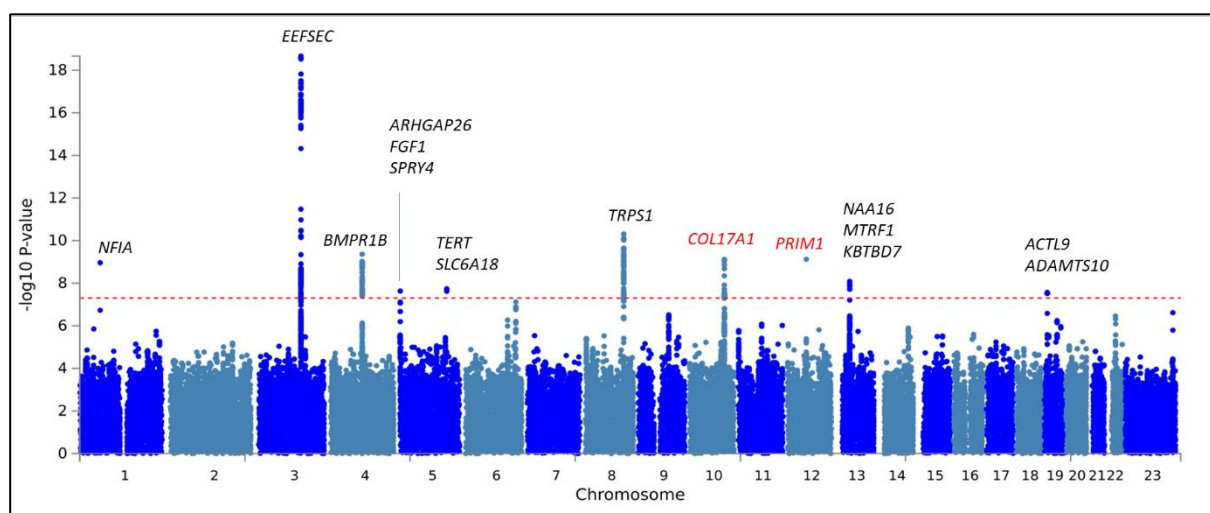
285

286

287

288

289



290 **Figure 1. Manhattan plot for significant genomic risk loci identified for polyps of female**

291 **genital tract.** On the Manhattan plot, the x-axis represents chromosomes, while the y-axis

292 represents  $-\log_{10}(\text{P-values})$  for the association of variants identified in polyps of the female

293 genital tract. The horizontal red dashed line represents the genome-wide significance threshold

294 ( $P < 5 \times 10^{-8}$ ). The prioritised genes for each locus are labelled on the top, and genes associated

295 with exonic missense variants are coloured red.

296 **Functional annotation of associated variants and gene prioritisation**

297 We performed a look-up in the Open Targets Genetics database to evaluate the

298 functional association of identified genomic loci and potentially mapped genes reflected by the

299 V2G score. We additionally highlighted those loci where the GWAS signal includes a missense  
 300 variant using annotation data from FUMA (Table 2). Full details of gene prioritisation together  
 301 with supporting evidence from FUMA eQTL and chromatin interaction mapping, can be found  
 302 in Supplementary Table 1.

303 **Table 2. Summary of gene prioritisation results**

Locus	Lead variant	Highest V2G score	Exonic variants	Biological evidence
3:128118711: A:G	rs1702136	<i>EEFSEC</i> <i>RPN1</i> <i>DNAJB8</i>	NA	<i>EEFSEC</i> potential endometrial cancer susceptibility gene (Kho <i>et al.</i> , 2021)
8:116869477: C:T	rs800578	<i>TRPS1</i>	NA	Potential tumour suppressor candidate in endometrial cancer (Liang <i>et al.</i> , 2012)
4:95731394:A: G	rs2865375	<i>BMPRI1B</i> <i>PDLIM5</i> <i>HPGDS</i>	NA	<i>BMPRI1B</i> is associated with female infertility. <i>BMPRI1B-AS1</i> facilitates endometrial cancer cell proliferation (Lai <i>et al.</i> , 2022). <i>Bmpr1B</i> mutant mice exhibit a failure in endometrial gland formation (Yi <i>et al.</i> , 2001)
12:57146069: G:T	rs2277339	<i>PRIMI</i> <i>STAT6</i> <i>HSD17B6</i>	<i>PRIMI</i>	Cell proliferation, DNA replication
10:105585753: C:T	rs193097753	<i>STN1</i> <i>SH3PXD2A</i> <i>SLK</i>	<i>COL17A1</i>	Cellular migration, cellular differentiation, extracellular matrix organisation

1:61592380:A: G	rs12751005	<i>NFIA</i>	NA	<i>NFIA</i> is involved in cellular growth and proliferation, tumour morphology. The gene is expressed in endometrial tissue (Humaidan <i>et al.</i> , 2012)
13:41869725: A:G	rs77478686	<i>NAA16</i> <i>MTRF1</i> <i>KBTBD7</i>	NA	Differential expression of <i>NAA16</i> is associated with the pathogenesis of endometriosis (She <i>et al.</i> , 2022)
5:142162633: A:G	rs7728894	<i>ARHGAP26</i> <i>FGF1</i> <i>SPRY4</i>	NA	<i>ARHGAP26</i> promotes ovarian cancer cell invasion and migration (Chen <i>et al.</i> , 2019)
5:1285974:A: C	rs7705526	<i>TERT</i> <i>CLPTMIL</i> <i>SLC6A18</i>	NA	<i>TERT-CLPTMIL</i> locus is a known susceptibility region for cancerous processes
19:8786624:A: G	rs2967684	<i>ACTL9</i> <i>NFILZ</i> <i>ADAMTS10</i>	NA	<i>ACTL9</i> locus associated with endometriosis (Rahmioglu <i>et al.</i> , 2023)

304

305 Based on coding variants in our GWAS signals, we were able to prioritise *PRIM1* and *COL17A1*  
306 in loci on chromosomes 12 and 10, respectively. *PRIM1* is a DNA primase involved in the  
307 initiation of DNA replication by synthesising RNA primers for Okazaki fragments during  
308 discontinuous DNA replication (Shiratori *et al.*, 1995). rs2277339 is also a cis-eQTL for *PRIM1*  
309 in endometrial tissue, which further supports *PRIM1* as a candidate gene in this locus.



310 *COL17A1*, on the other hand, is primarily involved in cellular migration, cellular differentiation,  
311 and extracellular matrix organisation (Jones *et al.*, 2020). While these genes have not yet been  
312 directly associated with FGT polyps, their biological functions, particularly those related to  
313 cellular proliferation, highlight their potential involvement in the development of polyps.

314 Several of the prioritised candidate genes had previously been associated with  
315 (endometrial) cancer - *EEFSEC*, *TRPS1*, *TERT/CLPTMIL*, one could be directly linked with  
316 endometrial biology (*BMPR1B*), and for the remaining, their biological significance in FGT  
317 polyps remains unclear.

### 318 **Gene-based associations of female genital tract polyps:**

319 To combine the joint effect of multiple genetic variants and increase the power to detect  
320 associations, we performed a MAGMA gene-based test implemented in FUMA. Eight genes  
321 were identified which passed the recommended threshold for significance ( $P = 2.6 \times 10^{-6}$ ,  
322 Bonferroni correction for association testing of 18,895 protein-coding genes): *EEFSEC*, *TERT*,  
323 *RUVBL1*, *BMPR1B*, *TRPS1*, *COL17A1*, *BETIL*, *WBP4* (Supplementary Figure 1,  
324 Supplementary Table 2). Majority of these associations mirror the genes prioritised in the single  
325 variant analysis, while the *BETIL* locus was not genome-wide significant in the single variant  
326 analysis and is thus novel. Previously, *BETIL* has been associated with endometrial cancer  
327 (Bateman *et al.*, 2017) and uterine fibroids (Cha *et al.*, 2011; Edwards *et al.*, 2013; Liu *et al.*,  
328 2018).

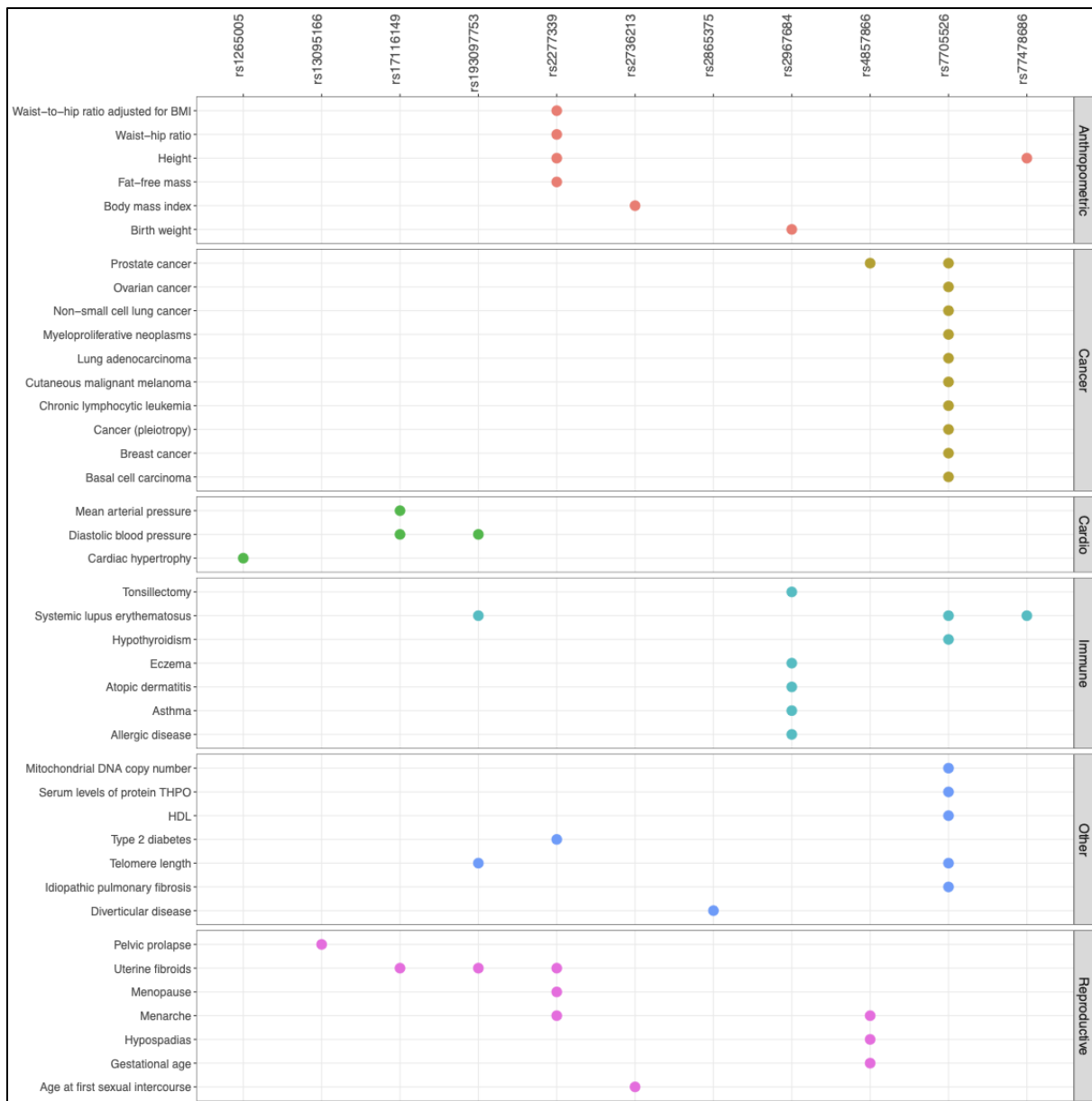
### 329 **GWAS catalogue look-up**

330 We searched the GWAS catalogue for associations between previously published  
331 phenotypic traits and candidate SNPs identified by the FUMA tool to gain additional insight  
332 into their potential biological roles (Fig. 2, Supplementary Table 3). Based on the GWAS  
333 catalogue look-up, the *TERT-CLPTMIL* (rs7705526) locus was clearly associated with cancers,

334 including the development of reproductive cancers such as ovarian cancer and prostate cancer.  
335 Potential involvement in cancerous processes was also observed for the *EEFSEC* locus  
336 (rs4857866), which was associated with prostate cancer and also nominally with endometrial  
337 cancer ( $P < 1 \times 10^{-6}$ ). Several identified signals were related to reproductive traits, including  
338 menarche (rs2277339-*PRIMI*), menopause (rs2277339-*PRIMI*), gestational age (rs4857866-  
339 *EEFSEC*) and uterine fibroids (rs2277339-*PRIMI*, rs193097753-*COL17A1*, rs17116149-  
340 *COL17A1*). The rs193097753 signal was additionally associated with the known risk factors  
341 for the development of FGT polyps, such as type 2 diabetes ( $P < 5 \times 10^{-8}$ ) and waist-to-hip ratio  
342 ( $P < 2 \times 10^{-11}$ ), further substantiating its functional significance.

343

344



345

346 **Figure. 2: GWAS catalogue look-up showing associations between genetic variants**  
 347 **associated with female genital tract polyps and other phenotypes.** The figure highlights the  
 348 genome-wide significant ( $P < 5 \times 10^{-8}$ ) association between FGT polyps genetic risk factors  
 349 (individual significant SNPs) and previously published phenotypic traits and disorders.

350 **Look-up of variants associated with endometrial cancer**

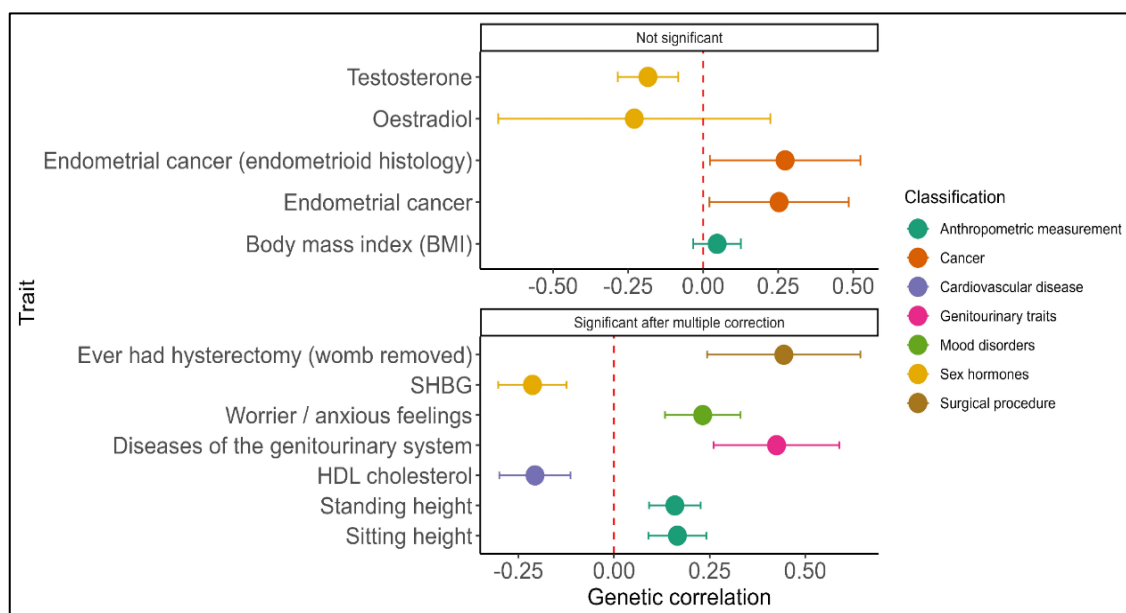
351 Since there is some overlap between FGT polyp genetic risk loci and those known to be  
 352 associated with (endometrial) cancer, we conducted a look-up of variants associated with  
 353 endometrial cancer (O'Mara *et al.*, 2018) in our GWAS data. Of the 19 SNPs queried, seven

354 were nominally significant ( $P < 0.05$ ) (rs9668337-*SSPN*, rs1740828-*SOX4*, rs17601876-  
355 *CYP19A1*, rs882380-*SNX11*, rs3184504-*SH2B3*, rs1129506-*EVI2A*, rs11263761-*HNF1B*) in  
356 the FGT polyp analysis as well (Supplementary Table 4). When querying the 10 lead variants  
357 associated with FGT polyp in the endometrial cancer summary statistics, we observed that three  
358 were nominally significant ( $P < 0.05$ ) in the endometrial cancer and also in the endometrial  
359 cancer with endometrioid histology studies (rs1702136-*EEFSEC*, rs2865375-*BMPR1B*,  
360 rs7705526-*TERT*).

### 361 **Genetic correlation between FGT polyps and other traits**

362 We conducted pairwise genetic correlation ( $r_g$ ) analyses to examine the relationship  
363 between polyps of the FGT and 1,335 different traits obtained from the Complex Traits  
364 Genetics Virtual Lab (CTG-VL, <https://genoma.io/>) and 3 additional endometrial cancer traits  
365 available in the GWAS catalogue (accession codes GCST006464, GCST006465 and  
366 GCST006466). After running the analysis and correcting for multiple testing, we identified a  
367 total of 11 significant ( $FDR < 0.05$ ) genetic correlations. Selected genetic correlations between  
368 polyps and various phenotypic categories such as anthropometrics, cardiovascular diseases,  
369 genitourinary traits, mood disorders, sex hormones and surgical procedures are displayed in  
370 Fig. 3, full results in Supplementary Table 5. As expected, a positive genetic correlation was  
371 observed with genitourinary traits and hysterectomy, emphasising that polyps are commonly  
372 occurring incidental findings in gynaecological disorders. Furthermore, sex hormone-binding  
373 globulin (SHBG) was negatively correlated with the risk of FGT polyps. We were not able to  
374 detect a significant genetic correlation between FGT polyps and estradiol level or BMI, known  
375 risk factors of FGT polyps, in our dataset. FGT polyps showed weak correlation with two of  
376 the three endometrial cancer phenotypes tested (endometrial cancer and endometrial cancer -  
377 endometrioid histology, both  $r_g$  0.25-0.27,  $se$  0.12-0.13), but these associations were only  
378 nominally significant ( $P = 0.03$ ) and did not pass multiple testing correction threshold.

379



380 **Figure 3: Results of genetic correlation analysis between endometrial polyps and other**  
 381 **relevant traits.**

382 The plot displays selected genetic correlations between female genital tract polyps and traits  
 383 from various phenotypic categories such as anthropometrics, cardiovascular traits,  
 384 genitourinary traits, mood disorders, sex hormones and surgical procedures that are colour-  
 385 coded. Significant associations after FDR multiple correction are shown in the lower panel,  
 386 while the upper panel shows selected traits of interest that do not pass the multiple testing  
 387 correction. The centre dot marks the estimated genetic correlation ( $r_g$ ) value, and error bars  
 388 indicate 95% confidence limits. The dotted red line indicates no genetic correlation.

### 389 **Associations between polyps of FGT and other diagnosis codes**

390 Additionally, to evaluate other phenotypes' association with FGT polyps, we performed  
 391 a phenome-wide phenotype association analysis using the EstBB data for disease codes. All  
 392 239 phenotypic associations with a  $P < 2.5 \times 10^{-5}$  (corresponding to a Bonferroni-corrected  
 393 threshold of 0.05/2020) are shown in Supplementary Table 6. The most significant association  
 394 of FGT polyps related to other phenotypes were: i) leiomyoma of uterus (D25), which are

395 uterine fibroids known to often occur together with EPs (Kinay *et al.*, 2016); ii) endometriosis  
396 (N80), which has been correlated with a high prevalence of endometrial polyps; iii) excessive,  
397 frequent and irregular menstruation (N92) which is one of the classical symptoms of EPs, and;  
398 iv) other noninflammatory disorders of uterus except cervix (N85) (Fig. 4). Moreover, women  
399 with a diagnosis for FGT polyps had significantly more diagnoses for benign neoplasms in  
400 several tissues, such as the uterus (D26), ovary (D27), unspecified female genital organs (D28  
401 and D39) and skin (D23) as well as malignant neoplasms of the uterus (C54) and skin (C44).  
402 In our analysis, nasal polyps (J33) and K62 (which includes anal/rectal polyps) were also  
403 nominally significant, which supports the idea of some systemic tissue overgrowth in epithelial  
404 tissues similar to endometrial polyps harbouring glandular and luminal epithelium.

405 For the highlighted associations, we queried the FinnGen Risteys R7 portal to see if we  
406 observe similar associations in the Finnish data. While the analysis in the Estonian data does  
407 not consider which diagnosis in the tested pair comes first, the FinnGen data survival analysis  
408 tests the association in both directions. Uterine leiomyoma, endometriosis, other benign  
409 neoplasm of uterus, benign neoplasm of ovary, malignant neoplasm of uterus, and carcinoma  
410 in situ of endometrium were significantly associated with FGT polyps in the survival analysis  
411 in FinnGen data, (Supplementary Table 7) both if the diagnosis occurred before, or after the  
412 FGT polyp diagnosis.

413 The results of phenotypic associations of FGT polyps are in concordance with the observed  
414 genetic associations and point to the fact that benign and cancerous processes share some  
415 mechanisms (such as cellular proliferation) on a genetic level.

416

417

418



442 with endometrial cancer. PheWAS analysis showed an increased prevalence of endometriosis,  
443 irregular and excessive menstruation, uterine fibroids and neoplasms of the uterus and other  
444 tissues in women with FGT polyps.

445 Even though most EPs are benign, there is still a certain level of risk these polyps may  
446 progress to malignant transformation (Antunes *et al.*, 2007; Lenci *et al.*, 2014). According to a  
447 systematic review and meta-analysis, 2.7%-3.6% of EPs can be considered malignant, with  
448 significant heterogeneity in prevalence among pre- and postmenopausal women (Lee *et al.*,  
449 2010; Sasaki *et al.*, 2018; Uglietti *et al.*, 2019). While we saw no statistically significant genetic  
450 correlation between FGT polyps and endometrial cancer, several of the mapped genetic risk  
451 loci/prioritised genes (rs1702136-*EEFSEC*, rs800578-*TRPS1*, rs2865375-*BMPR1B*, and  
452 rs7705526-*TERT-CLPTMIL*) have previously been associated with cancer either in GWAS or  
453 functional studies (Liang *et al.*, 2012; Wu *et al.*, 2014; Kho *et al.*, 2021; Dos Santos *et al.*,  
454 2022), most likely reflecting the fact that both benign and malignant tissue overgrowths may  
455 utilise the same cellular mechanisms to some extent. This is supported by the results of the  
456 associated diagnoses analysis where women with a diagnosis of FGT polyp also have more  
457 diagnoses related to benign and malignant neoplasms of the uterus, ovary, and skin. While the  
458 associations with reproductive tract neoplasms may arise due to incidental findings during  
459 diagnostic procedures, it cannot be ruled out that some women have a genetic predisposition to  
460 tissue overgrowth that may manifest in either benign or malignant neoplasms in certain tissues.  
461 Apart from a few small-scale reports (Unler *et al.*, 2016), there are no relevant studies  
462 evaluating the prevalence of different neoplasms among women with FGT polyps, therefore,  
463 our observations need to be confirmed and further evaluated in future studies.

464 We further saw significant associations of FGT polyps with uterine leiomyoma and  
465 endometriosis. Though the cellular composition, characteristics, and mechanism of origin of  
466 uterine fibroids and endometrial polyps differ from each other, they can occur concurrently due



467 to common risk factors like advanced age, obesity, and hormonal dysregulation (Kinay *et al.*,  
468 2016; Pavone *et al.*, 2018). Further, increased prevalence of endometriosis among women with  
469 FGT polyps emphasises the reported evidence of a higher prevalence of endometrial polyps in  
470 endometriosis patients (Shen *et al.*, 2011; Zhang *et al.*, 2018; Lin *et al.*, 2020; Peters *et al.*,  
471 2022).

472 Our genetic correlation analysis showed a significant negative correlation between  
473 SHBG levels and the risk of FGT polyps. In females, testosterone and oestrogen levels are  
474 regulated through their binding to SHBG, which helps maintain hormonal balance in the  
475 bloodstream. Lower levels of SHBG can result in elevated levels of free testosterone and free  
476 oestrogen. As EPs are oestrogen dependent, lower SHBG may directly stimulate EP  
477 development. Moreover, the expression of aromatase 450 enzyme in the EPs, converts free  
478 testosterone into oestrogen, further augments the stimulation of polyps' development (Filho *et*  
479 *al.*, 2007). Thus, our study's finding of the negative correlation of SHBG in FGT polyps  
480 concurs with this mechanism. Unexpectedly, our analysis did not reveal a genetic correlation  
481 between oestradiol and EPs. The discrepancy between circulating plasma levels and localized  
482 endometrial levels of oestradiol may underlie the lack of this correlation. In the literature, serum  
483 levels of oestradiol were shown to be significantly lower (five times) than the endometrial  
484 oestradiol concentration in the proliferative phase (Huhtinen *et al.*, 2012). The same trend was  
485 observed in women with abnormal uterine bleeding and hyperplasia (Cortés-Gallegos *et al.*,  
486 1975). Given that abnormal bleeding and hyperplasia are important symptoms of endometrial  
487 polyps, it suggests that localised oestradiol may play a more crucial role in EP development.  
488 Additionally, a study reported no significant difference in circulating oestradiol concentration  
489 among women with and without endometrial polyps (Cortés-Gallegos *et al.*, 1975), which  
490 further supports our finding. This can also explain the lack of correlation of FGT polyps with

491 BMI to some extent, as higher BMI is often associated with increased circulating oestradiol  
492 levels.

493 To understand how genetic variation impacts trait susceptibility, it is important to link  
494 associated genetic variants with specific genes and mechanisms. We employed a diverse range  
495 of data layers to map potential candidate genes for the identified genomic loci. Among others,  
496 *PRIMI* (rs2277339) and *COL17A1* (rs193097753) were prioritised since the GWAS signals  
497 tagged coding variants in these genes. In the literature, neither of these genes has been directly  
498 correlated with FGT development. The *PRIMI* gene is a key regulator of DNA replication  
499 during cellular proliferation process. It has been shown that the *PRIMI* expression was  
500 upregulated in breast tumour tissues compared to healthy tissues, and its inhibition led to  
501 tumour cell growth regression (Lee *et al.*, 2019). Furthermore, the *PRIMI*-induced tumour cell  
502 growth is stimulated by oestrogen through activation of the oestrogen receptor (ER) (Lee *et al.*,  
503 2019). As per one of the postulated hypotheses, polyp development is related to oestrogen  
504 stimulation with predominantly increased ER alpha (ER-alpha) and decreased progesterone  
505 receptors (PRs) A and B in the glandular epithelium. On the other hand, in the stromal cells of  
506 polyps, lower concentrations of ER and PR hinder the decidualization process, preventing them  
507 from shedding off during menstruation (Mittal *et al.*, 1996; Peng *et al.*, 2009; Nijkang *et al.*,  
508 2019). Since the *PRIMI* gene signalling is oestrogen-associated, this can provide important  
509 insights into the probable functional mechanisms underlying the development of polyps. The  
510 *COL17A1* gene is closely associated with epithelial tumour progression and invasiveness (Jones  
511 *et al.*, 2020), suggesting that a shared biological mechanism related to epithelial cell  
512 proliferation could be associated with the development of EPs. However, further functional  
513 assays on *PRIMI* and *COL17A1* genes are needed to confirm these associations.

514 While our study is a large-scale study, it had certain limitations. The cohort from  
515 Estonian and Finnish biobanks encompasses women with all kinds of reproductive polyps, and

516 our phenotype definition did not differentiate between the types/location of polyps to increase  
517 study sample size power. However, considering the overall occurrence of polyps in the general  
518 population, endometrial polyps would be more prevalent than the other types. A complete  
519 understanding of the transcriptome profile of endometrial polyps is still lacking, resulting in  
520 knowledge gaps regarding the interplay between genomic loci and the regulation of gene  
521 expression. By addressing these knowledge gaps, we can gain valuable insights into the genetic  
522 factors and gene expression patterns that contribute to the pathogenesis of endometrial polyps  
523 and potentially identify novel therapeutic targets for their management.

524 In conclusion, the first GWAS meta-analysis of FGT polyps highlights and clarifies the  
525 genetic mechanisms shared between EP development (tissue overgrowth) and cancerous  
526 processes. Furthermore, an analysis of associated diagnoses showed that women diagnosed with  
527 EPs have an increased prevalence of other diagnoses, such as endometriosis, uterine fibroids  
528 and both benign and malignant neoplasms, which could have implications for patient  
529 management and counselling.

530 **Supplementary data:** Supplementary data are available at Human reproduction online.

531 **Data availability:** FinnGen cohort level summary statistics can be accessed as described here:  
532 [https://www.finnngen.fi/en/access\\_results](https://www.finnngen.fi/en/access_results). Protocol for accessing the Estonian Biobank data is  
533 described here: <https://genomics.ut.ee/en/content/estonian-biobank>. GWAS meta-analysis  
534 summary statistics will be made available via the GWAS Catalogue (data upload pending).

535 **Author's roles:** Authors A.D.S.P., N.P.G, A.S., M.P and T.L. participated in the conception of  
536 the study, data analysis, and interpretation of the data, and writing the manuscript. V.R., J.D.,  
537 and R.M. contributed to data analysis and interpretation. T.L., M.S., M.P., V.R., J.D. and A.S.  
538 critically reviewed and provided feedback on each version of the manuscript and all the authors  
539 approved the final version. The EstBB Research Team provided the EstBB genotype and  
540 phenotype data.

541 **Acknowledgements:** We want to acknowledge the participants of the Estonian Biobank and  
542 the participants and investigators of the FinnGen study. We also acknowledge the Estonian  
543 Biobank Research team members Andres Metspalu, Tonu Esko, Mari Nelis, Georgi Hudjashov,  
544 and Lili Milani.

545 **Funding:** N.P.G. was supported by MATER Marie Sklodowska-Curie which received funding  
546 from the European Union's Horizon 2020 research and innovation programme under grant  
547 agreement No. 813707. T.L. was supported by the Estonian Research Council grant PSG776.  
548 This study was funded by European Union through the European Regional Development Fund  
549 Project No. 2014-2020.4.01.15-0012 GENTRANSMED. Computations were performed in the  
550 High-Performance Computing Center of University of Tartu. The study was also supported by  
551 the Estonian Research Council (grant no. PRG1076 and MOBJD1056) and Horizon 2020  
552 innovation grant (ERIN, grant no. EU952516).

553 **Conflict of interest:** All the authors declared no conflict of interest.

554

555 **References:**

556

557 Altaner S, Gucer F, Tokatli F, Guresci S, Ozdemir C, Puyan FO, Kutlu K. Expression of Bcl-2  
558 and Ki-67 in tamoxifen-associated endometrial polyps: comparison with  
559 postmenopausal polyps. *Onkologie* 2006;**29**:376–380.

560 Annan JJ, Aquilina J, Ball E. The management of endometrial polyps in the 21st century. *The*  
561 *Obstetrician & Gynaecologist* 2012;**14**:33–38.

562 Antunes A, Costa-Paiva L, Arthuso M, Costa JV, Pinto-Neto AM. Endometrial polyps in pre-  
563 and postmenopausal women: factors associated with malignancy. *Maturitas*  
564 2007;**57**:415–421.

- 565 Azevedo JM da R de, Azevedo LM da R de, Freitas F, Wender MCO. Endometrial polyps:  
566 when to resect? *Arch Gynecol Obstet* 2016;**293**:639–643.
- 567 Banas T, Pitynski K, Mikos M, Cielecka-Kuszyk J. Endometrial Polyps and Benign  
568 Endometrial Hyperplasia Have Increased Prevalence of DNA Fragmentation Factors 40  
569 and 45 (DFF40 and DFF45) Together With the Antiapoptotic B-Cell Lymphoma (Bcl-  
570 2) Protein Compared With Normal Human Endometria. *Int J Gynecol Pathol*  
571 2018;**37**:431–440.
- 572 Bateman NW, Dubil EA, Wang G, Hood BL, Oliver JM, Litzi TA, Gist GD, Mitchell DA,  
573 Blanton B, Phippen NT, *et al.* Race-specific molecular alterations correlate with  
574 differential outcomes for black and white endometrioid endometrial cancer patients.  
575 *Cancer* 2017;**123**:4004–4012.
- 576 Bettocchi S, Achilarré M, Ceci O, Luigi S. Fertility-Enhancing Hysteroscopic Surgery. *Semin*  
577 *Reprod Med* 2011;**29**:075–082.
- 578 Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park  
579 J, Hitz BC, Weng S, *et al.* Annotation of functional variation in personal genomes using  
580 RegulomeDB. *Genome Res* 2012;**22**:1790–1797.
- 581 Cha P-C, Takahashi A, Hosono N, Low S-K, Kamatani N, Kubo M, Nakamura Y. A genome-  
582 wide association study identifies three loci associated with susceptibility to uterine  
583 fibroids. *Nat Genet* 2011;**43**:447–450.
- 584 Chen X, Chen S, Li Y, Gao Y, Huang S, Li H, Zhu Y. SMURF1-mediated ubiquitination of  
585 ARHGAP26 promotes ovarian cancer cell invasion and migration. *Exp Mol Med*  
586 2019;**51**:1–12.
- 587 Dal Cin P, Vanni R, Marras S, Moerman P, Kools P, Andria M, Valdes E, Deprest J, Van de  
588 Ven W, Van den Berghe H. Four cytogenetic subgroups can be identified in endometrial  
589 polyps. *Cancer Res* 1995;**55**:1565–1568.

- 590 Doria PLS, Moscovitz T, Tcherniakovsky M, Fernandes CE, Pompei LM, Wajman M,  
591 Nimwegen AV, Haimovich S. Association of IGF-1 CA(n) and IGFBP3 rs2854746  
592 Polymorphisms with Endometrial Polyp Risk. *Biomed Res Int* 2018;**2018**:8704346.
- 593 Dos Santos GA, Viana NI, Pimenta R, Camargo JA de, Guimaraes VR, Romão P, Candido P,  
594 Ghazarian V, Reis ST, Leite KRM, *et al.* Pan-cancer analysis reveals that CTC1-STN1-  
595 TEN1 (CST) complex may have a key position in oncology. *Cancer Genet* 2022;**262**–  
596 **263**:80–90.
- 597 Dreisler E, Stampe Sorensen S, Ibsen PH, Lose G. Prevalence of endometrial polyps and  
598 abnormal uterine bleeding in a Danish population aged 20-74 years. *Ultrasound Obstet*  
599 *Gynecol* 2009;**33**:102–108.
- 600 Edwards TL, Hartmann KE, Velez Edwards DR. Variants in BET1L and TNRC6B associate  
601 with increasing fibroid volume and fibroid type among European Americans. *Hum*  
602 *Genet* 2013;**132**:1361–1369.
- 603 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human  
604 genome. *Nature* 2012;**489**:57–74.
- 605 Fatemi HM, Kasius JC, Timmermans A, Disseldorp J van, Fauser BC, Devroey P, Broekmans  
606 FJ. Prevalence of unsuspected uterine cavity abnormalities diagnosed by office  
607 hysteroscopy prior to in vitro fertilization. *Hum Reprod* 2010;**25**:1959–1965.
- 608 Filho AMB de B, Barbosa IC, Maia H, Genes CC, Coutinho EM. Effects of subdermal implants  
609 of estradiol and testosterone on the endometrium of postmenopausal women. *Gynecol*  
610 *Endocrinol* 2007;**23**:511–517.
- 611 Fung JN, Mortlock S, Girling JE, Holdsworth-Carson SJ, Teh WT, Zhu Z, Lukowski SW,  
612 McKinnon BD, McRae A, Yang J, *et al.* Genetic regulation of disease risk and  
613 endometrial gene expression highlights potential target genes for endometriosis and  
614 polycystic ovarian syndrome. *Sci Rep* 2018;**8**:11424.

- 615 Ghossaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, Fumis L, Miranda  
616 A, Carvalho-Silva D, Buniello A, *et al.* Open Targets Genetics: systematic identification  
617 of trait-associated genes using large-scale genetics and functional genomics. *Nucleic*  
618 *Acids Res* 2021;**49**:D1311–D1320.
- 619 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;**45**:580–  
620 585.
- 621 Hinckley MD, Milki AA. 1000 office-based hysteroscopies prior to in vitro fertilization:  
622 feasibility and findings. *JSLs* 2004;**8**:103–107.
- 623 Humaidan P, Van Vaerenbergh I, Bourgain C, Alsbjerg B, Blockeel C, Schuit F, Van Lommel  
624 L, Devroey P, Fatemi H. Endometrial gene expression in the early luteal phase is  
625 impacted by mode of triggering final oocyte maturation in recFSH stimulated and GnRH  
626 antagonist co-treated IVF cycles. *Hum Reprod* 2012;**27**:3259–3272.
- 627 Jones VA, Patel PM, Gibson FT, Cordova A, Amber KT. The Role of Collagen XVII in Cancer:  
628 Squamous Cell Carcinoma and Beyond. *Front Oncol* 2020;**10**:352.
- 629 Karayalcin R, Ozcan S, Moraloglu O, Ozyer S, Mollamahmutoglu L, Batioglu S. Results of  
630 2500 office-based diagnostic hysteroscopies before IVF. *Reprod Biomed Online*  
631 2010;**20**:689–693.
- 632 Kho PF, Wang X, Cuéllar-Partida G, Dörk T, Goode EL, Lambrechts D, Scott RJ, Spurdle AB,  
633 O’Mara TA, Glubb DM. Multi-tissue transcriptome-wide association study identifies  
634 eight candidate genes and tissue-specific gene expression underlying endometrial  
635 cancer susceptibility. *Commun Biol* 2021;**4**:1211.
- 636 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for  
637 estimating the relative pathogenicity of human genetic variants. *Nat Genet*  
638 2014;**46**:310–315.

- 639 Kınay T, Öztürk Başarır Z, Fırtına Tuncer S, Akpınar F, Kayıkçıoğlu F, Koç S. Prevalence of  
640 endometrial polyps coexisting with uterine fibroids and associated factors. *Turk J Obstet*  
641 *Gynecol* 2016;**13**:31–36.
- 642 Koel M, Võsa U, Jõeloo M, Läll K, Gualdo NP, Laivuori H, Lemmelä S; Estonian Biobank  
643 Research Team; FinnGen; Daly M, *et al.* GWAS meta-analyses clarify the genetics of  
644 cervical phenotypes and inform risk stratification for cervical cancer. *Hum Mol Genet.*  
645 2023;**32**:2103-2116.
- 646 Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, Reeve MP, Laivuori  
647 H, Aavikko M, Kaunisto MA, *et al.* FinnGen provides genetic insights from a well-  
648 phenotyped isolated population. *Nature* 2023;**613**:508–518.
- 649 Lai T, Qiu H, Si L, Zhen Y, Chu D, Guo R. Long noncoding RNA BMPR1B-AS1 facilitates  
650 endometrial cancer cell proliferation and metastasis by sponging miR-7-2-3p to  
651 modulate the DCLK1/Akt/NF-κB pathway. *Cell Cycle* 2022;**21**:1599–1618.
- 652 Laisk T, Lepamets M, Koel M, Abner E, Estonian Biobank Research Team, Mägi R. Genome-  
653 wide association study identifies five risk loci for pernicious anemia. *Nat Commun*  
654 2021;**12**:3761.
- 655 Lee SC, Kaunitz AM, Sanchez-Ramos L, Rhatigan RM. The oncogenic potential of endometrial  
656 polyps: a systematic review and meta-analysis. *Obstet Gynecol* 2010;**116**:1197–1205.
- 657 Lee W-H, Chen L-C, Lee C-J, Huang C-C, Ho Y-S, Yang P-S, Ho C-T, Chang H-L, Lin I-H,  
658 Chang H-W, *et al.* DNA primase polypeptide 1 (PRIM1) involves in estrogen-induced  
659 breast cancer formation through activation of the G2/M cell cycle checkpoint. *Int J*  
660 *Cancer* 2019;**144**:615–630.
- 661 Leeuw CA de, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of  
662 GWAS data. *PLoS Comput Biol* 2015;**11**:e1004219.



- 663 Lenci MA, Nascimento VAL do, Grandini AB, Fahmy WM, Depes D de B, Baracat FF, Lopes  
664 RGC. Premalignant and malignant lesions in endometrial polyps in patients undergoing  
665 hysteroscopic polypectomy. *Einstein (Sao Paulo)* 2014;**12**:16–21.
- 666 Liang H, Cheung LWT, Li J, Ju Z, Yu S, Stemke-Hale K, Dogruluk T, Lu Y, Liu X, Gu C, *et*  
667 *al.* Whole-exome sequencing combined with functional genomics reveals novel  
668 candidate driver cancer genes in endometrial cancer. *Genome Res* 2012;**22**:2120–2129.
- 669 Lin S, Xie X, Guo Y, Zhang H, Liu C, Yi J, Su Y, Deng Q, Zhu W. Clinical characteristics and  
670 pregnancy outcomes of infertile patients with endometriosis and endometrial polyps: A  
671 retrospective cohort study. *Taiwan J Obstet Gynecol* 2020;**59**:916–921.
- 672 Liu B, Wang T, Jiang J, Li M, Ma W, Wu H, Zhou Q. Association of BET1L and TNRC6B  
673 with uterine leiomyoma risk and its relevant clinical features in Han Chinese population.  
674 *Sci Rep* 2018;**8**:7401.
- 675 Mittal K, Schwartz L, Goswami S, Demopoulos R. Estrogen and progesterone receptor  
676 expression in endometrial polyps. *Int J Gynecol Pathol* 1996;**15**:345–348.
- 677 Nichols CA, Gibson WJ, Brown MS, Kosmicki JA, Busanovich JP, Wei H, Urbanski LM,  
678 Curimjee N, Berger AC, Gao GF, *et al.* Loss of heterozygosity of essential genes  
679 represents a widespread class of potential cancer vulnerabilities. *Nat Commun*  
680 2020;**11**:2517.
- 681 Nijkang NP, Anderson L, Markham R, Manconi F. Endometrial polyps: Pathogenesis, sequelae  
682 and treatment. *SAGE Open Med* 2019;**7**:2050312119848247.
- 683 O’Mara TA, Glubb DM, Amant F, Annibali D, Ashton K, Attia J, Auer PL, Beckmann MW,  
684 Black A, Bolla MK, *et al.* Identification of nine new susceptibility loci for endometrial  
685 cancer. *Nat Commun* 2018;**9**:3166.
- 686 Pal L, Niklaus AL, Kim M, Pollack S, Santoro N. Heterogeneity in endometrial expression of  
687 aromatase in polyp-bearing uteri. *Hum Reprod* 2008;**23**:80–84.

- 688 Pavone D, Clemenza S, Sorbi F, Fambrini M, Petraglia F. Epidemiology and Risk Factors of  
689 Uterine Fibroids. *Best Pract Res Clin Obstet Gynaecol* 2018;**46**:3–11.
- 690 Peng X, Li T, Xia E, Xia C, Liu Y, Yu D. A comparison of oestrogen receptor and progesterone  
691 receptor expression in endometrial polyps and endometrium of premenopausal women.  
692 *J Obstet Gynaecol* 2009;**29**:340–346.
- 693 Peters M, Mikeltadze I, Karro H, Saare M, Estonian Biobank Research Team, Salumets A,  
694 Mägi R, Laisk T. Endometriosis and irritable bowel syndrome: similarities and  
695 differences in the spectrum of comorbidities. *Hum Reprod* 2022;**37**:2186–2196.
- 696 Pujol-Gualdo N, Läll K, Lepamets M, Estonian Biobank Research Team, Rossi H-R, Arffman  
697 RK, Piltonen TT, Mägi R, Laisk T. Advancing our understanding of genetic risk factors  
698 and potential personalized strategies for pelvic organ prolapse. *Nat Commun*  
699 2022;**13**:3584.
- 700 Rahmioglu N, Mortlock S, Ghiasi M, Møller PL, Stefansdottir L, Galarneau G, Turman C,  
701 Danning R, Law MH, Sapkota Y, *et al.* The genetic basis of endometriosis and  
702 comorbidity with other pain and inflammatory conditions. *Nat Genet* 2023;**55**:423–436.
- 703 Raz N, Feinmesser L, Moore O, Haimovich S. Endometrial polyps: diagnosis and treatment  
704 options - a review of literature. *Minim Invasive Ther Allied Technol* 2021;**30**:278–287.
- 705 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A,  
706 Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, *et al.* Integrative analysis of 111  
707 reference human epigenomes. *Nature* 2015;**518**:317–330.
- 708 Sahoo SS, Aguilar M, Xu Y, Lucas E, Miller V, Chen H, Zheng W, Cuevas IC, Li H-D, Hitrys  
709 D, *et al.* Endometrial polyps are non-neoplastic but harbor epithelial mutations in  
710 endometrial cancer drivers at low allelic frequencies. *Mod Pathol* 2022;**35**:1702–1712.

- 711 Sasaki LMP, Andrade KRC, Figueiredo ACMG, Wanderley M da S, Pereira MG. Factors  
712 Associated with Malignancy in Hysteroscopically Resected Endometrial Polyps: A  
713 Systematic Review and Meta-Analysis. *J Minim Invasive Gynecol* 2018;**25**:777–785.
- 714 Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, Li Y, Lin S, Lin Y, Barr CL, *et al.* A  
715 Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the  
716 Human Genome. *Cell Rep* 2016;**17**:2042–2059.
- 717 She J, Su D, Diao R, Wang L. A Joint Model of Random Forest and Artificial Neural Network  
718 for the Diagnosis of Endometriosis. *Front Genet* 2022;**13**:848116.
- 719 Shen L, Wang Q, Huang W, Wang Q, Yuan Q, Huang Y, Lei H. High prevalence of endometrial  
720 polyps in endometriosis-associated infertility. *Fertil Steril* 2011;**95**:2722-2724.e1.
- 721 Shiratori A, Okumura K, Nogami M, Taguchi H, Onozaki T, Inoue T, Ando T, Shibata T, Izumi  
722 M, Miyazawa H. Assignment of the 49-kDa (PRIM1) and 58-kDa (PRIM2A and  
723 PRIM2B) subunit genes of the human DNA primase to chromosome bands 1q44 and  
724 6p11.1-p12. *Genomics* 1995;**28**:350–353.
- 725 Takeda T, Banno K, Kobayashi Y, Adachi M, Yanokura M, Tominaga E, Kosaki K, Aoki D.  
726 Mutations of RAS genes in endometrial polyps. *Oncol Rep* 2019;**42**:2303–2308.
- 727 Tanos V, Berry KE, Seikkula J, Abi Raad E, Stavroulis A, Sleiman Z, Campo R, Gordts S. The  
728 management of polyps in female reproductive organs. *Int J Surg* 2017;**43**:7–16.
- 729 Uglietti A, Buggio L, Farella M, Chiaffarino F, Dridi D, Vercellini P, Parazzini F. The risk of  
730 malignancy in uterine polyps: A systematic review and meta-analysis. *Eur J Obstet*  
731 *Gynecol Reprod Biol* 2019;**237**:48–56.
- 732 Unler GK, Gokturk HS, Toprak E, Erinanc OH, Korkmaz H. Does the Presence of Endometrial  
733 Polyp Predict Colorectal Polyp? *Am J Med Sci* 2016;**351**:129–132.

- 734 Vieira MDC, Vitagliano A, Rossette MC, Neto LC de A, Gallo A, Sardo ADS. Endometrial  
735 Polyps: Update Overview on Etiology, Diagnosis, Natural History and Treatment.  
736 *CEOG* 2022;**49**:232. IMR Press.
- 737 Vitale SG, Haimovich S, Laganà AS, Alonso L, Di Spiezio Sardo A, Carugno J, From the  
738 Global Community of Hysteroscopy Guidelines Committee. Endometrial polyps. An  
739 evidence-based diagnosis and management guide. *Eur J Obstet Gynecol Reprod Biol*  
740 2021;**260**:70–77.
- 741 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from  
742 high-throughput sequencing data. *Nucleic Acids Res* 2010;**38**:e164.
- 743 Watanabe K, Taskesen E, Bochoven A van, Posthuma D. Functional mapping and annotation  
744 of genetic associations with FUMA. *Nat Commun* 2017;**8**:1826.
- 745 Wu L, Wang Y, Liu Y, Yu S, Xie H, Shi X, Qin S, Ma F, Tan TZ, Thiery JP, *et al.* A central  
746 role for TRPS1 in the control of cell cycle and cancer development. *Oncotarget*  
747 2014;**5**:7677–7690.
- 748 Yi SE, LaPolt PS, Yoon BS, Chen JY, Lu JK, Lyons KM. The type I BMP receptor Bmpr1B is  
749 essential for female reproductive function. *Proc Natl Acad Sci U S A* 2001;**98**:7994–  
750 7999.
- 751 Zhang Y-N, Zhang Y-S, Yu Q, Guo Z-Z, Ma J-L, Yan L. Higher Prevalence of Endometrial  
752 Polyps in Infertile Patients with Endometriosis. *Gynecol Obstet Invest* 2018;**83**:558–  
753 563.
- 754
- 755
- 756
- 757

758 **Supplementary materials:**

759 **Supplementary figure 1: Genome-wide gene association analysis of our GWAS dataset**  
760 **using ‘Multi-marker Analysis of GenoMic Annotation (MAGMA) wherein input SNPs**  
761 **were mapped to 18895 protein-coding genes.** On the Manhattan plot, the y-axis represents  
762  $-\log_{10}(\text{P-values})$  for the association of variants identified in polyps of the female genital tract.  
763 The horizontal red dashed line represents the genome-wide significance threshold ( $P < 5 \times$   
764  $10^{-8}$ ) and prioritised genes for each locus are shown. Genes associated with exonic missense  
765 variants are denoted in red text.

766

767 **Supplementary Tables:**

768 ST 1 Genes mapped by expression quantitative trait loci (eQTL) and Chromatin interactions  
769 by Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA  
770 GWAS)

771 ST 2 Results of the Multi-marker Analysis of GenoMic Annotation (MAGMA) gene-based  
772 analysis

773 ST 3 Results of the GWAS catalogue lookup

774 ST 4 Look-up of the 10 lead variants in endometrial cancer summary statistics  
775 (<https://pubmed.ncbi.nlm.nih.gov/31040137/>)

776 ST 5 Genetic correlation between female genital tract polyps and other traits

777 ST 6 Phenotypes associated with Polyps of the female genital tract

778 ST 7 Results of survival analysis lookup in FinnGen data

779

780

781