## 1 *Article*

# 2 **Evidence of a causal relationship between genetic tendency to**  3 **gain muscle mass and uterine leiomyomata**

- 4 Eeva Sliz<sup>1,2</sup>, Jaakko Tyrmi<sup>1,2</sup>, Nilufer Rahmioglu<sup>6,7</sup>, Krina T. Zondervan<sup>6,7</sup>, Christian M.
- 5 Becker<sup>6</sup>, FinnGen, Outi Uimari<sup>3,4,5</sup>\*, Johannes Kettunen<sup>1,2</sup>\*
- <sup>1</sup> 6 Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Finland
- <sup>2</sup> 7 Biocenter Oulu, Oulu, Finland
- <sup>3</sup> 8 Department of Obstetrics and Gynecology, Oulu University Hospital, Oulu, Finland
- <sup>4</sup> PEDEGO Research Unit, University of Oulu and Oulu University Hospital, Oulu, Finland
- <sup>5</sup> Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu,
- 11 Finland
- <sup>6</sup> Oxford Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive
- 13 Health, University of Oxford, UK
- <sup>7</sup> Wellcome Centre for Human Genetics, University of Oxford, UK
- 15 \*Equal contribution of the last authors
- 16 Address for correspondence:
- 17 Eeva Sliz
- 18 Center For Life Course Health Research
- 19 Faculty of Medicine
- 20 Aapistie 5A
- 21 P.O.Box 5000, 90014 University of Oulu
- 22 Oulu, Finland
- 23 phone: +358 50 305 4865
- 24 e-mail: eeva.sliz@oulu.fi
- 25 Word count Introductory paragraph: 150
- 26 Word count Manuscript main body: 1627
- 27 Figures: 4
- 28 Tables: 1
- 29 Supplementary figures: 63
- 30 Supplementary tables: 9
- 31 References: 68
- 32 Key words: Uterine leiomyoma, GWAS, smooth muscle, muscle mass
- 33 *Summary statistics will be made available through the NHGRI-EBI GWAS Catalog with*
- 34 *GCSTxxxxxxx* upon publication.

**Uterine leiomyomata (UL) are the most common benign tumours of the female genital tract with an estimated lifetime incidence of up to 70%<sup>1</sup>** 36 **. To date, 7 genome-wide association studies (GWASs) have identified 35 loci predisposing to UL. To improve the understanding of the underlying genetic pathways, we conducted the largest genetic association study of UL to date in 426,558 European women from FinnGen and a previously published UL meta-GWAS2** 40 **. We identified 36 novel and replicated 31 previously reported loci. Annotations of the potential candidate genes suggest involvement of smooth muscle cell (SMC) differentiation and/or proliferation-regulating genes in modulating UL risk. Our results further advocate that genetic predisposition to increased fat-free mass may be causally related to UL risk, underscoring the involvement of altered muscle tissue biology in UL pathophysiology. Overall, our findings provide novel insights into genetic risk factors related to UL, which may aid in developing novel treatment strategies.** 

48 UL are hormone-driven benign neoplasms of the uterus composed mostly of SMC and 49 fibroblasts with a profound component of extracellular matrix (ECM). UL are present in 50 single or multiple numbers, with size ranging from millimeters to 20 cm or more in 51 diameter<sup>3</sup>. In 25-50% of women with ULs, the enlarged and deformed uterus can cause 52 symptoms such as heavy or prolonged menstrual bleeding resulting in anemia, reduced 53 fertility, and pregnancy complications<sup>4</sup>. Major risk factors for UL are family history, African 54 ancestry, increasing age up to menopause, and nulliparity<sup>5</sup>. Also, metabolic factors, including 55 hypertension and increased body mass index (BMI), have been reported in association with 56 . UL susceptibility<sup>6</sup>. Familial aggregation, twin studies, and disparity in prevalence between 57 different ethnic groups suggest that genetic factors modulate UL risk<sup>7-10</sup>.

58 Until recently, the focus in the genetics of UL has been on somatic rearrangements, and key 59 driver variations in *MED12*, *FH*, *HMGA2* and *COL4A6-COL4A5* have been reported<sup>11</sup>. To 60 add understanding to the heritable genetic underpinnings, we conducted two sets of fixed-61 effect, inverse-variance weighted meta-analyses with data from FinnGen (18,060 cases and 62 105,519 female controls) and 1) a previously published UL meta-GWAS<sup>2</sup> totaling 53,534 63 cases and 373,024 female controls in 'META-1' and 2) the UK Biobank totaling 33,244 64 cases and 311,271 controls in 'META-2'. META-1 was restricted to publicly available  $10,000$  variants from the previous meta-GWAS<sup>2</sup>, wheras the META-2 included 10,693,588 66 variants across the genome. In META-1, we identified 63 loci associating with UL at 67  $p \le 5x10^{-8}$  (Figure 1, Table S1); of these, 32 were novel (Table 1). In META-2, we identified 68 51 loci out of which 4 had not been associated with UL risk in prior GWASs or in META-1 69 (Figure 1, Table 1, Table S2). Regional association plots of the novel loci are presented in 70 Figures S1-S36. There was no evidence of inflation in the test statistics (Figure S37). LD 71 score (LDSC) regression-derived SNP-based heritability was estimated to be 0.099 (standard 72 error=0.01) on the liability scale. Characterization of the genome-wide results of META-2 73 suggested that the key UL associated variants were mostly intronic (Figure 2A). Genes 74 associated with UL risk (Figure 2B) were significantly enriched on pathways related to gonad 75 development, positive regulation of growth, and cellular senescence (Figure 2C) and they 76 were mostly expressed in uterus, cervix, and fallopian tube (Figure 2D). In two of the novel 77 loci, we found evidence that altered gene expression mediates UL association (Figures S38- 78 S40).

Previous GWAS findings<sup>2,12-17</sup> have indicated that genetic factors altering pathways involved 80 in estrogen signalling, Wnt signalling, transforming growth factor (TGF)-β signalling, and 81 cell cycle progression are associated with UL risk $18$ . The novel loci identified in this study 82 further underscore the involvement of pathways regulating SMC proliferation in modulation 83 of UL risk. Many of these pathways are interrelated: for example, both estrogen and

84 progesterone increase the secretion of Wnt ligands from myometrial or leiomyoma SMC, 85 which promotes cell proliferation and tumorigenesis via activation of β-catenin<sup>19</sup>. Steroid 86 hormones also influence the production of ECM via signalling through the TGF-β family of 87 ligands and receptors that are highly expressed in multiple fibrotic conditions<sup>20-26</sup> and 88 contribute to the fibrotic phenotype seen in UL<sup>24,25,27-29</sup>. We identified multiple loci with 89 potential candidate genes functioning in one or more of these pathways, and, in the following, 90 we describe some of our key findings with the focus on loci involved in regulation of SMC 91 proliferation.

92 A central finding is the novel association at 17p12 harbouring myocardin (*MYOCD*; Figure 93 S27). Myocardin is a transcription factor expressed in smooth muscle tissues, including most 94 prominently arteries and colon, but also uterus (Figure S41), and it is required for SMC 95 differentiation<sup>30</sup>. The expression of myocardin has been shown to be downregulated in UL 96 tissue compared with normal myometrium<sup>31</sup>. Also, it has been proposed that the loss of 97 myocardin function may be a key factor in driving SMC proliferation in  $UL<sup>31</sup>$ ; however, there 98 are no reported GWAS associations implicating myocardin until now. The lead variants near 99 *MYOCD* are intergenic variants with no strong evidence of altered regulatory consequences 100 (Table S3) and, thus, a possible association-driving mechanism remains inconclusive. We 101 identified another myocardin-related novel UL risk association near 'myocardin-induced 102 smooth muscle cell lncRNA, inducer of differentiation' (*MYOSLID*), a transcriptional target 103 of myocardin<sup>32</sup> (Figure S6). RegulomeDB<sup>33</sup> annotation provided robust evidence that the 104 association lead variant is a regulatory variant (Table S4) altering binding of multiple 105 transcription factors (Table S5), including Fos proto-oncogene (FOS) that has been shown to 106 be downregulated in UL<sup>34</sup>. To add yet another example of a myocardin-related UL risk loci, a well-established UL risk association at  $22q13.1^{2,12,16,35,36}$  locates near 'myocardin related well-established UL risk association at  $22q13.1^{2,12,16,35,36}$  locates near 'myocardin related transcription factor A' (*MRTFA*; also known as *MKL1*), a gene interacting with myocardin<sup>37</sup>. 109 Also, expression of *MRTFA/MKL1* has been shown to be downregulated in UL-related 110 deletions at  $22a^{38}$ . deletions at  $22q^{38}$ .

111 Others have suggested that loss of myocardin function may account for the differentiation 112 defects of human leiomyosarcoma cells during malignant transformation<sup>39</sup> downregulation 113 of myocardin resulted in lower expression of cyclin dependent kinase inhibitor 1A 114 (*CDKN1A*; also known as p21), a mediator of cell cycle G1 phase arrest, which facilitated cell cycle progression. We identified a novel UL association at 6p21.2 near *CDKN1A* (Figure 115 cell cycle progression. We identified a novel UL association at 6p21.2 near *CDKN1A* (Figure 116 S13); the association lead variants locate on intergenic region with possible regulatory 117 consequences (Table S6). Previous evidence suggests that *CDKN1A* is among the genes, the 118 expression of which correlates with UL size<sup>40</sup>. Another novel UL association at  $20q13.31$ 119 harbours RNA binding motif protein 38 (*RBM38*; Figure S29) that specifically binds to and 120 regulates the stability of *CDKN1A* transcripts<sup>37</sup>. In this locus, the UL risk-increasing rs13039273-C associates with lower *RBM38* expression in ovary  $(p=8.7 \times 10^{-6})$  (Table S7, 122 Figure S42: eOTL in uterus  $p=2.2 \times 10^{-3}$ ). Relevant to UL, a variant in the same locus Figure S42; eQTL in uterus  $p=2.2\times10^{-3}$ ). Relevant to UL, a variant in the same locus 123 (rs760640162) has been previously associated with 'pregnancy examination and test'  $(124 \text{ (p=5.97x10}^{-15})$  in the UK Biobank<sup>41,42</sup>. Interestingly, estrogen receptor (ER)α has been shown 125 to inhibit expression of myocardin<sup>31</sup>, suggesting that the ability of myocardin-CDKN1A-126 signalling to inhibit cell cycle progression may be impaired in tissues enriched with ERα. 127 Taken together our findings and previous evidence, it seems highly probable that 128 downregulation of myocardin-CDKN1A signalling increases the risk of UL.

129 We used an automated LDSC regression pipeline<sup>43</sup> to evaluate the genetic correlations ( $r_g$ ) of 130 UL with other traits (Figure 2; Table S8). The most robustly correlated traits were seen for 131 other gynecological traits, such as bilateral oophorectomy  $(r_g=0.74, p=1.05 \times 10^{-25})$ , hysterectomy  $(r_g=0.62, p=3.24x10^{-19})$ , age at menarche  $(r_g=-0.18, p=1.40x10^{-9})$ , excessive

frequent and irregular menstruation  $(r<sub>s</sub>=0.51, p=5.66x10^{-7})$ , having had menopause  $(r<sub>s</sub>=-0.33, p=5.66x10^{-7})$ 134  $p=2.31x10^{-5}$ ), and usage of hormonal replacement therapy (r<sub>g</sub>=0.20, p=5.54x10<sup>-5</sup>). Strong 135 correlations were also seen between UL and anthropometric and metabolic measures, including whole body fat-free mass  $(r_g=0.11, p=4.21x10^{-5})$ , high blood pressure  $(r_g=0.15, p=4.58x10^{-6})$ , and serum triglyceride level  $(r_g=0.16, p=4.80x10^{-5})$ . Genetic susceptibility for 137  $p=4.58\times10^{-6}$ ), and serum triglyceride level (r<sub>g</sub>=0.16, *p*=4.80x10<sup>-5</sup>). Genetic susceptibility for 138 UL was also correlated with genetic susceptibility for multiple psychiatric or mood-related UL was also correlated with genetic susceptibility for multiple psychiatric or mood-related traits, including depressive symptoms  $(r_g=0.20, p=3.31x10^{-5})$  or neuroticism score  $(r_g=0.14, p=3.31x10^{-5})$ 140  $p=3.88\times10^{-5}$ .

141 To estimate the causal relationships between UL and the key traits showing significant  $r_{\sigma}$ 142 with UL, we applied two-sample Mendelian randomization. Our results suggest that genetic 143 tendency to accumulate fat-free mass may be causally linked to higher risk of UL  $(144)$  ( $p=2.37x10^{-3}$ ; Figure 3; Table S9). In line with this, the higher impedance of whole body (i.e., 145 a bioelectrical measure used for estimating body composition; lower muscle mass leads to higher impedance) was causally associated with a lower risk of UL ( $p=1.05\times10^{-3}$ ). When 147 considering the null causal effect of whole-body fat mass on UL risk ( $p=0.46$ ), it seems 148 apparent that the nominally significant  $(p=8.40x10^{-3})$  causal effect of BMI on UL in the 149 present study and the previously reported associations of BMI with UL risk<sup> $44-46$ </sup> arise from the 150 increased lean body mass rather than fat mass. Although we did not observe horizontal 151 pleiotropy, the causal estimates were heterogenic for many key traits (Table S9), suggesting 152 that all the underlying genetic pathways may not function in a consistent manner – thus, these 153 results need to be interpreted with caution. However, in the leave-one-out sensitivity analyses 154 all causal estimates were consistently positive (higher fat-free mass was causally associated 155 with higher risk of UL) or negative (higher impedance was causally associated with lower 156 risk of UL) suggesting that there is no single variant driving the causal associations (Figures 157 S50-S63). Regarding most of the gynecological traits, the causal direction with UL was not 158 clear (Table S9), suggesting that the shared molecular factors predispose to multiple 159 gynecological complications simultaneously instead of one being causal to another. Of note, 160 we replicated the previously reported finding of genetic predisposition to UL being causal to 161 excessive menstrual bleeding<sup>2</sup>  $(p=1.07 \times 10^{-4})$ .

162 The discovered causal relationship between genetic tendency to accumulate fat-free mass and 163 UL risk provides a novel perspective on UL-related pathophysiology. UL are considered 164 estrogen-dependent, and UL have higher ERα expression compared with normal uterine 165 myometria<sup>47,48</sup>. ERs are expressed in a variety of tissues, including all musculoskeletal 166 tissues<sup>49</sup>. In females, muscle mass and strength are closely coupled with estrogen status: girls 167 begin to gain muscle mass after the onset of puberty<sup>50</sup> whereas in older age during 168 perimenopausal and postmenopausal periods, muscle strength declines considerably  $51,52$ . If 169 estrogen enhances muscle growth<sup>53</sup>, the observed causal relationship between fat-free mass 170 and UL risk could arise secondary to high estrogen contributing to the muscle growth. Many 171 of the newly identified candidate genes, however, directly regulate cell proliferation – thus, it 172 remains possible that the estrogen-rich environment, due to sexual maturity, may trigger 173 excess SMC growth resulting in ULs in women who are genetically susceptible to build up 174 muscle.

175 Currently, the only essentially curative treatment for UL is hysterectomy, which underscores 176 the high demand for the development of alternative effective therapies. Our findings double 177 the number of known UL risk loci and, thus, the herein presented results provide several 178 potential targets for translational research to develop pharmacologic interventions for UL. 179 Therapies targeted at myocardin-CDKN1A signalling or, considering the causal evidence, 180 other factors regulating muscle growth may hold the greatest potential.

#### 81**Table 1. Novel UL risk loci.**

The table reports novel distinct loci (more than 1Mb apart) containing at least one variant identified to be associated with UL at  $p<5x10^{-8}$  in a meta-analysis of

53,534 UL cases and 373,024 female controls from FinnGen and a previously published meta-GWAS of UL<sup>2</sup> (META-1) and in a meta-analysis of 33,244 UL

84cases and 311,271 female controls from FinnGen and the UK Biobank (META-2).



5 Chr, chromosome; Pos, position (build 38); EA, effect allele; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval; P, p-value; HetPVal, p-value for heterogeneity; FIN enr, Finnish enrichment (calculated as FIN AF / NFEE AF in the Genome Aggregation Database (gnomAD), where FIN AF is the Finnish allele frequency and NFEE AF is the non-Finnish-non-Estonian European allele frequency). 'Nearest gene(s)' reports the gene closest to the association lead variant. 'Candidate gene(s)' indicates the biologically most relevant gene within 1Mb window around the association lead variant.

89

#### 190 **Figure 1. A combined Manhattan plot of UL associations in META-1 and META-2.**

191 We conducted UL GWAS in FinnGen and, subsequently, two sets of meta-analyses with data 192 from a previously published UL meta-analysis<sup>2</sup> and the UK Biobank, adding up to 53,534 UL 193 cases and 373,024 female controls in META-1 (top) and 33,244 UL cases and 311,271 194 female controls in META-2 (bottom), respectively. Blue color denotes novel UL risk loci 195 identified in META-1, and red color indicates novel loci identified in META-2 that were not 196 associated previously with UL risk in prior GWASs or META-1. Black and grey colours 197 indicate odd and even chromosome numbers, respectively. The red dashed lines correspond 198 to the threshold for genome-wide significance  $(p<5x10<sup>-8</sup>)$ .



# 199 **Figure 2. Variant summary and gene set-based results using summary statistics from**  200 **META-2.**

201 A) The proportions of 'independent genome-wide significant variants' and 'variants in LD 202 with independent significant variants' having corresponding functional annotation. Bars are 203 colored according to  $-log_2$ (enrichment) relative to all variants in the reference panel. B) A 204 Manhattan plot of the gene-based test computed by  $MAGMA<sup>54</sup>$ . The input variants were Manhattan plot of the gene-based test computed by  $MAGMA<sup>54</sup>$ . The input variants were 205 mapped to 19,920 protein coding genes and, thus, significance was considered at  $p < 2.51 \times 10^{-6}$ 206 (0.05/19,920); only 87 genes with p-values below this threshold are plotted. Blue and red 207 colours indicate odd and even chromosome numbers, respectively. C) MAGMA<sup>54</sup> gene-set 208 enrichment analysis was performed for curated gene sets and GO terms available at MsigDB<sup>55</sup>. The plot shows the results for significantly enriched pathways ( $p_{\text{Bonferroni}}$ <0.05).<br>210 D) Results of MAGMA<sup>54</sup> tissue expression analysis testing a positive relationship between D) Results of  $MAGMA<sup>54</sup>$  tissue expression analysis testing a positive relationship between 211 tissue-specific gene expression profiles and disease-gene associations. Tissue-specific gene-212 expression data were from GTEx  $(v8)^{56}$ . All data plotted in Figure 2 A-D were produced 213  $\mu$ sing FUMA<sup>57</sup>.



#### 214 **Figure 3. Genetic correlations (rg) of UL with other traits.**

215 The analyses were completed using an automated LD score regression pipeline available at

216 http://ldsc.broadinstitute.org/<sup>43</sup>. Statistical significance was considered at  $p < 6x10^{-5}$  (0.05/830,

217 where 830 is the number of traits tested). The results of traits that showed significant  $r_g$  with

218 UL are plotted (more results are provided in Table S8). Traits are named as reported in the

219 LD Hub output.



#### 220 **Figure 4. Causal relationships between UL and body composition measures.**

221 We estimated causal relationships between UL and the key traits showing significant genetic 222 correlation with UL using two-sample Mendelian randomization implemented in <sup>223</sup> 'TwoSampleMR' R library<sup>58,59</sup>. Bi-directional causal estimates were obtained by using 224 summary statistics provided by the MRC Integrative Epidemiology Unit (IEU) GWAS 225 database and UL GWAS results obtained in FinnGen in the present study to extract genetic 226 instruments for other, mostly UKBB-based, traits and UL, respectively. LD pruning was 227 completed using European population reference, threshold of  $r^2=0.001$ , and clumping 228 window of 10kb. Statistical significance was considered at  $p < 3x10^{-3}$  (0.05/19, where 19 is the 229 number of traits tested; black diamonds). The plot shows the causal estimates obtained using 230 inverse variance weighted method for key body composition measures (more results are 231 provided in Table S9; the scatter plots for the body composition measures can be found in 232 Figures S43-S49, and the results of the leave-one-out sensitivity analyses are showin in



# 234 **Methods**

## 235 *Study populations*

236 **FinnGen** (www.finngen.fi/en) is a public-private partnership project launched in 2017 with 237 an aim to improve human health through genetic research. The project utilizes genome 238 information from a nationwide network of Finnish biobanks that are linked with digital health 239 records from national hospital discharge (available from 1968), death (1969-), cancer (1953- 240 ), and medication reimbursement (1995-) registries using the unique national personal identification codes. Ultimately, the data resource will cover roughly 10% of the Finnish identification codes. Ultimately, the data resource will cover roughly 10% of the Finnish 242 population. We studied data from 123,579 female participants (18,060 UL cases and 105,519 243 female controls) from FinnGen Preparatory Phase Data Freeze 5. UL cases were required to 244 have an entry of ICD-10: D25, ICD-9: 218, or ICD-8: 21899, and participants who had no 245 records of these entries were deemed as controls. The Coordinating Ethics Committee of the 246 Hospital District of Helsinki and Uusimaa (HUS) approved the FinnGen study protocol Nr 247 HUS/990/2017.

248 The FinnGen study is approved by Finnish Institute for Health and Welfare (THL), approval 249 number THL/2031/6.02.00/2017, amendments THL/1101/5.05.00/2017, 250 THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, 251 THL/1721/5.05.00/2019, Digital and population data service agency VRK43431/2017-3, 252 VRK/6909/2018-3, VRK/4415/2019-3 the Social Insurance Institution (KELA) KELA 253 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, and Statistics 254 Finland TK-53-1041-17.

255 The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data 256 Freeze 5 include: THL Biobank BB2017\_55, BB2017\_111, BB2018\_19, BB\_2018\_34, 257 BB\_2018\_67, BB2018\_71, BB2019\_7, BB2019\_8, BB2019\_26, Finnish Red Cross Blood 258 Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, 259 Biobank Borealis of Northern Finland\_2017\_1013, Biobank of Eastern Finland 1186/2018, 260 Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017, and 261 Terveystalo Biobank STB 2018001.

262 **FibroGENE** is consortium of conventional, population-based and direct-to-consumer cohorts 263 that was assembled to replicate and identify UL risk variants. In the study by Gallagher *et al*., 264 they studied data from 35,474 UL cases and 267,505 female controls including participants 265 from four population-based cohorts (Women's Genome Health Study, WGHS; Northern 266 Finland Birth Cohort, NFBC; QIMR Berghofer Medical Research Institute, QIMR; the UK 267 Biobank, UKBB) and one direct-to-consumer cohort (23andMe) Detailed descriptions of 268 cohorts and sample selections are available in Supplementary Methods<sup>2</sup>.

269 **The UK Biobank** is a large national and international health resource following the health 270 and wellbeing of 500,000 male and female volunteer participants, enrolled at ages from 40 to 271  $\,$  69  $\,$  60. The study began in 2006 with the aim to follow the participants for at least 30 years 272 thereafter. Information has been collected from participants during recruitment using 273 questionnaires on socioeconomic status, lifestyle, family history and medical history. 274 Participants have also been followed up for cause-specific morbidity and mortality through 275 linkage to disease registries, death registries, and hospital admission records. For this study, 276 altogether 220,936 women of European ancestry were considered. Based on both hospital-277 linked medical records and self-report (interview with research nurse), women with a history 278 of UL were selected as cases  $(n=15,184)$ , while controls  $(n=205,752)$  had no previous history 279 of UL. Informed consent was obtained from all participants. The UKBB project is approved 280 by the North West Multi-Centre Research Ethics Committee.

#### 281 *Genotyping, imputation, and quality control*

282 In FinnGen, genotyping of the samples was performed using Illumina and Affymetrix arrays 283 (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA). Sample 284 quality control (QC) was performed to exclude individuals with high genotype missingness 285 (>5%), ambiguous gender, excess heterozygosity (±4SD) and non-Finnish ancestry. 286 Regarding variant QC, all variants with low Hardy-Weinberg equilibrium (HWE) p-value 287 (<1e-6), high missingness (>2%) and minor allele count (MAC)<3 were excluded. Chip 288 genotyped samples were pre-phased with Eagle 2.3.5 with the number of conditioning 289 haplotypes set to 20,000. Genotype imputation was carried out by using the Finnish 290 population-specific SISu v3 reference panel with Beagle 4.1 (version 08Jun17.d8b) as<br>291 described in the following protocol: dx.doi.org/10.17504/protocols.io.nmndc5e. In post-291 described in the following protocol: dx.doi.org/10.17504/protocols.io.nmndc5e. In post-292 imputation QC, variants with imputation INFO<0.6 were excluded.

293 In UK Biobank, genotyping of samples was performed either on the Affymetrix UK BiLEVE 294 or Affymetrix UK Biobank Axiom array. The UK Biobank genotype data passed centralised

- quality control and was imputed by the UKBB team up to the HRC reference panel<sup>61</sup>.
- 296 *GWAS*

297 The UL GWAS in FinnGen was completed using the Scalable and Accurate Implementation 298 of Generalized (SAIGE) software<sup>62</sup>. The association models were adjusted for age, sex, the 299 first 10 genetic principal components, and genotyping batch, and only variants with minimum 300 allele count of 5 were included in the analysis.

301 Given the extensive relatedness in the UK Biobank data, a linear mixed model (LMM) was 302 utilised for association testing that accounts for population structure and model the related 303 individuals, including genotyping array, age and BMI as covariates as implemented in Bolt-304 LMM V2.3.

305 *Meta-analyses* 

306 Two sets of fixed-effect, inverse-variance weighted meta-analyses (implemented in  $307$  METAL<sup>63</sup>) were performed: the results obtained in FinnGen were meta-analyzed with 1) the 308 top 10,000 most significant variants associating with UL in a published GWAS<sup>2</sup> (META-1) 309 and 2) UL GWAS results obtained in the UKBB. Statistical significance was considered at 310 the standard genome-wide significance level  $(p<5x10^{-8})$ . Genomic inflation factor was 311 estimated using an automated LD score (LDSC) regression pipeline<sup>43</sup>.

312 *Characterization of association signals* 

313 We used a web-based platform  $FUMA<sup>57</sup>$  to perform functional annotations of the GWAS 314 results: we completed functional gene mapping and gene-based association and enrichment 315 tests using the genome-wide UL associations from META-2 (FinnGen and UKBB) and 316 predefined lead variants as reported in Table S2. FUMA identifies variants showing genome-317 wide significant association  $(p<5x10^{-8})$  with the study trait and, among the significant 318 variants, identifies variants in low LD  $(r^2<0.6)$  as 'independent significant variants' and 319 further identifies variants in LD  $(r^2>0.6)$  with the 'independent significant variants';  $320$  ANNOVAR<sup>64</sup> annotations are performed for all these variants to obtain information on the 321 functional consequences of the key variants.  $MAGMA<sup>54</sup>$ , also implemented in FUMA, was 322 used to perform gene-based association testing and gene-set enrichment analyses: gene-based 323 p-values were computed for protein-coding genes by mapping variants to genes and 324 subsequent enrichment analyses were performed for the significant genes using 4728 curated 325 gene sets and 6166 GO terms as reported in MsigDB<sup>55</sup>.

326 To further characterize the potential UL candidate gene(s) according to biological relevance, 327 we annotated all genes within ±1Mb window from the association lead variant. We explored 328 information provided by GenBank<sup>37</sup> and UniProt<sup>65</sup> to determine functions of the genes. To 329 complement the information available in these databases, a broad literature search was 330 performed to identify previous work published regarding to the genes of interest.

331 In each locus, we explored the associations of the lead variant with gene expression levels in 332 the Genotype-Tissue Expression (GTEx) Portal; GTEx was accessed during 02/10/2021- 333 02/11/2021. To further test if altered gene expression mediates UL risk associations, we used 334 a method proposed by Zhu et al.<sup>66</sup> as implemented in Complex Traits Genetics Virtual Lab (CTG-VL)<sup>67</sup>; we performed these tests using genome-wide UL results from META-2 and tissue-specific gene expression data (GTEx, V7) for uterus and whole blood. We further used tissue-specific gene expression data (GTEx, V7) for uterus and whole blood. We further used Regulome $DB^{33}$  to discover regulatory elements overlapping with the intergenic variants in the novel loci showing genome-wide significant association with UL risk. Here, significant 338 the novel loci showing genome-wide significant association with UL risk. Here, significant 339 results for the novel loci are reported.

340 *SNP-based heritability and genetic correlations* 

341 The SNP-based heritability ( $h_{SNP}^2$ ) of UL was estimated using LDSC regression implemented 342 in LDSC software<sup>68</sup>. Population prevalence of 0.30 (as in  $2$ ) and sample prevalence of 0.11 343 were used to estimate  $h^2_{SNP}$  on liability scale. We further applied LDSC to estimate genetic 344 correlation ( $r_g$ ) of UL with other traits using LD Hub<sup>43</sup>, a web interface for performing 345 automated LDSC regression. As recommended, we excluded the major histocompatilibity 346 complex region from the analyses<sup>43</sup>. We tested r<sub>g</sub> between UL and all 830 traits available in 347 the LD Hub database and, thus, statistical significance was considered at  $p < 6x10^{-5}$ 348 (0.05/830).

# 349 *Mendelian randomization*

350 To test for causal inferences between UL and the key traits showing significant  $r_g$  with UL, 351 we performed bi-directional two-sample Mendelian randomization. These analyses were completed using 'TwoSampleMR' R library<sup>58</sup> (https://mrcieu.github.io/TwoSampleMR/). To<br>353 avoid possible bias from overlapping samples, we extracted genetic instruments for UL from avoid possible bias from overlapping samples, we extracted genetic instruments for UL from 354 the GWAS results obtained in FinnGen, and for other traits from the GWAS database 355 provided by the MRC Integrative Epidemiology Unit (IEU) available 356 https://gwas.mrcieu.ac.uk/ and integrated in TwoSampleMR. LD pruning was completed 357 using European population reference, threshold of  $r^2$ =0.001, and clumping window of 10kb, 358 as set as default in 'clump\_data' function; the numbers of SNPs available for the analyses are 359 listed in Table S9. Inverse variance weighted (IVW) method was considered as a primary 360 analysis, and MR Egger estimates were derived in sensitivity analyses: the estimates were 361 required to be to a matching direction with the IVW-estimates and Egger intercepts were 362 evaluated to assess horizontal pleiotropy. Cochran's Q statistics were derived using 363 'mr\_heterogeneity' function to test for heterogeneity. To screen for highly influential variants 364 that could drive the association for example due to horizontal pleiotropy, we performed 365 leave-one-out analyses using 'mr\_leaveoneout' function. Statistical significance for the 366 causal effects was considered at  $p < 3x10^{-3}$  (0.05/19) to correct for 19 traits tested.

# 367 **Acknowledgements and funding**

368 The work was supported through The Sigrid Juselius Foundation (JK) and funds from the 369 Academy of Finland [grant numbers 297338 and 307247](JK), Novo Nordisk Foundation 370 [grant number NNF17OC0026062] (JK), and The Finnish Medical Association (OU).

371 The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and 372 UH 4386/31/2016) and eleven industry partners (AbbVie Inc, AstraZeneca UK Ltd, Biogen 373 MA Inc, Celgene Corporation, Celgene International II Sàrl, Genentech Inc, Merck Sharp & 374 Dohme Corp, Pfizer Inc., GlaxoSmithKline, Sanofi, Maze Therapeutics Inc., Janssen Biotech 374 Dohme Corp, Pfizer Inc., GlaxoSmithKline, Sanofi, Maze Therapeutics Inc., Janssen Biotech 375 Inc). Following biobanks are acknowledged for delivering biobank samples to Finngen: 376 Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki 377 Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland 378 (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-379 English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-380 US/Research\_and\_development/Finnish\_Clinical\_Biobank\_Tampere), Biobank of Eastern 381 Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-382 FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank 383 (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank 384 (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish 385 Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi).

386 This research has been conducted using data from UK Biobank, a major biomedical database, 387 (http://www.ukbiobank.ac.uk/) under project ID 9637 (Gallagher et al., 2019).

388 The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the

389 Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI,

390 NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were

391 obtained from the GTEx Portal on 02/10/2021-02/11/2021.

# 392 **Author contributions**

393 OU conceptualized the study. ES, NR, KTZ, CMB, and JK contributed to the analysis plan. 394 ES, JT and NR analyzed data and generated results. ES, JT, NR, KTZ, and OU interpreted 395 the results. ES and OU wrote the original manuscript. JK supervised the study, obtained 396 funding and provided additional study resources. All authors contributed to revising the 397 content of the manuscript and approved the final version.

# 398 **Competing interests**

399 KTZ: Competing financial interests: Scientific collaborations (grant funding) with Bayer AG, 400 Roche Diagnostics Inc, MDNA Life Sciences, and Evotec. Competing non-financial 401 interests: Board memberships of the World Endometriosis Society, World Endometriosis 402 Research Foundation, and research advisory committee member of Wellbeing of Women 403 UK.

404 CMB: Competing financial interests: Scientific collaborations (grant funding) with Bayer 405 AG, Roche Diagnostics Inc, MDNA Life Sciences, and Evotec. Scientific board Myovant; 406 IDDM Member ObsEva. Competing non-financial interests: Chair ESHRE Endometriosis 407 Guideline Development Group.

#### 408 **References**

- 409 1 Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D. & Schectman, J. M. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am. J. incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am. J. 411 *Obstet. Gynecol.* **188**, 100-107 (2003).
- 412 2 Gallagher, C. S. *et al.* Genome-wide association and epidemiological analyses reveal 413 common genetic origins between uterine leiomyomata and endometriosis. *Nat Commun* **10**, 414 4857, doi:10.1038/s41467-019-12536-4 (2019).
- 415 3 Abdul Ghaffar, N. A., Ismail, M. P., Nik Mahmood, N. M., Daud, K. & Abu Dzarr, G. A.<br>416 Huge uterine fibroid in a postmenopausal woman associated with polycythaemia: a case 416 Huge uterine fibroid in a postmenopausal woman associated with polycythaemia: a case report. *Maturitas* **60**, 177-179, doi:10.1016/j.maturitas.2008.03.013 (2008). 417 report. *Maturitas* **60**, 177-179, doi:10.1016/j.maturitas.2008.03.013 (2008).
- 418 4 Stewart, E. A. *et al.* Uterine fibroids. *Nat Rev Dis Primers* **2**, 16043, 419 doi:10.1038/nrdp.2016.43 (2016).
- 420 5 Stewart, E. A., Cookson, C. L., Gandolfo, R. A. & Schulze-Rath, R. Epidemiology of uterine<br>421 fibroids: a systematic review. *BJOG* 124, 1501-1512, doi:10.1111/1471-0528.14640 (2017). 421 fibroids: a systematic review. *BJOG* **124**, 1501-1512, doi:10.1111/1471-0528.14640 (2017).
- 422 6 Pavone, D., Clemenza, S., Sorbi, F., Fambrini, M. & Petraglia, F. Epidemiology and Risk<br>423 Factors of Uterine Fibroids. Best Pract. Res. Clin. Obstet. Gynaecol. 46. 3-11. 423 Factors of Uterine Fibroids. *Best Pract. Res. Clin. Obstet. Gynaecol.* **46**, 3-11, 424 doi:10.1016/j.bpobgyn.2017.09.004 (2018).
- 425 7 Van Voorhis, B. J., Romitti, P. A. & Jones, M. P. Family history as a risk factor for 426 development of uterine leiomyomas. Results of a pilot study. *J. Reprod. Med.* **47**, 663-669 427 (2002).
- 428 8 Luoto, R. *et al.* Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort 429 study. *Maturitas* **37**, 15-26 (2000).
- 430 9 Treloar, S. A., Martin, N. G., Dennerstein, L., Raphael, B. & Heath, A. C. Pathways to 431 hysterectomy: insights from longitudinal twin research. *Am. J. Obstet. Gynecol.* **167**, 82-88, 432 doi:10.1016/s0002-9378(11)91632-9 (1992).
- 433 10 Wise, L. A. *et al.* African ancestry and genetic risk for uterine leiomyomata. *Am. J.*  434 *Epidemiol.* **176**, 1159-1168, doi:10.1093/aje/kws276 (2012).
- 435 11 Mehine, M., Makinen, N., Heinonen, H. R., Aaltonen, L. A. & Vahteristo, P. Genomics of uterine leiomyomas: insights from high-throughput sequencing. *Fertil. Steril.* 102. 621-629. 436 uterine leiomyomas: insights from high-throughput sequencing. *Fertil. Steril.* **102**, 621-629, 437 doi:10.1016/j.fertnstert.2014.06.050 (2014).
- 438 12 Cha, P. C. *et al.* A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nat. Genet.* **43**, 447-450, doi:10.1038/ng.805 (2011). 439 susceptibility to uterine fibroids. *Nat. Genet.* **43**, 447-450, doi:10.1038/ng.805 (2011).
- 440 13 Eggert, S. L. *et al.* Genome-wide linkage and association analyses implicate FASN in 441 predisposition to Uterine Leiomyomata. *Am. J. Hum. Genet.* **91**, 621-628, 442 doi:10.1016/j.ajhg.2012.08.009 (2012).
- 443 14 Hellwege, J. N. *et al.* A multi-stage genome-wide association study of uterine fibroids in 444 African Americans. *Hum. Genet.* **136**, 1363-1373, doi:10.1007/s00439-017-1836-1 (2017).
- 445 15 Valimaki, N. *et al.* Genetic predisposition to uterine leiomyoma is determined by loci for 446 genitourinary development and genome stability. *Elife* **7**, doi:10.7554/eLife.37110 (2018).
- 447 16 Rafnar, T. *et al.* Variants associating with uterine leiomyoma highlight genetic background<br>448 shared by various cancers and hormone-related traits. Nat Commun 9. 3636. 448 shared by various cancers and hormone-related traits. *Nat Commun* **9**, 3636, 449 doi:10.1038/s41467-018-05428-6 (2018).
- 450 17 Sakai, K. *et al.* Identification of a novel uterine leiomyoma GWAS locus in a Japanese 451 population. *Sci. Rep.* **10**, 1197, doi:10.1038/s41598-020-58066-8 (2020).
- 452 18 Sepulveda, P. V., Bush, E. D. & Baar, K. Pharmacology of manipulating lean body mass.<br>453 Clin. Exp. Pharmacol. Physiol. 42, 1-13, doi:10.1111/1440-1681.12320 (2015). 453 *Clin. Exp. Pharmacol. Physiol.* **42**, 1-13, doi:10.1111/1440-1681.12320 (2015).

- 454 19 Ono, M. *et al.* Paracrine activation of WNT/beta-catenin pathway in uterine leiomyoma stem<br>455 cells promotes tumor growth. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17053-17058, 455 cells promotes tumor growth. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17053-17058, 456 doi:10.1073/pnas.1313650110 (2013).
- 457 20 Willert, K. & Nusse, R. Wnt proteins. *Cold Spring Harb. Perspect. Biol.* **4**, a007864, 458 doi:10.1101/cshperspect.a007864 (2012).
- 459 21 Borahay, M. A. *et al.* Estrogen Receptors and Signaling in Fibroids: Role in Pathobiology and 460 Therapeutic Implications. *Reprod. Sci.* **24**, 1235-1244, doi:10.1177/1933719116678686 461 (2017).
- 462 22 Massaad-Massade, L. *et al.* HMGA1 enhances the transcriptional activity and binding of the 463 estrogen receptor to its responsive element. *Biochemistry* **41**, 2760-2768, 464 doi:10.1021/bi011455j (2002).
- 465 23 Dhamad, A. E., Zhou, Z., Zhou, J. & Du, Y. Systematic Proteomic Identification of the Heat 466 Shock Proteins (Hsp) that Interact with Estrogen Receptor Alpha (ERalpha) and Biochemical 467 Characterization of the ERalpha-Hsp70 Interaction. *PLoS One* **11**, e0160312, 468 doi:10.1371/journal.pone.0160312 (2016).
- 469 24 Prossnitz, E. R. *et al.* Estrogen signaling through the transmembrane G protein-coupled 470 receptor GPR30. *Annu. Rev. Physiol.* **70**, 165-190, 471 doi:10.1146/annurev.physiol.70.113006.100518 (2008).
- 472 25 Ciarmela, P. *et al.* Growth factors and myometrium: biological effects in uterine fibroid and 473 possible clinical implications. *Hum. Reprod. Update* **17**, 772-790, 474 doi:10.1093/humupd/dmr031 (2011).
- 475 26 Peng, L. *et al.* Expression of insulin-like growth factors (IGFs) and IGF signaling: molecular 476 complexity in uterine leiomyomas. *Fertil. Steril.* **91**, 2664-2675, 477 doi:10.1016/j.fertnstert.2007.10.083 (2009).
- 478 27 Perrot, V. & Rechler, M. M. The coactivator p300 directly acetylates the forkhead<br>479 transcription factor Foxo1 and stimulates Foxo1-induced transcription. *Mol. Endocrinol*. 19, 479 transcription factor Foxo1 and stimulates Foxo1-induced transcription. *Mol. Endocrinol.* **19**, 2283-2298, doi:10.1210/me.2004-0292 (2005).
- 481 28 Guo, S. *et al.* Phosphorylation of serine 256 by protein kinase B disrupts transactivation by 482 FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 483 promoter activity through a conserved insulin response sequence. *J. Biol. Chem.* **274**, 17184- 484 17192, doi:10.1074/jbc.274.24.17184 (1999).
- 485 29 Lewis, T. D. *et al.* Ulipristal acetate decreases active TGF-beta3 and its canonical signaling in 486 uterine leiomyoma via two novel mechanisms. *Fertil. Steril.* **111**, 806-815 e801, 487 doi:10.1016/j.fertnstert.2018.12.026 (2019).
- 488 30 Chen, J., Kitchen, C. M., Streb, J. W. & Miano, J. M. Myocardin: a component of a molecular 489 switch for smooth muscle differentiation. *J. Mol. Cell. Cardiol.* **34**, 1345-1356, 490 doi:10.1006/jmcc.2002.2086 (2002).
- 491 31 Liao, X. H. *et al.* ERalpha inhibited myocardin-induced differentiation in uterine fibroids. 492 *Exp. Cell Res.* **350**, 73-82, doi:10.1016/j.yexcr.2016.11.007 (2017).
- 493 32 Zhao, J. *et al.* MYOSLID Is a Novel Serum Response Factor-Dependent Long Noncoding 494 RNA That Amplifies the Vascular Smooth Muscle Differentiation Program. *Arterioscler.*  495 *Thromb. Vasc. Biol.* **36**, 2088-2099, doi:10.1161/ATVBAHA.116.307879 (2016).
- 496 33 Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using 497 RegulomeDB. *Genome Res.* **22**, 1790-1797, doi:10.1101/gr.137323.112 (2012).
- 498 34 Raimundo, N., Vanharanta, S., Aaltonen, L. A., Hovatta, I. & Suomalainen, A. Downregulation of SRF-FOS-JUNB pathway in fumarate hydratase deficiency and in uterine 500 leiomyomas. *Oncogene* **28**, 1261-1273, doi:10.1038/onc.2008.472 (2009).

- 501 35 Edwards, T. L., Hartmann, K. E. & Velez Edwards, D. R. Variants in BET1L and TNRC6B<br>502 associate with increasing fibroid volume and fibroid type among European Americans. *Hum*. associate with increasing fibroid volume and fibroid type among European Americans. *Hum.* 503 *Genet.* **132**, 1361-1369, doi:10.1007/s00439-013-1340-1 (2013).
- 504 36 Masuda, T. *et al.* GWAS of five gynecologic diseases and cross-trait analysis in Japanese.<br>505 *Eur. J. Hum. Genet.* **28**, 95-107, doi:10.1038/s41431-019-0495-1 (2020). 505 *Eur. J. Hum. Genet.* **28**, 95-107, doi:10.1038/s41431-019-0495-1 (2020).
- 506 37 Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic*  507 *Acids Res.* **44**, D67-72, doi:10.1093/nar/gkv1276 (2016).
- 508 38 Mehine, M. *et al.* Integrated data analysis reveals uterine leiomyoma subtypes with distinct 509 driver pathways and biomarkers. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1315-1320, 510 doi:10.1073/pnas.1518752113 (2016).
- 511 39 Kimura, Y., Morita, T., Hayashi, K., Miki, T. & Sobue, K. Myocardin functions as an 512 effective inducer of growth arrest and differentiation in human uterine leiomyosarcoma cells. 513 *Cancer Res.* **70**, 501-511, doi:10.1158/0008-5472.CAN-09-1469 (2010).
- 514 40 Markowski, D. N. *et al.* HMGA2 and the p19Arf-TP53-CDKN1A axis: a delicate balance in 515 the growth of uterine leiomyomas. *Genes Chromosomes Cancer* **49**, 661-668, 516 doi:10.1002/gcc.20777 (2010).
- 517 41 Staley, J. R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. 518 *Bioinformatics* **32**, 3207-3209, doi:10.1093/bioinformatics/btw373 (2016).
- 519 42 Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-520 phenotype associations. *Bioinformatics* **35**, 4851-4853, doi:10.1093/bioinformatics/btz469 521 (2019).
- 522 43 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score<br>523 regression that maximizes the potential of summary level GWAS data for SNP heritability 523 regression that maximizes the potential of summary level GWAS data for SNP heritability<br>524 and genetic correlation analysis. *Bioinformatics* 33. 272-279. 524 and genetic correlation analysis. *Bioinformatics* **33**, 272-279, 525 doi:10.1093/bioinformatics/btw613 (2017).
- 526 44 Marshall, L. M. *et al.* Risk of uterine leiomyomata among premenopausal women in relation 527 to body size and cigarette smoking. *Epidemiology* **9**, 511-517 (1998).
- 528 45 Wise, L. A. *et al.* Influence of body size and body fat distribution on risk of uterine leiomyomata in U.S. black women. *Epidemiology* **16**, 346-354 (2005). 529 leiomyomata in U.S. black women. *Epidemiology* **16**, 346-354 (2005).
- 530 46 Lee, J. E. *et al.* Weight change and risk of uterine leiomyomas: Korea Nurses' Health Study.<br>531 Curr. Med. Res. Opin. **34**, 1913-1919, doi:10.1080/03007995.2018.1462783 (2018). 531 *Curr. Med. Res. Opin.* **34**, 1913-1919, doi:10.1080/03007995.2018.1462783 (2018).
- 532 47 Bakas, P. *et al.* Estrogen receptor alpha and beta in uterine fibroids: a basis for altered 533 estrogen responsiveness. *Fertil. Steril.* **90**, 1878-1885, doi:10.1016/j.fertnstert.2007.09.019  $(2008).$
- 535 48 Valladares, F. *et al.* Characterization of estrogen receptors alpha and beta in uterine 536 leiomyoma cells. *Fertil. Steril.* **86**, 1736-1743, doi:10.1016/j.fertnstert.2006.05.047 (2006).
- 537 49 Sipila, S. & Poutamo, J. Muscle performance, sex hormones and training in peri-menopausal 538 and post-menopausal women. *Scand. J. Med. Sci. Sports* **13**, 19-25, doi:10.1034/j.1600- 0838.2003.20210.x (2003).
- 540 50 Wood, C. L., Lane, L. C. & Cheetham, T. Puberty: Normal physiology (brief overview). *Best*  541 *Pract. Res. Clin. Endocrinol. Metab.* **33**, 101265, doi:10.1016/j.beem.2019.03.001 (2019).
- 542 51 Barros, R. P. & Gustafsson, J. A. Estrogen receptors and the metabolic network. *Cell Metab* 543 **14**, 289-299, doi:10.1016/j.cmet.2011.08.005 (2011).
- 544 52 Chidi-Ogbolu, N. & Baar, K. Effect of Estrogen on Musculoskeletal Performance and Injury<br>545 Risk. Front. Physiol. 9, 1834, doi:10.3389/fphys.2018.01834 (2018). 545 Risk. *Front. Physiol.* **9**, 1834, doi:10.3389/fphys.2018.01834 (2018).

- 546 53 Velders, M. & Diel, P. How sex hormones promote skeletal muscle regeneration. *Sports Med.* 547 **43**, 1089-1100, doi:10.1007/s40279-013-0081-6 (2013).
- 548 54 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set<br>549 analysis of GWAS data. *PLoS Comput. Biol.* 11. e1004219. 549 analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219, 550 doi:10.1371/journal.pcbi.1004219 (2015).
- 551 55 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739- 552 1740, doi:10.1093/bioinformatics/btr260 (2011).
- 553 56 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585, 554 doi:10.1038/ng.2653 (2013).
- 555 57 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and<br>556 annotation of genetic associations with FUMA. Nat Commun 8, 1826, doi:10.1038/s41467-556 annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467- 017-01261-5 (2017).
- 558 58 Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the 559 human phenome. *Elife* **7**, doi:10.7554/eLife.34408 (2018).
- 560 59 Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between<br>561 imprecisely measured traits using GWAS summary data. *PLoS genetics* 13, e1007081, 561 imprecisely measured traits using GWAS summary data. *PLoS genetics* **13**, e1007081, 562 doi:10.1371/journal.pgen.1007081 (2017).
- 563 60 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide<br>564 range of complex diseases of middle and old age. *PLoS Med.* 12. e1001779. 564 range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779, 565 doi:10.1371/journal.pmed.1001779 (2015).
- 566 61 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 567 **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 568 62 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in 569 large-scale genetic association studies. *Nat. Genet.* **50**, 1335-1341, doi:10.1038/s41588-018- 570 0184-y (2018).
- 571 63 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190-2191, 572 genomewide association scans. *Bioinformatics* **26**, 2190-2191, 573 doi:10.1093/bioinformatics/btq340 (2010).
- 574 64 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants 575 from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164, doi:10.1093/nar/gkq603 576 (2010).
- 577 65 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506- 578 D515, doi:10.1093/nar/gky1049 (2019).
- 579 66 Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex 580 trait gene targets. *Nat. Genet.* **48**, 481-487, doi:10.1038/ng.3538 (2016).
- 581 67 Cuéllar-Partida, G. *et al.* Complex-Traits Genetics Virtual Lab: A community-driven web 582 platform for post-GWAS analyses. *bioRxiv*, 518027, doi:10.1101/518027 (2019).
- 583 68 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity 584 in genome-wide association studies. *Nat. Genet.* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 585 586

#### 587 **Contributors of FinnGen**  588 **Steering Committee**  589 Aarno Palotie Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland 591<br>592<br>593 **Pharmaceutical companies**<br>Bridget Riley-Gills 593 Bridget Riley-Gills Abbvie, Chicago, IL, United States<br>594 Howard Jacob Abbvie, Chicago, IL, United States 594 Howard Jacob Abbvie, Chicago, IL, United States<br>595 Dirk Paul Astra Zeneca, Cambridge, United K<br>596 Heiko Runz Biogen, Cambridge, MA, United St 595 Dirk Paul Astra Zeneca, Cambridge, United Kingdom 596 Heiko Runz Biogen, Cambridge, MA, United States 597 Sally John Biogen, Cambridge, MA, United States 598 Robert Plenge Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States 599 Joseph Maranville Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States 600 Mark McCarthy Genentech, San Francisco, CA, United States<br>601 Julie Hunkapiller Genentech, San Francisco, CA, United States 601 Julie Hunkapiller Genentech, San Francisco, CA, United States 602 Meg Ehm GlaxoSmithKline, Brentford, United Kingdor 602 Meg Ehm GlaxoSmithKline, Brentford, United Kingdom<br>603 Kirsi Auro GlaxoSmithKline, Brentford, United Kingdom 603 Kirsi Auro GlaxoSmithKline, Brentford, United Kingdom<br>604 Simonne Longerich Merck. Kenilworth. NJ. United States 604 Simonne Longerich Merck, Kenilworth, NJ, United States<br>605 Caroline Fox Merck, Kenilworth, NJ, United States 605 Caroline Fox Merck, Kenilworth, NJ, United States<br>606 Anders Mälarstig Pfizer, New York, NY, United States 606 Anders Mälarstig Pfizer, New York, NY, United States 607 Katherine Klinger Sanofi, Paris, France 607 Katherine Klinger Sanofi, Paris, France<br>608 Deepak Raipal Sanofi, Paris, France 608 Deepak Raipal Sanofi, Paris, France<br>609 Eric Green Maze Therapeutics, S 609 Eric Green Maze Therapeutics, San Francisco, CA, United States<br>610 Robert Graham Maze Therapeutics, San Francisco, CA, United States 610 Robert Graham Maze Therapeutics, San Francisco, CA, United States 611 Robert Yang Janssen Biotech, Beerse, Belgium 611 Robert Yang Janssen Biotech, Beerse, Belgium<br>612 Chris O'Donnell Novartis, Basel, Switzerland Novartis, Basel, Switzerland 613 614 **University of Helsinki & Biobanks**  615 Tomi Mäkelä HiLIFE, University of Helsinki, Finland, Finland<br>616 Jaakko Kaprio Institute for Molecular Medicine Finland, HiLIFE 616 Jaakko Kaprio Institute for Molecular Medicine Finland, HiLIFE, Helsinki, Finland, Finland<br>617 Petri Virolainen Auria Biobank / University of Turku / Hospital District of Southwest Finland. 617 Petri Virolainen Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland<br>618 Antti Hakanen Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finla 618 Antti Hakanen Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland<br>619 Terhi Kilpi THL Biobank / The National Institute of Health and Welfare Helsinki, Finland 619 Terhi Kilpi THL Biobank / The National Institute of Health and Welfare Helsinki, Finland 620 Markus Perola THL Biobank / The National Institute of Health and Welfare Helsinki, Finland 620 Markus Perola THL Biobank / The National Institute of Health and Welfare Helsinki, Finland 621 Jukka Partanen Finnish Red Cross Blood Service / Finnish Hematology Registry and C<br>622 Helsinki, Finland Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, 622 Helsinki, Finland<br>623 Anne Pitkäranta<br>624 Helsinki Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, 624 Helsinki<br>625 Juhani Junttila 625 Juhani Junttila Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital<br>626 District, Oulu, Finland Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospita District, Oulu, Finland<br>Raisa Serpi Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland<br>Tarja Laitinen 629 Tarja Laitinen Finnish Clinical Biobank Tampere / University of Tampere / Pirkanmaa Hospital District, 630 630 Tampere, Finland<br>631 Veli-Matti Kosma 631 Veli-Matti Kosma Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District,<br>632 Kuopio, Finland 632 Kuopio, Finland<br>633 Urho Kujala 633 Urho Kujala Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland<br>Marco Hautalahti FINBB - Finnish biobank cooperative  $635$ <br> $636$ <br> $637$ <br> $638$ **Other Experts/ Non-Voting Members**<br> **Outi Tuovila** Business F 638 Outi Tuovila Business Finland, Helsinki, Finland<br>639 Raimo Pakkanen Business Finland, Helsinki, Finland Business Finland, Helsinki, Finland 640<br>641 641 **Scientific Committee**  642 **Pharmaceutical companies**  643 Jeffrey Waring Abbvie, Chicago, IL, United States<br>644 Bridget Riley-Gillis Abbvie, Chicago, IL, United States 644 Bridget Riley-Gillis Abbvie, Chicago, IL, United States<br>645 Ioanna Tachmazidou Astra Zeneca, Cambridge, United K 645 Ioanna Tachmazidou Astra Zeneca, Cambridge, United Kingdom<br>646 Chia-Yen Chen Biogen, Cambridge, MA, United States 646 Chia-Yen Chen Biogen, Cambridge, MA, United States<br>647 Heiko Runz Biogen, Cambridge, MA, United States 647 Heiko Runz Biogen, Cambridge, MA, United States<br>648 Shameek Biswas Celgene, Summit, NJ, United States/Bri 648 Shameek Biswas Celgene, Summit, NJ, United States/Bristol Myers Squibb, New York, NY, United States 649 Sarah Pendergrass Genentech, San Francisco, CA, United States 650 Julie Hunkapiller Genentech, San Francisco, CA, United States<br>651 Meg Ehm GlaxoSmithKline, Brentford, United Kingdor 651 Meg Ehm GlaxoSmithKline, Brentford, United Kingdom<br>652 David Pulford GlaxoSmithKline, Brentford, United Kingdom 652 David Pulford GlaxoSmithKline, Brentford, United Kingdom<br>653 Neha Raghavan Merck, Kenilworth, NJ, United States Merck, Kenilworth, NJ, United States















