

# Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis

C.S. Gallagher et al.<sup>#</sup>

Uterine leiomyomata (UL) are the most common neoplasms of the female reproductive tract and primary cause for hysterectomy, leading to considerable morbidity and high economic burden. Here we conduct a GWAS meta-analysis in 35,474 cases and 267,505 female controls of European ancestry, identifying eight novel genome-wide significant ( $P < 5 \times 10^{-8}$ ) loci, in addition to confirming 21 previously reported loci, including multiple independent signals at 10 loci. Phenotypic stratification of UL by heavy menstrual bleeding in 3409 cases and 199,171 female controls reveals genome-wide significant associations at three of the 29 UL loci: 5p15.33 (*TERT*), 5q35.2 (*FGFR4*) and 11q22.3 (*ATM*). Four loci identified in the meta-analysis are also associated with endometriosis risk; an epidemiological meta-analysis across 402,868 women suggests at least a doubling of risk for UL diagnosis among those with a history of endometriosis. These findings increase our understanding of genetic contribution and biology underlying UL development, and suggest overlapping genetic origins with endometriosis.

\*email: [netta\\_makinen@dfci.harvard.edu](mailto:netta_makinen@dfci.harvard.edu); [cmorton@bwh.harvard.edu](mailto:cmorton@bwh.harvard.edu). <sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

Uterine leiomyomata (UL), also known as uterine fibroids, are hormone-driven tumors with an estimated prevalence ranging from 20–77%<sup>1,2</sup>. Although the majority of UL are asymptomatic, about 25% of women with UL are symptomatic, and may experience heavy menstrual bleeding (HMB), abdominal pain, infertility, and increased risk of miscarriage<sup>3</sup>. Currently, the only essentially curative treatment is uterine extirpation via total hysterectomy. Known risk factors for UL include increasing age up to menopause, ethnicity (particularly African ancestry), family history of UL, nulliparity, and increased body mass index (BMI)<sup>4</sup>. Studies of familial aggregation and twins, as well as racial differences in prevalence and morbidity, suggest heritable factors influence risk for developing UL<sup>5–10</sup>. Recent GWAS have identified 26 loci significantly associated ( $P < 5 \times 10^{-8}$ ) with UL: 10q24.33, 11p15.5, and 22q13.1 in Japanese women<sup>11</sup>, 25 loci in white women of European ancestry<sup>12–15</sup>, including the three previously identified loci in Japanese women, and a distinct region at 22q13.1 in African American women<sup>16</sup>.

To define further the genetic architecture of UL, we perform a discovery meta-analysis of GWAS on UL across a total of 35,474 cases and 267,505 female controls of white European ancestry, which more than doubles the case sample size of previously reported GWAS<sup>11,12,14–16</sup>. The meta-analysis identifies eight novel loci significantly associated with UL ( $P < 5 \times 10^{-8}$ ) and confirms 21 previously reported European risk loci. Interestingly, HMB-limited UL GWAS reveals three of the 29 independent loci to be significantly associated with the co-occurrence of UL and HMB. Four loci identified in the meta-analysis are also reported to be associated with risk for endometriosis, which together with an epidemiological meta-analysis indicating an association between endometriosis and diagnosis of UL suggest overlapping genetic origins between the two highly common gynecologic diseases.

## Results

**UL GWAS meta-analysis.** Our discovery meta-analysis of GWAS on UL includes four population-based cohorts and one direct-to-consumer cohort of white European ancestry: Women's Genome Health Study (WGHS), Northern Finnish Birth Cohort (NFBC), QIMR Berghofer Medical Research Institute (QIMR), UK Biobank (UKBB), and 23andMe (Supplementary Methods, Supplementary Table 1). Imputation of genotypes was carried out using 1000 Genomes Project Phase 3 and Haplotype Reference Consortium (HRC) reference panels. UL phenotype in each cohort was analyzed in a logistic regression or linear mixed model assuming additive genetic effects with multivariate adjustment for age, BMI, and/or correction for population structure. After quality control metrics were applied, including exclusion of non-informative ( $MAF < 0.01$ ) and poorly imputed ( $r^2 < 0.4$ ) SNPs, we performed a fixed-effects, inverse-variance-weighted (IVW) meta-analysis across 35,474 cases with a clinical or self-reported history of UL and 267,505 unaffected female controls. Altogether 8,662,096 biallelic SNPs were analyzed and adjustments for genomic inflation performed (Supplementary Fig. 1, Supplementary Table 2). Through linkage disequilibrium score (LDSC) regression analysis, an estimated 89.5% of the genomic inflation factor ( $\lambda_{GC}$ ) of 1.12 was attributable to polygenic heritability (intercept = 1.02, s.e. = 0.0081). Overall, individual SNP-based heritability ( $h^2$ ) was estimated to be 0.0281 (s.e. = 0.0029) on the liability scale.

**Risk loci associated with UL.** We observe genome-wide significant associations ( $P < 5 \times 10^{-8}$ ) at 2505 SNPs across 29 independent loci (Table 1, Supplementary Fig. 2, Supplementary Table 3). The Manhattan plot is shown in Fig. 1. We identify eight

novel loci associated with UL (2p23.2, 4q22.3, 6p21.31, 7q31.2, 10p11.22, 11p14.1, 12q15, and 12q24.31), which include the following candidate genes of interest: *HMGAI*, *BABAM2*, and *WNT2*. *HMGAI* is a member of the high mobility group proteins and is involved in regulation of gene transcription<sup>17</sup>. Somatic rearrangements of *HMGAI* at 6p21 have been recurrently documented in UL, albeit at a much lower frequency than those of *HMGAI2*—another member of the high mobility group protein family<sup>18–20</sup>. *BABAM2* at 2p23.2 encodes a death receptor-associating intracellular protein that promotes tumor growth by suppressing apoptosis<sup>21</sup>. Associations at 7q31.2 containing *WNT2*, a member of the Wnt gene family, provide support for the previously suggested role of Wnt signaling in UL<sup>22,23</sup>.

Among 29 independent loci are 21 loci previously reported to be significantly associated with UL<sup>11–16</sup>. A number of identified loci harbor genes previously implicated in cell growth and cancer risk in different tissue types, including cervical cancer<sup>24</sup>, epithelial ovarian cancer<sup>25,26</sup>, breast cancer<sup>27,28</sup>, glioma<sup>29,30</sup>, bladder cancer<sup>31</sup>, and pancreatic cancer<sup>32–34</sup>. Specifically, seven independent loci contain well-characterized oncogenes and tumor suppressor genes from the Cancer Gene Census list in COSMIC<sup>35</sup>: *PDGFRA*, *TERT*, *ESR1*, *WT1*, *ATM*, *FOXO1*, and *TP53*.

Using approximate conditional analysis, we identify multiple distinct association signals for UL at 10 loci (at locus-wide significance,  $P < 1 \times 10^{-5}$ , Bonferroni correction) (Supplementary Table 4). Fine-mapping was conducted on all 43 distinct association signals arising from the 29 detected UL loci, revealing three association signals with a single variant in the 99% credible set (Fig. 2, Supplementary Table 5). The missense variant at 20p12.3 (rs16991615; E341K) maps to *MCM8*, a gene that encodes a protein involved in DNA double-strand break repair<sup>36</sup>. *MCM8* has also been implicated in length of reproductive lifespan, menopause, and premature ovarian failure<sup>37,38</sup>. Another variant (rs78378222) resides in the 3'UTR of *TP53* at 17p13.1, and has been shown to disturb 3'-end processing of *TP53* mRNA<sup>39</sup>. This variant has been associated with both malignant and benign tumor types<sup>39–41</sup>.

**UL GWAS limited by HMB.** HMB, one of the major symptoms of UL, is estimated to affect up to 30% of reproductive-aged women, having a considerable impact on a woman's quality of life. Thus, variants specifically associated with this symptom are of particular interest for drug target development. We performed a GWAS on UL limited by HMB using a linear mixed model across 3409 cases and 199,171 unaffected female controls from the UKBB (Supplementary Methods, Supplementary Fig. 3). We observe genome-wide significant associations ( $P < 5 \times 10^{-8}$ ) at three of the 29 independent UL loci: 5p15.33 (rs72709458, OR [95% CI] = 0.86 [0.81–0.91],  $P = 3.50 \times 10^{-8}$ ), 5q35.2 (rs2456181, OR [95% CI] = 0.87 [0.83–0.91],  $P = 4.20 \times 10^{-10}$ ), and 11q22.3 (rs1800057, OR [95% CI] = 0.66 [0.58–0.76],  $P = 2.80 \times 10^{-9}$ ) (Fig. 3, Supplementary Fig. 4, Supplementary Table 6). The lead SNP at 11q22.3, a missense variant in *ATM*, has been associated with increased risk of various cancers, such as breast cancer<sup>42,43</sup>, while the lead SNP at 5p15.33, an intronic *TERT* variant, has previously been implicated in gliomas<sup>44</sup>. The lead SNP rs2456181 at 5q35.2 resides near *FGFR4*, a gene encoding a cell-surface receptor for fibroblast growth factors involved in regulation of several pathways, including cell proliferation, differentiation, and migration.

**HMB GWAS.** A GWAS based solely on HMB across 9813 cases and 210,946 female controls reveals one genome-wide significant association at 11p14.1, one of the eight novel loci associated with

**Table 1 Overview of lead SNPs with significant associations at 29 independent loci in UL GWAS meta-analysis**

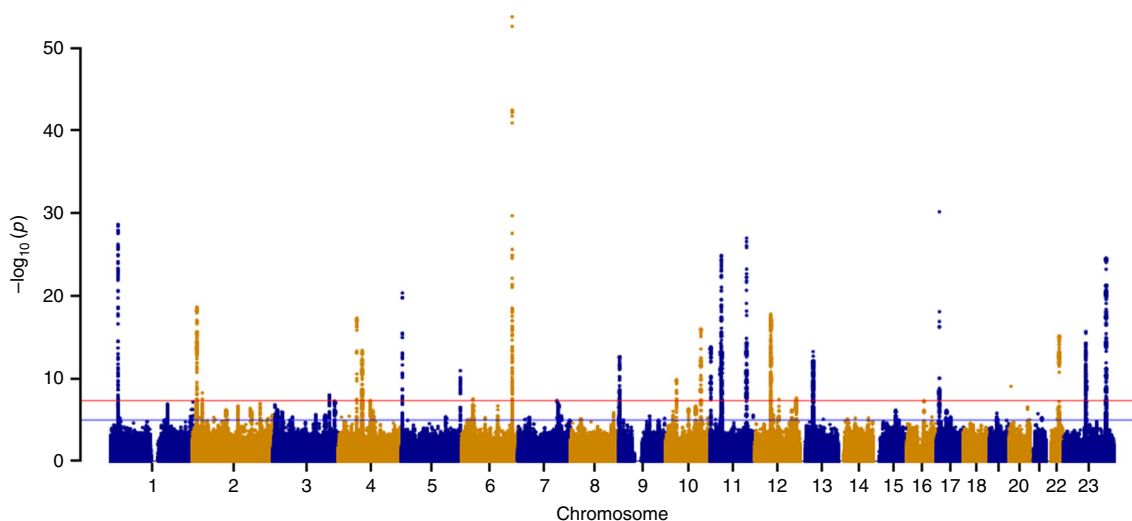
Locus	Lead SNP	RA	OA	RAF <sub>EUR</sub>	P <sub>Meta</sub>	OR (95% CI)	Gene(s) of interest <sup>a</sup>
1p36.12 <sup>b,c</sup>	rs7412010	C	G	0.15	2.4 × 10 <sup>-29</sup>	1.13 (1.11-1.16)	WNT4, CDC42
2p23.2	rs55819434	A	G	0.91	5.6 × 10 <sup>-09</sup>	0.92 (0.90-0.95)	BABAM2
2p25.1 <sup>b,c</sup>	rs35417544	T	C	0.69	2.3 × 10 <sup>-19</sup>	1.09 (1.07-1.10)	GREB1
3q26.2 <sup>c</sup>	rs35446936	A	G	0.24	1.0 × 10 <sup>-08</sup>	0.95 (0.93-0.96)	TERC
4q12 <sup>c</sup>	rs62323682	T	C	0.94	4.9 × 10 <sup>-18</sup>	0.87 (0.84-0.90)	LNX1, PDGFRA
4q13.3 <sup>c</sup>	rs12640488	A	G	0.52	4.0 × 10 <sup>-14</sup>	0.94 (0.92-0.96)	SULT1B1
4q22.3	rs4699299	T	C	0.69	4.7 × 10 <sup>-08</sup>	0.95 (0.94-0.97)	PDLIM5
5p15.33 <sup>c</sup>	rs72709458	T	C	0.23	4.7 × 10 <sup>-21</sup>	1.10 (1.08-1.13)	TERT
5q35.2 <sup>c</sup>	rs2456181	C	G	0.49	1.1 × 10 <sup>-11</sup>	0.94 (0.93-0.96)	ZNF346, UIMC1
6p21.31	rs116251328	A	T	0.02	3.0 × 10 <sup>-08</sup>	1.15 (1.09-1.21)	GRM4, HMGA1
6q25.2 <sup>b,c</sup>	rs58415480	C	G	0.84	1.9 × 10 <sup>-54</sup>	0.84 (0.82-0.86)	SYNE1, ESR1
7q31.2	rs2270206	A	C	0.16	4.6 × 10 <sup>-08</sup>	1.06 (1.04-1.09)	WNT2
9p24.3 <sup>c</sup>	rs10976689	A	G	0.60	2.4 × 10 <sup>-13</sup>	0.94 (0.93-0.96)	ANKRD15
10q24.3 <sup>c</sup>	rs9419958	T	C	0.13	1.1 × 10 <sup>-16</sup>	1.10 (1.08-1.13)	OBFC1, SLK
10p11.22	rs10508765	A	G	0.80	1.5 × 10 <sup>-10</sup>	1.07 (1.05-1.09)	ZEB1, ARHGAP12
11p15.5 <sup>c</sup>	rs547025	T	C	0.92	1.5 × 10 <sup>-14</sup>	1.13 (1.09-1.16)	RIC8A, BET1L
11p14.1 <sup>b</sup>	rs11031006	A	G	0.14	5.7 × 10 <sup>-15</sup>	0.91 (0.89-0.93)	FSHB
11p13 <sup>c</sup>	rs61889186	C	G	0.86	1.4 × 10 <sup>-25</sup>	0.89 (0.87-0.91)	WT1
11p13 <sup>c</sup>	rs2785202	C	G	0.55	6.9 × 10 <sup>-14</sup>	1.06 (1.05-1.08)	PDHX, CD44
11q22.3 <sup>c</sup>	rs149934734	T	C	0.03	1.1 × 10 <sup>-27</sup>	1.33 (1.26-1.40)	C11orf65, KDELDC2
12q13.11 <sup>c</sup>	rs2131371	A	C	0.28	1.6 × 10 <sup>-18</sup>	0.93 (0.91-0.94)	SLC38A2
12q15	rs11178393	T	C	0.89	3.3 × 10 <sup>-08</sup>	1.08 (1.05-1.10)	PTPRR
12q24.31	rs28583837	A	G	0.22	2.3 × 10 <sup>-08</sup>	0.94 (0.92-0.96)	PITPNM2
13q14.11 <sup>c</sup>	rs117245733	A	G	0.02	5.7 × 10 <sup>-14</sup>	1.31 (1.21-1.39)	FOXO1
17p13.1 <sup>c</sup>	rs78378222	T	G	0.99	7.1 × 10 <sup>-31</sup>	0.65 (0.60-0.70)	SHBG, TP53
20p12.3 <sup>c</sup>	rs16991615	A	G	0.07	8.8 × 10 <sup>-10</sup>	1.11 (1.07-1.14)	MCM8, TRMT6
22q13.1 <sup>c</sup>	rs4821939	A	T	0.20	7.8 × 10 <sup>-16</sup>	1.08 (1.06-1.10)	TNRC6B
Xp26.2 <sup>c</sup>	rs12392108	A	T	0.31	5.9 × 10 <sup>-46</sup>	1.13 (1.11-1.15)	RAP2C
Xq13.1 <sup>c</sup>	rs4360450	A	G	0.37	2.1 × 10 <sup>-18</sup>	1.08 (1.06-1.10)	MED12

SNP single-nucleotide polymorphism, RA risk allele, OA other allele, RAF<sub>EUR</sub> average risk allele frequency in European samples, OR odds ratio

<sup>a</sup>≤300 kb distant from association signal

<sup>b</sup>Loci previously associated with endometriosis

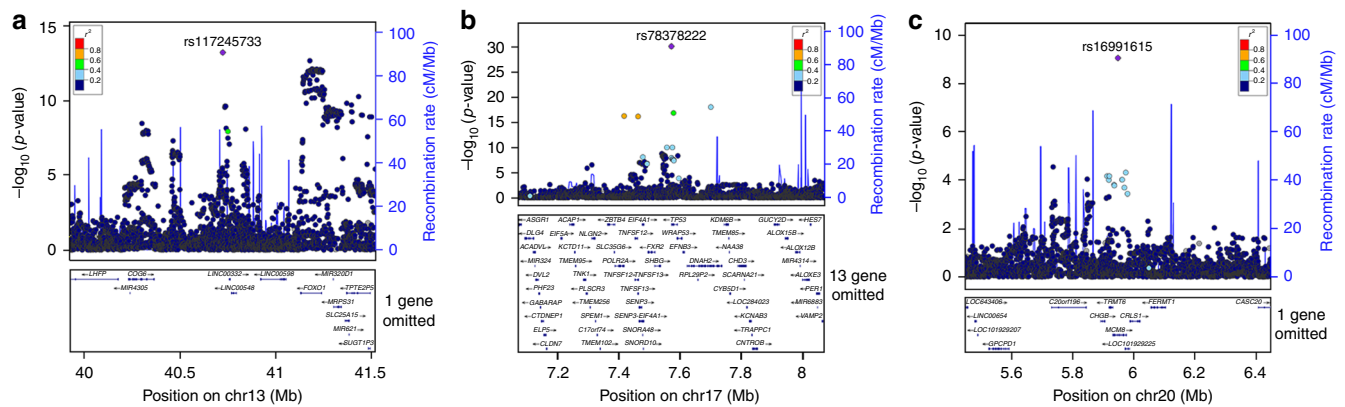
<sup>c</sup>Loci previously associated with UL



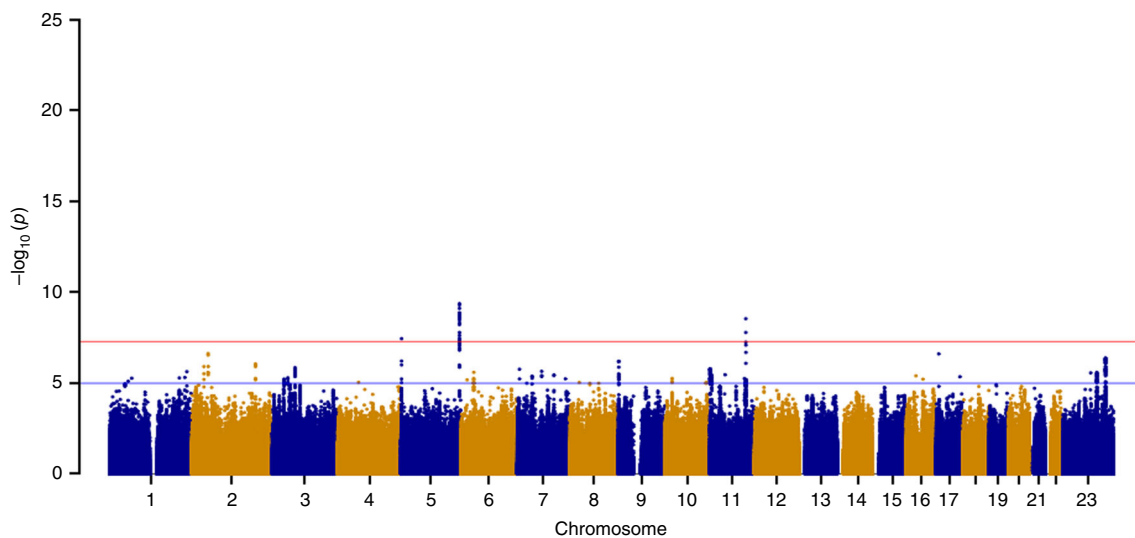
**Fig. 1** Manhattan plot for UL GWAS meta-analysis across all cohorts. Meta-analysis of GWAS including 302,979 women of white European ancestry across all cohorts identified 29 independent loci associated with UL. Red and blue horizontal lines indicate genome-wide significant ( $P < 5 \times 10^{-8}$ ) and suggestive ( $P < 1 \times 10^{-5}$ ) thresholds, respectively

UL (Supplementary Figs. 5 and 6). The lead SNPs for UL and HMB at 11p14.1 are in high LD, and the direction of the effect is the same (Supplementary Fig. 7). This locus has previously been associated with endometriosis, age at menarche, and follicle-stimulating/luteinizing hormone levels<sup>45-47</sup>. According to GTEx

(v7), the lead SNP for HMB (rs11031005) is a potential expression quantitative trait locus (eQTL) for *ARL14EP* in several tissue types, such as testis and thyroid. Mendelian randomization (MR) was used to assess the causality of genetic association between UL (exposure) and HMB (outcome). Interestingly, MR reveals that



**Fig. 2** Fine-mapping reveals three association signals with a single driver in 99% credible set. Association with UL is expressed as  $-\log_{10}(P$  value) for the three signals on chromosomes: (a) 13q14.11, (b) 17p13.1, and (c) 20p12.3. The labeled SNP represents the most significant SNP for each locus. SNP association  $P$ -value is shown on the y axis, while SNP position (with gene annotation) appears on the x axis. Each SNP is colored according to the strength of LD with the lead SNP. Regional association plots were produced in LocusZoom



**Fig. 3** Manhattan plot for GWAS on UL limited by heavy menstrual bleeding. GWAS across 202,580 women of white European ancestry identified three independent loci associated with UL limited by heavy menstrual bleeding. Red and blue horizontal lines indicate genome-wide significant ( $P < 5 \times 10^{-8}$ ) and suggestive ( $P < 1 \times 10^{-5}$ ) thresholds, respectively

genetic predisposition to UL is causally linked to an increased risk of HMB, with the  $\beta$  estimate of 0.26 being significant in the IVW model ( $P = 1.2 \times 10^{-12}$ ) in the absence of heterogeneity ( $P = 0.13$ ) (Supplementary Table 7). The MR Egger regression shows no significant directional pleiotropy (intercept = 0.01,  $P = 0.36$ ) supporting a causal relationship.

**Overlap of UL and endometriosis.** Interestingly, significant association signals are observed at several loci previously associated with endometriosis: 1p36.12 (rs7412010, OR [95% CI] = 1.13 [1.11–1.16],  $P = 2.43 \times 10^{-29}$ ), 2p25.1 (rs35417544, OR [95% CI] = 1.09 [1.07–1.10],  $P = 2.32 \times 10^{-19}$ ), 6q25.2 (rs58415480, OR [95% CI] = 1.19 [1.17–1.22],  $P = 1.86 \times 10^{-54}$ ), and 11p14.1 (rs11031006, OR [95% CI] = 1.10 [1.07–1.12],  $P = 5.65 \times 10^{-15}$ )<sup>45,48–50</sup>. LD is strong between UL and previously reported endometriosis lead SNPs<sup>45</sup> at all except one locus, 2p25.1 (Supplementary Table 8). In addition, the direction of effect is the same between the lead SNPs at 1p36.12. Using LDSC regression, we observe a moderate genetic correlation between UL and endometriosis in women with European ancestry ( $r_g = 0.39$ ,

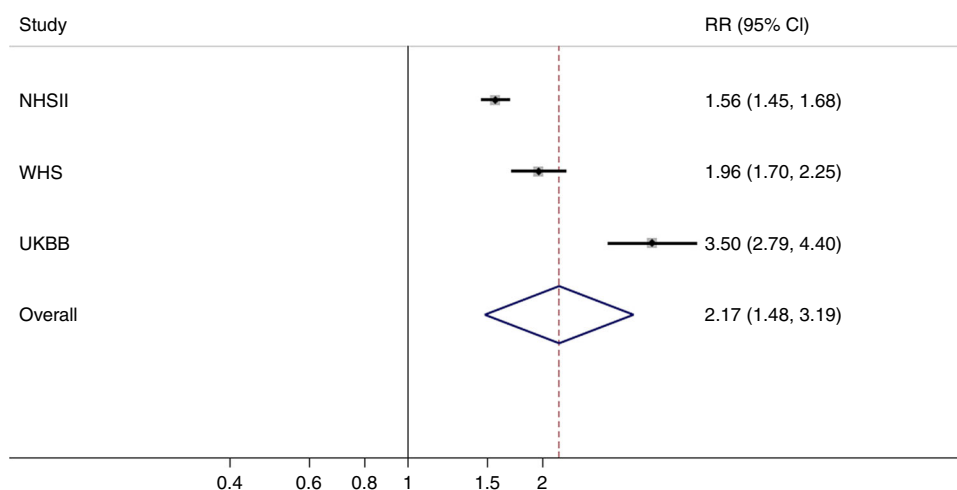
$s.e = 0.05$ ,  $P = 9.77 \times 10^{-13}$ ). Endometriosis has an earlier age-of-onset than UL, with a mean age of 25–29 years and 35 years, respectively. MR suggests that genetic predisposition to endometriosis (exposure) is causally linked to an increased risk of UL (outcome); the  $\beta$  of 0.36 is significant ( $P = 3.7 \times 10^{-3}$ ) in the IVW model (heterogeneity  $P = 9.5 \times 10^{-68}$ ) (Supplementary Table 7). Leave-one-out sensitivity analysis reveals that no single SNP alone drives the significant relationship between endometriosis and UL, but instead the relationship is accounted for by contributions from multiple variants across the genome (Supplementary Fig. 8). Given the high degree of heterogeneity, the effect sizes were estimated in a minimal set of SNPs that when used as a genetic instrument eliminate the heterogeneity (Supplementary Fig. 9). The effect size estimate ( $\beta = 0.12$ ) from the minimal set of variants remains significant ( $P = 4.3 \times 10^{-3}$ ) in the IVW model in the absence of heterogeneity ( $P = 0.23$ ). We also applied the MR pleiotropy residual sum and outlier (MR-PRESSO) global and distortion tests to adjust for variants causing significant bias in the estimates through horizontal pleiotropy. Outlier-adjusted estimates still provide significant evidence for a causal estimate of endometriosis on UL ( $\beta = 0.29$ ,  $P = 0.002$ ) (Supplementary Table 7).

Endometriosis, defined by ectopic growth of endometrial-like tissue outside the uterus, is a common inflammatory hormone-dependent disease that affects reproductive-aged women<sup>51</sup>. Although functional studies of relevant tissue are needed to confirm consequences of the variants in regulation of gene expression, each of the four observed overlapping genomic loci contains a gene(s) known to be involved in progesterone or estrogen signaling. *WNT4* at 1p36.12 encodes a secreted signaling factor that promotes female sex development, and regulates both postnatal uterine development and progesterone signaling during decidualization<sup>52,53</sup>. Recently, SNPs at 1p36.12 associated with a greater endometriosis risk have been suggested to act through *CDC42*, a gene that encodes a small GTPase of the Rho family<sup>54</sup>. *GREB1* at 2p25.1 is an early response gene in the estrogen receptor (ER)-regulated pathway, and promotes growth of breast and pancreatic cancer cells<sup>55,56</sup>. *ESR1* at 6q25.2 encodes the alpha subunit of the ligand-activated nuclear ER that regulates cell proliferation in the uterus<sup>57</sup>. *FSHB* at 11p14.1 encodes the biologically active subunit of follicle-stimulating hormone, which regulates maturation of ovarian follicles and release of ova during menstruation<sup>58,59</sup>.

**Epidemiological meta-analysis.** Given shared risk loci and genetic correlation of UL and endometriosis, we conducted an epidemiological meta-analysis including 402,868 women from three population-based cohorts: Nurses' Health Study II (NHSII), Women's Health Study (WHS), and UKBB (Supplementary Methods, Supplementary Table 9), to assess the likelihood of UL diagnosis among women who had or had not been diagnosed with endometriosis. Women with endometriosis had a significantly higher likelihood of UL diagnosis (multivariable-adjusted summary relative risk (RR) [95% CI] = 2.17 [1.48–3.19]) (Fig. 4). All cohort-specific analyses demonstrated at least a doubling of risk, suggesting a robust association (Table 2). However, biologically and statistically significant heterogeneity was observed in the pooling of effect size estimates in the meta-analysis ( $P < 1 \times 10^{-4}$ ) (Fig. 4). Therefore, absolute effect estimates need to be interpreted in the context of source populations. Heterogeneity could reflect various different population sampling and data collection factors among the three cohorts. First, the validity of self-reported diagnosis of endometriosis and to a lesser extent UL are known to be <75% in general population cohorts,

such as UKBB, compared to more highly validated self-assessment in the medical professional NHSII and WHS cohorts<sup>7</sup>. Second, endometriosis clinical definitions prior to the 1990s were more restrictive—often limited to the presence of endometrioma and/or “powderburn” superficial peritoneal lesions among adult women<sup>60</sup>. Subsequently, definitions have expanded to recognize a wide range of superficial peritoneal phenotypic presentations, as well as incidence among adolescents and young women<sup>61</sup>. It may be impactful therefore that the WHS participants were  $\geq 45$  years of age in 1992, while NHSII participants were  $\geq 25$  years of age in 1989, and UKBB participants were aged 40 to 69 in 2006. Thus, disease definitions varied during the peak calendar years of incidence among the cohorts, and in addition, while the NHSII were queried about endometriosis prospectively during their reproductive years, the WHS and UKBB cohorts were cross-sectionally asked to recall their gynecologic health experience decades earlier. It is also important to note that while WHS and NHSII participants were asked specifically about endometriosis diagnosis via questionnaire, the UKBB data collection included qualitative interviews during which endometriosis would be documented only when the participant herself raised it as a health issue. Those with mild symptoms or those past their reproductive years and thus past the moderate to severe life-impacting symptoms of the disease may have been less likely to offer endometriosis among the list of their health issues. This is supported by the low prevalence of endometriosis reported within the UKBB compared to WHS, NHSII, and other population-based estimates<sup>62</sup>. However, the UKBB participants (due to the qualitative interview structure and recall bias) and the WHS participants (due to recall bias) could have been more likely to choose to report endometriosis if they also suffered from UL together, resulting in diagnostic bias and consequently an inflation of effect estimates. Indeed, the population heterogeneity and differing potential for diagnostic bias by cohort fits with the observed differences among effect estimate magnitudes with the RRs and CI widths ordered from NHSII (RR = 1.56) to WHS (RR = 1.96) to UKBB (RR = 3.50) (Fig. 4).

**Bioinformatic analyses of UL risk SNPs and loci.** To estimate the genetic correlation between UL and various reproductive traits, as well as cardiometabolic traits/diseases, we performed LD Hub analysis for a total of 21 traits/diseases (Supplementary



**Fig. 4** Epidemiologic meta-analysis demonstrates endometriosis is associated with UL. Random-effects, inverse-variance-weighted meta-analysis was performed across the effect sizes and standard errors in 402,868 women from three cohorts (NHSII, WHS, and UKBB). Squares represent point estimates from individual studies, whiskers correspond to the 95% CIs, and the diamond represents results from the meta-analysis. There was evidence of significant heterogeneity based on Cochran's Q statistic ( $P < 1 \times 10^{-4}$ )



**Table 2 Multivariable-adjusted effect estimates of the association between endometriosis and UL among women in NHSII, WHS, and UKBB cohorts**

Cohort	n	UL cases	Age-adjusted	Multivariable-adjusted
Nurses' Health Study II <sup>a</sup>	102,545	10,714	1.61 (1.50-1.73)	1.57 (1.45-1.68)
Women's Health Study <sup>b</sup>	26,868	1,262	2.04 (1.78-2.34)	1.97 (1.71-2.26)
UK Biobank <sup>c</sup>	273,455	19,789	3.11 (2.86-3.37)	3.50 (2.77-4.38)

<sup>a</sup>Hazard ratios (95% confidence intervals [CIs]) from Cox regression models. Multivariable model was adjusted for age (continuous), age at menarche (<11, 11, 12, 13, 14-15, >15), infertility (yes, no), ancestry (White, Black, Hispanic, Asian, other), parity (nulliparous, 1, 2, 3, 4+), age at first birth (<25, 25-29, >29), time since last birth (<1, 1-3, 4-5, 6-7, 8-9, 10-12, 13-15, ≥16), age first oral contraceptive use (13-16, 17-20, 21-24, ≥25), BMI (<20, 20-21.9, 22-23.9, 24-24.9, 25-26.9, 27-29.9, ≥30), menstrual cycle length (<26, 26-31, 32-50, and >50 days), smoking (never, past, current), recent gynecologic/breast exam (no recent exam, recent exam), use of anti-hypertensive medications/diastolic blood pressure (no meds <65, no meds 65-74, no meds 75-84, no meds 85-89, no meds ≥90, meds <65, meds 65-74, meds 75-84, meds 85-89, meds ≥90), and physical activity (MET hours/week: <3, 3-9, 9-18, 18-27, 27-42, ≥42).

<sup>b</sup>Odds ratios (95% CI) from logistic regression models. Multivariable model was adjusted for age at baseline (continuous), age at menarche (<11, 11, 12, 13, 14-15, >15), ancestry (White, Black, Hispanic, Asian, other), parity (nulliparous, 1, 2, 3, ≥4), age at first birth (<25, 25-29, >29), oral contraceptive use (ever, never), BMI (<20, 20-21.9, 22-23.9, 24-24.9, 25-26.9, 27-29.9, ≥30), smoking (never, past, current), use of anti-hypertensive medications/diastolic blood pressure (no meds <65, no meds 65-74, no meds 75-84, no meds 85-89, no meds ≥90, meds <65, meds 65-74, meds 75-84, meds 85-89, meds ≥90), physical activity (never/rarely, <1 time/week, 1-3 times/week, ≥4 times/week), alcohol consumption (never/rarely, 1-3 drinks/month, 1-6 drinks/week, ≥1 drinks/day).

<sup>c</sup>Odds ratios (95% CI) from logistic regression models. Multivariable model was adjusted for age (<50, 50-55, 56-60, ≥60), age at menarche (<11, 11, 12, 13, 14-15, >15), ancestry (White, Black, Asian, other), parity (nulliparous, 1, 2, 3, ≥4), age at first birth (<25, 25-29, >29), oral contraceptive use (ever, never), BMI (<20, 20-21.9, 22-23.9, 24-24.9, 25-26.9, 27-29.9, ≥30), smoking (never, past, current), physical activity (never/rarely, 1-3 drinks/week, ≥4 drinks/week), alcohol consumption (never/rarely, 1-3/month, 1-4/week, ≥1/day), and menopausal status (premenopausal, postmenopausal).

Data 1). We observe significant correlations between increased risk of UL and earlier age of menarche ( $r_g = -0.16$ ,  $P = 3.7 \times 10^{-6}$ ), earlier age of first birth ( $r_g = -0.14$ ,  $P = 1.0 \times 10^{-3}$ ), increased levels of triglycerides ( $r_g = 0.13$ ,  $P = 1.9 \times 10^{-3}$ ), and increased BMI ( $r_g = 0.11$ ,  $P = 2.0 \times 10^{-3}$ ), as previously suggested by epidemiological studies<sup>63,64</sup>, illustrating that common genetic factors can predispose women to both risk factors related to, for example, adverse metabolic and cardiovascular disease risk and UL. Gene-set and tissue enrichment analyses across 8971 SNPs with suggestive ( $P < 1 \times 10^{-5}$ ) or significant ( $P < 5 \times 10^{-8}$ ) UL associations using DEPICT<sup>65</sup> reveal enrichments (false discovery rate (FDR) < 0.05) in gene sets, such as steroid hormone receptor (GO:0035258;  $P = 1.03 \times 10^{-5}$ ), hormone receptor binding (GO:0051427;  $P = 9.07 \times 10^{-5}$ ), and nuclear hormone receptor binding (GO:0035257;  $P = 1.53 \times 10^{-4}$ ) (Supplementary Data 2 and 3). The results are concordant with the hormone-driven nature of UL. We did not observe any cell/tissue types significantly enriched for the expression of the genes in the associated loci (Supplementary Fig. 10). To identify potential causal genes at UL risk loci, we used a summary-data based MR (SMR) method, including both eQTL and mQTL data from peripheral blood<sup>66,67</sup>. We identify 18 potential causal genes showing no significant heterogeneity in SMR ( $P_{\text{HEIDI}} > 5 \times 10^{-3}$ ), including *WNT4* (rs55938609,  $P_{\text{SMR}} = 6.92 \times 10^{-15}$ ), *GREB1* (rs35417544,  $P_{\text{SMR}} = 3.93 \times 10^{-19}$ ), *WT1* (rs12280757,  $P_{\text{SMR}} = 1.87 \times 10^{-18}$ ), and *FOXO1* (rs3924478,  $P_{\text{SMR}} = 5.76 \times 10^{-10}$ ) (Supplementary Data 4 and 5).

**FOXO1 expression in UL.** To explore potential functional significance, we examined expression of the FOXO1 protein, a transcription factor that plays an important role in cell proliferation, apoptosis, DNA repair, and stress response<sup>68</sup>. Interestingly, inactivation of *FOXO1* promotes cell proliferation and tumorigenesis in several hormone-regulated malignancies, such as prostate, breast, cervical, and endometrial cancers<sup>69-72</sup>. Conversely, we observe a significant increase in nuclear FOXO1 protein expression in UL compared to myometrial samples using immunohistochemistry on tissue microarrays (Supplementary Fig. 11). Patient-matched tumor-normal pairs show 1.69-fold higher ( $P = 0.01$ ; paired *t*-test) nuclear FOXO1 expression in UL, while the expression is as much as 2.32-fold greater ( $P = 1.52 \times 10^{-9}$ ; Welch's *t*-test) when all 335 UL are considered (Supplementary Fig. 12). These results are consistent with a previous study<sup>73</sup>, which showed phosphorylated (p) FOXO1 (pSer<sup>256</sup>) to be predominantly present in the nucleus in UL, but sequestered in the cytoplasm of myometrium. The concomitant increase of

p-FOXO1 and reduced expression of its interaction partner 14-3-3y in UL has been suggested to lead to impaired nuclear/cytoplasmic shuttling of p-FOXO1, which promotes cell survival<sup>73-75</sup>. We performed stratification of samples by genotype, revealing a statistically significant increase in FOXO1 levels of UL harboring the risk allele for rs6563799 (allelic dosage,  $P = 0.047$ ; homozygosity for risk allele,  $P = 0.035$ ) (Supplementary Figs. 11 and 13). An increase in FOXO1 levels of UL with the rs7986407 risk allele is also observed; however, the change is not statistically significant (Supplementary Figs. 11 and 13).

## Discussion

In our meta-analysis of GWAS on UL, we identify 29 genomic loci to be significantly associated with UL in women of white European ancestry, including eight novel and 21 previously reported loci. Candidate genes in the identified loci implicate pathways of estrogen and progesterone signaling (*ESR1*, *FSHB*, *GREB1*, *WNT2*, and *WNT4*), as well as cell growth (*FOXO1*, *PDGFRA*, *TERT*, *TERC*, and *TP53*) in predisposing women to UL. We do not confirm five of 26 previously identified loci reported to be significantly associated with UL<sup>12,14-16</sup>. Two of these loci, 3p24.1 and 16q12.1, are nominally significant ( $P < 1 \times 10^{-5}$ ) in our GWAS meta-analysis, but the remaining three loci (3q29, 17q25.3 and a distinct region at 22q13.1) do not reach nominal significance. Ancestral differences may explain the absence of the association originally identified in African American women in the genomic region at 22q13.1, while variation in phenotypic definitions<sup>12</sup> may underlie the two other loci.

Discovery of eight novel loci significantly associated with UL reveals several candidate genes of particular interest: *BABAM2*, *FSHB*, *HMGAI*, and *WNT2*. Because UL are benign tumors that rarely, if ever, develop into malignancy, the association between UL and multiple loci harboring well-known oncogenes and tumor suppressor genes is also worthy of note. Fine-mapping of the *TP53* locus identifies rs78378222 to be the most probable causal variant, which has been shown to disrupt the polyadenylation sequence in the 3'UTR of *TP53* and result in reduced expression of mRNA<sup>39</sup>. We also observe nuclear FOXO1 levels to be significantly elevated in UL when compared to myometrium. FOXO1 is a downstream target of the Akt signaling pathway that responds to hormone signaling through the progesterone receptor in UL and activates proliferative responses<sup>76</sup>.

HMB is one of the major debilitating symptoms of UL and can have a substantial impact on a woman's quality of life. Here, we report GWAS on both UL limited by HMB and solely on HMB, revealing potential targets for pharmacologic intervention:

*ARL14EP*, *ATM*, *TERT*, and *FGFR4*. In addition, MR analyses suggest that genetic predisposition to UL is causally linked to an increased risk of HMB. These results form a solid basis for further work to elucidate the mechanisms underlying UL-related HMB and towards tailored treatments of UL and HMB.

Biological overlap between UL and endometriosis, two highly common gynecologic diseases has long been suspected due to similarities in molecular mechanisms and progenitor cells. Our UL GWAS meta-analysis indicates that genes previously associated with endometriosis and involved in hormone-signaling pathways are also associated with UL (*WNT4/CDC42*, *GREB1*, *ESR1*, and *FSHB*). Overlap observed in the genetic etiology of endometriosis and UL led us to epidemiologically quantify the co-occurrence of these two diseases across three independent cohorts. The epidemiological meta-analysis indicates that women with a history of endometriosis are at elevated risk for reporting UL. Results from our MR analyses suggest that genetic predisposition to endometriosis is causally linked to increased risk of UL. Alternatively, given the discordance in the direction of allelic effects for the UL and endometriosis loci, our MR results may indicate a significant overlap in the underlying biology of the two diseases. Additional work is needed to better quantify the contribution of genetic effects to the directional relationship between endometriosis and UL. Results of which will enable us to quantify what portion of the MR results reflect the fundamental pathobiological overlap in these two diseases of the uterus. Further characterization of the mutual pathogenic mechanisms of UL and endometriosis has the capacity to discover not only a deeper understanding of the underlying biology, but also treatments for two diseases that cause significant morbidity in roughly one-third of the world's population.

## Methods

**Subjects.** For UL GWAS meta-analysis, four population-based cohorts (WGHS, NFBC, QIMR and UKBB) and one direct-to-consumer cohort (23andMe) from the FibroGENE consortium were included (Supplementary Table 1), resulting in 35,474 UL cases and 267,505 female controls of white European ancestry. Sample sizes were maximized using a basic, harmonizing phenotype definition to separate cases and controls solely based on either self-report or clinically documented UL history. Our large-scale epidemiologic analysis was comprised of three population-based cohorts (NHSII, WHS, and UKBB), totaling 402,869 women. HMB GWAS included the UKBB cohort, consisting of 220,759 women. Detailed descriptions of cohorts and sample selections are available in Supplementary Methods. All participants provided informed consent in accordance with the processes approved by the relevant jurisdiction for human subject research for each cohort: the Partners HealthCare System Human Research Committee (WHS/WGHS), the Ethical Committee of the Northern Ostrobothnia Hospital District (NFBC), the Human Research Ethics Committee at the QIMR Berghofer Medical Research Institute and the Australian Twin Registry (QIMR), the North West Multi-centre Research Ethics Committee (UKBB), Ethical and Independent Review Services (an external institutional review board; 23andMe), and the Institutional Review Boards at Harvard T.H. Chan School of Public Health and Brigham and Women's Hospital (Partners Human Research Committee) (NHSII).

**Genotyping.** Several different Illumina-based genotyping platforms (Illumina Inc., San Diego, CA, USA) were used: HumanHap300 Duo<sup>+</sup> chips or the combination of the Human-Hap300 Duo and iSelect chips (WGHS), Infinium 370cnvDuo array (NFBC), 317 K, 370 K, or 610 K SNP platforms (QIMR). Genotyping of participants in the UKBB was performed either on the Affymetrix UK BiLEVE or Affymetrix UK Biobank Axiom<sup>®</sup> array with over 95% similarity. Genotyping of participants in the 23andMe cohort was performed on various versions of Illumina-based BeadChips.

**Quality control and imputation.** Each cohort conducted quality control measures and imputation for their data. For WGHS, NFBC, QIMR, and 23andMe, all cases and controls with a genotyping call rate <0.98 were excluded from the study. Imputation was performed on both autosomal and sex chromosomes using the reference panel from the 1000 Genomes Project European dataset (1000 G EUR) Phase 3. Imputation was carried out using ShapeIT2 and IMPUTE2 softwares<sup>77,78</sup>. SNPs with call rates of <99% or with deviation from Hardy-Weinberg equilibrium ( $P \leq 1 \times 10^{-6}$ ) were excluded from further analyses. Population stratification for the data was examined with principal component analysis (PCA) using

EIGENSTRAT<sup>79</sup>. The four HapMap populations were used as reference groups: Europeans (CEU), Africans (YRI), Japanese (JPT), and Chinese (CHB). All observed outliers were removed from the study. UKBB data QC and imputation were performed centrally, prior to public release of the data<sup>80</sup>. Genotype data used in the present analyses were imputed up to the Haplotype Reference Consortium (HRC) panel. We applied additional quality control filters to exclude poorly imputed SNPs ( $r^2 < 0.4$ ) and SNPs with a MAF of <1%.

**Association analyses.** Using additive encoding of genotypes and adjusting for age, BMI, and/or the first five PCs, logistic regression analysis was performed in WGHS, NFBC, QIMR, and 23andMe cohorts and summary statistics were provided, including beta coefficients,  $\chi^2$  values, and standard errors, for genotyped and imputed SNPs. The UKBB association analyses were conducted using a linear mixed model (BOLT-LMM v.2.3.2)<sup>81</sup> adjusting for the two array types used, age and BMI (fixed effects), and a random effect accounting for relatedness between women. Effect size estimates ( $\beta$  and SE) from the linear mixed-model were converted to log-odds scale prior to meta-analysis. A fixed-effects, inverse-variance-weighted (IVW) meta-analysis on summary statistics was conducted using METAL<sup>82</sup> across all cohorts (Supplementary Data 6). A total of 8,662,096 SNPs were available from at least two of the five cohorts. A quantile-quantile plot of the results from meta-analysis across all GWAS cohorts is shown in Supplementary Fig. 1. Details on the overall genomic inflation factor and number of analyzed SNPs for each cohort are provided in Supplementary Table 2. For GWAS meta-analysis, independence of genetic association with UL was defined as SNPs in low LD ( $r^2 < 0.1$ ) with nearby ( $\leq 500$  kb) significantly associated SNPs. Individual loci correspond to regions of the genome containing all SNPs in LD ( $r^2 > 0.6$ ) with index SNPs. Any adjacent regions within 250 kb of one another were combined and classified as a single locus of association. All associated genomic regions were confirmed to have lead SNPs that were either directly genotyped or that met a rigorously high quality imputation threshold (INFO > 0.9) in at least two cohorts.

**Linkage disequilibrium score regression (LDSC).** Analysis of residual inflation in test statistics was conducted using univariate LDSC regression. Individual  $\chi^2$  values for each SNP analyzed in the GWAS meta-analysis was regressed onto LD scores estimated from the 1000 G EUR panel. Heritability calculations can be derived from analyzing the slope and y-axis intercept of the slope of the regression line. Percent impact of confounders, such as population stratification, on test statistic inflation are quantified as the LDSC ratio  $[(\text{intercept}-1)/(\text{mean } \chi^2-1)] \times 100\%$ . Remaining effects  $[(1-\text{LDSC ratio}) \times 100\%]$  represent the percentage of inflation attributed to polygenic heritability. Univariate LDSC regression was conducted using the LDSC software (<https://github.com/bulik/ldsc.git>). Adjustment of heritability ( $h^2$ ) calculations to the liability scale were performed by accounting for the prevalence of UL in the sample (~0.132) compared to the general population (~0.300). LDSC software was also used to estimate the genetic correlation between UL and endometriosis (Endo) using endometriosis GWA meta-analysis summary data from Sapkota et al.<sup>45</sup> consisting of only European cohorts. The heritability and LD score intercepts for both traits were computed, in this analysis with SNPs present in both datasets for LDSC regression again using LD scores from the 1000 G EUR panel. Genetic correlation between traits was estimated as the genetic covariance among SNPs  $/ \sqrt{h^2_{\text{UL}} \times h^2_{\text{Endo}}}$ .

**Approximate conditional analysis.** Approximate conditional analysis, implemented in GCTA<sup>83</sup>, was conducted to dissect distinct signals of association at each locus. Of note, where lead SNPs at adjacent loci mapped within 1 Mb of each other, loci were combined as a single region for conditional analysis, to account for potential LD between SNPs in different loci. GCTA makes use of meta-analysis association summary statistics (log-OR and corresponding standard error) and a reference panel of individual-level genotype data to obtain LD between all pairs of SNPs at a locus (or region) that approximates the covariance in effect estimates in a joint model. For these analyses, we made use of 5000 randomly selected white British women (of European descent) as reference. We used the -cojo-slc option to select index variants for each distinct association signal, at a locus-wide significance threshold of  $P < 10^{-5}$ , which is a conservative Bonferroni correction for the number of SNPs mapping to a locus. For loci with multiple distinct association signals, we obtained the conditional association summary statistics for each by conditioning on all other index SNPs at the locus (or region) using the -cojo-cond option.

**Fine-mapping distinct association signals.** For each distinct association signal, association summary statistics (log-OR and corresponding standard error) were extracted from the meta-analysis for all SNPs at the locus (or region). For loci with a single signal of association, we made use of association summary statistics from the unconditional meta-analysis. For loci with multiple signals of association, we made use of association summary statistics from the approximate conditional analysis. For each SNP  $j$ , we calculated an approximate Bayes' factor in favor of

association<sup>84</sup>, given by

$$\Lambda_j = \sqrt{\frac{V_j}{V_j + \omega}} \exp\left[\frac{\omega\beta_j^2}{2V_j(V_j + \omega)}\right], \quad (1)$$

where  $\beta_j$  and  $V_j$  denote the estimated log-OR and corresponding variance from the meta-analysis. The parameter  $\omega$  denotes the prior variance in allelic effects, taken here to be 0.04 for a disease outcome<sup>84</sup>. We then calculated the posterior probability,  $\pi_j$ , that the  $j$ th SNP is causal for the association signal, given by

$$\pi_{Cj} = \frac{\Lambda_j}{\sum_k \Lambda_k}, \quad (2)$$

where the summation is over all retained variants in the locus (or region). The 99% credible set for each signal was then constructed by: (i) ranking all variants according to their Bayes' factor,  $\Lambda_j$ ; and (ii) including ranked variants until their cumulative posterior probability of causality is at least 0.99.

**Heavy menstrual bleeding (HMB) GWAS.** The HMB GWAS was conducted using data from the UKBB cohort (Supplementary Methods). Both hospital-linked medical records and self-report were considered to identify women with a history of UL, while for HMB only hospital-linked medical records were taken into account. Controls had no previous history of either UL or HMB. Association analyses were performed using a linear mixed model (BOLT-LMM v.2.3.2)<sup>81</sup>. Effect size estimates ( $\beta$  and SE) from the linear mixed-model were converted to log-odds scale.

**Mendelian randomization (MR).** MR analyses were performed using the Two Sample Mendelian Randomization R package. GWAS summary statistics on HMB from the UKBB cohort were used to create outcome data for MR between UL (exposure) and HMB (outcome). To avoid overlap between samples in the exposure and outcome cohorts, we performed UL GWAS excluding all the HMB cases<sup>85</sup>. LD pruning was performed to confirm no duplication of exposure haplotypes or SNPs. Subsequently, data were harmonized to ensure the same reference alleles were used in exposure and outcome GWAS and that the variants were present in both GWAS datasets. Thirteen independent SNPs associated with UL from our GWAS meta-analysis were available in the HMB GWAS summary data to test for a causal effect of UL on HMB. There were too few significant SNPs available for HMB to test for a causal effect of HMB on UL.

GWAS summary statistics on endometriosis (with laparoscopy, without laparoscopy, and all self-reported endometriosis cases) from the WHS cohort were used to create outcome data for MR between UL (exposure) and endometriosis (outcome). To avoid overlap between samples in the exposure and outcome cohorts, WGHS was excluded from the UL GWAS for MR analysis. Twenty-two independent SNPs associated with UL were available in the endometriosis GWAS summary data to test for a causal effect of UL on endometriosis. For reverse causation model, summary statistics from seven GWAS listing 'endometriosis' as the phenotype of interest were available from the EMBL-EBI NHGRI GWAS catalog (Study Accession: GCST000797, GCST001894, GCST001720, GCST005906, GCST000916, GCST004549, GCST004873). Due to a low number of cases/controls or insufficient number of SNPs after LD pruning and data harmonizing, only one of the studies (GCST004549) was included in the analysis. Sixteen independent SNPs associated with endometriosis were available in our UL GWAS summary data to test for a causal effect of endometriosis on UL. The IVW model was used to test causality between exposure and outcome. In addition, the IVW (Q) method was used to test for heterogeneity, leave-one-out sensitivity analysis to identify the effect of individual SNPs, and MR Egger for horizontal pleiotropy. Due to heterogeneity in our initial MR estimates, we have now leveraged a similar approach to the one published in Corbin et al., 2016, to identify the minimum set of variants that when used as a genetic instrument eliminate heterogeneity<sup>86</sup>. We also conducted the MR-PRESSO test to identify and adjust for variants causing significant bias through horizontal pleiotropy<sup>87</sup>. MR-PRESSO method (1) applies a global test to evaluate whether horizontal pleiotropy is present, (2) calculates the causal estimates incorporating correction for the detected horizontal pleiotropy, and (3) applies a distortion test to evaluate if the causal estimate is significantly different after adjustment for outliers. We have reported the initial estimates along with the outlier-adjusted estimates as both the global and distortion tests showed significant results.

**Co-morbidity analyses.** Each cohort was analyzed individually with study-specific models chosen and covariates coded as appropriate for each cohort's data structure (Supplementary Methods). The study-specific effect estimates were combined using meta-analysis to obtain a summary RR. Between study heterogeneity was assessed with Cochran Q statistic and the  $I^2$  statistic<sup>88</sup>. Because heterogeneity among the studies was identified, we reported a random-effects IVW effect estimate based on the DerSimonian and Laird method<sup>89</sup>.

**LD Hub, gene-set, cell/tissue enrichment, and SMR analyses.** LD Hub analysis<sup>90</sup> was conducted using summary-level results data of UL GWAS meta-analysis to estimate the genetic correlation between UL and 21 different traits/diseases, including various reproductive traits and cardiometabolic traits/diseases that have

publicly available GWAS results on the LD Hub repository. Multiple-testing correction was performed ( $0.05/21 = 2.4 \times 10^{-3}$ ). For gene-set and cell/tissue enrichment, summary statistics from the set of 8971 SNPs with suggestive ( $P < 1 \times 10^{-5}$ ) or significant associations ( $P < 5 \times 10^{-8}$ ) were analyzed using the Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT) software<sup>65</sup>. Using the 1000 G EUR panel as a reference for LD calculations and the 'clumping' algorithm in PLINK<sup>91</sup>, we identified 104 independent loci at the suggestive threshold for DEPICT analyses (Supplementary Data 2). FDR  $< 0.05$  was considered statistically significant. For SMR analysis, SNPs present in at least two studies in the summary statistics were considered. The analysis was run using eQTL data from the CAGE blood dataset<sup>66</sup> and mQTLs from the LBC\_BSGS blood dataset<sup>67</sup>.

**FOXO1 immunohistochemistry and genotyping.** FOXO1 immunostaining was performed on two replicate tissue microarrays (TMAs) containing 335 UL and 36 myometrium tissue samples from 200 white women of European ancestry obtained from myomectomies and hysterectomies. Tissue cores on the replicate TMAs represent different regions of the same samples, which include corresponding tumor-normal tissue pairs from 35 women. Immunohistochemistry was carried out using the BOND staining system (Leica Biosystems, Buffalo Grove, IL) with a primary antibody dilution 1:100 (clone C29H4, Cell Signaling Technology, Danvers, MA) and hematoxylin as the counterstain. Immunostaining was analyzed using Aperio ImageScope software (Leica Biosystems). Each core was evaluated for the ratio of stain to counterstain taking into account variable cellularity between cores. Only nuclear labeling of the protein was evaluated. The average stain-to-counterstain ratio was compared between patient-matched UL and myometrium samples using a paired  $t$ -test (two-tailed), while an unpaired  $t$ -test (Welch's  $t$ -test, two-tailed) was applied to compare all UL and myometrium samples. Genomic DNA from 109 UL on the TMA was available for genotyping. These UL were genotyped for two SNPs with genome-wide significance at the 13q14.11 locus: rs6563799 and rs7986407. For each SNP, the average FOXO1 stain-to-counterstain ratio was compared across increasing dosage of the risk allele using a one-way analysis of variance test (two-tailed). We also performed an unpaired  $t$ -test to compare mean expression of UL homozygous for the risk variant against the other genotypes (Welch's  $t$ -test, two-tailed).  $P$ -values  $< 0.05$  were considered statistically significant.

**URLs.** For WHS see <http://whs.bwh.harvard.edu/>; for NFBC see <http://www.oulu.fi/nfbc/>; for QIMR see <http://www.qimberghofer.edu.au/>; for UK Biobank see <http://www.ukbiobank.ac.uk/>; for 23andMe see <https://research.23andme.com/>; for METAL see <http://csg.sph.umich.edu/abecasis/metal/>; for LDSC see <https://github.com/bulik/ldsc.git>; for DEPICT see <https://data.broadinstitute.org/mpg/depict/>; for SMR see <http://csgenomics.com/software/smr/>; and for PLINK see <http://pngu.mgh.harvard.edu/purcell/plink/>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The authors declare that the data supporting the findings of this study are available within the article and its Supplementary Information files. Summary statistics for the top 10,000 UL GWAS meta-analysis variants are provided in Supplementary Data 6. UL GWAS meta-analysis summary statistics (without 23andMe), UL GWAS limited by HMB and HMB GWAS summary statistics will be made available through the NHGRI-EBI GWAS Catalog <https://www.ebi.ac.uk/gwas/downloads/summary-statistics>. To request access to 23andMe GWAS summary statistics, please visit <https://research.23andme.com/dataset-access/>.

Received: 28 February 2019; Accepted: 10 September 2019;  
Published online: 24 October 2019

## References

- Stewart, E. A. Clinical practice. Uterine fibroids. *N. Engl. J. Med.* **372**, 1646–1655 (2015).
- Cramer, S. F. & Patel, A. The frequency of uterine leiomyomas. *Am. J. Clin. Pathol.* **94**, 435–438 (1990).
- Marino, J. L. et al. Uterine leiomyoma and menstrual cycle characteristics in a population-based cohort study. *Hum. Reprod.* **19**, 2350–2355 (2004).
- Pavone, D., Clemenza, S., Sorbi, F., Fambrini, M. & Petraglia, F. Epidemiology and risk factors of uterine fibroids. *Best Pr. Res Clin. Obstet. Gynaecol.* **46**, 3–11 (2018).
- Treloar, S. A., Martin, N. G., Dennerstein, L., Raphael, B. & Heath, A. C. Pathways to hysterectomy: Insights from longitudinal twin research. *Am. J. Obstet. Gynecol.* **167**, 82–88 (1992).



6. Vikhlyayeva, E. M., Khodzhaeva, Z. S. & Fantschenko, N. D. Familial predisposition to uterine leiomyomas. *Int. J. Gynecol. Obstet.* **51**, 127–131 (1995).
7. Marshall, L. M. et al. Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. *Obstet. Gynecol.* **90**, 967–973 (1997).
8. Luoto, R. et al. Heritability and risk factors of uterine fibroids—the Finnish Twin Cohort study. *Maturitas* **37**, 15–26 (2000).
9. Faerstein, E., Szkló, M. & Rosenshein, N. Risk factors for uterine leiomyoma: a practice-based case-control study. I. African-American heritage, reproductive history, body size, and smoking. *Am. J. Epidemiol.* **153**, 1–10 (2001).
10. Van Voorhis, B. J., Romitti, P. A. & Jones, M. P. Family history as a risk factor for development of uterine leiomyomas. Results of a pilot study. *J. Reprod. Med.* **47**, 663–669 (2002).
11. Cha, P. C. et al. A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nat. Genet.* **43**, 447–450 (2011).
12. Eggert, S. L. et al. Genome-wide linkage and association analyses implicate *FASN* in predisposition to uterine leiomyomata. *Am. J. Hum. Genet.* **91**, 621–628 (2012).
13. Gallagher, C. S. et al. Genome-wide association analysis identifies 27 novel loci associated with uterine leiomyomata revealing common genetic origins with endometriosis. Preprint at <https://www.biorxiv.org/content/10.1101/324905v1> (2018).
14. Rafnar, T. et al. Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. *Nat. Commun.* **9**, 3636 (2018).
15. Välimäki, N. et al. Genetic predisposition to uterine leiomyoma is determined by loci for genitourinary development and genome stability. *Elife* **7**, e37110 (2018).
16. Hellwege, J. N. et al. A multi-stage genome-wide association study of uterine fibroids in African Americans. *Hum. Genet.* **136**, 1363–1373 (2017).
17. Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nat. Rev. Cancer* **7**, 899–910 (2007).
18. Schoenberg Fejzo, M. et al. Translocation breakpoints upstream of the HMGIC gene in uterine leiomyomata suggest dysregulation of this gene by a mechanism different from that in lipomas. *Genes Chromosomes Cancer* **17**, 1–6 (1996).
19. Williams, A. J., Powell, W. L., Collins, T. & Morton, C. C. HMG(I)Y expression in human uterine leiomyomata. Involvement of another high-mobility group architectural factor in a benign neoplasm. *Am. J. Pathol.* **150**, 911–918 (1997).
20. Sornberger, K. S. et al. Expression of HMG(I)Y in three uterine leiomyomata with complex rearrangements of chromosome 6. *Cancer Genet. Cytogenet.* **114**, 9–16 (1999).
21. Chan, B. C. et al. BRE enhances in vivo growth of tumor cells. *Biochem Biophys. Res. Commun.* **326**, 268–273 (2005).
22. Ono, M. et al. Paracrine activation of WNT/ $\beta$ -catenin pathway in uterine leiomyoma stem cells promotes tumor growth. *Proc. Natl Acad. Sci. USA* **110**, 17053–17058 (2013).
23. Mehine, M. et al. Integrated data analysis reveals uterine leiomyoma subtypes with distinct driver pathways and biomarkers. *Proc. Natl Acad. Sci. USA* **113**, 1315–1320 (2016).
24. Shi, Y. et al. A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat. Genet.* **45**, 918–922 (2013).
25. Kuchenbaecker, K. B. et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat. Genet.* **47**, 164–171 (2015).
26. Phelan, C. M. et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* **49**, 680–691 (2017).
27. Haiman, C. A. et al. A common variant at the TERT-CLPTMIL locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.* **43**, 1210–1214 (2011).
28. Hamdi, Y. et al. Association of breast cancer risk in BRCA1 and BRCA2 mutation carriers with genetic variants showing differential allelic expression: identification of a modifier of breast cancer risk at locus 11q22.3. *Breast Cancer Res. Treat.* **161**, 117–134 (2017).
29. Shete, S. et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* **41**, 899–904 (2009).
30. Melin, B. S. et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* **49**, 789–794 (2017).
31. Figueroa, J. D. et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum. Mol. Genet.* **23**, 1387–1398 (2014).
32. Petersen, G. M. et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **42**, 224–228 (2010).
33. Wolpin, B. M. et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat. Genet.* **46**, 994–1000 (2014).
34. Zhang, M. et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget* **7**, 66328–66343 (2016).
35. Forbes, S. A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
36. Lutzmann, M. et al. MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Mol. Cell* **47**, 523–534 (2012).
37. He, C. et al. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.* **41**, 724–728 (2009).
38. AlAsiri, S. et al. Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J. Clin. Invest.* **125**, 258–262 (2015).
39. Stacey, S. N. et al. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).
40. Enciso-Mora, V. et al. Low penetrance susceptibility to glioma is caused by the TP53 variant rs78378222. *Br. J. Cancer* **108**, 2178–2185 (2013).
41. Diskin, S. J. et al. Rare variants in TP53 and susceptibility to neuroblastoma. *J. Natl Cancer Inst.* **106**, dju047 (2014).
42. Johnson, N. et al. Counting potentially functional variants in BRCA1, BRCA2 and ATM predicts breast cancer susceptibility. *Hum. Mol. Genet.* **16**, 1051–1057 (2007).
43. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
44. Kinnersley, B. et al. Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat. Commun.* **6**, 8559 (2015).
45. Sapkota, Y. et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat. Commun.* **8**, 15539 (2017).
46. Ruth, K. S. et al. Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur. J. Hum. Genet.* **24**, 284–290 (2016).
47. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
48. Uno, S. et al. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. *Nat. Genet.* **42**, 707–710 (2010).
49. Nyholt, D. R. et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat. Genet.* **44**, 1355–1359 (2012).
50. Albertsen, H. M., Chettier, R., Farrington, P. & Ward, K. Genome-wide association study link novel loci to endometriosis. *PLoS One* **8**, e58257 (2013).
51. Bulun, S. E. Endometriosis. *N. Engl. J. Med.* **360**, 268–279 (2009).
52. Biason-Lauber, A., Konrad, D., Navratil, F. & Schoenl, E. J. A WNT4 mutation associated with Mullerian-duct regression and virilization in a 46,XX woman. *N. Engl. J. Med.* **351**, 792–798 (2004).
53. Franco, H. L. et al. WNT4 is a key regulator of normal postnatal uterine development and progesterone signaling during embryo implantation and decidualization in the mouse. *FASEB J.* **25**, 1176–1187 (2011).
54. Powell, J. E. et al. Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339. *Hum. Mol. Genet.* **25**, 5046–5058 (2016).
55. Rae, J. M. et al. GREB1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res. Treat.* **92**, 141–149 (2005).
56. Rae, J. M. et al. GREB1 is a novel androgen-regulated gene required for prostate cancer growth. *Prostate* **66**, 886–894 (2006).
57. Bondesson, M., Hao, R., Lin, C. Y., Williams, C. & Gustafsson, J. A. Estrogen receptor signaling during vertebrate development. *Biochim Biophys. Acta* **1849**, 142–151 (2015).
58. Layman, L. C. et al. Delayed puberty and hypogonadism caused by mutations in the follicle-stimulating hormone beta-subunit gene. *N. Engl. J. Med.* **337**, 607–611 (1997).
59. Demeestere, I. et al. Follicle-stimulating hormone accelerates mouse oocyte development in vivo. *Biol. Reprod.* **87**, 1–11 (2012).
60. Missmer, S. A. & Cramer, D. W. The epidemiology of endometriosis. *Obstet. Gynecol. Clin. North Am.* **30**, 1–19 (2003).
61. Zondervan, K. T. et al. Endometriosis. *Nat. Rev. Dis. Prim.* **4**, 9 (2018).
62. Shafir, A. L. et al. Risk for and consequences of endometriosis: a critical epidemiologic review. *Best. Pr. Res. Clin. Obstet. Gynaecol.* **51**, 1–15 (2018).
63. Marshall, L. M. et al. A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata. *Fertil. Steril.* **70**, 432–439 (1998).
64. Uimari, O. et al. Uterine fibroids and cardiovascular risk. *Hum. Reprod.* **31**, 2689–2703 (2016).
65. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **19**, 5890 (2015).
66. Lloyd-Jones, L. R. et al. The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* **100**, 228–237 (2017).
67. McRae, A. et al. Identification of 55,000 Replicated DNA Methylation QTL. *Sci. Rep.* **8**, 17605 (2018).

68. Xing, Y. Q. et al. The regulation of FOXO1 and its role in disease progression. *Life Sci.* **193**, 124–131 (2018).
69. Jackson, J. G., Kreisberg, J. L., Koterba, A. P., Yee, D. & Brattain, M. G. Phosphorylation and nuclear exclusion of the forkhead transcription factor FKHR after epidermal growth factor treatment in human breast cancer cells. *Oncogene* **19**, 4574–4581 (2000).
70. Huang, H., Muddiman, D. C. & Tindall, D. J. Androgens negatively regulate forkhead transcription factor FKHR (FOXO1) through a proteolytic mechanism in prostate cancer cells. *J. Biol. Chem.* **279**, 13866–13877 (2004).
71. Goto, T. et al. Mechanism and functional consequences of loss of FOXO1 expression in endometrioid endometrial cancer cells. *Oncogene* **27**, 9–19 (2008).
72. Zhang, B., Gui, L. S., Zhao, X. L., Zhu, L. L. & Li, Q. W. FOXO1 is a tumor suppressor in cervical cancer. *GMR* **14**, 6605–6616 (2015).
73. Kovacs, K. A. et al. Involvement of FKHR (FOXO1) transcription factor in human uterus leiomyoma growth. *Fertil. Steril.* **94**, 1491–1495 (2010).
74. Lv, J. et al. Reduced expression of 14-3-3 gamma in uterine leiomyoma as identified by proteomics. *Fertil. Steril.* **90**, 1892–1898 (2008).
75. Shen, Q. et al. Overexpression of the 14-3-3gamma protein in uterine leiomyoma cells results in growth retardation and increased apoptosis. *Cell Signal* **45**, 43–53 (2018).
76. Hoekstra, A. V. et al. Progesterins activate the AKT pathway in leiomyoma cells and promote survival. *J. Clin. Endocrinol. Metab.* **94**, 1768–1774 (2009).
77. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
78. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
79. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
80. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
81. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
82. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
83. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
84. Wakefield, J. A. Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
85. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
86. Corbin, L. J. et al. BMI as a modifiable risk factor for type 2 diabetes: refining and understanding causal estimates using Mendelian randomization. *Diabetes* **65**, 3002–3007 (2016).
87. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018). Erratum in: *Nat Genet* **50**, 1196 (2018).
88. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
89. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin. Trials* **7**, 177–188 (1986).
90. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
91. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

## Acknowledgements

The authors thank all of the women and their families who participated in WGHS, NFBC, QIMR, UK Biobank, 23andMe, and NHSII, and acknowledge the Channing

Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School. This study was supported by the U.S. National Institutes of Health (NIH)/Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grant HD060530 to C.C.M. C.C.M. is also supported by the NIHR Manchester Biomedical Research Centre. N.M. acknowledges support from the Academy of Finland (295693) and Orion Research Foundation. H.R.H. is supported by NIH K22 CA193860. T.F. is supported by the NIHR Biomedical Research Centre, Oxford. S.E.M. is supported by the National Health and Medical Research Council (NHMRC) Fellowship Scheme (1103623). We thank the Specialized Histopathology Core of the Dana-Farber/Harvard Cancer Center for FOXO1 immunostaining. The Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support Grant P30 CA06516. Further acknowledgements are provided in Supplementary Note 1.

## Author contributions

C.S.G., S.A.M., K.T.Z. and C.C.M. designed the study. O.U., C.M.B., H.M., M.-R.J., J.E.B., S.E.M., D.R.N., P.A.L., J.N.P. and the 23andMe Research team contributed to phenotypic/clinical aspects of the cohorts. O.U., J.P.C., N.R., T.F., D.R.V.-E., T.L.E., F.D., V.K., P.M.R., S.D.G., S.E.M., G.W.M., D.R.N., D.A.H., J.Y.T., the 23andMe Research team, J.R.B.P., P.A.L., J.N.P., N.G.M., A.P.M., D.I.C. and K.T.Z. contributed to genotyping, quality control, imputation, and/or association analysis of the genotyping data. C.S.G. and N.R. performed the UL GWAS meta-analysis. N.R. conducted the HB GWAS. R.M.C., A.P.M. and D.I.C. provided statistical genetics advice. C.S.G., N.M., N.R., Z.R., S.M., G.W.M. and A.P.M. carried out or assisted with GWAS downstream analyses. C.S.G., H.R.H., O.U., N.S., N.R., K.L.T., J.E.B., S.A.M. and K.T.Z. contributed to large-scale epidemiologic analysis. N.M., C.S.G. and H.R.H. drafted the paper. G.W.M., N.G.M., A.P.M., D.I.C., S.A.M., K.T.Z. and C.C.M. provided critical comments on the paper, draft, and analysis. All authors read and approved the final paper.

## Competing interests

K.T.Z. and C.M.B. through Oxford University have research collaborations in benign gynecology with Bayer AG, Roche Diagnostics, Volition UK, and M DNA Life Sciences. D.A.H., J.Y.T., and members of the 23andMe Research Team are employees of 23andMe, Inc., and hold stock or stock options in 23andMe. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-019-12536-4>.

**Correspondence** and requests for materials should be addressed to N.M. or C.C.M.

**Peer review information** *Nature Communications* thanks Siddhartha Kar and Joellen Schildkraut for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019, corrected publication 2022

C.S. Gallagher<sup>1,30</sup>, N. Mäkinen<sup>2,30</sup>, H.R. Harris<sup>3,30</sup>, N. Rahmioglu<sup>4,30</sup>, O. Uimari<sup>5,6</sup>, J.P. Cook<sup>7</sup>, N. Shigeshi<sup>5</sup>, T. Ferreira<sup>4,8</sup>, D.R. Velez-Edwards<sup>9</sup>, T.L. Edwards<sup>10</sup>, S. Mortlock<sup>11</sup>, Z. Ruhioğlu<sup>12</sup>, F. Day<sup>12</sup>, C.M. Becker<sup>13</sup>, V. Karhunen<sup>13,14,15</sup>, H. Martikainen<sup>6</sup>, M.-R. Järvelin<sup>13,14,15,16,17</sup>, R.M. Cantor<sup>18</sup>, P.M. Ridker<sup>19</sup>, K.L. Terry<sup>20,21</sup>, J.E. Buring<sup>19</sup>, S.D. Gordon<sup>22</sup>, S.E. Medland<sup>23</sup>, G.W. Montgomery<sup>11,22</sup>, D.R. Nyholt<sup>22,24</sup>, D.A. Hinds<sup>25</sup>, J.Y. Tung<sup>25</sup>, the 23andMe Research Team, J.R.B. Perry<sup>12</sup>, P.A. Lind<sup>23</sup>, J.N. Painter<sup>23</sup>, N.G. Martin<sup>22</sup>, A.P. Morris<sup>4,7</sup>, D.I. Chasman<sup>19,31</sup>, S.A. Missmer<sup>21,26,31</sup>, K.T. Zondervan<sup>4,5,31</sup> & C.C. Morton<sup>2,27,28,29,31</sup>

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. <sup>2</sup>Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. <sup>3</sup>Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. <sup>4</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. <sup>5</sup>Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive Health, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. <sup>6</sup>Department of Obstetrics and Gynecology, Oulu University Hospital and PEDEGO Research Unit & Medical Research Center Oulu, University of Oulu and Oulu University Hospital, 90220 Oulu, Finland. <sup>7</sup>Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK. <sup>8</sup>Big Data Institute, Li Ka Shing Center for Health Information and Discovery, Oxford University, Oxford OX3 7LF, UK. <sup>9</sup>Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN 37203, USA. <sup>10</sup>Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37203, USA. <sup>11</sup>Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia. <sup>12</sup>MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. <sup>13</sup>Center for Life Course Health Research, Faculty of Medicine, University of Oulu, 90220 Oulu, Finland. <sup>14</sup>Unit of Primary Health Care, Oulu University Hospital, 90220 Oulu, Finland. <sup>15</sup>Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK. <sup>16</sup>Biocenter Oulu, University of Oulu, 90220 Oulu, Finland. <sup>17</sup>Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, Middlesex UB8 3PH, UK. <sup>18</sup>Department of Human Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90095, USA. <sup>19</sup>Division of Preventative Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>20</sup>Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. <sup>21</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. <sup>22</sup>Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. <sup>23</sup>Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia. <sup>24</sup>Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, QLD 4059, Australia. <sup>25</sup>23andMe, Mountain View, CA 94041, USA. <sup>26</sup>Department of Obstetrics, Gynecology, and Reproductive Biology, College of Human Medicine, Michigan State University, Grand Rapids, MI 49503, USA. <sup>27</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. <sup>28</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. <sup>29</sup>Manchester Centre for Audiology and Deafness, Manchester Academic Health Science Center, University of Manchester, Manchester M13 9PL, UK. <sup>30</sup>These authors contributed equally: C.S. Gallagher, N. Mäkinen, H.R. Harris, N. Rahmioglu. <sup>31</sup>These authors jointly supervised this work: D.I. Chasman, S.A. Missmer, K.T. Zondervan, C.C. Morton. A full list of consortium members appears at the end of the paper.

### the 23andMe Research Team

Michelle Agee<sup>25</sup>, Babak Alipanahi<sup>25</sup>, Adam Auton<sup>25</sup>, Robert K. Bell<sup>25</sup>, Katarzyna Bryc<sup>25</sup>, Sarah L. Elson<sup>25</sup>, Pierre Fontanillas<sup>25</sup>, Nicholas A. Furlotte<sup>25</sup>, Karen E. Huber<sup>25</sup>, Aaron Kleinman<sup>25</sup>, Nadia K. Litterman<sup>25</sup>, Matthew H. McIntyre<sup>25</sup>, Joanna L. Mountain<sup>25</sup>, Elizabeth S. Noblin<sup>25</sup>, Carrie A.M. Northover<sup>25</sup>, Steven J. Pitts<sup>25</sup>, J. Fah Sathirapongsasuti<sup>25</sup>, Olga V. Sazonova<sup>25</sup>, Janie F. Shelton<sup>25</sup>, Suyash Shringarpure<sup>25</sup>, Chao Tian<sup>25</sup>, Vladimir Vacic<sup>25</sup> & Catherine H. Wilson<sup>25</sup>