Genome-wide association analysis identifies 27 novel loci associated with uterine leiomyomata revealing common genetic origins with endometriosis

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

1) Department of Genetics, Harvard Medical School, Boston, MA, USA;

2) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA;

3) Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

4) Oxford Endometriosis CaRe Centre, Nuffield Department of Women's and Reproductive Health, University of Oxford, John Radcliffe Hospital, Oxford, UK;

5) Department of Obstetrics and Gynecology, Oulu University Hospital, Oulu, Finland, PEDEGO Research Unit, University of Oulu and Oulu University Hospital, Oulu, Finland, Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland;

6) Department of Biostatistics, University of Liverpool, Block F Waterhouse Building, Liverpool, UK;

7) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK;

8) Big Data Institute, Li Ka Shing Center for Health for Health Information and Discovery, Oxford University, Oxford, UK;

9) Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA;

10) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA;

11) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK;

12) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK;

13) Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland;

14) Unit of Primary Health Care, Oulu University Hospital, Oulu, Finland;

15) Biocenter Oulu, University of Oulu, Oulu, Finland;

16) Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Uxbridge, Middlesex, UK;

17) Department of Human Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA;

18) Division of Preventative Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA;

19) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia;

20) Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia;

21) Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia;

22) Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia;

23) 23andMe, Mountain View, CA, USA;

24) Division of Adolescent and Young Adult Medicine, Department of Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA;

25) Department of Obstetrics, Gynecology, and Reproductive Biology, College of Human Medicine, Michigan State University, Grand Rapids, MI, USA;

26) Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA;

27) Broad Institute of MIT and Harvard, Cambridge, MA, USA;

28) School of Psychological Sciences, University of Manchester, Manchester, UK

* Equal contribution

† Shared last authors

‡ Correspondence and material requests: cmorton@bwh.harvard.edu

Uterine leiomyomata (UL), also known as uterine fibroids, are the most common neoplasms of the reproductive tract and the primary cause for hysterectomy, leading to considerable impact on women's lives as well as high economic burden[1,2]. Genetic epidemiologic studies indicate that heritable risk factors contribute to UL pathogenesis[3]. Previous genome-wide association studies (GWAS) identified five loci associated with UL at genome-wide significance ($P < 5 \times 10^{-8}$)[4-6]. We conducted GWAS meta-analysis in 20,406 cases and 223,918 female controls of white European ancestry, identifying 24 genome-wide significant independent loci; 17 replicated in an unrelated cohort of 15,068 additional cases and 43,587 female controls. Aggregation of discovery and replication studies (35,474 cases and 267,505 female controls) revealed six additional significant loci. Interestingly, four of the 17 loci identified and replicated in these analyses have also been associated with risk for endometriosis – another common gynecologic disorder. These findings increase our understanding of the biological mechanisms underlying UL development, and suggest overlapping genetic origins with endometriosis.

UL are hormone-driven tumors that occur in 70-80% of all women by age 50[7]. Although the majority of UL are asymptomatic, about 25% of women with UL are symptomatic, and may experience excessive bleeding, abdominal pain, and infertility[2]. 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, and increased body mass index (BMI)[3]. Studies on familial aggregation and twins, as well as racial differences in prevalence and morbidity, suggest heritable factors influence the risk for developing UL[8-13]. To date, previous GWAS including up to approximately 1,600 cases have identified five

3

loci significantly associated ($P < 5$ x 10^{-8}) with UL: 10q24.33, 11p15.5 and 22q13.1 in Japanese women[4], 17q25.3 in white women of European ancestry[5], and a distinct region at 22q13.1 in African American women[6].

We conducted the largest discovery meta-analysis of GWAS on UL to date including four population-based cohorts of white European ancestry, increasing the case sample size almost 13-fold compared to previous studies (20,406 cases and 223,918 female controls): Women's Genome Health Study (WGHS), Northern Finnish Birth Cohort (NFBC), QIMR Berghofer Medical Research Institute (QIMR), and The UK Biobank **(Supplementary Notes, Supplementary Table 1)**. Ancestry in each cohort was verified by principal component analysis (PCA) of genotype data. UL phenotype in each cohort was analyzed in a generalized linear regression model assuming additive genetic effects with multivariate adjustment for age and BMI, and 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 meta-analysis. Altogether 8,292,290 biallelic SNPs were analyzed and adjustments for genomic inflation were performed **(Supplementary Fig. 1a, Supplementary Table 2)**. We identify a total of 1,167 SNPs across 24 loci with genome-wide significance ($P < 5$ x 10^{-8}) **(Supplementary Table 3)**. The Manhattan plot is shown in **Fig. 1**.

Replication of the peak SNPs was pursued in an independent direct-to-consumer cohort obtained from 23andMe, including 15,068 self-reported UL cases and 43,587 self-reported female controls of white European ancestry **(Supplementary Notes, Supplementary Fig. 1b, Supplementary Tables 1 and 2)**. We replicate associations at 17 of the 24 (71%) loci identified in our discovery-

phase meta-analysis (significance threshold; $P < 2.08$ x 10^{-3}) **(Table 1)**. Among independently replicated loci are all three loci previously reported to be associated with UL in Japanese women: 10q24.33 (rs9419958, odds ratio (OR) = 1.09, $P = 1.20$ x 10^{-9}), 11p15.5 (rs547025, OR = 1.15, $P = 4.77$ x 10^{-13}), and 22q13.1 (rs12484776, OR = 1.09, $P = 4.08$ x 10^{-12})[4].

Interestingly, significant association signals are also observed at several loci previously associated with endometriosis: 1p36.12 (rs223552, OR = 1.14, $P = 7.39$ x 10^{-21}), 2p25.1 (rs10929757, OR = 1.07, $P = 1.19$ x 10^{-12}), 6q25.2 (rs58415480, OR = 1.18, $P = 3.65$ x 10^{-32}), and 11p14.1 (rs11031006, OR = 1.08, $P = 2.91$ x 10^{-8})[14-17]. Endometriosis is another common hormone-dependent disease that affects reproductive-aged women, resulting from ectopic growth of endometrial tissue outside the uterine cavity[18]. Although functional studies of relevant tissue need to confirm the consequences of the variants in regulation of gene expression, each of the overlapping genomic loci contain 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[19,20]. Recently, SNPs increasing endometriosis risk at 1p36.12 have been suggested to act through *CDC42*, a gene which encodes a small GTPase of the Rho family [21]. *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[22,23]. *ESR1* at 6q25.2 encodes the alpha subunit of the ligand-activated nuclear ER that regulates cell proliferation in the uterus[24]. *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[25,26].

A number of replicated loci also harbor genes previously implicated in cell growth and cancer risk in different tissue types, including cervical cancer[27], epithelial ovarian cancer[28,29], breast cancer[30,31], glioma[32,33], bladder cancer[34], and pancreatic cancer[35-37]. Specifically, seven of the independent loci contain well-characterized oncogenes and tumor suppressor genes from the Cancer Gene Census list in COSMIC[38]: *PDGFRA, TERT, ESR1, WT1, ATM, FOXO1*, and *TP53*. FOXO1 is a transcription factor that plays an important role in cell proliferation, apoptosis, DNA repair, and stress response[39]. Inactivation of *FOXO1* promotes cell proliferation and tumorigenesis in several hormone-regulated malignancies, such as prostate, breast, cervical, and endometrial cancers[40-43]. We quantified and compared nuclear FOXO1 expression in 335 UL and 35 patient-matched normal myometrial samples using immunohistochemistry on tissue microarrays. In the patient-matched tumor-normal pairs, nuclear FOXO1 expression was 1.69-fold higher ($P = 0.01$) in UL (**Supplementary Fig. 2a**). When all 335 UL were taken into account, expression was increased as much as 2.32-fold ($P = 1.52 \times 10^{-9}$) in UL compared to myometrial samples (**Supplementary Fig. 2b**). A total of 109 patient samples were genotyped for rs6563799 and rs7986407 – two of the peak SNPs residing in the *FOXO1* locus. Stratification of samples by genotype reveals 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 Fig. 3a**). An increase in FOXO1 levels of UL with the rs7986407 risk allele is also observed; however, the change is not statistically significant (**Supplementary Fig. 3b**). Our results are consistent with a previous study reporting elevated FOXO1 expression in UL compared to matched myometrium[44], especially levels of phosphorylated (p) FOXO1 (pSer[256]). Normally, p-FOXO1 interacts with 14-3-3γ protein in the nucleus, resulting in translocation of FOXO1 to the cytoplasm and its transcriptional inactivation. Interestingly, Kovacs et al. showed p-FOXO1 to be

predominantly present in the nucleus in UL, but sequestered in the cytoplasm of myometrium[44]. The concomitant increase of p-FOXO1 and reduced expression of 14-3-3γ in UL has been suggested to lead to impaired nuclear/cytoplasmic shuttling of p-FOXO1, which promotes cell survival[44-46].

Next, we conducted a meta-analysis on a total of 8,602,260 SNPs across all cohorts, including a total of 35,474 UL cases and 267,505 controls (**Supplementary Fig. 1c, Supplementary Table 2**). Through linkage disequilibrium score (LDSC) regression analysis, an estimated 89.5% of the genomic inflation factor (λ_{GC}) of 1.12 was attributable to polygenic heritability (intercept = 1.02, s.e. = 0.0081). We observe genome-wide significant associations ($P < 5$ x 10^{-8}) at 2,045 SNPs across 27 independent loci (**Supplementary Fig. 4, Supplementary Table 4**). The Manhattan plot is shown in **Fig. 2**. In addition to 21 loci identified at the discovery stage, we observe six novel loci significantly associated with UL (**Table 2**). Among the novel, 'discovery' loci are three genes of interest: *HMGA1*, *BABAM2*, and *WNT2*. *HMGA1* is a member of the high mobility group proteins and is involved in regulation of gene transcription[47]. Somatic rearrangements of *HMGA1* at 6p21 have been recurrently documented in UL, albeit at a much lower frequency than those of *HMGA2* – another member of the protein family[48-50]. *BABAM2* encodes a death receptor-associating intracellular protein that promotes tumor growth by suppressing apoptosis[51]. Associations at the locus containing *WNT2* together with associations observed at the *WNT4* locus reinforce a possible role for Wnt signaling in UL. Overall, individual SNP-based heritability (h^2) was estimated to be relatively low, 0.0281 (s.e. = 0.0029) on the liability scale.

Gene-set and tissue enrichment analyses across 5,185 SNPs with suggestive ($P < 1$ x 10^{-5}) or significant ($P < 5$ x 10^{-8}) UL associations using DEPICT[52] reveal significant enrichments (false discovery rate (FDR) < 0.05) in gene sets, such as steroid hormone receptor (GO:0035258; $P =$ 8.09 x 10^{-6}), hormone receptor binding (GO:0051427; $P = 1.49$ x 10^{-4}), and nuclear hormone receptor binding (GO:0035257; $P = 9.71$ x 10^{-5}) **(Supplementary Tables 5 and 6)**. The results are concordant with the hormone-driven nature of UL. We also observe enrichment of genes associated with expression in female urogenital tissue **($P = 6.19$ x 10^{-4}) (Supplementary Fig. 5)**. To identify SNPs with likely regulatory function, we selected up to 30 of the most significant SNPs from each of the 27 loci identified in the meta-analysis of GWAS across all cohorts. Altogether 429 of 597 SNPs (72%) were present in the RegulomeDB[53], and 23 of these have a score < 3, indicating potential involvement in gene regulation **(Supplementary Table 7)**. Based on the RegulomeDB, two SNPs (rs498217 at 11p15.5 and rs1641528 at 17p13.1) are indicated as potential expression quantitative trait loci (eQTLs) in monocytes for *SCGB1C1* and *CD68*, respectively[54].

In summary, our GWA analyses uncovered 27 novel genomic loci associated with UL in women of white European ancestry. Many of the candidates fall into two categories: (1) genomic regions containing characterized tumor suppressors and oncogenes, and (2) genes involved in hormone signaling pathways previously associated with endometriosis. Biological overlap between two highly common gynecologic diseases, due to similarities in molecular mechanisms and progenitor cells, has long been suspected. Further characterization of the mutual pathogenic mechanisms has the capacity to direct 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.

Online Methods

Subjects

Four population-based cohorts (WGHS, NFBC, QIMR, and UK Biobank) and one direct-to-consumer cohort (23andMe) from the FibroGENE consortium were included in the study (**Supplementary Table 1**), resulting in 35,474 UL cases and 267,505 female controls of white European ancestry. Detailed descriptions of patient cohorts and sample selection are available in **Supplementary Notes**. For the current study, 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. All participants provided an informed consent in accordance with the processes approved by the relevant jurisdiction for human subject research for each cohort.

Genotyping

For the GWAS discovery stage, several different Illumina-based genotyping platforms (Illumina Inc., San Diego, CA, USA) were used: HumanHap300 Duo'+' chips or the combination of the Human-Hap300 Duo and iSelect chips (WGHS), Infinium 370cnvDuo array (NFBC), 317K, 370K, or 610K SNP platforms (QIMR). Genotyping of participants in the UK Biobank was performed either on the Affymetrix UK BiLEVE or Affymetrix UK Biobank Axiom® 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 (1000G EUR) Phase 3. Imputation was carried out using ShapeIt2 and IMPUTE2 softwares[55,56]. SNPs with call rates of < 99%, and SNPs showing 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[57]. 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. The UK Biobank data QC and imputation was handled by a dedicated team headed by the Wellcome Centre for Human Genetics, prior to public release of the data. 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 the first five principal components from EIGENSTRAT, WGHS, NFBC, QIMR, and 23andMe cohorts in both the discovery and replication stage performed logistic regression analysis and provided summary statistics, including beta coefficients, χ^2 values, and standard errors, for the genotyped and imputed SNPs. The UK Biobank association analyses were conducted using a linear mixed model (BOLT-LMM v.2.3.2)[58] adjusting for the two array types used, age and BMI (fixed effects) and a random effect adjusting for relatedness between women. Effect size estimates (β and SE) from the linear

10

mixed-model were converted to log-odds scale prior to meta-analysis. Two fixed-effects, inverse variance weighted meta-analyses on summary statistics were conducted using METAL[59], one in the discovery stage and the other across all cohorts. For discovery-phase meta-analysis, 8,292,290 biallelic SNPs for which data were available in at least two of the four cohorts were analyzed. A total of 8,602,260 SNPs were available from at least two of the five cohorts for the meta-analysis across all cohorts. Quantile-quantile plots of the results from meta-analysis of population-based cohorts and across all GWAS cohorts are 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.** Independence of genetic association with UL was defined as SNPs in low linkage disequilibrium (LD; $r^2 < 0.1$) with nearby ($\leq$ 500kb) 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.

FOXO1 immunohistochemistry and genotyping

FOXO1 immunostaining was performed on two replicate tissue microarrays (TMAs) containing 335 UL and 35 patient-matched 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 34 women. Immunohistochemistry was carried out using the BOND staining system (Leica Biosystems, Buffalo Grove, IL) with a primary antibody dilution of 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

11

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.

Linkage disequilibrium score regression (LDSC)

Analysis of residual inflation in test statistics was conducted using univariate LDSC regression. Individual χ^2 values for each SNP analyzed in the GWAS meta-analysis were regressed onto LD scores estimated from the 1000G 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 [((intercept – 1)) / ((mean χ^2 – 1))] * 100%. Remaining effects [(1 – LDSC ratio) * 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).

Gene-set and tissue enrichment analyses

Summary statistics from the set of 5,185 SNPs with suggestive ($P < 1$ x 10^{-5}) or significant associations ($P < 5$ x 10^{-8}) were analyzed for gene-set and tissue enrichment using the Data-driven Expression-Prioritized Integration for Complex Traits (DEPICT) software[52]. Using the 1000G EUR panel as a reference for LD calculations and the 'clumping' algorithm in PLINK[60], we identified 162 independent loci at the suggestive threshold for DEPICT analyses **(Supplementary Table 5)**. FDR < 0.05 was considered statistically significant.

URLs

WHS, http://whs.bwh.harvard.edu/; NFBC, http://www.oulu.fi/nfbc/; QIMR, http://www.qimrberghofer.edu.au/; UK Biobank, http://www.ukbiobank.ac.uk/; 23andMe, https://research.23andme.com/; METAL, http://csg.sph.umich.edu/abecasis/metal/; LDSC, https://github.com/bulik/ldsc.git; DEPICT, https://data.broadinstitute.org/mpg/depict/; RegulomeDB, http://www.regulomedb.org/; PLINK, http://pngu.mgh.harvard.edu/purcell/plink/

Supplementary Material

Supplementary Material includes Supplementary Notes, five figures, and seven tables.

Acknowledgements

Competing Interests statement

References

1.   Soliman, A. M., Yang, H., Du, E. X., Kelkar, S. S. & Winkel, C. The direct and indirect costs of uterine fibroid tumors: a systematic review of the literature between 2000 and 2013. *Am J Obstet Gynecol* **213,** 141-160 (2015).

2.   Stewart, E. A. Clinical practice. Uterine fibroids. *N Engl J Med* **372,** 1646-1655 (2015).

3.   Pavone, D., Clemenza, S., Sorbi, F., Fambrini, M. & Petraglia, F. Epidemiology and risk factors of uterine fibroids. *Best Pract Res Clin Obstet Gynaecol* **46,** 3-11 (2018).

4.   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).

5.   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).

6.   Hellwege, J. N. et al. A multi-stage genome-wide association study of uterine fibroids in African Americans. *Hum Genet.* **136,** 1363-1373 (2017).

7.   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 Obstet Gynecol* **188,** 100-107 (2003).

8.   Vikhlyaeva, E. M., Khodzhaeva, Z. S. & Fantschenko, N. D. Familial predisposition to uterine leiomyomas. *Int J Gynecol Obstet* **51,** 127-131 (1995).

9.   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).

10.   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).

11.   Luoto, R. et al. Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study. *Maturitas* **37,** 15-26 (2000).

12.   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).

13.   Faerstein, E., Szklo, 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).

14.   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).

15.   Albertsen, H. M., Chettier, R., Farrington, P. & Ward, K. Genome-wide association study link novel loci to endometriosis. *PloS One* **8,** e58257 (2013).

16.   Nyholt, D. R. et al. Genome-wide association meta-analysis identifies new endometriosis risk loci. *Nat Genet* **44,** 1355-1359 (2012).

17.   Sapkota, Y. et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nature Commun* **8,** 15539 (2017).

18.   Bulun, S. E. Endometriosis. *N Engl J Med* **360,** 268-279 (2009).

19.   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).

20.   Biason-Lauber, A., Konrad, D., Navratil, F. & Schoenle, 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).

21.   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).

22.   Rae, J. M. et al. GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast Cancer Res Treat* **92,** 141-149 (2005).

23.   Rae, J. M., et al. GREB1 is a novel androgen-regulated gene required for prostate cancer growth. *The Prostate* **66,** 886-894 (2006).

24.   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).

25.   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).

26.   Demeestere, I. et al. Follicle-stimulating hormone accelerates mouse oocyte development in vivo. *Biol Reprod.* **87,** 1-11 (2012).

27.   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).

28.   Kuchenbaecker, K. B. et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet* **47,** 164-171 (2015).

29.   Phelan, C. M. et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet* **49,** 680-691 (2017).

30.   Haiman, C. A. et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* **43,** 1210-1214 (2011).

31.   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).

32.   Shete, S. et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41,** 899-904 (2009).

33.   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).

34.   Figueroa, J. D. et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum Mol Genet* **23,** 1387-1398 (2014).

35.   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).

36.   Wolpin, B. M. et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet* **46,** 994-1000 (2014).

37.   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).

38.   Forbes, S. A. et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39,** D945-50 (2011).

39.   Xing, Y. Q. et al. The regulation of FOXO1 and its role in disease progression. *Life Sci* **193,** 124-131 (2018).

40.   Jackson, J. G., Kreisberg, J. I., 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).

41.   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).

42.   Goto, T. et al. Mechanism and functional consequences of loss of FOXO1 expression in endometrioid endometrial cancer cells. *Oncogene* **27,** 9-19 (2008).

43.   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).

44.   Kovacs, K. A. et al. Involvement of FKHR (FOXO1) transcription factor in human uterus leiomyoma growth. *Fertil Steril* **94,** 1491-1495 (2010).

45.   Lv, J. et al. Reduced expression of 14-3-3 gamma in uterine leiomyoma as identified by proteomics. *Fertil Steril* **90,** 1892-1898 (2008).

46.   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).

47.   Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nat Rev Cancer* **7,** 899-910 (2007).

48.   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).

49.   Williams, A. J., Powell, W. L., Collins, T. & Morton, C. C. HMGI(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).

50.   Sornberger, K. S. et al. Expression of HMGIY in three uterine leiomyomata with complex rearrangements of chromosome 6. *Cancer Genet Cytogenet* **114,** 9-16 (1999).

51.   Chan, B. C. et al. BRE enhances in vivo growth of tumor cells. *Biochem Biophys Res Commun* **326,** 268-273 (2005).

52. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **19,** 5890 (2015).

53. Boyle, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22,** 1790-1797 (2012).

54. Zeller, T. et al. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5,** e10693 (2010).

55. 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).

56. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9,** 179-181 (2011).

57. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38,** 904-909 (2006).

58. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47,** 284-290 (2015).

59. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190-2191 (2010).

60. 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).
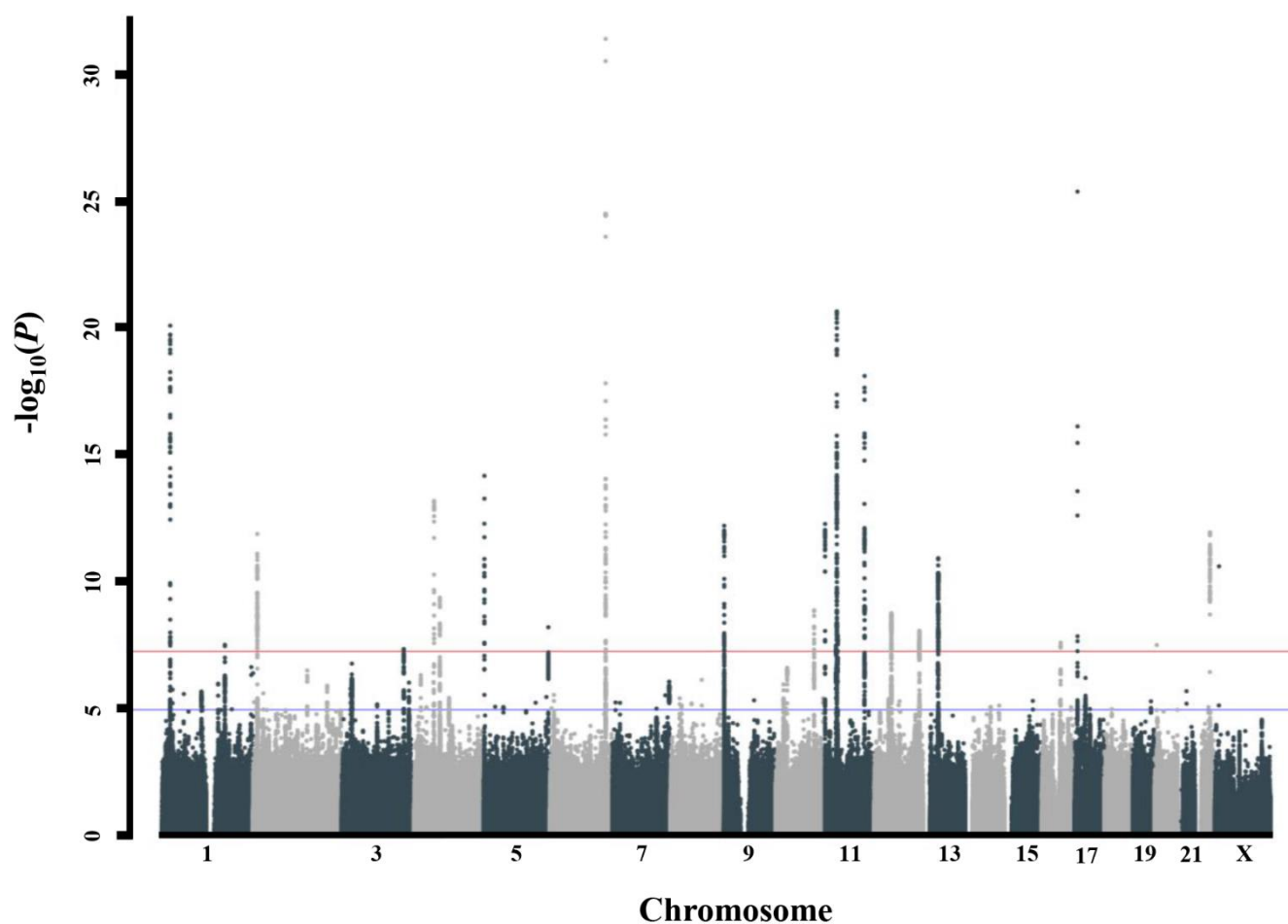
Fig. 1 Manhattan plot for discovery-phase meta-analysis. Meta-analysis of GWAS across 244,324 women of white European ancestry conducted in population-based cohorts identifies 24 independent loci associated with UL. Red and blue horizontal lines indicate genome-wide significant ($P < 5 \times 10^{-8}$) and suggestive ($P < 1 \times 10^{-5}$) thresholds.
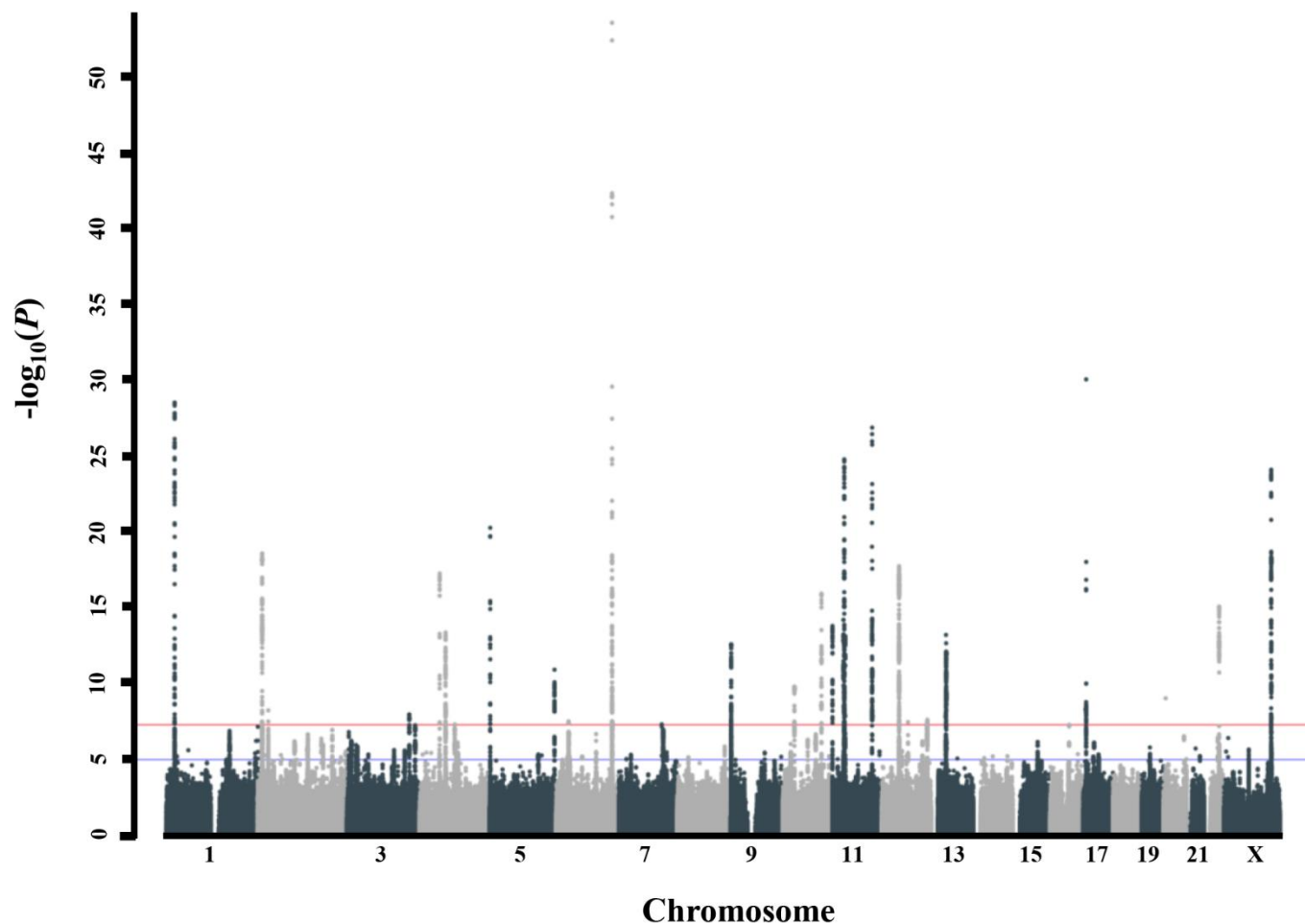
Fig. 2 Manhattan plot for meta-analysis across all cohorts from the FibroGENE consortium.

Meta-analysis of GWAS across 302,979 women of white European ancestry across all cohorts identifies 27 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.

Tables

Table 1. Seventeen loci identified in discovery-phase GWAS meta-analysis and replicated in an independent cohort from 23andMe

Locus	rsID	P_{Disc}	OR (95% CI)	$P_{23andMe}$	OR (95% CI)	Genes of interest[a]
1p36.12[b]	rs2235529	$7.39e^{-21}$	1.14 (1.11 - 1.17)	$2.55e^{-10}$	1.13 (1.09 - 1.17)	*WNT4, CDC42*
2p25.1[b]	rs10929757	$1.19e^{-12}$	1.07 (1.05 - 1.10)	$9.35e^{-07}$	1.07 (1.04 - 1.10)	*GREB1*
4q12	rs4864806	$5.95e^{-14}$	1.16 (1.12 - 1.21)	$3.27e^{-05}$	1.11 (1.06 - 1.17)	*LNX1, PDGFRA*
4q13.3	rs12640488	$3.83e^{-10}$	1.06 (1.04 - 1.09)	$1.38e^{-05}$	1.06 (1.03 - 1.09)	*SULT1B1*
5p15.33	rs2242652	$4.68e^{-13}$	1.10 (1.07 - 1.13)	$2.41e^{-09}$	1.12 (1.08 - 1.16)	*TERT*
5q35.2	rs2456181	$5.62e^{-09}$	1.07 (1.05 - 1.09)	$2.43e^{-04}$	1.05 (1.02 - 1.08)	*ZNF346, UIMC1*
6q25.2[b]	rs58415480	$3.65e^{-32}$	1.18 (1.15 - 1.21)	$2.31e^{-25}$	1.22 (1.18 - 1.27)	*SYNE1, ESR1*
10q24.3[c]	rs9419958	$1.20e^{-09}$	1.09 (1.06 - 1.12)	$4.58e^{-09}$	1.12 (1.08 - 1.17)	*OBFC1, SLK*
11p15.5[c]	rs547025	$4.77e^{-13}$	1.15 (1.11 – 1.19)	$1.47e^{-03}$	1.09 (1.03 - 1.14)	*RIC8A, BET1L*
11p14.1[b]	rs11031006	$2.91e^{-08}$	1.08 (1.05 - 1.12)	$4.23e^{-09}$	1.12 (1.08 - 1.17)	*FSHB, ARL14EP*
11p13	rs11031731	$2.04e^{-21}$	1.14 (1.11 - 1.17)	$4.36e^{-06}$	1.09 (1.05 - 1.13)	*WT1*
11p13	rs2553772	$1.20e^{-08}$	1.06 (1.04 - 1.08)	$8.75e^{-07}$	1.07 (1.04 - 1.10)	*PDHX, CD44*
11q22.3	rs149934734	$7.05e^{-19}$	1.34 (1.26 - 1.43)	$1.39e^{-10}$	1.31 (1.21 -1.42)	*C11orf65, KDELC2*
12q13.11	rs2131371	$1.63e^{-09}$	1.07 (1.05 - 1.09)	$5.60e^{-12}$	1.10 (1.08 - 1.14)	*SLC38A2*
13q14.11	rs1986649	$2.52e^{-10}$	1.31 (1.21 - 1.43)	$3.68e^{-05}$	1.27 (1.14 - 1.43)	*FOXO1*
17p13.1	rs78378222	$3.84e^{-26}$	1.64 (1.50 - 1.80)	$2.21e^{-07}$	1.38 (1.22 - 1.56)	*SHBG, TP53*
22q13.1[c]	rs12484776	$4.08e^{-12}$	1.09 (1.06 - 1.12)	$4.22e^{-05}$	1.07 (1.04 - 1.11)	*TNRC6B*

P_{Disc}, *P*-value in discovery-phase meta-analysis; $P_{23andMe}$, *P*-value in 23andMe cohort; OR, odds ratio; CI, confidence interval
Threshold for significance in replication set was $P < 2.08 \times 10^{-3}$.
[a] ≤ 300 kb distant from association signal
[b] Loci previously associated with endometriosis[15-18]
[c] Loci previously associated with UL[4]

Table 2. Six novel genome-wide significant loci identified in meta-analysis across all cohorts

Locus	rsID	P_{Meta}	OR (95% CI)	Genes of interest[a]
2p23.2	rs55819434	$5.59e^{-09}$	1.09 (1.06 – 1.12)	*BABAM2*
4q22.3	rs4699299	$4.72e^{-08}$	1.05 (1.03 – 1.07)	*PDLIM5*
6p21.31	rs116251328	$2.95e^{-08}$	1.15 (1.09 – 1.21)	*GRM4, HMGA1*
7q31.2	rs2270206	$4.64e^{-08}$	1.06 (1.04 – 1.09)	*WNT2*
10p11.22	rs10508765	$1.51e^{-10}$	1.07 (1.05 – 1.09)	*ZEB1, ARHGAP12*
12q15	rs11178393	$3.34e^{-08}$	1.08 (1.05 – 1.10)	*PTPRR*
OR, odds ratio; CI, confidence interval [a] $\leq$ 300 kb distant from association signal				