

Exploring potential causal genes for uterine leiomyomas: A summary data-based Mendelian randomization and FUMA analysis

Yuxin Dai¹, Xudong Liu², Yining Zhu³, Su Mao¹, Jingyun Yang^{4,5}, Lan Zhu¹

¹Department of Obstetrics and Gynecology, State Key Laboratory of Complex, Severe and Rare Diseases, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

²Medical Science Research Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China

³School of Mathematical Sciences, Fudan University, 220 Handan Road, Yangpu District, Shanghai, China

⁴Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

⁵Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

Corresponding to

Lan Zhu, MD

Department of Obstetrics and Gynecology

State Key Laboratory of Complex, Severe and Rare Diseases

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

National Clinical Research Center for Obstetric & Gynecologic Diseases

Peking Union Medical College Hospital

Chinese Academy of Medical Sciences and Peking Union Medical College

No.1 Shuaifuyuan Road

Dongcheng District, Beijing 100730

China

Tel: 86-21-69156204

E-mail: zhu_julie@vip.sina.com

Abstract

Objective: To explore potential causal genetic variants and genes underlying the pathogenesis of uterine leiomyomas (ULs).

Methods: We conducted the summary data-based Mendelian randomization (SMR) analysis and performed functional mapping and annotation using FUMA to examine genetic variants and genes that are potentially involved in the pathogenesis of ULs.

Both analyses used summarized data of a recent genome-wide association study (GWAS) on ULs, which has a total sample size of 244,324 (20,406 cases and 223,918 controls). For the SMR analysis, we performed separate analysis using CAGE and GTEx eQTL data.

Results: Using the CAGE eQTL data, our SMR analysis identified 13 probes tagging 10 unique genes that were pleiotropically/potentially causally associated with ULs, with the top three probes being ILMN_1675156 (tagging *CDC42*, $P_{\text{SMR}}=8.03\times 10^{-9}$), ILMN_1705330 (tagging *CDC42*, $P_{\text{SMR}}=1.02\times 10^{-7}$) and ILMN_2343048 (tagging *ABCB9*, $P_{\text{SMR}}=9.37\times 10^{-7}$). Using GTEx eQTL data, our SMR analysis did not identify any significant genes after correction for multiple testing. FUMA analysis identified 106 independent SNPs, 24 genomic loci and 137 genes that are potentially involved in the pathogenesis of ULs, seven of which were also identified by the SMR analysis.

Conclusions: We identified many genetic variants, genes, and genomic loci that are potentially involved in the pathogenesis of ULs. More studies are needed to explore the exact underlying mechanisms in the etiology of ULs.

Keywords: uterine leiomyomas; expression quantitative trait loci; summary

Mendelian randomization; genome-wide association study; functional mapping

Introduction

Uterine leiomyomas (ULs), also called myomas or uterine fibroids, are benign tumors in the smooth muscle tissue in myometrium[1, 2]. The overall prevalence of UL is about 70% in women of reproductive age, and approximately 25% of UL patients suffer from apparent clinical symptoms and require treatment[3]. ULs are the most prevalent benign tumor in female reproductive tract and the leading indication for hysterectomy. ULs represent a major cause of morbidity in women of childbearing age and account for excessive menstrual bleeding, pelvic pain or pressure, infertility, and pregnancy complications[4]. To date, the only definitive treatment for ULs, including the familial subtype, is hysterectomy, which creates a great challenge if fertility preservation is desired. ULs also cause tremendous economic burden. For example, the annual cost of ULs in the US alone, including direct medical costs and indirect financial losses, is estimated to be up to \$34.4 billion, higher than the combined cost of breast and colon cancer[5].

UL is a complex, multi-factorial gynecological benign disease with highly variable tumor size, tumor location and clinical manifestations. Many factors have been reported to be associated with the risk of ULs, including biological, demographic, reproductive and lifestyle factors[6-8]. Furthermore, previous studies also suggested that genetics plays an important role in the pathogenesis of ULs. For example, African-American women, or generally women with African origin, are more predisposed to develop ULs, with a prevalence as high as 80% [9], suggesting ethnicity-specific factors, potentially ethnicity-specific genetic structure, underlying

the pathogenesis of ULs. Familial clustering between first-degree relatives and twins was also observed as well as multiple inherited syndromes in which fibroid development occurred[10, 11]. Moreover, many GWAS and candidate gene studies have identified several genetic variants/loci associated with the susceptibility of ULs[12-17]. However, the role of putative risk factors and the underlying biological mechanisms underpinning ULs remain largely unclear, which has contributed to the slow progress in the development of effective treatment options for ULs. More studies are needed to explore genetic variants/genes that are potentially causally associated with ULs to better understand the pathogenesis of ULs.

Mendelian randomization (MR) uses genetic variants as the proxy to randomization. Recently, it has been widely adopted to explore pleiotropic/potentially causal effect of an exposure on the outcome (e.g., ULs)[18]. Confounding and reverse causation, which are commonly encountered in traditional association studies, can be greatly reduced by MR. This method has been successful in identifying gene expression probes or DNA methylation loci that are pleiotropically/potentially causally associated with various phenotypes, such as neuropathologies of Alzheimer's disease and severity of COVID-19[19, 20] .

In this paper, we attempted to prioritize genes that are potentially causally associated with ULs through a summary data-based MR (SMR) approach. We also performed functional mapping and annotation to further explore genetic variants and genomic loci that are potentially involved in the pathogenesis of ULs.

Methods

GWAS data for ULs

The GWAS summarized data for ULs were provided by a recent genome-wide association meta-analysis of ULs[14]. The results were based on meta-analyses of ULs using data from four population-based cohorts (Women's Genome Health Study, UK Biobank, Queensland Institute of Medical Research, and North Finnish Birth Cohort), with a total sample size of 244,324 (20,406 cases and 223,918 controls). Genotyping was done on different platforms, and imputation was performed using the reference panel from the 1000 Genomes Project European dataset (1000G EUR) Phase 3 or the Haplotype Reference Consortium (HRC) panel. For each cohort, logistic regression or linear mixed model association analysis was done, assuming an additive genetic model and adjusting for age, BMI, and/or the first five principal components, and/or array type, as appropriate. The GWAS summarized data can be downloaded at

http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST009001-GCST010000/GCST009158/.

eQTL data

The SMR analysis used cis-eQTL genetic variants as the instrumental variables (IVs) for gene expression. We performed separate SMR analysis using eQTL data from two sources. Specifically, we used the V7 release of the GTEx eQTL summarized data for uterus[21], which included 70 participants, and the CAGE eQTL summarized data for whole blood[22], which included 2,765 participants. The eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

SMR analysis

The SMR analysis was done using the software SMR. Detailed information regarding the SMR method was reported elsewhere[28]. The existence of linkage in the observed association was assessed using the heterogeneity in dependent instruments (HEIDI) test. $P_{\text{HEIDI}} < 0.05$ means rejection of the null hypothesis. That is, the observed association could be due to two distinct genetic variants in high linkage disequilibrium with each other. We adopted the default settings in SMR (e.g., $P_{\text{eQTL}} < 5 \times 10^{-8}$ and minor allele frequency [MAF] > 0.01) and used false discovery rate (FDR) to adjust for multiple testing. The SMR analytic process is illustrated in **Figure 1**.

FUMA analysis

To better understand the genetic mechanisms underlying ULs, we also conducted a FUMA analysis to functionally map and annotate the genetic association, again using the GWAS summary results of ULs. FUMA is an on-line platform that integrates information from multiple resources for easy implementation of post-GWAS analysis, such as functional annotation and gene prioritization[23]. It has two processes, SNP2GENE, which annotate SNPs regarding their biological functions and map them to genes, and GENE2FUNC, which annotates the mapped genes in biological contexts. In SNP2GENE, we performed both positional mapping and eQTL mapping using GTEx v8 of whole blood and uterus. We selected all types of genes in gene prioritization and adopted the default settings otherwise (e.g., maximum P-value of lead SNPs being 5×10^{-8} and r^2 threshold for independent significant SNPs being 0.6). In GENE2FUNC, we adopted the default settings (e.g., using FDR to correct for

multiple testing in the gene-set enrichment analysis).

Data cleaning and statistical/bioinformatical analysis was performed using R version 4.1.2 (<https://www.r-project.org/>), SMR (<https://cnsgenomics.com/software/smr/>), and FUMA (<https://fuma.ctglab.nl/>).

Results

Basic information of the summarized data

In the SMR analysis, the CAGE eQTL has a much larger number of participants than that of the GTEx eQTL data (2,765 vs. 70), so is the number of eligible probes (8,523 vs. 999). After checking allele frequencies among the datasets and LD pruning, there were more than 6 million eligible SNPs in each SMR analysis. In the FUMA analysis, about 8.6 million SNPs were used as the input. The detailed information was shown in

Table 1.

Pleiotropic association with UL

Using the CAGE eQTL data, our SMR analysis identified 13 probes tagging 10 unique genes that were pleiotropically/potentially causally associated with ULs, with the top three probes being ILMN_1675156 (tagging *CDC42*, $P_{\text{SMR}}=8.03\times 10^{-9}$), ILMN_1705330 (tagging *CDC42*, $P_{\text{SMR}}=1.02\times 10^{-7}$) and ILMN_2343048 (tagging *ABCB9*, $P_{\text{SMR}}=9.37\times 10^{-7}$; **Table 2**). There were three probes tagging *CDC42* (**Figure 2**) and two probes tagging *ABCB9* (**Figure 3**) that showed significant pleiotropic association with ULs. Using GTEx eQTL data, we did not identify any genes that were pleiotropically/potentially causally associated with ULs after correction for multiple testing (**Table 2**).

Functional mapping and annotation

FUMA analysis identified 106 independent significant SNPs, 33 lead SNPs (**Table S1-S3**), and 24 genomic risk loci (**Figure 4; Table S4**). In addition, FUMA identified 137 genes that are potentially involved in the pathogenesis of ULs (**Table S5**). These 137 genes are distributed in 20 genomic risk loci, with four genomic risk loci containing no identified genes (**Figure 4 & Table S5**). Of the 137 identified genes, 7 were also identified by SMR analysis, including *CDC42*, *SLC38A1*, *ABCB9*, *MPHOSPH9*, *SBNO1*, *MRPS31* and *CD68*. Expression of the prioritized genes in 30 tissues can be found in **Table S6** and **Figure S1**.

Gene-set enrichment analysis (GSEA) was undertaken to test the possible biological mechanisms of the 137 candidate genes implicated in ULs (**Table S7**). A total of 96 gene sets with an adjusted $P < 0.05$ were identified. We found enrichment signals related with uterine fibroids (adjusted $P=2.00 \times 10^{-51}$). In addition, we also found enrichment of sex-related signals such as GO_REGULATION_OF_GONAD_DEVELOPMENT (adjusted $P=0.023$), endometriosis (adjusted $P=7.3 \times 10^{-4}$), sex hormone-binding globulin levels (adjusted $P=0.003$), and sex hormone levels (adjusted $P=0.024$; **Table S7**).

Discussion

In this study, we conducted SMR and FUMA analysis to prioritize SNPs and genes to better understand the genetic mechanisms underlying ULs. We identified multiple genetic variants, genes, genomic risk loci and gene sets that may be involved in the pathogenesis of ULs. These findings provided helpful leads to a better understanding

of the pathogenesis of ULs and highlight potential therapeutic targets for the treatment of ULs.

Several probes tagging *CDC42* (cell division control protein 42 homolog) showed significant pleiotropic association with ULs using CAGE eQTL data (**Table 2**) in the SMR analysis using CAGE eQTL data. This gene was also identified by the FUMA analysis. *CDC42* is a member of the Rho family and is implicated in a variety of cellular functions including cell cycle progression, survival, transcription, actin cytoskeleton organization, and membrane trafficking[24]. *CDC42* has been linked to multiple human cancers and is involved in the initiation of many cellular responses during oncogenic processes, such as transition from epithelial to mesenchymal, cell-cycle progression, migration/invasion, tumor growth, angiogenesis, and oncogenic transformation[25, 26]. Several studies reported that *CDC42* might also play an essential role in the pathogenesis of fibroid. For example, genome-wide analysis revealed that the 1p36.12 region, where *CDC42/WNT4* is located, was associated with uterine fibroids[13, 14, 16]. Interestingly, the genetic variant rs10917151 in *CDC42/WNT4* seems to have ancestry-specific effect on the risk of uterine fibroids. Specifically, the A allele was associated with a reduced risk of uterine fibroids in women of African ancestry (OR=0.84) and an increased risk in women of European ancestry (OR=1.16)[13]. Since rs10917151 has been reported to be involved in hormone-related traits (e.g., endometriosis and endometrial cancer), it probably plays a role in the development of leiomyomas via influencing hormone metabolism[16]. Meanwhile, another study showed that the deregulation of *CDC42* influences

fibroblasts activation which is essential in the pathogenesis of ULs[27]. Given the fact that increased cellular proliferation is present in fibroid compared with the adjacent uterine tissue and the function of *CDC42* in influencing cell cycle, further investigation is needed to elucidate the role of *CDC42* in the development of leiomyoma and the potential of this gene as a promising target for the prevention and treatment of ULs.

We also found that two probes tagging *ABCB9* (ATP binding cassette subfamily B member 9) showed significant pleotropic association with ULs in the SMR analysis using CAGE eQTL data. This gene was also identified in the FUMA analysis. *ABCB9* belongs to the superfamily of ATP-binding cassette (ABC) transporters which fulfill diverse physiological functions in different cellular localizations ranging from the plasma membrane to intracellular membranous compartments[28]. *ABCB9*, located on 12q24.31, is an antigen processing-like (TAPL) transporter that has been found to be involved in the development and progression of various malignant tumors, such as ovarian cancer and non-small cell lung cancer[29, 30]. The genetic variant rs2270788 in *ABCB9* was found to be associated with both the risk and tumor size of ULs in African American participants[31]. *ABCB9* was downregulated in women with a high level of progesterone serum (>1.5 ng/ml), compared with women with a lower level of progesterone serum (<1.5 ng/ml)[32]. Since progesterone is a major promoter of leiomyoma development and growth[33], the role of *ABCB9* in fibroids in general, and its function in progesterone-driven growth of leiomyomas in particular, needs further exploration.

The recent GWAS study on ULs also performed two-sample Mendelian randomization analysis[14]. However, the MR analysis was different from our SMR analyses in that their objective was to examine the causality of genetic association between UL and heavy menstrual bleeding (HMB). Their study identified 29 independent loci for ULs, with 27 of them on the autosomal chromosomes while our FUMA analysis identified a total of 24 genomic loci on the autosomal chromosomes. The definition of genomic locus was different between their approach and FUMA: their genomic locus was defined as regions of the genome containing all SNPs in LD ($r^2 > 0.6$) with the index SNPs (independent SNPs, i.e., SNPs in low LD ($r^2 < 0.1$) with nearby (≤ 500 kb) significantly associated SNPs), with any adjacent regions within 250 kb of one another being combined and classified as a single locus. In FUMA, independent significant SNPs ($P < 5 \times 10^{-8}$ and independent from each other at $r^2 < 0.6$) were first identified. Then, all known SNPs in LD ($r^2 > 0.6$) with one of the independent SNPs were included, using the pre-calculated LD structure based on 1000G. As a results, SNPs that were not originally in the GWAS results could also be included. This may partly explain the difference in the findings.

Our study has limitations. The GWAS analysis was done in participants of European ancestry. As such, our findings might not be generalized to other ethnicities. The number of eligible probes used in SMR analysis was limited, especially in the analysis using GTEx eQTL. As a result, we could not rule out the possibility of missing some important genes that were not tagged in the eQTL data. In addition, the FDR approach to correct for multiple testing resulted in additional possibilities of

missing important genes. The HEIDI test was significant for some of the observed associations, implying the existence of horizontal pleiotropy (**Table 2**).

Conclusions

We identified many genetic variants, genes and genomic loci that are potentially involved in the pathogenesis of ULs. More studies are needed to explore the underlying mechanisms in the etiology of ULs.

Declarations

Ethics approval and consent to participate

This study used only publicly available data. As a result, ethics approval and consent to participate is not needed.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are publicly available as specified in the methods section of this paper. Specifically, the eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>, and the GWAS summarized data can be downloaded at http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST009001-GCST010000/GCST009158/.

Competing interests

No potential conflicts of interest were disclosed by the authors.

Funding

The study was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (NO. 2020-PT320-003). Dr. Jingyun Yang's research was supported by NIH/NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG033678, R01AG36042, U01AG61356, and 1RF1AG064312-01.

Authors' contributions

The manuscript was written through contributions of all authors. YD, JY, and LZ designed the study. YD, YZ and JY analyzed data and performed data interpretation. YD, XL and JY wrote the initial draft, and SM, JY and LD contributed to writing subsequent versions of the manuscript. All authors reviewed the study findings and read and approved the final version before submission.

Acknowledgements

Not applicable

Authors' information

Affiliations:

Department of Obstetrics and Gynecology, State Key Laboratory of Complex, Severe and Rare Diseases, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Yuxin Dai, Su Mao & Lan Zhu

**Medical Science Research Center, State Key Laboratory of Complex Severe and
Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of
Medical Science and Peking Union Medical College, Beijing, China**

Xudong Liu

**School of Mathematical Sciences, Fudan University, 220 Handan Road, Yangpu
District, Shanghai, China**

Yining Zhu

**Rush Alzheimer's Disease Center & Department of Neurological Sciences, Rush
University Medical Center, Chicago, IL, USA**

Jingyun Yang

References

1. Bulun SE: **Uterine fibroids**. *N Engl J Med* 2013, **369**(14):1344-1355.
2. Stewart EA, Laughlin-Tommaso SK, Catherino WH, Lalitkumar S, Gupta D, Vollenhoven B: **Uterine fibroids**. *Nat Rev Dis Primers* 2016, **2**:16043.
3. Stewart EA, Cookson CL, Gandolfo RA, Schulze-Rath R: **Epidemiology of uterine fibroids: a systematic review**. *BJOG* 2017, **124**(10):1501-1512.
4. Commandeur AE, Styer AK, Teixeira JM: **Epidemiological and genetic clues for molecular mechanisms involved in uterine leiomyoma development and growth**. *Hum Reprod Update* 2015, **21**(5):593-615.
5. Cardozo ER, Clark AD, Banks NK, Henne MB, Stegmann BJ, Segars JH: **The estimated annual cost of uterine leiomyomata in the United States**. *Am J Obstet Gynecol* 2012, **206**(3):211 e211-219.
6. Parazzini F: **Risk factors for clinically diagnosed uterine fibroids in women around menopause**. *Maturitas* 2006, **55**(2):174-179.
7. Wise LA, Laughlin-Tommaso SK: **Epidemiology of Uterine Fibroids: From Menarche to Menopause**. *Clin Obstet Gynecol* 2016, **59**(1):2-24.
8. Sparic R, Mirkovic L, Malvasi A, Tinelli A: **Epidemiology of Uterine Myomas: A Review**. *Int J Fertil Steril* 2016, **9**(4):424-435.
9. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM: **High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence**. *Am J Obstet Gynecol* 2003, **188**(1):100-107.
10. Luoto R, Kaprio J, Rutanen EM, Taipale P, Perola M, Koskenvuo M: **Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study**. *Maturitas* 2000, **37**(1):15-26.
11. Tomlinson IP, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, Leigh I, Gorman P, Lamlum H, Rahman S *et al*: **Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer**. *Nat Genet* 2002, **30**(4):406-410.
12. Cha PC, Takahashi A, Hosono N, Low SK, Kamatani N, Kubo M, Nakamura Y: **A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids**. *Nat Genet* 2011, **43**(5):447-450.
13. Edwards TL, Giri A, Hellwege JN, Hartmann KE, Stewart EA, Jeff JM, Bray MJ, Pendergrass SA, Torstenson ES, Keaton JM *et al*: **A Trans-Ethnic Genome-Wide Association Study of Uterine Fibroids**. *Front Genet* 2019, **10**:511.
14. Gallagher CS, Makinen N, Harris HR, Rahmioglu N, Uimari O, Cook JP, Shigesu N, Ferreira T, Velez-Edwards DR, Edwards TL *et al*: **Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis**. *Nat Commun* 2019, **10**(1):4857.
15. Hellwege JN, Jeff JM, Wise LA, Gallagher CS, Wellons M, Hartmann KE, Jones SF, Torstenson ES, Dickinson S, Ruiz-Narvaez EA *et al*: **A multi-stage genome-wide association study of uterine fibroids in African Americans**. *Hum Genet* 2017, **136**(10):1363-1373.
16. Rafnar T, Gunnarsson B, Stefansson OA, Sulem P, Ingason A, Frigge ML, Stefansdottir L, Sigurdsson JK, Tragante V, Steinthorsdottir V *et al*: **Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits**. *Nat Commun* 2018, **9**(1):3636.

17. Valimaki N, Kuisma H, Pasanen A, Heikinheimo O, Sjoberg J, Butzow R, Sarvilinna N, Heinonen HR, Tolvanen J, Bramante S *et al*: **Genetic predisposition to uterine leiomyoma is determined by loci for genitourinary development and genome stability.** *Elife* 2018, **7**.
18. Davey Smith G, Hemani G: **Mendelian randomization: genetic anchors for causal inference in epidemiological studies.** *Hum Mol Genet* 2014, **23**(R1):R89-98.
19. Liu D, Wang Y, Jing H, Meng Q, Yang J: **Mendelian randomization integrating GWAS and mQTL data identified novel pleiotropic DNA methylation loci for neuropathology of Alzheimer's disease.** *Neurobiol Aging* 2021, **97**:18-27.
20. Liu D, Yang J, Feng B, Lu W, Zhao C, Li L: **Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19.** *J Infect* 2021, **82**(1):126-132.
21. GTEx Consortium: **The GTEx Consortium atlas of genetic regulatory effects across human tissues.** *Science* 2020, **369**(6509):1318-1330.
22. Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, Zeng B, Bakshi A, Metspalu A, Dermitzakis M *et al*: **The Genetic Architecture of Gene Expression in Peripheral Blood.** *Am J Hum Genet* 2017, **100**(2):371.
23. Watanabe K, Taskesen E, van Bochoven A, Posthuma D: **Functional mapping and annotation of genetic associations with FUMA.** *Nat Commun* 2017, **8**(1):1826.
24. Huang QY, Lai XN, Qian XL, Lv LC, Li J, Duan J, Xiao XH, Xiong LX: **Cdc42: A Novel Regulator of Insulin Secretion and Diabetes-Associated Diseases.** *Int J Mol Sci* 2019, **20**(1).
25. Qadir MI, Parveen A, Ali M: **Cdc42: Role in Cancer Management.** *Chem Biol Drug Des* 2015, **86**(4):432-439.
26. Maldonado MDM, Dharmawardhane S: **Targeting Rac and Cdc42 GTPases in Cancer.** *Cancer Res* 2018, **78**(12):3101-3111.
27. Zheng LH, Cai FF, Ge I, Biskup E, Cheng ZP: **Stromal fibroblast activation and their potential association with uterine fibroids (Review).** *Oncol Lett* 2014, **8**(2):479-486.
28. Linton KJ: **Structure and function of ABC transporters.** *Physiology (Bethesda)* 2007, **22**:122-130.
29. Ween MP, Armstrong MA, Oehler MK, Ricciardelli C: **The role of ABC transporters in ovarian cancer progression and chemoresistance.** *Crit Rev Oncol Hematol* 2015, **96**(2):220-256.
30. Dong Z, Zhong Z, Yang L, Wang S, Gong Z: **MicroRNA-31 inhibits cisplatin-induced apoptosis in non-small cell lung cancer cells by regulating the drug transporter ABCB9.** *Cancer Lett* 2014, **343**(2):249-257.
31. Aissani B, Zhang K, Wiener H: **Follow-up to genome-wide linkage and admixture mapping studies implicates components of the extracellular matrix in susceptibility to and size of uterine fibroids.** *Fertil Steril* 2015, **103**(2):528-534 e513.
32. Labarta E, Martinez-Conejero JA, Alama P, Horcajadas JA, Pellicer A, Simon C, Bosch E: **Endometrial receptivity is affected in women with high circulating progesterone levels at the end of the follicular phase: a functional genomics analysis.** *Hum Reprod* 2011, **26**(7):1813-1825.
33. Kim JJ, Sefton EC: **The role of progesterone signaling in the pathogenesis of uterine leiomyoma.** *Mol Cell Endocrinol* 2012, **358**(2):223-231.

Figure legends

Figure 1. Flow chart for the SMR analysis.

A) SMR analysis using CAGE eQTL data from blood; and B) SMR analysis using GTEx eQTL data

eQTL, expression quantitative trait loci; GWAS, genome-wide association studies; LD, linkage disequilibrium; SMR, summary data-based Mendelian randomization; SNP, single nucleotide polymorphism

Figure 2. Pleiotropic association of *CDC42* with ULs using CAGE eQTL.

Top plot, grey dots represent the $-\log_{10}(P \text{ values})$ for SNPs from the GWAS of ULs, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

eQTL, expression quantitative trait loci; GWAS, genome-wide association studies; HEIDI, heterogeneity in dependent instruments; SMR, summary data-based Mendelian randomization; SNP, single nucleotide polymorphism; ULs, uterine leiomyomas

Figure 3. Pleiotropic association of *ABCB9* and *MPHOSPH9* with ULs using CAGE eQTL data.

Top plot, grey dots represent the $-\log_{10}(P \text{ values})$ for SNPs from the GWAS of ULs, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

eQTL, expression quantitative trait loci; GWAS, genome-wide association studies; HEIDI, heterogeneity in dependent instruments; SMR, summary data-based Mendelian randomization; SNP, single nucleotide polymorphism; ULs, uterine leiomyomas

Figure 4. Genetic risk loci identified by FUMA analysis using GWAS data on ULs.

Genomic risk loci are displayed in the format of ‘chromosome:start position-end

position' on the Y axis. For each genomic locus, histograms from left to right depict the size, the number of candidate SNPs, the number of mapped genes (using positional mapping and eQTL mapping), and the number of genes known to be located within the genomic locus, respectively.

eQTL, expression quantitative trait loci; GWAS, genome-wide association studies;

SNP, single nucleotide polymorphism; ULs, uterine leiomyomas

Table 1. Basic information of the eQTL and GWAS data.

Data Source	Total # of participants	Number of eligible genetic variants or probes
eQTL data		
CAGE	2,765	8,523
GTE_x	70	999
GWAS data for SMR analysis		
WGHS	3,375/9,465	-
NFBC	363/5,000	-
QIMR	1,484/3,701	-
UKBB	15,184/205,752	-
Total	20,406/223,918	CAGE: 6,198,856; GTE _x : 6,697,624
GWAS data for FUMA analysis	20,406/223,918	8,589,006

GWAS: genome-wide association studies; QTL, quantitative trait loci; WGHS, Women's Genome Health Study; QIMR, Queensland Institute of Medical Research; UKBB, UK Biobank, NFBC, North Finnish Birth Cohort

Table 2. The top hit probes identified in SMR analysis*.

eQTL data	Probe	Gene	CHR	Top SNP	P _{eQTL}	P _{GWAS}	Beta	SE	P _{SMR}	P _{HEIDI}	Q value
Uterus	ENSG00000198496.6	<i>NBR2</i>	17	rs2292595	4.34×10 ⁻²⁷	0.0002	0.0359	0.0100	0.0003	0.1417	0.2211
	ENSG00000155393.8	<i>HEATR3</i>	16	rs11642695	3.78×10 ⁻¹²	4.66×10 ⁻⁵	0.0871	0.0248	0.0004	0.5173	0.2211
	ENSG00000101751.6	<i>POLI</i>	18	rs4940321	1.31×10 ⁻⁸	5.47×10 ⁻⁵	-0.0960	0.0291	0.0010	0.4576	0.3252
	ENSG00000164535.10	<i>DAGLB</i>	7	rs13235365	5.58×10 ⁻¹³	0.0004	0.0404	0.0126	0.0014	0.9852	0.3505
	ENSG00000226752.3	<i>PSMD5-AS1</i>	9	rs4837796	3.37×10 ⁻²⁸	0.0031	0.0265	0.0093	0.0043	0.8854	0.7269
	ENSG00000164048.9	<i>ZNF589</i>	3	rs11718329	2.06×10 ⁻¹⁰	0.0019	0.0419	0.0150	0.0051	0.3164	0.7269
	ENSG00000164045.7	<i>CDC25A</i>	3	rs4511915	3.80×10 ⁻¹³	0.0035	0.0314	0.0115	0.0065	0.1342	0.7269
	ENSG00000188878.12	<i>FBF1</i>	17	rs9674908	1.20×10 ⁻⁸	0.0022	-0.0578	0.0213	0.0068	0.6307	0.7269
	ENSG0000027001.7	<i>MIPEP</i>	13	rs75783226	1.95×10 ⁻⁹	0.0025	0.0498	0.0185	0.0070	0.1324	0.7269
ENSG00000229759.1	<i>MRPS18API</i>	3	rs11130163	7.36×10 ⁻¹⁵	0.0052	0.0260	0.0099	0.0085	0.6590	0.7269	
Whole blood	ILMN_1675156	<i>CDC42</i>	1	rs2473290	7.00×10 ⁻¹¹⁸	2.82×10 ⁻⁹	0.0896	0.0155	8.03×10 ⁻⁹	1.56×10 ⁻⁸	6.84×10⁻⁵
	ILMN_1705330	<i>CDC42</i>	1	rs2473290	2.08×10 ⁻³²	2.82×10 ⁻⁹	0.1778	0.0334	1.02×10 ⁻⁷	3.83×10 ⁻⁵	0.0004
	ILMN_2343048	<i>ABCB9</i>	12	rs4148856	1.17×10 ⁻²³	1.96×10 ⁻⁸	-0.2064	0.0421	9.37×10 ⁻⁷	0.3185	0.0027
	ILMN_2343047	<i>ABCB9</i>	12	rs641760	9.79×10 ⁻¹⁷	7.64×10 ⁻⁹	-0.2509	0.0528	2.01×10 ⁻⁶	0.1106	0.0043

ILMN_1654421	<i>MPHOSPH9</i>	12	rs10772996	7.72×10^{-19}	6.34×10^{-8}	-0.2501	0.0541	3.73×10^{-6}	0.9425	0.0064
ILMN_1767642	<i>C11orf46</i>	11	rs12364889	1.81×10^{-34}	8.32×10^{-7}	0.1410	0.0309	5.10×10^{-6}	0.0977	0.0072
ILMN_2266948	<i>SLC38A1</i>	12	rs11183420	5.41×10^{-36}	2.33×10^{-6}	0.1401	0.0316	9.36×10^{-6}	0.0048	0.0107
ILMN_1691188	<i>UIMC1</i>	5	rs353491	4.49×10^{-17}	2.28×10^{-7}	0.2164	0.0490	1.00×10^{-5}	0.1816	0.0107
ILMN_2359907	<i>CD68</i>	17	rs56319762	2.76×10^{-51}	5.07×10^{-6}	0.1093	0.0250	1.26×10^{-5}	0.0157	0.0119
ILMN_1654552	<i>MRPS31</i>	13	rs7324090	5.69×10^{-10}	1.44×10^{-9}	-0.3700	0.0853	1.44×10^{-5}	0.3289	0.0122
ILMN_1706531	<i>ABCC5</i>	3	rs4074672	3.24×10^{-169}	1.61×10^{-5}	0.0552	0.0129	1.94×10^{-5}	0.2197	0.0150
ILMN_1738424	<i>CDC42</i>	1	rs2038106	7.64×10^{-30}	4.44×10^{-6}	0.1430	0.0337	2.21×10^{-5}	1.23×10^{-9}	0.0157
ILMN_1739943	<i>SBNO1</i>	12	rs1569068	1.30×10^{-10}	2.92×10^{-8}	0.3387	0.0807	2.73×10^{-5}	0.8568	0.0179

*We showed the top ten pleiotropic association for the SMR analysis using GTEx eQTL data. And all the significant pleiotropic association, after correction of multiple testing using FDR, in the SMR analysis using CAGE eQTL data. The GWAS summarized data were provided by the study of Gallagher et al. and can be downloaded at

http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST009001-GCST010000/GCST009158/. The CAGE and GTEx eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

P_{eQTL} is the P-value of the top associated cis-eQTL in the eQTL analysis, and P_{GWAS} is the P-value for the top associated cis-eQTL in the GWAS

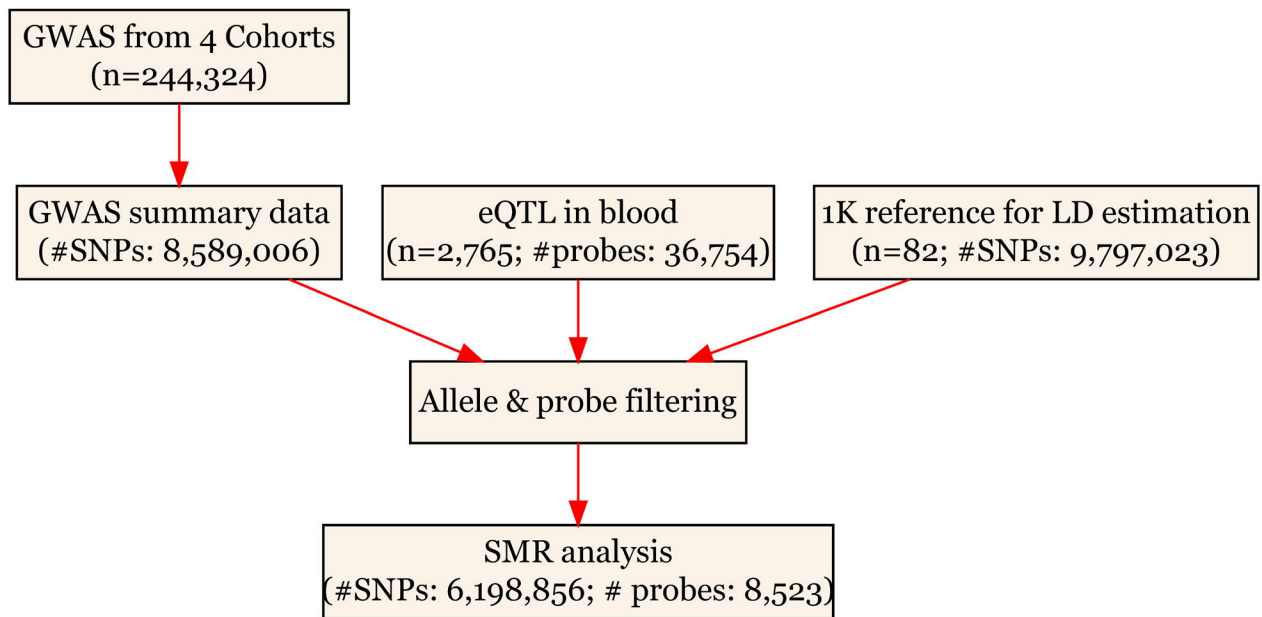
analysis. Beta is the estimated effect size in SMR analysis, SE is the corresponding standard error, P_{SMR} is the P-value for SMR analysis and P_{HEIDI} is the P-value for the HEIDI test.

FDR was calculated at $P=10^{-3}$ threshold.

Bold font means statistical significance after correction for multiple testing using FDR.

CAGE, Consortium for the Architecture of Gene Expression; CCT, central corneal thickness; CHR, chromosome; eQTL, expression quantitative trait loci; GTEx, Genotype-Tissue Expression; HEIDI, heterogeneity in dependent instruments; SNP, single-nucleotide polymorphism; SMR, summary data-based Mendelian randomization; FDR, false discovery rate; GWAS, genome-wide association studies

A



B

