Genome-wide association analysis identifies novel associations in uterine fibroids.

A. Shmygelska[1], N. Eriksson[1], J. Tung[1], J. Mountain[1], A. Kiefer[1], D. Hinds[1].

[1]23andMe, Inc, Mountain View, CA.

Introduction

Uterine leilomyomata, commonly known as uterine fibroids, are benign tumors derived from smooth muscle and fibrous tissue in the uterus, and are the leading cause of hysterectomy in the United States. The lifetime risk for a woman to develop fibroids has been estimated to be as high as 25%. Fibroids tend to grow under the influence of estrogen. The underlying causes of uterine fibroids are not well understood, but it is suggested by twin studies that approximately 55% of the variation in susceptibility to fibroids is genetic.

Methods

To investigate the genetic factors underlying uterine fibroids, we conducted a genome-wide association study (GWAS) of 4,121 cases and 12,252 controls of unrelated European ancestry individuals who have self-reported information on uterine fibroids from the 23andMe participant cohort. Samples were genotyped and imputed against 1000 Genomes reference haplotypes, a total of 8,058,452 SNPs met quality control criteria.

Uterine fibroids cases were defined as having said yes to the following question: "Have you ever been diagnosed with uterine fibroids?" The following table shows demographics of individuals included in the GWAS.

Group	Numbers	Age 0-30	Age 30-45	Age 45-60	Age > 60
case	4,121	5	534	1,779	1,811
control	12,252	110	4,656	4,080	3,443

Table 1. Demographics statistic for uterine fibroids phenotype.

Regression analyses were conducted in a set of responders, controlling for age, and population structure:

Uterine fibroids ~ age + pc.0 + pc.1 + pc.2 + pc.3 + pc.4 + genotype

The results have been adjusted for a genomic control inflation factor of 1.018.

Results

GWAS findings:

We report one novel genome-wide association and four suggestive associations. The most significant finding is the variant in spectrin repeat containing nuclear envelope 1 (SYNE1) gene (rs71575922: odds ratio=1.3, p-value = 4.8 x 10^{-10}), which is 18 kbp upstream of the ESR1 gene which encodes an estrogen receptor. Variants in ESR1 have been previously shown to be associated with breast [1] and endometrial cancer [2]. SYNE1 is expressed in skeletal and smooth muscle, and peripheral blood lymphocytes, that localizes to the nuclear membrane.

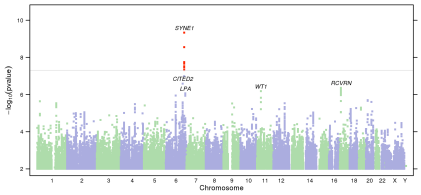

Figure 1. Manhattan Plot of the uterine fibroids GWAS.

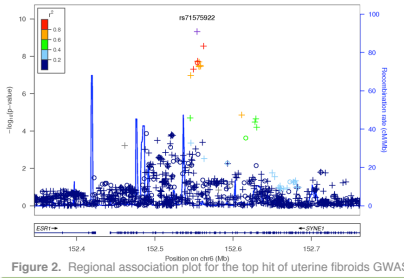

Figure 2. Regional association plot for the top hit of uterine fibroids GWAS.

Suggestive associations include the lipoprotein(a) (LP(a)) gene (rs1800769: odds ratio=1.2, p-value=8.8 x 10^{-7}), which has been shown to promote smooth muscle cells proliferation in transgenic rabbits [3] and rs12789861 (odds ratio=0.9, p-value=6.7 x 10^{-7}) upstream of the Wilms tumor protein (WT1) whose product interacts with estrogen receptor-alpha in breast cancer cells and downstream of the reticulocalbin 1 (RCN1) gene involved among other processes in muscle development and tumorigenesis [4]. We replicate three previously identified associations reported in women of Japanese ancestry (rs12484776 p-value=8.8 x 10^{-5}, rs2172873 p-value=0.012, and rs2280543 p-value=0.015) [5].

GWAS pathway analyses:

To evaluate how multiple genes in the same pathway may contribute to uterine fibroids we compared results from three pathway GWAS approaches: MAGENTA [6], ALIGATOR [7], and VEGAS [8]. Rather than focusing on a few SNPs and/or genes with the strongest evidence of disease association, by considering multiple contributing factors together, we potentially can improve the power to detect causal pathways and disease mechanisms.

We used a gene set of canonical pathways from the Molecular Signatures Database (MSigDB) containing 1320 gene sets compiled by domain experts [9].

Pathway	p-value
BIOCARTA_ETS_PATHWAY	**3.00E-04**
REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE	1.20E-03
REACTOME_SEMA4D_IN_SEMAPHORIN_SIGNALING	2.30E-03
BIOCARTA_TH1TH2_PATHWAY	6.30E-03
REACTOME_G_ALPHA1213_SIGNALLING_EVENTS	9.10E-03

Table 2. Pathway GWAS results for MAGENTA (p-value<0.01).

Pathway	p-value
ST_GA13_PATHWAY	**0**
REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE	**6.00E-04**
REACTOME_SEMA4D_IN_SEMAPHORIN_SIGNALING	**8.00E-04**
ST_P38_MAPK_PATHWAY	0.0024
SIG_CHEMOTAXIS	0.0032
ST_JNK_MAPK_PATHWAY	0.0032
PID_AMB2_NEUTROPHILS_PATHWAY	0.0046
PID_TELOMERASEPATHWAY	0.0052
REACTOME_G_ALPHA1213_SIGNALLING_EVENTS	0.0088

Table 3. Pathway GWAS results for ALIGATOR (p-value<0.01).

Pathway	p-value
REACTOME_IMMUNE_SYSTEM	**4.18E-12**
REACTOME_SIGNALING_BY_GPCR	**4.58E-11**
REACTOME_GPCR_DOWNSTREAM_SIGNALING	**1.96E-07**

Table 4. Pathway GWAS results for VEGAS gene ranking and hypergeometric testing (p-value<0.01).

All three pathway GWAS methods identified G-protein coupled receptor (GPCR) signaling as significant (MAGENTA p-value=9.1x10^{-3}, ALIGATOR p-value=0, VEGAS p-value=4.58x10^{-11}), particularly activation of G-alpha 13 by ligands such as lysophosphatidic acid (LPA) which is formed by oxidizing low density lipoprotein (e.g. LP(a)) have been identified as significant by both MAGENTA and ALIGATOR. Recently G-protein coupled receptors, namely GPCR10, have been implicated in proliferation of cultured primary human leiomyoma cells [10]. G-alpha 13 signaling has been shown to activate P38 MAPK and JNK MAPK which identified as significant by ALIGATOR.

Both MAGENTA (p-value = 2.3x10^{-3}) and ALIGATOR (p-value=8.0x10^{-4}) identified the semaphorin 4D signaling pathway as significant. Semaphorins are secreted and membrane-bound proteins that control axonal guidance in the nervous system, are widely expressed outside the nervous system and play a major signal transduction role in the regulation of cell-cell interactions. The SEMA4D receptor pathway in particular has been shown to control invasive growth in the epithelial cells [11]. In addition, SEMA4D and its receptors have been shown to play a role in tumor carcinogenesis in endometrial cancer [12].

Among pathways identified as significant by MAGENTA are the ETS pathway and cytokine Th1 Th2 pathway. Cytokine Th1/Th2 polymorphisms in IL4 and TNFA have been shown to be associated with increased risk for development of uterine fibroids [13]. ETS transcription factors, which are both targets and modulators of growth factor and steroid-dependent signal transduction and gene regulation, have been previously hypothesized to be important components of signaling pathways underlying the development and pathophysiology of leiomyomata [14].

Protein-protein interaction analysis:

Using the HitPredict database [15] of high confidence protein-protein interactions obtained from high-throughput experiments or derived from protein complex data, we examined the list of significant ALIGATOR GWAS genes with p-value<2.8x10^{-6}, which exceeds a Bonferroni-corrected threshold of p-value=0.05/17,787 autosomal genes. In our list of 48 significant genes, four interact with each other. ARHGEF11 and CDC42 are part of the SEMA4D pathway, MAP3K4 involved in MAPK pathway, both pathways are found significant by the pathway GWAS methods used.

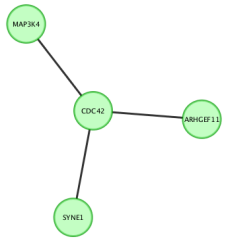

Figure 3. High-confidence protein-protein interactions identified among GWAS significant genes (p-value <2.8x10^{-6}).

Discussion and Conclusion

We pursued three different lines of investigation in order to understand etiology of uterine fibroids: GWAS, pathway based analysis of GWAS data, and protein-protein interaction analysis of ranked genes. In addition to work described here, we added body mass index (BMI) as a covariate for GWAS of European women and the variant within SYNE1 gene was still significant (p-value=1.3 x 10^{-17}). We also ran GWAS among women of African ancestry who are known to have 2-3x higher uterine fibroids prevalence than women of European ancestry (948 cases and 1,818 controls), and found a nearly significant hit in latrophilin-3 (LPHN3, rs6551725 p-value=10^{-7}) – a member of the latrophilin subfamily of G-coupled receptors. Methods presented converge on a set of pathways that are significant including G-coupled protein receptor and semaphorin signaling pathways. We believe that further analysis of growing 23andMe's customer base for this phenotype is very promising.

Acknowledgments

We thank other members of the research team at 23andMe, Inc. for productive discussions and comments.

References

[1] Gold, B., et al. "Estrogen receptor genotypes and haplotypes associated with breast cancer risk." Cancer Res. (2004) 64(24):8891-900.

[2] Doherty, J.A., et al. "ESR1/SYNE1 polymorphism and invasive epithelial ovarian cancer risk: an Ovarian Cancer Association Consortium study." Cancer Epidemiol Biomarkers Prev. (2010) 19(1):245-250.

[3] Ichikawa, T., et al. "Lipoprotein(a) promotes smooth muscle cell proliferation and dedifferentiation in atherosclerotic lesions of human apo(a) transgenic rabbits". Am. J. Pathol. (2002) 160 (1): 227–36.

[4] Fukuda, T., et al. "Distribution and Variable Expression of Secretory Pathway Protein Reticulocalbin in Normal Human Organs and Non-neoplastic Pathological Conditions." Histochem. Cytochem. (2007) 55(4):335-45.

[5] Cha, P.C., et al. "A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids." Nat Genet. (2011) 43(5):447-50.

[6] Segrè, A.V., et al. "Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits." PLoS Genet. (2010) 6(8). pii: e1001058.

[7] Holmans, P., et al. "Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder." Am J Hum Genet (2009) 85(10):13-24.

[8] Liu, J.Z., et al. "A versatile gene-based test for genome-wide association studies. " Am J Hum Genet (2010) 87: 139–145.

[9] Subramanian, A. et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles", Proc Natl Acad Sci U S A. (2005) 102 (43):15545-15550.

[10] Varghese, B.V., et al. "Loss of the repressor REST in uterine fibroids promotes aberrant G protein-coupled receptor 10 expression and activates mammalian target of rapamycin pathway." Proc Natl Acad Sci U S A. (2013) 110(6):2187-2192.

[11] Giordano, S., et al. "The semaphorin 4D receptor controls invasive growth by coupling with Met." Nat Cell Biol. (2002) 4(9):720-4.

[12] Neufeld, G., Kessler, O. "The semaphorins: versatile regulators of tumour progression and tumour angiogenesis", Nature Reviews Cancer (2008) 8, 632-645.

[13] Sosna, O., et al. "Th1/Th2 cytokine gene polymorphisms in patients with uterine fibroid." Folia Biol (Praha). (2010) 56(5):206-10.

[14] McCaffrey, R.L., et al. "Expression of ETS transcription factors in uterine myometrium and leiomyoma." (2004) University of Colorado Health Sciences Center, Denver CO.

[15] Patil, A., et al. "HitPredict: a database of quality-assessed protein-protein interactions in nine species," Nucl. Acids Res. (2011) Database Issue:D744-9.