Transcriptome and DNA methylome analyses reveal underlying mechanisms for the racial disparity in uterine fibroids

Emmanuel N. Paul, … , Hui Shen, Jose M. Teixeira


Graphical abstract

Find the latest version:

https://jci.me/160274/pdf
Transcriptome and DNA Methylome Analyses Reveal Underlying Mechanisms for the Racial Disparity in Uterine Fibroids.

Emmanuel N. Paul¹, Joshua A. Grey¹, Tyler J. Carpenter¹, Zachary B. Madaj², Kin H. Lau², Scott A. Givan², Gregory W. Burns¹, Ronald L. Chandler¹, Ganesa R. Wegienka⁴, Hui Shen³, Jose M. Teixeira¹*

¹Department of Obstetrics, Gynecology and Reproductive Biology, College of Human Medicine, Michigan State University, Grand Rapids, MI, USA
²Bioinformatics and Biostatistics Core, Van Andel Institute, Grand Rapids, MI, USA.
³Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA
⁴Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI, USA

*Corresponding Author:
Jose M. Teixeira, PhD, Michigan State University, 400 Monroe Ave NW, Grand Rapids, MI 49503. Email: teixei15@msu.edu

Acknowledgments:
These studies have been funded by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIH R01HD096259) to JMT.

The authors have no conflicts to declare.
Abstract:

Uterine fibroids (leiomyomas) affect Black women disproportionately in terms of prevalence, incidence, and severity of symptoms. The causes of this racial disparity are essentially unknown. We hypothesized that myometria of Black women are more susceptible to developing fibroids and examined the transcriptomic and DNA methylation profiles of myometria and fibroids from Black and White women for comparison. Myometrial samples cluster by race in both their transcriptome and DNA methylation profiles, whereas fibroid samples only cluster by race in the latter. More differentially expressed genes (DEGs) were detected in the Black and White myometrial sample comparison than in the fibroid comparison. Leiomyoma gene set expression analysis identified four clusters of DEGs, including a cluster of 24 genes with higher expression in myometrial samples from Black women. One of the DEGs in this group, VWF, was significantly hypomethylated at two CpG probes that are near a putative enhancer site in myometrial samples from Black women and in all fibroids and that correlate with VWF expression levels. These results suggest that the molecular basis for the disparity in fibroid disease between Black and White women could be found in the myometria before fibroid development and not in the fibroids themselves.

Keywords:

Leiomyoma, stress, myometrium, racial disparity, transcriptome, methylome
Introduction:

Uterine fibroids, also known as uterine leiomyomas, are the most common tumors in women during their reproductive years with an incidence of up to 80%, depending on race or ethnicity (1-3). Although benign, fibroids can significantly impact the quality of life for many women, with approximately 25% of women with fibroids seeking clinical care for their symptoms (4, 5). There is a strong racial disparity in fibroid disease incidence and severity; Black women are 2-3 times more likely to develop clinically significant fibroids (1, 6, 7), to have them at an earlier age (8, 9), and to report more and worse fibroid symptoms (most commonly, pain and heavy menstrual bleeding) (10).

Hysterectomy, which ends a woman’s fertility, is the only definitive treatment available for the disease, and fibroids are the most common indication for hysterectomy in the US (11). Short term therapies, including Lupron, SPRMs, and oral GnRH agonists (12-14), are available and can provide some relief for women with symptomatic fibroids, but their long-term use is mostly proscribed because of associated side-effects (15). The current gap in knowledge about the sources of racial disparity in incidence and severity of fibroids limits options for prevention or treatment. These limitations leave Black women with a greater healthcare burden overall in which they are forced to make decisions about treatment at younger ages than White women – decisions that have consequences for their fertility and the need for future treatments.

The causes of these racial disparities are unknown, but a meta-analysis demonstrated that chronic psychological stress was weakly associated with the prevalence of fibroids (16-18). Some evidence suggests that vitamin D (VitD) deficiency/insufficiency, which is endemic in Black women in the United States, could be
contributing to the disparity in both the incidence and severity of the disease (17, 19-21).

Previous studies have identified candidate factors and genetic differences associated with the racial disparity in disease incidence that could be involved (22), including polymorphisms of genes involved in estrogen signaling (23), retinoic acid pathway-related genes (24), and miRNA expression levels (25). However, the downstream pathways and mechanisms affected by any of these have not been completely elucidated. Most published fibroid studies of disease incidence and fibroid formation have concentrated on analyzing the uterine fibroids themselves. We used a different approach for the present study and analyzed the adjacent myometrial tissue in addition to the fibroid to determine the transcriptomic and epigenomic differences between the tissues associated with race. We analyzed myometrial samples from Black and White women with fibroids by comparing their transcriptomic and DNA methylation profiles across these race groups. Here, we show evidence that myometrial samples from Black women are different from those of White women, with higher expression of some stress- and fibroid-associated genes, suggesting mechanisms by which myometria of Black women may be more susceptible to uterine fibroid development.
Results:

RNA-seq analyses were performed on myometrial (MyoF) and *MED12*-mutant (*MED12*mt) fibroid (F) samples from Black (n=22) and White (n=24) women (Table S1) to identify DEGs that could be contributing to the racial disparity in uterine fibroid disease. To minimize sample variability, only tissue samples from patients with *MED12*mt fibroids, which are the most common genetic subtype of uterine fibroids (26), were used in the study. Unsupervised hierarchical clustering analysis of all expressed genes in F samples from Black and White women (Fig 1A) shows that the racial identities of the transcriptomes appear randomly scattered among the two main branches of the dendrogram. In contrast, when the same analysis is applied to the expressed genes in the MyoF samples from Black (n=18) and White (n=19) women (Fig 1B), the two main branches segregate comparatively well. Principal component (PC) regression analysis (Fig 1C and D) showed that PC1 in MyoF samples is significantly associated with race, which was not the case for the F samples nor for PC2 and PC3 in both the MyoF and F samples. Similarly, the PC plot of PC1 and PC2 shows a better separation by race in MyoF (Fig 1F) than in F (Fig 1E). Neither F nor MyoF samples showed significant separation by race in other component comparisons (Fig S1A and B).

Further analyses of the DEGs between Black and White transcriptomes revealed that more DEGs (FDR <0.05) were observed in the MyoF samples (Fig 2B, 1411 total DEGs, Table S2) compared with the F samples (Fig 2A, 198 total DEGs, Table S3). Three different genes that we thought could be differentially expressed in the tissues from Black and White women based on our previous studies (27, 28) or that are candidate genes gleaned from the literature that could be important for stress-related fibrosis (29-33) are
highlighted. Von Willebrand Factor (VWF) and Brain Derived Neurotrophic Factor (BDNF) were induced in Black MyoF samples compared to White MyoF samples but not in the Black F samples compared to White F samples. Cyclin D1 (CCND1), which we have previously reported upregulated in F compared to MyoF (27), was not significantly induced in MyoF from Black women compared to MyoF from White women. Gene Set Enrichment Analyses (GSEA) (Fig 2C and D) was performed with Hallmark gene sets (34, 35) to identify dysregulated pathways that could be important for the racial disparity in fibroid disease. Only 4 Hallmark gene sets involved in cell growth and cell cycle were enriched in Black F compared to White F samples (Fig 2C). In contrast, 23 Hallmark pathways were enriched in Black MyoF compared to White MyoF (Fig 2D), including known pathways that are enriched in uterine fibroids: TGFβ signaling, Wnt/β-catenin signaling, P53 pathway and myogenesis (36-38). Hypoxia and reactive oxygen species (ROS), stress-related pathways, were also enriched in MyoF from Black women. A total of 489 genes were down-regulated and 922 genes were up-regulated (Fig 2E) in MyoF samples compared from Black and White women. Only 97 genes were down-regulated and 101 were up-regulated in the same comparison with F samples (Fig 2E). Overlaps of the two comparisons were not significant.

To determine if the DEGs in MyoF from Black women compared to those from White women contain fibroid genes, we overlapped these DEGs with the DEGs of the F versus matching MyoF comparison. Over 15% of the DEGs identified in the MyoF samples from Black women compared with those of White women were also differentially expressed in F compared with MyoF (Fig 3A and B). Expression of BDNF was up-regulated in both the Black and White MyoF and the F and MyoF comparisons (Fig 3B).
VWF expression was only seen up-regulated in the Black and White comparison of MyoF samples, and CCND1 was up-regulated only in F compared to MyoF (Fig 3B).

We performed a clustering analysis of a leiomyoma gene set (39) with the addition of BDNF for MyoF and F samples in both race groups. The analysis showed four distinct clusters of genes, which we have called groups I, II, III, and IV (Fig 3C). Group I contained genes with higher expression in White MyoF samples compared to Black MyoF, Black F, and White F samples. Group II included genes, such as VWF, that showed higher expression in the Black MyoF samples. Gene expression in group III was tissue type dependent, with low expression in all MyoF samples compared to all F samples. Lastly, group IV had highly expressed genes in F samples, such as CCND1, with some genes also highly expressed in Black MyoF samples, such as BDNF.

We next compared the gene expression profiles of the MyoF and F samples from Black and White women with the White non-fibroid myometrial (MyoN, n=6) samples from a recently published study in our laboratory (27). In that study, MyoF transcriptomes appeared more like the transcriptomes of F samples than the MyoN samples, suggesting that the myometria from women with fibroids were prone to fibroid disease or were affected by the fibroids themselves. CCND1, VWF and BDNF expression levels were not significantly different between F from Black women and F from White women (Fig 3D-F and Table 1). CCND1 expression was significantly higher in F tissue samples from either race compared to those from MyoF and in MyoF samples compared to MyoN. Within the MyoF samples, Black women had a higher level of CCND1 expression than the White women (Fig 3D). VWF expression was significantly higher in MyoF (White and Black) samples compared to MyoN samples, but only the Black MyoF samples were significantly
higher than F samples of either race (Fig 3E). We also confirmed that VWF was upregulated in Black MyoF compared to White MyoF (Fig S1C) using data from a published study (40). BDNF expression was significantly increased in Black MyoF samples compared to MyoN samples and White MyoF. BDNF expression was not significantly different when Black or White F samples were compared to Black MyoF; however, BDNF expression was higher in F from either race compared to White MyoF.

Altered DNA methylation, both genome-wide and gene-specific, has been proposed as a possible mechanism for stress-related changes observed in a variety of systems (reviewed in (41)). Methylomes for MyoF and F samples from Black and White women (Table S4) were profiled using the Infinium MethylationEPIC array (EPIC) (42). MDS plots based on the 500 most variable EPIC DNA methylation probes by standard deviation revealed a significant difference in dimension 2 between Black and White patients in both F (n=16 Black, n=25 White) and MyoF (n=13 Black, 19 White) tissues (Fig 4A and B); these differences were independent of the patient’s actual age, their inferred age, and batch effects (data not shown). Comparison of the gene-associated differentially methylated regions (DMRs) containing differentially methylated CpGs in Black and White samples (Fig 4C and D, Tables S5 and 6) shows more total DMR-associate genes in the MyoF samples than in the comparison of F samples from Black and White women, reflecting the greater separation observed in Fig 4B compared to 4A. The number of hypomethylated DMR-associated genes was slightly greater in F than in the MyoF samples from Black women, but the opposite was true for hypermethylated DMRs with MyoF samples showing greater numbers of gene-associated DMRs in Black women than in F samples. Gene-associated DMRs were also identified that are either...
hypomethylated (n=54) or hypermethylated (n=539) in both F and MyoF from Black women (Fig 4C and D). We next determined which CpG probes showed the highest concordance in hypo- and hypermethylation between Black and White MyoF samples and all F and MyoF samples (Fig 4E and F, respectively) to identify genes associated with differential methylation that correlated with race and could be important for the disparity in fibroid disease. We identified 327 gene-associated probes that were hypomethylated and 495 that were hypermethylated by greater than 10% in both F compared to MyoF and in Black compared to White MyoF.

We observed two CpG probes associated with the VWF gene that were hypomethylated in both Black MyoF compared to White MyoF (Fig 5) and in F compared to MyoF (Fig S2A). These hypomethylated probes at approximately 60 kb upstream of the VWF transcription start site near a CTCF binding site (Fig 5A) identified by ChIP-seq in IMR-90 smooth muscle cells (43) that might be important for regulatory interactions with putative enhancer elements (44). The probes are also upstream of another gene on the opposite strand, CD9, which is similarly upregulated in Black MyoF compared to White MyoF samples (Fig S2B), suggesting coordinated regulation. One of these probes, which is hypomethylated in Black MyoF and in all F by more than the 10% cutoff is indicated in Fig 4E in blue. The mean beta values of both probes are significantly lower in Black MyoF than in White MyoF (Fig 5B). We also confirmed that the beta value of each probe is consistent with VWF expression in the RNA-seq results in each of the corresponding patient MyoF samples (Fig 5C).
Discussion:

There is a well-known racial disparity in uterine fibroid disease incidence and severity, in which Black women are more likely to have fibroids and have more and larger fibroids at diagnosis (45). In addition, Black women have higher rates of surgery for uterine fibroids than other women (46, 47). Previous studies investigating this disparity using transcriptomic or DNA methylation have been reported, but results did not show correlation by race with either gene expression (48) or DNA methylation (28, 49). By studying the race difference in both the fibroid tissue and the adjacent myometrium, we have discovered transcriptomic and DNA methylation differences between Black and White women that could be driving or contributing to the racial disparity in the disease.

We observed that while RNA-seq myometrial cluster 1 had a statistically significantly higher proportion of Black women and cluster 2 had a higher proportion of White women, the clusters were not strictly defined by patient race (Fig 1B). This suggests that shared exposures or experiences amongst cluster members could likely explain the differences between the clusters and can likely provide insight into fibroid formation. Although in the current study patients had 1-3 fibroids evenly distributed by race, Black women are known to be at higher risk to develop multiple and larger fibroids than White women (45), and a field effect by the fibroids on the surrounding myometrium could influence the transcriptomic and epigenetic profiles of the tissue. This caveat could confound our interpretation of the results and represent a limitation of this study. Expansion of the study by collect and comparing myometrial samples from three different cohorts of Black women: (1) without fibroids, (2) and with only one fibroid, and (3) with multiple fibroids as confirmed by pathology, could be dispositive.
We have highlighted the differential expression of three genes (BDNF, CCND1, and VWF) in our results that we think could play key roles in uterine fibroid etiology or pathogenesis. CCND1, a cell cycle regulator that promotes cellular proliferation and is frequently overexpressed in human cancers (50), was previously identified by us (27, 28) and others (51-53) as a central gene whose expression is upregulated in F compared to both MyoF and MyoN. CCND1 expression was not significantly dysregulated in MyoF from Black women compared to that of White women despite the fact that its expression was increased 1.5-fold. We hypothesize that the increased expression of CCND1 in myometria of Black women is still biologically important in the fibroid disparity given the trend for increasing expression in F>MyoF>MyoN and speculate that our study was underpowered to detect a significant difference in the comparison of myometria from Black and White women. Serum levels of VWF are known to be elevated in Black subjects (29, 30), which provides us with a degree of confidence that these results are biologically meaningful. VWF, an endothelial and tumor progression marker (54), was upregulated in MyoF patients compared with myometrium from patients without fibroids, but was not differentially expressed in the F tissues (27). This result suggests that vascularization in MyoF from Black women could be higher than in MyoF from White women and similar in F from both, after the tumors are established. Indeed, microvessel density in F has been reported to be lower than in adjacent MyoF (31).

We showed that BDNF expression is higher in Black MyoF and in both F compared to White MyoF, the significance of which is not clear. BDNF has been associated with psychological stress (32), and serum BDNF appears elevated after acute psychosocial stress (33). However, another study has reported a conflicting correlation between BDNF
expression and stress (55), which adds a degree of uncertainty to attributing a possible role to BDNF in fibroid biology. Black women are known to suffer more from socio-economic stress than other races (56) and Black individuals experience more factors that negatively impact their health (e.g., perceived personal discrimination, perceived structural racism). Through Weathering (the accumulation of negative effects) and John Henryism (the cost of coping with these insults), their individual health is further eroded. Additionally, community level factors (e.g., toxicants) both directly impact health status and modify the relationship between individual level factors and individual health status (57, 58), suggesting that stress-associated genes could play a role in the establishment of uterine fibroids, particularly in Black women.

Interestingly, 3 genes, ARL2, AIP and CMYA5 from a total of 33 in a peptide ancestry informative markers study of Black and White F tissues (59) were found in our DEG list from the Black MyoF comparison with White MyoF. The pathological significance of these genes to fibroid biology is not known, and further analyses are needed. The AIP gene is particularly interesting since it encodes for the aryl hydrocarbon receptor interacting protein, mutation of which has been associated with pituitary adenoma development and acromegaly (reviewed in (60, 61)). It is possible that higher AIP expression in Black myometrial samples could be associated with higher levels of exposure to pollutants, and it is tempting to speculate that it provides the missing link between the higher incidence of fibroids in Black women with living in an urban environment. More studies are needed to explore this possible link in the general context of the role urban environmental pollutant exposures have on fibroid development and growth.
Psychological stress and depression have also been shown to induce remodeling of the chromatin landscape, which affects gene expression (62, 63). Additionally, transgenerational epigenetic reprogramming has been associated with in utero and parental stress (reviewed in (64)) suggesting a repeating or even perpetual mechanism for maintaining health disparities in a variety of disease states in marginalized or oppressed communities. In our study, we showed that MyoF from Black women compared with those from White women have a similar number of hypomethylated gene-associated DMRs as in the F comparison. In contrast, more hypermethylated gene-associated DMRs were observed in the MyoF samples from Black women compared to those from White women than in F. In both cases, the number of overlapping DMRs was relatively small. Further study is needed to determine the biological importance of these striking and contrasting results.

Although CCND1 and BDNF gene regions were not differentially methylated, two probes upstream of VWF, were hypomethylated in MyoF from Black patients compared with MyoF from White patients and in F samples compared with MyoF, results that were correlated with RNA-seq results. Moreover, GSEA of the MyoF samples showed pathways known to be upregulated during stress, (e.g., reactive oxygen species (reviewed in (65)), hypoxia (66, 67)), and uterine fibroid tissues, (e.g., Wnt/β-catenin signaling (68), TGFβ signaling (27, 69) and myogenesis (70)). The results of this study provide a compelling rationale to include adjacent myometrial tissues in addition to the fibroid tissues in studies designed to better understand the etiology of fibroids and the racial disparity in the incidence of the disease. These data also suggest that the
relationships between various exposures and epigenetic patterns in the fibroids and other uterine tissue could provide important information in fibroid etiology.
Methods

Tissue Collection

Fibroids (F) samples and matched myometrial (MyoF) from White and Black pre-menopausal women (aged 37–52) were obtained following total hysterectomy. All patients who participated to the study gave consent to donate tissue for this study through the Spectrum Health Biorepository. Human samples were processed as previously described (27). Briefly, samples were washed with phosphate-buffered saline, minced, and immersed in RNALater (Sigma, Saint Louis, MO, USA) and stored at 4 °C for RNA-seq analyses. The remaining tissue pieces were immediately flash frozen and stored at −80 °C for methylation analysis. MED12 mutation in the fibroids was determined by PCR amplification followed by Sanger sequencing as described in our previous study (27).

RNA Isolation

Total RNA was isolated from frozen tissues stored in RNALater. Tissues were homogenized in TRIzol reagent (Thermo Fisher Scientific, Fairlawn, NJ, USA) and RNA was isolated following the manufacturer’s instructions. Table S6 contains all the metadata available for these samples. Isolated RNA was stored at −80 °C in nuclease-free water. Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Fair-lawn, NJ, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) instruments were used to measure RNA concentration and quality, according to the manufacturers’ protocols. RNA integrity values of ≥7.5 were required for sequencing.
Library Preparation and Sequencing

High quality RNA samples from White F, Black F, White MyoF and, Black MyoF were submitted to the Van Andel Research Institute (VARI) Genomics Core for library preparation and paired-end (2 × 100 bp) RNA-sequencing on an Illumina NextSeq 6000 instrument (Illumina Inc., San Diego, CA, USA). Libraries were prepared using a Kapa RNA HyperPrep kit with ribosomal reduction, pooled, and sequenced on flowcells to yield approximately 50–60 million reads/sample. Raw fastq files were deposited in the NCBI Gene Expression Omnibus (GSE207350).

RNA Seq Analysis

New reads and reads from samples in previously published studies (27, 28, 71), Black F and matching MyoF (GSE128229), Black and White F and MyoF (GSE135446), and MyoN (GSE169255), were trimmed for quality and adapters using TrimGalore (version 0.6.5), and quality trimmed reads were assessed with FastQC (version 0.11.7). Trimmed reads were mapped to Homo sapiens genome assembly GRCh38 (hg38) using STAR (version 2.7.9a) (72). Reads overlapping Ensembl annotations (version 99) were quantified with STAR prior to model-based differential expression analysis using the edgeR-robust method. Genes with low counts per million (CPM) were removed using the filterByExpr function from edgeR (72). Consensus clustering plots were made using the median-centered, normalized counts for the 5,000 most variable genes based on median absolute deviation and the ConsensusClusterPlus package (version 1.56.0) (73) with the parameters, ‘reps=1000’, ‘pItem=0.8’, ‘pFeature=0.8’ and ‘distance=spearman’. Briefly, these settings resample the data 1000 times using 80% of the samples and features each
time, then finds the consensus clustering based on hierarchical clustering of each
resampling using (1-Spearman correlation) as distance. Scatterplots of two selected
principal components was constructed with the pca function of the PCAtools R package
(version 2.5.13) to verify sample separation prior to statistical testing. Generalized linear
models were used to determine if principal components were significantly associated with
race by tissue. Genes were considered differentially expressed if their respective edgeR-
robust false discovery rates (FDR) corrected p-values were less than 0.05. Differential
expression was calculated by comparing F(B) versus F(W), MyoF(B) versus MyoF(W), or
F versus MyoF. Differentially expressed genes (DEGs) were visualized with volcano plots
and heatmaps generated using the EnhancedVolcano (version 1.6.0) and pheatmap
(version 1.0.12) packages in R. Downstream analyses of RNA-seq results were
completed using the clusterProfiler (version 3.16.1) (74) package in R with an FDR p-
value cutoff of 0.05. Gene set enrichment analyses were conducted with all expressed
genes using the 50 Hallmark gene sets collection (H) (34) downloaded from the Molecular
Signatures Database (MSigDB) (35). The top enriched GSEA terms were shown in the
figures. Venn diagrams were constructed to visualize overlapping genes between groups
or gene sets using the eulerr package (version 6.1.1). DOSE r package (version 3.14.0)
was used to generate the leiomyoma gene set heatmap using DOID:127 (75). Box plots
of the log2(CPM +1) values were generated using the R package ggplot2 (version 3.3.5).
Independent MyoF samples (n=4 MyoF White and n=4 MyoF Black) from another
leiomyomas study (40) were used to confirm our RNA-seq analysis (GSE193320).
DNA Isolation and DNA Methylation Analysis

DNA was isolated from snap-frozen myometrium and fibroid tissue samples (Table S4) and hybridized to the Infinium MethylationEPIC array and analyzed essentially as previously described (28). Raw IDAT files were deposited in NCBI GEO (GSE207350). Additional raw IDAT files from our previous report (28) GSE120854 and GSE135446 were added. Briefly, raw IDAT files were processed using R package SeSAMe (version 1.12.7) and the openSesame pipeline with noob background correction, non-linear dye bias correction, and non-detection masking (42, 76). DNA methylation was measured in beta values for each probe calculated as a quantitative percentage of methylated signals over both methylated and unmethylated probe signals. Matched samples and self-identified race were confirmed using EPIC SNP probes (Figures S3A and S3B) with a published model (42), and cellular composition of the samples was determined by promoter methylation of MIR200C/141 (77) and ACTA2 (78) (Figure S3C).

Multi-dimensional scaling analysis of the top 500 most variable CpG probes was performed using R package minfi (version 1.40.0). Differentially methylated regions (DMRs) were called using R package SeSAMe by first modeling the variation in DNA methylation in each CpG probe using race as the independent variable. Neighboring CpG probes that showed consistent variation in methylation were then merged into DMRs. Visualization of DNA methylation in probes across a genomics region was generated using UCSC genome browser tracks from available track hubs including GeneHancer regulatory elements (44) and gene interactions and CTCF ChIP-seq peaks in IMR-90.
Probe annotation of Illumina EPIC array (human reference genome (NCBI build GRCh38/hg38)) was downloaded from (42).

Statistical Analysis

Bioinformatic statistics were performed using the listed packages in R (version 4.0.2). Differentially expressed genes were identified as those having a Benjamini–Hochberg FDR corrected p <0.05 (79). Data with unequal variances were log transformed, and homogeneity of variances verified before completion of analyses. Hypergeometric testing was performed using the phyper function from the stats package (version 4.0.2) and the Likelihood ratio tests were done with lrtest function using the lmtest package (version 0.9-40). Comparison of two means was performed with a two-sided student t test, and significance was determined at p <0.05 after confirming normal distribution using Graphpad Prism (version 9.3.1). CpG beta correlations with expression were also performed with Prism.

Study approval

The use of human tissue specimens was approved by the Spectrum Health Systems and Michigan State University Institutional Review Boards (MSU IRB Study ID: STUDY00003101, SR IRB #2017-198) as secondary use of biobank materials.

Author contributions

Experimental design (ENP, RLC, HS, JMT), collected data and performed experiments (ENP, JAG, TJC), analyzed data (ENP, JAG, TJC, ZBM, KHL, SAG, JMT),
wrote/reviewed manuscript (ENP, JAG, TJC, ZBM, KHL, SAG, GWB, RLC, GRW, HS, JMT).

Acknowledgments

We would like to thank the patients who consented for the study, the Spectrum Health Systems Universal Biorepository staff, and the Van Andel Research Institute Genomics Core for their help with the study. This work was supported in part through computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University.
References:


Figure 1. Race-based clustering of RNA-seq results from myometrial and fibroid samples. Whereas unsupervised hierarchical k=2 means clustering of RNA-seq results from *MED12*mt fibroid samples (n= 24 F White and n= 22 F Black) is not significantly associated with race (Fisher’s exact test, p=0.76) (A), the results from myometrial samples (n= 19 MyoF White and n= 18 MyoF Black) show an association with race (Fisher’s exact test, p=8.3x10^{-4}) (B). The length of each leaf in the dendrograms indicates degree of dissimilarity. Race of each sample is color coded as indicated. The clusters
are shown above the race. Spearman bootstrap analyses (1000x) are shown below each dendrogram for each sample. Y axis of bootstrap columns indicates stability of clustering. Boxplots of the first 3 principal components (PC) of RNA-seq results from Black and White F (C) and MyoF (D). The value of each sample represents the individual gene expression values transformed by the rotation matrix estimated from the PCA. Only PC1 in MyoF are significantly different (likelihood ratio test, *p=0.03). PC plot analyses for PC1 vs PC2 of F (E) and MyoF (F) shows samples from White and Black women in two-dimensional space. Each dot represents one sample.
Figure 2. Differential gene expression by race in F and MyoF samples. Fibroids (F) from n=22 Black and n=24 White women, and matching myometrium (MyoF) from n=18 Black and n=19 White women, were compared. Volcano plots showing the up- and down-regulated genes with a false discovery rate (FDR) p-value < 0.05 depicted as red dots in F (A) and MyoF (B). Gene set enrichment analysis of expressed genes using Hallmark biological processes in MSigDB from the comparison of the F (C) or MyoF (D) samples from Black and White women. Gene count and significance level are shown by the size and color of each circle. Venn diagrams illustrate the overlap of the down- and up-regulated genes.
regulated genes (E) between Black MyoF compared to White MyoF samples and Black F compared with White F samples. Hypergeometric tests of the overlaps between the two comparisons of the 33 up-regulated and 30 down-regulated sets of overlapping genes were not significant (p > 0.99 for both).
Figure 3. Transcriptomic similarities in fibroids (F) from all women and Black women’s matched myometria (MyoF). Venn diagrams illustrate the overlapping down- (A) and up-regulated (B) differentially expressed genes (DEGs) between MyoF from Black patients (n=18) and MyoF from White patients (n=19) and between F (Black and White combined, n=46) and MyoF (Black and White combined, n=37). Hypergeometric testing revealed that the overlaps were significant, with p=9.0x10^{-05} for the down-regulated genes and p=1.9x10^{-14} for the up-regulated genes. (C) Heatmap of the...
leiomyoma gene set from Disease Ontology with added *BDNF*, using the average log₂(CPM +1) of each group: MyoF White, MyoF Black, F White (n=24), and F Black (n=22). Color gradient represents gene expression levels as z-scores. Boxplot of *CCND1* (D) *VWF* (E) and *BDNF* (F) of myometrium from White non-fibroid patients (MyoN) (n=6), MyoF from White (n=19) or Black (n=18) patients and F from White (n=24) or Black (n=22). Gene expression is shown as log₂(CPM +1). FDR p-values for each comparison are reported in the Table 1.
Figure 4. DNA methylation profiles of Black and White fibroids (F) and matched myometria (MyoF). DNA methylation in the samples was determined using the Illumina MethylationEPIC Beadchip microarray. Multidimension scaling plots of beta values for fibroids (F) from Black (n=16) and White (n=25) women (A) and matching myometrial samples (MyoF) from Black (n=13) and White (n=19) women (B). Each dot represents an individual sample. Significant methylation differences by race were determined by the likelihood ratio test; p=2.7x10^{-5} for F and p=2.2x10^{-16} for MyoF. Gene-associated differentially methylated regions (DMRs) for the Black and White race comparisons of MyoF and F samples are shown in a Venn diagram as hypomethylated (C) or hypermethylated (D), with the overlap in the circles indicating shared DMRs. Hypergeometric testing of the overlaps in C and D revealed that they were statistically
significant (p-values = 2.6x10^{-9} and 2.3x10^{-66}, respectively). The congruent hypomethylated (E) and hypermethylated (F) EPIC CpG probes in the Black and White MyoF and the MyoF and F comparisons (>10%) are shown plotted with randomly displayed gene names. A CD9/VWF-associated probe is highlighted in blue.
Figure 5. VWF gene hypomethylation correlates with increased gene expression.

University of California Santa Cruz (UCSC) genome browser view of the VWF gene locus.
(A), coupled with the DNA methylation beta values for this region, shown as a heatmap (columns, CpGs; rows, samples grouped by race). Top UCSC tracks include locations CTCF ChIP-seq peaks in IMR-90 cells from ENCODE (in orange) and predicted cis-regulatory elements provided by GeneHancer. The heatmap shows the CpG methylation in the VWF gene. Box and whisker plots (B) show the median and range of beta values in Black and White MyoF for the two probes identified in A. Means for probes are significantly different by t test (p=0.01 for both). (C) Beta values (X axis) and VWF expression (Y axis) show that beta value is negatively correlated with expression in the MyoF samples ($r^2=0.19$ for cg14954762 and $r^2=0.21$ for cg25926078, p=0.03 for both).
Table 1: False Discovery Rate (FDR) for \textit{CCND1}, \textit{VWF} and \textit{BDNF} boxplots in Fig 3.

NS: non-significant change, FDR>0.05

<table>
<thead>
<tr>
<th>Comparison</th>
<th>\textit{CCND1}</th>
<th>\textit{VWF}</th>
<th>\textit{BDNF}</th>
</tr>
</thead>
<tbody>
<tr>
<td>MyoF(W) vs. MyoN</td>
<td>6.25E-03</td>
<td>1.16E-03</td>
<td>NS*</td>
</tr>
<tr>
<td>MyoF(B) vs. MyoN</td>
<td>1.88E-03</td>
<td>7.29E-11</td>
<td>2.35E-02</td>
</tr>
<tr>
<td>F(W) vs. MyoN</td>
<td>1.10E-12</td>
<td>5.90E-04</td>
<td>NS</td>
</tr>
<tr>
<td>F(B) vs. MyoN</td>
<td>1.16E-11</td>
<td>1.37E-04</td>
<td>1.02E-03</td>
</tr>
<tr>
<td>MyoF(B) vs. MyoF(W)</td>
<td>NS</td>
<td>1.50E-06</td>
<td>2.41E-02</td>
</tr>
<tr>
<td>F(W) vs. MyoF(W)</td>
<td>1.00E-19</td>
<td>NS</td>
<td>6.71E-10</td>
</tr>
<tr>
<td>F(B) vs. MyoF(W)</td>
<td>9.22E-18</td>
<td>4.28E-02</td>
<td>6.69E-05</td>
</tr>
<tr>
<td>F(W) vs. MyoF(B)</td>
<td>9.98E-08</td>
<td>1.46E-04</td>
<td>NS</td>
</tr>
<tr>
<td>F(B) vs. MyoF(B)</td>
<td>1.83E-06</td>
<td>2.35E-02</td>
<td>NS</td>
</tr>
<tr>
<td>F(B) vs. F(W)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>