

# Full Proposal Week 7 Revised

## Research Proposal

*by* Janis Corona

---

**Submission date:** 03-Mar-2019 02:16PM (UTC-0600)

**Submission ID:** 1086858880

**File name:** ResearchProposalWeek7.docx (38.85K)

**Word count:** 6889

**Character count:** 36262

Janis Corona

BIOL 51200: Research in Biotechnology

## Week 7 Assignment

March 2, 2019

## Gene Expression Analysis of the Ubiquitous Genotypes Associated with Uterine Leiomyoma

## Development in Healthy Tissue using Bioconductor

In this research project, the top genetic markers for heterogenetic risk in developing uterine leiomyomas (UL) or alternatively uterine fibroids (UF) will be examined and analyzed in the data made available for gene expression using the Gene Expression Omnibus (GEO) online data repository. There

are many genome wide association studies (GWAS) on the many single nucleotide polymorphisms (SNP).

associated with uterine fibroids, the studies have been fine tuned to heterogenous differences between races of European Americans, Japanese, Chinese, African Americans, Australians, White females from Australia or the United Kingdom, and Saudi Arabian females (Edwards, T. et al., 2013; Liu, B. et al., 2018; Hellwege, J. et al., 2017; Eggert, S. et al., 2012; Rafnar, T. et al., 2018).

Many **UL** and **UF** research studies define uterine leiomyomas and their synonym uterine fibroids as benign tumors in the uterine myometrium or similarly as benign growths in the smooth muscle tissue of the myometrium (Eggert, S. et al., 2012; Bondagji, N. et al., 2018). Some of the known risk factors for developing a UF are age at menarche, alcohol consumption, child bearing age, family history of UF, race, and obesity (Hellwege, J. et al., 2017; Eggert, S. et al., 2012; Rafnar, T. et al., 2018). It is also known that UF and UL are estrogen responsive and that discontinuing hormone therapy make UF symptoms and tumor growth recur (Rafnar, T. et al., 2018). There is a risk of developing a UL if the UL patient also has thyroid dysregulation, kidney cancer, stage III or higher endometrial cancer, or endometrial cancer with the genotype rs10917151 of the CDC42/WNT4 gene (Rafnar, T. et al., 2018). It is also known that MED12 is the only gene to have a causal relationship in having a UF or UL (Bandagji, N., et al, 2017). The

knowledge of how uterine fibroids develop is still unknown and many GWAS studies have sought to find gene targets along SNPs of highly up or down regulated genes in differential studies between normal uterine tissue and UL or UF tissue (Zhang, D. et al., 2012; Hodge, J. et al., 2012).

The most ubiquitous SNP genotypes highlighted in these GWAS population specific studies are the rs2280543 genotype belonging to the Bet1 Golgi Vesicular Membrane Trafficking Protein like gene called BET1L and the rs12484776 genotype belonging to the trinucleotide repeat containing 6B gene called TNRC6B (Edwards, T. et al., 2013; Rafnar, T. et al., 2018; Liu, B. et al., 2018, Bondagji, N. et al., 2017). These genotypes have been shown in separate population specific studies to associate to the number of UL or UF one patient has (rs2280543, BET1L) and the size of the UF one person has (rs12484776, TNRC6B) in European American, Japanese, and Han Chinese populations (Edwards, T. et al., 2013; Liu, B. et al., 2018). Saudi Arabian populations found that TNRC6B only poses a risk of developing a UF or UL (Bondagji, N. et al., 2017). Two studies by separate researchers Rafnar, T. et al., (2018) examining UL in Europeans from the United Kingdom and Iceland and Aissani, B. et al., 2015 studying UL in European Americans found that BET1L genotype rs2280543 is not associated with UL. However, two other separate studies by Eggert, S. et al., 2012 and Edwards, T. et al., 2013 found that the BET1L genotype rs2280543 is associated with UL risk for white women and European Americans.

A study on European Americans by Edwards, T. et al. (2013) found that rs2280543 of BET1L associated with what part of the uterus a UL formed in European American populations, such as in the uterine wall (intramural), under the endometrium (submucosal), or under the mucosal layer of the uterus (subserous). This same phenotype of BET1L genotype rs2280543 is also found to be significant in the Han Chinese population (Liu, B. et al., 2018).

In a particular study on white races of Australian and European origin, additional SNPs for fatty acid synthase (FASN) and rs4247357 of coiled-coil domain containing 57 gene called CCDC57 have been

found to have a genome-wide significance for UF or UL in white populations while not showing significance in Arab populations (Eggert, S. et al., 2012; Bondagji, N.S., et al., 2017).

There is insufficient evidence to include these same genotypes as biomarkers for UF in the African American females possibly due to misclassification of fibroid by the self-reporting of UF in control groups used in this study (Aissani, B. et al., 2015; Hellwege, J. et al., 2017). Because UF diagnosis is only reported if symptomatic and most cases of UL or UF are asymptomatic as only 20-33% of patients with UL show symptoms such as pain in the pelvis and heavy bleeding (Bondagji, N. et al., 2017; Eggert, S. et al., 2012). The gene target found to be an exclusive heterogenetic risk of UL in African American populations is the rs739187 SNP of cytohesin-4 or CYTH4; when it is expressed low in thyroid tissue there is a risk for developing UF for African American females (Hellwege, J. et al., 2017).

There is also a study by Eggert, S., et al., 2012<sup>6</sup> on white females from European and Australian data of women with UL and also included sisters and family members with UL. In this study there was a genome wide significance level of risk of UL with coiled-coil domain containing 57 gene called CCDC57. The study also found that fatty acid synthase or FASN plays a role in risk of UL in white females.<sup>Citation Needed</sup>

When excluding studies on heterogeneity of UL or UF, Hodge, J. et al., 2012 found that the putative gene HGMA2 of the high mobility group on chromosome 12 being over expressed in UL and as the most significant altered gene.<sup>7</sup> This same study also suggested that due to the most variation in clustering around patient demographics than clustering of t(12;14) and non-t(12;14)<sup>5</sup><sup>8</sup> that there is reason to believe that race plays a role in risk for UL development.<sup>Citation Needed</sup>

Another study that excluded race as a determinant in gene expression analysis of UL is the study by Zhang, D., et al., (2012). In this study on differential gene expression, the four phases of menstruation were analyzed. This was to see when the best time for implantation of a fertilized ova to produce an embryo would occur. This study was not race specific to the uterus samples gathered at different stages of the gene sample extraction. But measured for high variation of genes expressed and collected those

genes to find the most significant ones. The study identified the chromosomes of the genes most expressed as chromosomes 4, 9, and 14. Many of the top gene SNPs from the GWAS samples were gathered from most expressed genes along a region of a chromosome. Then, only those genes were further analyzed in determining the significance of risk for UL in each race by significantly high gene expression of SNPs in UL cases (Aissani, B., et al., 2015; Eggert, S., et al., 2012; Bondagji, N., et al., 2017; Edward, T., et al., 2013). The currently significant SNP genes found associated with UL among all of the population studies researched are BET1L on chromosome 11, TNRC6B on chromosome 22, FASN on chromosome 17, CYTH4 chromosome 22, CCDC57 on chromosome 17, HGMA2 on chromosome 12, and MED12 on chromosome X or 23 (Aissani, B., et al., 2015; Eggert, S., et al., 2012; Bondagji, N., et al., 2017; Edward, T., et al., 2013; Hodge, J., et al., 2012; Hellwege, J. et al., 2017; Liu, B., et al., 2018; Rafnar, T., et al., 2018). Zhang, D., et al., (2012) found chromosomes 4, 14, and 9 to be in healthy uterine tissue capable of impregnation; these chromosomes are not from the SNP located genotypes along chromosomes 11, 12, 17, 22, and 23 that have regions associated with developing UF among varying populations. Thus, it makes sense to further study these genotypes associated with UF except for the MED12 gene that is already been proven causal to UF (Bondagji, N., et al., 2017). The CDC4/WNT4 genotypes are excluded because they are only found to be associated with UF or UL in patients who have endometrial cancer, and this research focus is on UF development in healthy people (Rafnar, T., et al., 2018).

The design of this research will be to examine the gene expression data collected from the GEO data repository of eight separate studies involving healthy human uterine myometrial tissue and human UF or UL tissue (Robtree JS, Jelinsky SA, Harris HA, Choe SE et al, 2009; Delaney MA, Wan YW, Kim GE, Creighton CJ et al., 2017; Hoffman PJ, Milliken DB, Gregg LC, Davis RR et al., 2004; Miyata T, Sonoda K, Tomikawa J, Tayama C et al., 2015; Quade BJ, Mutter GL, Morton CC, 2004; Teixeira, JM, 2018; Vanharanta S, Pollard PJ, Lehtonen HJ, Laiho P et al., 2006; Zavadil J, Ye H, Liu Z, Wu J et al, 2010). The R

biostatistics software Bioconductor will be used for machine learning on the GEO data. Aiming to find significant SNPs associated with UF among all the GEO gene samples of UF and healthy donors that are either in chromosomes 11, 12, 17, and 22 or that are the top genotypes of CCDC57, BET1L, TNRC6B, FASN, HGMA2, or CYTH4. Methods similar to Zhang, D., et al., 2012 and Eggert, S., et al., 2012 will be used for developing differential expression MA plot for highly up or down regulated genes between groups of UF and healthy samples, box plots per gene, QQ plot (with top SNPs being far from the diagonal), PCA analysis of variable genes, link analysis of top genes to their closest genes to elucidate the functional inhibition that may or may not occur, the highest expressed up or down regulated genes between the healthy and UF samples, and other analytics as needed using only the packages in Bioconductor. Proving there is a significant association with UF development in healthy people using gene expression analysis of a multiracial pool of samples totaling 169 of 78 healthy myometrial tissue and 91 UF tissue is the aim of this research proposal. Using statistics and machine learning on 169 samples it is possible to find the most highly expressed genes between the two groups of non-UF and UF tissue, then use machine learning to develop a network of linked genes that impact other genes along the chromosomes associated with UF development. Using a training set of 70% of the 169 samples randomly selected and a test set of the other 30% will give a training set of 118 samples and a testing set of 51 samples. The model built produces accuracy of a certain threshold like 90%, then it will be assumed a good measure of which genotypes are associated with UF. Models could be built for linear regression, random forest, k-nearest neighbor, and other regression models like Bayesian P has packages not exclusive to Bioconductor to operate predictive analytics on the data gathered. If the accuracy compare across all models is below 50%, then the original chromosomal markers need to be identified for the hundreds of genes left out when originally searching for SNPs associated with UF as some studies suggested (Bondagji, N., et al., 2017; Aissani, B., et al., 2015).

## References

8

Aissani, B., Zhang, K., and Wiener, H. (2015). Evaluation of GWAS candidate susceptibility loci for uterine

16

leiomyoma in the multi-ethnic NIEHS uterine fibroid study. *Frontiers in Genetics*, 6, 241.

DOI:10.3389/fgene.2015.00241

49

This research study does a meta-analysis between the NIEHS-UFA National Institute of

48

Environmental Health Study, The RFTS or Right from The Start and BioVu study on European

Americans, the studies on Japanese, Chinese, Africans, and other ethnic studies published to

analyze the significant gene markers in all the SNP genome wide association studies associated

with uterine fibroids. The study found that SNPs in TNRC6B labeled rs139909 and rs138089 in

17

the NIEHS-UFA study to be significantly associated with uterine leiomyomas (UL) of European

Americans. Also, finding that rs12484776 to be significant in the RTFS study for tumor size and

risk. The African American population didn't show this gene target of SNPs to be significant. The

BET1L UL risk was reproduced in the Han Chinese, Japanese, and European American

populations but not the African American population. This article also suggested THDS7B and

the ECM or extracellular matrix components are factors in uterine fibroid risk and growth. The

study discussed that there could be misclassification in the control group of uterine fibroids in

the African population based on the self-reported method in the BWHS or Black Women's

Health Study. If that misclassification was true, then there could be a new significant finding in

the TNRC6B and BET1L genotypes being associated with tumorigenesis in UL among African

American populations as they have shown for Chinese, Japanese, and European populations.

Bioconductor, version 3.8, (2019). Bioconductor: Open Source Software for Bioinformatics. Retrieved

March 3, 2019 from: <https://www.bioconductor.org/install/>

This is an open source and free software to analyze genomic and other biological data using biostatistics. It operates with R software version 3.5.0 or higher as this version of Bioconductor is version 3.8. There are more than 1600 biostatistics packages available using this software for 28. R is also an open source free software for statistical analysis and machine learning. R is easily able to be updated and when compared to Python it is not as package version dependent, nor platform dependent like Python2 and Python3 platforms. R is easier to upload with less time spent downloading and configuring packages that need to be rolled back or updated. The coding of R is similar to python and other coding languages.

45

21

Bondagji, N., Morad, F., Al-Nefaei, A., et al. (2017). Replication of GWAS loci revealed the moderate effect of TNRC6B locus on susceptibility of Saudi women to develop uterine leiomyomas. *Journal of Obstetrics and Gynaecology*, 43(2):330-338. DOI:10.1111/jog.13217

18

This study used Saudi Arabian women with uterine leiomyomas (UL) and without UL to study the putative genes shown in Japanese and European populations to be a risk factor for UL. The results showed that the SNPs for the alleles in TNRC6B as rs12484776 is a risk factor in Arab populations of UL and that the SNPs rs2280543 in BET1L, as well as rs7913069 of LOC102724351 and rs1056836 of CYP1B1 also increase the risk of UL in Arab populations. The study reports on the current findings on UL such as the gene MED12 is the only gene directly shown to have a causal relationship to UL. Also, TNRC6B and BET1L are risk factors for UL in Japanese populations and in developing UL in European populations, and that FASN and CCDC57 are risk factors for UL in Europeans and Australian populations. This study also examined the location of the fibroid as being in the uterine wall (intramural), below the mucosal layer of the uterus (subserous), or below the endometrial layer of the uterus (submucosal). Most of this population had more than one fibroid (79%) compared to one fibroid (21%). The study admits to not being large enough

given its roughly 100 samples of UL and non-UL case and control groups for applying the findings to the entire Arab population. Results also did not analyze whether the rs12464776 SNP of TNRC6B is a biomarker for the number of fibroids a patient has, the location of each fibroid a patient has within the layers of the uterus, or the size of the fibroid in each patient due to perceived statistical errors in running those tests. The study also reports that although TNRC6B was found to be a risk factor in UL in Arab populations it is a possibility that the missing gene expression data and linked gene region could be the real pathogenesis of UL because this study only examined the five genes shown in previous UL studies to be risk factors of UL in Japanese and European populations. And these five SNPs were screened for frequencies while ignoring the millions of other genes that showed variations in humans so the results could be omitting other genes that might be risk factors of UL in Arab populations.

1

Crabtree JS, Jelinsky SA, Harris HA, Choe SE et al, (2009). Comparison of human and rat uterine

leiomyomata: identification of a dysregulated mammalian target of rapamycin pathway. *Cancer Research*, 69(15):6171-8. PMID:19622772

This database was published August 20, 2009 and is of rat and human uterine samples. It uses

14

the Gene Expression Omnibus (GEO) database found at <https://www.ncbi.nlm.nih.gov/geo/> by

entering the GEO Access ID of GSE13319. The platform this particular data sequence uses is the platform GPL570. There are 23 healthy human myometrial tissue samples and 23 human uterine leiomyoma tissue samples. Included with the human samples are the samples of rat myometrial

34

and uterine leiomyoma from rat tissue. The human tissue samples are from the Affymetrix

Human Genome U133 Plus 2.0 Array. GPL570 platform was made public on November 7, 2003.

All the gene expression arrays of the human samples in this data will be added to the human samples from seven other GEO data series of human healthy myometrial control groups and

human uterine leiomyoma (UL) case groups. After adding the other GEO data of seven other studies, there will be a total of 78 healthy myometrial samples and 91 UL samples for a total of 169 human samples.

18

Delaney MA, Wan YW, Kim GE, Creighton CJ et al., (2017). A Role for Progesterone-Regulated sFRP4

24

Expression in Uterine Leiomyomas. *Journal of Clinical Endocrinology and Metabolism*, 102(9):3316-3326. PMID:28637297

This is a Gene Expression Omnibus (GEO) data series made public on June 6, 2018 and last updated on June 8, 2018. GEO can be found at this link: <https://www.ncbi.nlm.nih.gov/geo/>.

The GEO Access ID is GSE95101 using GEO platform GPL13376. The study using this sample found secreted frizzled-related protein 4 (sFRP4) as a protein product of progesterone receptor (PR) activation during the proliferative phase of the menstrual cycle. The tissue is from frozen hysterectomy samples during different phases of menstruation, not all from the same sample at separate menstrual phases. The leiomyoma tissue samples are from the same hysterectomy sample of normal myometrial tissue that is next to the leiomyoma. The tumors were 3-6 cm in diameter and the healthy tissue was not more than 1 cm away from the nearest fibroid. The 17 healthy myometrial samples and the 17 uterine leiomyoma (UL) samples will be an addition to the database of healthy and UL genes for gene expression analysis. After adding the other GEO data of seven other studies, there will be a total of 78 healthy myometrial samples and 91 UL samples.

18  
29  
33  
20

Edwards, T., Hartmann, K., Edwards, D. (2013). Variants in BET1L and TNRC6B associate with increasing fibroid volume and fibroid type among European Americans. *Human Genetics*, 132(12).

DOI:10.1007/s00439-013-1340-1

This research used the RFTS – Right from the Start study and BioVu DNA data specifically at the BET1L, TNRC6B, and SLK gene targets discovered in a Japanese population study on uterine fibroids to analyze the European American population risk, size, and location of uterine fibroids. It found that there wasn't significance in the SLK gene target due to MAF or minor allele frequency below threshold value. However, it did confirm the significance of BET1L in uterine location of fibroid in layers of the uterus as either subserous (below the mucosal layer of the uterus) or intramural (in the uterine muscle). Confirmation that an associated risk of uterine fibroid with both genotypes of BET1L rs2280543 and TNRC6B rs12484776 in European populations was one conclusion of this study. Another conclusion of this study was that TNRC6B genotype rs12484776 is significant as a biomarker for the size of UL in European populations.

42

9

Eggert, S., Huyck, K., Somasundaram, P., et al., (2012). Genome-wide linkage and association analyses implicate FASN in predisposition to uterine leiomyomata. *American Journal of Human Genetics*, 91(4): 621–628. DOI:10.1016/j.ajhg.2012.08.009

This study used white female sisters having medically diagnosed uterine leiomyomas (UL) and other family members totaling 385 pairs of sisters from 261 families with 1,103 individuals to search for gene targets in UL that have cytogenetic abnormalities. Two other studies involving sisters, twins, and moms was also used in this study by adding the data to this study's analysis of gene targets and abnormalities in UL. The gene Fatty acid Synthase (FASN) was found in prostate, breast, colon cancers, and also found in UL. Research from this study showed that impairing this gene's receptors can slow down colon and breast cancer. FASN is highly expressed in hormone sensitive cells, found to be regulated in transcriptional and post-transcriptional levels. Primary transcription factor is sterol-regulating-element-binding-transcription-factor 1 (SREBP-1) activated downstream of growth hormone and hormone receptors. Inhibiting FASN

7

7

has led to cancer cell line apoptosis, tumor growth ceasing, and has shown little or minimal effects on surrounding normal cells. There is a connection between FASN and the P13K/Akt signaling pathway which is a commonly dysregulated pathway in human cancers. The major and minor allele rs4247357 of gene coiled-coil domain containing 57 or CCDC57 was genotyped from the healthy and fibroid tissues of 20 tumors from 12 women of which half had the major allele and the other half had the minor allele. In order to satisfy a genome-wide association significance level the P value of .05 is not good enough and having even a P value of  $10^{-4}$  or .0005 is not significant enough to mark a SNP as a gene target for UL risk. After analyzing the CCDC57 alleles and using linkage analysis, a 35 Mb genomic region was found containing hundreds of genes that possibly pose a risk for UL. Candidate SNPs found on chromosome 17 in this region are FASN, CCDC57, and SLC16A3. The rs4247357 SNP of CCDC57 was found to be at a genome-wide association significance level for UL risk in white patients using the Finding Genes for Fibroids (FGFF) study, the Women's Genome Health Study (WGHS) and the Australian cohort study.

Hellwege, J. N., Jeff, J. M., Wise, L. A., Gallagher, C. S., Wellons, M., Hartmann, K. E., ... Velez Edwards, D. R. (2017). A multi-stage genome-wide association study of uterine fibroids in African Americans. *Human Genetics*, 136(10), 1363–1373. DOI:10.1007/s00439-017-1836-1

The research in this study focused exclusively on African Americans using the 23andMe.com database and the genome wide association studies (GWAS) database. This study found an increase in uterine fibroid (UF) risk in the SNP rs739187 of cytohesin 4 (CYTH4). Results showed a lower predicted gene expression in the thyroid tissue was significant in determining UF risks in African Americans. The study recognizes thyroid problems such as overt hypothyroidism, thyroid nodules, and thyroid cancer are associated with UF and that genes HMGA2 and PLAG1 are

associated with UF and strongly correlated with thyroid tumors. None of the BET1L, TNRC6B, or SLK genotype SNPs for UF risk in the Japanese population were found to be a risk for UF in African American populations. Also, women not at risk of developing fibroids (age beyond menopause) and self-reporting of UF were part of the control group which may have affected the results due to misclassification. UF cases are reported only when symptomatic or clinically obvious, which according to Aissani, B., et al., 2015 is in 33% of patients who have UF. This measure of symptomatic UF changes to 50% for Liu, B., et al., 2018 and changes to 20-25% symptomatic UF for Eggert, S., et al., 2012. So, it is clear that given those measures for UF patients that are seeking help for their UF, as many as 80% of other UF populations do not either know they have a UF or are not reporting they have a UF. Therefore, it is likely misclassification of UF occurred in this self-reporting control group for UF. This case study implies heterogenetic risk factors for UF by presenting evidence that the genotypes of TNRC6B does not associate with risk of UF in African populations when it has been shown in Chinese, Japanese, Saudi, European, and other white races to be associated with risk of UF. At the same time provides contrapositive evidence of a genotype that is exclusively associated with risk of UF in African Americans while not being associated with UF risk in Chinese, Japanese, Saudi, European, and other white races. That genotype is rs739187 of CYTH4, and this study showed it is significant in being associated with risk for UF in African American populations. This study also confirms the decrease in angio-tension for renal homeostasis (AGT) in thyroid tissue and high ALDH2 (involved in alcohol metabolism) as risk factors for UF in African American populations.

1

Hodge, J., Kim, T., Dreyfuss, J., Somasundaram, P., et al., (2012). Expression profiling of uterine

leiomyomata cytogenetic subgroups reveals distinct signatures in matched myometrium:

transcriptional profiling of the t(12;14) and evidence in support of predisposing genetic

heterogeneity. *Human Molecular Genetics*, 21. 102312–2329. DOI:10.1093/hmg/dds051

This study used nine female patients having more than one uterine leiomyoma (UL) or multiple

uterine leiomyomas (MUL) to do a paired comparison study of the recurrent chromosome

5

abnormalities of t(12;14) and non-t(12;14) genomic regions of UL. The study confirmed the

putative gene HGMA2 of the high mobility group being over expressed in UL and as the most

significant altered gene. Using unsupervised PCA the variation between patients was greater

5

than the t(12;14) UL and non-t(12;14) UL. Also, patient variation was a greater classifier than

5

del(7q) UL and t(12;14) UL. This suggests personalized UL treatment because of the

heterogeneity in patient demographics having UL. The patients were of different racial origin in

one cohort which differs from the other UL studies looking for genome wide associations in each

race. And all patients had more than one UL each. There are 7.5% of UL having translocation

5

differences in the 12q14-15 region of the genome, specifically at t(12;14)(q14-15;q23-24). The

samples were weighted according to the percent of cells in the sample which controlled patient

5

demographic variability. The results of this study created a gene expression profile for the

subgroup t(12;14) of UL. This study made it clear that genetic heterogeneities exist in the

pathogenesis of UL and the first step to treating UL is to identify the genetic profiles for UL. The

way the UL gene is listed is different from other studies that list the chromosome location and

the base pair location along that region showing high expression in the genes between control

and patient groups. The other studies used a genome expression omnibus type of identification

or one recognized by the National Center for Bioinformatics Information like Encode or UCSC

genomic data repositories. It used fluorescence in situ hybridization and mentioned HGMA2 is

on chromosome 12 and responsible for allowing transcription factors to reach their target genes

and restricted mostly to proliferative embryonic tissue that include the derivatives of the

mesenchymal tissue. This includes the myometrium from which uterine fibroids are found to grow. HGMA2 has been found expressed in the myometrium in vivo.

12

Hoffman PJ, Milliken DB, Gregg LC, Davis RR et al., (2004). Molecular characterization of uterine fibroids and its implication for underlying mechanisms of pathogenesis. *Fertility and Sterility*, 82(3):639-

49. PMID:15374708

14

This resource is data collected from the Gene Expression Omnibus (GEO) archive located at:

<https://www.ncbi.nlm.nih.gov/geo/>. The data was published to GEO October 17, 2003. The GEO

platform ID is GPL96 and the GEO Access ID is GSE593. In this data there are five healthy myometrial samples and five uterine leiomyoma (UL) samples used to compare gene expression 41 using the samples with a fold change greater than 1.5 or less than -1.5. The results that were significant were those that had a P value less than or equal to 0.05. The human gene samples of healthy myometrial and UL gene samples will be added to the database of the other seven GEO data sets of healthy human and UL human gene samples. In total there will be 169 human samples for comparing genes differentially expressed the most between 78 healthy myometrial samples and 91 UL samples.

23

Liu, B., Wang, T., Jiang, J., Li, M., Ma, W., Wu, H., Zhou, Q. (2018). Association of BET1L and TNRC6B with uterine leiomyoma risk and its relevant clinical features in Han Chinese population. *Scientific Reports*, 8,7401. DOI:10.1038/s41598-018-25792-z

The female case and control groups for uterine fibroid (UF) or synonymously uterine leiomyoma (UL) in this study are not related and all from the Han Chinese population. The genes BET1L and TNRC6B were found to be risk factors of UL in Han Chinese women. SNPs associated with these two genes also showed significance in the number of fibroids and the size of the fibroids. This 16

study found the SNP for TNRC6B called rs12484776 with genotypes GG, GA, and AA (most common genotype is AA) is significantly associated with the size of the UL. It also found that the SNP rs2280543 associated with BET1 has two genotypes TT and CT (most are CC) that are significantly associated with the number of ULs one patient has. Neither of these gene targets are associated with gene expression in the uterus using eQTL-expression quantitative trait loci from gtexportal.org/home or the gtex database that uses healthy tissue samples for gene expression mapping. This study also compares itself to the Japanese, Saudi Arabian, and European population studies on UL from 2011-2017.

11

Miyata T, Sonoda K, Tomikawa J, Tayama C et al., (2015). Genomic, Epigenomic, and Transcriptomic Profiling towards Identifying Omics Features and Specific Biomarkers That Distinguish Uterine

Leiomyosarcoma and Leiomyoma at Molecular Levels. *Sarcoma* 2015. 412068. PMID: 27057136

47

The gene expression omnibus (GEO) repository of data located at

37

<https://www.ncbi.nlm.nih.gov/geo/> was accessed for the data of healthy myometrial and uterine leiomyoma (UL) gene data. This specific study also includes cancerous leiomyomas called leiomyosarcomas that will not be included. The GEO Access ID is GSE68295 using GEO platform GPL6480 and the data was published to GEO on February 9, 2017. In the GSE68295 data series, the samples are separately identified as UL, uterine leiomyosarcoma (cancerous UL), or healthy myometrial or uterine tissue. This is how all the data is separated in the other seven GEO resources used. In each GEO series prepended with 'GSE' and followed by the ID number of that study, there will be a list of separate data sets that identify what type of data it is. These data sets begin with 'GSM' and are followed by their data set ID number. In this study, the GSM IDs are clearly explained as to what data it is. For instance, GSM1667147 is titled, 'uterine leiomyoma tissue from case 4,' and GSM1667145 is titled, 'uterine normal myometrium tissue'

from case 2.' There are three healthy and three UL samples totaling six samples to add to the other seven GEO data on healthy myometrium gene samples and UL gene samples. The six gene samples of leiomyosarcoma will not be added to the data needed for research on gene expression of most differentially expressed genes between normal and UL tissue. When all of the eight GEO resources used are extracted for human normal myometrium and human UL gene samples is completed, the database for studying UL gene expression will total 169 gene samples of 78 healthy myometrium and 91 uterine leiomyomas.

19

Quade BJ, Mutter GL, Morton CC, (2004). Comparison of Gene Expression in Uterine Smooth Muscle

Tumors. Gene Expression Omnibus. GEO Accession ID: GSE764. Retrieved March 2019 from:

4

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>

46

The above data is from the Gene Expression Omnibus (GEO) repository with link provided above. The data was published January 1, 2004 with no listed research article associated with it judging from the GEO citation information missing. The GEO Access ID is GSE764 using GEO platform GPL80. This data does not have an attached study that was published. The data for this GEO series has seven uterine leiomyoma (UL) gene samples and four healthy myometrial gene samples all from humans as are the other GEO resources except the one using rats. This data is similar to another data series in comparing tissue from not only UL and healthy myometrial tissue but also using cancerous leiomyosarcoma and uterine cancerous myometrial tissue. The data sets all clearly identify which data sets beginning with 'GSM' and followed by the data set ID number are either 'Myo' for healthy myometrial gene samples, 'Leio' for UL gene samples, 'LMS' for leiomyosarcoma, or 'exULMS' for extra cancerous uterine leiomyosarcoma. Only the data sets that begin with 'Myo' or 'Leio' will be used. From this GEO database there will be gene samples from four healthy myometrial and seven uterine leiomyomas. This data contributes to

the accumulation of data on healthy uterine and uterine leiomyoma gene samples for researching high differentially expressed genes between normal and UL human tissue. There will be a total of 78 healthy or normal myometrial gene samples and 91 UL gene samples after all eight of the GEO data studies are collected.

R, (2019). CRAN: Comprehensive R Archive Network. R, version 3.5.2, for Windows 64-bit Operating System. Retrieved March 3, 2019 from: <https://cran.cnr.berkeley.edu/>

This source is the R statistical programming software that will be used to install Bioconductor, and access statistical and machine learning package libraries as needed to conduct research on gene expression data from the Gene Expression Omnibus (GEO) online data repository. This version is for the Windows 64-bit operating system and is 79 megabytes in size. Bioconductor has to be installed from within R once installed.

Rafnar, T., Gunnarsson, B., Stefansson, O., et al., (2018). Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. *Nature Communications*, 9:3636. DOI:10.1038/s41467-018-05428-6

The research done in this article involved a meta-analysis of two genome wide association studies (GWAS) of uterine leiomyomas using Icelandic and English European females. The patients with uterine leiomyomas (UL) are the case group and the volunteers without UL are the control group. Using two separate studies, one study was on genes expressed in cancers and other benign tumors that are also expressed in UL. The other study was on the putative loci regions associated with hormone related diseases and the changes in these loci in UL. This research elucidated the relationship that hormones have on uterine fibroid growth and made explicit the common genes being expressed between UL, cancer and benign tumors elsewhere

in the body. The information about hormonal therapy to treat symptoms of estrogen responsive leiomyomas before hysterectomy can cause the symptoms to recur was found in this article. The TNRC6B and the BET1L genes were found to also be associated with uterine leiomyomas in this study. But the BET1L gene found in Japanese populations and the CYTH4 gene found in African American populations were not found to be associated with uterine leiomyomas in this study on European women. This study on Europeans confirmed one of the endometrial cancer genes associated with uterine leiomyomas is r10917151 of CDC42/WNT4 and was able to exclude the other seven genes previous genome wide association studies found to be associated with uterine leiomyomas and cancer. They also show in a table of polygenic risk scores that there is a significant association of ULs in patients with thyroid cancer (R-squared = 21% and P value: 3.0\*10^-5), endometriosis Stage III and IV (R-squared = 11% and P value = 4.1\*10^-3), and kidney cancer (R-squared = 10% and P value = 2.43\*10^-3). This research on Europeans, shows a connection between thyroid dysfunctions and thyroid cancer that the research done on African American populations also found to be associated with uterine fibroid risk.

27

Teixeira, JM, (2018). Integrated epigenome, exome and transcriptome analyses reveal molecular subtypes and homeotic transformation in uterine fibroids [DNA methylation]. Gene Expression Omnibus, GEO Accession ID: GSE120854. Retrieved March 2019 from:

17

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>

This source is from the Gene Expression Omnibus data repository on gene expression (GEO). The data can be accessed with GEO Access ID: GSE120854. This data series was published October 5, 2018 on GEO and the citations to any published article using this data were missing. The findings replicate similar findings on HGMA2 in another research article by Hodge, J. et al., (2012). All gene samples are from heterogeneous ethnic origins. The study looked at MED12, HGMA2,

HGMA2hi, HGMA1, HGMA1hi, and HOXA13. The GEO summary says that HOXA13 was found to be highly up regulated in uterine fibroid tissue. In this data series there are 34 total samples of 10 normal myometrial gene samples and 24 UL gene samples. All of these gene samples will be added to the collection of data on normal and UL gene samples. When all GEO data is collected there will be 78 normal myometrial gene samples and 91 UL tissue samples.

22

Vanharanta S, Pollard PJ, Lehtonen HJ, Laiho P et al., (2006). Distinct expression profile in fumarate-hydratase-deficient uterine fibroids. *Human Molecular Genetics*, 15(1):97-103. PMID:16319128

This reference is used for the data collected from the Gene Expression Omnibus data repository on gene expression (GEO). GEO can be accessed at

10

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>. The GEO Access ID is GSE2724 using GEO platform GPL96 with this data made publicly available December 20, 2005. In this data series there are seven uterine leiomyoma (UL) samples and 11 healthy myometrial gene samples that will be used to add to the data of normal and UL gene samples in studying gene expression between the two groups. This data series has labeled the UL samples to include an 'm' for fibroid tissue, and an 'n' for normal myometrial tissue. The 'm' is actually for mutated fumarate hydrase (FH) as this study was examining the mutated FH in uterine fibroids (UF) to find a connection of FH expression in UL or its synonym UF. When this data is added to the data from the other GEO data on normal and UL gene samples there will be a total of 169 gene samples comparing 78 normal uterine tissue to 91 UL tissue. The completed database of normal and UL gene samples will be used for data analysis.

3

Zhang, D., Sun, C., Ma, C., et al. (2012). Data mining of spatial-temporal expression of genes in the human endometrium during the window of implantation. *Reproductive Sciences*, 19(10). 1085-98. DOI:10.1177/1933719112442248

This article uses data mining to extract gene expression data from the microarrays of the  
19 ArrayExpress gene expression data repository and the Gene Expression Omnibus (GEO) database. Both are available online. In total this research gathered 45 samples of four different stages of fertility in women during the implantation window of fertility. This study and the study by Hodge, J. et al., 2012 are two studies that perform gene expression analysis independent of race of uterine leiomyoma (UL) donor tissue origin and of having only a UL. Almost 200 potential biomarkers were discovered and tested to predict the likelihood of a woman capable of becoming pregnant to do just that. The experimental study of how the samples were collected that were uploaded into GEO and ArrayExpress was detailed and bioinformatic statistics was conducted using a software called GeneSifter. This software was used to perform differential gene expression analysis and validate for fold-change of at least 10 so that the most highly  
39 changed genes could be compared. Principal Component Analysis (PCA) was done to evaluate the multivariate data complexities as well as hierarchical clustering after PCA on the data files to conditionally categorize the genes. Afterwards, a software program called Ingenuity was used to compare the fold-change values of genes at least equal to two for the genes that are potential biomarkers along the canonical pathways the Ingenuity Pathway Alignment library produced.  
31 This was verified by Fisher Exact Test for P values of at most 0.05. Network analysis was then used on the top 35 genes having at least a fold-change value of 10 from the GeneSifter analysis done earlier to visualize the network of connections each gene has to other genes using the algorithm generated by Ingenuity software. To validate the data, genes that were selected for validation were either significantly different from the same gene in another array, had a very  
26 low P value, or a high fold-change in the previous analysis. Statistical analysis of the genes being validated was performed using the regular expression algorithm, the Least Significant Difference (LSD), and one-way analysis of variance. Those genes resulting in a P value less than or equal to

.05 were labeled statistically significant. The chromosomes 4, 9, and 14 were found to be the chromosomes with the most differentially expressed genes.

3

Zavadil J, Ye H, Liu Z, Wu J et al, (2010). Profiling and functional analyses of microRNAs and their target gene products in human uterine leiomyomas. PLoS One 2010 Aug 24;5(8):e12362. PMID: 20808773

This is a Gene Expression Omnibus (GEO) data series of information published July 24, 2010.

10

GEO can be accessed at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> and using the GEO Access ID of GSE23112 in the search box will give a list of the data sets in this study. The platform used in GEO is GPL96. The study examines gene expression qualities of miRNA and mRNA in tumorigenesis and tumor regulation of uterine fibroids (UF) also known as uterine leiomyomas (UL). In total this data contributes 10 samples of five healthy uterine myometrial gene samples and five UL gene samples. After the addition of these data sets to the other seven GEO data sets on normal and UL gene samples, there will be a database of 78 healthy uterine myometrial gene samples and 91 UL gene samples.

# Full Proposal Week 7 Revised Research Proposal

## ORIGINALITY REPORT



## PRIMARY SOURCES

- |   |  |     |
|---|--|-----|
| 1 | <a href="http://repositorio.unesp.br">repositorio.unesp.br</a>   | 1 % |
| 2 | Submitted to The Chicago School of Professional Psychology   | 1 % |
| 3 | <a href="http://static.pubmed.gov">static.pubmed.gov</a>   | 1 % |
| 4 | <a href="http://elifesciences.org">elifesciences.org</a>   | 1 % |
| 5 | <a href="http://hmg.oxfordjournals.org">hmg.oxfordjournals.org</a>   | 1 % |
| 6 | Submitted to Australian Catholic University<br>Student Paper   | 1 % |
| 7 | Stacey L. Eggert, Karen L. Huyck, Priya Somasundaram, Raghava Kavalla et al.<br>"Genome-wide Linkage and Association Analyses Implicate FASN in Predisposition to Uterine Leiomyomata", The American Journal of Human Genetics, 2012 | 1 % |

- 8 journal.frontiersin.org 1 %  
Internet Source
- 9 espace.library.uq.edu.au <1 %  
Internet Source
- 10 www.frontiersin.org <1 %  
Internet Source
- 11 Hideaki Tsuyoshi, Yoshio Yoshida. "Molecular biomarkers for uterine leiomyosarcoma and endometrial stromal sarcoma", Cancer Science, 2018 <1 %  
Publication
- 12 Submitted to Universiti Sains Malaysia <1 %  
Student Paper
- 13 genepi.qimr.edu.au <1 %  
Internet Source
- 14 www.spandidos-publications.com <1 %  
Internet Source
- 15 www.ucl.ac.uk <1 %  
Internet Source
- 16 Bailing Liu, Tao Wang, Jue Jiang, Miao Li, Wenqi Ma, Haibin Wu, Qi Zhou. "Association of BET1L and TNRC6B with uterine leiomyoma risk and its relevant clinical features in Han Chinese population", Scientific Reports, 2018 <1 %

- 17 link.springer.com <1 %  
Internet Source
- 18 www.happyhormonecottage.com <1 %  
Internet Source
- 19 www.biomedcentral.com <1 %  
Internet Source
- 20 erc.endocrinology-journals.org <1 %  
Internet Source
- 21 onlinelibrary.wiley.com <1 %  
Internet Source
- 22 Gayun Chan-Smutko. "Familial Renal Cell Cancers and Pheochromocytomas", Principles of Clinical Cancer Genetics, 2010 <1 %  
Publication
- 23 obgyn.onlinelibrary.wiley.com <1 %  
Internet Source
- 24 Vladimir Wolf, Guo Ke, A.M Dharmarajan, Wolfgang Bielke, Livia Artuso, Susanne Saurer, Robert Friis. "DDC-4, an apoptosis-associated gene, is a secreted frizzled relative", FEBS Letters, 1997 <1 %  
Publication
- 25 Xavier Santamaria, Aymara Mas, Irene Cervelló, Hugh Taylor, Carlos Simon. "Uterine <1 %

stem cells: from basic research to advanced cell therapies", Human Reproduction Update, 2018

Publication

---

- 26 spandidos-publications.com <1 %  
Internet Source
- 27 "2019 | Scientific Abstracts", Reproductive Sciences, 2019 <1 %  
Publication
- 28 lexicall.widged.com <1 %  
Internet Source
- 29 Julien Levy, Cecilia Tamborindeguy, Giridhar Athrey, Douglas C. Scheuring, Jeffrey W. Koym, J. Creighton Miller. "Transcriptome of Russet Norkotah and its clonal selection, TXNS278", BMC Research Notes, 2018 <1 %  
Publication
- 30 Nabeel Saleem Bondagji, Fatima Amanullah Morad, Afnan Abed Abdullah Al-Nefaei, Imran Ali Khan et al. " Replication of GWAS loci revealed the moderate effect of locus on susceptibility of Saudi women to develop uterine leiomyomas ", Journal of Obstetrics and Gynaecology Research, 2017 <1 %  
Publication
- 31 www.earticle.net <1 %  
Internet Source

- 
- 32 discovery.ucl.ac.uk **<1 %**  
Internet Source
- 
- 33 G. Suo. "Murine Xenograft Model for Human Uterine Fibroids: An In Vivo Imaging Approach", Reproductive Sciences, 09/01/2009 **<1 %**  
Publication
- 
- 34 Internet Source **<1 %**
- 
- 35 Liping Xia, Yan Liu, Yan Fu, Shengyi Dongye, Dewei Wang. "Integrated analysis reveals candidate mRNA and their potential roles in uterine leiomyomas", Journal of Obstetrics and Gynaecology Research, 2017 **<1 %**  
Publication
- 
- 36 www.ighz.edu.pl **<1 %**  
Internet Source
- 
- 37 academic.oup.com **<1 %**  
Internet Source
- 
- 38 Virgil S. Bideau, Angela T. Alleyne. "A preliminary study of fatty acid synthase gene and the risk of uterine leiomyoma in an Afro-Caribbean female population", Meta Gene, 2019 **<1 %**  
Publication
- 
- 39 Dan Zhang, Cuixiang Sun, Chengbin Ma, **<1 %**

Haiyan Dai, Wei Zhang. "Data Mining of Spatial–Temporal Expression of Genes in the Human Endometrium During the Window of Implantation", Reproductive Sciences, 2012

Publication

40

"Abstracts", Journal of Thoracic Oncology, 2009

<1 %

Publication

41

gxreferences.blogspot.com

Internet Source

<1 %

42

Submitted to Grand Canyon University

Student Paper

<1 %

43

Molecular and Integrative Toxicology, 2015.

Publication

<1 %

44

"Encyclopedia of Cancer", Springer Nature, 2017

Publication

<1 %

45

Ghada M. A. Ajabnoor, Nesma Amin Mohammed, Babajan Banaganapalli, Layla Saleh Abdullah et al. "Expanded Somatic Mutation Spectrum of MED12 Gene in Uterine Leiomyomas of Saudi Arabian Women", Frontiers in Genetics, 2018

Publication

<1 %

46

Jian Shang, Fan Wang, Pengfei Chen, Xiaobing Wang, Feng Ding, Shi Liu, Qiu Zhao. "Co-

<1 %

expression Network Analysis Identified COL8A1 Is Associated with the Progression and Prognosis in Human Colon Adenocarcinoma", Digestive Diseases and Sciences, 2018

Publication

---

- 47 M. Huss. "Introduction into the analysis of high-throughput-sequencing based epigenome data", *Briefings in Bioinformatics*, 2010 <1 %
- Publication
- 48 Zhang, Kui, Howard Wiener, and Brahim Aissani. "Admixture mapping of genetic variants for uterine fibroids", *Journal of Human Genetics*, 2015. <1 %
- Publication
- 49 Aissani, Brahim, Kui Zhang, and Howard Wiener. "Evaluation of GWAS candidate susceptibility loci for uterine leiomyoma in the multi-ethnic NIEHS uterine fibroid study", *Frontiers in Genetics*, 2015. <1 %
- Publication
- 

Exclude quotes

Off

Exclude matches

Off

Exclude bibliography

Off

# Full Proposal Week 7 Revised Research Proposal

## GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

17 /20

### Instructor

It is obvious that you have devoted a lot of energy into investigating what is already known about this topic in the literature. The background information that you have pulled together is vast and will give you a good foundation for your project.

The portion of the proposal that deals with your research, however, is less well developed. I think that you imply your intentions (and if I "read between the lines," I can make an educated guess about what you propose), but unfortunately the research approach does not come across as clearly as you may have intended. Some of the necessary concepts I believe are included in the context of your annotations in the bibliography, but the full integration is missing in the proposal section itself. In short, the "biology" part of the project is defined pretty well, but the "data" part of the project is less clear.

You will need to more clearly address the following:

1.) You intend to build your own data set, drawing from individual entries in GEO, correct? Assuming that's the case, some unanswered questions:

A.) What are characteristics of the data in each of the individual sets (what info is known for each individual)? Are the individual sets compatible?

B.) What procedure will be used for building this new set? This will be an important part of the method you will need to detail.

2.) Once you have your data to work with, it seems that you intend to use machine learning to build some models. However, it is less explicit what you are aiming to learn in each of these analyses, and perhaps the problem comes from not having a fully defined data set. Will you only be looking at differences in gene transcription, without regards to other diagnosis/tissue information? Or are some other clinical values going to be used to stratify?

3.) How is your proposed project complementary to the existing literature but not redundant? It will be important to make clear that you are not just re-producing what is known. I do not mean to imply that there is no novelty to your project, but the uniqueness has not been conveyed.

Finally, please make sure to clean up APA-format related items in the future.

---

PAGE 1

---



### **Comment 1**

Start a new sentence here -- otherwise it currently is a run-on sentence.



### **Comment 2**

These are all females, correct? If you re-word this sentence, up front you can make clear that you are considering/discussing females from different race groups/isolated geographic populations. Otherwise, some are clearly stated as females while other groups in your list are not.



### **Comment 3**

For in text citations, get in the habit of providing author last names only (without the initial) -- then have the last name and initial on your reference page.

Also, make sure that you are adhering to formatting rules. If there are less than six authors on a work, all author names must be listed the first time (you cannot jump directly to using "et al." the first time, per APA rules).

See [https://owl.purdue.edu/owl/research\\_and\\_citation/apa\\_style/apa\\_formatting\\_and\\_style\\_guide](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide) for help.



### **Comment 4**

If these terms are synonymous, then you should explain once that the terms are interchangeable and then subsequently only use one of the two. Later in this paragraph, you switch between using the different terms and that can be confusing for a reader -- it implies that they are two distinct concepts.



### **Comment 5**

In what circumstance would females be receiving hormone therapy? This is not ubiquitous for all females -- you cannot assume that the reader would understand why hormone therapy would be in use (to then be discontinued).



## Comment 6

Make sure that you are using the correct format (APA) when making these citations in the text. See [https://owl.purdue.edu/owl/research\\_and\\_citation/apa\\_style/apa\\_formatting\\_and\\_style\\_guide.html](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide.html)



## Citation Needed

Cite Source:

Please use the link below to find links to information regarding specific citation styles:  
[http://www.plagiarism.org/plag\\_article\\_citation\\_styles.html](http://www.plagiarism.org/plag_article_citation_styles.html)



## Comment 7

You have made clear the source for information given in this sentence. If the rest of the paragraph is also from the same source, it needs to be associated with the reference, too. References placed at the end of a paragraph imply that the entire paragraph (or back to a point in the paragraph where another citation is given) is supported by the same source. However, giving a source at the start of a paragraph does not imply all information in that entire paragraph is from the same source.



## Comment 8

What does this mean? Can you put the same information in more common terms that a larger readership would understand?



## Citation Needed

Cite Source:

Please use the link below to find links to information regarding specific citation styles:  
[http://www.plagiarism.org/plag\\_article\\_citation\\_styles.html](http://www.plagiarism.org/plag_article_citation_styles.html)



## Fragment

Fragment



## Comment 9

Right now it is not clear -- are these currently stored in GEO as unique data sets? Will the starting point of your work include needing to access each and then build into a master set for analysis? Or is there already a single set that includes all of these study outcomes rolled together? If they are separate, that helps explain some of the novelty in what you propose. However, if there is already one existing set of collected information, how are you proposing something new here?

## Fragment

Fragment



### Comment 10

I'm not sure that I understand this statement. If you have samples that are designated as "healthy," you cannot make an assessment of UF development (by definition, as the samples are not designated as UF). Are you saying you would like to identify factors that are different between healthy and UF, so that potentially disease could be predicted?



### Comment 11

Here you have given only a partial description of the data you intend to use (in terms of number of individuals represented). For each individual sample, what information is present? What platform was used to gather gene expression data (is it all microarray? RNA-seq)? Is there additional clinical data available? What kind of information will be there for you to start stratifying samples in your analyses?



### Comment 12

How is this different from the research that has been completed previously? Is the power in merging of different data sets? Will your analyses be comparable to previous work (so you can see if your results are similar trends) or completely unique?



### Comment 13

How will you make sure that you have both UF and non-UF represented in each?



### Comment 14

Why would these be helpful in terms of this project? What would be the utility of each type of analysis?



### Comment 15

As you are working on these different phases of your analyses, you will need to think about organization, and how it will be broken down in the context of your written report -- this will be important for your prospectus as well.

Also, have you started thinking about how you will share the process as well as results of your analyses? What kind of tables and figures will be the best way to communicate to your audience?



### Comment 16

As an example, here is a reference where there are only three authors. The first time you cite this source in the text, it should not be given as "Aissani, et al." in order to follow APA format.



## Comment 17

My understanding of what you have proposed is that you would like to perform a meta analysis as well. How will your meta analysis be distinct from (and not a reproduction of) this study?

PAGE 7

---



## Comment 18

This is not the correct format, in APA, to report author names. Please reference [https://owl.purdue.edu/owl/research\\_and\\_citation/apa\\_style/apa\\_formatting\\_and\\_style](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style) for help.

PAGE 8

---

PAGE 9

---

PAGE 10

---

PAGE 11

---

PAGE 12

---

PAGE 13

---

PAGE 14

---

PAGE 15

---

PAGE 16

---

PAGE 17

---

PAGE 18

---

PAGE 19

---

PAGE 20

---

PAGE 21

---

---

LOGICAL

3 / 5

## Logical Progression of Thought

---

5 (5)	Excellent progression of the content with logical synthesis of evidence
4 (4)	Content displays logical progression with appropriate synthesis of evidence
3 (3)	Content adequately structured with some synthesis of evidence
2 (2)	Content partially organized, evidence disjointed
1 (1)	Content not well organized, evidence lacking

---

INFORMATION

5 / 5

## Synthesis of Information from Scientific Literature

---

5 (5)	Exemplary use of the literature to support the proposal; literature used exemplifies current state of knowledge and is used critically
4 (4)	Literature used supports the proposal and is accurately portrayed, but contains some unnecessary information and/or is not a complete representation of the literature
3 (3)	Literature used appropriate to the project topic and adequate comprehension of material displayed, but conceptual detail is lacking; adequate but limited review of literature.
2 (2)	Literature used appropriate to the project topic but evidence of comprehension lacking; inadequate review of the literature
1 (1)	Literature used not always appropriate to the project topic; no evidence of comprehension; insufficient review of the literature

---

BIBLIOGRAPHY

4 / 5

## Annotated Bibliography

---

5 (5)	Exemplary report of at least 10 references (mostly primary sources, supplemented by additional resources) with full citation (APA format) and summary of how each will support the project
4 (4)	Report of at least 10 references (even number of primary and secondary sources) with full citation (APA format) and summary
3 (3)	Report of 10 sources, but not in correct format or incomplete summary

2 Less than 10 sources given, but correct in format, summary  
(2)

1 Insufficient number of sources, inadequate information provided  
(1)

## GRAMMAR

5 / 5

Spelling, Grammar, Mechanics

---

5 Less than 3 spelling, grammatical or mechanical errors  
(5)

4 No more than 5 spelling, grammatical and/or mechanical errors  
(4)

3 Fewer than 8 spelling, grammatical and/or mechanical errors  
(3)

2 Less than 10 spelling, grammatical and/or mechanical errors  
(2)

1 More than 10 spelling, grammatical and/or mechanical errors  
(1)