Week 8 Prospectus Outline

by Janis Corona

Submission date: 09-Mar-2019 01:42PM (UTC-0600)
Submission ID: 1090514730
File name: ResearchProposalProspectusFinalVersion.docx (27.22K)
Word count: 2107
Character count: 12166

Prospectus for Meta-Analysis of the Ubiquitous Genotypes Associated with Human Uterine Leiomyoma Development in Healthy Tissue using Bioconductor

Janis Corona

Lewis University

Prospectus for Gene Expression Analysis of the Ubiquitous Genotypes Associated with Uterine Leiomyoma Development in Healthy Tissue using Bioconductor

I. **Introduction**. This section describes the Ubiquitous Genotypes Currently Associated with Uterine Fibroids and explains what they are and the impact they have on lives.

A. **Description of uterine leiomyomas (UL)**. This sub-section describes who UL affect, the known risk factors for UL, and what known SNP or genotypes are currently found to be significantly associated with UL.

a. **UL described in populations**. This sub-sub section describes the population demographic risks, symptoms of UL, asymptomatic patients' impact on reproducible research, and the most popular treatment for UL being hysterectomy (Hellwege et al., 2017; Eggert, et al., 2012; Rafnar et al., 2018; Bondagji et al., 2018; Liu et al., 2018; Hodge et al., 2012).

b. **Significant genotypes for UL**. This sub-sub-section lists the SNPs found to be significantly associated with UL development in women of child bearing age. These genotypes associated with UL are: BET1L, TNRC6B, HGMA2, FASN, CCDC57, and CYTH4 (Hellwege et al., 2017; Eggert et al., 2012; Rafnar et al., 2018; Bondagji et al., 2018; Liu et al., 2018; Hodge et al., 2012).

B. **Chromosomal location of UL associated SNPs**. This sub-section lists and describes each chromosome associated with regions of SNPs that house the most common SNPs significantly associated with UL.

a. **Chromosome 11**. This sub-sub-section describes the BET1L genotype SNP that is described as having various associations with UL such as which uterine layer a UL is originating from or how many UL are in one uterus making the UL patient have multiple UL (Liu et al., 2018; T. Edwards, Hartmann, & D.

Edwards, 2013; Rafnar et al., 2018). This genotype, BET1L, was tested for significance in association with UL in studies on other race demographics and found not significant in certain races (Bondagji et al., 2017; Aissani, Wang, & Wiener, 2015; Rafnar et al., 2018).

b. **Chromosome 12**. This sub-sub-section describes the location of HGMA2 and how it is associated with high expression in UL samples (Hodge at al., 2012). Another study stated HGMA2 to be a factor in tumorigenesis from studies done in 1988 researching HGMA2 and tumor formation (Aissani et al., 2015).

c. **Chromosome 17**. This sub-sub-section describes the two genes with SNPs on chromosome 17 named CCDC57 and FASN that are associated with UL in Europeans (Eggert et al., 2012; Aissani et al., 2015). Another study tested these two gene SNPs and found no significance in UL for Saudi Arabian populations (Bondagji et al., 2017).

d. **Chromosome 22**. This sub-sub-section describes the two genes that are found on Chromosome 22 to be significant in UL in specific races. For the first gene SNP called TNRC6B, it is found to be significant in Chinese, Japanese, Europeans, European Americans, and Saudi Arabians (Rafnar et al., 2018; Liu et al., 2018; Edwards et al., 2013; Aissani et al., 2015; Bondagji et al., 2017). TNRC6B was not found to be significant in African Americans (Hellwege et al., 2017). CYTH4, one of the other genes on Chromosome 22 found to be significant in UL for African Americans will be described in this sub-sub-section as well as TNRC6B (Hellwege et al., 2017).

II. **Materials and Methods**. This section will describe the material sources used and the type of software used to analyze the data. The ubiquitous SNPs found most differentially expressed

between non-UL and UL samples of gene expression data will be used to build a model (D. Zhang, Sun, Ma, Dai, & W. Zhang, 2012; Eggert et al., 2012). Link analysis will be generated linking influential genes used in their SNP regions of the chromosomes and the separate population studies found to be significant in UL association (Hellwege et al., 2017; Eggert et al., 2012; Rafnar et al., 2018; Bondagji et al., 2018; Liu et al., 2018; Hodge et al., 2012).

A. **Gene Expression Omnibus (GEO) data of UL and healthy uterine samples.** This sub-section explains the use of GEO to access eight studies that list their gene expression data on GEO for sharing (Crabtree et al, 2009; Delaney et al., 2017; Hoffman et al., 2004; Miyata et al., 2015; Quade, Mutter, & Morton, 2004; Teixeira, 2018; Vanharanta et al., 2006; Zavadil et al, 2010). This sub-section will analyze those gene expression samples using the 'in situ oligonucleotide' or microarray platforms derived from five separate GEO studies (Hoffman et al., 2004; Vanharanta et al., 2006; Zavadil et al., 2010; Crabtree et al., 2009; Quade et al., 2004). The table of all the microarray or 'in situ oligonucleotide' technology type samples will aggregate to 51 non-UL and 50 UL observations for analyzing. Then, two separate GEO studies that used the 'oligonucleotide beads' or bead chip technology for RNA extraction will be analyzed (Delaney et al., 2017; Teixiera, 2018). The bead chip studies will aggregate to 27 non-UL and 41 UL samples for comparing similarities to the microarray data when analyzing differential gene expression. Bootstrapped simulations will be made to independently and identically sample with replacement to produce a much larger data set of each technology type data as either microarray or bead chip. From each of those much larger data sets the focus will be on differential gene expression between non-UL and UL samples to find the relation of the ubiquitous genotypes mapped next to the other runner up genotypes not further explored in the population studies.

B. **R Statistical software for statistical analysis**. This sub-section describes the packages used in R, what R is, and the types of algorithms used such as linear, logistic, binomial, etc. to explore the data (R, 2019). This is expected to generate quick calculations for the top genes expressed in samples so that further tests can be performed on those specific genes.

C. **Bioconductor for biostatistics in R**. This sub-section describes the packages used in Bioconductor and what functionality each package from Bioconductor contributed to the biostatistical analysis of this research, such as to build heatmaps, M-A plots, and so on for analyzing the GEO data (Bioconductor, 2019).

III. **Results**. This section will display the results and provide a detailed summary of each result and what it means.

A. **Analyzing the biostatistical data from R**. This sub-section will explain the meaning of the scatter points or box plots developed using the R statistical packages. Any plots, tables, or other visualizations for analyzing the data before making predictions on the population from the GEO data set of samples will be displayed in this section.

B. **Machine Learning with Bioconductor**. This sub-section will display the visualizations of the tables and/or graphical plots created in Bioconductor and explain the meaning of each visualization as it is displayed.

IV. **Discussion**. This section will synthesize all the information from the results gathered and bring it back to answering the research or show results interpreted as a solution to the question of which ubiquitous genotypes are associated with UL in healthy females. The conclusion will be within the end of this section and will interpret the results. The focus of this research was to contribute to the study of UL development in healthy tissue using GEO data across all races, as other studies on UL development used race specific studies of UL

development to find genotypes that were ubiquitous across all the race specific studies. The distinct difference between this study and previous studies is using non-race specific gene expression samples instead of solely samples from one specific race to find or not find those same ubiquitous genotypes to be associated with UL development. Leaving an open question for further analysis of the next set of five top genotypes below these ubiquitous genotypes in UL development that may also be associated with UL development the previous population studies excluded. Right now, the idea is that UL development is currently unique to each race and personalized treatment for UL in each race is needed (Hodge et al., 2012; Hellwege et al., 2017; Liu et al., 2018; Edwards et al., 2013).

V. **Supplementary Material**. This section will provide the database of samples used from GEO as one table of UL and non-UL samples identified as either microarray or bead chip technology. In this section, there will also be a link to view the code created in Bioconductor and R to produce the results in the Materials and Methods section and the Results section.

VI. **References**. This section will display all of the references used in this research that include data resources, research articles, and software used. If any other types of data are used, such as supplemental articles or review articles published in peer review journals, then they will also be listed in this section.

References

Aissani, B., Zhang, K., & Wiener, H. (2015). Evaluation of GWAS candidate susceptibility loci for uterine leiomyoma in the multi-ethnic NIEHS uterine fibroid study. *Frontiers in Genetics*, 6, 241. DOI:10.3389/fgene.2015.00241

Bioconductor, version 3.8, (2019). Bioconductor: Open Source Software for Bioinformatics. Retrieved March 3, 2019 from: https://www.bioconductor.org/install/

Bondagji, N.S., Morad, A.F., Al-Nefaei, A.A., IKhan, I. A., Elango, R., Abdullah, L.S., … Shaik, N.A. (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

Crabtree, J.S., Jelinsky, S.A., Harris, H.A., Choe, S.E., Cotreau, M.M., Kimberland, M.L., … Walker, C.L. (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

Delaney, M.A., Wan, Y.W., Kim, G.E., Creighton, C.J., Taylor, M.G., Masand, R., … Anderson, M.L. (2017). A Role for Progesterone-Regulated sFRP4 Expression in Uterine Leiomyomas. *Journal of Clinical Endocrinology and Metabolism,* 102(9), 3316-3326. PMID:28637297

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

Eggert, S., Huyck, K., Somasundaram, P., Kavalla, R., Stewart, E., Lu, A., … Morton, C. (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

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

Hodge, J.C., Kim, T., Dreyfuss, J.M., Somasundaram, P., Christacos, N.C., Rouselle, M., … Morton, C.C. (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

Hoffman, P.J., Milliken, D.B., Gregg, L.C., Davis, R.R., & Gregg, J.P. (2004). Molecular characterization of uterine fibroids and its implication for underlying mechanisms of pathogenesis. *Fertility and Sterility*,82(3), 639-49. PMID:15374708

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

Miyata, T., Sonoda, K., Tomikawa, J., Tayama, C., Okamura, K., Maehara, K., … Nakabayashi, K. (2015). Genomic, Epigenomic, and Transcriptomic Profiling towards Identifying Omics Features and Specific Biomarkers That Distinguish Uterine Leiomyosarcoma and Leiomyoma at Molecular Levels. *Sarcoma 2015*. PMID: 27057136

Quade, B.J., Mutter, G.L., & Morton, C.C. (2004). Comparison of Gene Expression in Uterine Smooth Muscle Tumors. Gene Expression Omnibus. GEO Accession ID: GSE764. Retrieved March 2019 from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi

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/

Rafnar, T., Gunnarsson, B., Stefansson, O.A., Sulem, P., Ingason, A., Frigge, M.L., … Stefansson, K. (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

Teixeira, J.M. (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: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi

Vanharanta, S., Pollard, P.J., Lehtonen, H.J., Laiho, P., Sjoberg, J., Leminen, A., … Aaltonen, L.A. (2006). Distinct expression profile in fumarate-hydratase-deficient uterine fibroids. *Human Molecular Genetics*, 15(1), 97-103. PMID:16319128

Zhang, D., Sun, C., Ma, C., Dai, H., & Zhang, W. (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

Zavadil, J., Ye, H., Liu, Z., Wu, J., Lee, P., Hernando, E., … Wei, J.J. (2010). Profiling and functional analyses of microRNAs and their target gene products in human uterine leiomyomas. PLoS One, 5(8). PMID: 20808773

Week 8 Prospectus Outline

ORIGINALITY REPORT

31%	28%	26%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	helda.helsinki.fi Internet Source	3%
2	Submitted to Rutgers University, New Brunswick Student Paper	2%
3	Submitted to University of Westminster Student Paper	2%
4	ethesis.helsinki.fi Internet Source	2%
5	mdanderson.influuent.utsystem.edu Internet Source	2%
6	Submitted to Australian Catholic University Student Paper	2%
7	www.ucl.ac.uk Internet Source	2%
8	Submitted to AUT University Student Paper	2%
9	link.springer.com	

	Internet Source	2%
10	Jiri Zavadil, Huihui Ye, Zhaojian Liu, JingJing Wu et al. "Profiling and Functional Analyses of MicroRNAs and Their Target Gene Products in Human Uterine Leiomyomas", PLoS ONE, 2010 Publication	1%
11	journal.frontiersin.org Internet Source	1%
12	Submitted to University of Sheffield Student Paper	1%
13	www.biochem.ualberta.ca Internet Source	1%
14	erc.endocrinology-journals.org Internet Source	1%
15	msanderlab.org Internet Source	1%
16	obgyn.onlinelibrary.wiley.com Internet Source	1%
17	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	
18	health.usnews.com Internet Source	1%
19	Zhang, Kui, Howard Wiener, and Brahim Aissani. "Admixture mapping of genetic variants for uterine fibroids", Journal of Human Genetics, 2015. Publication	1%
20	"2019 | President's Plenary Session", Reproductive Sciences, 2019 Publication	1%
21	www.canberra.edu.au Internet Source	<1%
22	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 Publication	<1%
23	embomolmed.embopress.org Internet Source	<1%
24	Hodge, J. C., T.-M. Kim, J. M. Dreyfuss, P. Somasundaram, N. C. Christacos, M. Rousselle, B. J. Quade, P. J. Park, E. A. Stewart, and C. C. Morton. "Expression profiling of uterine leiomyomata cytogenetic subgroups	<1%

reveals distinct signatures in matched myometrium: transcriptional profilingof the t(12;14) and evidence in support of predisposing genetic heterogeneity", Human Molecular Genetics, 2012.
Publication

---

25 Karmon, A. E., E. R. Cardozo, B. R. Rueda, and A. K. Styer. "MicroRNAs in the development and pathobiology of uterine leiomyomata: does evidence support future strategies for clinical intervention?", Human Reproduction Update, 2014. <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

Week 8 Prospectus Outline

GRADEMARK REPORT

FINAL GRADE

14/15

GENERAL COMMENTS

Instructor

The biggest deficiency right now is that you did not account for an overall introduction section, and that means that your research question does not have a place to be defined at the start -- it will be important to lead with that as your work on your project moving forward.

As long as you have been able to successfully access and are aware of what kind of data are available through GEO, I think it is a strong plan to move forward working with these data and not worry about BioVU.

PAGE 1

PAGE 2

Comment 1

This section seems to be a background (literature review) focused on ULs. It is important information to provide, and I think it will be useful for the reader to know the genomic locations (chromosomal placements) of genes associated with UL -- I like that organization.

However, before this "Background" chapter, you should have your first chapter "Introduction" which will give an overview of your entire project. It should introduce your topic (UL, noting that there are some known genetic associations with disease outset), note the dataset you want to work with, make clear your research objective/thesis statement, and then overview how you will assess. Thus, the Introduction will give an overview of your entire project so that it is very clear, upfront, what you are attempting to accomplish. Otherwise, right now you do not have a place to really make clear to your audience what you are wanting to accomplish in your paper.

PAGE 3

PAGE 4

Comment 2

You might want to consider splitting this into two sections. In one, give a full description of the data, comparing and contrasting the two different types of data available (from the different platforms) -- this will be descriptive of what you are working with. In the second, make it clear

the method by which you are joining independent records (different studies) into the larger working data sets -- this is a more procedural item. Both are important explanations prior to then explaining how you will then analyze the data.

PAGE 5

Comment 3

It will be important to explain -- just make sure that you focus on how these components are used in the context of your research specifically, rather than getting caught up in a theoretical explanation of the software itself. The same recommendation should be applied to your section on Bioconductor.

Comment 4

This section is the hardest to forecast (since you do not yet have data). The general layout as given here is sufficient for now, but you may want to revise in the future, depending on what your actual results show.

PAGE 6

Comment 5

It will be important to make clear your objective in the first chapter (before you have presented all of your work) so that ultimately your discussion section can evaluate whether the objective was achieved, and to then assess the quality of the outcome, discuss limitations, consider future directions, etc.

Comment 6

Assuming that your database is large enough, you'll likely want to share it as a link to data, also. Trying to format into a table that can easily be read in the context of the document will most likely be a challenge.

PAGE 7

PAGE 8

PAGE 9

CONTENT 5 / 5

Content

5 (5)	Exemplary demonstration of understanding the topic
4 (4)	Very good demonstration of understanding the topic
3 (3)	Adequate demonstration of understanding the topic
2 (2)	Limited demonstration of understanding the topic
1 (1)	Inadequate demonstration of understanding the topic

TOPIC 4 / 5

Topic Development

5 (5)	Main purpose clearly evident throughout, with all aspects clearly connected, and demonstration of extensions that will be made from the existing literature
4 (4)	Above average expression of the main purpose, ideas well defined and connected with above average demonstration of knowledge, extensions that will be made.
3 (3)	dequate expression of main purpose, connection of ideas, and allusion to extensions that will be made
2 (2)	Expression of the main purpose and included concepts inconsistent and somewhat disconnected; extensions that will be made stated, but not clearly tied to publication record
1 (1)	Inadequate expression of main purpose; ideas unclear and disjointed; not clear how new contribution will be made

FORMAT 5 / 5

Format

5 (5)	Detailed outline provided, with appropriate distribution of information into sections and sub-sections with an explanation of information to be included in the section; clear indication of where references will be incorporated
4 (4)	Outline provided with appropriate distribution of information; summary of information and references to be included provided
3 (3)	Adequate outline of project, but details are sparse, summaries brief, and not all references incorporated

2 (2)	Outline lacks detailed breakdown of concepts; summaries lacking or missing; limited incorporation of references
1 (1)	Inadequate outline of project; summaries of content not provided; references not incorporated