1. Fixed the run-on sentence in comment 1 of feedback on page 1
2. Explained the estrogen responsiveness of UL on as being an antagonist to UL growth, that UL can be inhibited by taking estrogen blockers but not cured once treatment stops, page 1
3. Comments 3:21, clarify if SNP or gene expression being examined in proposal, not interchangeable. Ultimately, I am comparing gene expression data to determine if those ubiquitous genes show high in differential expression in the microarray samples and not the actual SNPs of those genes.
4. Explained how the R package, caret, is going to be used to partition the data into randomly placed samples of training or testing sets for analyzing the data and building a model to predict UL risk
5. Explained the models being potentially tested on the training and testing data in R for predicting UL risk on the testing data
6. Changed the focus in the proposal and prospectus to be on the genes and not the SNP genotypes of those genes
7. Clarified the visualizations desired for this project using the Gviz package of Bioconductor in the Materials and Methods section of the Prospectus
8. What is known about differences in gene expression of UL patients? CYTH4 if expressed low in the thyroid is an indicator of UL risk in AAs. BET1L, FASN, HMGA2, CCDC57 and TNRC6B expressed more or less in UL patients? In the studies the MAF of the alleles of SNPs to the ubiquitous genes was used to test for significance of the SNPs not genes associated with UL risk.
9. What are the meta fields or identifiers in the data, what information on patients is known? The data uses two platforms now that one data series was excluded (GSE764). There are four data series that use GPL96/GPL570 similarly the same use of Affymetrix Human Genome U133 Array, and the other series uses GPL6480. The only data platform that uses sequence data, cytoband, and chromosome location is the GPL6480 platform, the other identifiers are the alternate IDs in UCSC ENSEMBL, Gen Bank, NCBI, orthologue, gene function, gene name, species sample derived, annotation date, and other IDs.
10. This data is exclusively human samples, the series that included rat samples labeled those samples as Rat\_Myometrium\_Donor or Rat\_Leiomyoma\_Donor in GEO and were excluded by sample listing in the series.
11. Link directly to NCBI GEO in annotated bibliography comments 32:39, <https://www.ncbi.nlm.nih.gov/geo/info/linking.html> using: [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE**xxx**](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSExxx) **for GSE series and** [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL**xxx**](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPLxxx) **for Platform**
12. Eliminate PMID or use DOI instead on Miyata and Vanharanta resources
13. Included GSE764 (Quade, Mutter, & Morton, 2004) only where it was a limitation. Other limitations were in not having attached sequence IDs to the gene IDs in the four other data sets other than the GSE68295.
14. Removed the GEO resource as I needed to include the GSE resources specifically included in the master data set
15. Keep the references SNP/genotypes to genes because the focus of this research is on genes having those SNPs associated with UL risk.