Janis Corona

BIOL 59000: Data Science Project for Life Sciences

Week 5 Presentation Outline

July 7, 2019

Presentation Outline

I. Title Slide: Meta-Analysis of the Genes Ubiquitously Associated with Human Uterine Leiomyoma Development in Healthy Humans Using the Gene Expression Omnibus Data

A. Background: Description of Uterine Leiomyoma (UL)

1. UL Described in Populations: Discuss indicators for increased risk of UL in otherwise healthy human females

1. Map of the Genes ubiquitous to UL risk in populations across the world

Significant Genes for UL: Discuss the significant genes found to have genotypes associated with UL in various populations

B. Explain Gene Regulation

1. Show image of Gene transcription in the cell (not the image of the cell). Explain how gene expression is regulated

2. Show image of Gene transcription and translation in image of the picture of a cell

3. Overview of Gene regulation concepts. One slide to show the term, notes will explain each

a. Gene transcription: Chromosome, chromatin, histone, nucleosomes, introns and exons post-transcription, moving to cytoplasm

b. Gene translation: ribosomes translate RNA to proteins, cis-acting elements, promotors of genes, repressors of genes, 5’ to 3’ translation region, recoil by RNA binding proteins upstream, downstream, along coding region, millions of base pairs away, external stimuli effects gene expression.

c. Factors that affect gene expression such as stress due to chemical or environmental factors can inhibit the transcription at the epigenomic level and can alter translation of RNA outside of the nucleus.

C. Linkage Disequilibrium (LD) and give some images of the LD analysis some reviews have shown (Rafnar, Eggeret, Cha, Hellwege).

1. Eggert

2. Cha

3. Rafnar

4. Hellwege

D. Chromosomal Location of UL genes: Use the Gviz images of the chromosomes to show the location of the genes that are in the neighborhood of the genes found to have significance in UL risk within populations.

1. Chromosome 11

2. Chromosome 12

3. Chromosome 17

4. Chromosome 22

II. Methods

1. Discuss the Geo Data of UL and non-UL samples
2. Discuss the R software chosen to combine the data sets, the findings, limitations, and the reasons for choosing R
3. Discuss how Bioconductor for R added to this project

III. Display Data Results

1. Display the Tables on the top 10 genes expressed most and the full names of their abbreviations as well as the additional six genes ubiquitous to current UL risk studies
2. Display the figures of the top expressed genes on the chromosome, some additional plots outside of the report that showed scatter plots of the genes in Lattice and ggplot2
3. Discuss the predictive outcomes from the algorithms chosen, why those algorithms were chosen

IV. Conclusions

1. Discuss Results
2. Discuss Limitations
3. Discuss further work that could be done to establish genetic relationships between UL and non-UL gene expression samples in the Gene Expression Omnibus (GEO) microarray samples

V – References: Cite all references used in the power point presentation