Jose L. Moreno
BENG 185
Luis Arturo Soto
6/16/2017

Final Project:

****Note I changed my project topic since I first proposed it

Background:

The human immunodeficiency virus (HIV) is a lentivirus that causes HIV infection and over time acquired immunodeficiency syndrome (AIDS). HIV itself can be categorized into different subtypes with two prevalent strains; HIV-1 and HIV-2. HIV-1 and HIV-2 have many similarities including gene expression, modes of transmission, and consequences; AIDS. However, it's been shown that HIV-2 has a lower rate of transmission as well as lower likely-hood of progression to AIDs. The virus attacks a specific type of immune system cell in the body, known as CD4 helper T cells. Attacking these cells makes it harder for your immune system to fight back. So, studying these types of cells is essential to understand how HIV works.

My hope for this project is to try to understand whether gene expression along with gene functionality play a role in driving factors of lower transmissibility and progression when comparing the two different types of HIV strands (HIV-1 and HIV-2) in T cells.

Definition:

Question: Which genes are co-expressed on the HIV-1 strand but not in the HIV-2 strand and functionally, how are these co-expressions related to the pathology pathway of the virus and relate back to having lower transmissibility and progression in HIV-2.

Hypothesis: Considering the fact that these pathogenic viruses are the same but have diverged overtime into different strands of the same virus, some genes will be expressed in one strand that play a functional role in infection that isn't seen in the other strand. These functional differences in gene expression will play a role in transcription regulation, cell cycle, and infection to better understand how the strands transmit and progress.

Assumptions:

- MicroArray Analysis is accurate
- Biological procedure was conducted correctly
- The **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) gives reliable functional annotations for genes

Strategy:

Using microarray data from Devadas, Krishnakumar paper titled, "Analysis of Host Gene Expression Profile in HIV-1 and HIV-2 Infected T-Cells.", I am going to extract gene expressional data for genes up/down regulated in HIV-1 and HIV-2. With the gene expression data in hand, I would need to map back the microarray ID's back into gene IDs (A_23_P31135 -> ACAT2) using DAVID; since that is how the data is given. Once I confirmed that this method worked, I used the expression excel sheet provided in the paper supplement. With the data neatly organized in an excel format, I extracted the most important columns to create an expressional data table and a functional data table. These tables were loaded into MySQL database labeled: HIV_1_expression, HIV_2_expression, HIV_1_functions, HIV_2_functions. Once the tables were loaded, I was able to write 2 scrips named **express_analysis.py** and **functional_analysis.py.** The first script was used

to extract raw gene expressional data for both the HIV-1 infected T cell and the control; same thing was done for HIV-2. Since I was going through the expressional data table for both strands, I was also able to make a list of unique up and down regulated genes in both HIV-1 and HIV-2 as well as give expressional data for each unique gene. The second script (functional_analysis.py), was used in order to extract Ontology of each unique gene in the lists I created. Ontology lets me put genes into 3 categories:

molecular function

molecular activities of gene products

cellular component

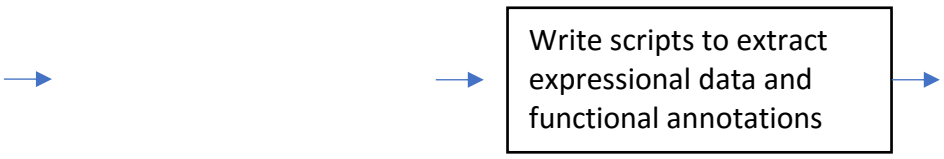
where gene products are active

biological process

pathways and larger processes made up of the activities of multiple gene products.

This is a standard way of organizing gene functions and the Ontologies for each gene was confirmed using Gene Ontology Consortium web interface. Next, I wanted to get a better understanding of what these genes did in a less broad format. Using DAVID, I was able to input up/down regulated unique genes from HIV-1 and HIV-2 and was able to get a better understanding functionally what every unique gene from each strain role was. I was also able to analysis genes that are shared by both strains and explore what role they played in the underlining mechanism involved in different cellular processes.

FlowChart:



Obtain expressional and functional data from paper

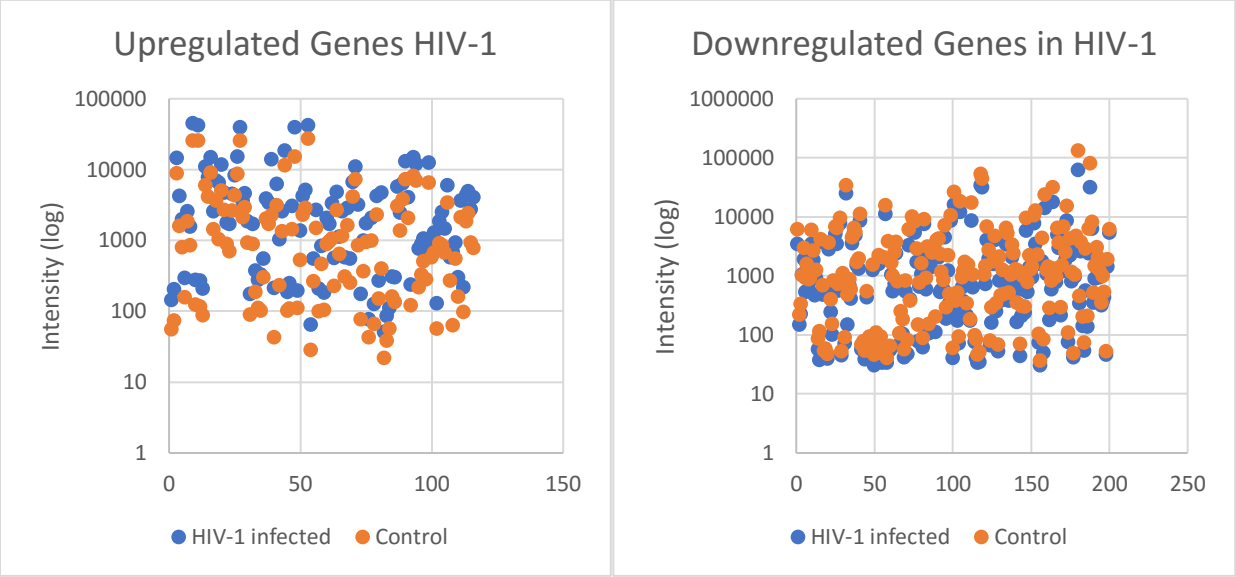
Load Tables with both expression and functional data

Plot and Analysis

Use DAVID and Gene Ontology Consortium to confirm functions and do further analysis

Results:

Figure #1: Up/Down Regulation of unique genes in either of the two HIV strains



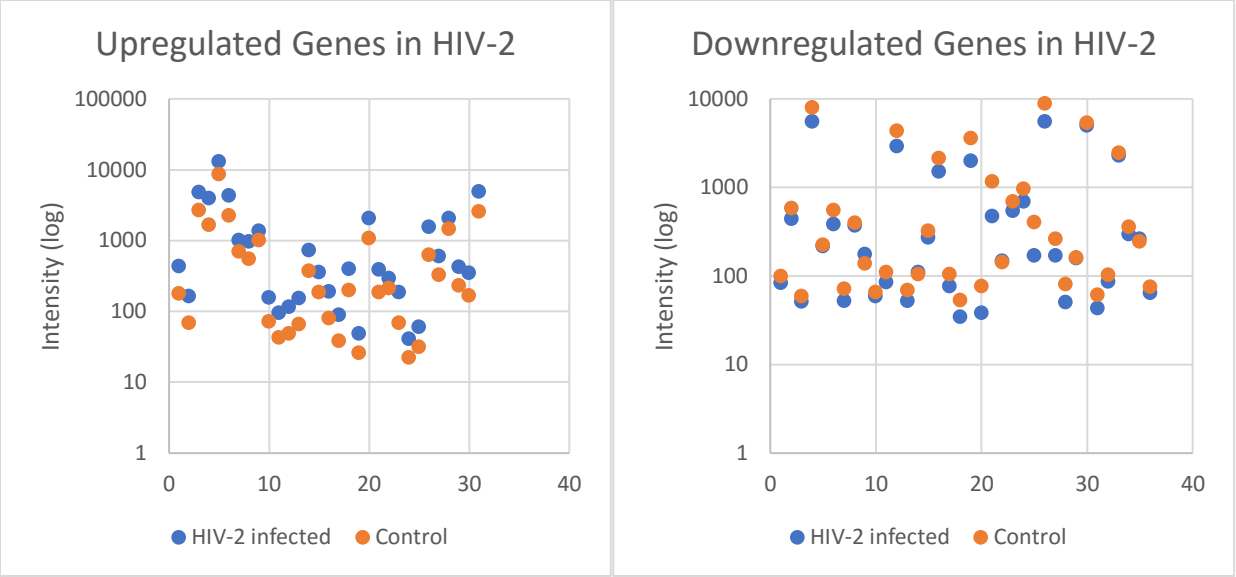
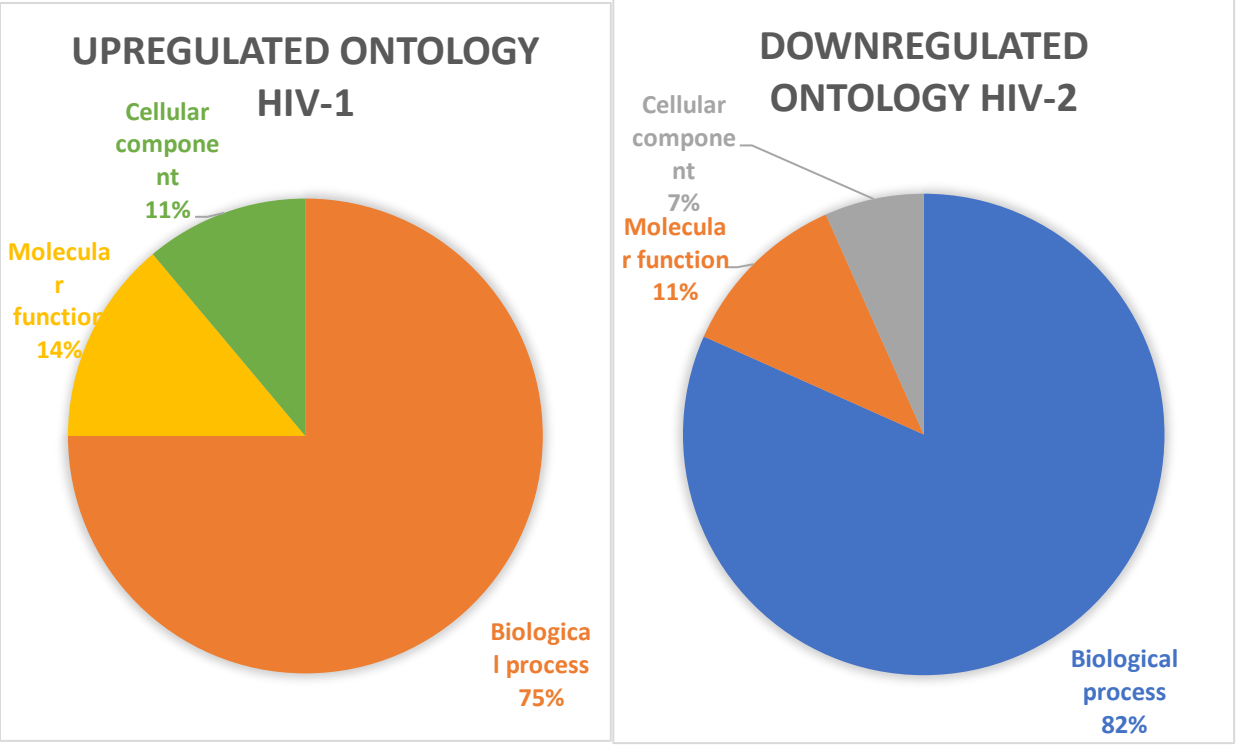


Figure #1: Gene expression data confirming up/down regulation of genes unique to each strain of HIV. Orange being the control and right above/below each orange dot is its up/down HIV infected sample expressional level (blue dots).

Figure #2: Ontological Analysis of Genes in HIV-1 and HIV-2



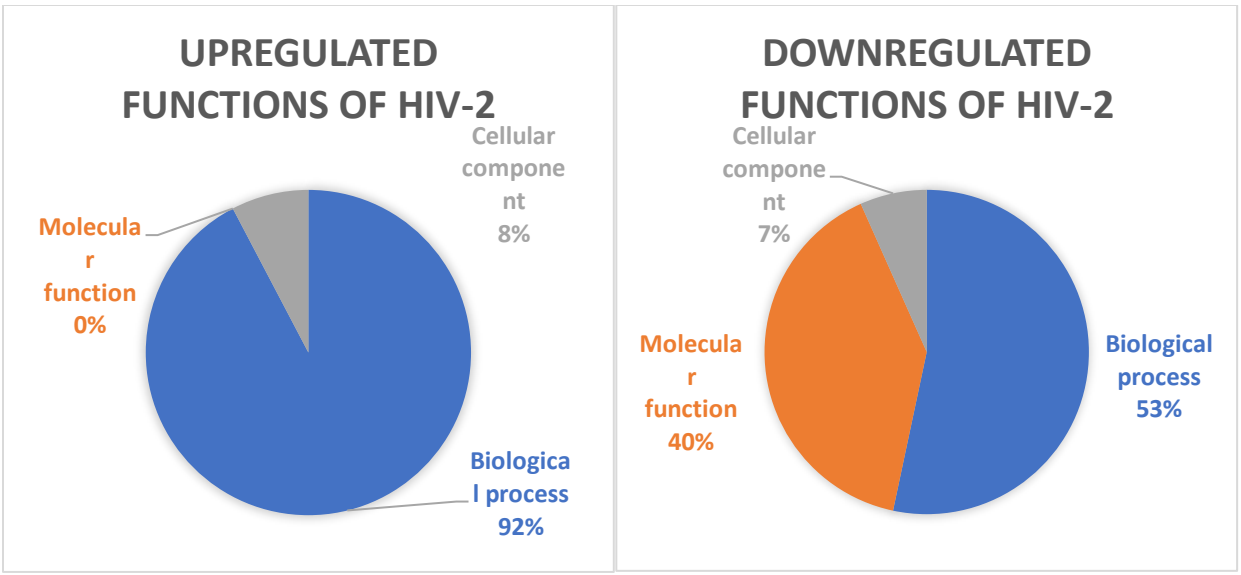
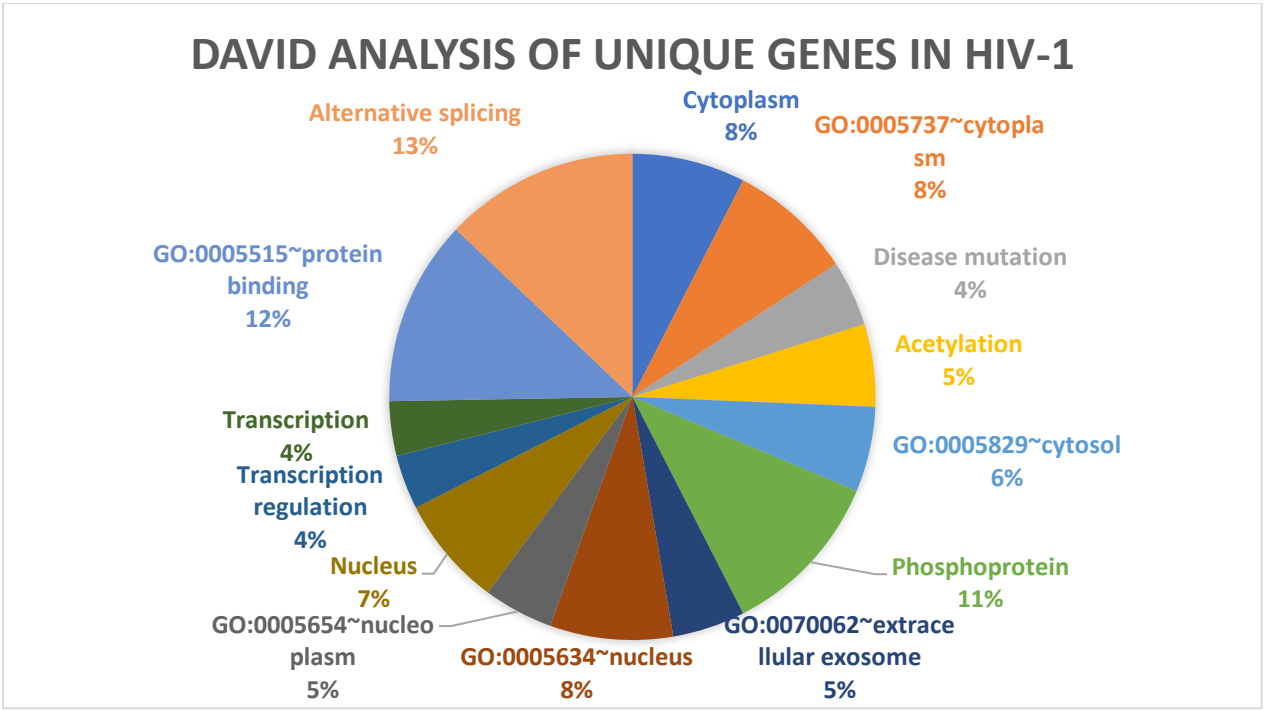


Figure #2: Classification of each gene with its Ontology in both strains of HIV.

Figure #3: DAVID Analysis



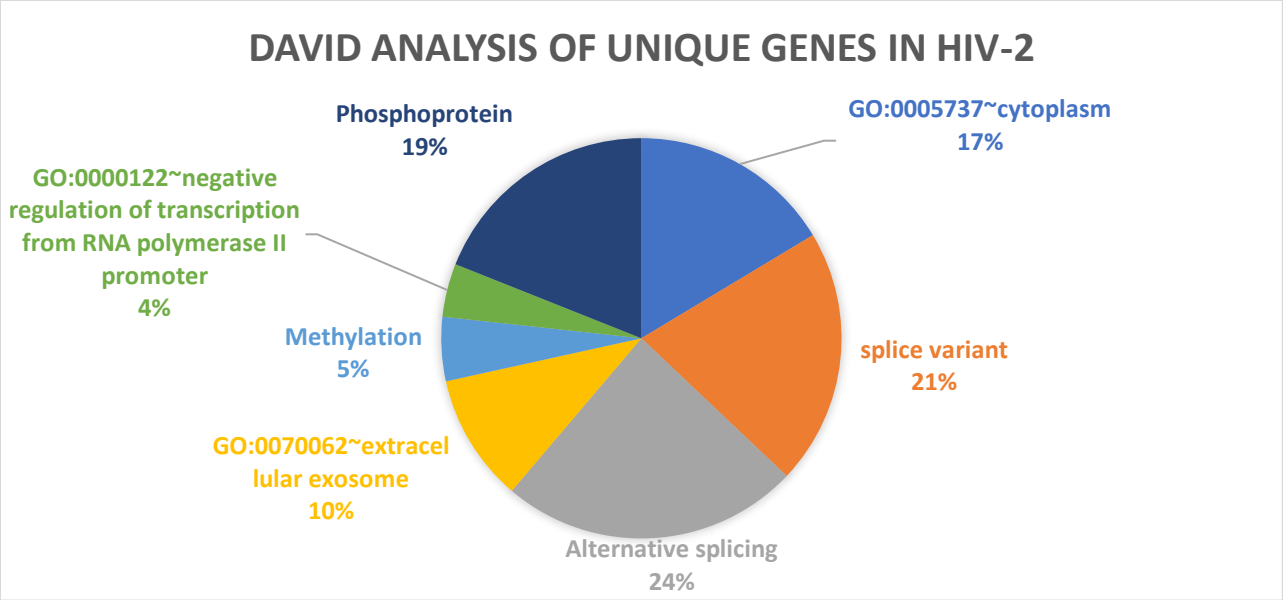


Figure #3: More specific annotation of the gene functions unique to each HIV strain

Table #1:

Shared Genes Between HIV-1 and HIV-2
FLJ41481
RRAS2 (role in activating signal transduction pathways that control cell proliferation)
ANXA8L2 (Overexpression of this gene has been associated with acute myelocytic leukemia)
ARHGDIG (role in modulating the activation of GTPases by inhibiting the exchange of GDP for GTP.)
PRO0628

Discussion:

From Figure #1 we were able to confidentially state that the genes unique to each strain of HIV were either up/down regulated based of gene expression. Figure #2, we were able to get a perspective on what each gene does functionally by categorizing them into the 3 groups; biological process, cell component, and molecular function. I went more into depth by

standardizing each gene into a more meaningful function using DAVID (Figure #3). And finally, identifying the genes shared by both strains of HIV.

My DAVID analysis did support my hypothesis as we can see that genes unique to HIV-1 were more prone to transcriptional regulation, disease mutation, and acetylation. All 3 of these categories play a role, I believe, in the HIV-1 progression being much faster than HIV-2. When we looked at HIV-2 we see that genes unique to HIV-2 are more prone to methylation, alternative splicing, and regulation of RNA polymerase. HIV-2 can be characterized by a longer asymptomatic stage due to the genes discussed. These groups all play a role in the decreased progression of HIV-2 into AIDs as many of these genes repress transcriptional expression. Genes shared by the two strains weren't surprising as RRAS2 effects cell proliferation of cells and ANXA8L2 being associated with leukemia.

Transmissibility is something I can't answer with my analysis alone. To address that specific topic, I believe we would need to do in vivo studies to see which strain is more transmissible.

*** All code is on my github as well as the excel sheet I used for analysis

github: https://github.com/j2moreno/BENG185

References:

Devadas, Krishnakumar, Santanu Biswas, Mohan Haleyurgirisetty, Owen Wood, Viswanath Ragupathy, Sherwin Lee, and Indira Hewlett. "Analysis of Host Gene Expression Profile in HIV-1 and HIV-2 Infected T-Cells." Plos One 11.1 (2016): n. pag. Web.

Nyamweya, Samuel, Andrea Hegedus, Assan Jaye, Sarah Rowland-Jones, Katie L. Flanagan, and Derek C. Macallan. "Comparing HIV-1 and HIV-2 Infection: Lessons for Viral Immunopathogenesis." Reviews in Medical Virology 23.4 (2013): 221-40. Web.