***Comparing Alignment-based Methods to Alignment-free Methods of RNAseq Sample Quantification.*** *Syed Monis Ali [1\*], Chandler Russell [1\*], Jung-Han Kim [2], James Denvir [2†], Nicole Garrison [3†]. [1] Department of Arts and Science, Bluefield State College; [2] Department of Biomedical Sciences, Marshall University; [3] Department of Biomedical Sciences, West Liberty University.*

RNA sequencing (RNAseq) is a next generation sequencing/high throughput method which measures gene expression levels using counts of sequenced mRNA fragments. Quantification of reads that map to specific transcripts is an essential step in differential expression analysis. There are two major computational approaches to accomplish this process – alignment-based and alignment-free methods. In this study, we compared an alignment free method (salmon) and an alignment-based method (HISAT2) to determine if there are differences in the resulting count tables and our ability to identify differentially expressed genes between these two methods. Both programs performed similarly for genes expressed at higher levels within the samples. However, the alignment free program tended to have difficulty quantifying genes expressed at a low level, possibly resulting in loss of rare transcripts in the resulting pathway analyses. *(Supported by NIH Grant P20GM103434 to the West Virginia IDeA Network for Biomedical Research Excellence and NIDDK/NIH 1R15DK113604 (JH Kim, PI))*

**Testing Minimum Assembly Read Depth in SARS-CoV-2 Sequencing Analysis.** *Christopher Carver [1\*], Sharmel Cavender [2\*], Peter Stoilov [3], Pete Perrotta [4], Ryan Percifield [5], James Denvir [6†], Nicole Garrison [7†]. [1] Department of Science and Math, Glenville State College; [2] Department of Biology, University of Charleston; [3] Department of Biochemistry, West Virginia University; [4] Department of Pathology, West Virginia University; [5] Department of Biology, West Virginia University; [6] Department of Biomedical Sciences, Marshall University; [7] Department of Biomedical Sciences, West Liberty University.*

Adequate sequencing read depth allows accurate mapping and identification of new mutations and new variants in emerging pathogens. Currently, tracking and recognizing variants or new mutations of the SARS-CoV-2 virus is a public health priority. However, it is unclear which read depth level is necessary to accurately determine variants. Variants are typically identified by using the highest read depth available to get an acceptable mutation calling count. We used SARS-CoV-2 samples collected within the borders of West Virginia to investigate the effect of different initial read depths on the variant calling process. We were able to generate a series of input files representing subsets of our data (5000-10000 reads) using a custom python script and applied a standard pipeline of programs to align our data and call variants. Our research allowed us to determine the minimum allowed read depth required to maintain the mutation count. Mutation calls frequency increased with higher numbers of input reads. Understanding the effect of smaller starting read numbers on variant calling will allow other investigators tracking SARS-CoV-2 to potentially scale up sequencing efforts or speed up sequence analysis pipelines. *(Supported by NIH Grant P20GM103434 to the West Virginia IDeA Network for Biomedical Research Excellence and WV-CTSI NIGMS.NIH U54GM10)*

**The Effect Diet and Genetics Have on Gene Expression in Pancreas Tissue in Mice.** *Gage Hough [1\*]*, *Dalton McGeeney [2\*]*, *Jung-Han Kim [3]*, *James Denvir [3†], Nicole Garrison [4†]. [1] Department of Computer of Science, Fairmont State University; [2] Department of Biology, Glenville State College; [3] Department of Biomedical Sciences, Marshall University; [4] Department of Biomedical Sciences, West Liberty University.*

Diabetes is a national health issue that can be investigated using a next generation sequencing approach. In order to better understand the effect of diet and genetics on pancreas function, we used RNAseq data collected from tissue of TALLYHO (TH) and wildtype mice that were fed either normal or high fat diets. Raw Illumina sequence data was filtered and trimmed for quality using FastQC and trimmomatic before being compared to a reference genome in a guided transcriptome mapping using HISAT2. We used the DEseq2 pipeline to identify genes being differentially expressed in each treatment. We found genes that were differentially expressed between the two mouse lines during both diet experimental conditions. There was a difference in gene expression seen during application of a high fat diet when compared to the control diet. These results show activated metabolic pathways when eating a high fat diet and unique pathways in TH mice. These results will assist researchers investigating genetic mechanisms underlying diabetes and devise preventive medication or lifestyle choices. *(Supported by NIH Grant P20GM103434 to the West Virginia IDeA Network for Biomedical Research Excellence and NIDDK/NIH 1R15DK113604 (JH Kim, PI))*

**Comparison of *de-novo* and Reference Guided Assembly Methods in a SARS-CoV-2 Sequencing Analysis Pipeline.** *Malik Khan [1\*], Nicholas Lemon [1\*], Peter Stoilov [2], Pete Perrotta [3], Ryan Percifield [4], James Denvir [5†], Nicole Garrison [6†]. [1] Department of Biology, Shepherd University, [2] Department of Biochemistry, West Virginia University; [3] Department of Pathology, West Virginia University; [4] Department of Biology, West Virginia University; [5] Department of Biomedical Sciences, Marshall University; [6] Department of Biomedical Sciences, West Liberty University.*

When sequencing genomic data, like that of the SARS-CoV-2 virus, accurate assembly is an essential step in identifying variants. There are different computational approaches to genome assembly – reference guided and *de-novo*. Reference based assembly is done by first aligning the nucleotide sequence reads to a reference genome and calling the consensus sequence. The second method is to create a *de-novo* assembly of the reads without comparing to a reference genome. The reference guided method is thought to be reliable in tracking variants closely related to an original reference strain, so we compared this against the *de-novo* assembly method to see if there is a difference in variants detected downstream. Using SARS-CoV-2 samples taken from across West Virginia, we generated a table of viral genomes aligned with the reference genome (using bwa), and then created a consensus sequences for each sample. For the second method, we created *de-novo* assemblies (using the SPAdes assembler) and compiled consensus sequences for each sample. We then subjected each set of consensus sequences to variant calling and lineage assignment with pangolin and compared the resulting assignments. Assembling sequences *de-novo* may reduce bias introduced by the reference sequence, however the impact on SARS-CoV-2 analysis should be minimal due to the small genome size and degree of overlap of most samples to the reference. *(Supported by NIH Grant P20GM103434 to the West Virginia IDeA Network for Biomedical Research Excellence and and WV-CTSI NIGMS.NIH U54GM10)*