**Descriptive Title of Project**

Validation of Gene Expression Processing and Analysis Pipeline, XPRESSpipe, using TCGA Data

**\*Research Use Statement (RUS)**

Advances in RNA-seq have led to their widespread use in biomedical research. However, methods used often vary between protocols in a given lab or consortium often vary. Additionally, novice users of RNA-seq data may have limited training on correct methodologies for effective RNA-seq library processing and analysis. In response to this need for a unification of RNA-seq in silico methodologies, we created the XPRESSyourself suite of bioinformatics tools. To date, this includes pipelines for single-end and paired-end RNA-seq and ribosome profiling, along with a suite of quality control and analysis tools in the pipeline and in a toolkit within the suite called XPRESStools. We also introduce several new tools, mostly pertaining to ribosome profiling, in this suite of pipelines and tools. As the general design of pipelines were designed to conform to TCGA specifications, we wish to use a sampling of TCGA RNA-seq samples, run them through our pipeline, and compare the results to the FPKM counts tables hosting publicly on TCGA for that given sample. We also will create vignettes of how to use the tool using the publicly available data to aid users in using the toolset.

**I am requesting permission to use cloud computing to carry out the research as described in my Research Use Statement.** Yes

**\*Non-technical summary** [*What is a "non-technical summary of your Research Use Statement?"*](https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?from=newprj_project&page=help&topic=public-statement#public-statement&from=newprj_project)

 RNA-seq is a method by which one can profile the changes in gene expression in a cell throughout different conditions. Several methods exist for processing this data. In an effort to unify the computational methodologies for handling this data, we developed an automated, easy to use pipeline, as well as several new analytical tools in try to solve these issues. We designed the pipeline to mirror that of the TCGA processing pipeline. In order to validate the operation of this pipeline, we wish to use TCGA raw RNA-seq data, process it with our pipeline, and compare the output to the processed RNA-seq files publicly available in TCGA.

**\*Choose your Signing Official (SO): Applicant organization**

\*Institution Name University of Utah

Department Biochemistry

Division

\*Street 1 15 N Medical Drive East, Rm 5200

Street 2

\*City Salt Lake City

\*State Utah

\*ZIP/Postal code 84112

\*Country USA

**Create Decryption Password**.

The files distributed through the dbGaP system are encrypted. A password is required for decrypting downloaded files. Please provide a decryption password for the project.

Valid passwords must be at least 8 ASCII characters long and must contain at least 3 of the following 4 characters:

* upper case letters
* lower case letters
* numbers
* non-alphanumeric characters

\*Password for project: RutterXPRESS!

\*Password confirmation:

**RUS**

A RUS is a brief description of the applicant's proposed use of dbGaP dataset(s). The RUS will be reviewed by all NIH Institutes and Centers responsible for data covered by this Data Access Request. Please note that if access is approved, you agree that the RUS, along with your name and institution, will be included on the dbGaP website to describe your research project to the public.

Please make it clear whether you plan to combine requested datasets with other datasets outside of dbGaP, and, if so, whether you plan to analyze these datasets independently or together. If you do plan to combine datasets in any way, please describe your plan and also please discuss whether it creates any additional risks to participants. If you are focusing on outcomes or hypotheses that were not the focus of the primary study (or studies), please describe the outcomes you propose to examine.

Investigators do not need to submit a new project request unless the dataset will be used for research outside of the scope of the approved Research Use Statement

Please enter your RUS in the area below. The RUS should be one or two paragraphs in length and include research objectives, the study design, and an analysis plan (including the phenotypic characteristics that will be tested for association with genetic variants). If you are requesting multiple datasets, please describe how you will use them. Examples of RUS can be found at [GDS website](https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/). Please limit your RUS to 4500 characters.

**Non-technical summary**

Please enter below a non-technical summary of your RUS suitable for understanding by the general public (written at a high school reading level or below). Please limit your non-technical summary to 1300 characters.