

Download SPEAQeasy Example

July 8, 2020

Contents

| | | |
|----------|---|----------|
| 1 | Clone this repository | 1 |
| 2 | Pull example FASTQ data for use in SPEAQeasy | 1 |

1 Clone this repository

First, make sure the main repository is cloned to your system, as it contains code and data which will be used in the upcoming analysis.

```
git clone git@github.com:LieberInstitute/SPEAQeasy-example.git
```

This repository contains genotype and phenotype data for the example set of 42 human samples, as well as the associated main outputs from the SPEAQeasy pipeline. For the purposes of this example, the main outputs include a single VCF file of genotype calls for all samples, and `RangedSummarizedExperiment` objects containing gene, exon, and exon-exon junction counts.

2 Pull example FASTQ data for use in SPEAQeasy

You are encouraged to optionally reproduce our SPEAQeasy results before moving on to the identity resolution and differential expression steps. Those interested in running SPEAQeasy on the raw FASTQ files must download these files. Note that this is an optional step, as we provide the relevant SPEAQeasy outputs for those interested simply in how to use these outputs in subsequent analyses.

We provide a bash script, which utilizes synapse to download the publicly available FASTQ files. Run this script and specify a directory where you wish to place the FASTQ files:

```
bash pull_data/pull_fastq_data.sh [/destination_dir]
```

This script also writes a `samples.manifest` file to the same directory. This text file is used by SPEAQeasy to find the input FASTQ files and associate them with a unique ID.