Requirements for bulk RNA-seq and scRNA-seq to push to MySQL databases:

**Bulk:**

1. Store in a .rdata or .RData or .Rdata object
2. The R object must **only** contain a dds object including the count matrix and the coldata (metadata) associated with that count matrix
3. The dds object should be named “dds” when loaded into the R environment
4. Column headers accepted: treatment, media, day, cell\_line, mutation, dosage, differentiation, condition
   1. If the data has a column that doesn’t fit into one of the above headers, let us know and we can add it

**scRNA-Seq:**

1. Store in an .rdata, .RData, .Rdata object
2. The .RDS, .rds or .Rds should contain the data in a Seurat object
3. All metadata should include separate columns for day, treatment, etc.
4. Please make sure the Seurat object has been processed through the basic vignette up to calling clusters (please see: <https://satijalab.org/seurat/articles/pbmc3k_tutorial.html>). We have written a basic tutorial for building a Seurat object: 1.0\_Run\_Seurat.R
5. Column headers accepted: treatment, media, day, cell\_line, differentiation, method, phenotype, location
   1. If the data has a column that doesn’t fit into one of the above headers, let us know and we can add it

All dds and Seurat objects must be placed in their respective locations on the ec2 server in this location:

/home/ec2-user/bulk

/home/ec2-user/sc

After data is munged and is read to be pushed to the RDS dbs, it will be automatically saved here:

/home/ec2-user/adj\_data/bulk

/home/ec2-user/adj\_data/sc

To run the code do it in screens on the server housing the complex app:

1. Login to bastion: ssh -i rshiny.pem -o IPQoS=throughput -i rshiny.pem ubuntu@3.88.186.250
2. Then login to the server: ssh -i .ssh/rshiny.pem [ubuntu@10.9.3.185](mailto:ubuntu@10.9.3.185)
3. Make sure you’re in the home directory (type cd ~ to make sure)
4. screen, enter
5. Rscript 1.0\_munge\_count\_meta\_data\_SQL.R
6. After the above is finished running, run:

Rscript 2.0\_push\_count\_meta\_data\_SQL.R

1. To clean up files (meaning move datasets that have been pushed to the RDS:

./cleanup.sh

**Questions/Comments:**

1. **We will push new datasets 1/month**
2. **We will need to cross-check the data to make sure data is correctly specified (metadata and count matrix).**