

## Instructions on How to Run This Code

This notebook is designed for the **Al Jord** team at the **Centre for Genomic Regulation**. All components in this repository operate within a *Virtual Machine* (either Docker or Singularity), ensuring consistent functionality regardless of when or where it is downloaded, as long as the working directory is the root of this repository.

---

### Accessing the HPC Cluster

To access the HPC Cluster, you may require assistance from **Emyr James**, the head of Scientific IT as of 2024-2025. Contact him first to grant you access to the cluster and then follow these steps to set you up:

1. **(Optional for external access)** Install Forticlient VPN:
  - Select *Add New Connection* and choose *SSL-VPN*.
  - Set ***https://vpn.crg.es:10000/sslvpn/*** as the *Remote Gateway*.
  - Use your CRG credentials for the *Username* and *Password*.
2. **SSH Client Setup:**
  - For *MacOS* or *Linux*, open Terminal and run:  
`ssh login1.hpc.crg.es -l <your-CRG-username>`  
Enter your CRG password when prompted.
  - For *Windows*, open *Powershell* or *Terminal* as Administrator and run:  
`wsl --install`  
`wsl --set-default-version 2`  
`wsl --install Ubuntu-20.04`  
Follow the installation steps for Ubuntu. Then, run:  
`ssh login1.hpc.crg.es -l <your-CRG-username>`  
Enter your CRG password when prompted.

You will now be in your personal folder within the lab team directory.

---

### Download the Repository

Clone the repository to your HPC folder:

```
git clone https://github.com/andresgordoortiz/24CRG_ADEL_MANU_OOCYTE_SPLICING.git
cd 24CRG_ADEL_MANU_OOCYTE_SPLICING
```

---

Explore the repository's folders:

```
ls
```

**Note:** These folders correspond to those in *Isilon*. Files can be uploaded or downloaded through Isilon as needed.

## Running the Analysis

Run the following pipelines to execute the complete analysis, from downloading samples to generating the final report and Excel tables:

```
# Important: you must pass a suitable VASTDB database as absolute path to run the pipelines
sbatch workflows/full_processing_pipeline_fmndko.sh /users/mirimia/projects/vast-tools/VASTDB
sbatch workflows/full_processing_pipeline_pladb.sh /users/mirimia/projects/vast-tools/VASTDB
sbatch workflows/full_processing_pipeline_spire.sh /users/mirimia/projects/vast-tools/VASTDB
```

**Important:** Ensure the workflow is provided with a valid path to the VASTDB database. The code above should run smoothly since Manu keeps a copy of it in his folder but, if not available, download the Mm2 database as follows:

```
# This will download the VASTDB for the mouse assembly.
mkdir VASTDB
wget https://vastdb.crg.eu/libs/vastdb.mm2.23.06.20.tar.gz
tar -xzf vastdb.mm2.23.06.20.tar.gz -C VASTDB
```

And then run the pipelines with the new VASTDB:

```
sbatch workflows/full_processing_pipeline_fmndko.sh $(pwd)/VASTDB
sbatch workflows/full_processing_pipeline_pladb.sh $(pwd)/VASTDB
sbatch workflows/full_processing_pipeline_spire.sh $(pwd)/VASTDB
```

After a few hours the analysis should finish, but you can check the estatus of your query using:

```
squeue -u <your-CRG-user>
```

After it has finished, run the R Report (it will take an hour or so)

```
sbatch scripts/R/run_notebook.sh
```

Signed: Andrés Gordo Ortiz