

Single cell genomics workshop

19th of April 2017, KU Leuven

Alejandro Sifrim

alejandro.sifrim@kuleuven.be

Introduction

In this tutorial we will plunge into the exciting world of single cell genomic data analysis. First we will look at how our data structures look like and at ways to quickly get important summary statistics out of them. Afterwards we will perform copy-number analysis on these files.

Preparation

To run these exercises I've created a virtual machine with all the necessary software and data. All you need to do is create a virtual machine with the provided image, for this we will use "free" software named VirtualBox.

Steps:

1. Install Virtualbox
2. Download image from: <http://bit.ly/2pPP3T9>
3. Create new virtual machine with image (use at least 1GB of RAM).

This will require at least 6.3 GB of hard drive space.

4. Boot up the virtual machine

The user credentials for the virtual machine are:

- Username: sc1
- Password: sc1

Never use such ludicrous credentials in real life!

Data

Data can be found in your home folder:

```
/home/sc1/data/
```

Checkout the contents of the data folder.

► Spoiler

Questions:

- Is this single end or paired-end data?
- What is (roughly) the average coverage of the samples provided?
- Bonus: What is obviously weird about these single cell data?

► Spoiler

CNV Calling

[Ginkgo](#) has been installed on this virtual machine. This is an interactive

tool to explore CNVs from single-cell low coverage genomic data. In this tutorial we will use the web interface but keep in mind that these analyses can be performed from the command line which can be useful for automation in larger analysis pipelines.

1) To access Ginkgo open a browser and go to:

```
localhost/ginkgo
```

If you're having problems reaching localhost, set the network settings for the virtual box to "not attached".

You can also use the [public Ginkgo server](#) for small batches of data without having to perform a local install. As to not destroy their server please use the version installed in the virtual machine.

2) Read the instructions and format the data as to fit the Ginkgo input format.

► Spoiler

3) Once your data is converted import it into Ginkgo.

4) Explore Ginkgos "Advanced Parameters". Stay of the "Experimental features".

5) Run Ginkgo for 500kb bins with variable and fixed bin size. Observe the difference (use different browser tabs for convenience).

6) Run Ginkgo for 250kb bins and 2.5Mb bins. Observe the difference (use different browser tabs for convenience).

Questions

- Which cells seem to belong to the same cell type?
- Do you see heterogeneity?
- Look at coverage uniformity across cells. Is there a particular trend here? Can you explain it?
- Are there any weird looking cells? Look at these cells under different resolutions.
- How does their read count frequency plot look like?
- What could be a potential explanation for this?
- Observe the copy number profiles, do you believe the automatically chosen ploidy?