

Homework Week 3

Friederike Duendar and Luce Skrabanek

ANGSD Course 2019

Questions

1. A somatic human cell contains about 6 picograms of DNA. How much DNA does a sperm cell contain? **(1pt)**
2. How many human cells will you need to obtain 10 micrograms of DNA, as is requested by many sequencing protocols? **(1pt)**
3. Describe one typical cause for DNA loss during DNA extraction **(1pt)**.
4. Describe two functions of the adapters that are typically added during Illumina's library preparations. Which types of sequences do they often entail? **(2pts)**
5. Components of what kind of lab equipment have to be integrated in any type of Illumina sequencer? **(1pt)**
 - agarose gel
 - centrifuge
 - microwave
 - microscope
 - scale

Exercises

Similarly to last week, please make sure to document all the steps you take, including making folders etc.

1. Write a for-loop to download all `fastq` files of WT biological replicate no. 2 of the Gierlinski data set (`UNIX`). Try to have a solution that's as generally applicable as possible. **(3pts)**
 - [ENA](#) accession number: ERP004763 (feel free to download from another entry point to the SRA)
 - You can download the summary of the sample information, including the URLs for the fastq files, by clicking on the `TEXT` link on the **Project** page. You can also use `wget` to download this file, after you have clicked the link once (the URL needs to be generated for each dataset).

- Use the mapping information available at <http://dx.doi.org/10.6084/m9.figshare.1416210> to figure out which ENA accession names you need to obtain the samples of WT biological replicate no 2. Use the UNIX commands you know to generate a file with the sample names and then use that list as input to your for-loop.
- 2. Why are there multiple `fastq` files per sample? What does each file represent? **(1pt)**
- 3. Count the number of reads stored in each FASTQ file and keep a note of the results (UNIX). The `zcat` command allows you to see the contents of a gzipped file. **(2pts)**

Project

1. Identify and write down one or two biological or technical questions of interest for your **project**. **(1pt)**

Compile the `.Rmd` file and send both the `.Rmd` and the HTML files to angsd_wmc@zohomail.com by Saturday night.