Cellranger Workflow

Files and requirements - What and where to get them?

e!Ensembl → fasta files [Gallus_GRCg7b.dna.toplevel.fa]
→ gtf files [Gallus_GRCg7b.110.fa]

10x GENOMICS → application [cellranger-arc]

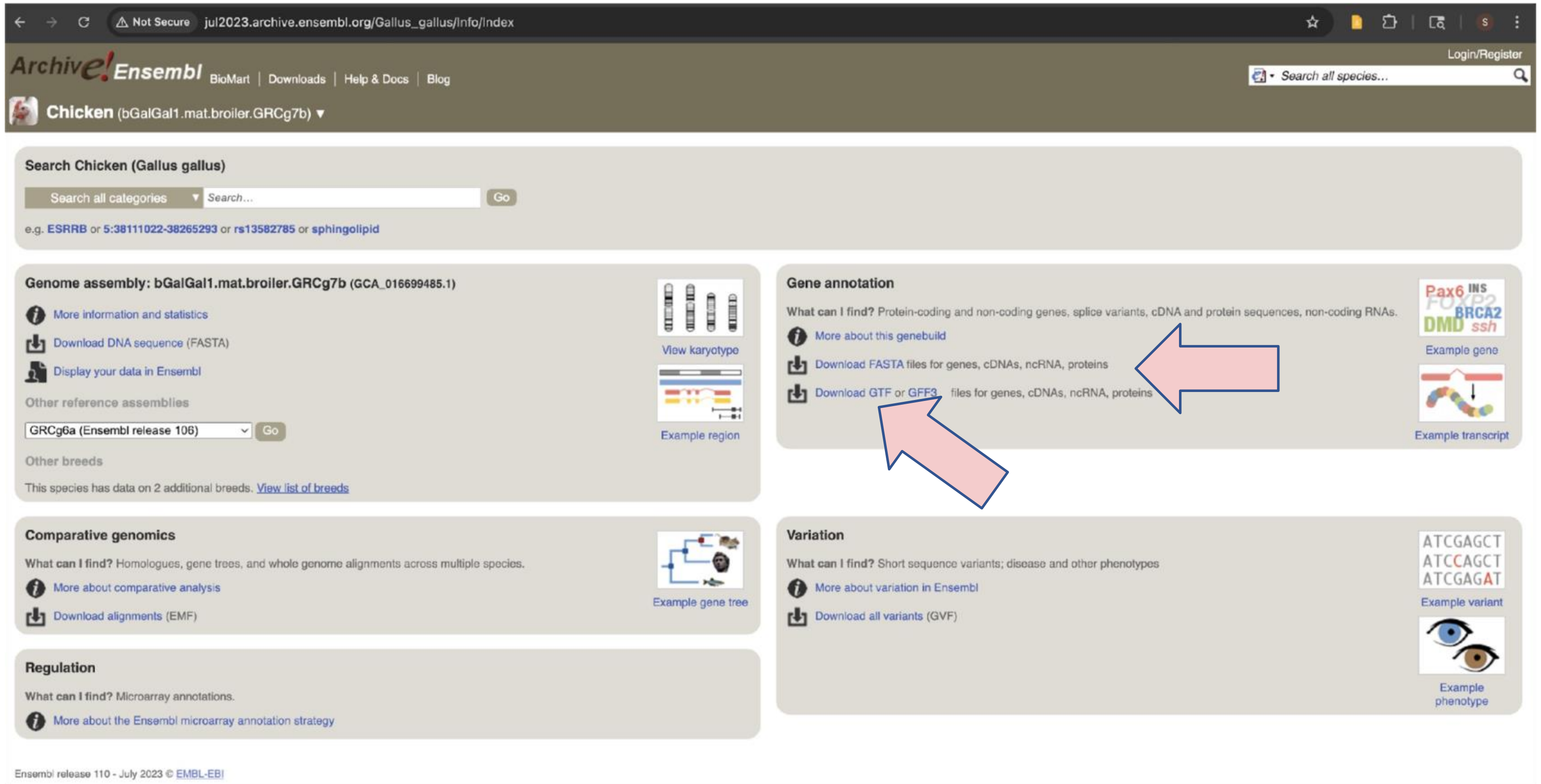
Novogene → fastq files [raw data]

→ terminal [powerful computer]

File Requirements:

- Genome
- GTF

I have used the release **110** from Ensembl (**bGalGal1.mat.broiler.GRCg7b**)



jul2023.archive.ensembl.org/Gallus_gallus/Info/Index

Chicken (bGalGal1.mat.broiler.GRCg7b)

Search Chicken (Gallus gallus)

e.g. ESRRB or 5:38111022-38265293 or rs13582785 or sphingolipid

Genome assembly: bGalGal1.mat.broiler.GRCg7b (GCA_016699485.1)

- More information and statistics
- Download DNA sequence (FASTA)
- Display your data in Ensembl

Other reference assemblies

GRCg6a (Ensembl release 106)

Other breeds

This species has data on 2 additional breeds. View list of breeds

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild
- Download FASTA files for genes, cDNAs, ncRNA, proteins
- Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis
- Download alignments (EMF)

Variation

What can I find? Short sequence variants; disease and other phenotypes

- More about variation in Ensembl
- Download all variants (GVF)

Regulation

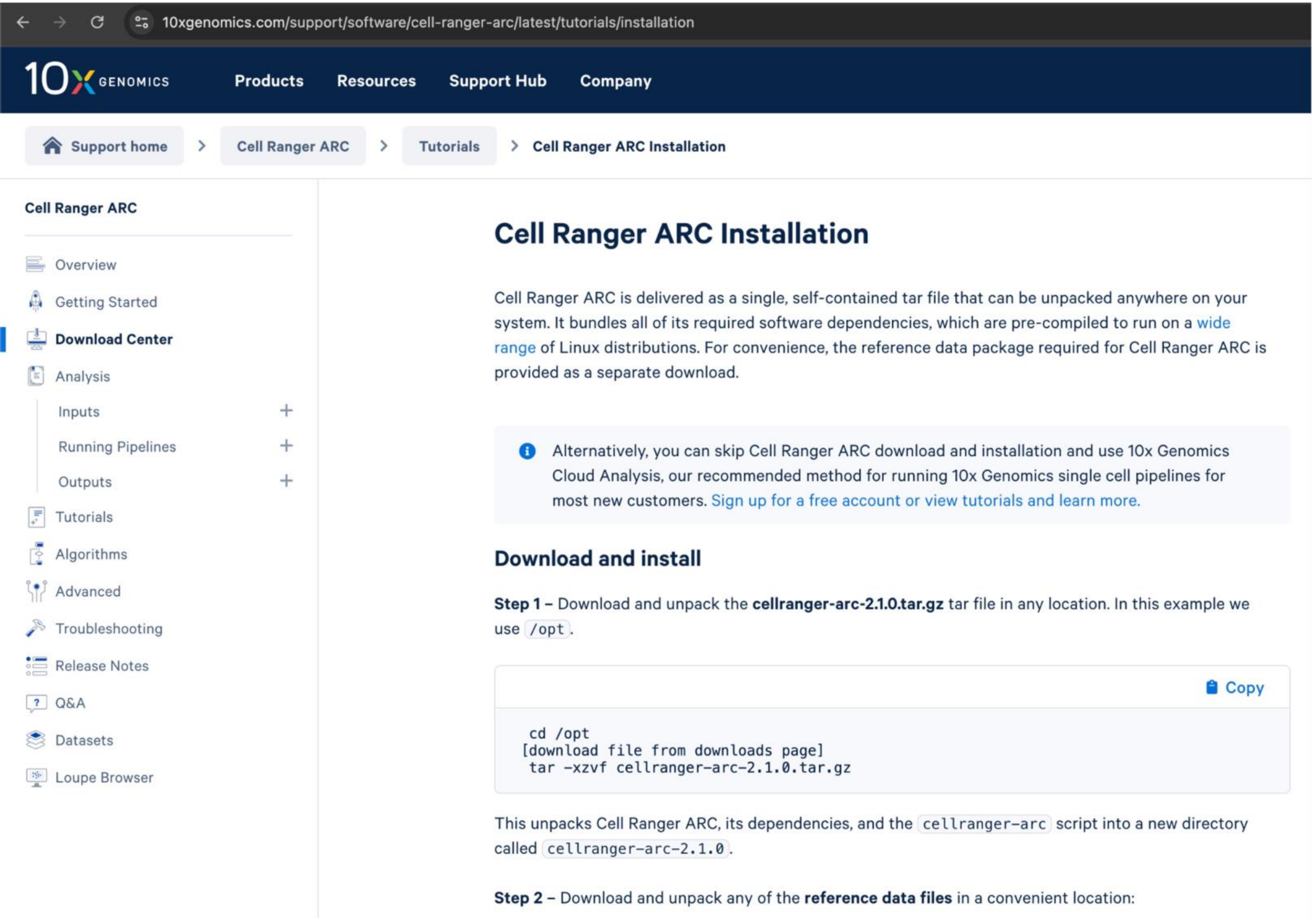
What can I find? Microarray annotations.

- More about the Ensembl microarray annotation strategy

Ensembl release 110 - July 2023 © EMBL-EBI

File Requirements:

- cellranger-arc



Cell Ranger ARC Installation

Cell Ranger ARC is delivered as a single, self-contained tar file that can be unpacked anywhere on your system. It bundles all of its required software dependencies, which are pre-compiled to run on a wide range of Linux distributions. For convenience, the reference data package required for Cell Ranger ARC is provided as a separate download.

Alternatively, you can skip Cell Ranger ARC download and installation and use 10x Genomics Cloud Analysis, our recommended method for running 10x Genomics single cell pipelines for most new customers. Sign up for a free account or view tutorials and learn more.

Download and install

Step 1 – Download and unpack the **cellranger-arc-2.1.0.tar.gz** tar file in any location. In this example we use `/opt`.

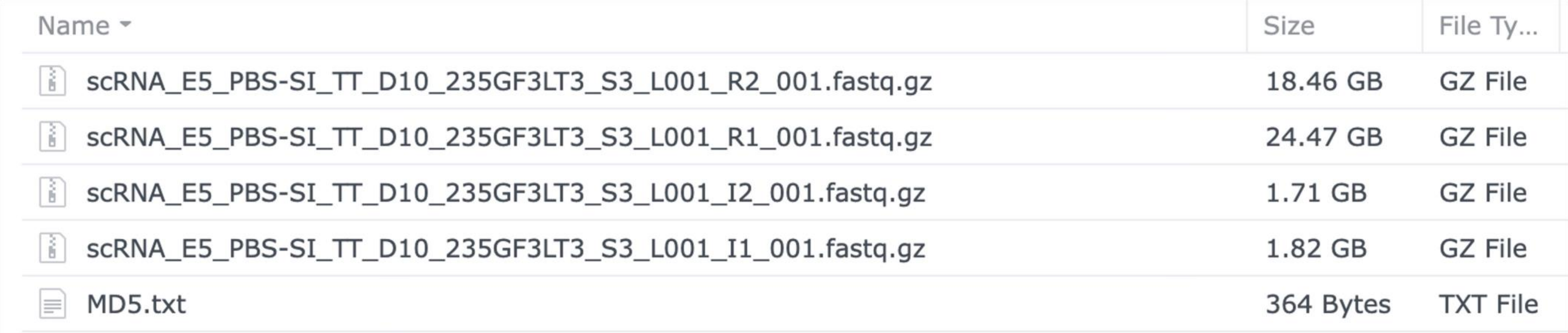
```
 cd /opt
[download file from downloads page]
 tar -xzvf cellranger-arc-2.1.0.tar.gz
```

This unpacks Cell Ranger ARC, its dependencies, and the `cellranger-arc` script into a new directory called `cellranger-arc-2.1.0`.

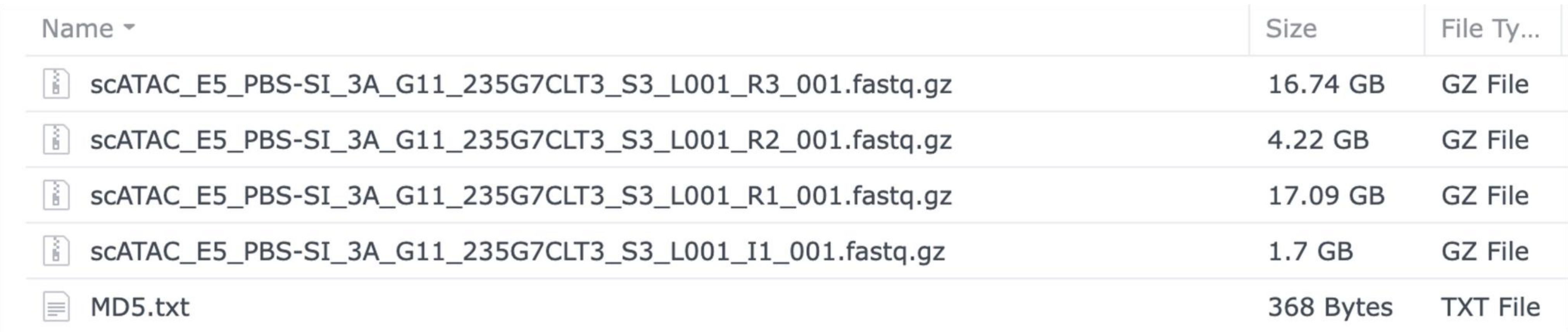
Step 2 – Download and unpack any of the **reference data files** in a convenient location:

File Requirements:

- raw data fastq files



Name	Size	File Ty...
scRNA_E5_PBS-SI_TT_D10_235GF3LT3_S3_L001_R2_001.fastq.gz	18.46 GB	GZ File
scRNA_E5_PBS-SI_TT_D10_235GF3LT3_S3_L001_R1_001.fastq.gz	24.47 GB	GZ File
scRNA_E5_PBS-SI_TT_D10_235GF3LT3_S3_L001_I2_001.fastq.gz	1.71 GB	GZ File
scRNA_E5_PBS-SI_TT_D10_235GF3LT3_S3_L001_I1_001.fastq.gz	1.82 GB	GZ File
MD5.txt	364 Bytes	TXT File



Name	Size	File Ty...
scATAC_E5_PBS-SI_3A_G11_235G7CLT3_S3_L001_R3_001.fastq.gz	16.74 GB	GZ File
scATAC_E5_PBS-SI_3A_G11_235G7CLT3_S3_L001_R2_001.fastq.gz	4.22 GB	GZ File
scATAC_E5_PBS-SI_3A_G11_235G7CLT3_S3_L001_R1_001.fastq.gz	17.09 GB	GZ File
scATAC_E5_PBS-SI_3A_G11_235G7CLT3_S3_L001_I1_001.fastq.gz	1.7 GB	GZ File
MD5.txt	368 Bytes	TXT File

Terminal

File Requirements:

- cellranger-arc binary
- raw data fastq files
- fasta files
- gtf files

```
> cellranger-arc[mkref]   + fasta files + gtf files = cellranger reference

> cellranger-arc[counts] + cellranger reference + fastq files = filtered h5 file
```

Step 1: cellranger Reference

Make Reference file input for Chicken Genome

```
{
   organism: "chicken"
   genome: ["GRCg7b"]
   input_fasta: ["./Gallus_gallus.bGalGal1.mat.broiler.GRCg7b.dna.toplevel.fa"]
   input_gtf: ["./Gallus_gallus.bGalGal1.mat.broiler.GRCg7b.110.gtf"]
}
```

txt.file

Run cellranger command to create reference using the file input

```
cellranger-arc mkref \
  --config=1Pre-ref.config
```

code

Step 2: Cellranger count

Library csv input for Chicken Genome

```
 $ Library_E4FGF.csv
fastqs,sample,library_type
./FGF_Samples/ATAC_fastq,E4_FGF,Chromatin Accessibility
./FGF_Samples/RNA_fastq,E4_FGF,Gene Expression

 $ Library_E5FGF.csv
./FGF_Samples/ATAC_fastq,E5_FGF,Chromatin Accessibility
./FGF_Samples/RNA_fastq,E5_FGF,Gene Expression
```

csv.files

cellranger-arc count for Chicken Genome

```
#!/bin/bash
# File: Count.sh
list=(E4FGF E5FGF)
for i in ${list[@]} do
cellranger-arc count --id=Multiomix_${i} \
 --reference=./GRCg7b \
 --libraries=./Library_${i}.csv \
 --localcores=2 \
 --localmem=8 \
 --create-bam=true
echo "Finish RNA-ATAC_${i}"
done
```

Code: bash Count.sh