

*Jeffrey Granja and Ryan Corces*

---

# ***ArchR: Comprehensive scATAC analysis in R***



---

## ***Contents***

---

<b>List of Tables</b>	<b>5</b>
<b>List of Figures</b>	<b>7</b>
<b>Preface</b>	<b>9</b>
<b>1 CHAPTER 1</b>	<b>11</b>
1.1 What is an ArrowFile / ArchRProject? . . . . .	11
1.2 Getting Set Up . . . . .	13
<b>2 Creating Arrow Files</b>	<b>15</b>
2.1 QC TSS Scores by Unique Fragments . . . . .	16



---

## ***List of Tables***

---



---

## ***List of Figures***

---





---

## ***Preface***

---

This is the ArchR documentation.



# 1

---

## CHAPTER 1

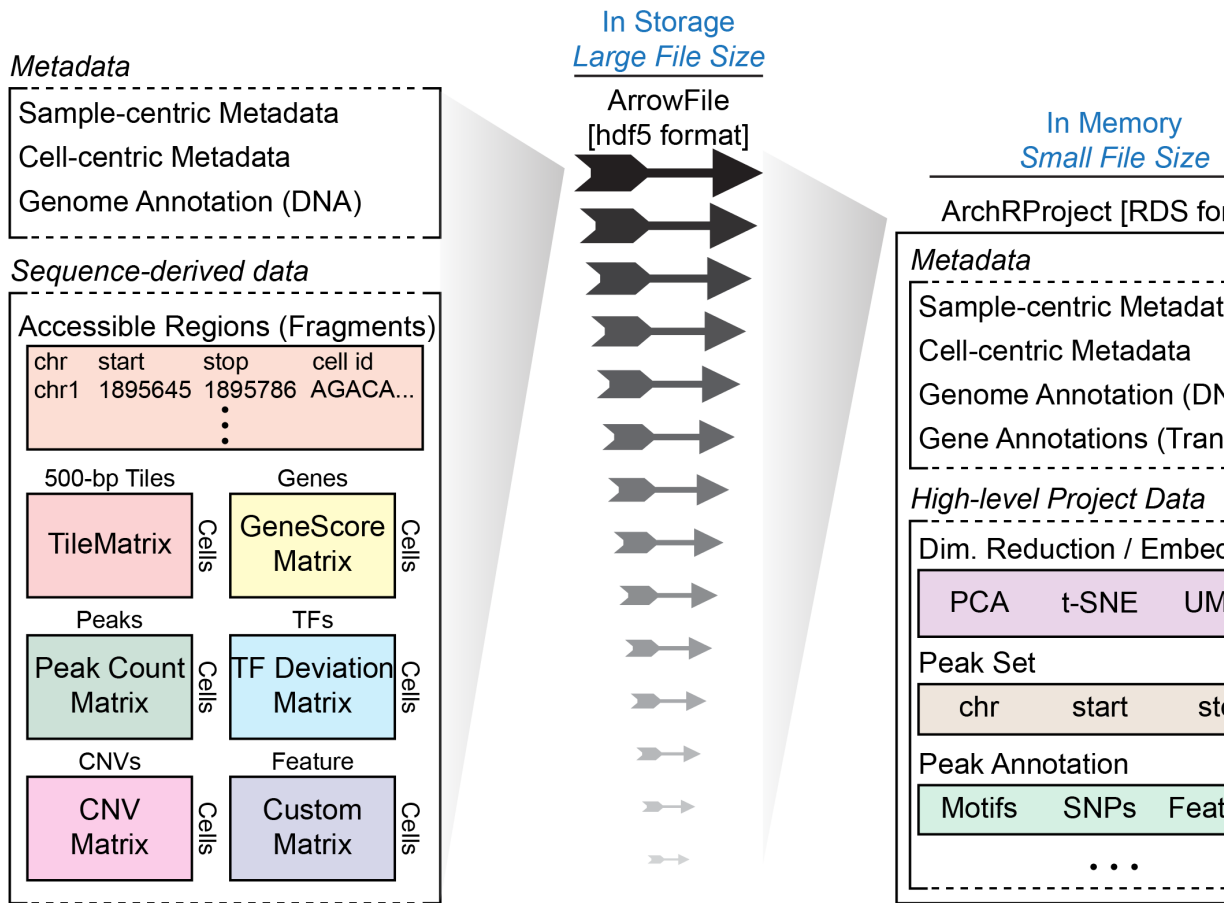
---

The following tutorial shows the basics of setting up and interacting with an `ArchRProject` using a gold-standard downsampled dataset of hematopoietic cells [Granja\\* et al. Nature Biotechnology 2019](#). This tutorial and all of the accompanying vignettes assume that you are running ArchR **locally**. **QQQ** Once all of these steps work for you, feel free to [set up ArchR to work in a cluster environment](#). This tutorial does not explain every detail of every step but rather shows you how to perform the most common analyses. Please see the [Vignettes section](#) for more details on each major analytical step and all of the major features of ArchR.

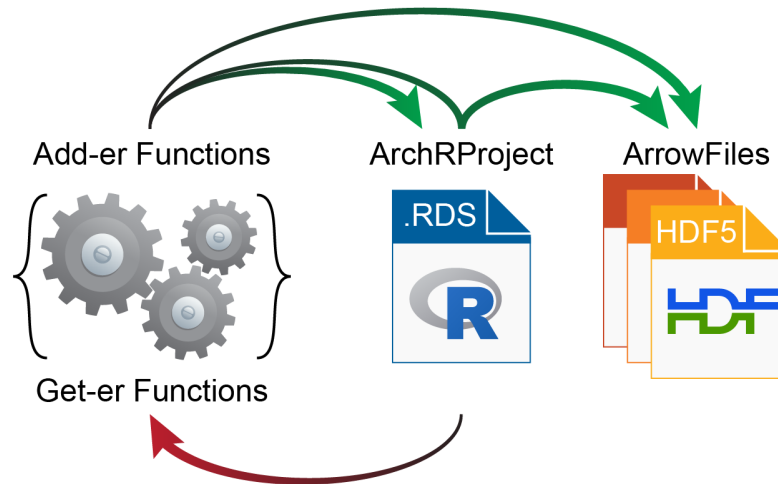
---

### 1.1 What is an `ArrowFile` / `ArchRProject`?

The base unit of an analytical project in ArchR is called an `ArrowFile`. Each `ArrowFile` stores all of the data associated with an individual sample (i.e. metadata, accessible fragments, and data matrices). Here, an “individual sample” would be the most detailed unit of analysis desired (for ex. a single replicate of a particular condition). During creation and as additional analyses are performed, ArchR updates and edits each `ArrowFile` to contain additional layers of information. It is worth noting that, to ArchR, an `ArrowFile` is actually just a path to an external file stored on disk. More explicitly, an `ArrowFile` is not an R-language object that is stored in memory. Because of this, we use an `ArchRProject` object to associate these `ArrowFiles` together into a single analytical framework that can be rapidly accessed in R. This `ArchRProject` object is small in size and is stored in memory.



Certain actions can be taken directly on ArrowFiles while other actions are taken on an ArchRProject which in turn updates each associated ArrowFile. Because ArrowFiles are stored as large HDF5-format files, “get-er” functions in ArchR retrieve data by interacting with the ArchRProject while “add-er” functions either (i) add data directly to ArrowFiles, (ii) add data directly to an ArchRProject, or (iii) add data to ArrowFiles by interacting with an ArchRProject.



---

## 1.2 Getting Set Up

The first thing we do is set up our working directory, load our gene and genome annotations, and set the number of threads we would like to use. Depending on the configuration of your local environment, you may need to modify the number of threads used below in `addArchRThreads()`. By default ArchR uses half of the total number of threads available but you can adjust this manually as you see fit. If you are using windows, the usable threads will automatically be set to 1 because the parallel processing in ArchR is build for Unix-based operating systems.

For the purposes of this tutorial, we provide the gene and genome annotations but you can create your own using the `createGeneAnnotation()` and `createGenomeAnnotation()` functions. See the [Gene and Genome Annotations vignette](#) for more information.

```
#Load R Libraries  
library(ArchR)
```

```
#Create a new folder and set this as the working directory for tutorial and
wd <- "ArchR_tutorial"
dir.create(wd, showWarnings = FALSE, recursive = TRUE)
setwd(wd)

#Load genome annotations. Available annotations are for "Hg19", "Hg38", "Mm10"
data("geneAnnoHg19")
data("genomeAnnoHg19")
geneAnno <- geneAnnoHg19
genomeAnno <- genomeAnnoHg19

#Set Default Threads for ArchR Functions
addArchRThreads(threads = floor(parallel::detectCores()/2)) #QQQ I favor in
```

## 2

### *Creating Arrow Files*

For this tutorial, we will download a collection of fragment files. Fragment files are one of the base file types of the 10x Genomics analytical platform (and other platforms) and can be easily created from any BAM file. See [the ArchR input file types vignette](#) for information on making your own fragment files for input to ArchR. Once we have our fragment files, we provide their paths as a character vector to `createArrowFiles()`. During creation, some basic metadata and matrices are added to each ArrowFile including a “TileMatrix” containing insertion counts across genome-wide 500-bp bins (see `addTileMatrix()`) and a “GeneScoreMatrix” that is determined based on weighting insertion counts in tiles nearby a gene promoter (see `addGeneScoreMatrix()`). These gene activity scores are described in more depth in the [Gene Activity Score vignette](#).

```
#Get Tutorial Data ~2.2GB To Download (if downloaded already ArchR will bypass)
inputFiles <- getTutorialData("Hematopoiesis")

#Create Arrow Files (~10-15 minutes) w/ helpful messages displaying progress
#For each sample, this step will:
# 1. Read accessible fragments.
# 2. Calculate QC Information for each cell (TSS Enrichment, Nucleosome info)
# 3. Filter cells based on QC parameters.
# 4. Create a genome-wide TileMatrix using 500-bp bins.
# 5. Create a GeneScoreMatrix using the provided geneAnnotation.
ArrowFiles <- createArrowFiles(
  inputFiles = inputFiles,
  sampleNames = names(inputFiles),
  geneAnnotation = geneAnno,
  genomeAnnotation = genomeAnno,
```

```
filterTSS = 4,  
filterFrag = 1000,  
addTileMat = TRUE,  
addGeneScoreMat = TRUE  
)
```

---

## 2.1 QC TSS Scores by Unique Fragments

Since this was plotted prior to creation of an ArchRProject we go to `QualityControl/scATAC_BMMC_R2/scATAC_BMMC_R2-TSS_by_Unique_Frags.pdf` for the plot below.

---

## 2.2 QC Fragment Size Distribution

Since this was plotted prior to creation of an ArchRProject we go to `QualityControl/scATAC_BMMC_R2/scATAC_BMMC_R2-Fragment_Size_Distribution.pdf` for the plot below.