# Day 1 overview

- SSH keys and logging on to AWS
- Review library prep and sequencing
- Basic Bash/Intro to Git
- Library QC
- Vim and Vimtutor

# Logging on to the super computer

- On the website (https://biodatasci.colorado.edu/shortread/sr2023/) under Day 1, open and follow the worksheet named **SSH Keys and AWS**

- You must
  1) Be logged into your GitHub account
  2) Have a terminal application open (see worksheet for details)

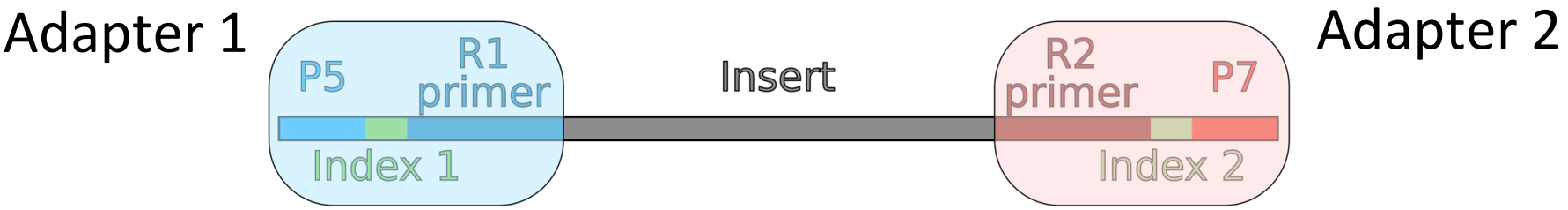
- After adding your SSH keys, log into the AWS:

```
ssh <github_username>@18.216.90.72
```

- If you need help, flag us with a red sticky note
- If you are successfully logged on, put up a green sticky note

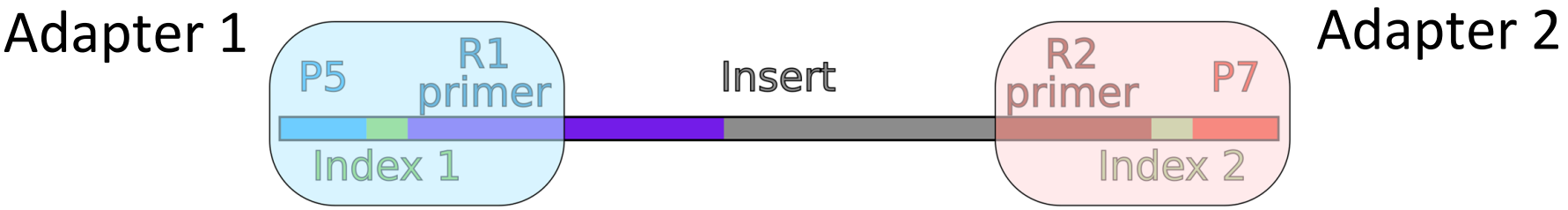
# Library prep and sequencing

# Anatomy of a library



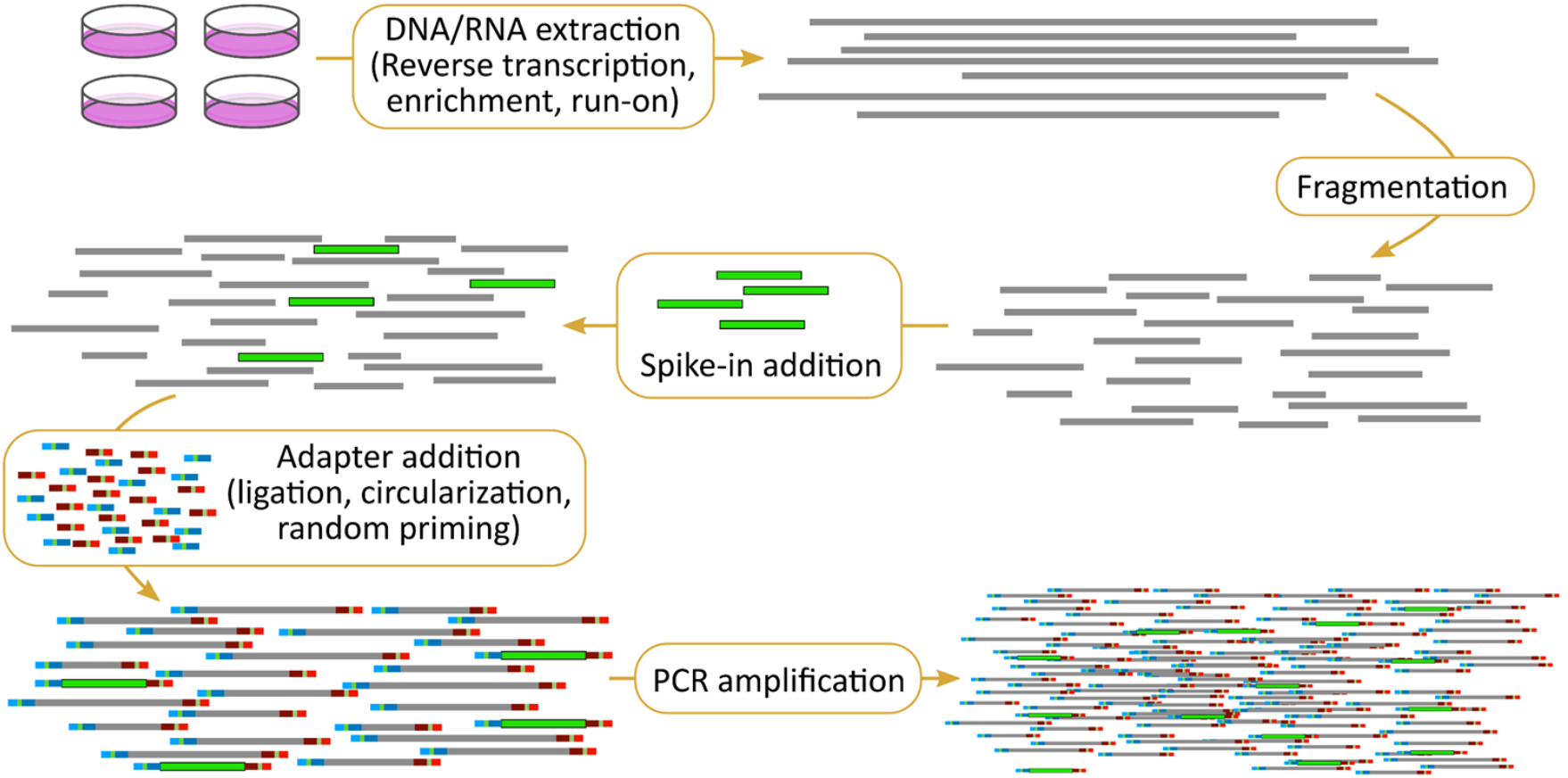

| | |
|---|---|
| P5/P7 | Ends that attach to flow cell |
| Index 1/2 | ID sequences for multiplexing samples |
| R1/R2 primers | Sequencing primers |
| Insert | Fragment of sample DNA/cDNA |

# Anatomy of a library




| | |
|---|---|
| P5/P7 | Ends that attach to flow cell |
| Index 1/2 | ID sequences for multiplexing samples |
| R1/R2 primers | Sequencing primers |
| Insert | Fragment of sample DNA/cDNA |
| Read | Sequenced portion of fragment |

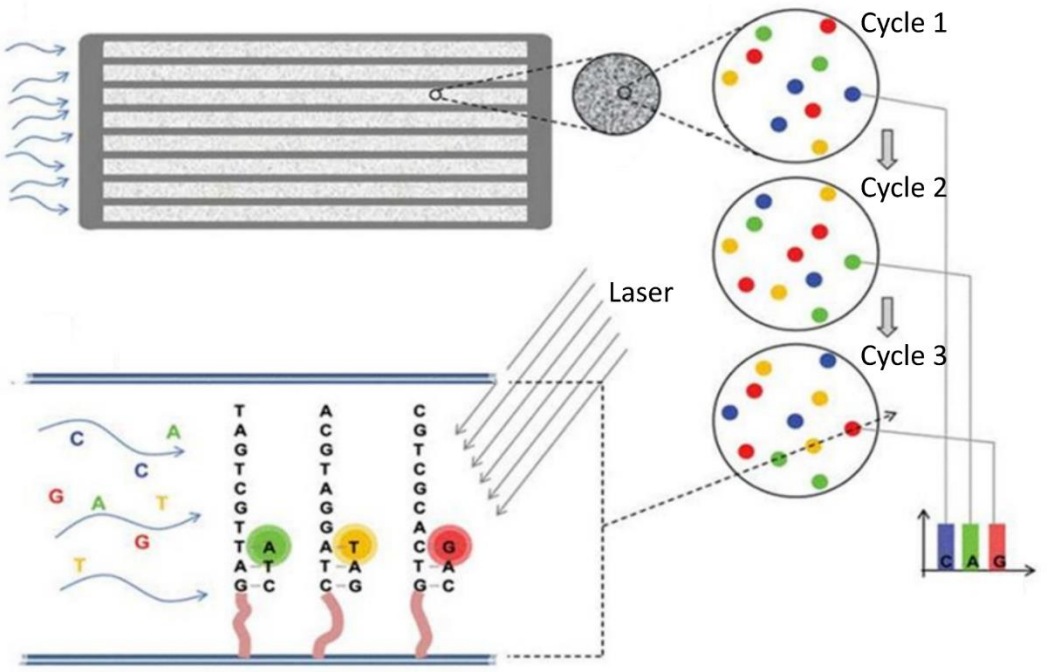
# Creating libraries



# Illumina sequencing technology

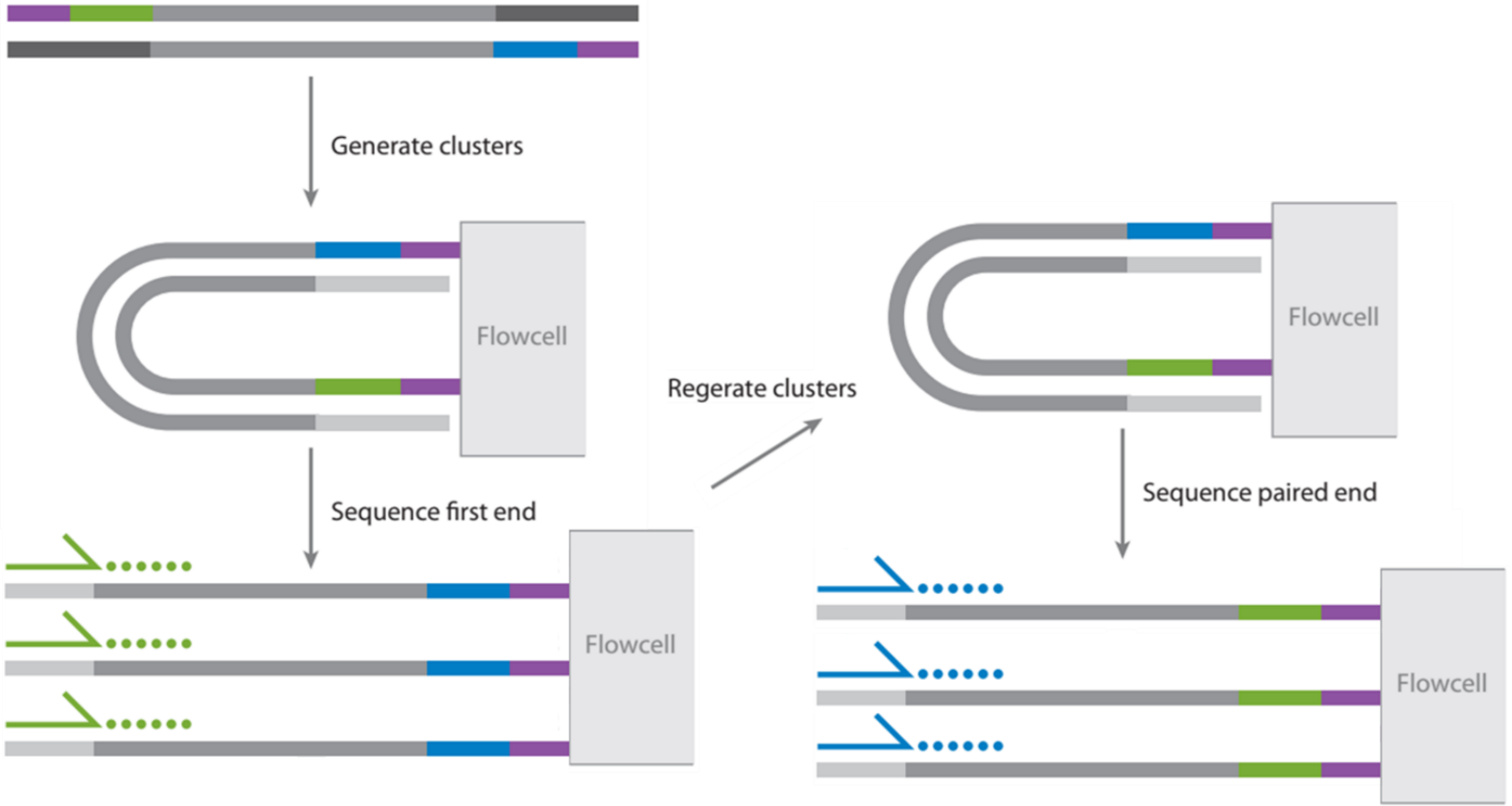
Imaging a slide (flow cell) with millions/billions of DNA clusters by cycling in fluorescent nucleotides

Sequencing:



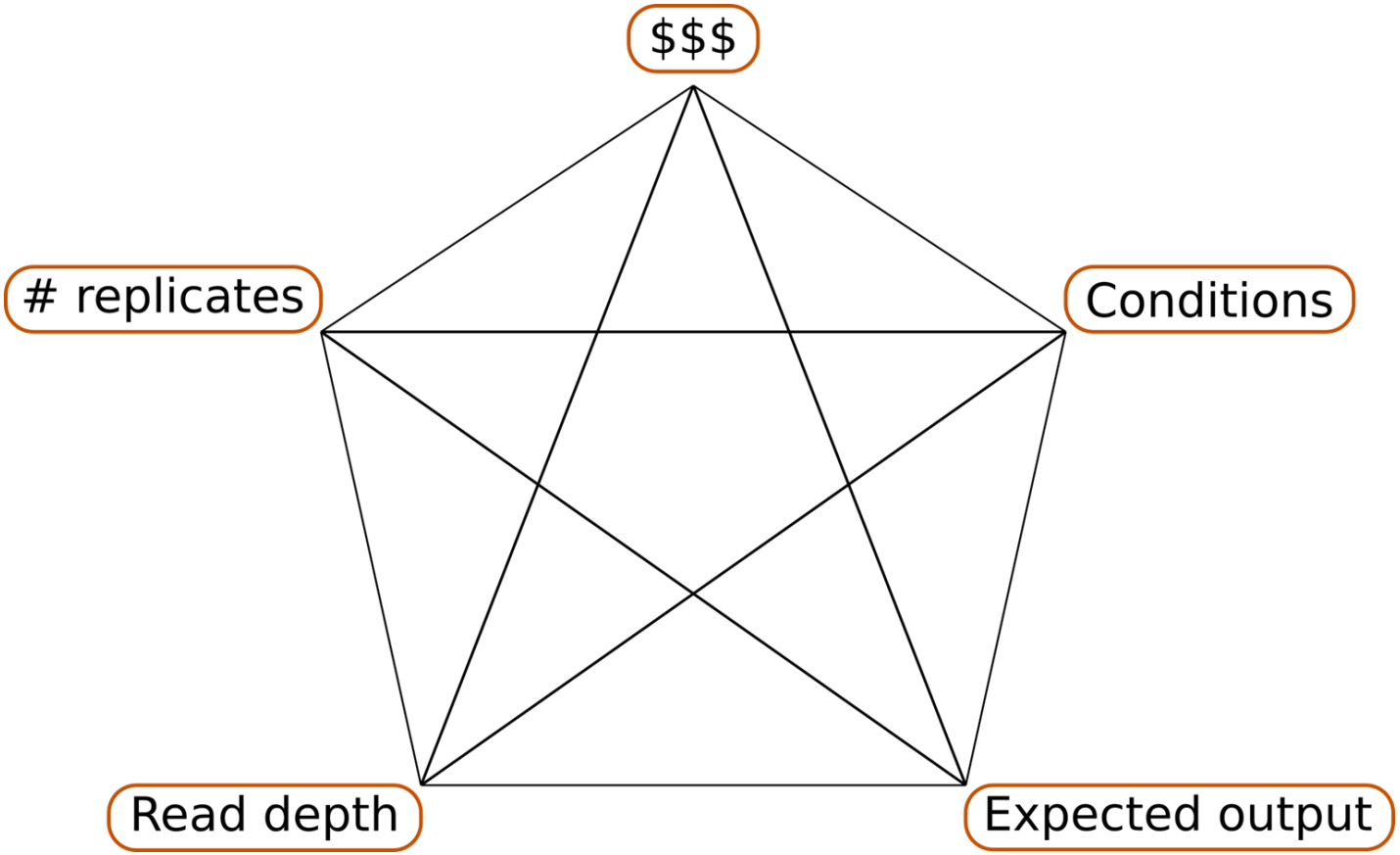

http://tucf-genomics.tufts.edu/home/ordering

# Single/paired end sequencing




Mardis. https://doi.org/10.1146/annurev-anchem-062012-092628

# Designing a sequencing experiment



# Basic Bash and Git/GitHub

- On the website (https://biodatasci.colorado.edu/shortread/sr2023/) under Day 1, open and follow the worksheet named **Git, GitHub, and basic Bash**

- You'll interact with the sr2023 GitHub repository. We will use this repository throughout the class

  This repo is at https://github.com/Dowell-Lab/sr2023

- If you need help, flag us with a red sticky note

- If you've cloned the repo and finished the worksheet, put up a green sticky note

Break

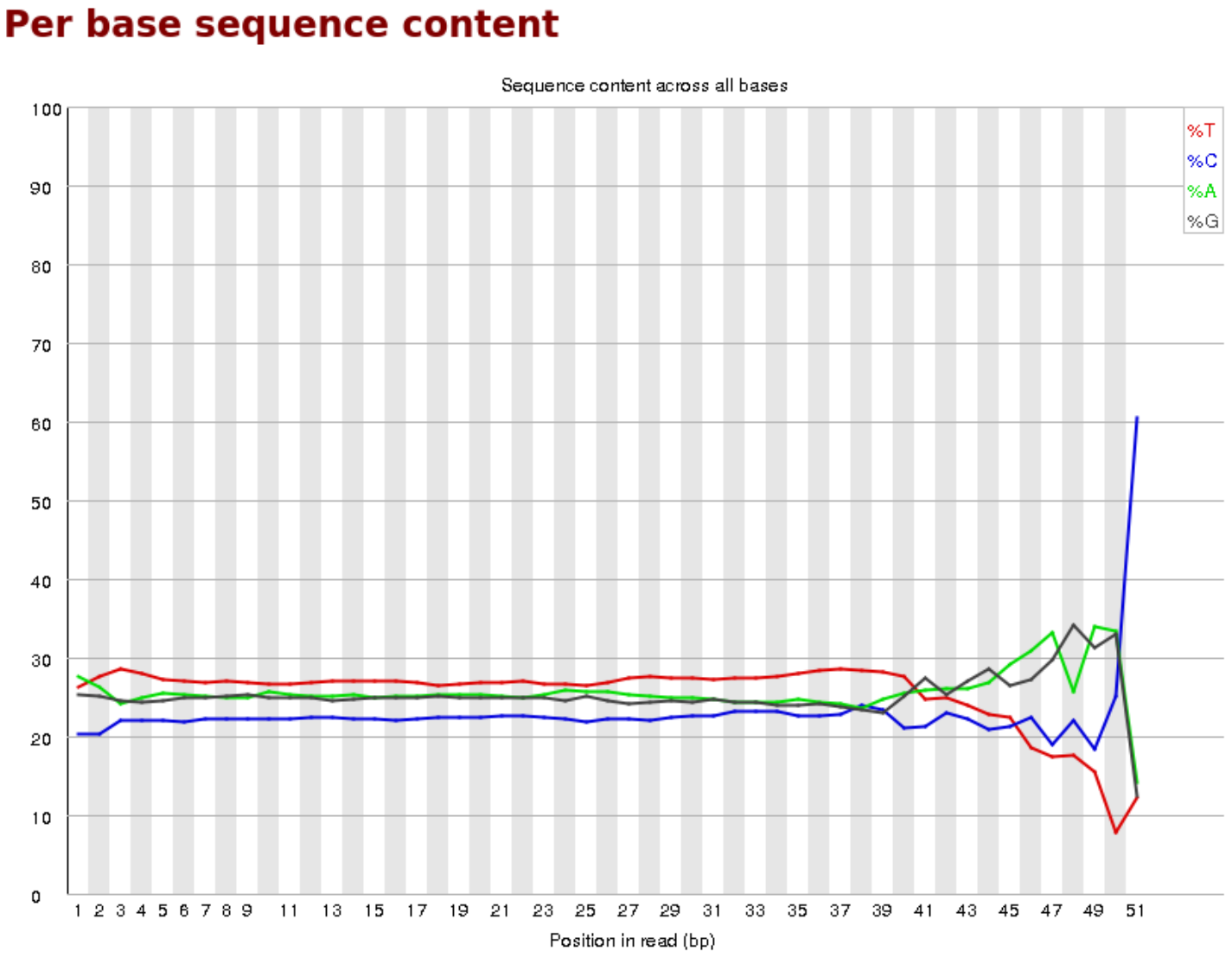
(If you haven’t gotten onto the AWS, please log into Google Shell now)

# Library QC

# FastQC

Base diversity

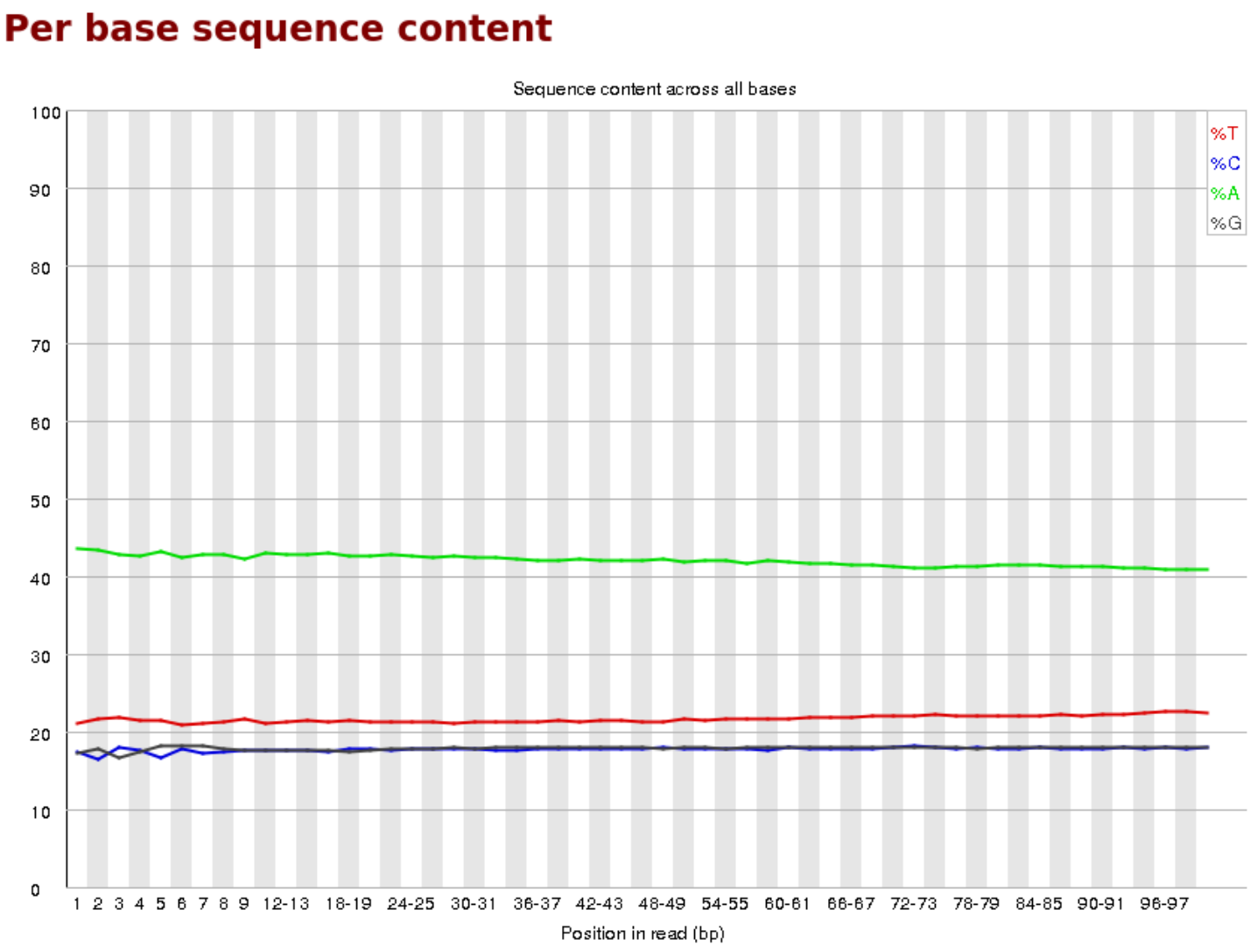
Complexity



# FastQC

Base diversity

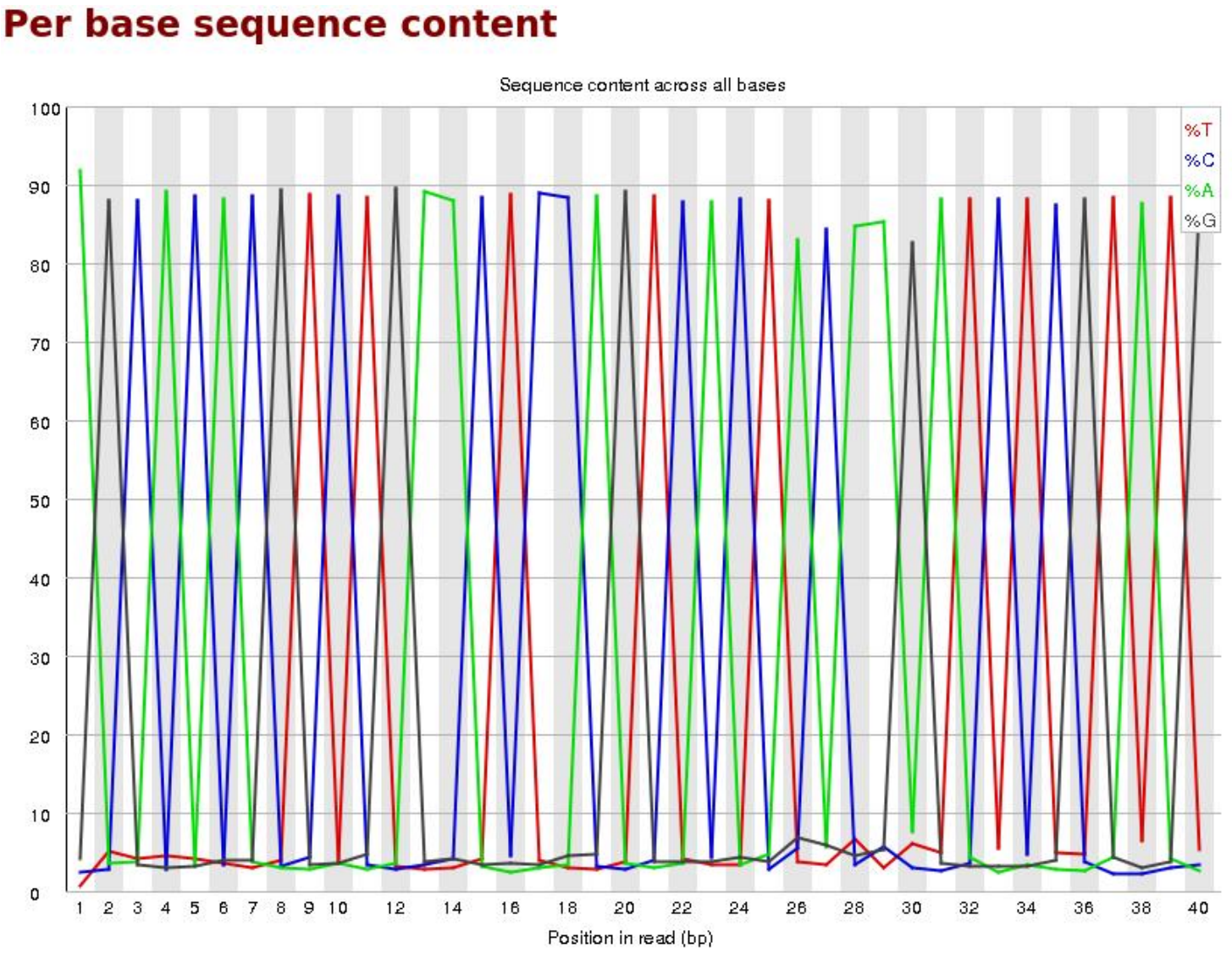
Complexity



# FastQC

Base diversity

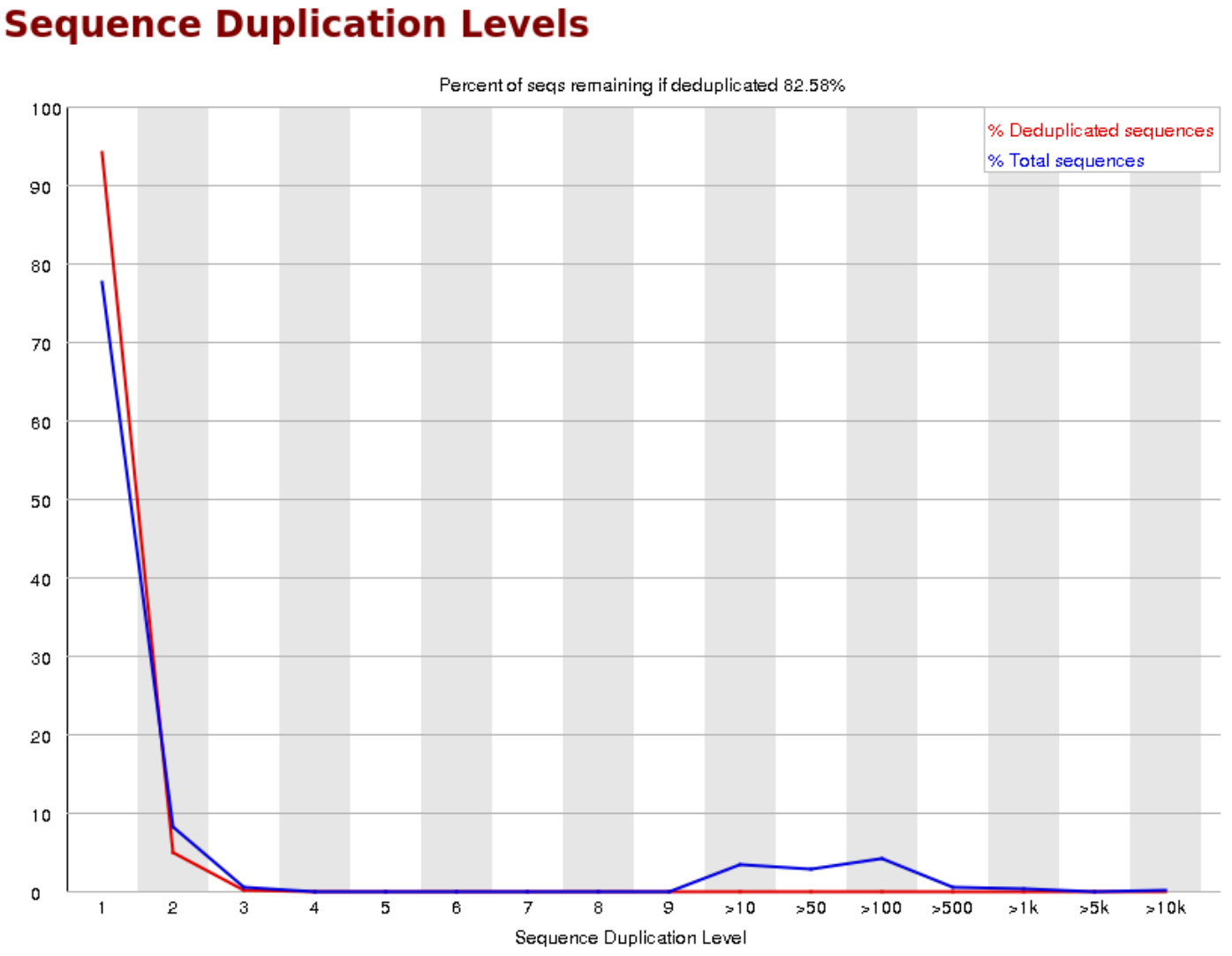
Complexity



# FastQC

Complexity

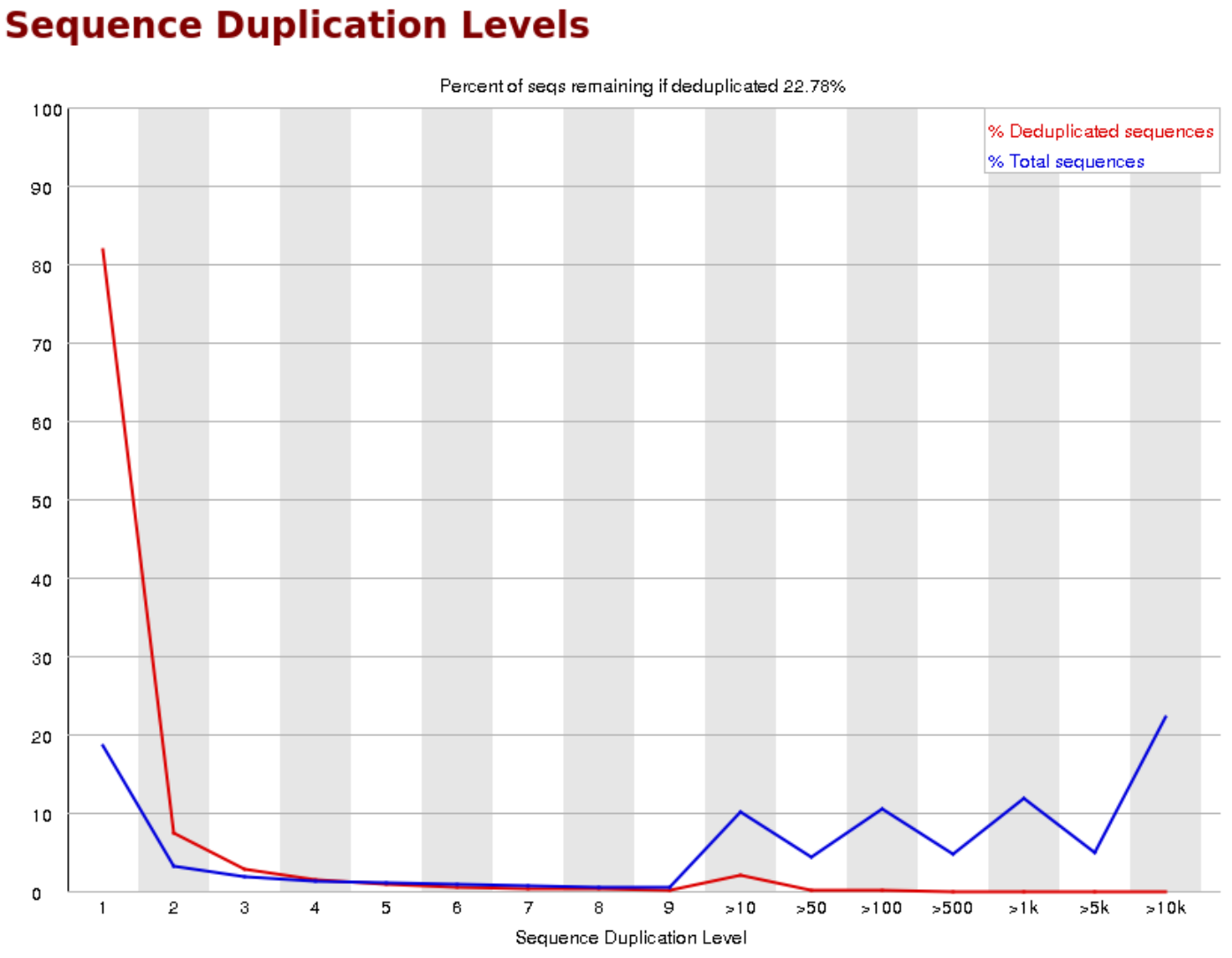
Duplication



# FastQC

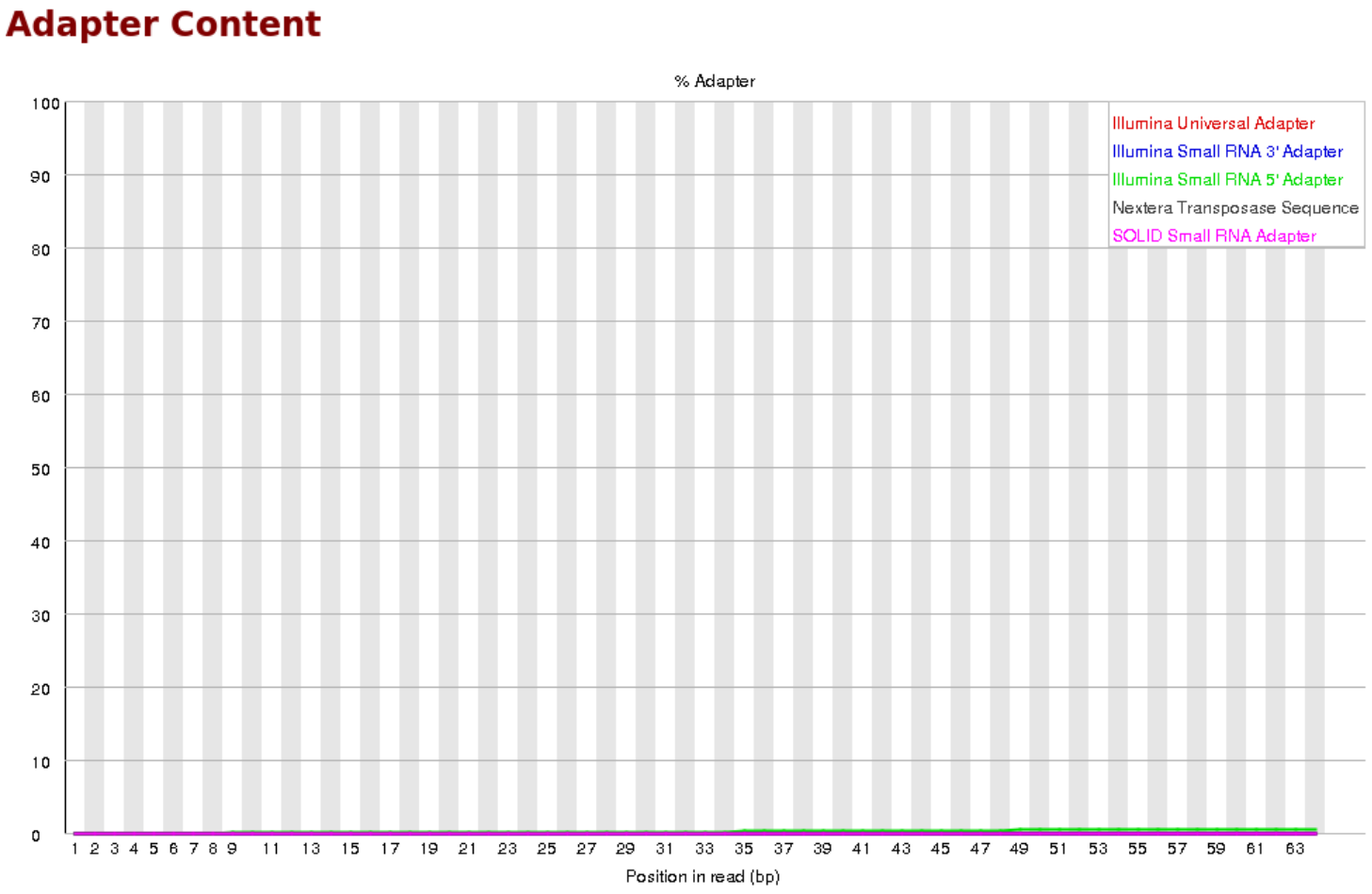
Complexity

Duplication



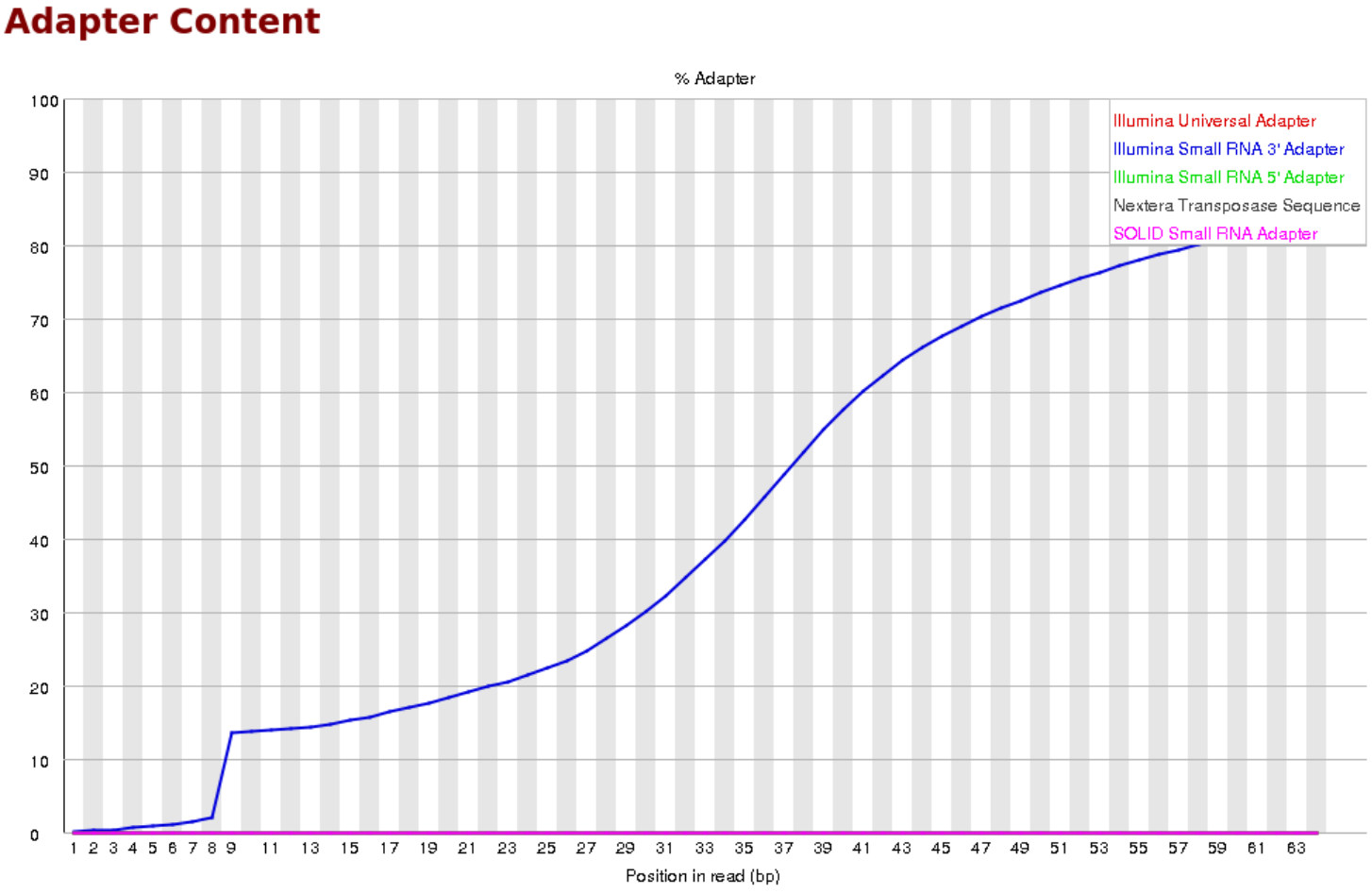
# FastQC

Adapter Contamination



# FastQC

Adapter Contamination



# VIM and vimtutor

- What is VIM?
  - Text editor – read, write and save text files
  - Entirely keyboard-based
  - You cannot use your mouse to move the cursor!
- vimtutor is on every linux system and teaches you how to use vim – open it now

```
~$ vimtutor
```

# Variables in Bash (see worksheet)

- On the website (https://biodatasci.colorado.edu/shortread/sr2023/) under Day 1 (or from the GitHub repo), open and follow the worksheet named **Creating a variable in bash**

- If you need help, flag us with a red sticky note

- If you've cloned the repo and finished the worksheet, put up a green sticky note