

Bioinformatics for Disease

Table of contents

Preface	3
1 Molecular Sequencing: Concepts and Technologies	4
I Data skills	5
2 Common data visualization techniques	7
II DNA	8
3 Aligning DNA sequences	9
4 DNA quality control	10
5 Variant calling	11
III RNA	12
6 Aligning RNA sequences	13
7 RNA quality control	14
8 Differential expression analysis	15
9 Gene set enrichment analysis (GSEA)	16
IV Infectious Disease	17
V Heritable Disease	18

VI Cancer	19
10 Cancer drivers	20
11 Driver enrichment	21
12 Driver enrichment	22
VII Appendices	23
13 Reference versions	25
13.1 Downloading genome reference files	25
13.2 Additional Resources	25
14 Useful databases	26
14.1 Downloading genome reference files	26
14.2 Additional Resources	26
References	27

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

1 Molecular Sequencing: Concepts and Technologies

Part I

Data skills

In summary, this book has no content whatsoever.

1 + 1

[1] 2

2 Common data visualization techniques

Part II

DNA

3 Aligning DNA sequences

4 DNA quality control

Here I'm listing some useful reference material.

Chapter [14](#)

5 Variant calling

Part III

RNA

6 Aligning RNA sequences

7 RNA quality control

8 Differential expression analysis

9 Gene set enrichment analysis (GSEA)

Part IV

Infectious Disease

Part V

Heritable Disease

Part VI

Cancer

10 Cancer drivers

11 Driver enrichment

12 Driver enrichment

Part VII

Appendices

Here I'm listing some useful reference material.

Chapter [14](#)

13 Reference versions

13.1 Downloading genome reference files

Species	Version	Command
human	GRCh38/hg38	<code>wget https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz</code>
human	GRCh37/hg19	<code>wget https://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/hg19.fa.gz</code>
mouse	GRCm39/mm39	<code>wget https://hgdownload.soe.ucsc.edu/goldenPath/mm39/bigZips/mm39.fa.gz</code>
mouse	GRCm38/mm10	<code>wget https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz</code>

13.2 Additional Resources

1. [Broad Institute](#)

14 Useful databases

14.1 Downloading genome reference files

Species	Version	Command
human	GRCh38/hg38	wget https://hgdownload.soe.ucsc.edu/goldenP
human	GRCh37/hg19	wget https://hgdownload.soe.ucsc.edu/goldenP
mouse	GRCm39/mm39	wget https://hgdownload.soe.ucsc.edu/goldenP
mouse	GRCm38/mm10	wget https://hgdownload.soe.ucsc.edu/goldenP

14.2 Additional Resources

1. [Broad Institute](#)

References