

# **Bioinformatics for Disease**

Andrew Lynch

# Table of contents

<b>Preface</b>	<b>3</b>
<b>1 Molecular Sequencing Concepts and Technologies</b>	<b>4</b>
<b>I Data skills</b>	<b>5</b>
<b>2 Common data visualization techniques</b>	<b>7</b>
<b>II DNA</b>	<b>8</b>
<b>3 Aligning DNA sequences</b>	<b>9</b>
<b>4 DNA quality control</b>	<b>10</b>
<b>5 Variant calling</b>	<b>11</b>
<b>III RNA</b>	<b>12</b>
<b>6 Aligning RNA sequences</b>	<b>13</b>
<b>7 RNA quality control</b>	<b>14</b>
<b>8 Differential expression analysis</b>	<b>15</b>
<b>9 Gene set enrichment analysis (GSEA)</b>	<b>16</b>
<b>IV Infectious Disease</b>	<b>17</b>
<b>V Cancer</b>	<b>18</b>
<b>10 Cancer drivers</b>	<b>19</b>
<b>11 Driver enrichment</b>	<b>20</b>

<b>12 Driver enrichment</b>	<b>21</b>
<b>VI Appendices</b>	<b>22</b>
<b>13 Reference versions</b>	<b>24</b>
13.1 Downloading genome reference files . . . . .	24
13.2 Additional Resources . . . . .	24
<b>14 Useful databases</b>	<b>25</b>
14.1 Downloading genome reference files . . . . .	25
14.2 Additional Resources . . . . .	25
<b>References</b>	<b>26</b>

# 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**

# **Cancer**

## **10 Cancer drivers**

## **11 Driver enrichment**

## **12 Driver enrichment**

# **Part VI**

## **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	wget https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz
human	GRCh37/hg19	wget https://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/hg19.fa.gz
mouse	GRCm39/mm39	wget https://hgdownload.soe.ucsc.edu/goldenPath/mm39/bigZips/mm39.fa.gz
mouse	GRCm38/mm10	wget https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz

## 13.2 Additional Resources

1. [Broad Institute](#)

# 14 Useful databases

## 14.1 Downloading genome reference files

Species	Version	Command
human	GRCh38/hg38	wget <a href="https://hgdownload.soe.ucsc.edu/goldenP">https://hgdownload.soe.ucsc.edu/goldenP</a>
human	GRCh37/hg19	wget <a href="https://hgdownload.soe.ucsc.edu/goldenP">https://hgdownload.soe.ucsc.edu/goldenP</a>
mouse	GRCm39/mm39	wget <a href="https://hgdownload.soe.ucsc.edu/goldenP">https://hgdownload.soe.ucsc.edu/goldenP</a>
mouse	GRCm38/mm10	wget <a href="https://hgdownload.soe.ucsc.edu/goldenP">https://hgdownload.soe.ucsc.edu/goldenP</a>

## 14.2 Additional Resources

1. [Broad Institute](#)

## References