

生命科学データ解析をはじめの前に読む本

ゼロから手を動かして学ぶ!

Linuxの基本から解析環境の構築、スパコンの利用方法まで

目次

● はじめに	3
● 本書の使い方	10

第0章

『生命科学データ解析をはじめの前に読む本』を読む前に 14

1 本書のねらい	14
2 学習の進め方	16
3 本書で一貫して大切にしてほしいこと	19

Part I Colabで体験するデータ解析の世界

とにかく真似して動かす

第1章

とりあえず体験してみよう 21

1 Colabに触れて基本的な使い方をマスターしよう	22
1 Colabとは? メリットと基本的な使い方	22
2 Colab上でPythonを実行する	28
3 Colab上でRを実行する	28
4 セッションとランタイム	30
2 データの読み込みから仮説検定までをRで体験しよう	31
1 R体験: 簡単なデータの読み込み, 集計	31
2 R体験: 生物統計検定の初歩	35
3 生命科学データ解析の航海に向けて	38
1 RとPythonの選び方・使い分けの考え方	38
2 生命科学データ解析をするうえで必要な知識	40
Column なぜ生命科学データ解析でCSV/TSVが好まれExcelが使われないのか	42
Column Colabのターミナルからプログラムを実行する	43

第2章

ファイルや文字列を扱う基本コマンドを覚えよう

45

1 Linuxの基本知識とコマンドの使い方を知ろう	46
1 Linuxの簡単な歴史と用語の解説	46
2 Linuxコマンドを使う, その前に	48
3 Linuxコマンド入門 その1	49
4 引数とオプション	53
2 ディレクトリやファイルを自在に扱えるようになろう	58
1 Linuxコマンド入門 その2	58
2 絶対パスと相対パス	60
3 Linuxコマンド入門 その3	62
4 Linuxコマンド入門 その4	64
3 Linuxの機能とコマンドを組み合わせよう	70
1 出力のリダイレクト	70
2 プレース展開で効率的に文字列を生成する	73
実践演習 アミノ酸配列FASTAファイルを使ったミニ解析演習	74

第3章

「動かない!」を乗り越える問題解決術

79

1 エラーメッセージとの向き合い方を知ろう	80
1 エラーメッセージの読み解き方	81
2 原因を突き止め修正する	83
2 エラーの解決策を探せるようになろう	85
1 効果的な検索テクニック	85
2 生成AIの活用	86
3 信頼できる情報源を見極める	87
4 質問力を磨く	90

第4章

生命科学データ解析を体験しよう

91

1 ヒトY染色体上のPAM配列を探せ!	92
1 公共データベースからデータを取得	93
2 チェックサムでデータの同一性を確認	95

3 圧縮ファイルの扱い方	97
4 FASTA形式の中身を見る、配列を検索し数える	98
2 coding領域にあるSNPを探せ!	103
1 GTFとBED: ゲノムアノテーションのファイル形式	104
2 TSV, CSVの加工コマンド	106
3 BEDファイルの取得	109
4 2つのBEDファイルの共通項を見つける	111
Column ゲノム情報の「住所」を正しく扱うために	116

第5章

面倒な作業を自動化させよう

119

1 一連のコマンドをまとめた解析レシピを作ろう	120
1 研究のためのシェルスクリプト超速習 その1	121
2 Shebangの概念と実行権限の付与	122
2 より柔軟なスクリプトのための要素を学ぼう	126
1 研究のためのシェルスクリプト超速習 その2	126
2 研究のためのシェルスクリプト超速習 その3	131
実践演習 生命科学データの自動ダウンロードと解析	137
Column 生命科学データ解析に役立つ便利なその他の応用コマンド	143

Part II ローカル環境での実践

自分のデータで試してみる

第6章

自分のPCの環境を整えよう

147

1 ターミナルを準備して開発環境を整えよう	148
1 ローカル環境の必要性	148
2 ターミナルの準備	149
3 ファイルシステムの探索と操作	153
4 Visual Studio Codeのインストール	154
2 パスを通して自分のコマンドやツールを管理しよう	156
1 パスを通す	156
2 シンボリックリンク	163
Column とてもありがたいパッケージマネージャー	166

第7章 環境を誰かと共有しよう

169

- 1 Condaをインストールして仮想環境を構築しよう 170
 - 1 仮想化は特別なことじゃない 170
 - 2 パッケージ管理と環境構築を始める前に 171
 - 3 Condaのインストール 172
 - 4 仮想環境の基本操作 174

 - 2 仮想環境による解析環境を共有しよう 180
 - 1 仮想環境の書き出し 180
 - 2 仮想環境の再構築 181
- Column** conda install と pip install, 混ぜるな危険? 184

第8章 他人が作ったプログラムを再利用しよう

187

- 1 VS Codeを導入して快適なローカル環境を整えよう 188
 - 1 VS Codeの導入と基本設定 188

 - 2 GitHubを使って他人のプログラムを再利用しよう 192
 - 1 GitHubを使ったコードの取得と環境構築 192
 - 2 プログラムを通じてWebサービスと対話するしくみ API 200
- Column** AIとともにコードを書く時代のこれから 203

Part III 未来への羅針盤

今後の研究に活かす

第9章 スパコン・計算サーバーへの接続

205

- 1 スパコン利用までの流れを知ろう 206
 - 1 なぜスパコン/計算サーバーを使うのか? 206
 - 2 スパコンの利用申請から最初のログインまで 207

2 スパコンの使い方を身につけよう	214
1 ログインノードと計算ノードの違い, 基本的なマナー	214
2 モジュールの読み込み	215
3 ログアウトとファイル転送	217
4 スパコンを使った生命科学データの解析	219
Column スパコンと仲良くなろう	224

第10章

そして次のステップへ：広がるデータ解析の世界

227

1 これまでの学習から広がるデータ解析の航海に出よう	228
1 これから挑むバイオインフォマティクス分野の俯瞰	228
2 生命科学における機械学習の俯瞰	233
2 さらに学習を深めるためのリソース	237
Column 生成AIが進歩した後も残り続ける人間の技術はなんだろう？	239
Column 生命科学データ解析は今後どうなっていくのだろう？	240

● 付録 Linuxコマンドの働きとUsage一覧表	242
● おわりに	247
● 索引	249