

RNA-seq 入門

NGS の基礎から de novo 解析まで 準備編・実践編

生物情報学を必ずしも専門としない生物研究者が、RNA-seqで得られる網羅的な遺伝子発現データを解析し、生物学的な情報を抽出するための、基礎的技術と考え方を身に付けることを目的としたコース。

基生研・ゲノムインフォマティクス トレーニングコースの理念

- For biologists
- 「基礎力」と「即戦力」をバランスよく習得

RNA-seq 入門

準備編（2月22～23日）

- UNIX基本コマンド
- エディタとスクリプト
- R入門
- NGS基本データフォーマット
- クオリティコントロールとNGS基本ツール
- UNIXによるテキストファイル処理
- 生物情報解析システムの紹介
- 演習

実践編（3月8～9日）

- NGS 基本フォーマット・ツール復習
- NGS データ可視化ツール
- 統計学入門
- RNA-seq パイプライン：基礎
- RNA-seq パイプライン：transcript ベース
- RNA-seq パイプライン：genome ベース
- RNA-seq パイプライン：de novo
- 多変量解析
- 機能アノテーションとGO解析
- 演習

「準備編」の位置づけ

- RNA-seq 等NGS解析技術習得の「準備編」
UNIX, R, NGSの基本
- 日々ビッグデータ化する近年の生物学研究全般を進める
上での必須スキルの習得



Goal of 準備編

- インフォマティクスに対する心的障壁を取り除く
- ゲノムインフォマティクスの基礎的技術と考え方を身に付ける
 - UNIXコマンドラインの操作や環境に慣れる
 - タブ区切りテキストを処理する程度の簡単なプログラミングを学ぶきっかけをつかむ
- 独習するための基盤を身に付ける
 - 今後独習する為に必要な基礎的なスキル
 - 今後何を学べば良いかの指針を得る
- インフォマティクス専門家と対話できる程度の基礎知識を身に付ける

Biologist が身に付けるべき 6つのinformatics スキル

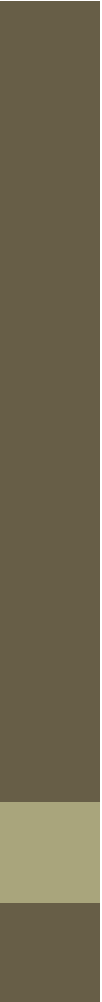
- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース



今回みっちり学びます

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース

Why Learning UNIX?

- 最先端の学術ソフトウェアはUNIX用に開発
 - 大量データ解析に対応
 - 高い信頼性、安定性、汎用性
- 

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース

UNIX + 初歩的でプログラミングで一気に広がる解析の幅


- 解析の自動化
- テキストデータの処理
 - 特定のプログラムのinputファイルの準備
 - outputデータの加工
 - 塩基配列やアミノ酸配列自体が文字列である
- 解析パイプラインの構築（本格的な解析プログラムまで開発可能）



準備編でRを学びます
実践編で統計基礎を勉強します

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース

統計的な考え方と技術

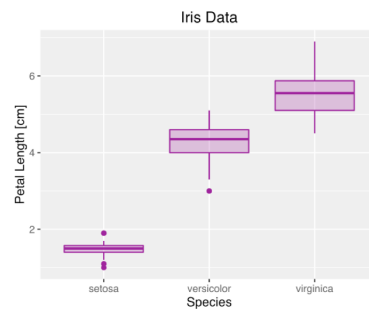
- 大規模な-omics データは統計的な考え方なしでは適切に扱えない。
- 統計処理やデータ可視化、データマイニングの道具に習熟する。ーRが最適
- 実践編では統計的な考え方の基礎から勉強します。
- データ解析の際に有用なだけでなく、実験デザインを組む際にも不可欠です。

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術 
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール 
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース 

“Graphs force us to note the unexpected; nothing could be important.”

-- by John Tukey (1915-2000)



Exploratory Data Analysis (EDA)

- 初級) UNIXの基礎
- 初級) 統計的な考え方と技術
- 初級) 業界標準のツール
- 初級) データ可視化
- 中級) 初歩的なプログラミング
- 中級) データベース

- バイオデータベースを使う
- データベースを作成・操作する

オススメ勉強法

- コマンドやプログラムは自分で試してみる。copy & pasteでなくタイピングすること。（熊楠メソッド）
- 気軽に質問する。講師はもちろん、隣や前後の受講生にも。その一方で、ヘルプやマニュアルドキュメントをうまく活用する。
- 自分の研究との接点を常に意識する。自分の研究に応用する。



南方熊楠 (1867-1941)

RNA-seq 入門

準備編（2月22～23日）

- UNIX基本コマンド
- エディタとスクリプト
- R入門
- NGS基本データフォーマット
- クオリティコントロールとNGS基本ツール
- UNIXによるテキストファイル処理
- 生物情報解析システムの紹介
- 演習

実践編（3月8～9日）

- NGS 基本フォーマット・ツール復習
- NGS データ可視化ツール
- 統計学入門
- RNA-seq パイプライン：基礎
- RNA-seq パイプライン：transcript ベース
- RNA-seq パイプライン：genome ベース
- RNA-seq パイプライン：de novo
- 多変量解析
- 機能アノテーションとGO解析
- 演習

