

改訂の序（本書の改訂方針）

2019年に上梓した「RNA-Seq データ解析 WET ラボのための鉄板レシピ」(通称 RNA-Seq レシピ本)は、お披露目の場となったその年の日本分子生物学会年会で1番の売れ行きとなり、多数の研究室において座右の書として活用していただいている。その翌年の2020年4月から発足したわが研究室(bonohulab)においても、公共データベースから RNA-Seq データをメタ解析する研究に取り組む第一歩として、この RNA-Seq レシピ本が大活躍している。

光陰矢のごとしで、その RNA-Seq レシピ本も出版から3年以上が経った。Annual updateによる更新情報により出版後のフォローもあったものの、根本的に見直そうという話になった。それは、2020年代に入り、RNA-Seq 解析はシングルセル解析方面だけでなく、一般的なモデル生物以外の多様な非モデル生物種（有用物質をつくる植物や昆虫、魚類、菌類など）の配列解析にも広く用いられるようになってきたからである。それらの生物を対象としたデータ解析を念頭に改訂を行った。

まずは、データ解析環境として旧版では Mac を中心に解説していたが、現在 bonohulab では Docker や Windows (WSL2) 上で普通に実行している状況をかんがみ、それらの解析環境整備に関して陥りがちなポイントを中心に解説した (Chapter 1)。しかしながら RNA-Seq データ解析のコアな技術は 2010 年代に出来上がっているため、旧版のよいところはそのままに情報の更新を行った (Chapter 3 ~ 6, 9)。

また、複数のサンプルに対する発現値データを解釈する手法である階層クラスタリングと主成分分析 (PCA) に関して、その実際的な手順を詳細に記述した (Chapter 7)。

さらに、有用なデータが多数収められている公共データベースから必要なデータを取得し、それらをメタ解析するという、bonohulab で行ってきた手法に関してその解析プロトコールを手前味噌ながら詳細に紹介した (Chapter 8)。

最後に、論文投稿の際に求められることが一般的になつたデータ登録&公開に関しても、新規モデル生物種で必要となる DDBJ の Transcriptome Shotgun Assembly (TSA) に登録するやり方について詳細にアップデートした (Chapter 10)。

本書は、RNA-Seq データ解析の一連の流れ（ワークフロー）に沿つて章立てがなされているものの、それぞれの Chapter は独立している。なので、気になったところから読み進めていってほしい。料理と同じく、本書をベースに自らのアレンジを加えて、読者の皆さまが自身の「レシピ」を確立していくことを祈っています。

2023年7月 梅雨明けの酒都西条にて
坊農秀雅

本書のX(旧Twitter) のハッシュタグは
#RNaseqRecipe

序(本書の編集方針)

RNA-Seq は高値の花でなくなった。1サンプル数万円の現実的な価格となり、その結果、1ファイルあたり数 GB の塩基配列情報が簡単に得られるようになっている。しかし、塩基配列情報をどういったツールを使って定量し、発現量としたらよいのか？ なんとか定量できたとしても、それをどう料理して図や表として可視化すればよいのか？ 論文発表する際の味付けはどうしたらよいのか？ 何かいいスパイスはないものか？ 手法が新しいだけに確立した方法に乏しく、悩みは尽きない。確かに、インターネット検索すればいろいろと情報は出てくる。しかしながら、そのうちのどれを信じて使ったらよいのかよくわからないのが正直なところだろう。生命科学データ解析業界は改廃が激しく、現時点で検索して探し当てたレシピでちゃんと動くのだろうか？ また、なんとか動いたとしてもそれが一番よく使われている解析方法なのか、もっと良い料理方法があるのではないか？ そこで、そのような悩みに応えるために RNA-Seq データ解析のレシピとなる本を作ろうと画策した。

本書では、次世代シーケンサー (NGS) からの配列解析全般を解説する類書とは一線を画し、多くの研究で頻繁に用いられるようになった配列解読による発現定量手法である RNA-Seq だけに的を絞った。インターネット上に多くのコンテンツを発信している第一線のバイオインフォマティクス研究者のみならず業界関係者にお願いし、コマンドラインの定番ツールはもちろん、有用なウェブツールや商用アプリケーションにいたるまで紹介した。それに加えて、データ解析の前段階として、データ解析環境を準備したり、配列解読を外注したり、公共データベースからデータ入手するところも省略せずに、RNA-Seq に関わるさまざまなステージで問題となるであろう部分について解説している。データ解析が終わって論文投稿する際に必要な Sequence Read Archive (SRA) へのデータ登録に関しても、実際に登録受付業務を行なっている国立遺伝学研究所 DNA Data Bank of Japan (DDBJ) の担当者に執筆いただいている。例えるなら、RNA-Seq や周辺技術のエキスパートたちに最新のデータ料理方法や調理器具、果てはオリジナルスパイスにいたるまで、深く解説してもらったということになるのだろう。

本書をベースにアレンジを加えて、皆さまが自分なりの RNA-Seq データ解析ワークフローを築き上げ、RNA-Seq データ解析の達人になられることを心待ちにしています。

2019年11月 ようやく秋らしくなってきた三島にて
坊農秀雅

本書の Twitter のハッシュタグは
#RNAseqRecipe