

序 (本書の編集方針について)

2020年現在、ヒトゲノム1人当たりの解読コストが千ドル(約11万円)を切り、ヒト個人のゲノム配列が大量に解読される時代になった。医学にかかわる研究者としては、それらの配列情報が収められた公共データベース(DB)にアクセスして、個人ゲノム間の「違い」、すなわちバリエーション情報を自らの研究に役立てたいのが筋だろう。

しかしながら、バリエーション情報を参照するためにはたくさんのハードルが待ち構えている。まずは、個人情報保護の問題で生の配列データは直接入手できない。それらを手に入れるためにはヒトに関するデータの利用手続きが一般的に必要なのだ。また、仮に配列データをもっていたとしてもそれらをデータ解析しないといけない。どういったプログラムで、そしてどのコンピューターでやればいいのか。それらすべてを1人の研究者でやろうとすると、かなりの手間がかかる。そして、そのような煩雑さがゆえに公共DBにあるそれらのデータを利用することを諦めてしまうかもしれない。

そこで、達人たちにヒトバリエーション情報の利用手段を解説してもらって、それらをレシピにすることを考えた。そのやり方は、『RNA-Seq データ解析 WET ラボのための鉄板レシピ』(坊農秀雅/編, 羊土社, 2019)として出版したRNA-Seq データ解析のレシピ本を踏襲している。生配列データを直接入手できなくても、データ解析済みの個人情報紐づかない「丸めた」データであれば、実は誰でもいつでも利用可能なのである。そのようなデータ利用に関する現状把握を含めて、まずは世界で使われているバリエーションデータの利用手段をChapter 1で紹介する。そして続くChapter 2では、日本人のバリエーションデータを検索する手段を、その開発や維持にかかわっている現場の研究者に解説してもらった。最後にChapter 3では、がん細胞でみられる体細胞変異のバリエーションのデータベースを検索する手段を若手のがん生物学研究者に紹介してもらった。

公共DBからバリエーション情報を検索する際の基本レシピ集として本書を活用することで、自分なりのカスタマイズされたデータ解析手法をみ出し、読者の皆さまがバリエーション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。

2020年9月 鏡山のふもと、酒都西条にて
坊農秀雅

本書の Twitter のハッシュタグは
[#VariantRecipe](#)