データサイエンスのテクノロジーで生命現象を読み解く
Biological Big Data to Knowledge, using Data Science

データサイエンスを用いた生命情報解析

次世代シーケンサおよび質量分析機から出力される計測データをハイスループットに解析する情報科学的手法の開発を行っています。近年、計測技術の発展により、生物学において算出される電子データは増加の一途をたどっており、大量の生物学データを標準的な方法で処理することがすでに困難な課題となっています。加えて、異なる次元のデータを統合し、従来モデル化が難しいデータに対しても関連性を見出すためには、ビッグデータ解析技術や機械学習の最新の成果（データサイエンス）を取り入れて情報解析を行うことが不可欠になっています。また、大量のゲノムデータの中から生物学的な意味や関連性を見出すには大規模にデータを集約させ、分散処理を行う必要があります。将来的なクラウド運用を見据えて、Hadoop/Sparkといったクラウドで標準的な分散基盤や深層学習のライブラリを用いた生命情報の解析基盤を開発しています。

研究領域の紹介

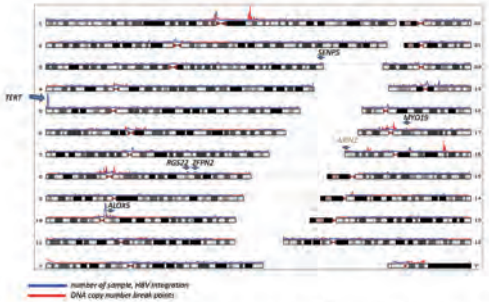
次世代シーケンサおよび質量分析機の応用範囲は多岐に渡りますが、以下のような領域で研究を行い、同時にソフトウェアを開発しています。

(1) がんゲノミクス
(2) タンパク質の転写後修飾の解析
(3) エピトランスクリプトーム(RNA修飾)解析

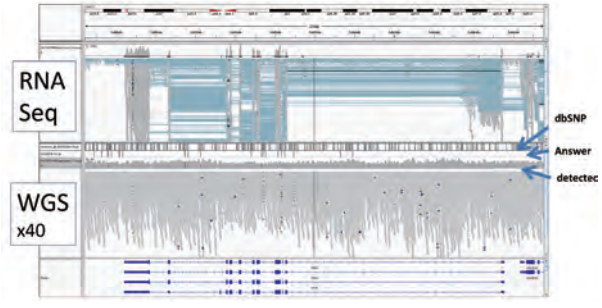
Biological Data Science

With the development of sequencing technology, electronic data yields in biology have been steadily increasing, and it is already a challenging task to process large volumes of data with conventional methods. In addition, in order to extract knowledge from multi modal big data, (ex. Multi-omics data) it is necessary to incorporate the latest Data Science technology, such as cloud computing and machine learning. We are developing cloud based NGS analysis pipeline using Hadoop / Spark, popular cloud computing framework, and deep learning library.
Our research include following:

(1) Cancer genomics
(2) Proteomics and post translational modification
(3) epitranscriptome (RNA modifications) analysis



1 B型肝炎ウィルスの挿入部位（青）とコピー数変異部位（赤）のゲノム位置
Hepatitis B Virus (HBV) integration sites (blue) and DNA copy number break points (red) on human genome



2 Hadoopを用いたRNAシーケンスおよび全ゲノムシーケンス解析結果
RNA Sequencing and Whole genome sequencing using Hadoop

講師
上田 宏生
Hiroki UEDA, Lecturer
専門分野：情報生命科学、がんゲノミクス、機械学習
Specialized field：Computational Biology, Cancer Genomics, Machine Learning
E-mail：ueda@genome.rcast.u-tokyo.ac.jp