

## 数万検体規模のトランスクリプトーム解析を達成する情報解析基盤

国立がん研究センター 白石友一

シーケンス技術の革新により、NCBI, EBI, DDBJ などの公共レポジトリに膨大な量のゲノム・トランスクリプトームシーケンスデータの蓄積が加速度的に進んでいる。また、コンテナ仮想化技術、クラウド計算技術の革新により、大規模なデータを効率的に再現可能な形で解析、またデータの効果的な共有・管理を行う技術についても大きな革新が続いている。さらに recount2 や ARCHS4 など、発現量などの基本統計量を一次処理し、簡便に利用できるデータベースも数多く到来してきており、大規模なシーケンスデータ解析のための環境はこれまでになく整備されつつある。

本発表では、トランスクリプトーム、特にスプライシング異常という切り口から、我々が開発している種々の統計解析手法解析、情報解析基盤について紹介する。さらに数万検体規模のトランスクリプトームデータの解析を通じて獲得された生物・医学的な知見の概説を行う。