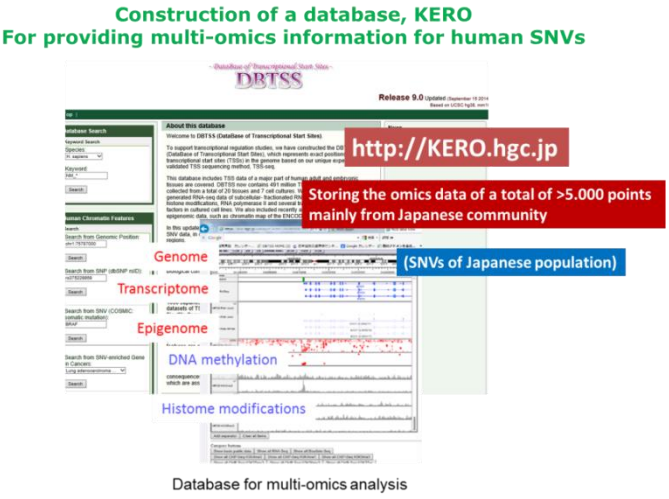
**成果報告④** **10月5日17:20—17:40**

# 疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オミクスデータの統合

菅野　純夫　（東京大学大学院新領域創成科学研究科　教授）

我々の研究グループでは、KERO-DB(Kashiwa Encyclopedia of Regulatory Omics Database)を開発し、ヒトのマルチオミクスデータの統合を試みている。本データベースは、特に、疾患に関連したヒトゲノム多型・変異に生物学的機能注釈を与えることを目的に、様々な目的で産出された、ゲノム変異位置、近傍におけるエピゲノム（ヒストン修飾、DNA メチル化パターン)、トランスクリプトーム情報（発現量、スプライスパターン）をヒトゲノム情報に統合している。データの統合は、現在、がんゲノム解析を志向したものを中心に行っているが、最終的には疾患の別を超えたデータの統合を目指す。現在は、ヒト培養細胞系あるいはマウスをはじめとする動物モデル系から得られたオミクスデータを中心にデータベースに収載している。これにより臨床検体で集積が乏しいエピゲノム情報を充実させ、同時に生物学的機能解析の実践の場としてのモデル系におけるオミクス情報を整備することを目指している。将来的にはヒト臨床検体から直接得られた多層オミクスデータを直接関連付ける予定である。



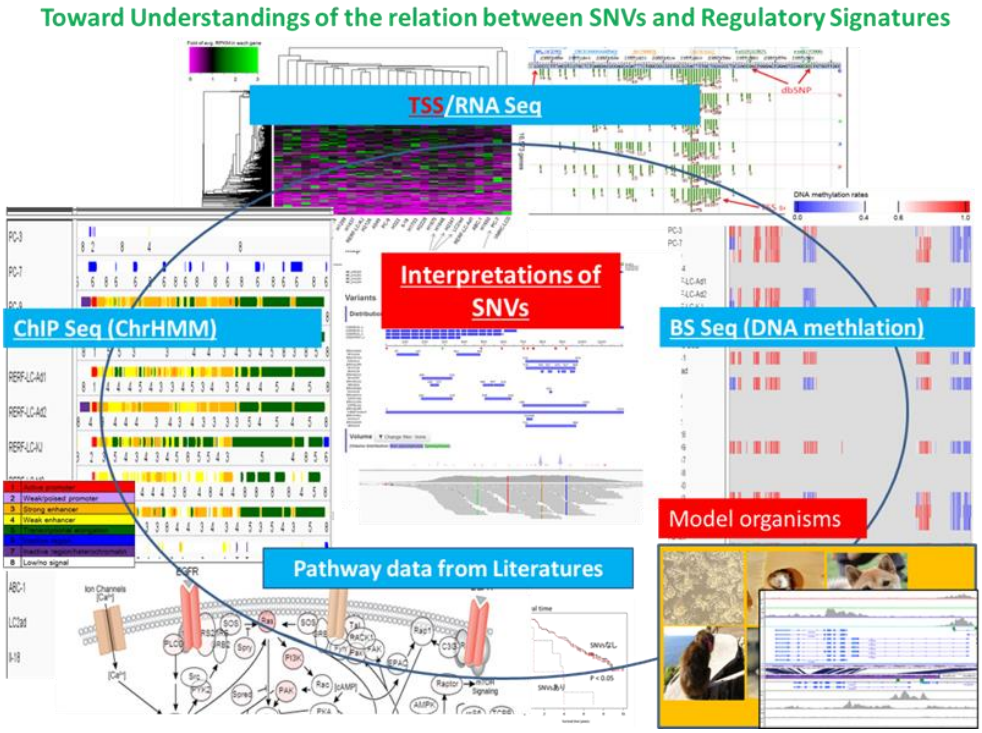
Database for multi-omics analysis

現在、公開されているデータの中でも、特に肺腺癌培養細胞の多層オミクスデータに特徴を持つ。同一のがん種に由来する培養細胞２６種について、全ゲノム配列、抑制性および活性性のヒストン修飾６種類、Methyl C Seq による DNA メチル化データ、転写開始点データ、トランスクリプトームデータ、miRNA データが体系的に記載されたものであり、各細胞のがんドライバーパターンについて、そのエピゲノム、トランスクリプトームの状態を横断的に検索することが可能なデータとなっている。他にも、マウス免疫担当細胞を種々の刺激源で刺激した詳細な経時計測系において、多階層オミクスデータを収集、データベースとして収載している（転写開始点デー

タ、ヒストン修飾等について計 400 計測点)。がん研究においても、がん免疫療法の急速な展開から、免疫担当細胞の多層オミクスの理解は大きな注目を集める分野となっている。本データベースでは、さらに多くのモデル生物種についての多層オミクスデータを文部科学省新学術領域「ゲノム支援」を中心としたゲノム解析支援活動を通じて収集している。現在、これらをモデルデータに、ヒト細胞系へのデータの写像を行ったデータベースの構築を行うべくデータを成形し、データベースとしての公開を逐次開始している段階である。現在までにマウスを中心に 10 系列のデータを公開しており、次年度にさらに拡張を行う予定である。

また、一方でゲノムデータについては、本データから参照する形で東北メディカルメガバンク、長浜コホート、日本ファーマコゲノミクスコンソーシアムにより産出された日本人多型データ約 5000 人分が無償、自由に閲覧可能となっている。

さらに、上記のモデルデータについての RDF 化によるデータ統合プラットフォームの整備を進めている。昨年度から着手した整備計画をさらに発展させる形で、格納されたデータについて、統合データベースの観点からの検索可能性を考慮した RDF 化によるデータフォーマット、検索の統一化を図っている。




参考文献

Suzuki A, Wakaguri H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, Suzuki Y, Nakai K. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. Nucleic Acids Res. 2015 (Database issue) 43 (D1): D1-D5.