

課題名 (タイトル) :

次世代シーケンスデータの大規模解析

利用者氏名 : 沼田 興治

理研での所属研究室名 : 筑波研究所 バイオリソースセンター バイオリソース関連研究開発プログラム
動物変異動態解析技術開発チーム

報告内容

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

細胞の遺伝情報を解析する DNA シーケンサーはここ数年でデータのスループットが飛躍的に増大した。一般に次世代シーケンサー、高速シーケンサーなどと呼ばれ、原理や機種によって異なるが、一度の解析で数十~数百億塩基のデータが排出され、これらのデータを全ゲノムワイドで解析するためには膨大な計算コストをとらなければならない。本研究課題では、主にマウスの生殖細胞におけるゲノムの化学修飾状態(シトシンのメチル化や、ヒストンの修飾)や、遺伝子の発現状態の変化を明らかにするために高速シーケンス解析をおこない、そのデータマイニングに RICC をもちいる。

2. 具体的な利用内容、計算方法

本課題でもちいるデータは、マウスゲノムの領域毎の遺伝子発現量やゲノム修飾状態を反映した定量的な情報である。細胞種の違いによる遺伝子発現量やゲノム修飾状態の変化のパターンを、ゲノム領域毎に追跡していくことを目的としている。そのために、特定のゲノム領域(数百万から数千万の領域)に対応するデータを高速に抽出する処理を RICC によっておこなう。

3. 今後の計画・展望

現在、複数の細胞種をもちいたデータの取得と、得られたデータの予備的な解析をおこなっている。この結果をふまえて次年度以降に RICC を利用した高速な解析をおこなう予定である。