

課題名 (タイトル) :

## ExRAT 並列化

利用者氏名 : ○角田 達彦、森園 隆、中村 英二

理研での所属研究室名 :

横浜研究所 ゲノム医科学研究センター 統計解析・技術開発グループ 情報解析研究チーム

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  
現在の医学を推進するゲノム解析の最先端の方法である、ゲノムワイド関連解析を当センターは行っており、未知の疾患関連遺伝子の発見に続々と成功しているが、さらに複数の遺伝子が疾患へのリスクを上昇させることが大変重要であると考えられる。それらの遺伝子の組合せを探索する場合、全染色体上の 50 万マーカー間の組合せを、数千人から数万人規模で調べる必要があり、またマーカー間の関係を加味するために、データを大量にサンプリングする必要があるが、それらは膨大な計算量を必要とし、次世代計算機「京」での実行が期待される。そこでそのようなプログラムを開発し、並列性が得られることを検証するために、RICC を利用した。
2. 具体的な利用内容、計算方法  
ロジスティック回帰分析を、50 万×50 万組合せに行う方法と、importance サンプリングを用いて疑似 permutation test を行い、より厳密な p 値を求める方法との 2 種類を MPI 並列で実装し、シミュレーションで一たで計算を行った。
3. 結果  
両方のプログラムとも 8192 並列まで到達し、台数効果も得られることがわかった。また細かなチューニングを追加し、高速化の進捗が見られた。
4. まとめ  
次世代計算機「京」での実行への目安が得られ、RICC の能力の高さも実感できた。
5. 今後の計画・展望  
より台数効果を上昇させるために、綿密なプロファイリングと、さらなる並列性の向上を目指す。
6. RICC の継続利用を希望の場合は、これまで利用した状況 (どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか) や、継続して利用する際に行う具体的な内容  
これまでの並列化に加え、データ分散や仕様を細かく詰めることによって、さらなる速度向上を狙う予定である。
7. 一般利用で演算時間を使い切れなかった理由
8. 利用研究成果が無かった場合の理由  
上記の目的に使用したため、論文という形での成果ではないが、実装と並列度達成という成果が得られた。