（別添）　平成 27 年度調整費繰越分による実施内容

I. 成果の概要（総括研究報告）

本課題は、バイオバンク等に収集された DNA サンプル等を用いたゲノム解析を実施し、疾患関連遺伝子研究及び薬剤関連遺伝子研究のための基盤情報を提供する事を目的としている。

平成 27 年度調整費繰越分により、バイオバンク・ジャパンで収集された心筋梗塞 1,293 例の全ゲノムシークエンス解析を実施した。全ゲノムシークエンスデータの取得は、AMED の方針に従い民間企業に委託した。各実験プロセスにおいて、委託企業が作成したデータを理化学研究所が設定する品質管理基準を用いて評価を行った。その過程で、委託企業が生成するシークエンスデータの品質が低い事が明らかとなり、バリアントコールの精度を確認するため、同一のゲノム DNA を使用した大規模なターゲットシークエンスによるバリアント検証の準備を開始した。また、委託企業が、理化学研究所が設定した品質基準を満たすデータの生成に非常に時間を要したため、全ゲノムシークエンスデータの取得は、当初の予定（平成 28 年 9 月末納品）より大幅に遅延し、平成 29 年 3 月末に品質検査基準を満たしたデータの納品が完了した。

平成 27 年度調整費による遺伝性乳がん卵巣がん症候群関連遺伝子のターゲットシークエンス解析については、平成 27 年度に実施したターゲットシークエンスデータをもとに解析を進め、女性乳がん 7,051 名、女性対照群 11,241 名において、合計 1,781 個のバリアントを検出した。各バリアントの関連解析では、BRCA1/2 に存在する 3 つのレアバリアントがゲノムワイドに有意な関連を示し、かつオッズ比は 20 以上であった。これらのバリアントは、過去の乳がんの大規模 GWAS においても同定されておらず、シークエンス関連解析におけるレアバリアントの重要性を示した。

さらに、全ゲノムシークエンス解析のため、新たに計算機クラスタを導入し、平成 28 年 7 月より稼働開始した。この計算機クラスタを用いて、平成 27 年度に納入されたバイオバンク・ジャパンの心筋梗塞 500 例の全ゲノムシークエンスの粗データについて解析を実施した。

The aim of this project is to perform various genomic researches using DNA samples collected at Biobank Japan and provide basic information for susceptibility genes of disease and drug responses.

We carried out whole genome sequencing of 1,293 myocardial infarction cases collected at Biobank Japan. Acquisition of whole genome sequencing data was entrusted to company according to AMED policy. In each experiment process, RIKEN evaluated the company's data using the quality control criteria. During this process, it became clear that the quality of the sequence data generated by the company is low. Therefore, we started to prepare the validation analysis to confirm the accuracy of called variants by using the same genomic DNA. Moreover, since the company took long time to meet the quality control criteria, acquisition of whole genome sequencing data was largely delayed from the planned schedule. As a result, acquisition of whole genome sequencing data was completed at the end of March, 2017.

For target sequence study of hereditary breast cancer genes, we performed clinical annotation of called variants using the sequencing data obtained last fiscal year. We identified a total of 1,781 variants in 7,051 female breast cancer cases and 11,241 female controls. In the single variant association analysis, we identified three rare variants in BRCA1/2 that showed a genome-wide significant level of association ($P < 5x10E-08$) with odds ratios more than 20. These variants have not

been identified even in the large-scale GWAS of breast cancer, indicating the importance of rare variants in sequence-based genomic research.

In addition, we introduced a new computer cluster for whole genome sequencing analysis and started operation in July. Using this computer cluster, we analyzed whole genome sequence data of 500 myocardial infarction cases obtained last fiscal year.