

## 研究開発課題別事後評価結果

### 1. 研究開発課題名

大規模ゲノム疫学研究の統合情報基盤の構築

### 2. 代表研究者名

京都大学大学院医学研究科 附属ゲノム医学センター センター長・教授 松田文彦

### 3. 研究実施概要

本課題では、日本国内で実施されている多数の疫学(コホート)研究が独自に収集、解析されたデータの統合を可能にし、今後予防医学研究の主流となる多施設共同のゲノム疫学研究の情報基盤を構築することを目指し、ながはま0次コホート研究をモデルケースとして設計と実装を進めた。

#### 1) ゲノム疫学研究の統合基盤構築

複数拠点にわたる様々なゲノム疫学研究を統合するためには、データディクショナリの定義とその統合的な管理が重要であり、マスターサーバを中心とする分散系システムとして構築した。具体的な設計・開発にあたっては、京都大学が実施するながはま0次コホート研究をモデルケースに、データフローや匿名化、試料のロジスティック等の手順を洗い出し、一般化した後、運用パフォーマンスやセキュリティに配慮しつつシステムとして実装した。なお、オントロジーについては、既存オントロジーの調査と Resource Description Framework (RDF)化のための実装方法の検討にとどまり、実装には至っていない。

#### 2) データ公開

ゲノムブラウザ仕様のインタラクティブなウェブサイト「Human Genomic Variation DB」を独自に構築し、各SNPのジェノタイプ頻度やマイクロアレイとの関連解析(eQTL)の結果を公開した。0次コホートで得られた一部のデータは、「NBDCヒトデータベース」への提供手続きを進めている。

#### 3) データベースシステムの提供

本課題で開発されたデータベースシステムに、共同研究として実施するゲノムコホート研究情報を格納し、さらに複数の研究から得られるデータとの統合を進めるためデータの受け入れを実施中である。また、東北メディカル・メガバンクに対し、データベース全体のシステムパッケージを提供し、さらに、そのシステムの構築を支援した。

#### 4) 人材育成

平成 24 年度と 25 年度に、若手研究者を対象に、次世代シーケンサー解析トレーニングコース（3～4 日間）を実施し、さらに疾患ゲノム解析シンポジウムの実施を開催するなど、バイオインフォマティクス・統計遺伝学の若手人材の育成に務めた。

#### 5) PHR (Personal Health Record) の情報収集

データベースに蓄積された生命分子情報、環境・生活習慣情報に用いた PHR を展開するには、これらに加えて複数の医療機関に蓄積されている複数の疾患関連の臨床情報を迅速かつ効率よくデータベースに取り込む必要がある。まずは一医療機関から一疾患の情報を確実に抽出し、データベースへ登録するシンプルモデルとして、京都大学病院の医療情報の Electronic Health Record (EHR) を介して収集することを試みた。取得可能なデータは文字列で記載された情報が多く、データ登録用のウェブフォームに、その文字列情報から数値やカテゴリなどへの変換を支援する機能を開発した。

#### 4-1. 当初計画の達成度

本課題は、1) ゲノム疫学研究の統合基盤構築、3) データベースシステムの提供、4) 人材の育成、5) PHR の情報収集について、当初の計画通り達成している。但し、2) データ公開に関しては、「Human Genomic Variation DB」を構築・公開し、メタデータの公開はされたが、当初計画していた NBDC ヒトデータベースからのデータ公開に関しては、関係各所との調整中のため、データ提供が遅れており、予定されていた研究開発期間中の公開には至らなかった。統合化のための準備はできているが、データ統合化の実現まではまだ時間を要する状況にある。

#### 4-2. 研究開発成果の公開および利用の状況等

SNP のジェノタイプ頻度やマイクロアレイとの関連解析(eQTL)を公開するウェブブラウザタイプのウェブサイト「Human Genomic Variation DB」が平成 25 年 11 月より公開されている。このサイトへの月間ユニーク IP アクセス数は 2 千件程度となっており、広く利用されている。

#### 4-3. 研究開発成果によるライフサイエンス分野のデータ活用への波及効果

本課題の成果は、ゲノム疫学研究の進展や、ヒトの健康を維持するための高齢化社会実現のための重要な情報源であり、今後データが公開されれば、付加価値の高い情報提供および貢献が期待できるものとする。

#### 4-4. 広報・アウトリーチ活動等

論文発表、学会講演・発表などで研究成果の発表を行っている。さらに、展示会において、

デモ等を実施している。

## 5. 総合評価

本課題は、臨床の現場に即してコホートデータベース開発における様々な問題を解決しようとしており、開発された統合データベースシステムを他のコホート研究へ提供を行っている。個人情報が多い課題であり、データの公開には慎重な取り組みが必要だが、研究成果の早期公開が望ましい。さらに、近い将来、本データベースシステムが標準システムへ発展し、コホートデータを中心とする臨床疾患情報とゲノム情報の統合化データベースの普及により、研究情報が効率的に共有、連携されることでデータ駆動型の臨床治療や予防医学の発展へ寄与することが期待される。