

疾患解析から医療応用を実現する DB開発

東大医、東大病院、東海大学、日立製作所

1-1 2008/04現在での進捗

- 標準SNP DBの構築 (版07/08完了)
- SNP品質管理、及び、遺伝統計解析技術の開発 (開発中)
- Case-control DBの構築(版完了)
- ALSリシーケンシングDB (版完了)
- レポジトリ環境の整備 (実施中)
- 詳細データ公開方針案(倫理検討、アクセス権)の作成

1-2

標準DBの構築

目的: 品質管理、健常対照者データを目的としたDBの構築

内容:

- 健常者500名以上の30-100万SNPの遺伝子型頻度、アレル頻度、ハーディーワインバーグ平衡検定値、Call rate等
- 連鎖不平衡値 (D' , r^2)、ハプロタイプ頻度
- SNPのアノテーション (機能、染色体上位置、同義/非同義等)

データの品質管理:

- 検体のQC(欠損率、ヘテロ接合度), サンプル構成のQC(サンプル重複、混血検体の同定), SNPのQC(欠損率)

現在:

- DBにはAffymetrix500K: 459検体、Affymetrix6.0(900K): 198検体、Illumina317K: 200検体を蓄積。逐次データを追加。
- 健常者459人分、500K SNPのタイピングに基づき、各種解析を終え、論文投稿中。

目的: 遺伝統計以外の人も解析結果を有効利用できるDBの構築

内容:

- 30-100万SNPの遺伝子型頻度、アレル頻度、ハーディーワインバーグ平衡検定値、Call rate等
- P-value (genotype, allele) Additive risk model, recessive model, dominant model のP-value, OR, 95% CI, AICなどの遺伝統計値
- ハプロタイプもしくはSNPの組み合わせに関する疾患関連性の統計値
- SNPのアノテーション (機能、染色体上位置、同義/非同義など)

表示:

- 上記の値を、テーブルとグラフで表示

現在:

- 版完了
- グラフィカルモデリングのPath consistencyアルゴリズムを用いたゲノムワイドなSNP間相互作用検出方法の検討

1-4

データ公開に関する倫理面の検討

- 1) Genotype, Allele frequency data、計算結果について
 - 個人情報を含まない研究成果のため、論文がアクセプトされ次第、あるいはDBに登録12ヶ月後に公開。

- 2) Genotype data, raw dataについて
 - GWASデータ公開方針の最終案がほぼ完成。研究倫理審査委員会による審査、データアクセス委員会による審査等を経て提供される案。

 - 統合DBの倫理問題検討委員会を開いて検討、了承いただく。

2-1 リシークエンスDB進捗と今年度の予定

- 筋萎縮性側索硬化症 (ALS) のリシークエンスDBの構築
(版完了)
 - 目的:
網羅的なmutationの情報とそれに付随する臨床情報を搭載し、臨床現場にも役立つDBの構築を目的
 - 内容:
疾患関連遺伝子に関する患者の独自のリシークエンスデータ(41検体)と付随する臨床情報
論文から収集したmutation情報と臨床情報(症状の種類、発症年齢など)
- その他
Mutationの遺伝子機能情報、立体構造情報など
- 今後の予定: パーキンソン病への拡張

- デモ

3-1 今年度のGWAS-DB予定

- ケース・コントロールGWAS-DBのインターフェースの拡充 (genotype callの表示など)。
- 第2ステージ (再現性確認と高精度マッピング) のタイピング結果の蓄積、表示へ対応するためにDB拡張。
- パニック障害、子宮内膜症 (厚生科研) ほかのケースコントロールデータをDBに追加搭載及び、外部機関産出データを搭載予定。
- 外部の機関が産出したデータの受け入れのため遺伝統計処理部分の計算パイプラインの構築。
- GWAS多重検定の問題回避手法の開発とマイニング手法の開発。

4

年次計画

