統合データベースプロジェクトにおける ヒトゲノムバリエーションデータベース

○小池　麻子[1]　西田　奈央[2]　吉田　真希子[1]　井ノ上　逸朗[3]　辻　省次[4]　徳永　勝士[2]

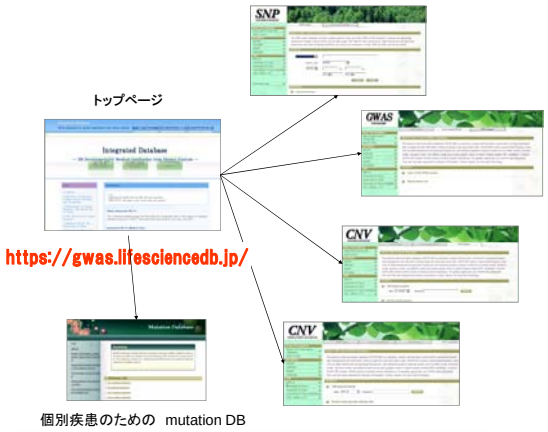
[1]日立製作所　中央研究所, [2]東京大学大学院医学系研究科, [3]遺伝研, [4]東京大学医学部附属病院

HITACHI
Inspire the Next

要旨

近年の遺伝子多型・変異解析技術の発展に伴って表出したデータ管理の問題に応えるために、我々（東大医、東大病院、遺伝研、日立製作所）は、データの半永続的な一元管理と、研究者間の情報共有による個別疾患研究促進、及び、遺伝統計学研究の促進を目指し、ヒトゲノムバリエーションデータベースを構築し、データのサブミッションを広くお願いしている(https://gwas.lifesciencedb.jp/)。本データベースは、GWASの研究を対象としたDBと、個別の疾患を対象としたmutation DBからなる。前者として、健常対照群のSNPデータを登録した標準SNP DB、ゲノムワイドなSNP関連解析結果を登録したGWAS DB、健常対照群のコピー数変異を登録したCNV DB、及びコピー数変異の関連解析結果を登録したCNV関連解析DBを構築している。これらのDBでは、研究概要、品質基準、各種頻度データや遺伝統計解析結果を登録・表示するだけでなく、エクソン情報, miRNA情報など、解析に有効な他の情報と計算結果を重ね合わせて可視化する機能も備えており、 GWAS研究者だけでなく個別疾患の研究者にも広く利用して頂けるDBを目指している。

一方、mutation DBでは筋萎縮性側索硬化症などの神経変性疾患について、疾患ごとに変異と臨床情報を登録しており、臨床現場でも役立つDBを目指している。



個別疾患のための　mutation DB

1. 統合データベースにおける ゲノムワイド関連解析データベース（GWAS-DB）

実施内容: 標準SNPのDBの構築とcase-control GWAS DBの構築と データの受入れ。

目的:

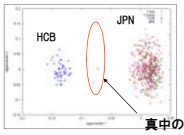
1)実験結果・解析結果のDB化による半永続的な集約的データ管理

2)GWAS以外の研究者にも情報を提供することによる疾患研究の促進

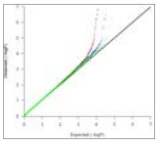
3)研究者間の情報共有による遺伝統計研究の促進

2. SNPの品質管理

- SNPの品質管理
  - Call rate
  - Hardy-Weinberg 平衡検定値
- 検体の品質管理
  - Call rate
  - ヘテロ接合度
- 集団の品質管理
  - HAPMAP dataとの比較
    （主成分分析、多次元尺度法など）
  - qq-plot



PCAの例（Eigenstrat）



qq-plotの例

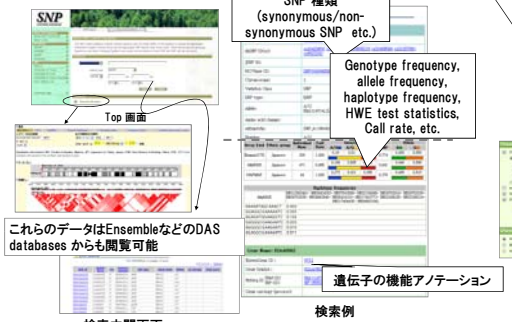
3. 標準SNPのDBの構築

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、Hardy-Weinberg平衡検定値、Call rate等
連鎖不平衡値（D', r2)、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）

現在のデータ数:
DBにはAffy500K 450検体　Affy6.0 200検体を蓄積。
逐次データを追加。

4. 標準DBの画面例



検索中間画面　　検索例

5. GWAS (case-control) DBの構築

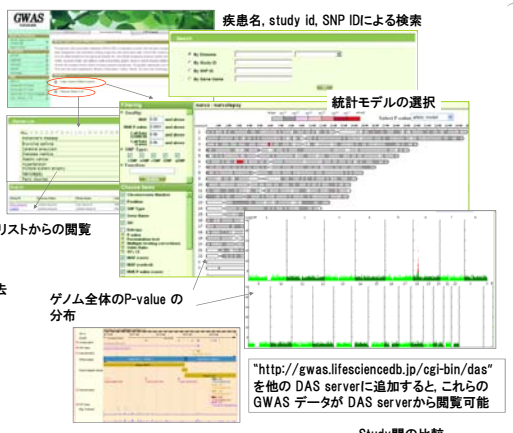
内容:

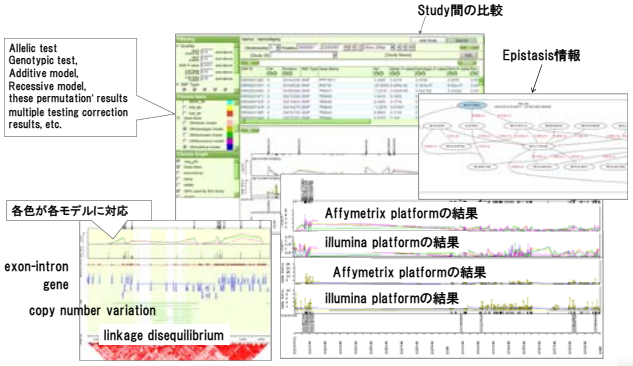
- 30-100万SNPの遺伝子型頻度、アレル頻度、ハーディーワインバーグ平衡検定値、Call rate等
- genotypic test, allelic test, additive risk model, recessive model, dominant model のP-value, OR, 95% CI, AICなどの遺伝統計値
- Bonferroni, FDR, などでの多重検定の補正値
- Haplotypes と表現型との関係性
- SNPのアノテーション（機能、染色体上の位置、同義語/非同義のSNPなど）

現在9 studiesが公開中。

内部用DBでは40studies以上が登録・利用。

6. GWAS (case-control) DBの画面例

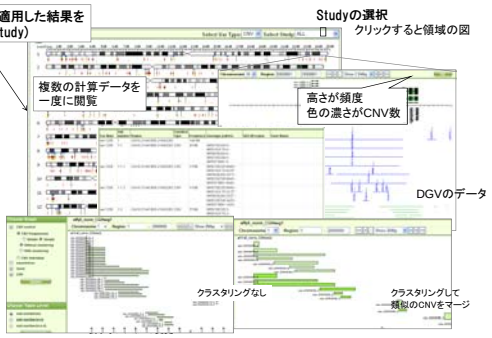




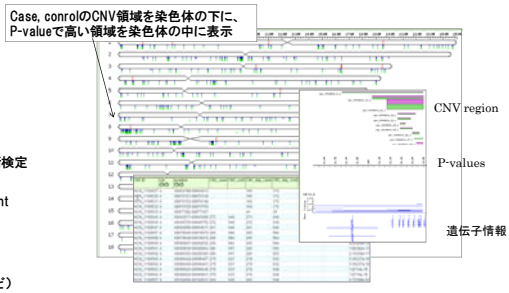
他の情報との比較例　　Study間の比較の例

データ登録者の解析結果、もしくは、plink, haploview,などを用いた遺伝統計解析結果

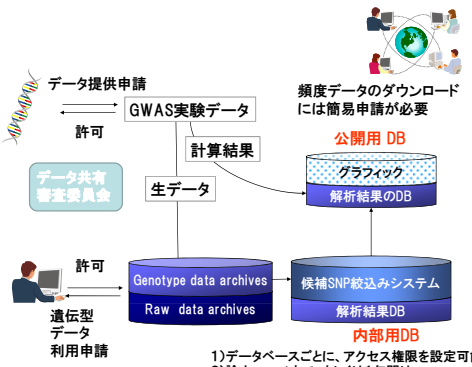
7. CNVデータの画面例



8. CNV association DBの画面例



9. レポジトリーシステムの概要



1)データベースごとに、アクセス権限を設定可能。
2)論文acceptまで、もしくは1年間は
データ未公開で、グループ内での閲覧は可能

10. Mutation DBの概要

目的:
1)　遺伝子変異と臨床情報の両方を俯瞰
2)　疾患の機序の理解
3)　将来的には臨床現場で役に立つDBに

内容:
・文献から収集した変異情報と臨床情報
・サブミッションされた変異情報と臨床情報

ALS (Amyotrophic lateral sclerosis) database
筋萎縮性側索硬化症
PD (Parkinson disease)
ALD (Adrenoleukodystrophy disease)
副腎白質ジストロフィー
HSP (Hereditary spastic paraplegia)
家族性痙性対麻痺

本GWAS データベースはレポジトリー型のデータベースです。皆様の GWAS データのサブミッションをお待ちしております。
gwas@lifesciencedb.jp

謝辞

謝辞:本研究は、文部科学省統合データベース整備事業の研究費で行われた。

Licensed under a Creative Commons表示2.1日本 license
(c)統合DB –PJ 疾患解析グループ