

# 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 ゲノムワイド関連解析 (GWAS) 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 (構築中、 版08/02完了予定)
- レポジトリ環境の整備 (実施中)

# 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ク平衡検定値、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, サンプル構成のQC, SNPのQC

現在：

- DBにはAffy500K 450検体 Illumina317K 200検体を蓄積。逐次データを追加。
- 健常者390人分、50万SNPのタイピングに基づき、各種解析を終え、論文作成中。

# 1-3

## 標準DBの構築

SNPの検索 (アクセッション番号、染色体上の位置、機能、疾患との関連性などで検索可能)

Genome Browserを利用して、他のデータベースコンテンツと同時に表示

SNPのゲノム上の位置、SNPの種類(同義/非同義など)

Genotype 頻度、アレル頻度、HWE検定値、Call rateなど

対応する遺伝子のアノテーション情報

The screenshot displays the 'SNP control database' interface. At the top, there are navigation links (link01 to link04) and a 'Link' menu. The main content area shows search results for a specific SNP. A callout box points to the SNP details, including its genomic position, type, and alleles. Another callout box points to a table of genotype data, which includes columns for array kind, ethnic group, individual number, call rate, and genotype frequencies (G/G, G/T, T/T) along with HWE and allele frequencies. A third callout box points to the gene ontology information for the associated gene, 'catenin (cadherin-associated protein), alpha 3'. A fourth callout box points to the OMIM ID for the gene.

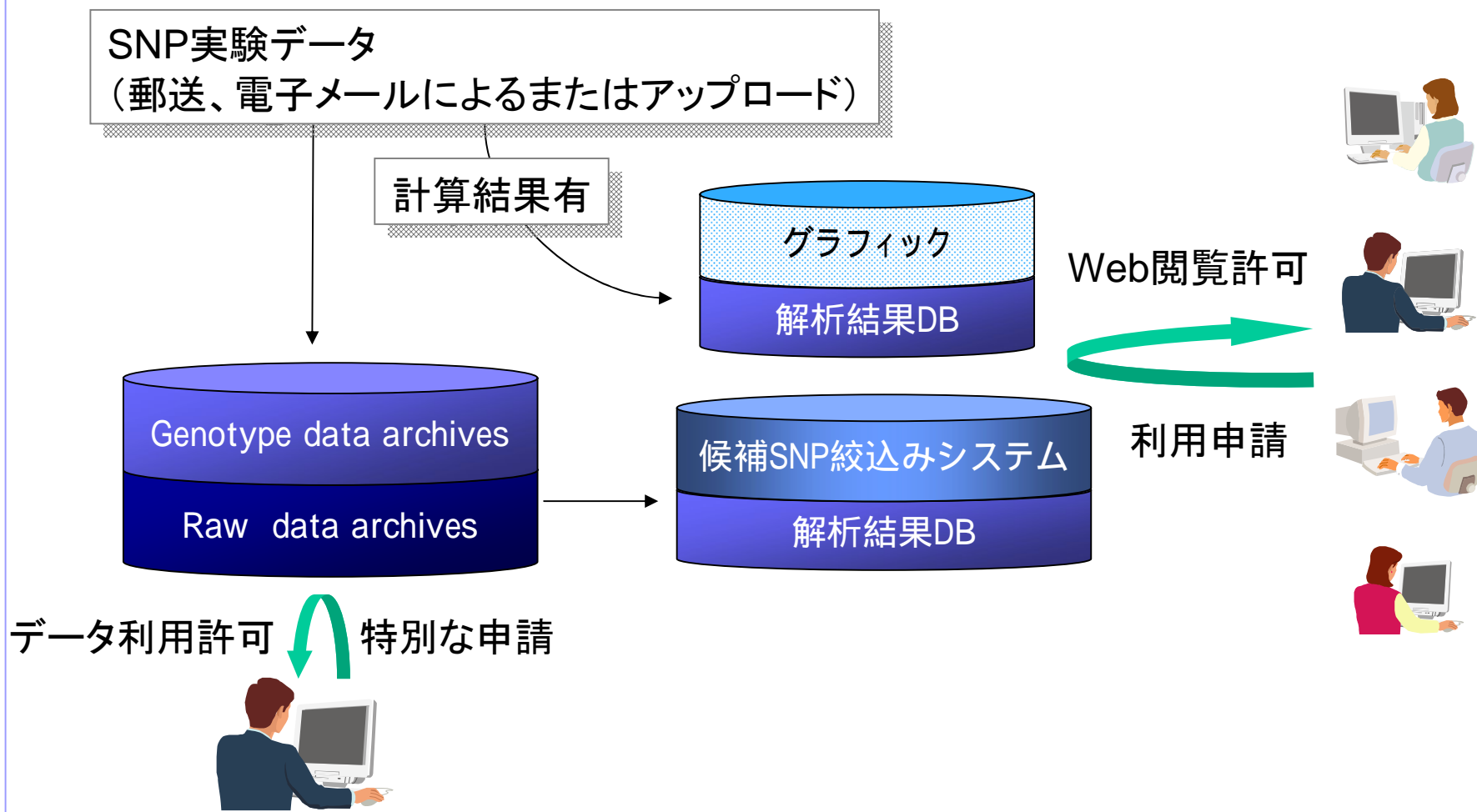
Array kind	Ethnic group	Individual Num.	Call Rate	Genotype detail			HWP	Allele	
				G/G	G/T	T/T		G	T
Amy600K	Japanese	477	0.976	0.803	0.354	0.041	0.804	0.780	0.220
Human317K	Japanese	200	1.000	0.555	0.405	0.04	0.180	0.760	0.240
HAPMAP	Japanese	44	0.978	0.477	0.409	0.113	0.732	0.680	0.320

## 2-1 今年度のGWAS-DB予定

- 簡易なインターフェース(計算結果の表示のみ)のケース・コントロールGWAS-DBは来年2月に完了予定。
- ナルコレプシー、パニック障害、多系統萎縮症(ゲノム4領域)、脳動脈瘤(CREST)、子宮内膜症(厚生科研)についてのケースコントロールデータをDBに搭載予定。
- 90万SNP (+CNV) 200検体の標準データ追加 (本年11月)。
- 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。
- 外部の機関が産出したデータの受け入れのためのレポジトリシステムの構築。

# 2-2 レポジトリシステムの概要

## データ登録



## 2-3

# ゲノム4領域でのGWASデータベースとの切り分け

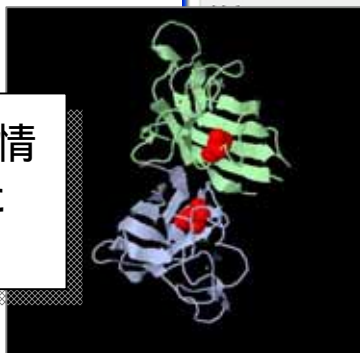
- ゲノム4領域からのGWASデータは一部
- ゲノム4領域でのGWASデータベースはSNPタイピングを行う班員の情報共有が主目的（重複実験の回避、多数の目によるデータのQC）
  - タイムリーなデータ共有（論文アクセプト前を含む）
  - QCは最低限、生データとAllele・Genotype頻度、HWE検定値のみの簡易解析結果を蓄積
- 統合DBでのGWASデータベースはSNPタイピングを行う研究者のみでなく機能解析を行う研究者への情報提供を目的
  - 外部研究者による特別申請に対して、生データを提供するシステム
  - 半永続的なメンテナンスを目指したDBシステム構成
  - 統計遺伝に不慣れな研究者もデータを活用できるよう、統計遺伝解析を一通り行った結果をグラフィック表示、ダウンロード機能を付加
  - QCによる解析精度の向上

## 3-1 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(本年10月 版完了予定)
- 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
- パーキンソン病のリシークエンスDBを来年2月に完了予定。既に、リシークエンスデータ(240検体)と200文献弱のmutation-臨床データを収集済み。

# 3-2 リシークエンスDB進捗

リシークエンスによる新規mutation情報 + 臨床情報と文献から抽出した Mutation情報 + 臨床情報



ADAR2 detail information - Microsoft Internet Explorer

ALS mutation database

Top Overview of mutations Help Link1 Link2

ADAR2 adenosine deaminase, RNA-specific, B1 (RED1 homolog rat)

**Detail information**

Gene Symbol ADAR2

Full name adenosine deaminase, RNA-specific, B1 (RED1 homolog rat)

Synonym RED1; ADAR2; ADAR2a; ADAR2b; ADAR2d; ADAR2g; DRABA2; DRADA2; ADAR2a-L1; ADAR2a-L2; ADAR2a-L3

Genome position

Links

Mutation Search

**Sequence information**

0 1

P37962 NP\_00929.1 P37962-2 NP\_00931.1 P37962-3

Original exp

Active sites

Multifunctional domains

Synuclein

a.a.:

```

MK1PRMTPGQDFRNSLRGGRNPKYFAMDIEDENKSSGSDYKFNRLDNVSPKDGSTPGQEGSLSNGGGGG
SKYRLKRRKTPQPYLPMALMQLNEIKPGLQYLLSOTGPHAPLFYMSYEVNDYFEGSDPTKXKAKLHAKEA
LAMQRTLSYNTDFTSDDADFPDILMDFETPKAEPPFFYVSGNGDGFSSCGQLSLSSAPYPASLADPPLPLPFF
    
```

ALS関連遺伝子配列情報 + variation情報

gene	Genomic	AA change	homo/hetero	no. of patients	clinical characteristic	FALS/SALS	age on set	duration(y/m)	onset site	years until initiation of artificial respirator
SOD1	chr21 G31953966T	A4S		1/29(29: SOD1 mutation)		ALS	34	>3 y #1	L leg #2	
SOD1	chr21 G31953966T	A4T		14		FALS	44 ± 11	1.2 y	Lower limb/ upper limb	
SOD1	chr21 G31953966T	A4T		1/29(29: SOD1 mutation)		ALS	21	20 m	L lower limb	
SOD1	chr21 G31953966T	A4T			rapid progression course	FALS			Lower limb	
SOD1	chr21 C31953967T	A4V	Hetero	61/148(148: SOD1 mutation)	rapid progression course	FALS	25	12 m	R arm /leg	5 m(non- invasive)
SOD1	chr21 C31953967T	A4V	Hetero	61/148(148: SOD1 mutation)	rapid progression course	FALS	26	>3 y	Bulbar	7 m(non- invasive) 9 m (invasive)
SOD1	chr21 C31953967T	A4V		2/34(34: FALS)	rapid progression course	FALS	50	8 m	Upper limbs proximal	
SOD1	chr21 C31953967T	A4V		2/34(34: FALS)	rapid progression course	FALS	57		Upper limbs distal	



# 4

## 年次計画

簡易なインターフェースのcase-control DBは  
08/02に完了予定。  
今年度中に遺伝子型頻度と遺伝統計解析結果  
まで公開予定。

研究項目	1年目	2年目	3年目	4年目
GWAS DB・解析手法開発	← 一次スクリーニング用DBシステム →	← 二次スクリーニング用DBシステム →	← システムの公開と改良 →	← 公開と改良 →
標準SNP DB構築	← 標準SNP - DBの構築 →	← システムの構築 →	← システムの構築 →	← 公開と改良 →
リシーケンシング DB・解析手法開発	← 変異疾患・臨床情報DBの構築 →	← システムの構築 →	← システムの公開と改良 →	← 公開と改良 →

DB構築はベータ版完了(07/08)、GWASデー  
タの追加と外部データの受け入れシステム  
の構築。(08/03に完了)  
今年度中に遺伝子型頻度と遺伝統計解析結  
果まで公開予定。

ALSのデータベースが(07/10)  
に完成予定。  
Raw data 以外は今年度末公開  
予定

PD のデータベースが(08/02)  
に完成予定。