

ライフサイエンス分野の統合データベース整備事業

# 疾患解析から医療応用を 実現するDB開発

2007/04/27

東京大学大学院医学系研究科  
徳永 勝士

# 1-1. 背景

HapMap PJの完了と配列解析技術の飛躍的向上により  
ゲノムワイドなSNPタイピング、将来的には大規模な  
リシーケンシングが可能

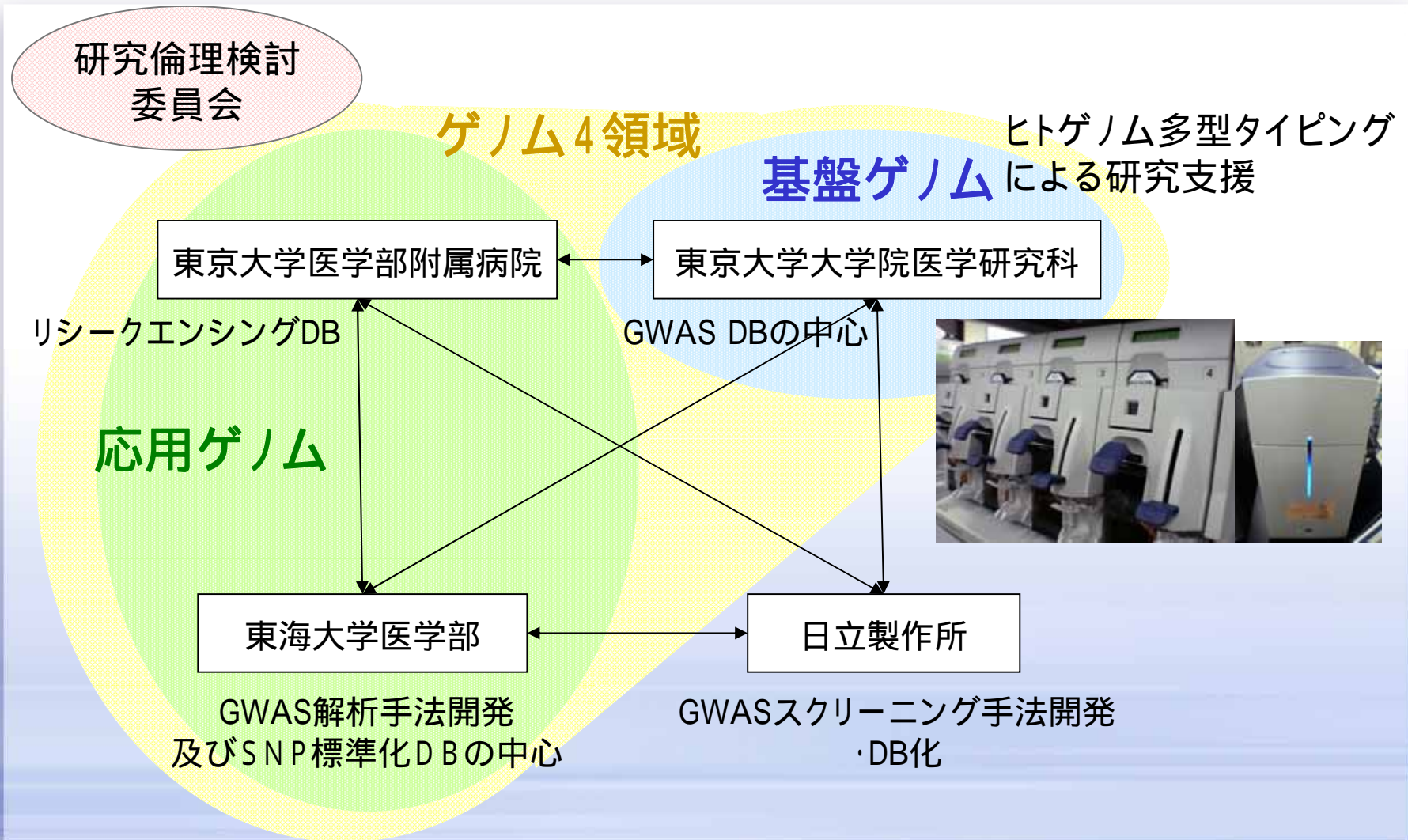


疾患関連遺伝子のゲノムワイドな探索と  
膨大な実験結果・解析結果のDB化と  
公開が重要な課題

## < 国内外の研究状況 >

- ・NIH NCBI: dbGaP公開
- ・英国、アイスランド: 疫学・診療情報と遺伝子情報をDB化
- ・理研 30万人のバイオバンクプロジェクト

# 1-2. 研究体制



# 2-1. 研究開発概要

目的: 疾患の遺伝要因の解明と、個別化医療実現に向け、ゲノム情報と臨床情報を統合したDB構築と解析手法の開発

1) 50万～100万種のSNPタイピングによるゲノムワイド関連解析 (GWAS)のDB開発と解析手法・システムの開発

[DB公開 2008～]

2) リシークエンシングによる疾患関連遺伝子の変異情報と付随臨床情報のDB構築と解析手法・システムの開発

[DB公開 2009～]

[中核機関側 2010～]

連鎖解析・関連解析データ、およびゲノム多様性の  
デポジトリシステムの構築

## 2-2. 研究対象と整備状況

### <ゲノムワイド関連解析:GWAS>

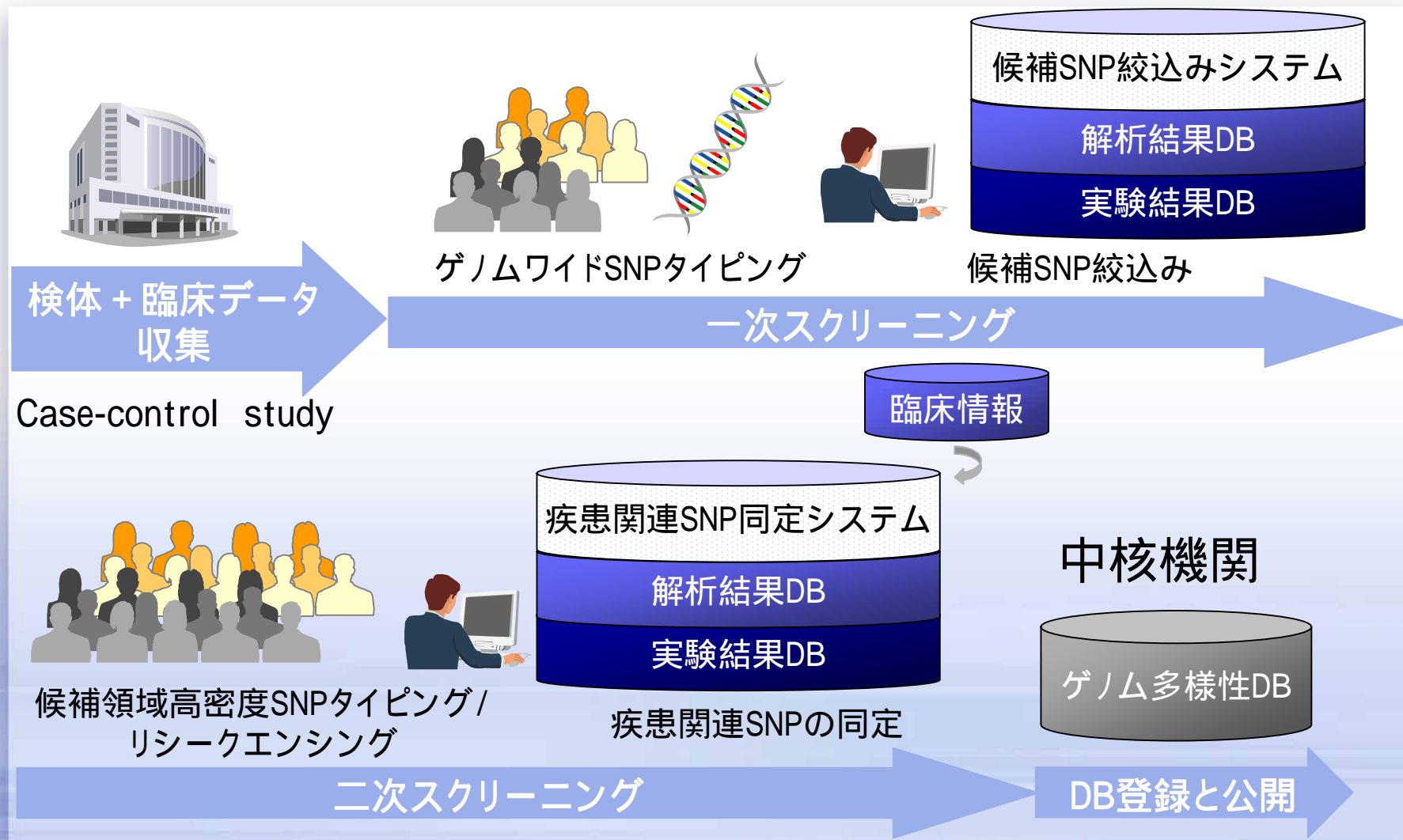
- ・現在4疾患（多系統萎縮症、パニック障害、脳動脈瘤、ナルコレプシー）のSNPタイピングは完了し簡易DB化。変形性関節症など今後も対象疾患は増加

### <リシーケンシング>

- ・筋萎縮性側索硬化症(ALS)、パーキンソン病は完了し簡易DB化。認知症、痙性対麻痺など、今後も対象疾患は増加。解析は独自設計のリシーケンシングマイクロアレイを利用(~80遺伝子)。

[ゲノム実験解析は文科省特定ゲノム、CREST、厚労省科研費などにより実施]

# 2-3. ゲノムワイド関連解析 (GWAS)



# 2-4. GWAS 一次、二次スクリーニング解析

## < SNP品質解析 >

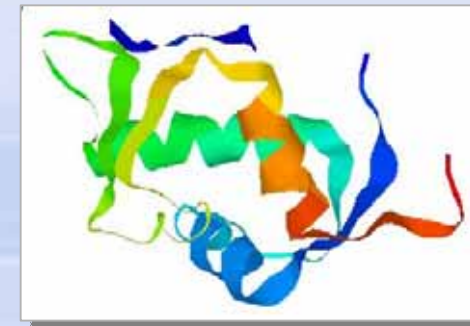
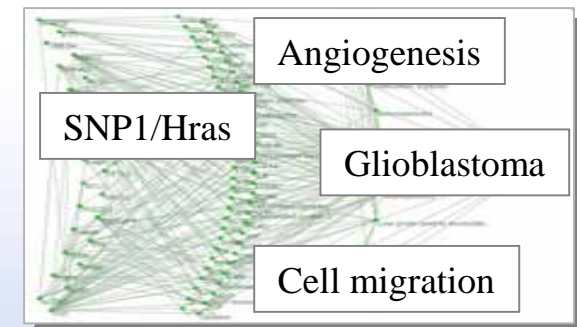
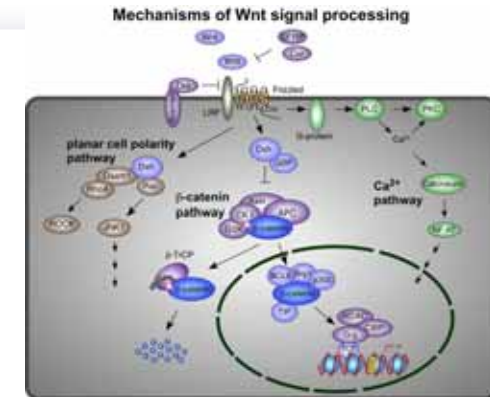
- ・ SNP標準DBを構築し、低品質のSNPを除外

## < 統計遺伝学的解析 >

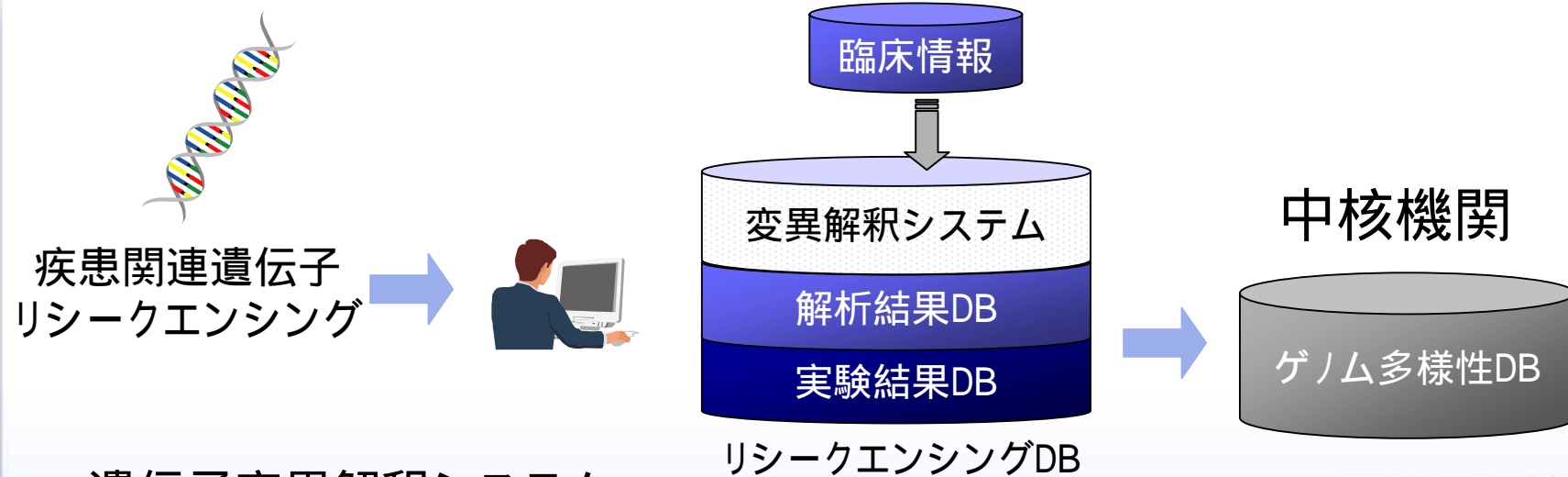
- ・ 単点解析、ハプロタイプ解析など

## < その他の解析 >

- ・ Case群, control群を最適に判別できる臨床情報 + SNP組合わせの選択
- ・ Pathwayや蛋白質相互作用情報の利用
- ・ 遺伝子アノテーション情報、および文献情報の利用
- ・ 転写制御領域、立体構造情報の利用

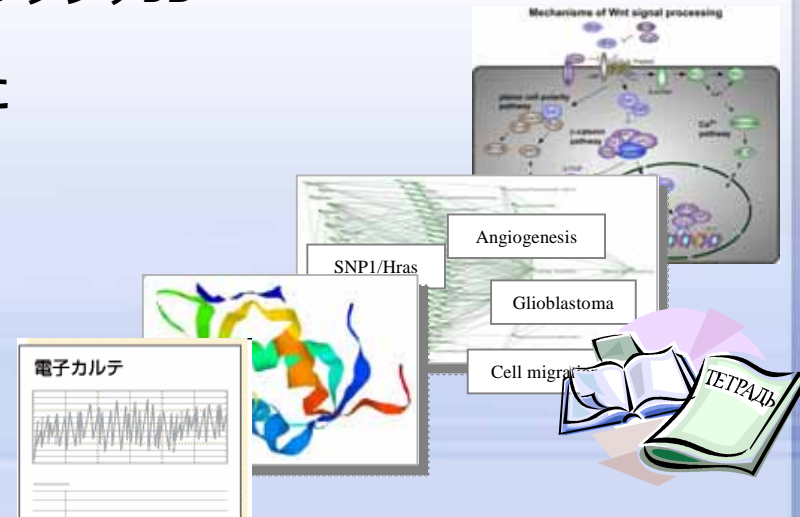


# 2-5. リシーケンシングのDB



## < 遺伝子変異解析システム >

- ・ pathway, 文献, 立体構造などを考慮した遺伝子変異の解釈
- ・ 臨床情報も利用し、疾患と遺伝子情報の因果関係を予測
- ・ 前向き研究として、遺伝子情報と予後の関係を予測





# 3-1. 年次計画

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	システムの公開と改良
標準SNP DB構築	標準SNP - DBの構築	システムの公開と改良		
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		システムの公開と改良	

おわり

---



## 補足： リシークエンシング

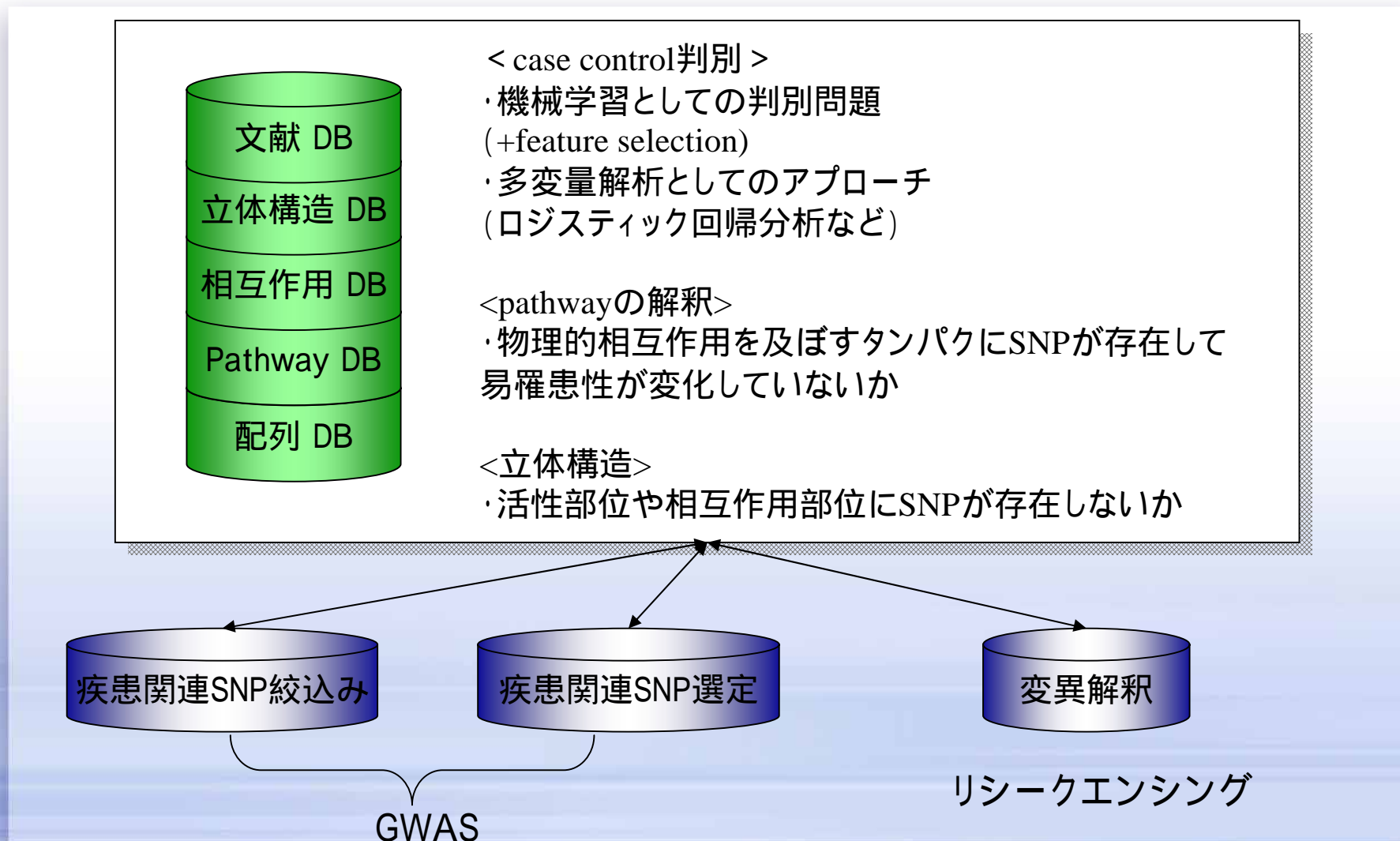
### ・解析した主な患者件数

家族性パーキンソン病(PD)	240例,
筋萎縮性側索硬化症(ALS)	41例(家族性8例, 孤
発性33例),	
痙性対麻痺(FSP)	85例,
副腎白質ジストロフィー(ALD)	20例,
家族性アルツハイマー病(AD)	10例,

## 補足： 連携について

- ・GWASのデポジトリーのように、半永久的に維持すべき項目については中核機関と連携の下、システムを構築する。
- ・ゲノム多様性DBは、中核機関の計算機に搭載し、その他のDB(発現など)を利用し複雑な条件検索に対応可能に。
- ・GWASスクリーニング及びリシークエンシングのマイニングに利用する文献情報処理では、特定の疾患名称などに特化して用語を収集するため、収集した用語は中核機関の辞書構築にフィードバック。
- ・RESTやSOAPを使った仮想的なDBの統合方式については、中核機関と連携しながら必要に応じて開発。

# 補足： 共通データ・解析手法の利用



# 補足： 開発すべき技術課題

## 1. 疾患関連SNPスクリーニング技術

### 1-1 統計的アプローチ

グラフィカルモデリングなどを利用したSNP組み合わせによる偽陽性の低減など

### 1-2 Bioinformatics的なアプローチ

<case control判別>

・機械学習としての判別問題

(+feature selection)

・多変量解析としてのアプローチ (ロジスティック回帰分析など)

<pathwayの解釈>

・物理的相互作用を及ぼす蛋白にSNPが存在して易罹患性が変化していないか

<立体構造>

・活性部位や相互作用部位にSNPが存在しないか、もしくは、  
配列情報からfunctional SNPらしさを予測など

## 2. RESTやSOAPを利用した仮想的な統合DB技術及び統合フォーマット

(必要に応じて)

## 3. 文献情報処理技術による機能・疾患関連SNP情報の自動収集

# 補足：スクリーニング システムイメージ

### SNP位置テーブル

SNP1	Chr 11	522243	a.a置換有	ras
SNP2	Chr 2	10223		intergenic
...				

### 臨床情報テーブル

患者1	情報A	情報B	....
患者2			
...			

### Pathway data

SNP1	VEGF signaling
SNP1	T cell receptor signaling
...	

### 遺伝子アノテーション情報

SNP2	Cell cycle
SNP3	Cell death
...	

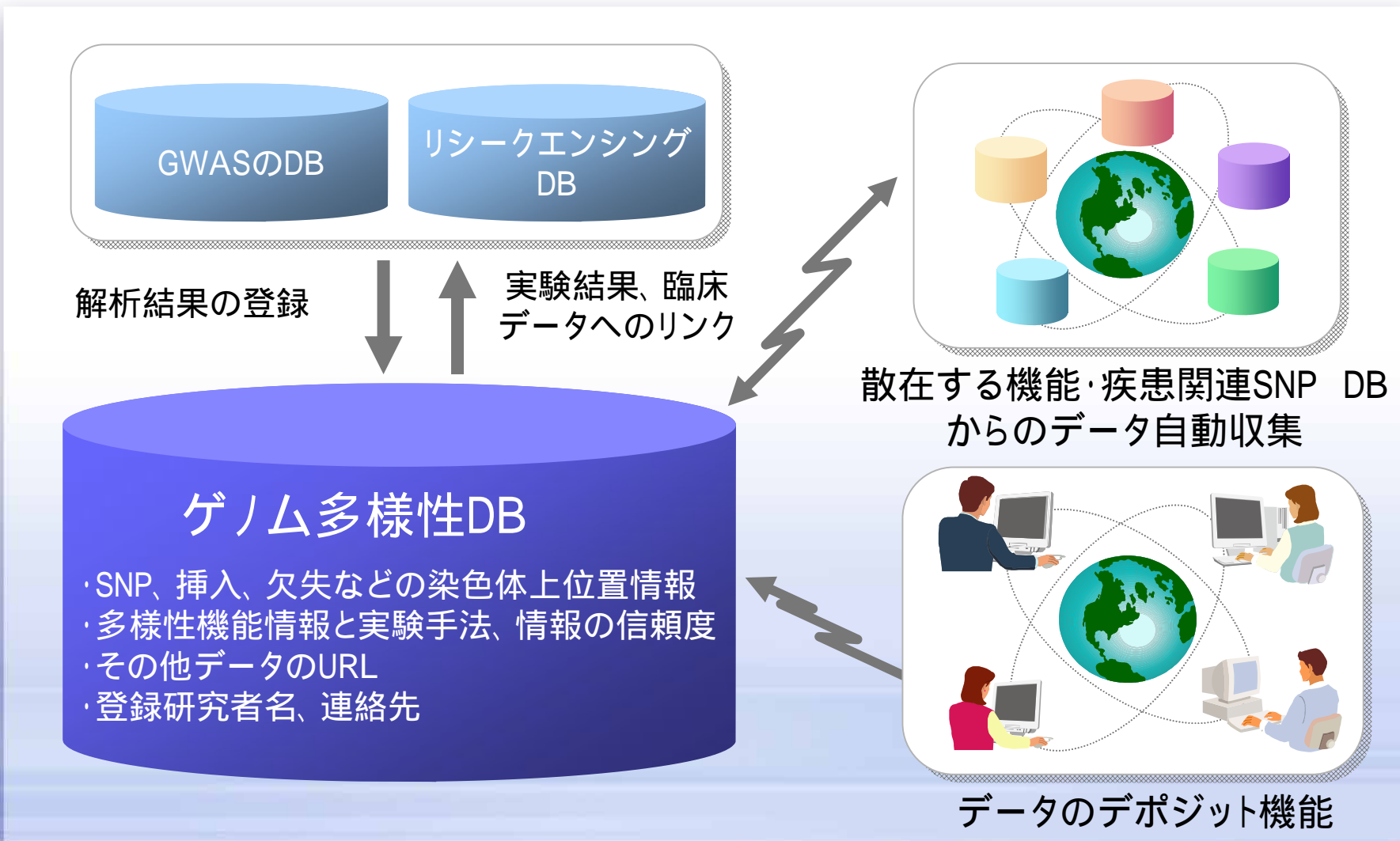
### 蛋白質相互作用データ

SNP2	SNP10	免疫沈降
SNP3	SNP12	Y2H
...		

# 補足

## 中核機関:

# ゲノム多様性データベースの構築





# 補足 中核機関: GWASのデポジトリデータベース

