資料 3

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 2009/01現在での進捗

- 標準SNP DB　ー開発完了公開中　逐次データ追加
- SNP品質管理、及び、遺伝統計解析技術の開発 ー開発中
- Case-control DBの構築　ー基本機能実装完了公開中　高機能化実装中
- CNV標準DBの構築　ー基本機能実装完了　高機能化実装中
- ALSリシークエンシングDB ー基本機能実装完了
- パーキンソン病リシークエンシングDB ー開発中
- レポジトリー環境の整備　ー実施中
- 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:

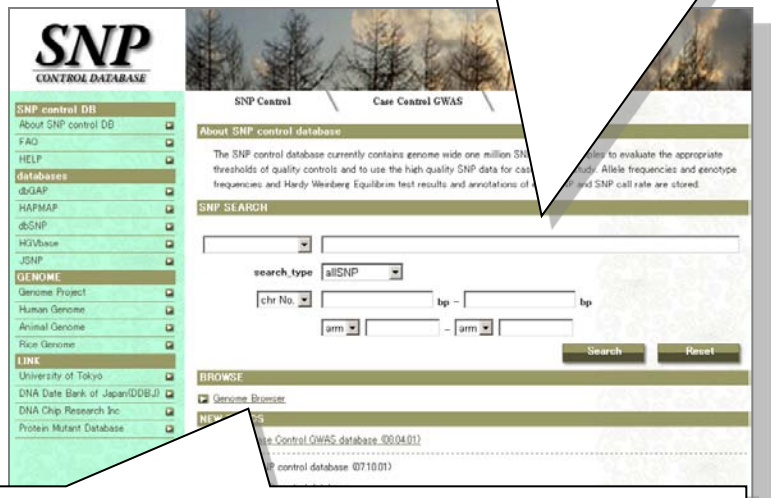
- 検体のQC(欠損率、ヘテロ接合度), サンプル構成のQC(サンプル重複、混血検体の同定), SNPのQC(欠損率)

現在:

- DBにはAffy500K 約500検体、Affy6.0 約200検体、Illumina317K 約200検体を蓄積。逐次データを追加。

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、ハプロタイプ頻度、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

SNP search

SNP ID: NRS6663840	
dbSNP ID(rs):	rs6663840
dbSNP ID(ss):	
JSNP ID:	
HGVbase ID:	
Chromosome:	1
Variation Class	SNP
SNP type:	
Allele:	
Amino acid change:	
Affymetrix:	
Illumina:	A/G

Array kind	Ethnic group	Individual Num.	Call Rate	Genotype detail A/A	A/G	G/G	HWP	Allele A	Allele G
Illumina317K	Japanese	200	1.000	0.1	0.51	0.3	0.774	0.450	0.550
Affy500K	Japanese	471	0.965	0.188	0.505	0.305	0.641	0.440	0.560
HAPMAP	Japanese	44	1.000	0.272	0.431	0.295	0.376	0.490	0.510

Haplotype frequencies	
Affy500K	NRS12563491- NRS9424283- NRS7543006- NRS2154068- NRS6702916- NRS6702935- NRS6703035- NRS6663840- NRS9424310- NRS17403773- NRS2298225- NRS2298224- NRS17404435- NRS6683156;
AAAAATAGCAAACT	0.403
GGAGGCGAAAAATC	0.393
AGAGATGGAAGGTC	0.104
GGGGGCGAAGAATC	0.033
AGGGGCGAAGAATC	0.015
AGAGGCGAAAAATC	0.011

Gene Name: KIAA0562	
EntrezGene ID :	9731
Gene Symbol :	KIAA0562
Refseq ID (NM-ID) (NP-ID)	NM_014704 NP_055519
Gene ontology (process):	

1-4 Case-control DBの構築

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 (2DF, 1DF), 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）

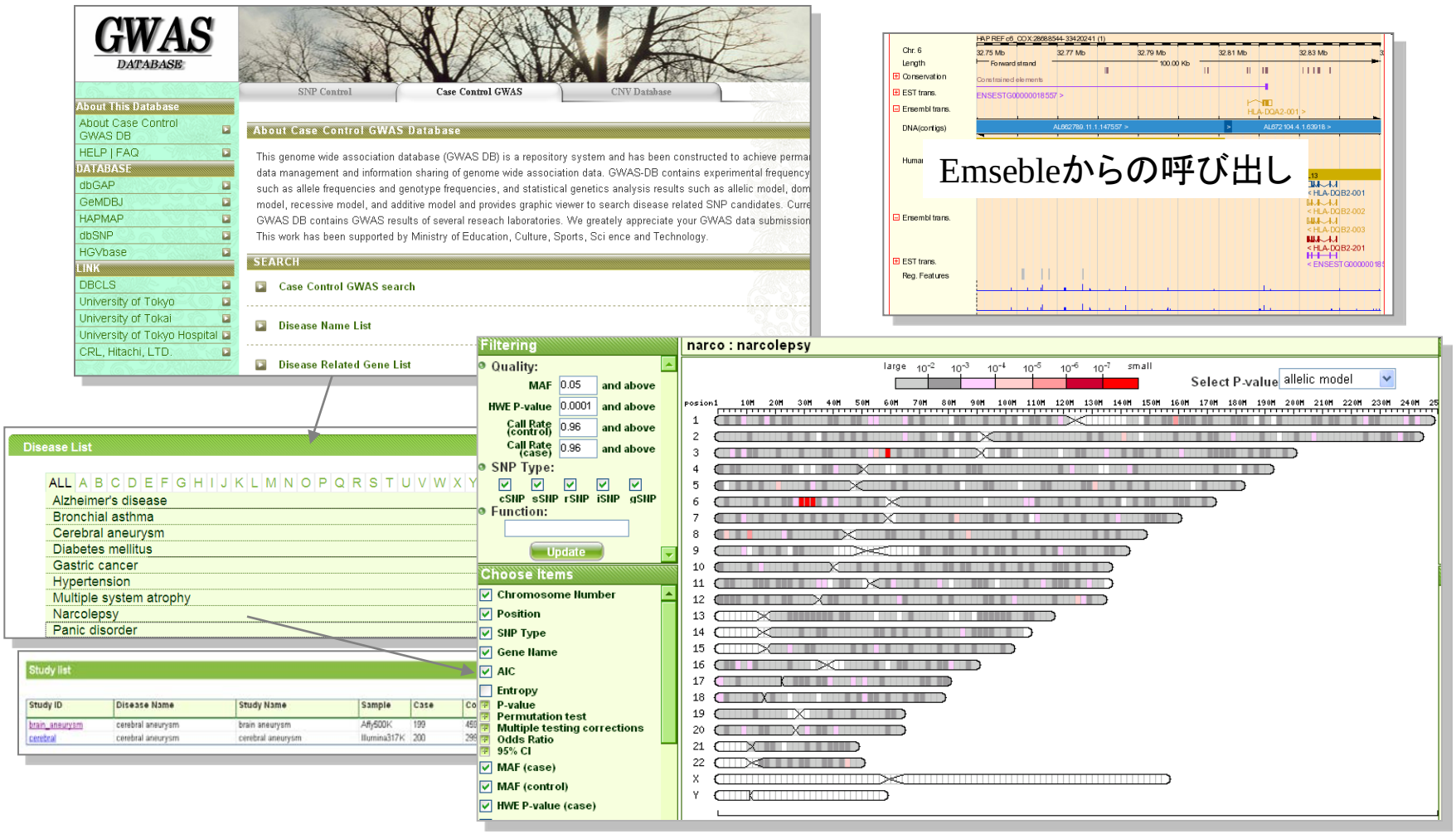
表示：

- 上記の値を、テーブルとグラフで表示
- プラットフォーム間の比較表示、既知疾患情報との比較表示が可能

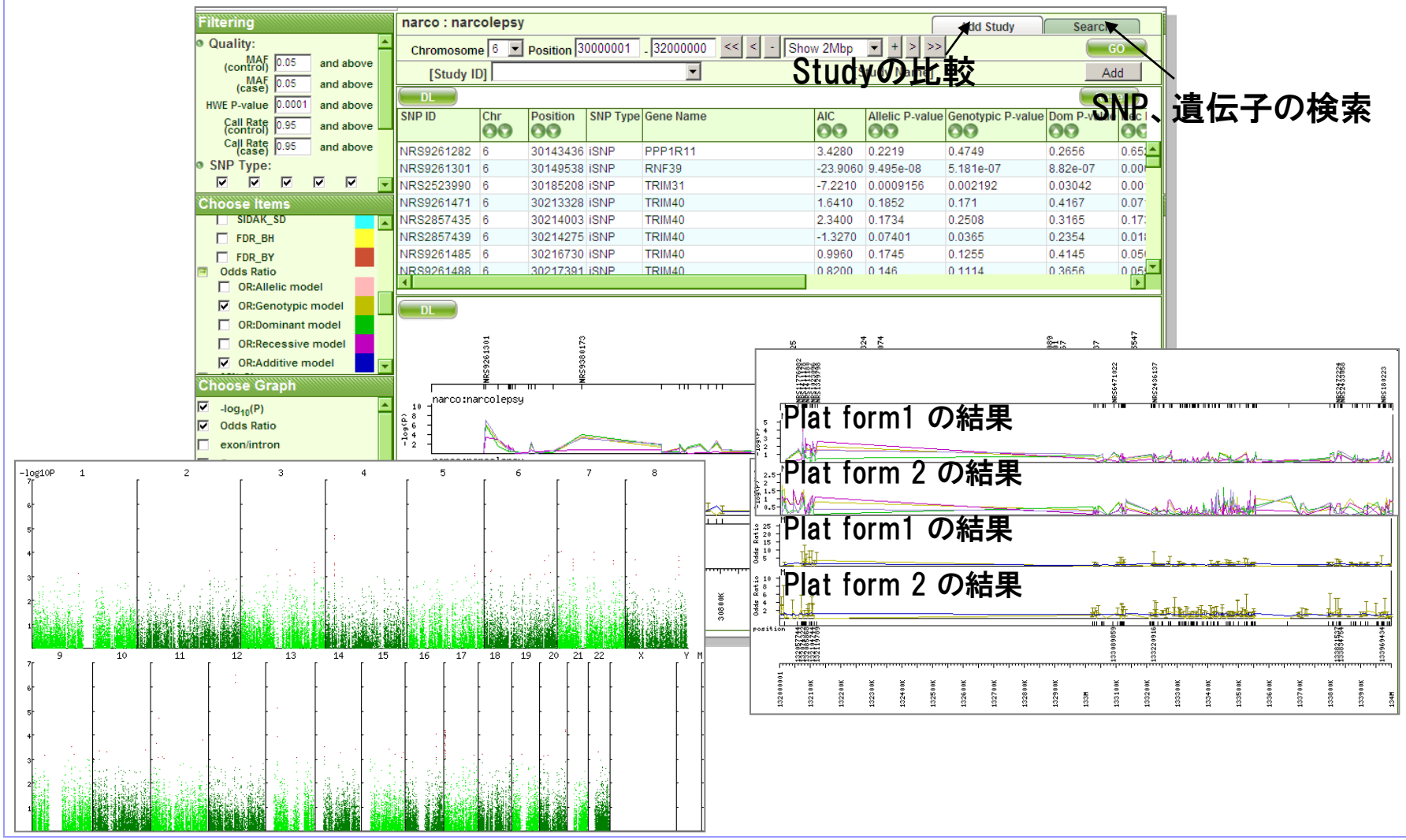
現在：

- 12疾患中8疾患を公開中
- 機能追加部分を実装中

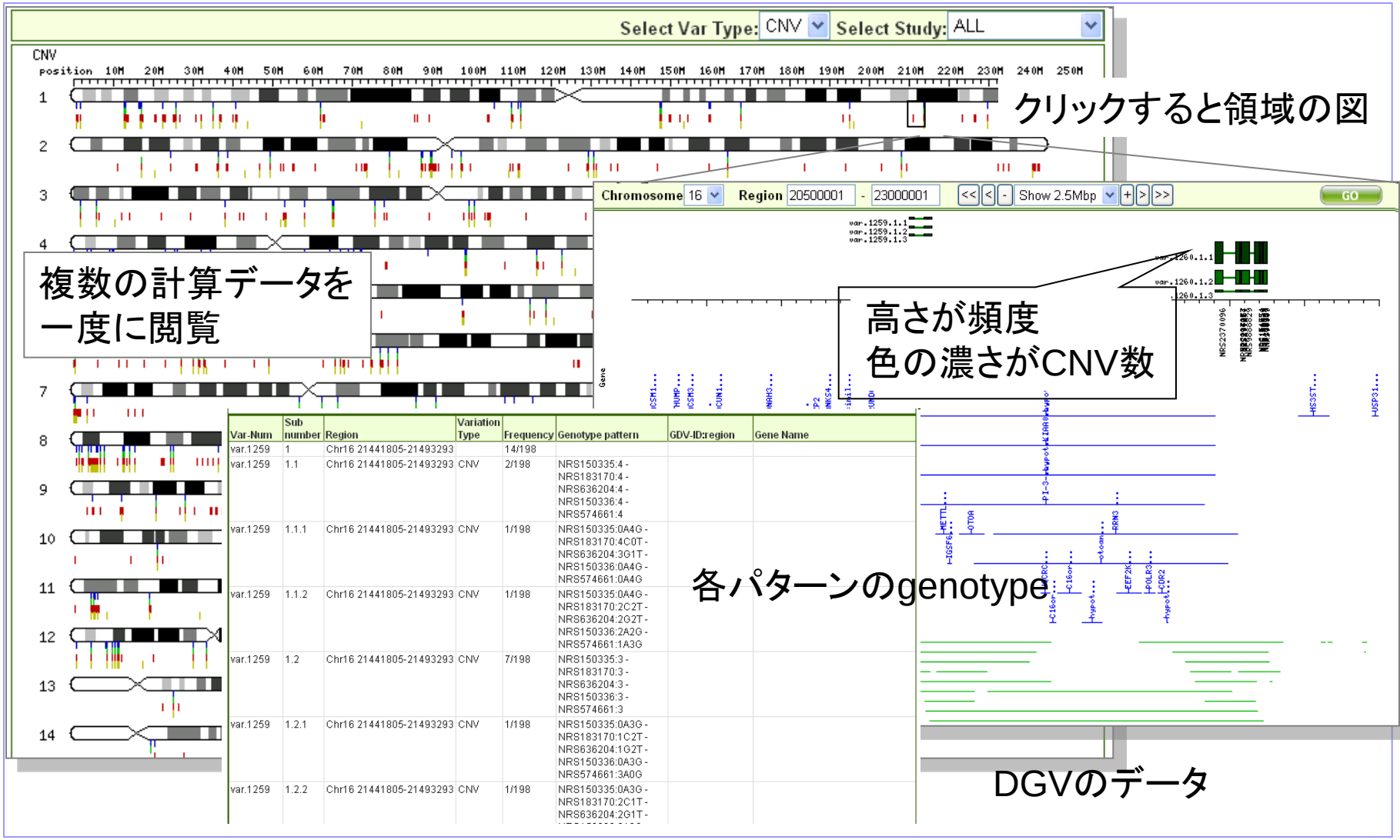
1-5 Case-control DBの画面例



1-6 Case-control DBの機能追加例



1-7 CNVのデータベースの表示例



Var-Num	Sub number	Region	Variation Type	Frequency	Genotype pattern	GDV-ID:region	Gene Name
var.1259	1	Chr16 21441805-21493293		14/198			
var.1259	1.1	Chr16 21441805-21493293	CNV	2/198	NRS150335:4 - NRS183170:4 - NRS636204:4 - NRS150336:4 - NRS574661:4		
var.1259	1.1.1	Chr16 21441805-21493293	CNV	1/198	NRS150335:0A4G - NRS183170:4C0T - NRS636204:3G1T - NRS150336:0A4G - NRS574661:0A4G		
var.1259	1.1.2	Chr16 21441805-21493293	CNV	1/198	NRS150335:0A4G - NRS183170:2C2T - NRS636204:2G2T - NRS150336:2A2G - NRS574661:1A3G		
var.1259	1.2	Chr16 21441805-21493293	CNV	7/198	NRS150335:3 - NRS183170:3 - NRS636204:3 - NRS150336:3 - NRS574661:3		
var.1259	1.2.1	Chr16 21441805-21493293	CNV	1/198	NRS150335:0A3G - NRS183170:1C2T - NRS636204:1G2T - NRS150336:0A3G - NRS574661:3A0G		
var.1259	1.2.2	Chr16 21441805-21493293	CNV	1/198	NRS150335:0A3G - NRS183170:2C1T - NRS636204:2G1T -		

1-8 データの受け入れと再配布

- 頻度情報、統計解析結果については一般公開
- 統合データベースプロジェクト　ヒトゲノム多型データ共有方針 を、「特定ゲノム」の方針を参考にして、中核機関と共同で作成済み
- データ受け入れ時のデータ利用の申請書とデータの利用に関する同意書とセキュリティ手順書を作成済み
- 検討委員会を招集し、年度末までに決定予定

2-1 今年度のGWAS-DB予定

- SNP間相互作用の蓄積、グラフ表示機能。
- SNPごとのシグナル強度のグラフ表示などのGWAS-DBのインターフェースの拡張。
- 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。
- 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ーシステムの構築。

3-1 リシークエンスDB進捗と今年度の予定

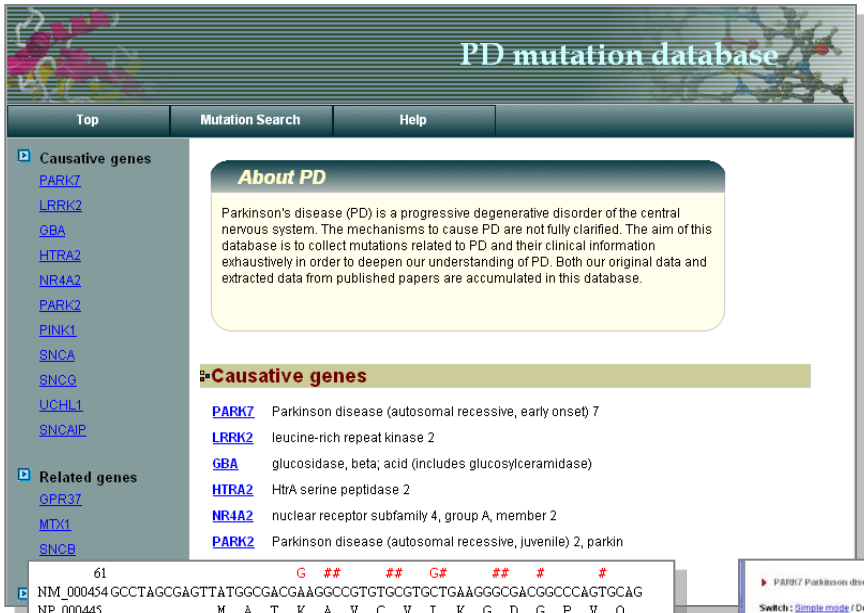
- パーキンソン病のリシークエンスDBの構築とALSリシークエンスDBのインターフェースの改良

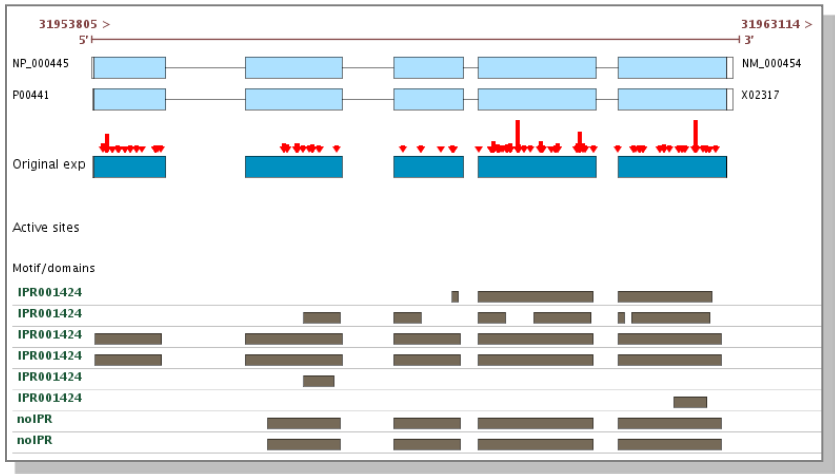
- 目的:
  - 網羅的なmutationの情報とそれに付随する臨床情報を搭載し、臨床現場にも役立つDBの構築を目的
- 内容:
  - 疾患関連遺伝子に関する患者の独自のリシークエンスデータと付随する臨床情報
  - 論文から収集したmutation情報と臨床情報(症状の種類、発症年齢など)

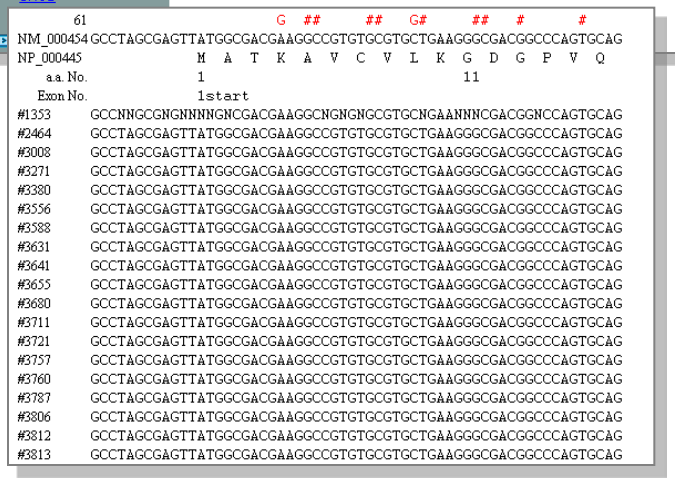
  その他
  - Mutationの遺伝子機能情報、立体構造情報など

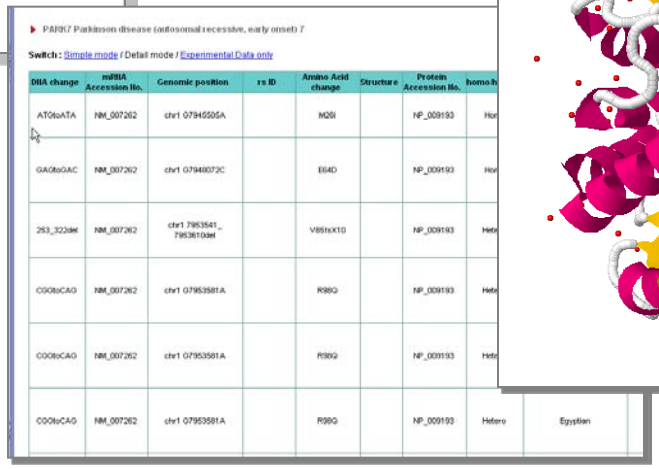
- 現在:ALSは構築完了、パーキンソン病は構築中
- 今年度の予定:パーキンソン病の実験データの追加

3-2 リシークエンスDBの画面例









4 来年度の計画

- CNVの検出手法の改良とCNVのcase-control DBの構築
- より多くのGWASデータの受け入れ、登録と公開
- 一次データと二次データの関係性の表示、発現データなどの他実験データなどとの重ね合わせ表示など、GWAS DBの拡張
- SNP相互作用部分の計算手法の開発
- 新たに定める倫理規定、及び手順に従ったデータの預け入れと再配布サービスの開始
- 海外他機関との連携の検討
- パーキンソン病リシークエンスDBの仕上げ
- 副腎白質ジストロフィーリシークエンスDB 、痙性対麻痺リシークエンスDBの構築