

# 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 2008/01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 (構築中、 版08/02完了予定)
- ALSリシークエンシングDB (表示機能完了)
- 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

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)

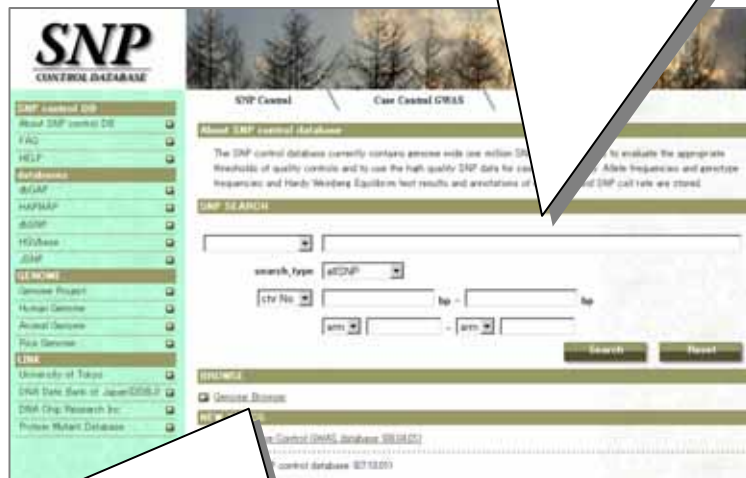
データの品質管理:

- 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 459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。

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 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:

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

Array kind	Ethnic group	Individual Num.	Call Rate	A	G	detail G/G	HWP	Allele A	G
Illumina317K	Japanese	200	1.000	0.188	0.812	0.3	0.774	0.450	0.550
Affy500K	Japanese	471	0.965	0.188	0.812	0.305	0.641	0.440	0.560
HAPMAP	Japanese	44	1.000	0.272	0.728	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
AGAGATGGAAAGGTC	0.104
GGGGCGAAGAATC	0.033
AGGGCGAAGAATC	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):

対応する遺伝子のアノテーション情報

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:

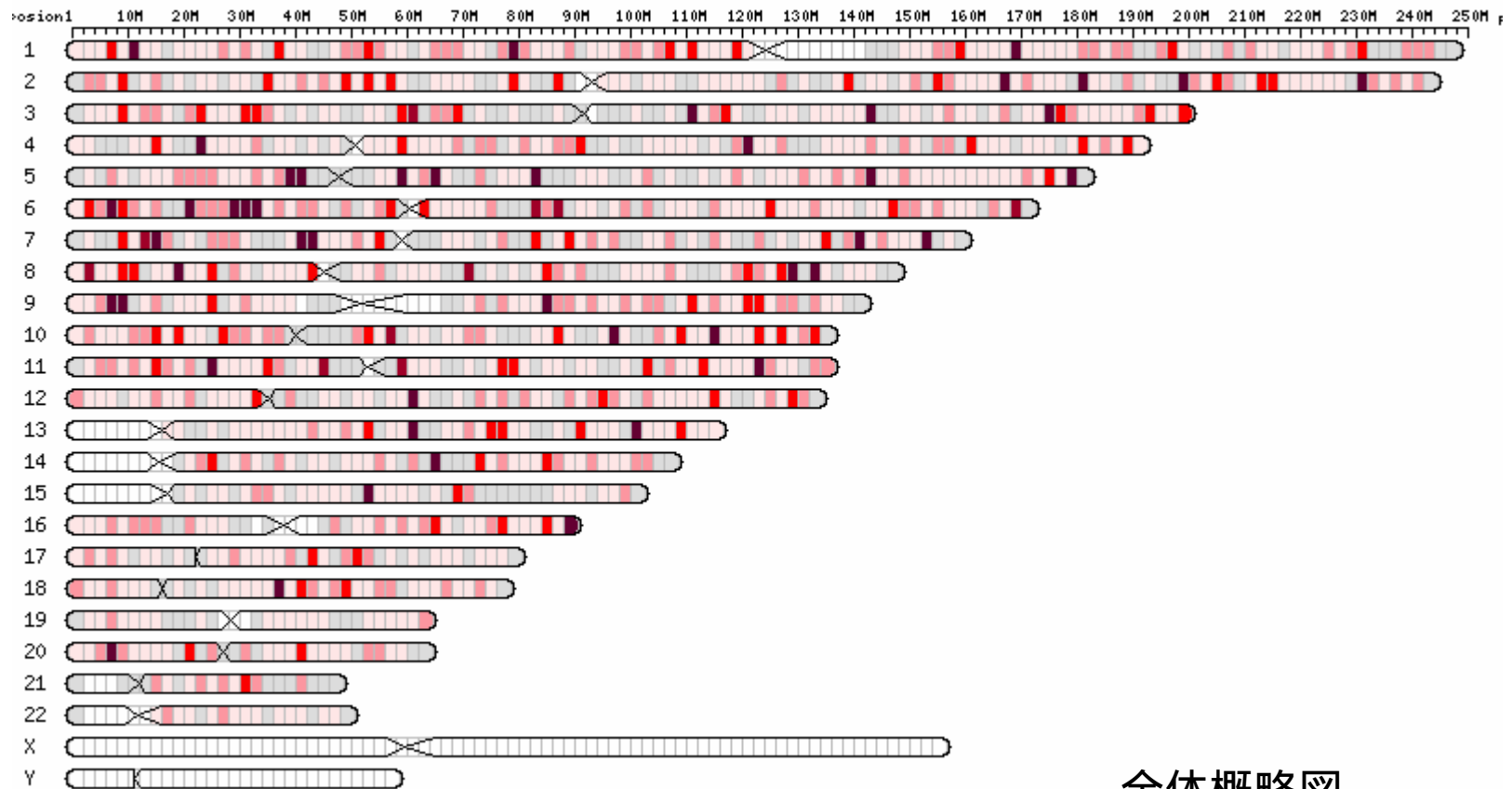
- 上記の値を、テーブルとグラフで表示

現在:

- 現在DB構築中。08/02に 版完了予定
- 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

# 1-5

## Case-control DBの構築



全体概略図  
P-valueごとの色分け



## 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は今年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 (今月中)。
- 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。

## 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  
(現在データ登録機能以外は実装済み)
- 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: データ簡易登録機能の実装



# 3-2 リシークエンスDB進捗



```

a.a.:NP_000445P00441
  EV G Q V E S C      NY H R C R Q
MATAKAVLKGDPVQGLINFEQWENGPQKVVGGIKGLTEGLHGHVHE
  S C V A F K VA Y TN G
PCDNTA CTSACDFRNPFSVHYGGPKDRERHVCDELGVTADRDGVADVSI
  S FLY NTAG KTKKX HDGGRPKX  -H Y RGE FT IT
EDSVISLSGDRLIGRTLIVYRRKADLLKGGNEESTKTNAGNSRLACGV1
  
```

```

mRNA : NM_000454
GTTTGGGCCAGAGTGGCCGAGCCCGAGTCTGGCCTATAAAGTACTC
GCGGAGCGGGGTGCTGGTTTGGCTGTAAGTCTCTCGCAGGCTCTGGGT
TTTCGTTGCGAGTCTCGGAACCCAGACCTCGGCGCTGGCCTAGCGAGTTAT
  T T T A T T C C G C
GGCAGCAGAGCGCTGCTGGCTGCTGAAGGGCAGCGCCCGCTGCAAGGCA
  G
TTCATCAATTTTCGACCAAGCAAGTAATGACCAGTCAAAGCTCTGGCGA
  
```

ALS mutation database

Top Mutation Search Help

Table of contents  
[Detail information](#)  
[Sequence information](#)  
[Mutation Search](#)  
[Mutation & Clinical information](#)  
[Multiple alignments](#)  
[3D structure](#)  
[Structure information](#)  
[Overview of mutations](#)  
[Reference information](#)

Causative genes  
[ADARF1](#)  
[CNTF](#)  
[CSTH1](#)  
[TARDPF](#)  
[RNL1V](#)  
[ANO](#)

SOD1 superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))

**Detail Information**

Gene Symbol	SOD1
Full name	superoxide dismutase 1, soluble (amyotrophic lateral sclerosis 1 (adult))
Synonym	ALS, ALS1, IPOA, SOD, homodimer
Genome position	chromosome: 21; Location: 21q22.1[21q22.11 31953806 .. 31963115 (strand: +)]
Links	<a href="#">Entrez</a> <a href="#">UCSC</a>

**Sequence Inform**

NP\_000445 P00441

NM\_000454 X02317

Original exp

Active sites

Motif/domains

IPR001424  
IPR001424  
IPR001424

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

ALS関連遺伝子  
 配列情報 + variation情報

MS change	mRNA Accession No.	Genomic position	rs ID	Amino Acid change	Structure	Protein Accession No.	Source hetero.	Population	No. of families	No. of patients with mutation
GCC>C	NM_000454.012	chr21:31953807		Asp		NP_000445	ALS		1	
GCC>A	NM_000454.013A	chr21:31953807		Asp						
GCC>A	NM_000454.013A	chr21:31953807		Asp						
GCC>A	NM_000454.013A	chr21:31953807		Asp						

Clinical characteristics	FALS/TALS	Sex (M/F)	Age on set	Ethnicity	Inheritor type	Onset site	Years since mutation of official register	PubMed
	ALS	M	34	<2 y		Leg/42		PMID: 1628103: Phe mutant SOD1 in hand
	FALS		44.511	12 y		Lower limb/upper limb		PMID: 1628103: The clinical phenotype of wild-type SOD1 in hand
	ALS	M	21	20 m		Lower limb		PMID: 1628103: Phe mutant SOD1 in hand
rapid progression course	FALS					Lower limb		PMID: 1628103: Phe mutant SOD1 in hand
rapid progression course	FALS	M	25	12 m		Arm/leg		PMID: 1628103: Phe mutant SOD1 in hand

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

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のデータベース完了。  
Raw data 以外は今年度末公開  
予定