

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)

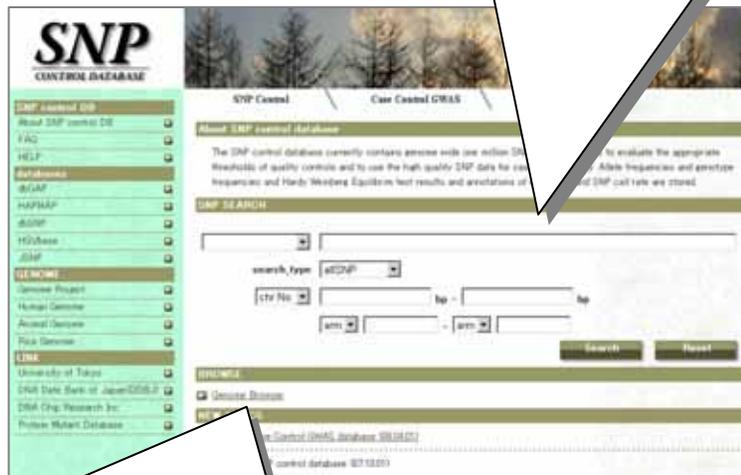
データの品質管理:

- 検体のQC(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	0.505 0.305	0.641	0.440	0.560
HAPMAP	Japanese	44	1.000	0.272	0.728	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
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:

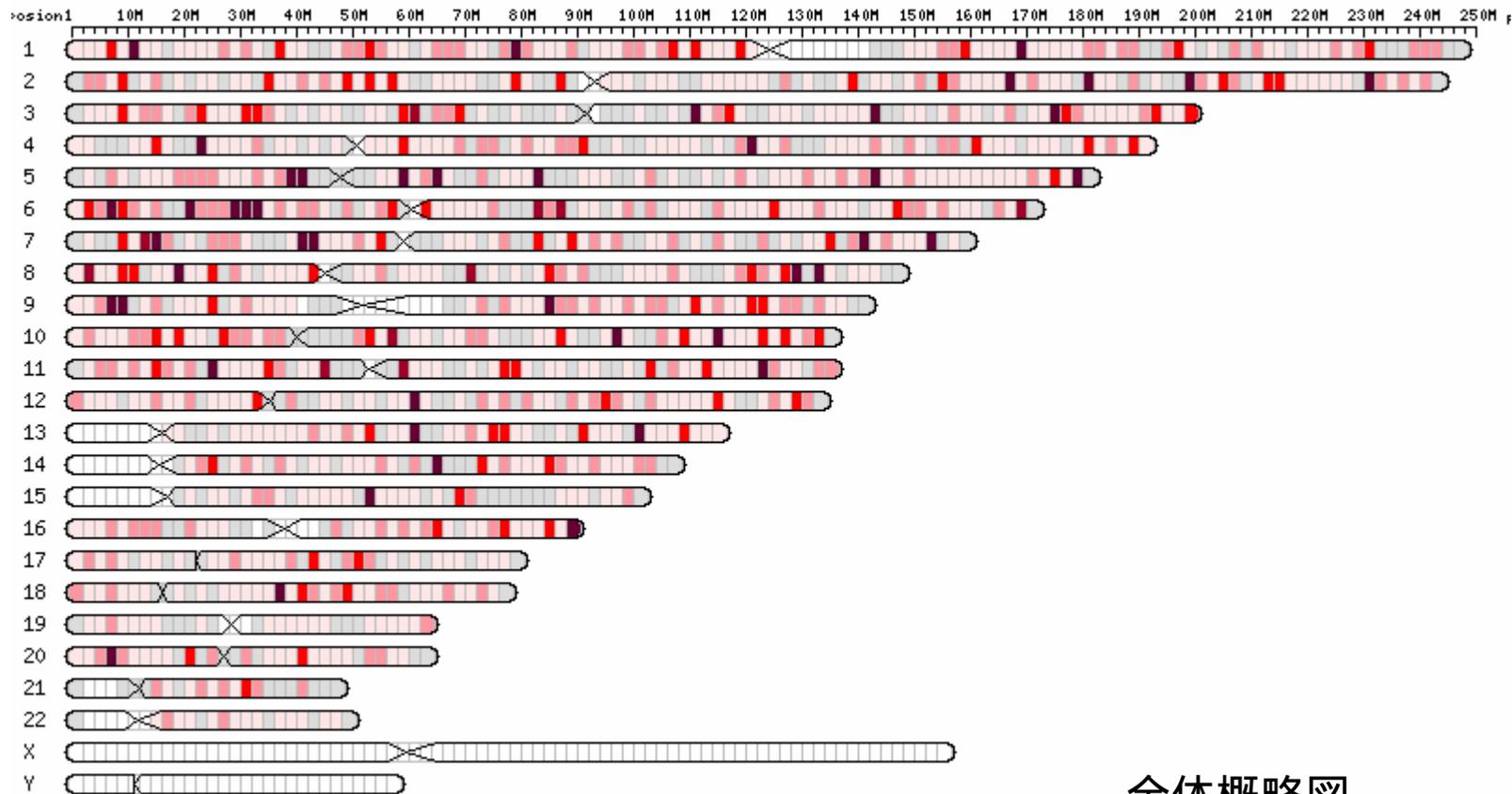
- 上記の値を、テーブルとグラフで表示

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

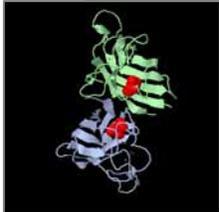
2-1 今年度のGWAS-DB予定

- 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
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
  M A T K A V C V L K G D G P V Q G I I N F Q K E S N G P Q V V G S I K G L T E G L I G H G H V H E
  S E V G Q V F E S C C V A F K Y A V T N G
  P C D N T A C T S A C P H R N P L S E V H G G P D R R H V C D L G V T A D R D C V A D V S I
  S F L V N T A G K T O K E H G G G R R K X - N Y R G E F T I T
  E D S V I S L S G D R C I I G R T L V Y R R K A D D L K G G N E E S T K T G N A S R L A C G V I
  
```

```

mRNA : NM_000454
  G T T T G G G C C A G A G T G G C G A G G C C G A G G T C T G G C C T A T A A A G T A G T C
  G C G G A G A G G G G T G C T G G T T T G C G T A G T C T C T G C A G G T C T G G G T
  T T C G T T G C A G T C T C G G A A C C A G A C C T C G G C T G G C C T A G C G A G T T A T
  G G C A C A A G C G C T G C G T G C T G A A G G G C A G C G G C C A G C A G G G C A
  T T C A T C A A T T T G C A C C A G G A A G A A T A T G C A C T G A A G G T C T G G G A
  
```

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 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
[ADAR1](#)
[CNTF](#)
[CSTH1](#)
[TARDP](#)
[RNF19](#)
[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	Entrez UCSC

Sequence Inform

NP_000445 P00441

NM_000454 X02317

Original exp

Active sites

Motif/domains

IPR001424
IPR001424
IPR001424

ALS関連遺伝子
配列情報 + variation情報

MS change	RefSeq 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_000454.0121	chr21:031903867		Asp		NP_000445.0151			1	
GCC>A	NM_000454.013A	chr21:031903867		Asp						
GCC>A	NM_000454.013A	chr21:031903867		Asp						
GCC>A	NM_000454.013A	chr21:031903867		Asp						

Clinical characteristics	FALS/TALS	Sex (M/F)	Age on set	Ethnicity	Inheritance type	Onset site	Years since mutation of official register	PubMed
	FALS	M	34	<10 yr		Leg/42		PMID: 1628103: Fam mutant SOD1 in hand Sato T, Nakashita T, Yi Z, Acker F, Sagar F, He Nakagawa M, Haraoka Neurology. 2005; 65:1
	FALS		44.511	12 y		Lower limb/upper limb		PMID: 1625215: The clinical phenotype of hand Rigbi L, Vangelopoulos Robinson W. Arch Neurol. 2005; 62
	FALS	M	21	20 m		Lower limb		PMID: 1622103: Fam mutant SOD1 in hand Sato T, Nakashita T, Yi Z, Acker F, Sagar F, He Nakagawa M, Haraoka Neurology. 2005; 65:1
rapid progression course	FALS					Lower limb		PMID: 1622051: mutat associated with hand Nakano R, Inohara T, Nishiguchi N, Tsutsi S. Neurosci Lett. 1996; 2
rapid progression course	FALS	M	25	12 m		Arm/leg		PMID: 1450030: SOD dismutase gene in amyotrophic lateral sclerosis Anderson RM, Sosa RL, D'Souza RC, London Amyotroph Lateral Scl

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 →	← 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。 →	← 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 →

ALSのデータベース完了。
Raw data 以外は今年度末公開
予定