

# 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

## 2009年5月現在での進捗

- ・ 標準SNP DB: 開発完了公開中、逐次データ追加
- ・ SNP品質管理および遺伝統計解析技術: 標準機能実装完了、新規開発中
- ・ Case-control DBの構築: 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: 基本機能実装完了、高機能化実装中
- ・ ALSリスクエンシングDB: 実装完了
- ・ パーキンソン病リスクエンシングDB: 開発中
- ・ 家族性痙性対麻痺リスクエンシングDB: 開発中
- ・ レポジトリ環境の整備: 実施中
- ・ 詳細データ公開方針案(倫理面、アクセス権)を作成

# 1-1 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)

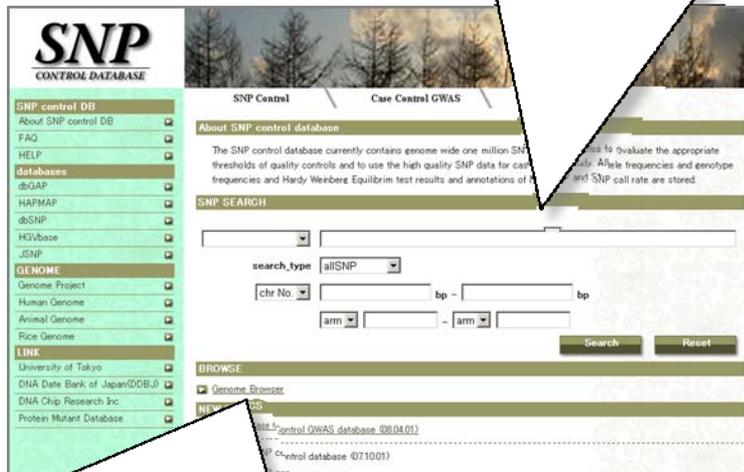
**データの品質管理:**

- ・ 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。逐次データを追加予定

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): **SNPのゲノム上の位置、SNPの種類(同義/非同義など)**

JSNP ID:

HGVbase ID: [chr1:6663840](#)

Chromosome: 1

Variation Class: SNP

SNP type:

Allele:

Amino acid change:

Affymetrix:

Illumina:

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/G	HWP	Allele	
Illumina317K	Japanese	200	1.000	0.189	0.51	0.3	0.774	0.450 0.550
Affy500K	Japanese	471	0.965	0.189	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

Array kind	Haplotype frequencies
Affy500K	NRS12563491- NRS9424283- NRS7543006- NRS2154068- NRS6702916- NRS6702935- NRS6703035- NRS6663840- NRS9424310- NRS17403773- NRS2298225- NRS2298224- NRS17404435- NRS6683156;
	AAAAATAGCAAACT 0.403
	GGAGGCGAAAAATC 0.393
	AGAGATGGAAAGTC 0.104
	GGGGCGAAGAATC 0.033
	AGGGCGAAGAATC 0.015
	AGAGGCGAAAAATC 0.011

Gene Name: KIAA0562

EntrezGene ID : [9731](#)

Gene Symbol : [KIAA0562](#)

Refseq ID (NM-ID) [NM\\_014704](#)  
(NP-ID) [NP\\_055519](#)

Gene ontology (process):

対応する遺伝子のアノテーション情報

目的: ゲノム医学、統計遺伝学の専門家以外も解析結果を有効利用できるDB

内容:

- ・ 30-100万SNPの遺伝子型頻度、アレル頻度、Hardy-Weinberg平衡検定値、Call rate等
- ・ アレル頻度、遺伝子型頻度比較のP-value, 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:

- ・ 15疾患中8疾患を公開中
- ・ 機能追加部分を実装中
- ・ Shrinkage法および閾値モデルを応用した多因子疾患の病因解析システムの開発

# 2-2 Case-Control DBの画面例

**GWAS DATABASE**

SNP Central | Case Control GWAS | GWAS Database

**About This Database**

About Case Control GWAS DB

**DATABASE**

- dbGAP
- dbMDBJ
- HAPMAP
- dbSNP
- RCVtree
- LINK
- DBCLS
- University of Tokyo
- University of Tsukuba
- University of Tokyo Hospital
- CRI, Iihachi, LTD

**SEARCH**

- Case Control GWAS search
- Disease Name List
- Disease Related Gene List

This genome wide association database (GWAS DB) is a repository system and has been constructed to achieve permanent data management and information sharing of genome wide association data. GWAS-DB contains experimental frequency data such as allele frequencies and genotype frequencies, and statistical genetics analysis results such as allelic model, dominant model, recessive model, and additive model and provides graphic viewer to search disease related SNP candidates. Current GWAS DB contains GWAS results of several research laboratories. We greatly appreciate your GWAS data submission. This work has been supported by Ministry of Education, Culture, Sports, Science and Technology.



**Disease List**

ALL | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y

- Alzheimer's disease
- Bronchial asthma
- Cerebral aneurysm
- Diabetes mellitus
- Gastric cancer
- Hypertension
- Multiple system atrophy
- Narcolepsy
- Panic disorder

**Study list**

Study ID	Disease Name	Study Name	Sample	Case	Con
brain_aneurysm	cerebral aneurysm	brain aneurysm	Affy500K	199	459
cerebral	cerebral aneurysm	cerebral aneurysm	Illumina317K	200	299

**Filtering**

**Quality:**

- MAF: 0.05 and above
- HWE P-value: 0.0001 and above
- Call Rate (control): 0.96 and above
- Call Rate (case): 0.96 and above

**SNP Type:**

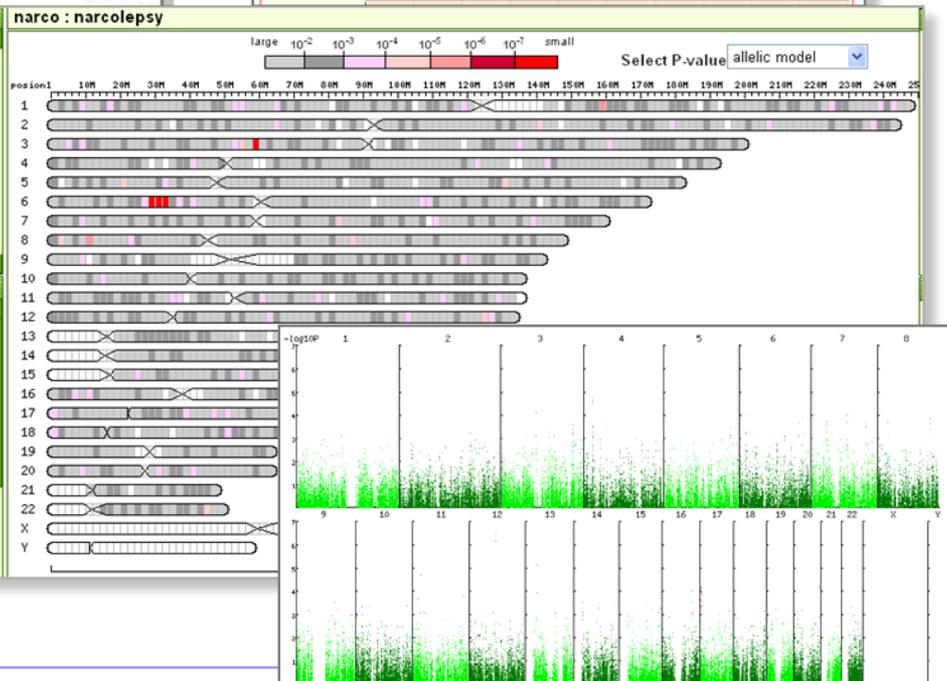
- cSNP
- eSNP
- rSNP
- iSNP
- qSNP

**Function:**

Update

**Choose Items**

- Chromosome Number
- Position
- SNP Type
- Gene Name
- AIC
- Entropy
- P-value
- Permutation test
- Odds Ratio
- Multiple testing corrections
- 95% CI
- MAF (case)
- MAF (control)
- HWE P-value (case)



# 2-3

# Case-Control DBの機能追加例

narco : narcolepsy

Chromosome 6 Position 30000001 . 32000000

[Study ID]

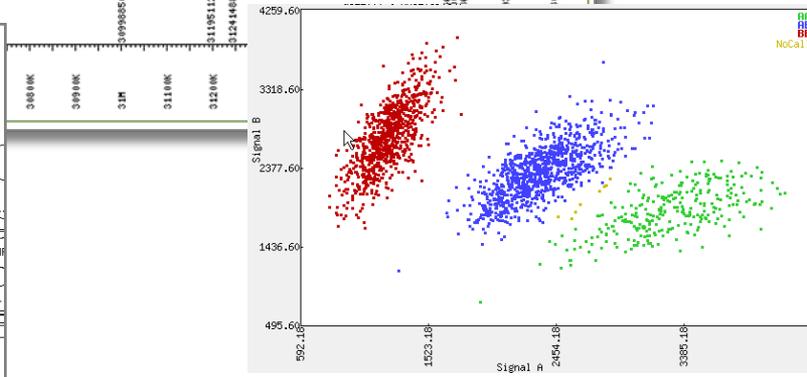
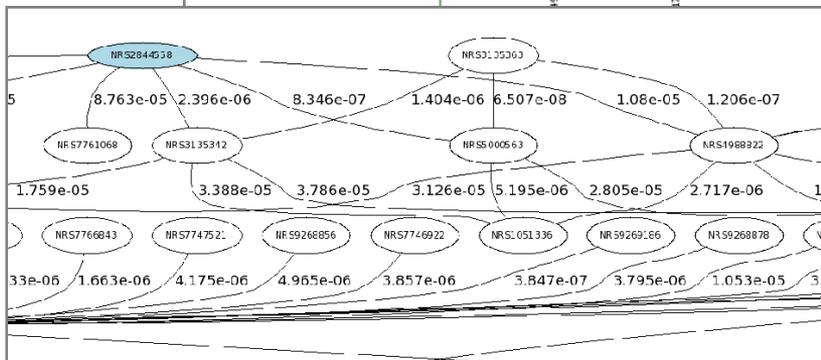
DL

SNP ID	Chr	Position	SNP Type	Gene Name	AIC	Allelic P-value	Genotypic P-value	Dom P-value	Rec P-value
NRS9261282	6	30143436	ISNP	PPP1R11	3.4280	0.2219	0.4749	0.2656	0.65
NRS9261301	6	30149538	ISNP	RNF39	-23.9060	9.495e-08	5.181e-07	8.82e-07	0.00
NRS2523990	6	30185208	ISNP	TRIM31	-7.2210	0.0009156	0.002192	0.03042	0.00
NRS9261471	6	30213328	ISNP	TRIM40	1.6410	0.1852	0.171	0.4167	0.07
NRS2857435	6	30214003	ISNP	TRIM40	2.3400	0.1734	0.2508	0.3165	0.17
NRS2857439	6	30214275	ISNP	TRIM40	-1.3270	0.07401	0.0365	0.2354	0.01
NRS9261485	6	30216730	ISNP	TRIM40	0.9960	0.1745	0.1255	0.4145	0.05
NRS9261488	6	30217391	ISNP	TRIM40	0.8200	0.146	0.1114	0.3656	0.05

DL

Studyの比較

SNP、遺伝子の検索



# 2-4

# Case-Control DBの機能追加例

**GWAS DATABASE**

Case Control GWAS

SEARCH

Case Control GWAS search

Disease Name List

Disease Related Gene List

**Disease Related Gene List**

ALL A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Panic disorder

**Related Gene list**

Disease Name : Panic disorder

Gene Symbol	SNP ID	Chr	Position Start-end	Pubmed ID ; Title ; Author	Judge	Methods	case m	case n	case mratio	cont m	cont n	cont mratio	P. value
TSPO	<a href="#">+ Show SNP ID</a>	22	41877478-41889191	PMID: 16511838 TITLE: Evidence that variation in the peripheral benzodiazepine receptor (PBR) gene influences susceptibility to panic disorder. AUTHOR: "Nakamura K, Yamada K, Iwayama Y, Toyota T, Furukawa A, Takimoto T, Terajama H, Iwahashi K, Takei N, Minabe Y, Sekine Y, Suzuki K, Iwata Y, Pillai A, Nakamoto Y, Ikeda K, Yoshii M, Fukunishi I, Yoshikawa T, Mori N."	0	resequencing	1	28	3.600	0	0	0.000	0
TSPO	<a href="#">+ Show SNP ID</a>	22	41877478-41889191	PMID: 16511838 TITLE: Evidence that variation in the peripheral benzodiazepine receptor (PBR) gene influences susceptibility to panic disorder. AUTHOR: "Nakamura K, Yamada K, Iwayama Y, Toyota T, Furukawa A, Takimoto T, Terajama H, Iwahashi K, Takei N, Minabe Y, Sekine Y, Suzuki K, Iwata Y, Pillai A, Nakamoto Y, Ikeda K, Yoshii M, Fukunishi I, Yoshikawa T, Mori N."	0	resequencing	1	28	3.600	0	0	0.000	0
TSPO	<a href="#">+ Show SNP ID</a>	22	41877478-41889191	PMID: 16511838 TITLE: Evidence that variation in the peripheral benzodiazepine receptor (PBR) gene influences susceptibility to panic disorder. AUTHOR: "Nakamura K, Yamada K, Iwayama Y, Toyota T, Furukawa A, Takimoto T, Terajama H, Iwahashi K, Takei N, Minabe Y, Sekine Y, Suzuki K, Iwata Y, Pillai A, Nakamoto Y, Ikeda K, Yoshii M, Fukunishi I, Yoshikawa T, Mori N."	0	resequencing	1	28	3.600	0	0	0.000	0

Genotype Counts (control)

Effective Individuals (case)

Effective Individuals (control)

2haplotype

3haplotype

Epistasis

Update Table

Update Table and Image

**Choose Graph**

Related Gene

OMIM

SNPs used by this Study

dbSNP

MicroSatellite

CNV

Update Table

Update Table and Image

DL

NRS1107554 22 40997538 iSNP hypothetical gene supported by BC03

NRS134898 22 41012501

panic: -log(p) vs position

panic: -log(p) vs position

panic: -log(p) vs position

panic: -log(p) vs position

Related Gene Odds Ratio vs position

TSPO

既知SNP/mutation 情報を登録・閲覧

# 3-1 CNV DBの表示例

Select Var Type: CNV Select Study: ALL

Position 10M 20M 30M 40M 50M 60M 70M 80M 90M 100M 110M 120M 130M 140M 150M 160M 170M 180M 190M 200M 210M 220M 230M 240M 250M

1  
2  
3  
4  
7  
8  
9  
10  
11  
12  
13  
14

Chromosome 16 Region 20500001 - 23000001 Show 2.5Mbp GO

var-1259.1.1  
var-1259.1.2  
var-1259.1.3

var-1260.1.1  
var-1260.1.2  
var-1260.1.3

Gene  
CSRL...  
RMP...  
CSRL...  
CINL...  
NRG3...  
P2  
NRS4...  
S17...  
URK1  
HES1...  
HGP31...

複製の計算データを一度に閲覧

クリックすると領域の図

高さが頻度  
色の濃さがCNV数

各パターンのgenotype

DGVのデータ

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	NR8150335:0A4G- NR8183170:2C2T- NR8636204:2G2T- NR8150336:2A2G- NR8574661:1A3G		
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-		

# 4-1 リシークエンスDB進捗と今年度の予定

## 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:

- ・ 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
- ・ 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)
- ・ その他に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



- ・ 統合データベースプロジェクト ヒトゲノム多型・変異データの公開・共有方針を、京都大(加藤和人先生)と中核機関の協力を得て作成済み  
データの 카테고리:
  - A. 頻度データ、統計解析結果
  - B. GWAS遺伝子型データ
  - C. GWAS生データ
  - D. リシーケンシングデータ
- ・ 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
- ・ 6月に倫理検討委員会にて検討

# 今年度の計画

- ・ CNVの検出手法の改良とCNVのcase-control DBの構築
- ・ 2次データの登録、表示やEpistasis表示の高機能化などのGWAS DBの改良
- ・ Epistasis検出の計算手法の開発
- ・ 新たに定める倫理規定およ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DBとの連携の検討
- ・ パーキンソン病リシークエンスDBの構築完了
- ・ 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