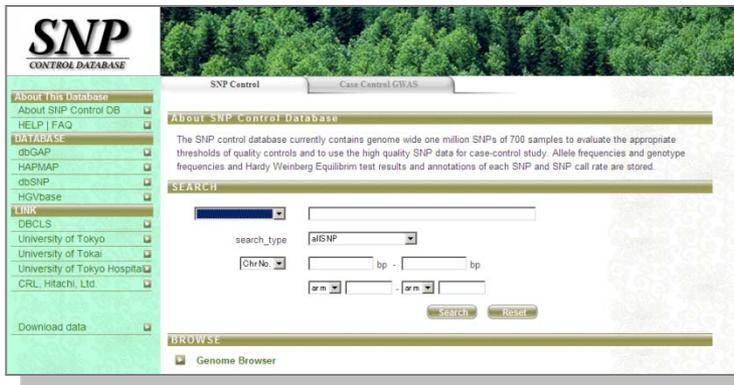


# 疾患解析から医療応用を実現する DB開発

東大医、東大病院、東海大学、日立製作所

# データベースのトップ画面

## 標準SNP database

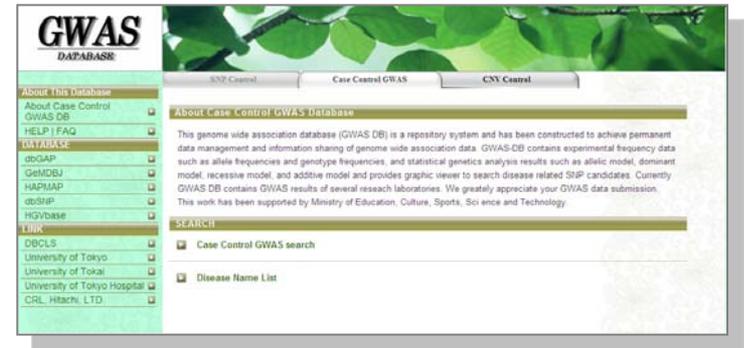


<https://gwas.lifesciencedb.jp/>



Mutation database

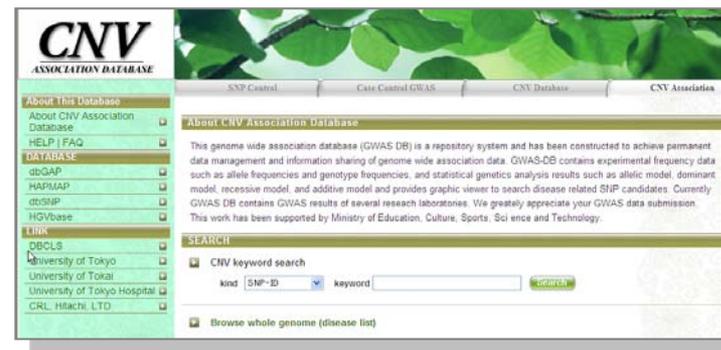
## Case-control database



## 標準CNV database



## CNV case-control database



# 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検体登録

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	Genotype detail			HWP	Allele	
				A/A	G/G	A/G		A	G
Illumina317K	Japanese	200	1.000	0.188	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

Array kind	Haplotype frequencies
Affy500K	NRS12563491- NRS9424283- NRS7543006- NRS2154068- NRS6702916- NRS6702935- NRS6703035- NRS6663840- NRS9424310- NRS17403773- NRS2298225- NRS2298224- NRS17404435- NRS6683156;
	AAAAATAGCAAAC 0.403
	GGAGGCGAAAAATC 0.393
	AGAGATGGGAAGTC 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:

- 1)実験結果・解析結果のDB化による **半永続的な集約的データ管理**
- 2)GWAS以外の研究者にも情報を提供することによる  
**疾患研究の促進** (計算結果と頻度情報のみ公開)
- 3)研究者間の情報の共有による**遺伝統計研究の促進**

## コンテンツ:

- ・ 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)

## 現在:

- ・ 18疾患 (25 studies) を内部用DBに登録しており、そのうち9疾患 (10 studies)を公開中。 17量的形質 (17 studies)を内部用DBに登録。
- ・ 機能追加部分を実装中



# 1-5

## Case-control DBの画面例

**Filtering**

Quality:

- MAF (control) 0.05 and above
- MAF (case) 0.05 and above
- HWE P-value 0.0001 and above
- Call Rate (control) 0.95 and above
- Call Rate (case) 0.95 and above

SNP Type:

- SNP
- InDel
- STR
- CNV

Choose items:

- SIDAK\_SD
- FDR\_BH
- FDR\_BY
- Odds Ratio
- OR: Allelic model
- OR: Genotypic model
- OR: Dominant model
- OR: Recessive model
- OR: Additive model

narco : narcolepsy

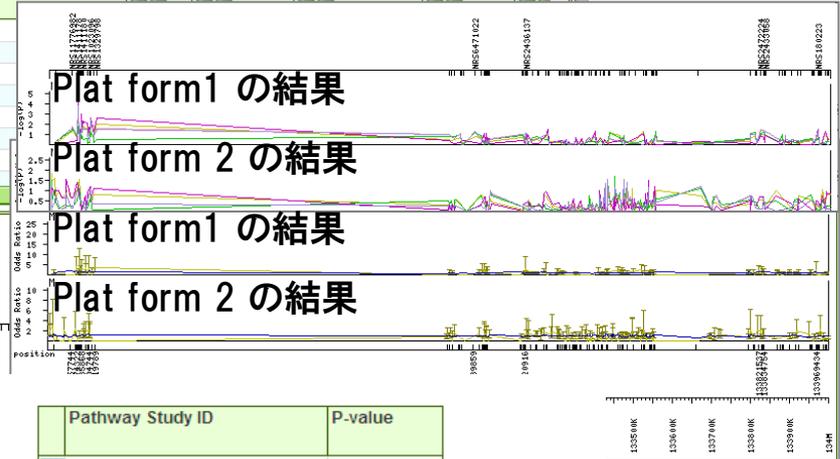
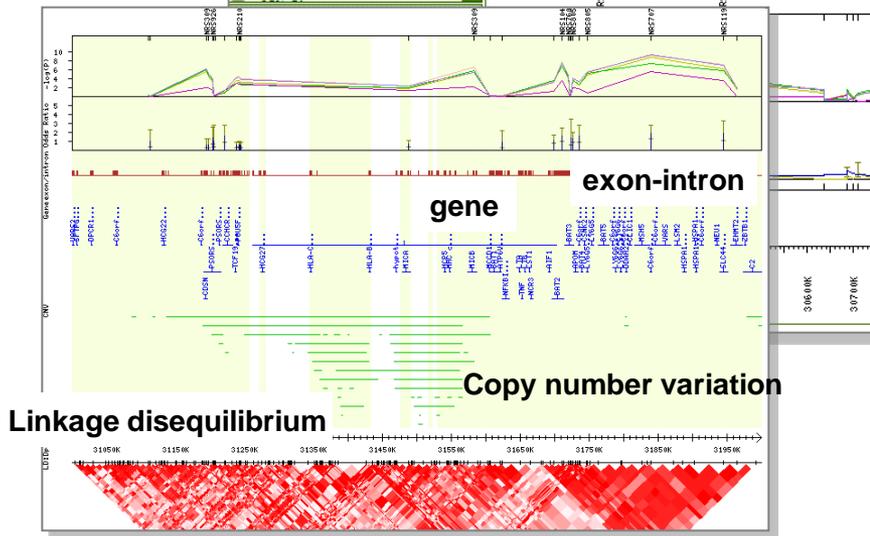
Chromosome 6 Position 30000001 - 32000000

[Study ID]

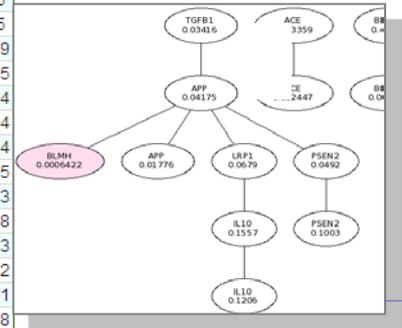
Studyの比較

SNP、遺伝子の検索

SNP ID	Chr	Position	SNP Type	Gene Name	AIC	Allelic P-value	Genotypic P-value	Dom P-value
NRS9261282	6	30143436	ISNP	PPP1R11				
NRS9261301	6	30149538	ISNP	RNF39				
NRS2523990	6	30185208	ISNP	TRIM31				
NRS9261471	6	30213328	ISNP	TRIM40				
NRS2857435	6	30214003	ISNP	TRIM40				
NRS2857439	6	30214275	ISNP	TRIM40				
NRS9261485	6	30216730	ISNP	TRIM40				
NRS9261488	6	30217391	ISNP	TRIM40				

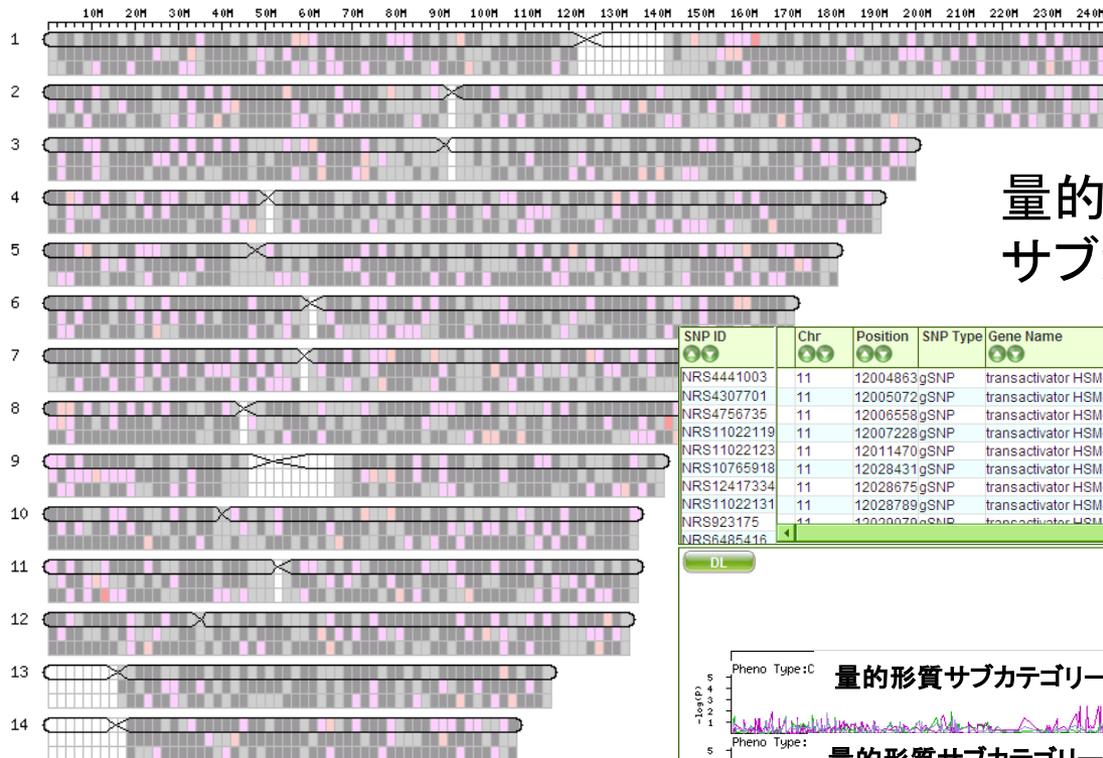


Pathway Study ID	P-value
Glypican 1 network	4.2251e-07
Glypican pathway	6.84988e-07
TRAIL signaling pathway	4.5801e-06
"Cell Cycle, Mitotic"	7.1857e-06
Metabolism of carbohydrates	4.43064e-05
IFN-gamma pathway	4.98883e-05
Glucose metabolism	7.16107e-05
Glucose uptake	0.000125719
Class I PI3K signaling events	0.000162725
Regulation of cytoplasmic and	0.000242294
TGF-beta receptor signaling	0.000242294
Regulation of nuclear SMAD2/3	0.000242294
TNF receptor signaling pathway	0.000390605
Signaling in Immune system	0.000451453
Glucose is carried across the	0.000491088
IL-1-mediated signaling events	0.006147643
Class A/1 (Rhodopsin-like rece	0.008425562
M Phase	0.008759591
Signalling by NGF	0.009839738

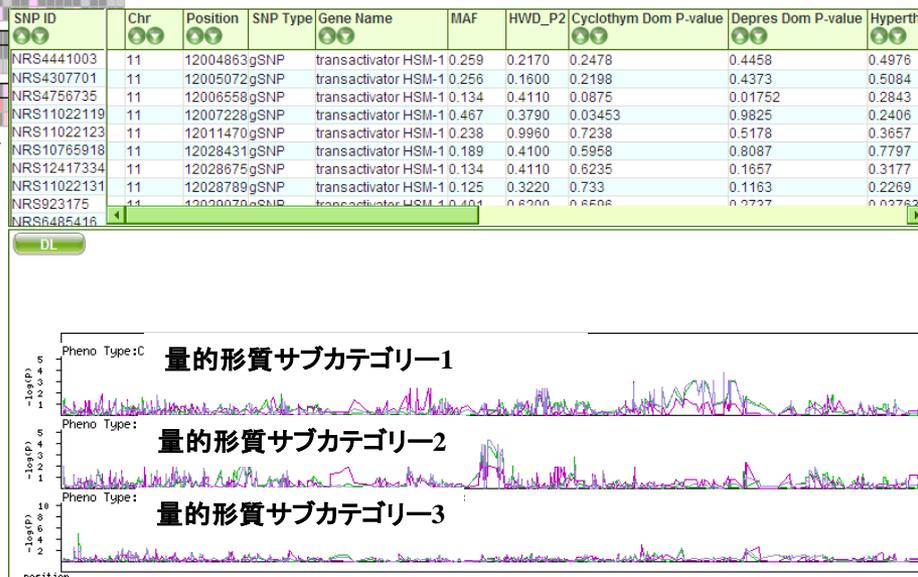


# 1-6

## 量的形質の関連解析結果の表示



量的形質解析結果の  
サブカテゴリー並列表示例

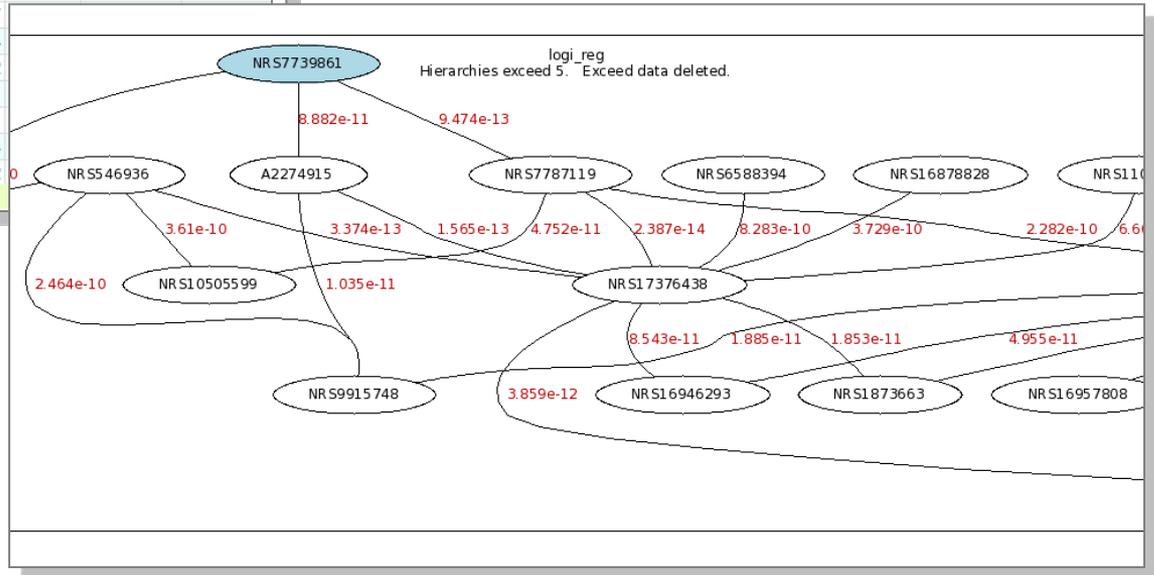


# 1-7

## Epistasisの登録と表示

Int Kind	Gene A/SNP-A	Chr	Pos	Gene B/SNP-B	Chr	Pos	Weight
logi_reg	NRS7787119	7	43687934	NRS17376438	10	77725266	2.387e-14
logi_reg	A2274915	9	14250138	NRS17376438	10	77725266	1.565e-13
logi_reg	NRS546936	6	44757114	NRS17376438	10	77725266	3.374e-13
logi_reg	NRS7739861	6	667716	NRS7787119	7	43687934	9.474e-13
logi_reg	NRS17824132	8	73317985	NRS16957808	13	100303160	1.546e-12
logi_reg	NRS17824132	8	73317985	NRS3902307	12	100045651	3.694e-12
logi_reg	NRS17376438	10	77725266	NRS6035066	20	18464588	3.859e-12
logi_reg	NRS7787119	7	43687934	NRS6008536	22	46842252	4.935e-12
logi_reg	NRS7787119	7	43687934	NRS6085297	20	5791479	5.276e-12
logi_reg	NRS1945652	11	80905066	NRS16957808	13	100303160	9.108e-12
logi_reg	NRS17824132	8	73317985	NRS17102227	14	64391947	9.725e-12
logi_reg	A2274915	9	14250138	NRS9915748	17	51518243	1.035e-11
logi_reg	NRS7787119	7	43687934	NRS9915748	17	51518243	1.167e-11
logi_reg	NRS1945652	11	80905066	NRS11090824	22	47038271	1.305e-11
logi_reg	NRS1945652	11	80905066	NRS17102227	14	64391947	1.305e-11
logi_reg	NRS17824132	8	73317985	NRS16956024	13	100303160	1.305e-11
logi_reg	NRS4682115	3	113923788	NRS17824132	8	73317985	1.305e-11
logi_reg	NRS17376438	10	77725266	NRS1873663	20	18464588	1.305e-11
logi_reg	NRS17824132	8	73317985	NRS9915748	17	51518243	1.305e-11
logi_reg	NRS7787119	7	43687934	NRS16956024	13	100303160	1.305e-11
logi_reg	NRS4700811	5	178966007	NRS17824132	8	73317985	1.305e-11

### Logistic regressionをはじめとした Epistasisの登録、表示



**目的:** 遺伝統計以外の人も解析結果を有効利用できるCNV DBの構築

**種類:**

- CNV control DB: 健常者のCNV DB
- CNV case control DB: CNVのcase control 解析

**表示:**

- 上記の値を、テーブルとグラフで表示
- 開始、終止位置が微妙に異なるCNVも多く検出されるため、類似CNVをクラスタリングしてマージし、CNVのパターン一覧を見やすくしたオプションも用意

**現在:**

- CNV control DBには健常者の約160名のデータを登録・公開中
- CNV case control DBには、内部用DBに5疾患登録



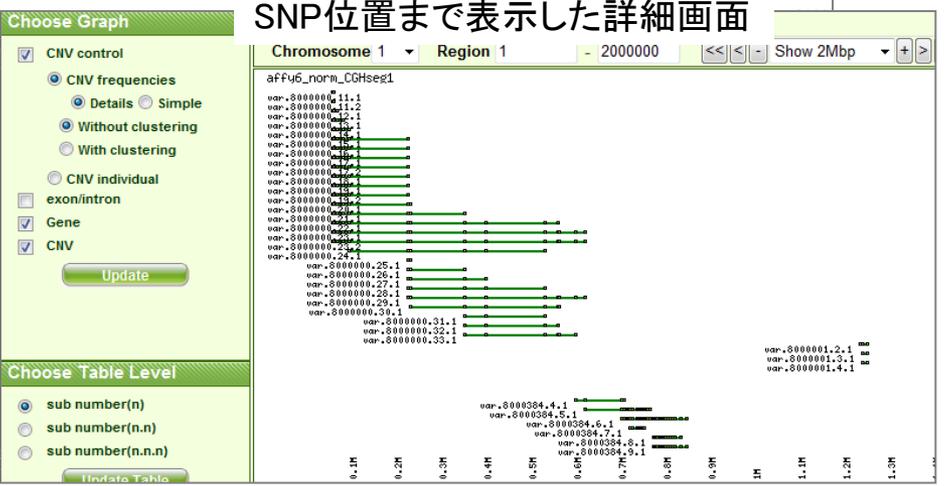
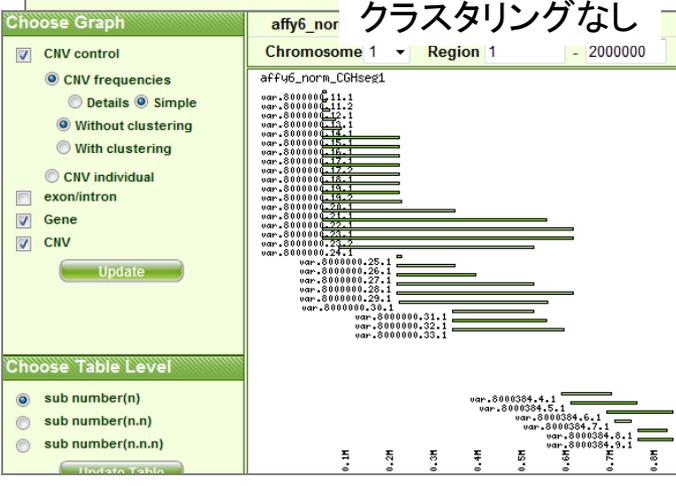
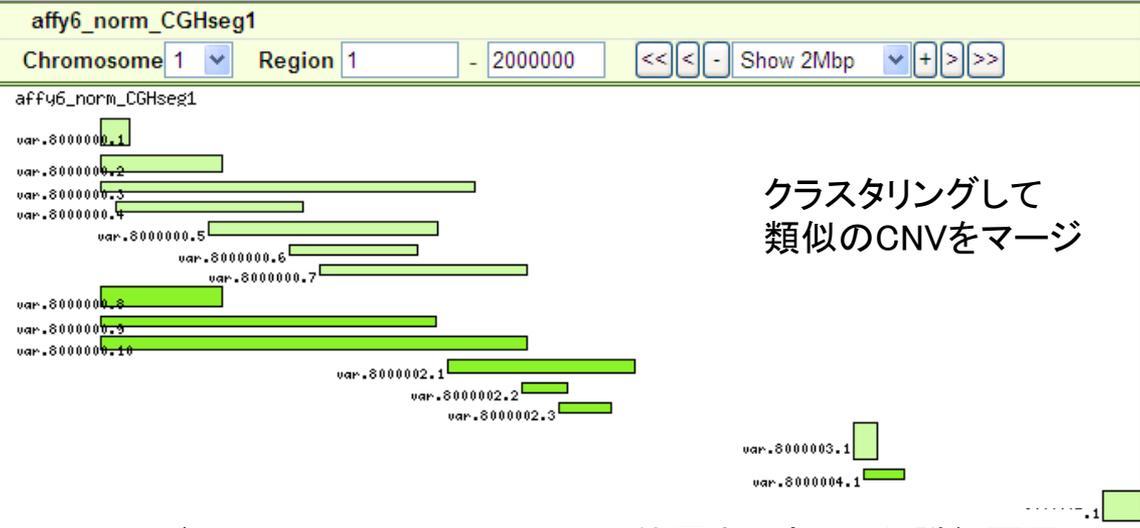
# 1-10 CNVのデータベースの表示例

**Choose Graph**

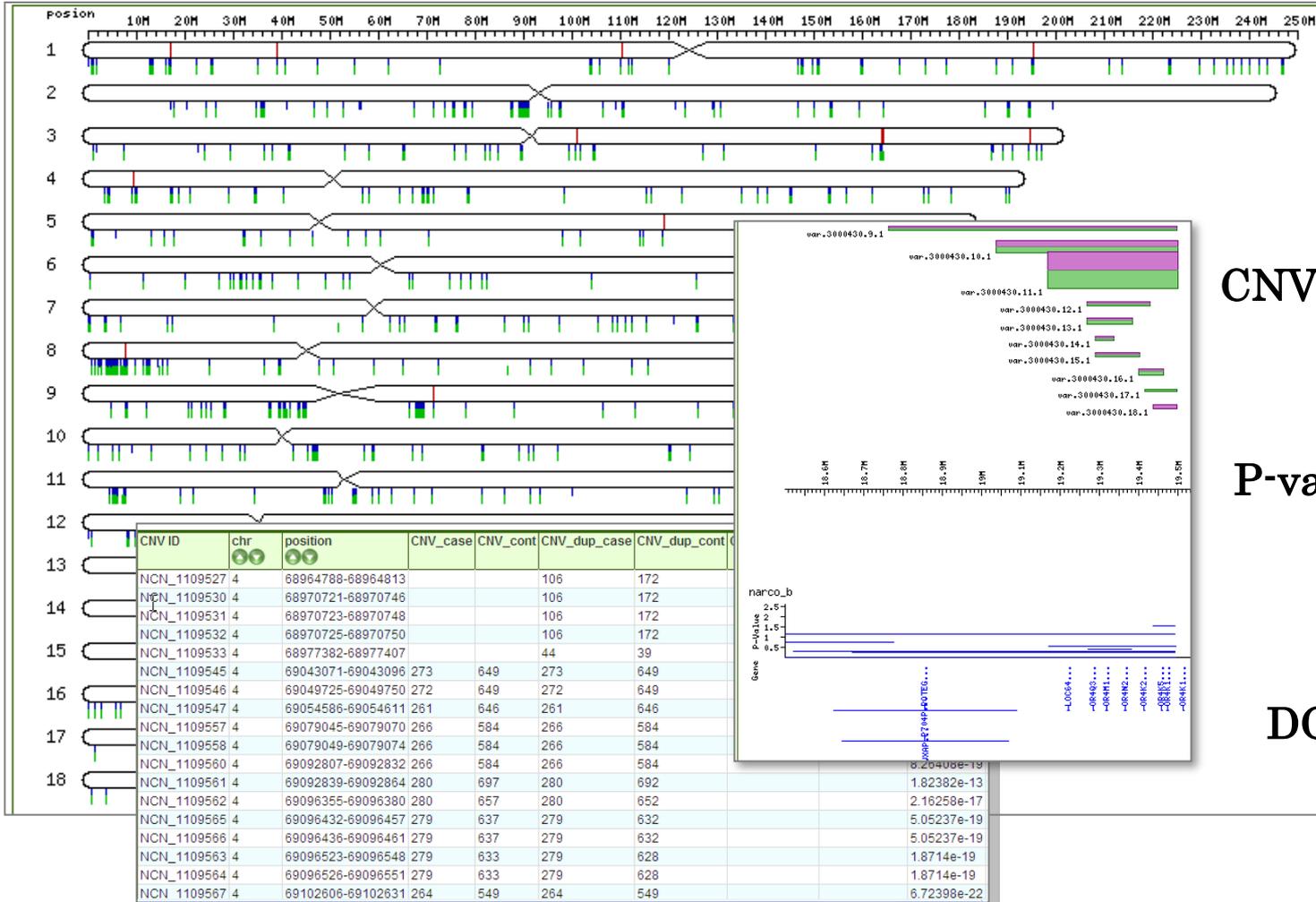
- cnv control
- Details  Simple
- Without clustering
- With clustering
- exon/intron
- Gene
- CNV

Update

表示オプションの変更



# 1-11 CNV case control DBの表示例



CNV region

P-values

DGV

# 1-12 データの受け入れと再配布

- 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データ共有検討委員会で決定。
- データ共有審査委員会を発足し、適切な倫理審査の手続きを行った上で、研究者からのデータのリポジトリを行うとともに、データの利用申請者への再配布を行うしくみを開始。

Integrated Database  
DB development for medical applications from disease analysis

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

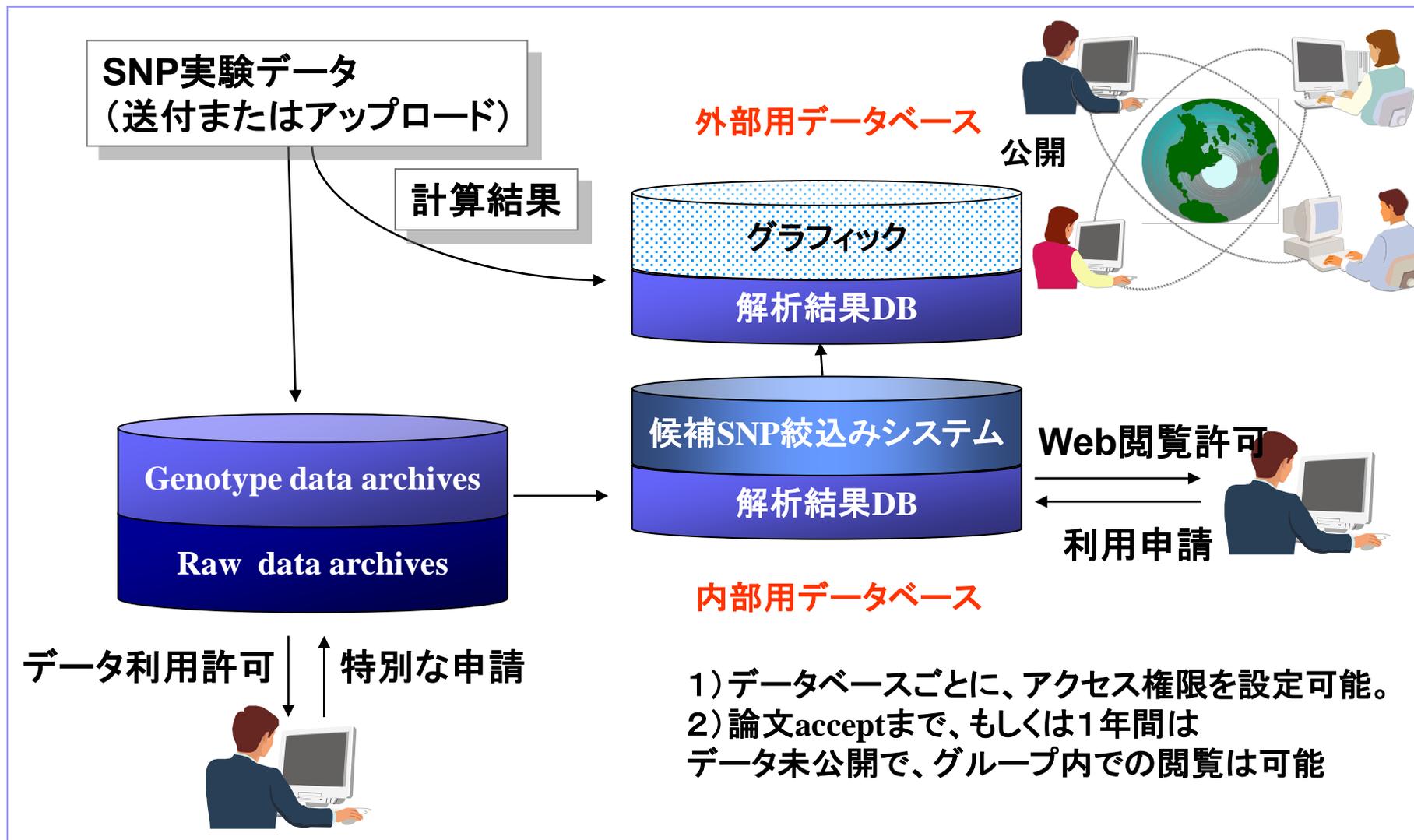
現在、疾患関連遺伝子の探索においては非常に多くでヒト遺伝子に関する情報が増加しており、これらの大量データの有効活用が急務である。文部科学省委託研究開発事業「統合データベースプロジェクト」はこのようなデータが広く共有されることを目的としており、様々なデータに即けて協働を進めている。本方針は、「統合データベースプロジェクト・疾患解析DBグループ」(代表機関 国立大学法人東京大学) (以下、「疾患解析DBグループ」) におけるゲノムワイドな多型情報および関連解析情報(GWAS) および疾患関連遺伝子のリークエーションデータについての共有方針について述べたものである。データの内容に応じて以下のカテゴリーのデータを3つのレベルに分類し、それぞれのレベルに対する共有方針を定める。なお、データのデータベースへの提供とアクセスについては、「疾患解析DBグループ」内にデータ共有審査委員会(以下、「データ共有審査委員会」)を設置し審査を行う。委員会メンバーには「疾患解析DBグループ」の外側の委員を招聘し定めることとし、科学者・倫理面・その他の観点から検討を行う。

<本共有方針の概要>  
データのカテゴリー/レベルとデータ使用時の手続き等について

レベル	1	2	3
データの種類	カテゴリー1(臨床データ、解析結果)	カテゴリー2 (臨床レベルでのGWAS情報)	カテゴリー3 (GWAS関連子型データ)
リークエーションデータ	カテゴリー1 (公知のリークエーションデータおよび発見データ)	カテゴリー2 (リークエーションデータおよび発見データ)	(提供なし)
データ使用時の手続き	ウェブサイトで、閲覧可能。但しカテゴリー1のデータを複製する場合はレベル2と同様の申請を要する。	共有、閲覧、連絡先、使用目的、e-mailアドレス(原則、所属機関から発行されたアドレス)を申請する。	データアクセス申請書を提出し、許可を受ける。また、使用期間に応じ、データ使用報告書の提出を要する。

データの安全な取り扱いについて

# 1-13 Repository systemの概要



# 2-1 Resequencing DB (mutation DB) の概要

- ・ 神経性疾患のMutation DB目的:
  - 1) 遺伝子変異と臨床情報の両方を俯瞰
  - 2) 疾患の機序の理解
  - 3) 将来的には臨床現場で役に立つDBに

- ・ 現在公開している疾患:

- ALS (Amyotrophic lateral sclerosis) database  
筋萎縮性側索硬化症
- PD (Parkinson disease)
- ALD (Adrenoleukodystrophy disease)  
副腎白質ジストロフィー
- HSP (Hereditary spastic paraplegia)  
家族性痙性対麻痺

The image displays two screenshots of mutation databases. The top screenshot is for the ALD mutation database, showing a navigation menu with options like 'About mutation database', 'Causative genes', 'Related genes', and 'Link'. The bottom screenshot is for the HSP mutation database, showing a similar navigation menu and a list of causative genes. A text box on the right provides an 'About HSP mutation database' description.

**ALD mutation database**

**HSP mutation database**

**About HSP mutation database**

Hereditary spastic paraplegia (HSP) is a group of inherited degenerative disorders that are characterized by progressive spasticity in the lower limbs. The mechanisms that cause HSP are not fully clarified. The aim of this database is to collect mutations related to HSP and their clinical information exhaustively in order to deepen our understanding of HSP. Both our original data and extracted data from published papers are accumulated in this database.

**Causative genes**

- **NIPA1** non imprinted in Prader-Willi/Angelman syndrome 1
- **SPG20** spastic paraplegia 20 (Troyer syndrome)
- **BSCL2** Bernardinelli-Seip congenital lipodystrophy 2 (scopin)
- **HSPD1** heat shock 80kDa protein 1 (chaperonin)
- **KIF5A** kinesin family member 5A
- **L1CAM** L1 cell adhesion molecule

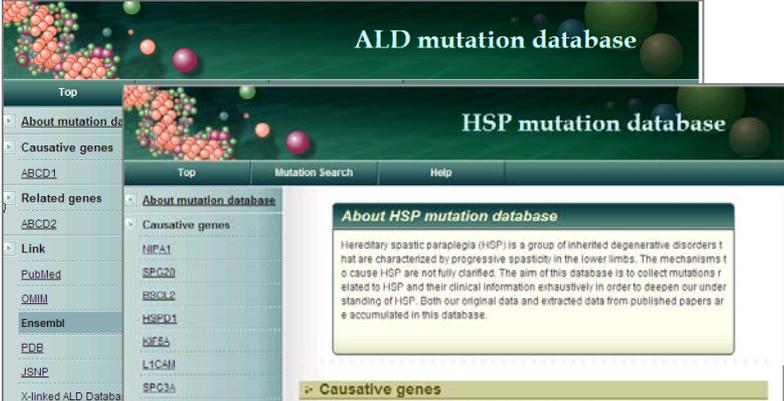


- ALS contents
  - 107 文献から638エントリーを抽出 (178 unique variants)
    - familial ALS / sporadic ALS: 70% entries are related to familial ALS
  - 新たに実験的に発見した変異+臨床情報
  - サブミッションしてもらった変異+臨床情報
- PD contents
  - 222 文献から1508エントリーを抽出 (472 unique variants)
  - 新たに実験的に発見した変異+臨床情報
- HSP contents
  - 103 文献から988エントリーを抽出(347 unique variants)
- ALD contents
  - 49 文献から274エントリーを抽出(157 unique variants)
  - サブミッションしてもらった変異+臨床情報

## 2-4 Mutation and clinical informationの項目

- Genomic position of variations
  - mRNA/amino acid sequence accession including the variation position
  - Homo/Hetero
  - Num of patients with/without mutations
  - Num of controls with/without mutations
  - Rate of patients and controls
  - Odds ratio, CI, chi-square P-value
  - Clinical characteristics
  - Sex
  - Age on set
  - Duration
  - Disease type
  - Onset site
  - Years until initiation of respirator
- 変異表記はHGVS nomenclature  
に従った表記と原論文での表記  
の併用
- Clinical information terms  
are different in among diseases

# 2-5 リシークエンスDBの画面例



**ALD mutation database**

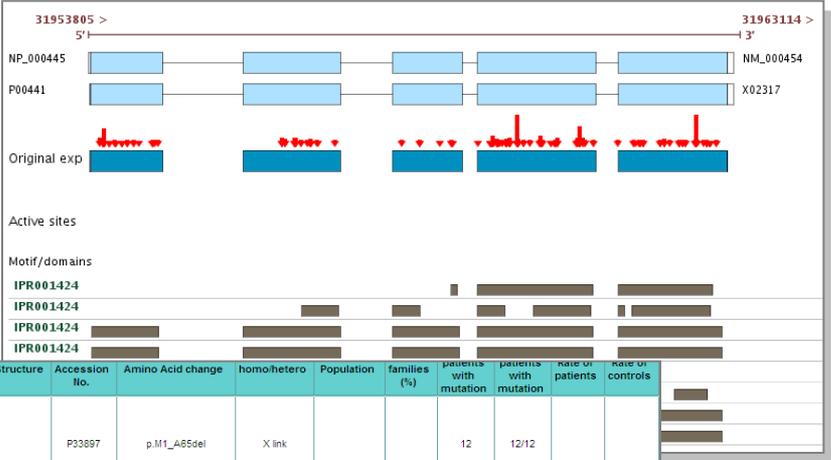
**HSP mutation database**

**About HSP mutation database**

Hereditary spastic paraplegia (HSP) is a group of inherited degenerative disorders that are characterized by progressive spasticity in the lower limbs. The mechanisms that cause HSP are not fully clarified. The aim of this database is to collect mutations related to HSP and their clinical information exhaustively in order to deepen our understanding of HSP. Both our original data and extracted data from published papers are accumulated in this database.

**Causative genes**

- ABCD1
- ABCD2
- SPG20
- BGCL2
- HSPD1
- HPS5
- L1CAM
- SPG3A



31953805 > 5' 3' 31963114 >

NP\_000445 NM\_000454

P00441 X02317

Original exp

Active Sites

Motif/domains

IPR001424

IPR001424

IPR001424

a.a.: P33897

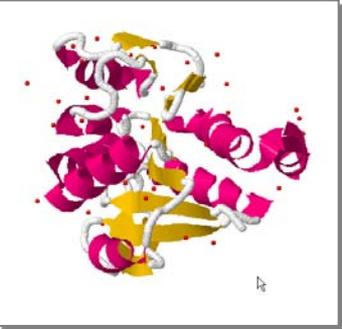
```

1  MPVLSRPFWRGNTLKRTAVLLALAAAYGAHKVYVPLVRQCLAPARGLQAPA
51  GEPTQEASGVAAAKAGMNRVFLQRLLWLLRLLFPRVLCRETGLLALHSAA
101 IVSRTELSVYVARLDGRLARCIVRKDPRAFGWQLLQWLLIALPATFVNSA
151 IRYLEGQLALSFRSRLVAHAYRLYFSQQTYRVSNMHGRLRNPDQSLTED
201 VVAFAAASVAHLYSNLTKPLLDVAVTYSYLLRAARSRGAGTAWPSAIGLV
251 VFLTANVLRAPFKPGELVAEEARRKGGELRYMHSRVVANSSEEIFYGGHE
301 VELALLQRSYQDLASQINLILLERLRYVVMLEQFLMKYVWSASGLLHVAVP
351 IITATGYSSEDAEAVKKAALKKEEELVSRTEAFTIARNLLTAADATE
401 RIMSSYKVEVELAGYTVARVHEMFQVFEDVQRCHFKRPRELEDAQAGSGTI
    
```

mRNA: BC015541

```

1  GGACGCGCCTGGTGCCCGGGGAGGGGGCGCCACCGGGGAGGAGGAGGAG
51  GAGAAGGTGGAGGAGGAAGAGACGCCCCCTCTGCCCGAGACCTCTCAAGGC
101 CCTGACCTCAGGGGCCAGGGCAGTACAGGACAGGAGAGCCAAAGTTCCTC
151 CACTTGGGCTGCCCGAAGAGGGCCGGCCCTGGAGGGCCCTGAGCCACC
201 GCACCAAGGGCCCCAGCACCCCGGGGGCCATAAAGCGACAGTCTCAGG
251 GGCCATCGCAAGGTTTCCAGTTGCCTAGACAAACAGGCCAGGGTCAGAGC
301 AACAACTCTCCAGCCACTGCTCACTGCTGCCCCAGGCCAGCCAGCCCC
351 AGTCCTCAGCGGCAGCCAGCCAGGTGACATGCCGGTGTCTCCAGGCC
401 CCGGCCCTGGCGGGGAACACGCTGAAGCGCACGGCGTGTCTCGGCC
    
```



(original description)	Structure	Accession No.	Amino Acid change	homo/hetero	Population	families (%)	patients with mutation	patients with mutation	rate of patients	rate of controls
M1_A65del		P33897	p.M1_A65del	X link			12	12/12		
M1_A65del		P33897	p.M1_A65del	X link						
W10X		P33897								

Age on set	Onset duration (y/m)	Examination age	Disease type	Initial symptom	Onset site	Reference (title)	journal, year, volume, page)	Comments
			AMN			<a href="#">PMID: 11738005</a> - ABCD1 translocation-inhibitor mutation demonstrates genotype-phenotype correlation for AMN.	<a href="#">O'Neill GN, Aoki M, Brown RH Jr. Neurology. 2001; 57:1956-62.</a>	personal data in same family(A) below
late 10's		40	AMN	stiffness	leg	<a href="#">PMID: 11738005</a> - ABCD1 translocation-inhibitor mutation demonstrates genotype-phenotype correlation for AMN.	<a href="#">O'Neill GN, Aoki M, Brown RH Jr. Neurology. 2001; 57:1956-62.</a>	(A)
			eALDIAMN			<a href="#">PMID: 8551200</a> - Mutational and protein analysis of patients and heterozygous women with X-linked adrenoleukodystrophy.	<a href="#">Feigenbaum V, Lombard-Pandolf G, Guillon S, Sarde CO, Mandel JL, Aubourg P. Am J Hum Genet. 1996; 58:1135-44.</a>	
			ccALD			<a href="#">PMID: 7825602</a> - Spectrum of mutations in the gene encoding the adrenoleukodystrophy protein.	<a href="#">Ligtenberg MJ, Kemp S, Sarde CO, van Geel BM, Kijner WJ, Barni PC, Mandel JL, van Oost BA, Bohule PA. Am J Hum Genet. 1996; 58:44-50.</a>	