

# ヒトゲノムバリエーション データベースの進展

東京大学大学院医学系研究科  
徳永勝士

共同研究機関

東京大医附属病院、国立遺伝学研究所、日立製作所

辻 省次

井ノ上逸郎

小池麻子



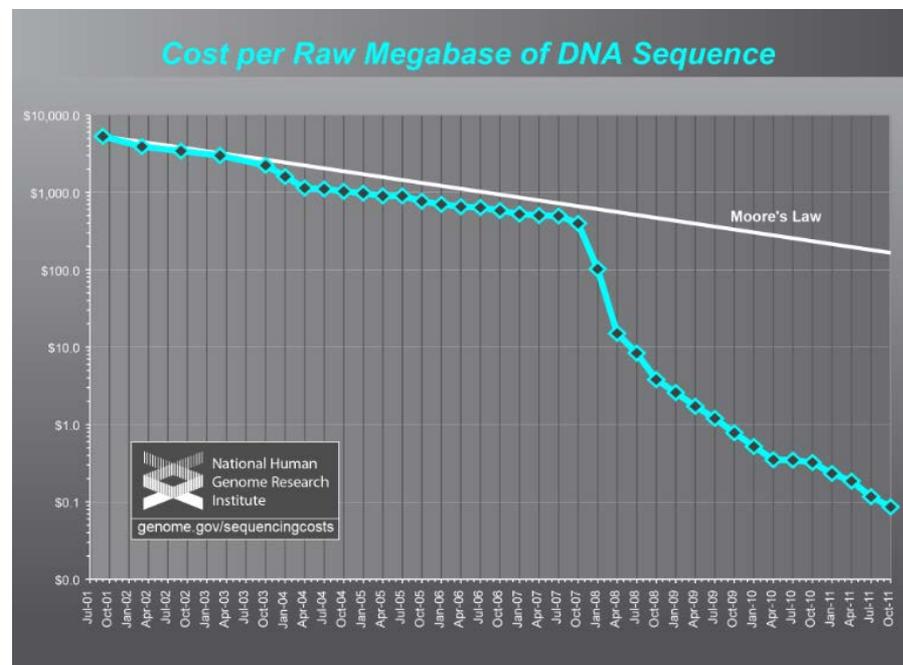
# ゲノムワイド関連解析と次世代シーケンサー解析

HapMap、dbSNPの充実やSNPタイピング技術の飛躍的向上により大規模なゲノムワイド関連解析が多数進行



Published Genome-Wide Associations through 12/2012  
>1,400 published GWA reports at  $p \leq 5 \times 10^{-8}$

次世代シーケンサーによる解析の低価格化・大容量化により、Exome、WGSでの疾患遺伝子探索が興隆



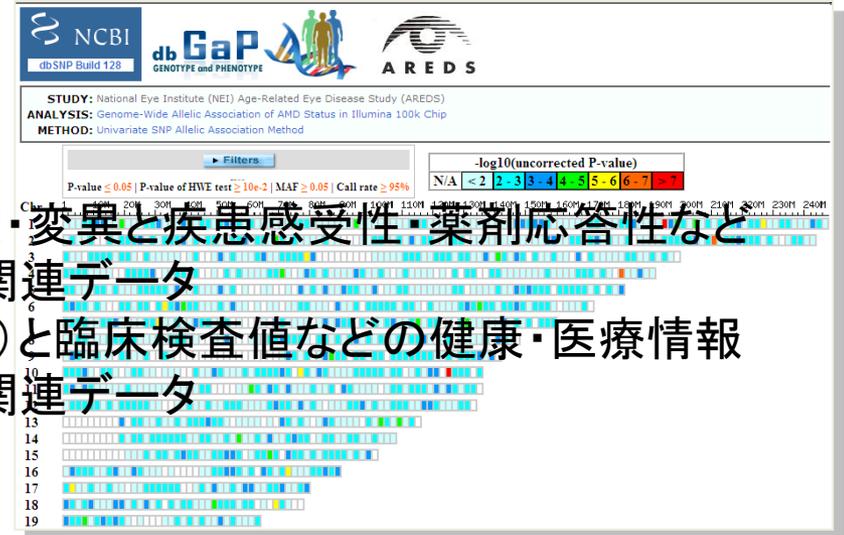
ゲノム解析コストの推移

大量データの管理・共有化の問題

# 海外でのGWAS/NGSデータ共有化

## 1) 米国NCBI

- dbSNP SNP情報を蓄積
- dbVAR 構造多型のデータを蓄積
- dbGaP GWAS、NGS結果を含むゲノム多型・変異と疾患感受性・薬剤応答性などを対象としたgenotype-phenotype関連データ
- ClinVar ゲノム多型・変異(dbSNP, dbVAR)と臨床検査値などの健康・医療情報を対象としたgenotype-phenotype関連データ



## 2) ヨーロッパEBI

- EGA GWAS、NGS結果を含むゲノム多型・変異と疾患感受性・薬剤応答性などを対象としたgenotype-phenotype関連データ



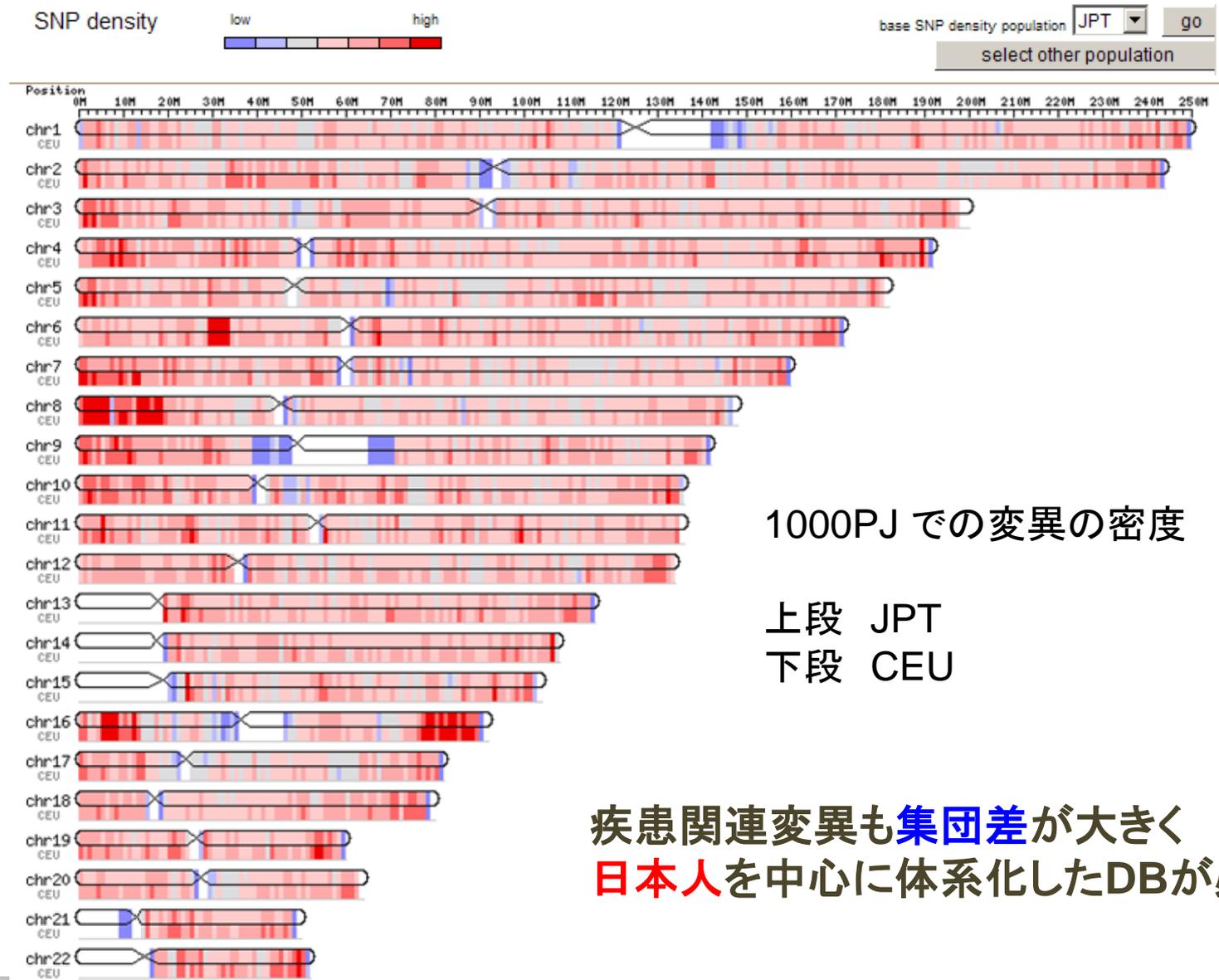
## 3) HGVS (human genome variation society)

- LSDB (locus specific database) 各種遺伝病の原因変異データ

## 4) その他、PJ単位のDB

- 1000 genome PJ
- 国際がんゲノムコンソーシアムなど

# Human Variation DB – 分布の集団間比較



# 原発性胆汁性肝硬変 (PBC) に見る 日本人とヨーロッパ系集団の共通性と異質性

日本人でも再現された感受性遺伝子

Gene	OR	P value
<b>HLA-DQB1</b>	<b>1.94</b>	<b>8.30 × 10<sup>-13</sup></b>
<b>IL7R</b>	<b>1.47</b>	<b>4.13 × 10<sup>-4</sup></b>
<b>IKZF3</b>	<b>1.44</b>	<b>6.77 × 10<sup>-4</sup></b>
<b>STAT4</b>	<b>1.35</b>	<b>8.11 × 10<sup>-4</sup></b>
<b>RAD51L1</b>	<b>1.07</b>	<b>0.0023</b>
<b>CXCR5</b>	<b>1.42</b>	<b>0.0027</b>
<b>PLCL2</b>	<b>1.07</b>	<b>0.0030</b>
<b>rs6974491</b>	<b>1.33</b>	<b>0.010</b>
<b>IL12RB2/SCHIP1</b>	<b>1.12</b>	<b>0.013</b>
<b>NFKB1</b>	<b>1.35</b>	<b>0.019</b>
<b>CD80</b>	<b>1.48</b>	<b>0.022</b>
<b>DENND1B</b>	<b>1.14</b>	<b>0.046</b>

日本人で見出された遺伝子

<b>TNFSF15</b>	<b>1.56</b>	<b>2.84 × 10<sup>-14</sup></b>
<b>PAU2AF1</b>	<b>1.39</b>	<b>2.38 × 10<sup>-8</sup></b>

日本人で再現されなかった遺伝子

Gene	OR	P
<b>IRF5/TNPO3</b>	<b>1.08</b>	<b>0.068</b>
<b>rs11117432</b>	<b>1.21</b>	<b>0.10</b>
<b>SPIB</b>	<b>1.08</b>	<b>0.10</b>
<b>RPS6KA4</b>	<b>1.12</b>	<b>0.10</b>
<b>CLEC16A</b>	<b>1.10</b>	<b>0.15</b>
<b>TNFRSF1A</b>	<b>1.12</b>	<b>0.16</b>
<b>IL12A</b>	<b>1.02</b>	<b>0.37</b>
<b>MAP3K7IP1 (TAB1)</b>	<b>1.29</b>	<b>0.43</b>
<b>MMEL1</b>	<b>1.07</b>	<b>0.78</b>
<b>TNFAIP2</b>	<b>1.22</b>	<b>0.98</b>

Th, B, Immune system

# ヒトゲノムバリエーションDBの目的

変異・疾患・臨床情報を整理・体系化し、成果・情報を俯瞰可能とすると共に、健常者のゲノム多様性情報を提供する

→ 効率的な疾患遺伝子の探索に役立てる

実施内容:

次世代シーケンサーおよび、その他の解析法(GWASを含む)によって新規発見される多型・変異情報の受け入れと研究者間の情報共有

文献情報を含め過去に産出された疾患感受性、薬剤反応性などに関わる多型・変異情報の収集とDB化

HLAハプロタイプごとの変異を登録し、HLA遺伝子群の多様性と疾患感受性、薬剤過敏症などの関係を俯瞰可能に

健常者データについては、phasingやハプロタイプ推定、必要に応じて1000 genome PJデータ、GWAS 健常者データも用いて遺伝子型推定を行い、SNP, in/del, CNVなど各種多型・変異のアリル頻度、ハプロタイプ頻度を計算・公開

# 開発データベース

**Human Variation DB**

Human Variation DB | HLA Database | SNP Control | Case Control GWAS | CNV Database

### HLA DATABASE

About This Database  
About HLA Database  
HELP | FAQ

DATABASE  
dbGAP  
GeMDBJ  
JSNP  
HAPMAP  
dbSNP  
HGvbase

LINK  
HGVRD top-page  
HGVRD data sharing policy  
NBDC  
DBCLS  
University of Tokyo  
National Institute of Genetics  
University of Tokyo Hospital  
CRL, Hitachi, LTD.

SEARCH

Search by gene name  
keyword |

Search by disease name  
keyword |

Search by genomic position  
Chro. |  Region |

Search by SNP number

SEARCH & BROWSE

Study info

Sequences

Multiple alignment Japanese Haplotype

Select multiple studies  
Select reference study

Base level view | HLA-A |

Pair-alignment

First |  | Second |  |

Multi-fasta view  
Select multiple studies

NGSなど  
対応DB

**SNP CONTROL DATABASE**

Human Variation DB | About SNP Control DB | HELP | FAQ

DATABASE  
dbGAP  
GeMDBJ  
JSNP  
HAPMAP  
dbSNP  
HGvbase

LINK  
HGVRD top-page  
HGVRD data sharing policy  
NBDC  
DBCLS  
University of Tokyo  
National Institute of Genetics  
University of Tokyo Hospital  
CRL, Hitachi, LTD.

DOWNLOAD data

**GWAS DATABASE**

Human Variation DB | About Case Control GWAS DB | HELP | FAQ

DATABASE  
dbGAP  
GeMDBJ  
JSNP  
HAPMAP  
dbSNP  
HGvbase

LINK  
HGVRD top-page  
HGVRD data sharing policy  
NBDC  
DBCLS  
University of Tokyo  
National Institute of Genetics  
University of Tokyo Hospital  
CRL, Hitachi, LTD.

SEARCH & BROWSE

Case co  
Browse

Disease  
Browse

Genome  
Browse

**CNV ASSOCIATION DATABASE**

Human Variation DB | HLA Database | SNP Control | Case Control GWAS | CNV Database | CNV Association

About CNV Association Database

This copy number variation association database (CNV association DB) is a repository system and has been constructed to achieve permanent data management and information sharing of CNV data. CNV-DB contains CNV region data. Currently CNV DB contains CNV results of one research laboratory. We greatly appreciate your CNV association data submission. This work has been supported by Ministry of Education, Culture, Sports, Science and Technology.

SEARCH & BROWSE

CNV keyword search  
kind [SNP-ID] | keyword |

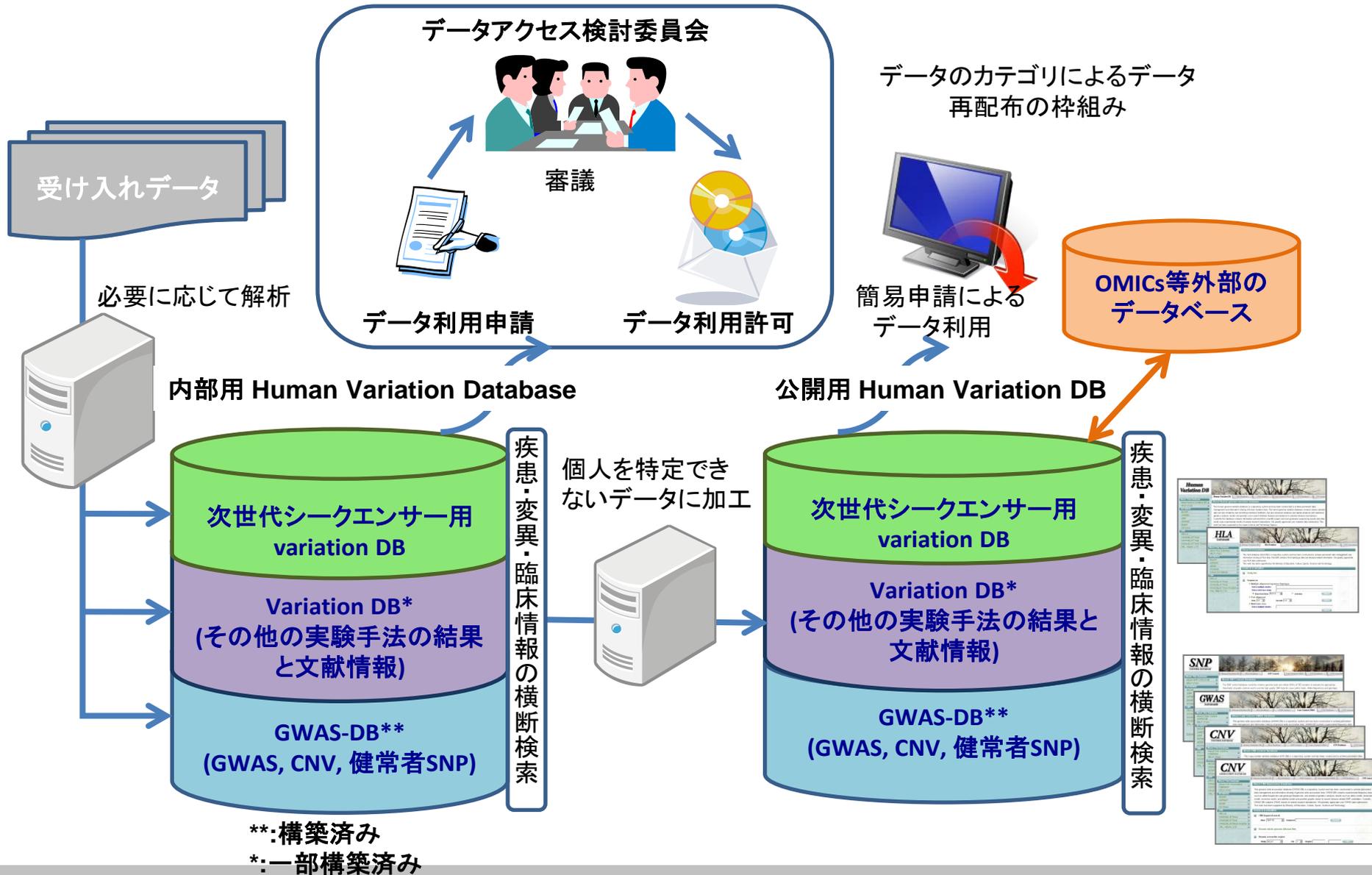
Browse whole genome (disease list)

Browse across the region  
Study [panic-plus-b] | Chr |  | Region |  |

Genome Browser (GBrowse)

GWAS  
対応DB

# DBシステムの概要



# 標準SNP DB

- 標準 SNP DB: 健常者のSNPデータ (GWAS チップ用) のデータ  
Affy500K 約500検体、Affy6.0 約600検体, Affy Axiom 約420検体、Illumina OMNI-2.5 約420検体

Contents:

- ・ 30-250万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
- ・ genotypic test, allelic test, additive risk model, recessive model, dominant model のP-value, OR, 95% CI, AICなどの遺伝統計値
- ・ SNPのアノテーション

**SNP**

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)

Database: DBCLS, University of Tokyo, University of Tokai, University of Tokyo Hospital, CRL, Hitachi, Ltd.

search\_type: s[SNP]

ChrNo: bp bp

Search Submit

Genome Browser

**SNP search**

SNP ID: [NRS6663840](#)

dbSNP ID(rs): [rs6663840](#)

dbSNP ID(ss): [ss16429890](#) [ss19129725](#) [ss19855219](#) [ss20488566](#) [ss23157850](#) [ss9823292](#)

JSNP ID:

HGVbase ID: [SNP006996858](#)

Chromosome: 1

Variation Class: SNP

SNP type: iSNP

Allele: A/G  
NM\_014704.2[A/G]:forward

Amino acid change:

Affymetrix: SNP\_A-1960639 A/G

Illumina: A/G

Array kind	Ethnic group	Individual Num.	Call Rate	Genotype detail			HWP	Allele	
				A/A	A/G	G/G		A	G
Illumina317K	Japanese	200	1.000	0.19	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  
GGAGGC GAAAAATC 0.393  
AGAGATGGAAGGTC 0.104  
GGGGCGGAAGAATC 0.033  
AGGGCGGAAGAATC 0.015  
AGAGGC GAAAAATC 0.011

**Gene Name: KIAA0562**

EntrezGene ID: [9731](#)

Gene Symbol: [KIAA0562](#)

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

Gene ontology (process):

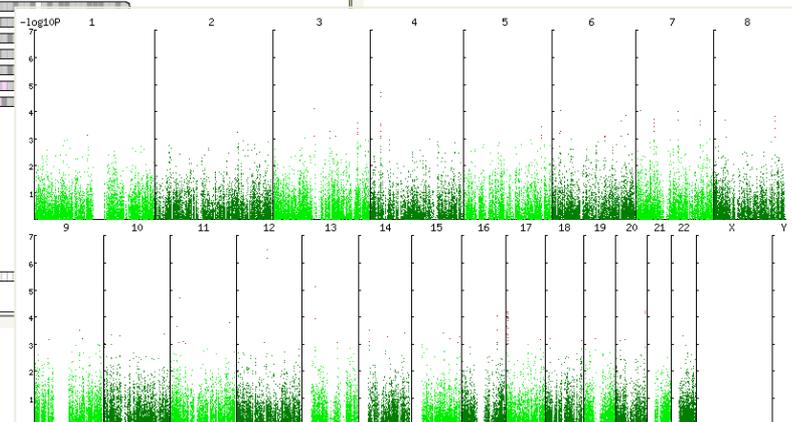
検索結果例

# SNP-based GWAS DB

GWAS DB: GWASデータ  
19疾患/28スタディー(内部用DB登録)  
17形質(内部用DB登録)  
14疾患/16スタディー(公開データ)

Contents:

- 30-100万SNPの遺伝子型頻度、アレル頻度、ハーディー・ワインベルク平衡検定値、Call rate等
- P値(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



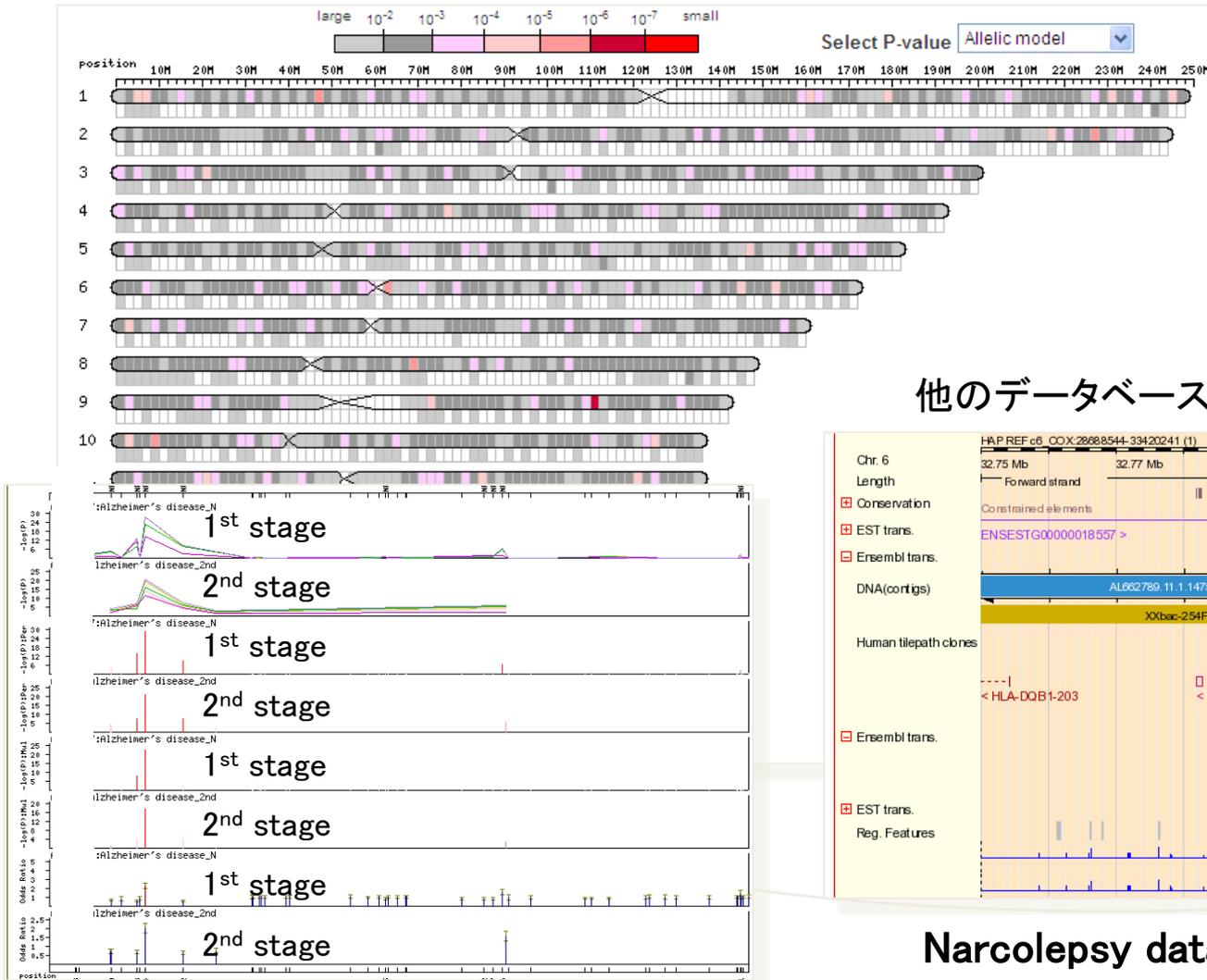
健常者SNPデータ (GWAS チップ用)

Affy500K 約500検体、Affy6.0 約600検体, Illumina OMNI-2.5 約420検体など



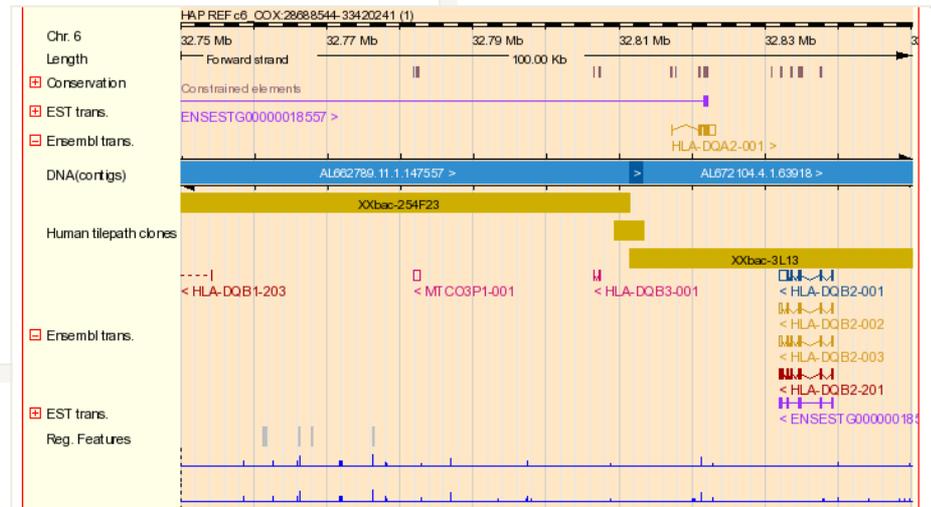
# SNP-based GWAS DB – 比較機能

## 複数の実験結果の比較



上が1st stage  
下が2nd stage

## 他のデータベースからの呼び出し (DAS)



Narcolepsy data view from Ensembl



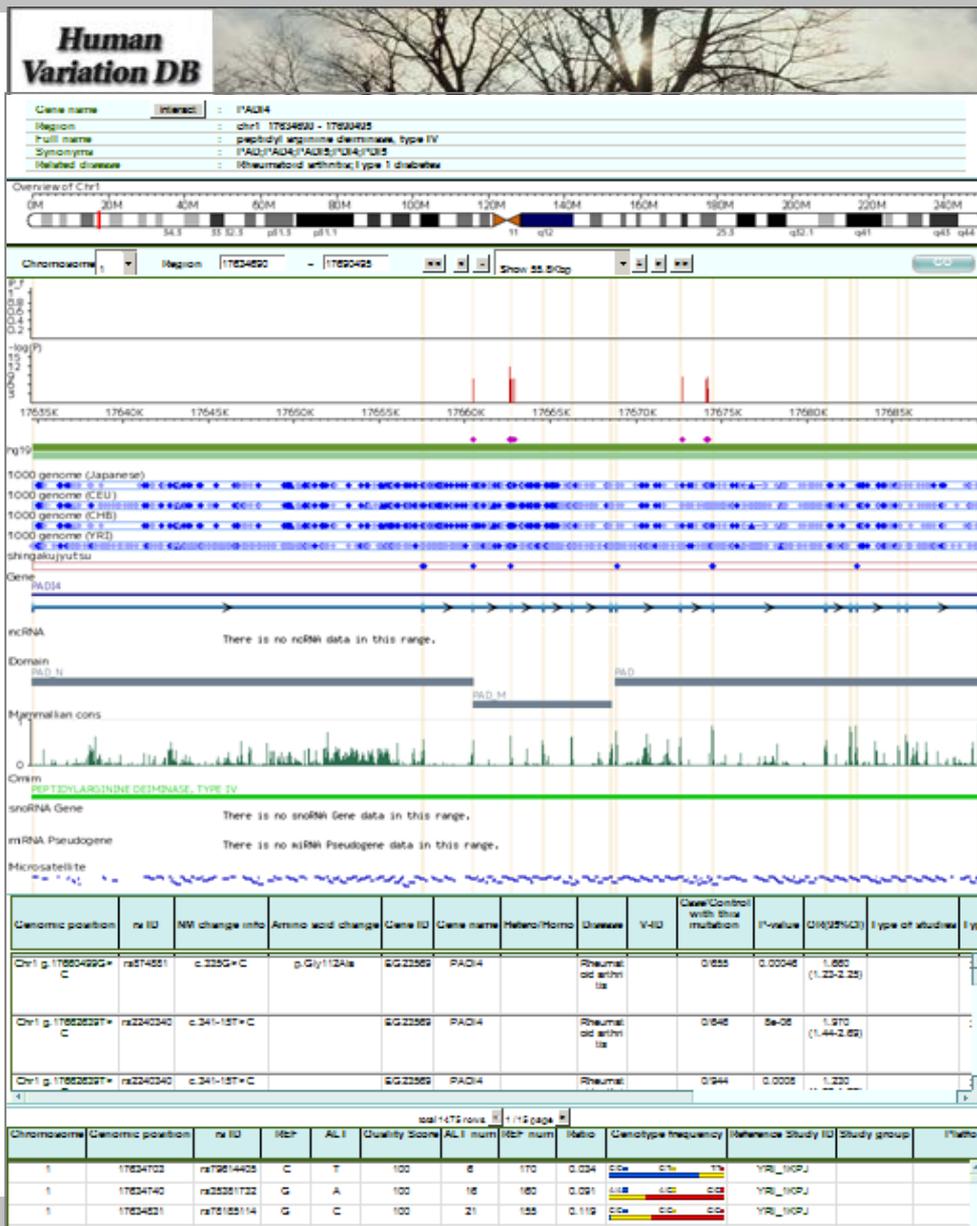
# Human Variation DB トップページ

The screenshot shows the Human Variation DB homepage. On the left is a navigation menu with sections: ABOUT THIS DATABASE (About Human Variation DB, HELP | FAQ), DATABASES (dbGAP, GeMDBJ, JSNP, HAPMAP, dbSNP, HGVDbase), and LINKS (NBDC, DBCLS, University of Tokyo, National Institute of Genetics, University of Tokyo Hospital, CRL, Hitachi, LTD.). The main content area has a header with 'Human Variation DB' and sub-headers for HLA Database, SNP Control, Case Control GWAS, CNV Database, and CNV Association. Below is an 'About Human genome variation database' section with a descriptive paragraph. A 'SEARCH' section follows, containing five search methods: 'Search by gene name' (keyword: EOC), 'Search by disease name' (keyword: ), 'Search by genomic position' (ChrNo. dropdown, Region: -), 'Search by SNP number' (kind: SNP-ID dropdown, keyword: ), and 'Interaction view by gene name' (keyword: ). Each search method has a 'Search' button. Below the search section is a 'BROWSE' section with six options: 'Browse by disease name', 'Browse by gene name', 'Browse by chromosome', 'Overview of SNP density', 'Summary of all diseases', and 'Genome Browser (GBrowse)'. Two callout boxes are overlaid on the page. The first callout box points to the search methods and contains the text: '遺伝子検索 疾患検索 領域検索等が可能'. The second callout box points to the browse options and contains the text: '遺伝子名, 疾患名でのブラウズ 集団ごとの変異分布の鳥瞰図 疾患ごとの変異分布の鳥瞰図 DASとしての図'.

遺伝子検索  
疾患検索  
領域検索等が可能

遺伝子名,  
疾患名でのブラウズ  
集団ごとの変異分布の鳥瞰図  
疾患ごとの変異分布の鳥瞰図  
DASとしての図

# Human Variation DB – 領域表示 (1)



遺伝子名

(何を表示するか  
選択可能)

P-value

Reference Genome

Domain情報

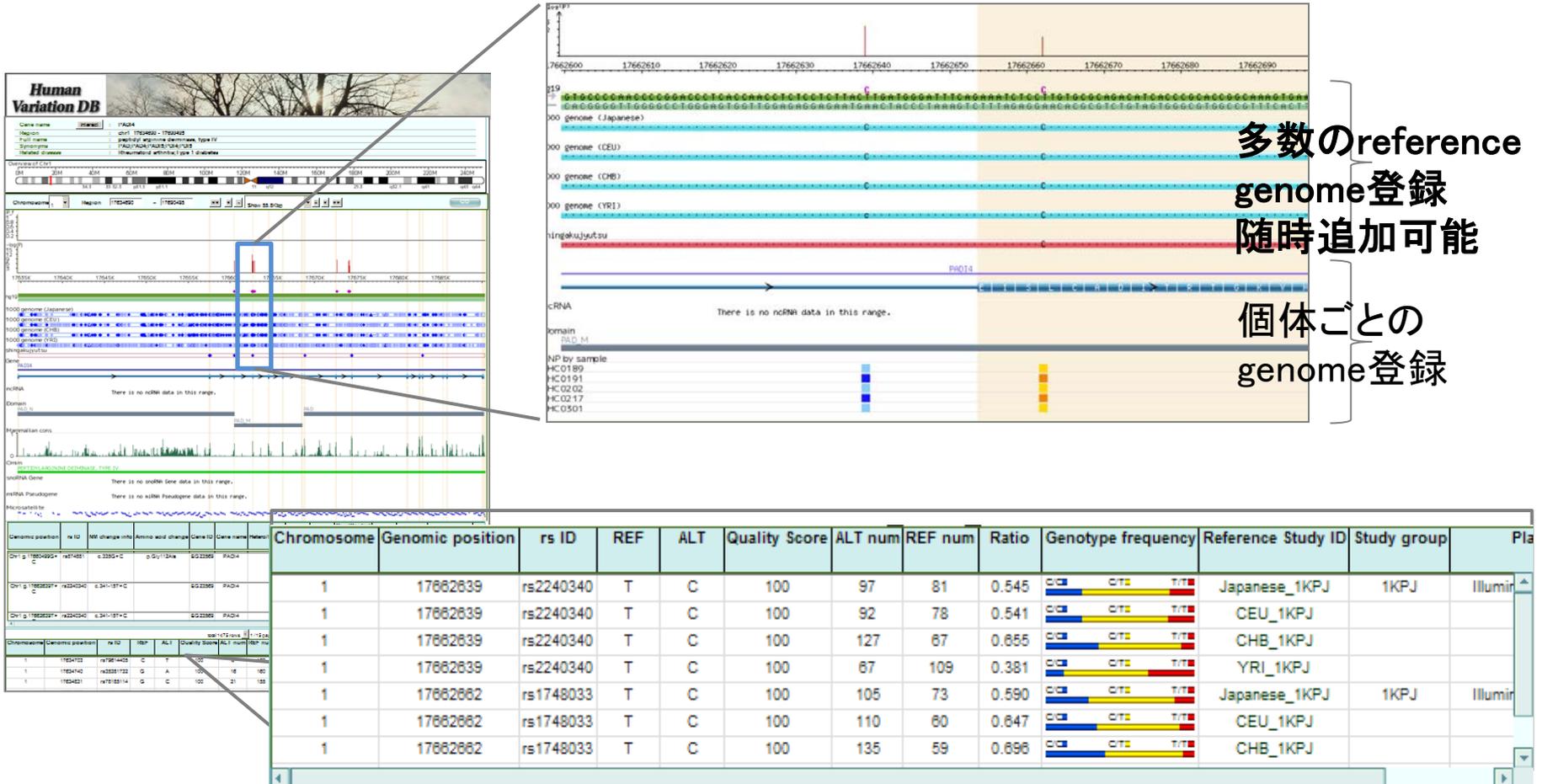
Conservation score

miRNA等情報

変異詳細情報

Reference情報(アレル頻度)

# Human Variation DB – 領域表示 (2)



Reference genomeのアリル頻度情報

# Human Variation DB – 領域表示 (3)

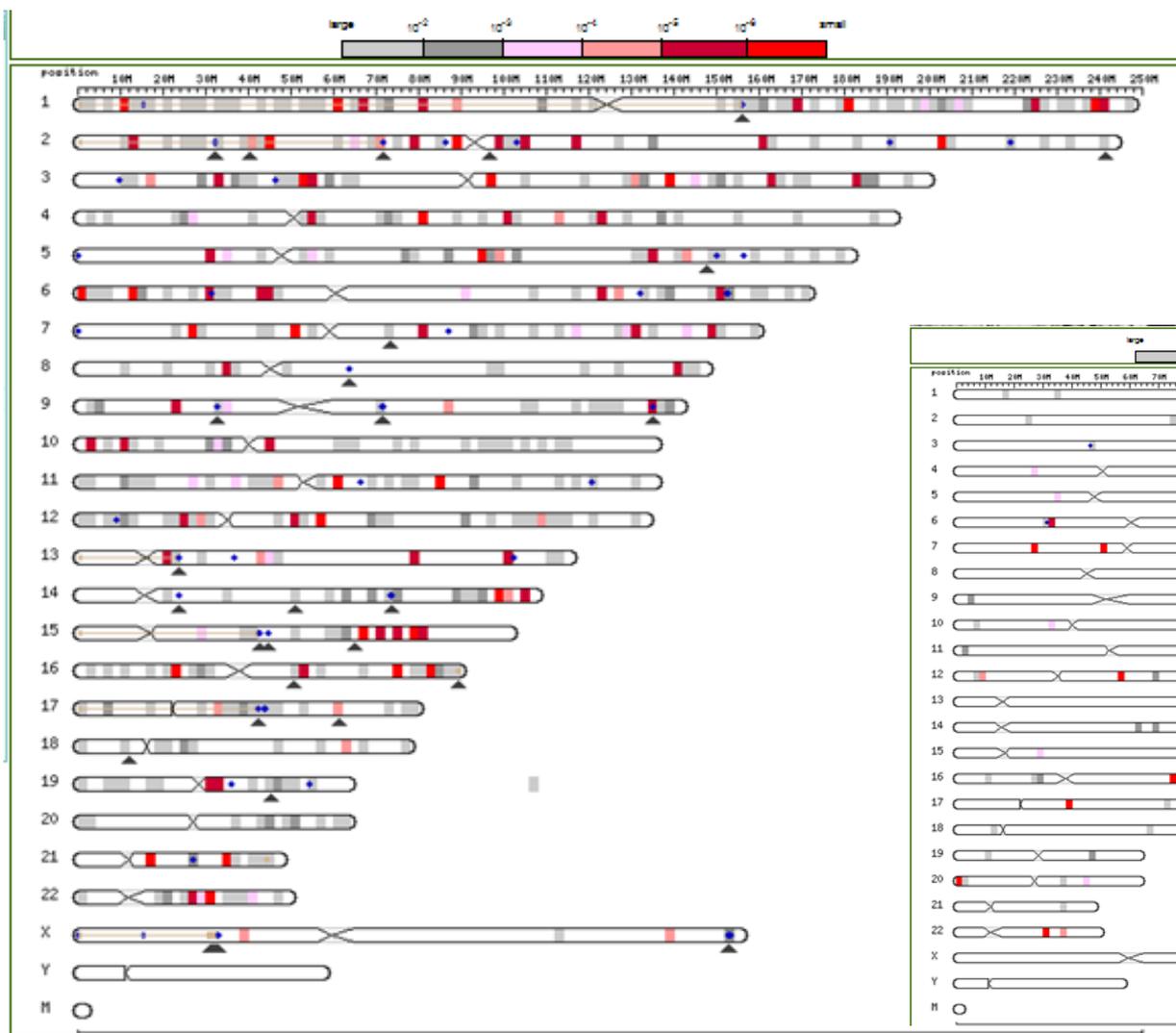
変異のゲノム上の位置、 SNPの種類、アミノ酸置換情報、症例-対照関連P値、オッズ比、実験手法、臨床情報等

NGS解析データも、文献データも同時に表示

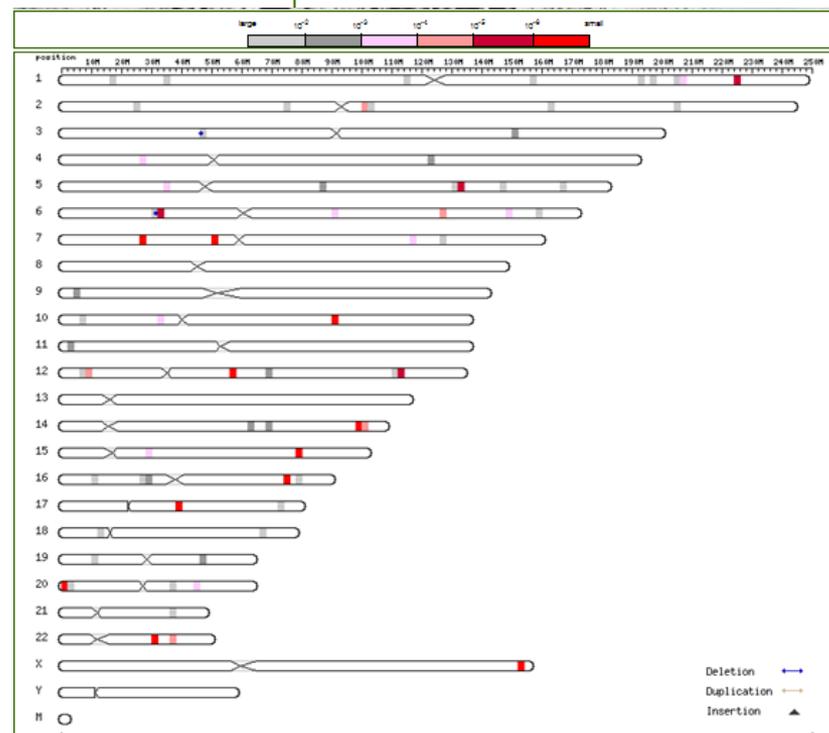


Genomic position	rs ID	NM ID	NM change info	NP ID	NP change info	Amino acid change	Gene ID	Gene name	var_type	SNP type	Hetero/Homo	Ethnic	N_fa
C		387.2		519.2									
Chr1 g.17862683T>C	rs2240340	NM_012387.2	c.341-15T>C	NP_036519.2			EG23569	PADI4	mutation	iSNP			
Chr1 g.17862682T>C	rs1748033	NM_012387.2	c.349T>C	NP_036519.2	p.Leu117=	TTG>CTG	EG23569	PADI4	mutation	sSNP		Japanese	
Chr1 g.17862682T>C	rs1748033	NM_012387.2	c.349T>C	NP_036519.2	p.Leu117=	TTG>CTG	EG23569	PADI4	mutation	sSNP		Japanese	

# Human Variation DB – 疾患ごとの変異分布



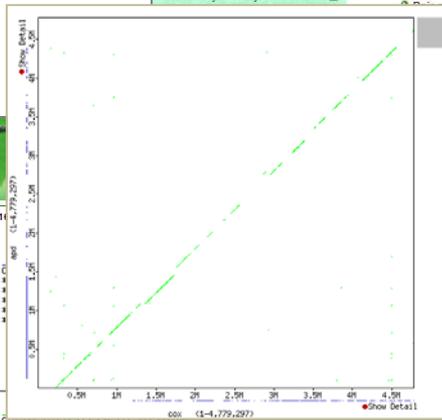
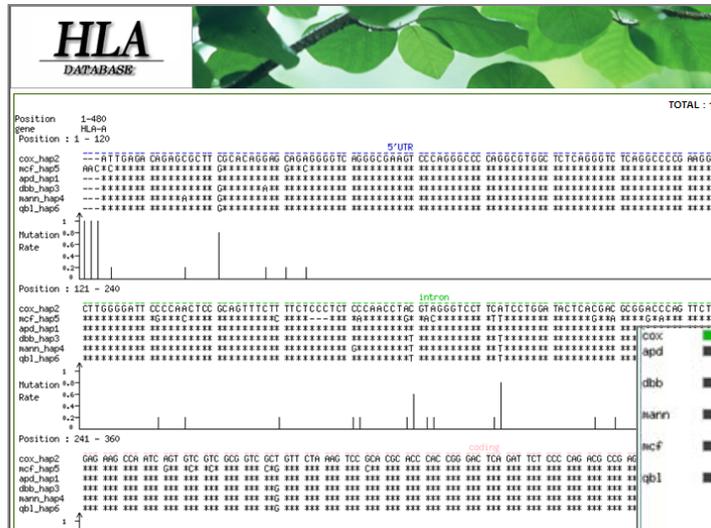
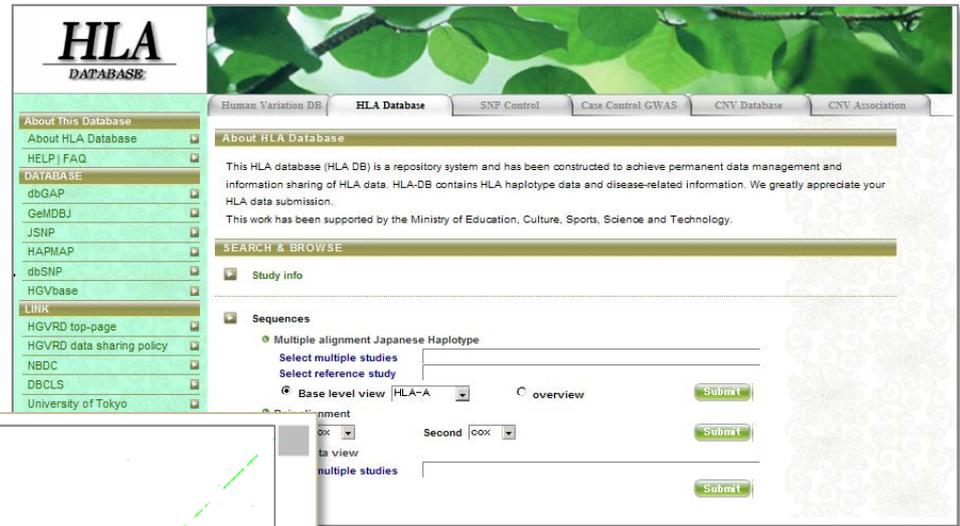
登録している全ての遺伝子での変異分布



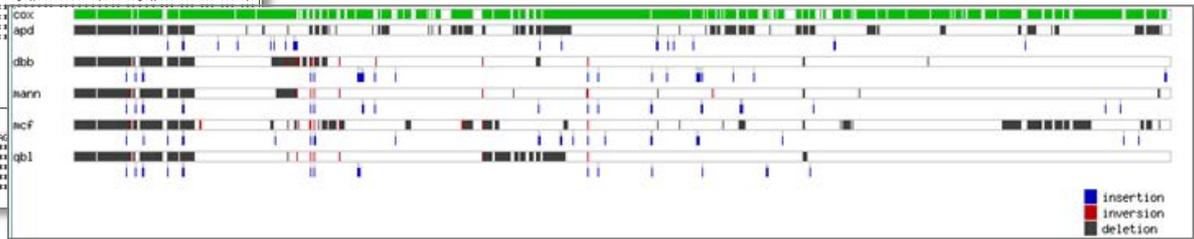
ある疾患での変異分布

# HLA DB

- HLAのハプロタイプごとの変異の登録
- HLAの多型と疾患感受性、免疫応答性、薬剤過敏症との関連を俯瞰可能



異なるHLA型間での相同性 (遺伝子ごと)



異なるHLA型間の相同性 (HLA領域全体)

HLA型間の塩基配列の違い

# NGSと文献登録データ

## 疾患原因変異にも特徴がある！

遺伝子のどの部位に生じても疾患発症

遺伝子の特定ドメインに生じた場合に発症

フレームシフト変異がホモ接合になった場合に発症

遺伝子の特定の位置で特定の置換をした場合に発症

集団により発症リスクが異なる遺伝子などなど

体系化することにより

疾患の**発症機序**の理解

新規変異の**表現型への影響**の予測への貢献を期待

## データ登録数の現状

### ➤ NGS公開データ

健常者: 1000 genome data exome 98検体

### ➤ NGS内部登録データ

健常者: exome 21検体、68検体、健常者: HLA 1検体、HLA 33検体 (セルライン)

### ➤ NGS内部登録準備データ

疾患遺伝子: 4遺伝子変異 (新規)  
2遺伝子変異 (既知)

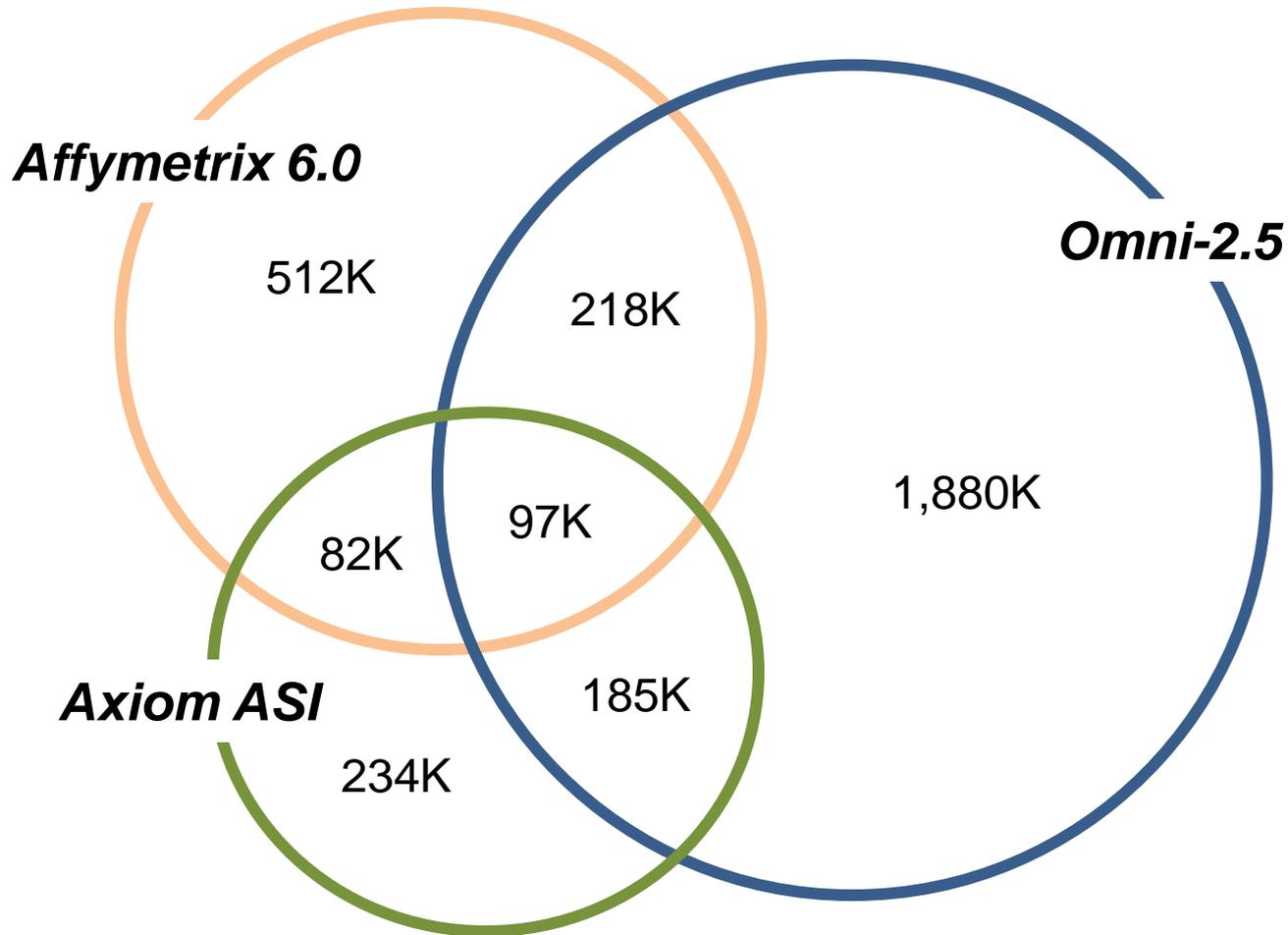
### ➤ 文献公開データ

Common disease, 神経変性変異データを  
中心に15,000変異と付随情報の登録

### ➤ 文献内部データ

25,000変異と付随情報の登録

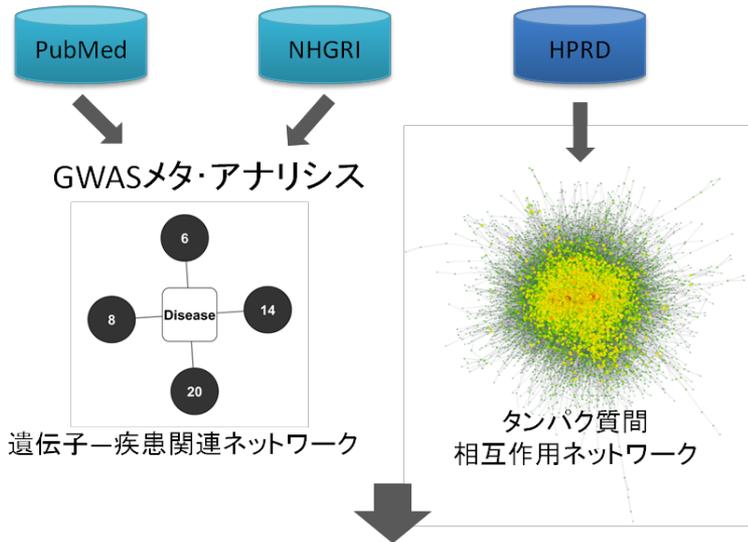
# SNP プラットフォーム間のimputation



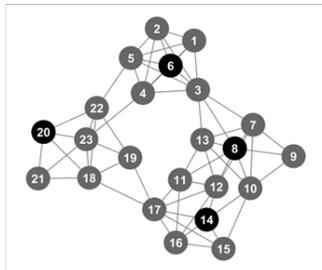
日本人健常者約420検体のAxiom ASI、Omni-2.5、Affymetrix 6.0解析データを用い、いずれかのデータから他の解析結果の推定が可能

# メタ解析・ポストGWAS解析手法の開発

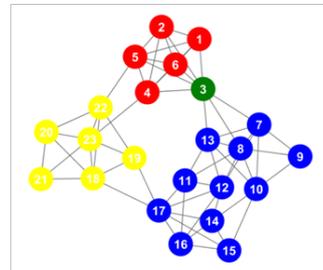
網羅的データベース検索



ネットワーク解析

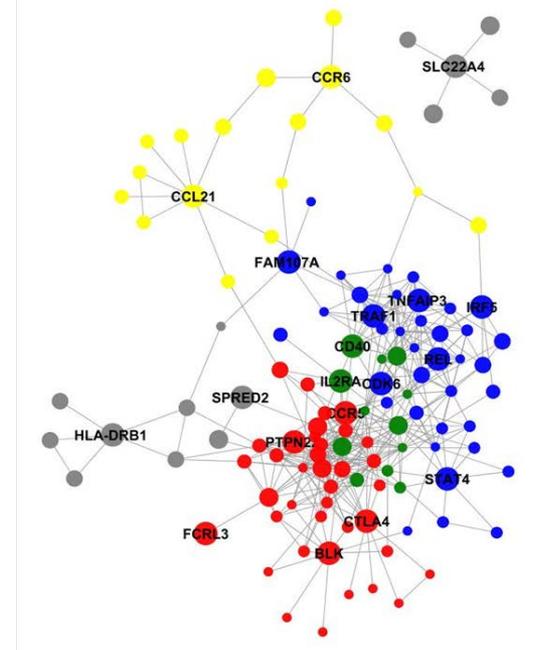


リシーケンス遺伝子選定



生物学的経路同定

関節リウマチでのケーススタディー



リシーケンス候補遺伝子: ZAP70, IL2RB, IL2

免疫学的経路: 白血球活性化・分化、  
パターン認識受容体シグナル伝達、  
ケモカイン・ケモカインレセプター

Nakaoka et al. PLoS ONE 2011

GWASメタ解析の統計値、蛋白質間相互作用、random walk with restart (RWR) algorithm、階層的クラスタリング法を用いて、疾患関連候補遺伝子をランキングすると共に、パスウェイの疾患関連機能モジュールを同定する方法を確立



# NBDCヒトデータ共有ガイドライン

ヒトデータを公開・共有するための運用ルールとしてのガイドラインを策定  
<http://humandbs.biosciencedbc.jp/guidelines/data-sharing-guidelines>

公的資金を用いて産生されたヒトに関するデータ一般に適用  
データの適正に管理しつつ、収集したデータをなるべく広く共有  
公開の有無、アクセス制限のレベルによって以下の4つに分類

**オープンデータ**: アクセスに制限を設けることなく、利用することが可能な公開データ。頻度データ、統計解析結果、すでに発表された論文の参照データなど。

**制限公開データ**: 利用者、利用目的等を明らかにしたうえで、関連研究に実績のある研究者が研究のために利用可能な公開データ。個人毎の解析結果など。

**公開待機データ**: 論文発表や知的財産権取得等、データ提供者による成果公開の後、オープンデータあるいは制限公開データとして公開。

**匿名化前・公開留保データ**: 各プロジェクトや実施機関が保有する匿名化前のデータ。

# 今後の展望

**Human Variation DB** の充実: 次世代シーケンサーなどによる新規多型・変異情報と、文献からの多型・変異情報を情報量、機能ともに一層充実させる

Human Variation DB とオミックスデータ等との連携を強化し、変異の表現型への影響を解釈・予想できるような知識型DBへ発展させる

データの受け入れ、再配布に関し、NBDCによる新たなガイドラインに基づいた運用を開始する

ゲノム支援、新学術、その他のプロジェクト、外部機関からの **NGSデータ**登録、**GWASデータ**登録のための活動を継続する

海外の関連するデータベースとの連携をはかる

GWAS Central, Neuroscience Information Framework など