

統合データベース整備事業

疾患解析から医療応用を実現する DB 開発

21年度 研究成果報告書

平成22年3月

東京大学	大学院医学系研究科	徳永 勝士
東京大学	医学部附属病院	辻 省次
東海大学	医学部	井ノ上 逸朗
株式会社	日立製作所	小池 麻子

本報告書は、文部科学省の科学技術試験研究委託事業による委託業務として、東京大学大学院医学系研究科、東京大学医学部附属病院、東海大学医学部、及び株式会社日立製作所が共同で実施した、平成21年度の「疾患解析から医療応用を実現するDB開発（契約書第1条で定めた委託業務題目）」の成果を取りまとめたものです。

目次

1. 委託業務の目的	1
2. 平成 21 年度（報告年度）の実施内容	1
2.1 実施計画	1
(1) 標準 SNP DB の構築	1
(2) GWAS DB の構築	1
(3) リシークエンスによる臨床情報・ゲノム情報 DB の構築	2
2.2 実施内容（成果）	2
(1) 標準 SNP DB の構築	2
(2) GWAS DB の構築	4
(3) リシークエンスによる臨床情報・ゲノム情報 DB の構築	12
(4) プロジェクトの総合推進	15
(5) DBアクセス状況について	15
2.3 成果の外部発表	15
2.4 活動（運営委員会の活動等）	15
2.5 実施体制	16
2.6 整備実績一覧	17
別紙	21
学会等発表実績	

1. 委託業務の目的

ゲノムワイドな SNP タイピングおよび疾患原因・関連遺伝子のリシーケンスを行い、臨床情報とゲノム・遺伝子情報との関連性を解析してデータベース (DB) 化するとともに、希望者には所定の手続きもしくは審査を経て、データの再配布を行う。この DB をより多くの研究者等が利用することにより、疾患の遺伝要因の解明や、遺伝子診断、疾患の分子疫学等の研究が促進され、個別化医療の実現が加速されることを目的とする。このため、東京大学大学院医学系研究科、東京大学医学部附属病院、東海大学、日立製作所が共同して以下の3つの業務を行う。

(1)標準 SNP DB の構築 (東京大学大学院医学系研究科が主担当として DB を構築、東海大学が統計遺伝学手法を分担)

(2)GWAS (ゲノムワイド関連解析) DB の構築 (東京大学大学院医学系研究科が主担当として DB を構築、東海大学が統計遺伝学的手法等および日立製作所が、外部機関が産出した GWAS データ処理と Bioinformatics 的手法を分担)

(3)リシーケンスによる臨床情報・ゲノム情報 DB の構築 (東京大学医学部附属病院が主担当として第1次 DB を構築、東京大学大学院医学系研究科が第2次 DB の構築および日立製作所がマイニング手法と Bioinformatics 的手法等を分担)

2. 平成21年度(報告年度)の実施内容

2.1 実施計画

(1) 標準 SNP DB の構築

ゲノムワイドな関連解析 (GWAS) では、データ解析に使用する検体、SNP についての品質管理が重要である。平成19年度、東京大学大学院医学系研究科においては、健常日本人約500例からの50万種のSNPおよび約200例からの90万種のSNPについて標準アレル頻度、遺伝子型頻度、ハプロタイプ頻度などを登録した標準化データベースを構築した。平成20年度は、東京大学大学院医学系研究科においては、当該機関で引き続き産出する対象集団のSNPデータと解析データの追加登録と共に、データベースのインターフェースの拡充を行った。平成21年度、東京大学大学院医学系研究科においては、引き続き解析データの追加登録を行う。東海大学は、平成19年度に健常日本人について、50万種：約500検体、90万種：約200検体のSNPデータを用い、品質管理に必要となる適切な基準を作成した。平成20年度においても、新たなGWASのプラットフォームにおいて産出されたSNPについて、品質管理に必要とある適切な基準を作成した。東海大学はH21年度も引き続き、新たに産出されるデータについて、各種遺伝統計値、及び、他機関が産出するデータを基に、健常日本人の品質管理となる基準値の検討を行う。

(2) GWAS DB の構築

—GWAS 第一ステージ DB の構築と疾患関連 SNP 探索手法の研究開発—

GWAS は第1ステージ (探索) のタイピング、第2ステージ (再現性確認と高精度マッピング) のタイピングからなるが、平成19年度は、東京大学大学院医学系研究科においては、8疾患について第1ステージの解析とDB化 (スタディ・デザイン及び、SNP ごとの call rate、アレル頻度、関連解析結果などの遺伝統計値を登録) を行い、日立製作所は機能情報などを用いた SNP 絞込み手法を開発した。平成20年度は、東京大学大学院医学系研究科において、DB の拡張と解析結果を表示するインターフェースの拡充、及び、データベース検索の高速化などを行うとともに、新たな4種の疾患について GWAS データを登録した。Copy number variation(CNV) の GWAS の基礎となる CNV 標準データベースを構築した。平成20年度は、日立製作所においては、外部から受け入れる GWAS データに関して一連の基本的な遺伝統計的解析を行うパイプラインの構築、及び登録システムの構築と Bioinformatics 的候補 SNP 絞込み手法を開発した。また、東海大学は、平成19, 20年度に、統計遺伝学手法の開発として多変量解析と閾値もモデルを組み合わせた SNP 間相互作用の手法開発を行った。

平成21年度は、東京大学大学院医学系研究科において表示機能の拡張、および、第1ステージの結果と第2ステージ結果との関係性表示機能などのDB機能拡張とインターフェースの

更なる拡充を行う。日立製作所は、外部から受け入れる GWAS データに関する SNP 間相互作用や Copy number variation を含む応用的な面での遺伝統計的解析を行うパイプラインの構築を構築するとともに、平成 20 年度に引き続き Bioinformatics 的候補 SNP 絞込み手法を開発する。

平成 21 年度、東海大学は引き続き、統計遺伝学手法の開発を行う。具体的には、疾患の発症リスクに対する、複数の SNP の寄与度を同時に評価する手法として、Bayesian shrinkage 法のアルゴリズムをソフトウェア化する。本手法は、stepwise ロジスティック回帰法など、従来広く用いられてきたものと比較して、はるかに多くの SNP データを解析することが可能である。

(3) リシークエンシングによる臨床情報・ゲノム情報 DB の構築

ーリシークエンス DB の臨床情報・ゲノム情報 DB の構築と解析手法の開発ー

東京大医学部附属病院で産出される疾患関連遺伝子のリシークエンスによる遺伝子・ゲノム変異情報とそれに付随する臨床情報を DB 化すると共に（第 1 次 DB）、遺伝子・ゲノム変異情報と臨床情報との関連などをマイニング手法および Bioinformatics 的手法を用いて解析する。東京大学医学部附属病院が主担当としてリシークエンスデータの第 1 次 DB を構築し、東京大学大学院医学系研究科は、外部の DB の有用なデータをインポートする機能を開発する(第 2 次 DB)。平成 19 年度は筋萎縮性側索硬化症(ALS)について東京大医学部附属病院が第 1 次 DB を東京大学大学院医学系研究科は第 2 次 DB の構築を行った。平成 20 年度は本データベースのインターフェースを充実化すると共に、パーキンソン病について東京大医学部附属病院と東京大学大学院医学系研究科により第 1 次 DB と第 2 次 DB の基本機能を実装した。平成 21 年度、パーキンソン病について、東京大学大学院医学系研究科は第 2 次 DB の残り部分であるデータ登録部の拡張実装を行い、東京大医学部附属病院はゲノム変異情報と臨床情報を拡充させるとともに、インターフェースの拡充を行う。

また、新たに、平成 21 年度から 2 年計画で、家族性痙攣性対麻痺と副腎白質ジストロフィーのリシークエンスデータベースの第 1 次 DB を東京大医学部附属病院が、第 2 次 DB を東京大学大学院医学系研究科が構築する。

2.2 実施内容(成果)

(1) 標準 SNP DB の構築

①標準 SNP DB の構築のための統計遺伝学手法の開発（東海大学実施）

平成 19 年度は、健常日本人約 700 例データ（タイピングに用いたプラットフォームは Affymetrix 500K アレイ、Affymetrix 6.0 おおおよび、Illumina 317K）を用い、データの品質管理のための基準を設定したが、平成 20 年度は、新たに追加された健常日本人 460 例について、GoldenGate アッセイでタイピングを行った 2,300SNP のデータについて各種遺伝統計値を計算した。また、東京大学大学院医学系研究科が産出したデータを基に、健常日本人の品質管理に必要な基準値の検討を行い、データのクリーニングを行った。

平成 21 年度は、新たに追加した健常日本人 223 名について、TaqMan® SNP ジェノタイピングアッセイで 22SNP のタイピングを行い、データの品質管理に必要な基準値の検討を行った。まず、コール率 80%を下回る 10 例を削除した（図 1 参照）。また、重複または潜在的な血縁者をデータ内に有するサンプル 4 例を削除した。SNP については、19 年度に Illumina 317K でタイピングした 200 例（東海大学で管理しているもののみ）、および 20 年度に GoldenGate アッセイでタイピングした 460 例についても、TaqMan® SNP ジェノタイピングアッセイで再度タイピングを行い、以前に得たタイピングデータと今回のそれを照合して一致率を算出したが、いずれも 95%以上という高い一致率を示したため（表 1 参照）、削除の対象とはしなかった。

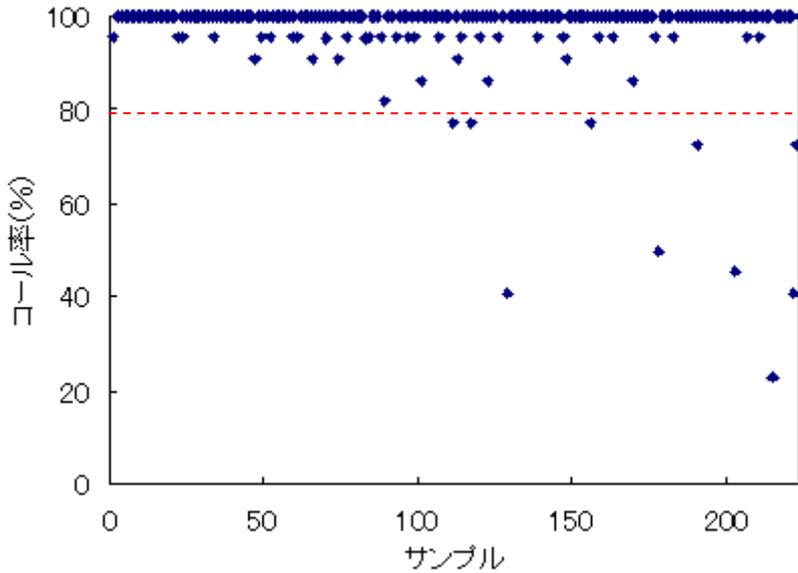


図 1. サンプルの品質管理
 コール率について、基準を満たさないサンプル（図中の赤い点線より下に位置する点）はデータから削除した。

表 1. SNP の品質管理
 同一のサンプルを異なるプラットフォームでタイピングしたが、いずれも高い一致率 (> 95%) を示した。

SNP	In first stage, HumanHap300 vs TaqMan				In second stage, GoldenGate vs TaqMan			
	Missing rate (%)		Number compaired	CR (%)	Missingrate (%)		Number compaired	CR (%)
	HH	TM			GG	TM		
rs7550260	0	4.2	454	99.8	0.1	2.6	829	99.1
rs1367878	0	0.8	470	98.9	0.6	0.7	842	99.6
rs9864101	0.4	2.5	460	99.8	0.6	2.3	827	98.9
rs2703888	0.6	1.5	464	99.6	0	1.9	836	99.3
rs7702812	0.8	0.6	467	98.7	0.2	1.7	836	99.6
rs1565873	0	1.3	468	99.4	0	1.3	841	99
rs10872573	0	1.1	469	99.8	0	1.7	838	99.4
rs215939	0	1.3	468	99.1	0	1.5	839	99.3
rs6570836	0.8	1.9	461	99.6	0.3	1.4	837	99.5
rs7781293	0	1.3	468	99.8	0.3	0.4	846	99.8
rs4628172	0.2	1.3	467	99.8	0.2	3.5	819	97.3
rs2389409	0	4.4	453	98.5	0.3	5.2	803	95.4
rs3176292	0.2	1.7	465	99.6	0.6	0.8	841	99.5
rs6470572	0	1.1	469	99.1	0.1	1.3	840	99.4
rs1930095	0	1.1	469	99.6	0.1	1	843	99.8
rs268300	0	1.1	469	99.6	0.3	0.7	844	96.9
rs7923449	0.2	1.1	468	99.8	0.1	1.3	840	99.9

rs945335	0	2.3	463	98.3	0.2	1.1	841	98.3
rs1150229	0	1.1	469	98.7	0.3	1.5	836	98.6
rs7209819	0.4	2.1	462	99.6	0	1	844	99.3
rs1978503	0	1.3	468	100	0	1	844	99.4
rs873286	0	2.5	462	99.4	0.1	2.3	831	99.2
Average (\pm S.D)	-	-	-	99.4 \pm 0.5	-	-	-	98.9 \pm 1.1

②標準 SNP DB の構築（東京大学大学院医学系研究科実施）

平成 19 年度に構築した健常者の SNP DB について、中核機関への DB サーバー移植を行い、Affymetrix 500K 日本人健常者 459 検体、Affymetrix6.0 198 検体について公開した。また、様々な SNP-ID に対応できるように検索機能を拡張した。標準 SNP DB に含まれる遺伝子型頻度、アレル頻度、call rate, Hardy-weinberg 平衡検定値などの基本情報を Das server にバッチ形式で登録できるように環境を整えた。平成 21 年度は、既に構築してある標準 SNP DB に登録するため (https://gwas.lifesciencedb.jp/spndb/snp_top.php から公開)、健常者約 1,300 人分（提供元：東京大、三重大、新潟大ほか）の主成分分析による集団構造の解析、及び、Hardy-weinberg 平衡検定などの遺伝統計解析を行った。関係論文が公開され次第追加データ分の公開を行う。

（2）GWAS DB の構築

①GWAS DB の構築と手法開発（東京大学大学院医学系研究科及び日立製作所実施）

GWAS は第 1 ステージ（探索）のタイピング、第 2 ステージ（バリデーション）のタイピングからなるが、平成 19 年度では第 1 ステージの解析と解析結果の DB 化を行い、東京大学医学部が DB 構築を行い、日立製作所が外部データである遺伝子位置情報や OMIM 情報などのインポート部分を実装した。本 DB は、SNP ごとの遺伝子型頻度、アレル頻度、call rate、Hardy-Weinberg 平衡検定値などの基本情報とともに、genotypic model, allelic model, additive risk model, recessive model, dominant model など主な遺伝統計値を登録可能であり、遺伝統計値の染色体全体での map 表示機能を備え、copy number variation(CNV), OMIM などの他の情報と共にグラフ表示できる機能を DB に搭載し、疾患関連候補 SNP の絞込みを可能としている。平成 20 年度は本 DB に対し、SNP 間の相互作用の登録機能、SNP 間ネットワーク表示、ユーザの実験データの登録・表示機能等を追加するとともに、各検索の高速化のためにインデックスとテーブル構成の見直しを行った。（東京大学大学院医学系研究科実施）また、新たに健常者の CNV のデータを登録するために、CNV の検出手法の比較検討を行うと共に（日立製作所実施）、CNV を蓄積・閲覧するための CNV-DB を構築した。（東京大学大学院医学系研究科実施）また、外部機関が産出するデータ向けには、基本的な遺伝統計的計算を行うパイプライン、及び、登録システムの構築を行った。（日立製作所実施）

平成 21 年度は、ユーザ登録等を簡易に Web で行う機能の追加（図 2-1）、第 1 ステージの結果と第 2 ステージ結果との関係性表示機能の追加（図 2-2）、データダウンロード時の詳細設定機能の追加（図 2-3）、snoRNA, miRNA pseudogene の表示の追加（図 2-4）や、閾値以下の p-value を持つ SNPs を統計的に有意に多く含む pathway の登録と表示機能の追加（図 2-5）などの機能拡張を行った。（東京大学大学院医学系研究科実施）

また、CNV 標準データベースの Copy Number Variation のクラスタリング機能(図 3-1)、個体ごとの表示機能(図 3-2)などの機能拡充を行った。また、Copy Number Variation の case-control 解析の結果を登録するための CNV case-control DB を新たに構築した(図 3-3)。上記 DB に、新規に、膝関節症、パニック障害、多系統萎縮症の GWAS 結果と、情緒的気質、自閉的性質、人格に関する 20 種類の量的形質の GWAS 結果を登録した。（東京大学大学院医

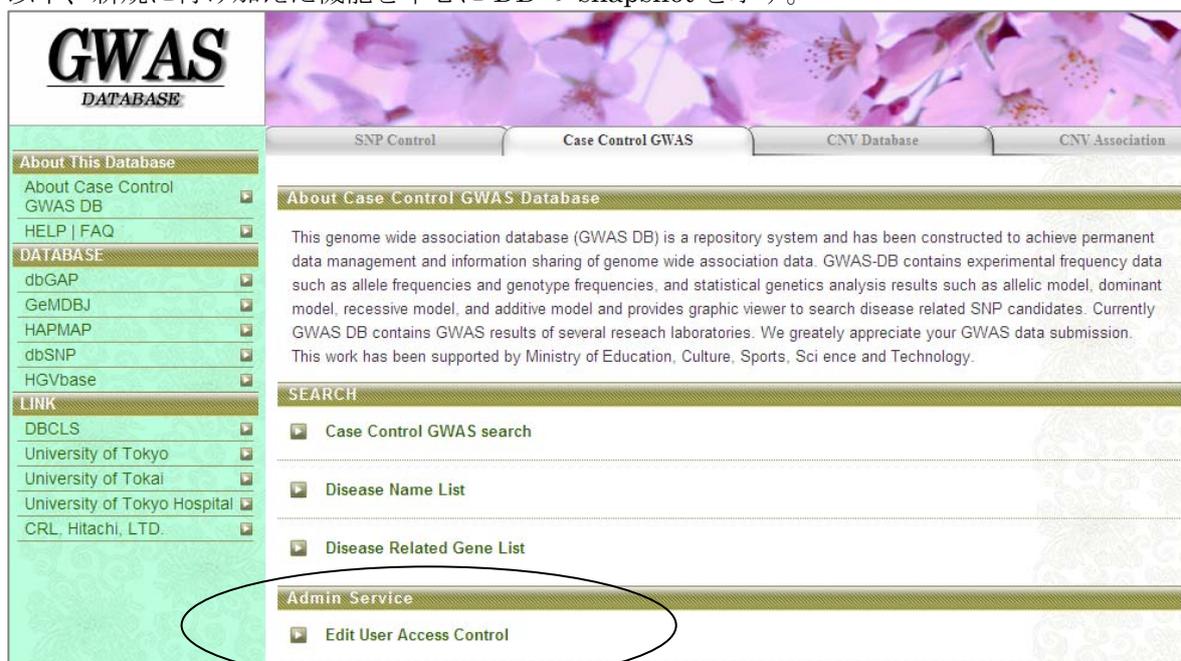
学系研究科実施)

Circular binary segmentation, Hidden markov model, Gaussian process modeling method などのアルゴリズムについて HapMap データを利用した同一検体 CNV の再現性や親子データにおける CNV の一致度など計算することにより、検出精度を評価した。また、CNV の start-end 位置の個体間の些細な実験的な検出のずれを、ずれた領域に含まれる probe 数を考慮してクラスタリングすることにより、集団としての CNV 領域を決定した。(日立製作所実施)

外部から受け入れる GWAS データの解析自動化の一環として、CNV の検出、及び、CNV を利用した case-control の遺伝統計的解析を行うパイプラインの構築を行った。また、特徴量の多いデータセットに適した集団学習法の一つである Random forest を拡張し、Random forest 中での SNP の同時出現頻度を利用した SNP 絞り込み手法、及び、SNP 間相互作用の検出を行う手法を開発し、同様に解析のパイプライン化を行った。(日立製作所実施)

また、東京大学大学院医学系研究科を中心として、また、データ公開およびデータの共有に関する方針案および関連する各種書類案を作成し、データアクセス倫理検討委員会において審議して最終案を決定した。この方針と書類一式を Web 上で公開するとともに、データアクセス倫理審査委員会が新たに発足した。(図 4-1, 図 4-2)

以下、新規に付け加えた機能を中心に DB の snapshot を示す。



ユーザー登録及び、アクセスコントロー

図 2-1 GWAS-DB ユーザ登録機能

上が第1ステージの結果

下が第2ステージの結果

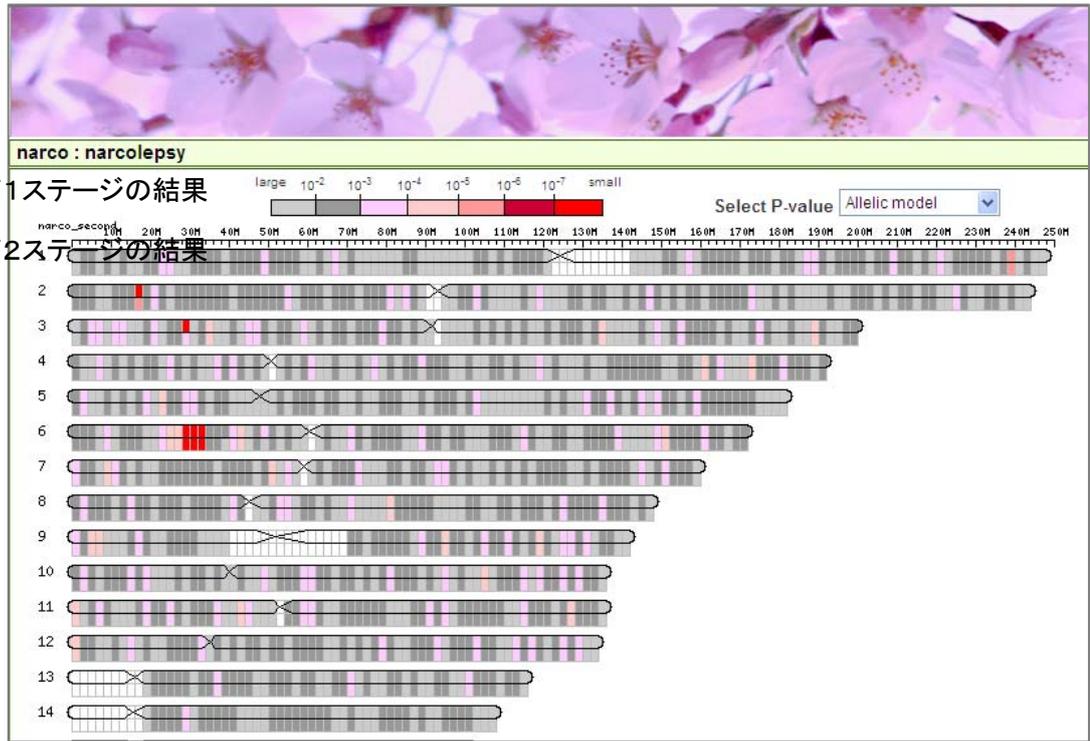


図 2-2 GWAS-DB 第一、第二ステージ結果の同時表示画面

Download narco : narcolepsy

Region: all Chromosome 1 Position [] - []

SNP-ID specified file: [] 参照...

Filtering	Item	Check if you want to download the item
Quality:	Chromosome Number	<input checked="" type="checkbox"/>
MAF (control) 0.05 and above	Position	<input checked="" type="checkbox"/>
MAF (case) 0.05 and above	SNP Type	<input checked="" type="checkbox"/>
HWE p-value 0.0001 and above	Gene Name	<input checked="" type="checkbox"/>
Call Rate (control) 0.95 and above	AIC	<input checked="" type="checkbox"/>
Call Rate (case) 0.95 and above	P-value	
SNP Type:	Allelic model	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> all SNP kinds	Genotypic model	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> cSNP	Dominant model	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> sSNP	Recessive model	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> rSNP	Additive model	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> iSNP	Permutation test	
<input checked="" type="checkbox"/> gSNP	Allelic model	<input checked="" type="checkbox"/>
Function:	Genotypic model	<input type="checkbox"/>
Minor allele homozygosity:	Dominant model	<input type="checkbox"/>
Value: [] and below	Recessive model	<input type="checkbox"/>
	Additive model	<input type="checkbox"/>
	Multiple test corrections	
	BONF	<input checked="" type="checkbox"/>
	HOLM	<input type="checkbox"/>

Download

図 2-3 ダウンロード時の詳細設定機能の追加

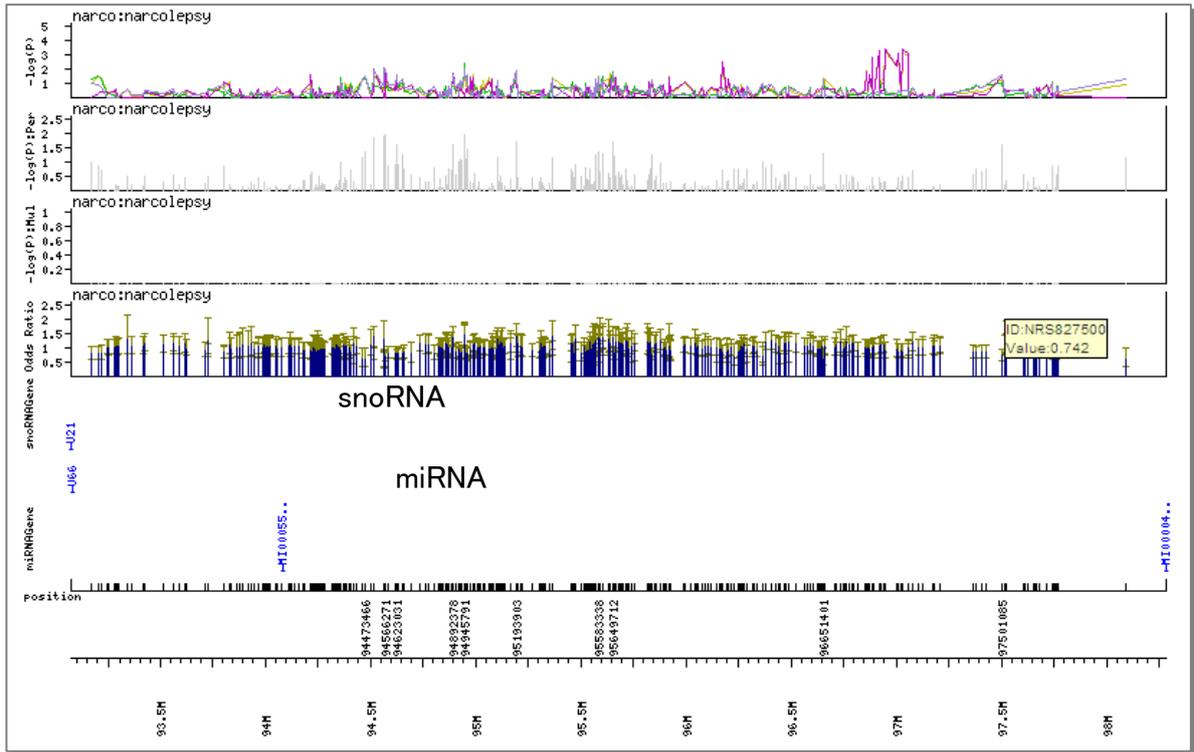


図 2-4 snoRNA, miRNA の登録・表示機能

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

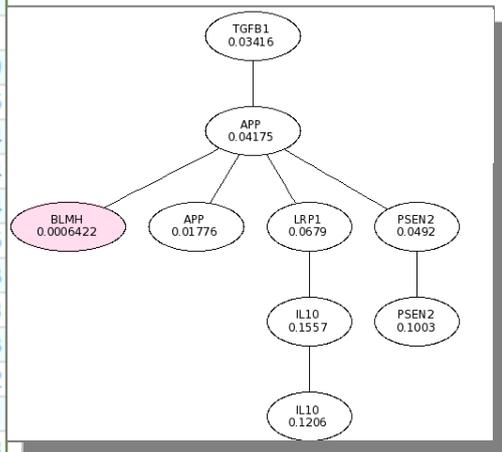


図 2-5 GWAS 結果に対する pathway 解析結果の登録・表示機能

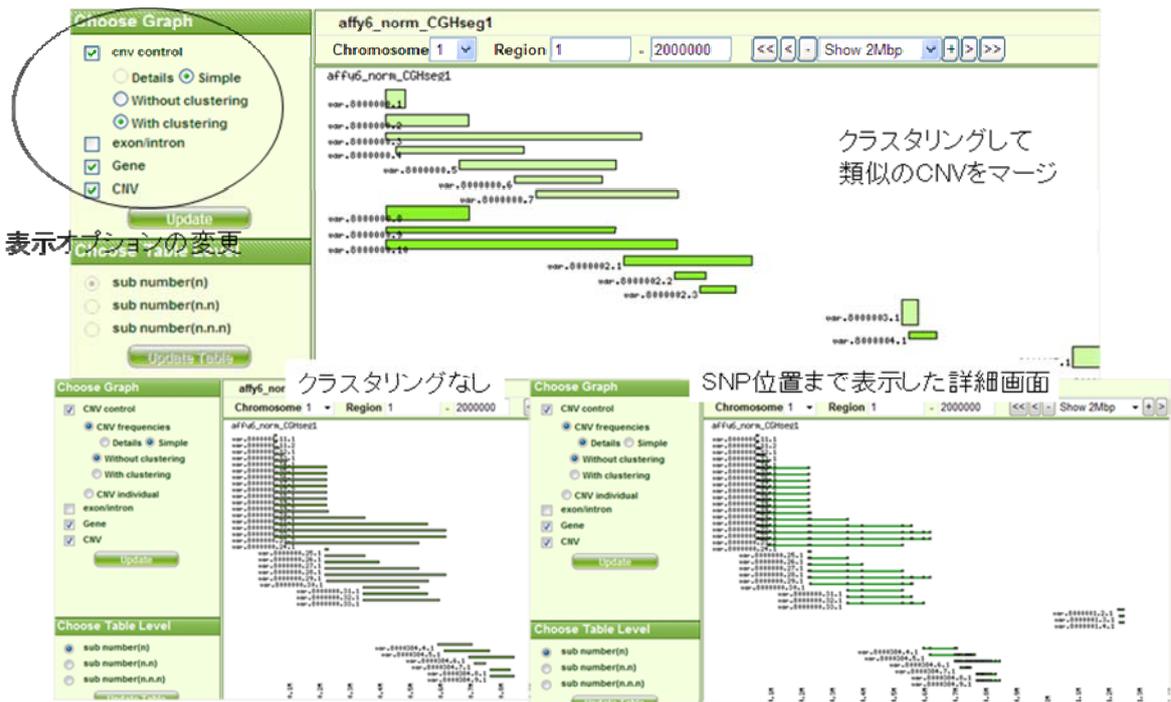


図 3-1 CNV 標準データベースの snapshot

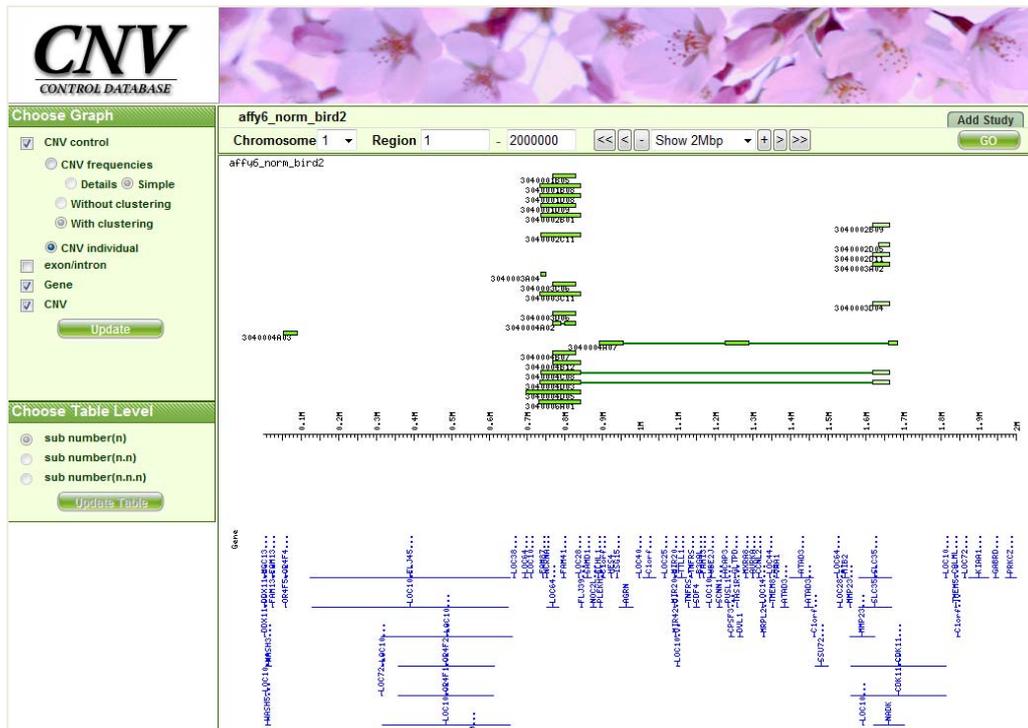


図 3-2 CNV 標準データベースの個体ごとの CNV 表示

Duplication, Deletionの選択

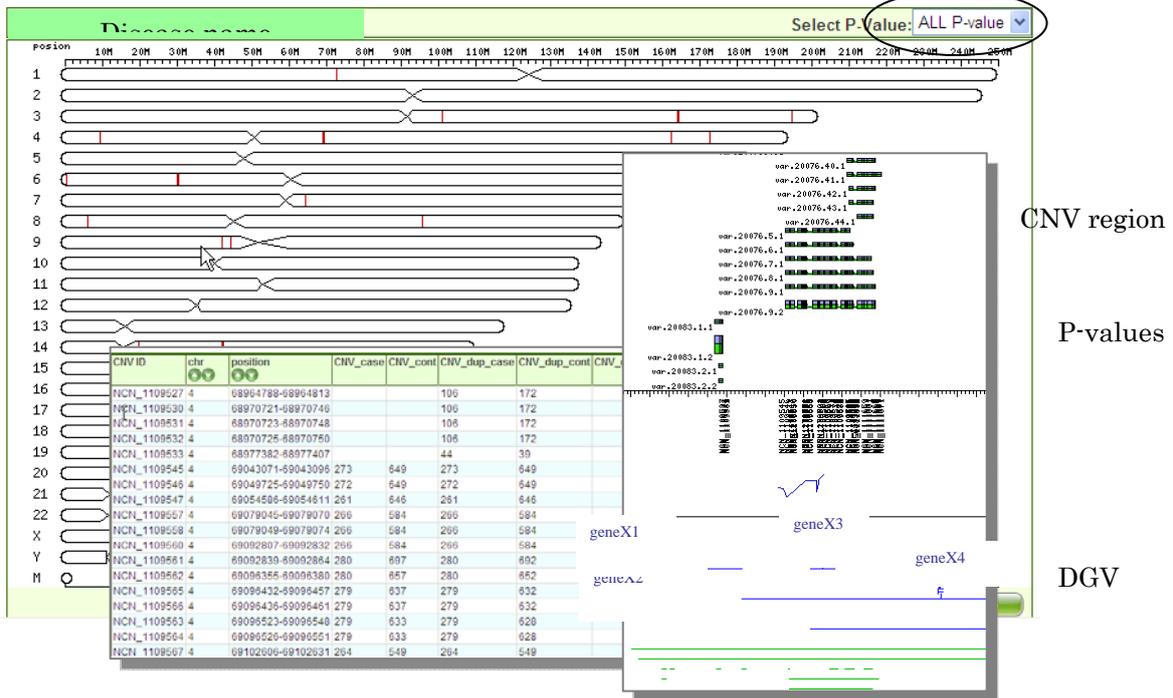


図 3-3 CNV case-control DB の snapshot

Integrated Database
DB development for medical applications from disease analysis

English

統合データベースプロジェクト ヒトゲノムバリエーションデータベース共有方針

現在、疾患関連遺伝子の探索においては非常に勢い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、様々なデータの共有に向けて施策を進めている。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シークエンシング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6カテゴリーのデータを3つのレベルに分類して、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」の外部の委員を複数含めることとし、科学面・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
データの種類の	GWASデータ	カテゴリ-A (頻度データ、統計解析結果)	カテゴリ-B (個体レベルでのCNV情報)
	リシークエンシングデータ	カテゴリ-E (公知のリシークエンシングおよび実用データ)	カテゴリ-F (リシークエンシングおよび実用データ)
データ使用時の手続き等	ウェブサイトにおいて、閲覧可能。但しカテゴリ-Aのデータを大量取得する場合はレベル2と同様の申請を要する。	氏名、職名、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、使用期間に応じて、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

図 4-1 データ共有方針（日本語）の snapshot

Integrated Database																		
DB development for medical applications from disease analysis																		
Japanese																		
<p>Integrated Database Project Human Genome Variation Research Database (HGVRD) Sharing Policy</p> <p>At present, the volume of information concerning human genes is increasing at an astounding rate in the discovery of disease-associated genes, and efficient use of such a huge volume of data has become an urgent challenge. The Integrated Database Project, a research and development project commissioned by the Ministry of Education, Culture, Sports, Science and Technology, is aiming at extensive sharing of these data, and implementing measures to enable various data to be shared. This Policy defines the data-sharing policy of the "Integrated Database Project - Disease Analysis DB Group" (Lead by K. Tokunaga, University of Tokyo) (hereafter, "Disease Analysis DB Group") concerning the genome-wide SNP and association study (GWAS) data and the resequencing data for disease-associated genes. These data are classified into six categories according to the data contents and further into three levels for data access, and a separate sharing policy is established for each level. The Data Sharing Review Board of Disease Analysis DB Group ("DSRB") is established within the Disease Analysis DB Group to review deposit of the data to database and access to the data. Members of the Board include multiple members from outside the Disease Analysis DB Group, who will examine each case from scientific, ethical and other perspectives.</p> <p><Outline of This Sharing Policy></p> <p>Data Category/Level and Procedures for Data Use</p> <table border="1"> <thead> <tr> <th colspan="2">Level</th> <th>1</th> <th>2</th> <th>3</th> </tr> </thead> <tbody> <tr> <td rowspan="2">Type of Data</td> <td>GWAS data</td> <td>Category A (Frequency data and statistical values)</td> <td>Category B (Individual-level copy number variation [CNV] information)</td> <td>Category C (GWAS genotype data) Category D (GWAS raw data)</td> </tr> <tr> <td>Resequencing data</td> <td>Category E (Publicly available)</td> <td>Category F (Resequencing data and</td> <td>(Not applicable)</td> </tr> </tbody> </table>					Level		1	2	3	Type of Data	GWAS data	Category A (Frequency data and statistical values)	Category B (Individual-level copy number variation [CNV] information)	Category C (GWAS genotype data) Category D (GWAS raw data)	Resequencing data	Category E (Publicly available)	Category F (Resequencing data and	(Not applicable)
Level		1	2	3														
Type of Data	GWAS data	Category A (Frequency data and statistical values)	Category B (Individual-level copy number variation [CNV] information)	Category C (GWAS genotype data) Category D (GWAS raw data)														
	Resequencing data	Category E (Publicly available)	Category F (Resequencing data and	(Not applicable)														

図 4-2 データ共有方針（英語）の snapshot

②GWAS DB の統計遺伝学手法および解析ツールの開発（東海大学実施）

GWAS は第 1 ステージ（探索）のタイピング、第 2 ステージ（再現性確認と高精度マッピング）のタイピングからなるが、平成 19 年度では、3 疾患の関連解析に用いる統計遺伝学手法の開発を行った。また平成 20 年度には、グラフィカルモデリングなどの統計的にノード間の依存関係を予測する枠組みを用い、SNP 間の相互作用を検出する手法の開発を行った。

平成 21 年度は、引き続き統計遺伝学手法の開発を行った。具体的には、疾患の発症リスクに対する、複数の SNP の寄与度を同時に評価する手法として、Bayesian shrinkage 法のアルゴリズムをソフトウェア化し、シミュレーション実験により、本手法の有効性を検証した。

シミュレーション実験に用いた仮想データセットのサイズは 16 × 480（それぞれ SNP、サンプルの数に相当）、SNP のアレル頻度はすべて 0.5 とした。また、16SNP のうち、SNP6 と SNP14 が相互作用を示しつつ、量的形質としての発症リスクに寄与するものとした。開発したツールを、上記のデータセットに適用したところ、表 2 に示すように、モデルが正しく推定されており、この結果は本手法の有効性を示唆するものである。

表 2. Bayesian shrinkage 法の有効性の検証

解析の結果、仮想データセットを発生させた際のパラメータ（SNP の組み合わせなど）を正しく推定している。

$2*\log(L_1/L_0)$	σ_e^2	σ_g^2	Marker(s) in the model
48.734204	95.841471	7.211297	6 14
39.924525	96.732207	7.120908	9 14
36.931467	97.111046	6.721468	8 14
33.115573	98.907721	4.885727	14
32.216619	95.454627	8.571798	8 9 14
28.646645	95.853858	8.691143	5 14 16
25.925930	98.742687	4.758265	12 14
24.435480	98.817457	5.098832	6 9
23.193275	98.846928	5.474342	10 14
23.184568	99.088710	4.429338	3 14
22.899752	99.174043	4.373345	13 14
22.440157	99.208932	4.333612	11 14
22.150224	99.246825	4.255357	4 14
20.509343	99.258172	4.846660	14 15
19.354183	99.645031	3.964763	14 16
19.315728	99.625065	4.000905	1 14
17.851743	98.228258	5.507904	3 12 14
16.250350	100.149147	3.469943	9 12
14.878450	100.219337	3.704009	1 6
14.592027	100.492716	2.995522	9 10

:

また、stepwise ロジスティック回帰法など、従来広く用いられてきたものと比較して、より多くの SNP で構成された複雑なモデルの評価が可能であり、各遺伝子型効果の事前情報を考慮した、より柔軟な解析も可能である。

(3) リシークエンシングによる臨床情報・ゲノム情報 DB の構築

①リシークエンス DB の臨床情報・ゲノム情報 DB の構築と解析手法の開発（東京大学大学院医学系研究科、東京大学医学部附属病院、日立製作所が実施）

臨床現場で役立つことを目的としたリシークエンスデータベースの構築を行っている。平成 19 年度は、ALS（筋萎縮性側索硬化症）に関するリシークエンスデータベースを構築した。平成 20 年度は、ALS の DB 拡張と共に、新たにパーキンソン病のリシークエンスデータベースの構築を行った。

平成 21 年度は、PD（パーキンソン病）については、実験データの追加（東京大学医学部附属病院実施）と文献データの追加（日立製作所実施）を行い、DB を 3 月に公開した。また、家族性痙性対麻痺、及び、副腎白質ジストロフィーについては、リシークエンス DB の全体の DB 構築（既知配列、配列 2 次構造表示、蛋白質 3 次元構造表示、患者配列表示、オルソログス配列表示、既知文献情報表示など）と、実験配列情報の DB 化を東京大学医学部附属病院実施、UniProt の 2 次構造データ、Entrez Gene からの遺伝子名情報、Interpro によるドメイン構造など外部データベースから本 DB に登録すべき情報の取り込みスキーム構築を東京大学大学院医学系研究科が実施した（図 5-1, 5-2, 5-3, 5-4）。また、家族性痙性対麻痺、及び、副腎白質ジストロフィー関連遺伝子の mutation (deletion, insertion を含む) と、家族性痙性対麻痺、及び、副腎白質ジストロフィー病との関係性に関する文献各疾患 100 本以上から、mutation の位置、頻度、家系情報と共に、発症年齢、どのような症状か等の臨床情報をまとめ、データベース化した。mutation の頻度に関して、オッズ比、P 値などの統計解析を行った。更に、これらの文献情報、及びリシークエンス実験データの mutation について、近傍配列をゲノムにマッピングすることによりゲノム上の位置あわせを行った。（mutation は、どの build version の参照ゲノムを基準にするか、どのアクセッション番号の mRNA を基準にするか、もしくは coding region を基準にするのか、UTR を含めて考えるか等、複数定義があるため慣習的な名前の位置が不明確であり、曖昧性を排除するためには位置の特定が必要）家族性痙性対麻痺、及び、副腎白質ジストロフィーに関係する配列について、既知の蛋白質立体構造がないものについては、蛋白質立体構造の 3 次構造予測を行い、mutation 位置が蛋白構造上どこにあるか明示可能とした。また、domain, motif 位置なども同定し、新規 mutation が与えられたとき、どのような遺伝子機能に影響があるか検討可能とした。更に、主な生物種の orthologus sequence の multiple alignment 等の解析を実施し、mutation と進化の関係の検討を可能とした。（日立製作所実施）

以下、リシークエンス DB の snap shot である。

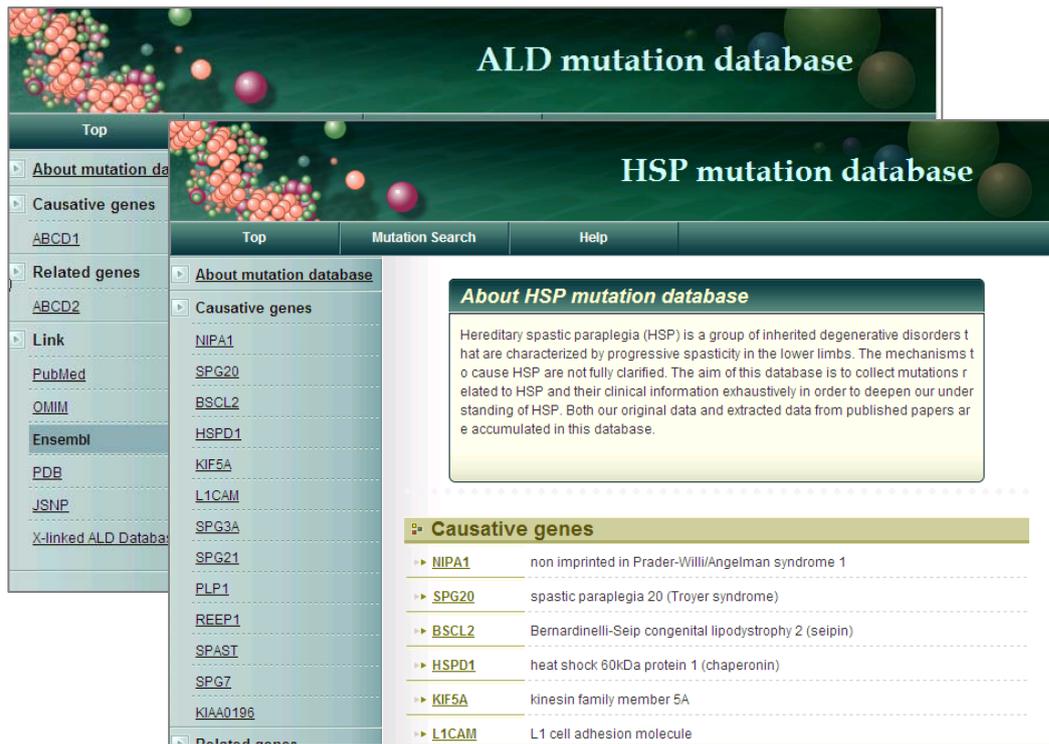


図 5-1 ALD, HSP リシークエンス DB トップ画面



図 5-2 ALD リシークエンス DB 検索結果画面—配列表示 (赤字が変異、#が2つ以上の変異を持つ部位)

(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								
A19fs		P33897								

age at onset	onset duration (y/m)	examination age	Disease type	Initial symptom	Onset site	Reference (title)	Journal, year, volume, page	Comments
late 10's		40	ALM	stiffness	leg	PMID : 11728802 : ABCD1 translation-initiator mutation demonstrates genotype-phenotype correlation for ALM.	O'Neill GN, Aoki M, Brown RH Jr. Neurology. 2001; 57:1950-62.	personal data in same family(A) below (A)
			cALD/ALM			PMID : 8651200 : Mutational and protein analysis of patients and heterozygous women with X-linked adrenoleukodystrophy.	Feigenbaum V, Lombard-Platel G, Guibouk S, Sarde CO, Mandel JL, Aubourg P. Am J Hum Genet. 1995; 58:1135-44.	
			ccALD			PMID : 7826602 : Spectrum of mutations in the gene encoding the adrenoleukodystrophy protein.	Ligtenberg MJ, Kemp S, Sarde CO, van Geel SM, Kester WJ, Barm PC, Mandel JL, van Oost BA, Sultus SA. Am J Hum Genet. 1995; 56:44-50.	

図 5-3 ALD リシーケンス DB 文献情報表示画面

Structure information

Exon	Start AA	End AA	Structure
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1	745	Chain
1	92	112	Transmembrane region
1	131	151	Transmembrane region
1	238	258	Transmembrane region
2	333	353	Transmembrane region
5	473	493	Transmembrane region
1, 2, 3	94	386	Domain
5, 6, 7, 8, 9, 10	474	700	Domain
6	507	514	Nucleotide phosphate-binding region
1	67	186	Region of interest
1	214	214	Glycosylation site
1	13	13	Sequence variant
1	88	88	Sequence variant
1	90	90	Sequence variant
1	98	98	Sequence variant
1	99	99	Sequence variant
1	103	103	Sequence variant
1	104	104	Sequence variant
1	104	104	Sequence variant
1	105	105	Sequence variant
1	105	105	Sequence variant
1	107	107	Sequence variant
1	108	108	Sequence variant
1	108	108	Sequence variant
1	113	113	Sequence variant
1	113	113	Sequence variant

図 5-4 ALD リシーケンス DB 構造情報表示画面

(4) プロジェクトの総合的推進

随時ミーティング等（全体会議が4回、部分会議が5回）を開き、参加機関連携の下、上記プロジェクトを推進した。

全体会議：2009/6/5, 6/19, 7/27, 2010/2/10（出席者：東大徳永、西田、東大病院 辻、東海大学 井ノ上、日立 小池、議題：データの受け入れ方法と、データクオリティーチェックについて。）

部分会議：2009/5/8, 7/6, 9/29, 10/8、（出席者：東大徳永、西田、日立小池、議題：CNV DB, CNV association DBのデータベーススキーマとインターフェースについて）

2010/3/5（出席者：東大辻、後藤、高橋、福田、日立小池、吉田、議題：ALD, HSPリシークエンスDBの実験データ追加方法について）

(5) DB アクセス状況について

GWAS DB の訪問者数は公開用 DB で 2008 年度（8 月から公開のため 12 月までの 5 ヶ月）は 1136 名（88,973hit）、2009 年度は 4,646 名（178,531hit）、2010 年度は 1 月-3 月までで 2,178 名（85,434hits）であり、少なくとも約半分のアクセスが国内のアカデミアからである。内部用データに関しては、複数の疾患部会に利用していただき、使用に関する感想及び要望を伺っているが、様々な情報が掲載されるということでおおむね良好である。また、海外の主要なデータベースである HGV baseG2P database おおおよび European Genotype Archive(EGA)の担当者より DB 連携の打診を受けており、今後、海外連携も視野に入れていきたい。まだ、十分な広報活動ができていないため、今後は、学会活動、論文発表等で積極的に本 DB の存在をアピールし、データの登録数を充実させていきたい。

2.3 成果の外部への発表

別紙参照。

2.4 活動（運営委員会等の活動等）

運営のための各種委員会：データ公開および GWAS 生データ（個体毎の遺伝子型およびタイピング生データ）を含むデータの共有に関する方針案および関連する各種書類案を作成し、データアクセス倫理検討委員会において審議して最終案を決定した。この方針と書類一式を Web 上で公開するとともに、データアクセス倫理審査委員会を新たに発足した。また、検体のゲノム多型・変異解析情報および臨床情報は各々の疾患の研究グループ（パニック障害、アルツハイマー病、肝炎、1 型糖尿病など、本 PJ とは独立に存在している）から提供いただいている。

（疾患解析 DB 開発委員会）

2009/6/5, 7/27 議題：統合 DB プロジェクト「疾患解析 DB 開発」データの公開および共有の方針案についての検討、出席者：位田（京都大）、井ノ上（東海大）、加藤（京都大）、辻（東京大）、徳永（東京大）、福嶋（信州大）、武藤（東京大）、米本（東京大）、田中（文科省）、箕輪（統合 DB 中核機関）、下川（東京医科歯科大）、井戸（東京医科歯科大）、小池（日立）、西田（東京大）

（データアクセス倫理検討委員会）

2010/2/10 議題：レベル 3 のデータ提供およびデータ利用の希望者に対するデータベース共有委員会が行う具体的な手順について、出席者：位田（京都大）、井ノ上（東海大）、加藤（京都大）、辻（東京大）、徳永（東京大）、武藤（東京大）、米本（東京大）、田中（文科省）、箕輪（統合 DB 中核機関）、小池（日立）、西田（東京大）

2.5 実施体制

別表1 平成21年度に於ける実施体制

研 究 項 目	担当機関等	研究担当者
(1) 標準 SNP DB の構築	東京大学大学院医学系研究科 東京大学大学院医学系研究科 東海大学医学部 東海大学医学部	◎ 徳永勝士 西田奈央 井ノ上逸朗 成田 暁
(1)GWAS DB の構築 －GWAS 第一ステージ疾患関連 SNP 探索手法の研究開発	東京大学大学院医学系研究科 東京大学大学院医学系研究科 東海大学医学部 東海大学医学部 (株) 日立製作所 (株) 日立製作所 (株) 日立製作所	◎ 徳永勝士 西田奈央 井ノ上逸朗 成田暁 小池麻子 吉田真希子 橋詰 明英
(3)リシーケンシングによる臨床情報・ゲノム情報 DB の構築 －リシーケンス DB の解析手法の開発	東京大学医学部附属病院 東京大学医学部附属病院 東京大学医学部附属病院 東京大学医学部附属病院 東京大学大学院医学系研究科 東京大学大学院医学系研究科 (株) 日立製作所 (株) 日立製作所	○ 辻省次 後藤順 高橋祐二 福田陽子 徳永勝士 西田奈央 小池麻子 吉田真希子

注1. ◎：課題代表者、○：サブテーマ代表者

注2. 本業務に携わっている方は、全て記入

2.6 整備実績一覧

(1) 保有データ情報

※ 貴機関グループ内で保有するデータに関して、以下の内容を記述して下さい。

(1-1) データの種類

①生物種	Homo sapiens
②試料・ライブラリー等の種類、数	ナルコレプシー約 200 検体、パニック障害 約 750 検体、脳動脈瘤 200 検体、そのほか、多系統萎縮症 200 検体、B型肝炎、C型肝炎、老年性アルツハイマー、若年性アルツハイマー、膵炎、膵臓ガン、膝関節症、糖尿病など合計約 4500 検体の 50-90 万 SNP の遺伝子型データと約 1500 検体の 1000SNP の遺伝子型データ（但し、同一の検体の異なるプラットフォームの結果を含む） リシーケンスデータ ALS に関連する遺伝子の mutation 情報及び、臨床情報
③測定方法	Affymetrix, Illumina の 50-90 万の SNP タイピングセット
④データの内容	検体の性別、疾患情報などの基本情報、genotype データ、genotype calling 前の画像生データ
⑤その他、特記事項	

(1-2) データソース

① 現在のデータ量	ナルコレプシー約 200 検体、パニック障害 約 750 検体、脳動脈瘤 200 検体、そのほか、多系統萎縮症 200 検体、B型肝炎、C型肝炎、老年性アルツハイマー、若年性アルツハイマー、膵炎、膵臓ガン、膝間接症、糖尿病など合計約 4500 検体の 50-90 万 SNP の遺伝子型データと約 1500 検体の 1000SNP の遺伝子型データ（但し、同一の検体の異なるプラットフォームの結果を含む）そのほか、GeMDBJ の健常者合計 3200 検体、疾患 4160 検体の数万から 30 万の遺伝子型頻度データ
②データ区分	<input checked="" type="checkbox"/> 自前 <input checked="" type="checkbox"/> 第三者 <input type="checkbox"/> 文献データ <input checked="" type="checkbox"/> 計算結果等の二次データ <input type="checkbox"/> その他（下欄に詳細を記述） ※複数選択可。二次データのみの保有は不可。
③将来の増加の見込み	厚労科研、CREST のデータも収集予定。
④権利関係	所有者（各データの所持者） 公開（ <input type="checkbox"/> 可 <input type="checkbox"/> 否 <input checked="" type="checkbox"/> その他 [genotype frequency data, や解析結果公開可能、個々の genotype data は限定された研究者に開示可能]） GWAS-DB, GWAS 標準 DB、CNV 標準 DB は既に公開 リシーケンス DB ALS, Parkinson 病について公開 CNV case control DB については、論文化後に公開 ※既に公開している場合は URL を「⑤その他、特記事項」に記述 ※権利関係が未解決で、プロジェクト期間内に解決の見込みがある場合は、解決のための手立て等を「⑤その他、特記事項」に詳述
⑤その他、特	GWAS-DB: http://gwas.lifesciencedb.jp/cgi-bin/gwasdb/gwas_t

記事項	op.cgi GWAS 標準 DB: http://gwas.lifesciencedb.jp/snpdb/snp_top.php CNV 標準DB: http://gwas.lifesciencedb.jp/cgi-bin/cnvdb/cnv_top.cgi ALS: https://reseq.lifesciencedb.jp/resequence/SearchDisease.do?targetId=1 Parkinson: https://reseq.lifesciencedb.jp/resequence/SearchDisease.do?targetId=2
-----	--

(1-3) データの管理状況

①更新頻度等の管理状況、体制	内部用データベースは東京大学大学院においてあり、サーバー管理は日立製作所がおこなっているが、常駐 SE がいるわけではない。更新は新たなデータが提供されたとき。 公開用データベースは中核機関のサーバーに搭載 raw データは東京大学医学系大学院にて管理 倫理審査委員会開催後は、配布用 raw データについては中核機関のサーバでも管理予定。
②その他、特記事項	

※ 既にデータベースを保有している場合は、以下についても記述して下さい。

(1-4) データベース関係

①DB 管理者数	2
②キュレータ・アナレータ数	0
③データ構造	Relational DB
④DB 管理ソフト	Mysql
⑤サーバの OS	Redhat enterprise linux ES v4
⑥サーバ規模	Dell powerEdge 2900
⑦DB へのアクセス数	2008 年度 (8 月から公開のため 12 月までの 5 ヶ月) は一意の訪問者数が 1,136 名 (88,973hit)、2009 年度は 4,646 名 (178,531hit) 2010 年度は 1 月-3 月までで 2,178 名 (85,434hits)
⑧独立 IP 数	2008 年度 696 2009 年度 2,846, 2010 年度 1,215
⑨その他、特記事項	

(2) データ (又は DB) の連結、統合化整備 (※試験的、限定的公開済みのものも含む。)

通番	データ (又は DB) の名称 ※URL があれば記述	公開 / 未公開	概要 (データの種類 (生物種)・数量 (kB 等)、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述) ※ 公開している場合は、開始年月、利用状況 (平均利用者数、アクセス数、ダウンロード数等の数値的指標で記述) ※ 必要に応じて画面コピー等の図表添付可
1	標準 SNP DB	公開	日本人健常者の 30 万 SNP 約 200 検体、50 万 SNP 約 500

			<p>検体、90万 SNP 約 200 検体の genotype frequency, allele frequency, Hardy-weinberg 平衡検定値、ハプロタイプ頻度など。</p> <p>今後の計画：データを随時登録していく。</p>
2	GWAS DB	公開	<p>SNP ごとの genotype frequency, allele frequency, call rate, Hardy-weinberg 平衡検定値、genotypic model, allelic model, additive risk model, recessive model, dominant model など主な遺伝統計値を登録している。copy number variation, OMIM などの他の情報と共に上記計算結果をグラフ表示することが可能である。内部用、公開用の 2 つの DB がある。</p> <p>進捗：システムを構築し、ナルコレプシー、脳動脈瘤、多系統萎縮症などを登録し、一部のデータは公開している。</p> <p>今後の計画：ユーザーフレンドリーになるように、インターフェース周りの改良を行うとともに、量的形質用に DB とインターフェースの拡張を行う。</p> <p>また、学会発表、論文等により、データの submission を広く呼びかけていく。</p>
3	標準 CNV DB	公開	<p>日本人健常者 約 200 名の copy number variation を解析し、登録。</p> <p>今後の計画：データ登録数を増やす。</p>
3	CNV case control DB	未公開	<p>アルツハイマー、パニック障害などについて、CNV ベースの case control 解析を行い、登録。</p> <p>今後の計画：データ登録数を増やし、論文採択後に公開</p>
4	ALS リシークエンス DB	公開	<p>ALS (筋萎縮性側索硬化症) に関するリシークエンスデータベースであり、東京大医学部附属病院で産出した ALS 関連遺伝子のリシークエンスデータ及び臨床データのほか、フルペーパーから抽出した mutation の位置、頻度、家系情報と共に、発症してから何年で人工呼吸器をつけたか、どのような症状か等の臨床情報、更には、蛋白質の 2 次構造、3 次構造などのデータも登録している。</p> <p>今後の計画：学会等で広く呼びかけてデータ登録数を増やす。</p>
5	PD リシークエンス DB	公開	<p>パーキンソン病に関するリシークエンスデータベースであり、東京大医学部附属病院で産出した ALS 関連遺伝子のリシークエンスデータ及び臨床データのほか、フルペーパーから抽出した mutation の位置、頻度、家系情報と共に、発症してから何年間生存したか、どのような症状か等の臨床情報、更には、蛋白質の 2 次構造、3 次構造などのデータも登録している。</p> <p>今後の計画：学会等で広く呼びかけてデータ登録数を増やす。</p>
6	ALD リシークエンス DB	未公開	<p>副腎白質ジストロフィー(ALD)に関するリシークエンスデータベースであり、東京大医学部附属病院で産出した ALD 関連遺伝子のリシークエンスデータ及び臨床データのほか、フルペーパーから抽出した mutation の位置、頻度、家系情報と共に、発症してから何年間生存したか、</p>

			<p>どのような症状か等の臨床情報、更には、蛋白質の2次構造、3次構造などのデータも登録している。</p> <p>進捗：システムの基本機能はほぼ完成している。実験データを追加する。</p> <p>今後の計画：学会等で広く呼びかけてデータ登録数を増やす。</p>
7	HSP リシークエンス DB	未公開	<p>家族性痙性対麻痺(HSP)に関するリシークエンスデータベースであり、東京大医学部附属病院で産出したHSP関連遺伝子のリシークエンスデータ及び臨床データのほか、フルペーパーから抽出したmutationの位置、頻度、家系情報と共に、発症してから何年間生存したか、どのような症状か等の臨床情報、更には、蛋白質の2次構造、3次構造などのデータも登録している。</p> <p>進捗：システムの基本機能はほぼ完成している。実験データを追加する。</p> <p>今後の計画：学会等で広く呼びかけてデータ登録数を増やす。</p>

(3) DB基盤システム、ツール等開発成果物の整備（※試験的、限定的公開済みのものも含む。）

通番	DB基盤システム、ツール等の名称	公開／未公開	<p>概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）</p> <p>※ プログラムプロダクトに限らず、データ形式共通化、標準化のための仕様書、共通規約等のドキュメントについてもリリースしているものは対象とする。（リリース済みドキュメントは参考として目次一覧、抜粋を添付）</p> <p>※ 必要に応じて画面コピー等の図表添付可</p>
1	疾患-SNPネットワークマップ作成のためのプログラム	未公開	<p>疾患-SNPネットワークマップをPC (path consistency) アルゴリズムを利用して計算する。</p>

(4) その他の成果物 ((2)、(3) に該当しないもの)

通番	名称	公開／未公開	<p>概要</p> <p>※ 必要に応じて画面コピー等の図表添付可</p>

別紙

学会等発表実績

委託業務題目：「疾患解析から医療応用を実現する DB 開発」

機 関 名：東京大学大学院医学系研究科・東京大学医学部附属病院・東海大学医学部・日立製作所中央研究所

1. 学会等における口頭・ポスター発表（表の左のマーク：委託費で執行されたもののうち、100%本委託費によるものは◎、一部外部資金も入っているものは○）

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期
◎疾患解析から医療応用を実現する DB 開発-ゲノムワイド関連解析データベースの構築 (ポスター)	小池麻子、西田奈央、井ノ上逸朗、辻省次、徳永勝士	統合 DB プロジェクトシンポジウム	2009/06/12
◎疾患解析から医療応用を実現する DB 開発-リシーケンスデータベースの構築- (ポスター)	吉田真希子、高橋裕二、後藤順、徳永勝士、辻省次	統合 DB プロジェクトシンポジウム	2009/06/12
◎Comparisons of copy number variation detection methods (poster)	A. Koike, N. Nishida, K. Tokunaga	Cold Spring Harbor Meeting, Personal Genome	2009/09/14-17
◎統合データベースプロジェクトにおけるコピー数多型データベース (poster)	小池麻子、西田奈央、井ノ上逸朗、辻省次、徳永勝士	日本人類遺伝学会第 54 回大会	2009/09/23-26
◎統合データベースプロジェクトにおけるゲノムワイド関連解析・コピー数変異データベース開発 (口頭発表)	小池麻子	日本人類遺伝学会第 54 回大会	2009/09/23-26
◎SNP アレイ解析と疾患ゲノム研究 (口頭発表)	西田奈央	日本人類遺伝学会第 54 回大会	2009/09/23-26
◎筋萎縮性側索硬化症 (ALS) 遺伝子変異データベース (ポスター)	吉田真希子、高橋裕二、福田陽子、後藤順、辻省次	日本人類遺伝学会第 54 回大会	2009/09/23-26
○日本人集団のゲノムワイド関連解析からみえてきた脳動脈瘤感受性遺伝子多型 (口頭発表)	秋山 康一、中岡 博史、安野勝史、成田 暁、田嶋 敦、羽田 明、井ノ上 逸朗	日本人類遺伝学会第 54 回大会	2009/09/23-26
◎ Path consistency (PC) アルゴリズムを応用した遺伝子間相互作用解析 (口頭発表)	成田 暁、田嶋 敦、安野勝史、宮下 哲典、桑野 良三、井ノ上 逸朗	日本人類遺伝学会第 54 回大会	2009/09/23-26

○臨床情報と 9p21 領域の遺伝的多型情報を用いた脳動脈瘤感受性に関する多変量ロジスティック回帰分析 (口頭発表)	中岡 博史、高橋 朋子、崔泰林、田嶋 敦、井ノ上 逸朗	日本人類遺伝学会第 54 回大会	2009/09/23-26
◎ A Variation Database for Amyotrophic Lateral Sclerosis (poster)	M. Yoshida, Y. Takahashi, A. Koike, Y. Fukuda, J. Gotoh, S. Tsuji,	The 59th Annual Meeting of the American Society of Human Genetics	2009/10/20-24
◎ Databases for copy number variations (poster) in Japanese Integrated Database Project	A. Koike, N. Nishida, I. Inoue, S. Tsuji, K. Tokunaga	The 59th Annual Meeting of the American Society of Human Genetics	2009/10/20-24
○ Genome-wide association studies for complex diseases and therapy responses. (Oral)	K. Tokunaga	The 9th Annual Meeting to East Asian Union of Human Genetics Societies	2009/11/19
◎ Variation databases for neurodegenerative disorders (Oral)	A. Koike, M. Yoshida, Y. Takahashi, J. Gotoh, S. Tsuji	International Symposium on Applied Genomics 2009	2009/12/3-4
◎ Genome-wide Association Database in Japanese Integrated Database Project. (Oral)	K. Tokunaga	International Symposium on Applied Genomics 2009	2009/12/3-4
◎ A Variation Database for Amyotrophic Lateral Sclerosis (poster)	吉田真希子、高橋裕二、福田陽子、後藤順、辻省次	分子生物学会	2009/12/9-12

2. 学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期
○Narcolepsy is strongly associated with the TCR alpha locus.	Hallmayer J, Faraco J, Ling L, Hesselson S, Winkelmann J, Kawashima M, Mayer G, Plazzi G, Nevsimalova S, Bourgin P, Hong S, Honda Y, Honda M, Hogl B, Longstreth WT, Montplaisir J, Kemlink D, Eiben M, Chen J, Musone SL, Akana M, Miyagawa T, Duan J, Desautels A, Erthardt C, Hesla PE, Poli F, Frauscher B, Jeong J-H, Lee S-P, Ton TGN, Kvale M, Kolesar L, Dobrovolna M, Nepom GT, Salomon D, Wichmann HE, Rouleau GA, Gieger C, Levinson DF, Gejman PV, Meitinger T, Young T, Peppard P, Tokunaga K, Kwok P-Y, Risch N, and Mignot E	Nature Genetics 41(6):708-711.	2009 Jun
◎ Genome-wide association database developed in the Japanese Integrated Database Project.	A. Koike, N. Nishida, I. Inoue, S. Tsuji, K. Tokunaga	Journal of Human Genetics 54(9):543-546.	2009 Sep
○ Genome-wide association of IL28B with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C	Tanaka Y, Nishida N, Sugiyama M, Kurosaki M, Matsuura K, Sakamoto N, Nakagawa M, Korenaga M, Hino K, Hige S, Ito Y, Mita E, Tanaka E, Mochida S, Murawaki Y, Honda M, Sakai A, Hiasa Y, Nishiguchi S, Koike A, Sakaida I, Imamura M, Ito K, Yano K, Masaki N, Sugauchi F, Izumi N, Tokunaga K, Mizokami M	Nature Genetics 41(10):1105-1109.	2009 Oct

○ Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse.	Nakaoka H, Inoue I	Journal of Human Genetics 54: 615-623.	2009 October
○ Replication of a genome-wide association study of panic disorder in a Japanese population.	Otowa T, Tanii H, Sugaya N, Yoshida E, Inoue K, Yasuda S, Shimada T, Kawamura Y, Tochigi M, Minato T, Umekage T, Miyagawa T, Nishida N, Tokunaga K, Okazaki Y, Kaiya H, Sasaki T.	Journal of Human Genetics 55(2):91-96.	2010 Feb
◎ゲノムワイド関連解析データの公開と共有	小池麻子, 徳永勝士	実験医学	2009年7月
◎ゲノムワイド関連解析(GWAS)の進展(特集編集)	徳永勝士	MSD(メディカル・サイエンス・ダイジェスト)	2009年7月
○疾患感受性遺伝子とゲノムワイド関連解析	西田奈央, 徳永勝士	治療学	2009年3月
○テーラーメイド医療をめざした疾患感受性遺伝子のゲノムワイド探索	西田奈央, 徳永勝士	遺伝子医学MOOK	2009年9月
○疾患関連遺伝子を探し出すためのSNP解析	西田奈央, 徳永勝士	肝胆膵2009	2009 Dec