

平成 23 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業「統合化推進プログラム」
平成 23 年度採択 研究代表者

徳永勝士
東京大学大学院医学系研究科・教授

ヒトゲノムバリエーションデータベースの開発

§1. 研究実施体制

(1)「ヒト多型・変異 DB の開発」グループ

- ① 研究代表者: 徳永勝士 (東京大学大学院医学系研究科、教授)
- ② 研究項目
 - ・ゲノムワイド関連解析のデータベースの機能拡張とデータの収集
 - ・次世代シーケンサー用 variation DB の構築
 - ・HLA 領域における多型・変異の DB 構築

(2)「ヒト多型・変異データの体系化」グループ

- ① 主たる共同研究者: 辻 省次 (東京大学医学部附属病院、教授)
- ② 研究項目
 - ・新学術領域研究「パーソナルゲノム情報に基づく脳疾患メカニズムの解明」で産出した神経変性疾患の新規変異の整理と登録
 - ・新学術領域研究の「ゲノム支援」で産出した変異データの整理と登録

(3)「統計遺伝学的計算手法の開発」グループ

- ① 主たる共同研究者: 井ノ上 逸朗 (国立遺伝学研究所、人類遺伝研究部門、教授)
- ② 研究項目
 - ・新学術「ゲノム支援」で産出された変異データのクオリティーコントロールを行うとともに登録を行う。
 - ・新学術以外で産出された変異データの収集と整理・登録
 - ・HLA ハプロタイプの配列決定およびDB登録
 - ・GWAS メタ解析手法の開発と適用

(4)「多型・変異関連の計算手法の開発」グループ

①主たる共同研究者:小池 麻子 (日立製作所 公共システム事業部 主任技師)

②研究項目

- ・GWAS データの解析と運用預け入れ・データ再配布と解析・登録。
- ・GWAS データの解析パイプラインの新プラットフォーム対応や解析機能追加などの機能拡張を行う。
- ・次世代シーケンサー以外の実験結果、及び、文献から抽出した変異に関する **variation DB** の構築とデータ収集。
- ・次世代シーケンサーで利用するための日本人健常者の標準ゲノム構築に関する計算手法の開発。

§2. 研究実施内容

1) 研究の目的

近年の高速大容量の SNP タイピング技術の確立により、数十万～数百万の SNP を利用したゲノムワイドな関連解析が可能となり、ここ6年ほどは世界レベルで大規模なゲノムワイド関連解析が実施され、多くの疾患関連 SNP を同定するという成功を収めた。一方、ここ2・3年ほどは、次世代シーケンサーの革新的な高スループット化に伴い、リシーケンシングによるゲノムワイド疾患関連変異の探索が可能となりつつあり、新規疾患関連変異の発見も相次いでいる。これらの実験データは大量かつ貴重であることから、データの永続的な管理と共に、研究者間での情報の共有化は重要である。

2006-2010年度の統合DBプロジェクトにおいては、GWAS-DBを構築しデータの集約的な管理を行うとともに、データの預け入れと再配布を行う枠組みを構築・運用し、研究者間のデータの共有化を図ってきた。本プロジェクトにおいては、Human variation DBを構築し、GWASと共に次世代シーケンサーによって産出された変異データについてもデータの預け入れと再配布を行い、データの散逸を防ぐと共に研究者間のデータの共有化を促進することを目的とする。更に、過去産出された変異も含め、変異と疾患との関係を体系化していく。

2) 概要

GWAS-DBをベースに引き続きGWASデータの受け入れと再配布を行うと共に、GWAS-DBのimputation機能などの機能追加を行う。また、新たにHuman variation DB, HLA DBを構築し、次世代シーケンサーにより検出された新規変異データを登録すると共に、過去に報告された変異についても文献などから情報を抽出し、複数の疾患に共通して関与する遺伝子、同一疾患における同一遺伝子変異箇所依存した症状の差異、及び、原因・関連遺伝子(変異)の集団間差異などが俯瞰できるよう、疾患・変異・臨床情報を整理・体系化し横断的に検索できるようにする。

また、次世代シーケンサーにおける実験手法、解析手法の差によるクオリティーコントロールを行う仕組みを整える。更に、新規に産出される健常者ゲノムデータと共に、1000人ゲノムのデータ、及び、GWASの健常者データなども利用し、次世代シーケンサーによる変異解析に利用できる日本人健常者の標準的なゲノム配列を整備・提供する。

3) 進捗状況

Human Variation DB及び、HLA DBを構築し、公開すると共に、学会等でデータの登録を呼びかけている。また、GWAS-DBについては、前プロジェクトから引き続き、データの預け入れと再配布を実施している。

4) 研究成果

4-1) Human variation DBの構築

Human variation DBでは、次世代シーケンサーによる変異データ、その他の実験手法による変異データ、また、文献から抽出した変異データを登録するためのDBを構築した。次世代シーケンサーからのデータは、データのQuality Controlのために、変異を抽出した際のプログラム

名称、パラメータ等の計算の詳細についても登録できるようにしている。文献情報については、実験の種類、case-control GWAS での P-value、オッズ比、臨床情報についても登録できるようにしている。また、reference genome としては、hg19 の他、主に 1000 ゲノムの日本人の健常者データから計算した reference データも同時に表示できるようにしている。

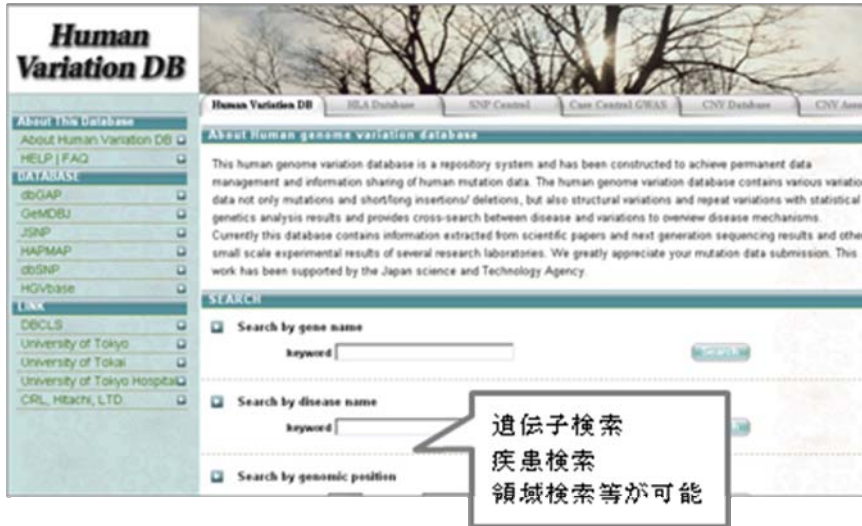


図 1. Human variation DB のトップページ

検索は、遺伝子名、疾患名、ゲノム領域などから可能。

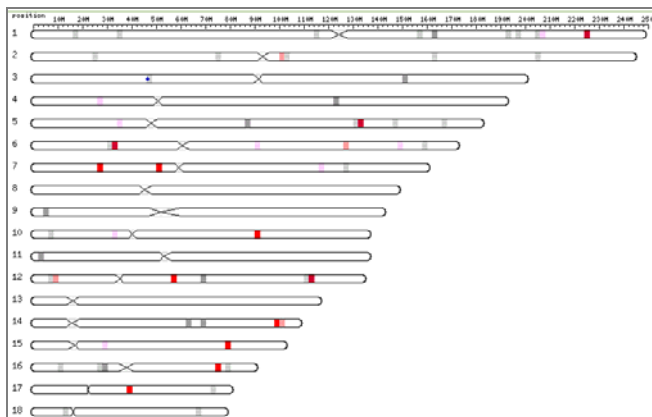


図 2 ある疾患の既知感受性遺伝子の全ゲノム上での位置
赤が濃いほど P-value が低い

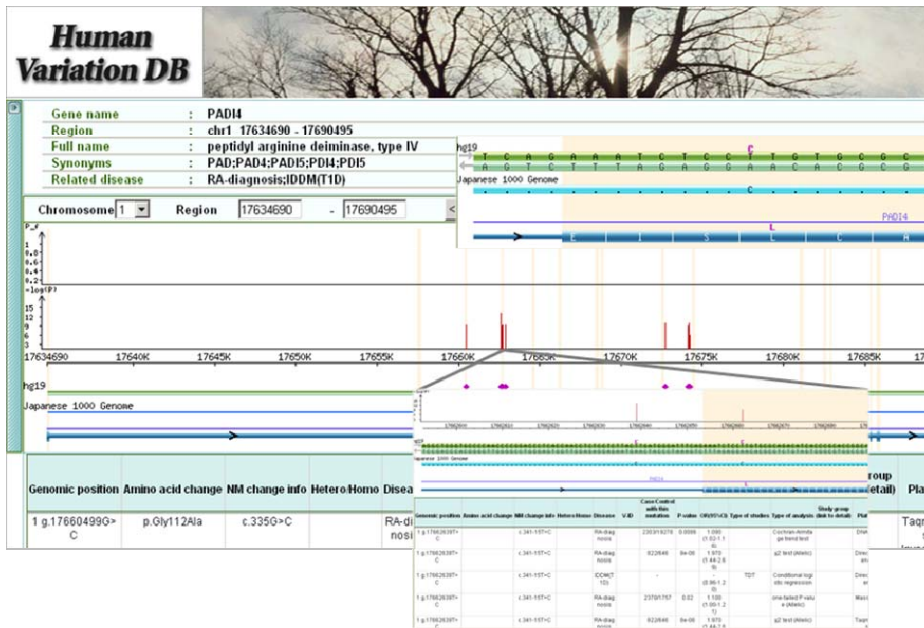


図3 領域ごとの変異情報（遺伝子名で検索した結果）

赤い棒が高いほど、P-value が低い。Reference genome も複数表示可能

Human variation 内部用 DB には、健常者データとしては新学術領域研究「ゲノム支援」で産出された exome 68 検体+ 1000genome data exome 98 検体の変異情報が、疾患データとしては、新学術「パーソナルゲノム情報に基づく脳疾患メカニズムの解明」で産出された 4 遺伝子変異（新規）+2 遺伝子変異（既知）が登録されている。一方、公開用 DB では、健常者の reference genome は 1000 genome data を基にして作成した変異情報を公開しており、また、文献から抽出した Common disease、神経変性変異のデータを中心に 2500 変異と遺伝統計情報を登録・公開している。

データの quality control については新学術領域研究と連携し、次世代シーケンサーによるデータとキャピラリーシーケンサーによるデータを比較しつつ、さらに公開されている trio data を用いた変異との一致率を利用した計算手法の最適化を行っている。

4-2) HLA DB の構築

HLA 領域については、アレルの組み合わせであるハプロタイプ情報が重要であるため、他領域とは別に HLA DB として構築している。現時点では、HLA-A, HLA-B, 等のローカスごとにハプロタイプのマルチプルアライメントを計算して表示すると共に、ハプロタイプ間の保存性も併せて表示している。また、ハプロタイプ間のペアワイズアライメントを行い、ハプロタイプ同士の類似性を表示できるようにしている。複数のローカスにおけるハプロタイプ情報が重要であるため、将来的にはそれらの情報も登録できるようにしていく。現在、公開データの他、新学術のデータとして1検体の HLA 領域内変異データを内部用 DB に登録している。

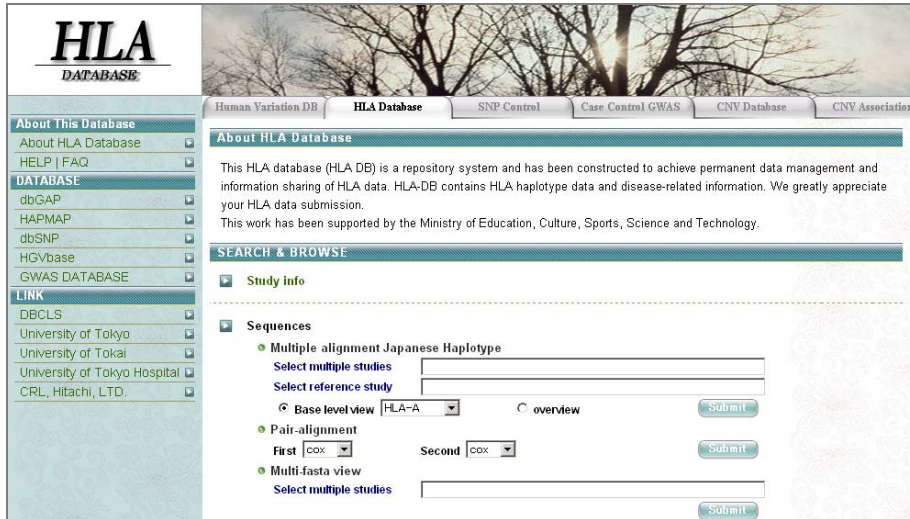


図 4 HLA DB トップページ

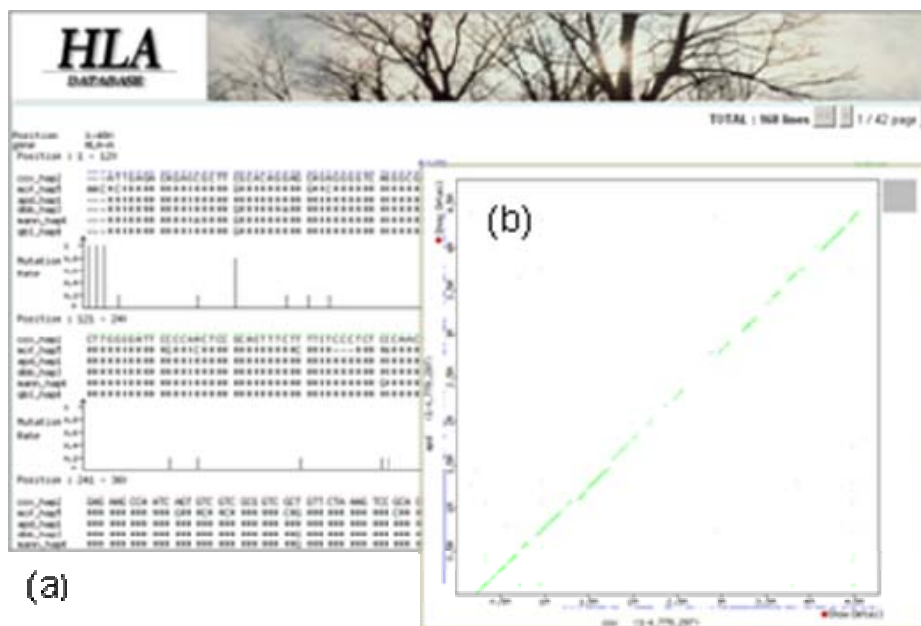


図 5 (a) HLA ハプロタイプの塩基レベルの表示 (b) 異なるハプロタイプ間の相同性比較

4-3) GWAS-DB, SNP-DB, CNV control DB, CNV association DB の機能追加と運用

GWAS-DB, SNP-DB, CNV control DB, CNV association DB には、各種モデルの遺伝統計解析結果とアレル、遺伝型頻度情報、CNV 領域などを登録している。現在内部用 GWAS-DB には 19 疾患 (26 studies) , 17 形質を登録しており、公開用 DB には 11 疾患を登録している。SNP-DB は 700 検体の健常者 SNP チップのデータを、CNV control DB は健常者 160 検体の CNV 情報を、CNV association DB は内部用 DB に 5 疾患、公開用 DB に 1 疾患を登録している。

これらの DB については、前プロジェクトから引き続きデータの預け入れと再配布を行った。GWAS-DB に関しては 1000genome data を利用した imputation などの機能追加も行った。

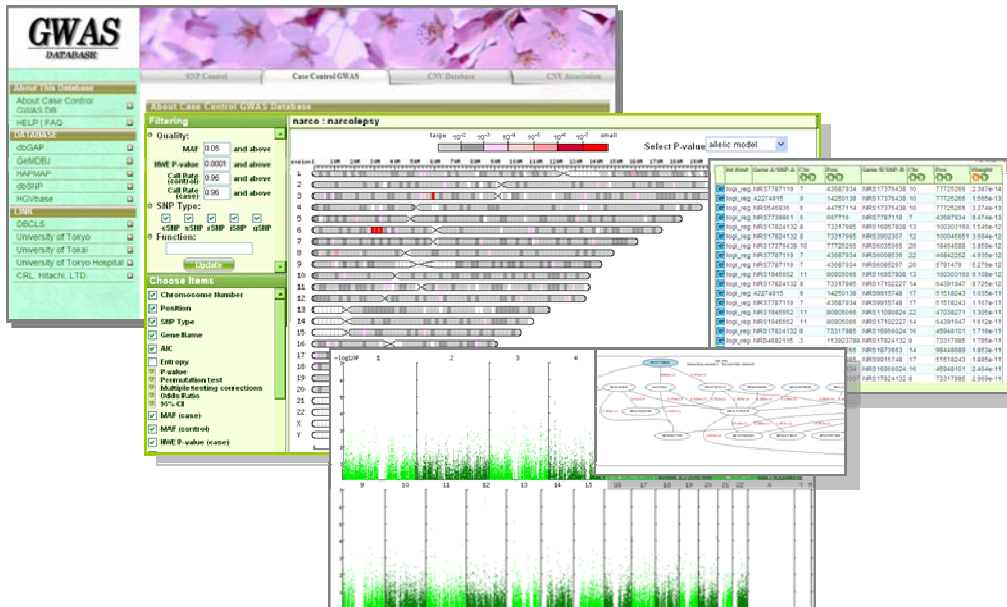


図 6 GWAS DB の画面例

4-4) 国際連携等について

EBI, Gene2phen などと、連携に関する話し合いを学会の際やメールを通して続けている。

5) 今後の見通し等

新学術領域に関わらず、幅広くデータの受け入れを行えるように学会等で広報活動を行う共に、次世代シーケンサーのデータのアーカイブについては DDBJ 等と連携できるように働きかける。

§3. 成果発表等

(3-1) データベースおよびウェブツールの構築と公開

① 公開中のデータベース・ウェブツール等

データベース名: GWAS-DB

概要: 数十万から数百万の SNP チップによる疾患関連解析の頻度情報と遺伝統計解析の結果を登録し、GUI により結果を閲覧できるようにしている。

公開日: H20 年 8 月 26 日

URL: http://gwas.lifesciencedb.jp/cgi-bin/gwasdb/gwas_top.cgi

アクセス数:

公開日から H24 年 3 月 31 日現在 uniq IP address 数 16719 (総 IP 数 25947) 回

データベース名: SNP-DB

概要: 数十万から数百万の SNP チップを用いて解析した健常者のアレル・遺伝子型頻度の情報とハーディーワインバーグ平衡検定値を登録し、GUI により結果を閲覧できるようにしている。

公開日: H20 年 8 月 26 日

URL: http://gwas.lifesciencedb.jp/snpdb/snp_top.php

アクセス数: GWAS-DB と同一ログ

データベース名: CNV control DB

概要: 健常者の数十万から数百万の SNP チップデータから検出した CNV を登録し、GUI により結果を閲覧できるようにしている。

公開日: H22 年 1 月 18 日

URL: http://gwas.lifesciencedb.jp/cgi-bin/cnvdb/cnv_top.cgi

アクセス数: GWAS-DB と同一ログ

データベース名: CNV association DB

概要: 健常者の数十万から数百万の SNP チップデータから検出した CNV のケースコントロール解析の結果を登録し、GUI により結果を閲覧できるようにしている。

公開日: H23 年 12 月 01 日

URL: http://gwas.lifesciencedb.jp/cgi-bin/cccdb/ccc_top.cgi

アクセス数: GWAS-DB と同一ログ

データベース名: Human Variation DB

概要: 次世代シーケンサーによって検出された健常者変異、疾患関連変異、及び、文献に発表されている疾患関連変異を抽出・登録し、GUI により閲覧できるようにしている。また、日本人健常者の参照配列情報も閲覧できるようにしている。

公開日: H24 年 2 月 22 日

URL: http://gwas.lifesciencedb.jp/cgi-bin/hvdb/hv_top.cgi

アクセス数: GWAS-DB と同一ログ

データベース名: HLA DB

概要: 次世代シーケンサーによって検出された HLA 領域における健常者変異、疾患関連変異、及び、文献に発表されている疾患関連変異を抽出・登録し、GUI により閲覧できるようにしている。

公開日: H24 年 2 月 22 日

URL: http://gwas.lifesciencedb.jp/cgi-bin/hladb/hla_top.cgi

アクセス数: GWAS-DB と同一ログ

(3-2) 原著論文発表

① 発行済論文数(国内(和文) 1 件、国際(欧文) 14 件):

② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0 件、国際(欧文) 5 件)

③ 論文詳細情報

1. Tanaka Y, Kurosaki M, Nishida N, Sugiyama M, Matsuura K, Sakamoto N, Enomoto N, Yatsushashi H, Nishiguchi S, Hino K, Hige S, Itoh Y, Tanaka E, Mochida S, Honda M, Hiasa Y, Koike A, Sugauchi F, Kaneko S, Izumi N, Tokunaga K, Mizokami M. “Genome-wide association study identified ITPA/DDRKG1 variants reflecting thrombocytopenia in pegylated interferon and ribavirin therapy for chronic hepatitis C.” *Hum. Mol. Genet.* 20(17): 3507-3516, 2011. (DOI. 10.1093/hmg/ddr249)
2. Kurotaki N, Tasaki S, Mishima H, Ono S, Imamura A, Kikuchi T, Nishida N, Tokunaga K, Yoshiura K, and Ozawa H. “Identification of novel schizophrenia loci by homozygosity mapping using DNA microarray analysis” *PLoS ONE* 6(5): e20589, Epub 2011 May.
3. Kawamura Y, Otowa T, Koike A, Sugaya N, Yoshida E, Yasuda S, Inoue K, Takei K, Konishi Y, Tanii H, Shimada T, Tochigi M, Kakiuchi C, Umekage T, Liu X, Nishida N, Tokunaga K, Kuwano R, Okazaki Y, Kaiya H, Sasaki T. “A genome-wide CNV association study on panic disorder in a Japanese population” *J. Hum. Genet.* 2011 Dec; 56(12): 852-856. (doi: 10.1038/jhg.2011.117)
4. Yoshida M, Koike A. “SNPInterForest: a new method for detecting epistatic interactions.” *BMC Bioinformatics.* 12: e469, Epub 2011 Dec. (DOI. 10.1186/1471-2105-12-469)
5. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl E, Kurreman F, Nishida N, Ohmiya H, Myouzen K, Takahashi M, Sawada T, Nishioka Y, Yukioka M, Matsubara T, Wakitani S, Teshima R, Tohma S, Takasugi K, Shimada K, Murasawa A, Honjo S, Matsuo K, Tanaka H, Tajima K, Suzuki T, Iwamoto T, Kawamura Y, Tanii H, Okazaki Y, Sasaki T, Gregersen PK, Padyukov L, Worthigton J, Siminovitch KA, Lathrop M, Taniguchi A, Takahashi A, Tokunaga K, Kubo M, Nakamura Y, Kamatani N, Mimori T, Plenge RM, Yamanaka H, Momohara S, Yamada R, Matsuda F, and Yamamoto K: Meta-analysis of

genome-wide association studies identifies multiple novel loci associated with rheumatoid arthritis in the Japanese population. Nat. Genet. 2012 Mar 25. [Epub ahead of print] (doi: 10.1038/ng.2231)

6. Matsukawa T, Asheuer M, Takahashi Y, Goto J, Suzuki Y, Shimozawa N, Takano H, Onodera O, Nishizawa M, Aubourg P and Tsuji S. Identification of novel SNPs of ABCD1, ABCD2, ABCD3 and ABCD4 genes in patients with X-linked adrenoleukodystrophy (ALD) based on comprehensive resequencing and association studies with ALD phenotypes. *Neurogenet.* 12: 41-50, 2011 (doi:10.1007/s10048-010-0253-6)
7. Ishiura H, Fukuda Y, Mitsui J, Nakahara Y, Ahsan B, Takahashi Y, Ichikawa Y, Goto J, Sakai T, Tsuji S. Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1. *Neurogenet.* 12:117-21, 2011 (doi:10.1007/s10048-010-0271-4)
8. Matsukawa T, Wang X, Liu R, Onuki Y, Kubota A, Hida A, Kowa H, Fukuda Y, Ishiura H, Mitsui J, Takahashi Y, Aoki S, Takizawa T, Shimizu J, Goto J, Proud CG and Tsuji S. Adult-onset leukoencephalopathies with vanishing white matter with novel missense mutations in EIF2B2, EIF2B3, and EIF2B5. *Neurogenet.* 12:259-61, 2011 (DOI 10.1007/s10048-011-0284-7)
9. Majounie E, Renton AE, Mok K, et al. C9ORF72 hexanucleotide repeat expansion in sporadic ALS and FTD around the world. *Lancet Neurol.* (in press)
10. Ishiura H, Takahashi Y, Mitsui J, Yoshida S, Kihira T, Kokubo Y, Kuzuhara S, Ranum LPW, Tamaoki T, Ichikawa Y, Date H, Goto J, and Tsuji S. C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. *Arch. Neurol.* (in press)
11. Hashimoto Maeda M, Mitsui J, Soong B-W, Takahashi Y, Ishiura H, Hayashi S, Shirota Y, Ichikawa Y, Matsumoto H, Arai M, Okamoto T, Miyama S, Shimizu J, Inazawa J, Goto J and Shoji Tsuji. Increased gene dosage of myelin protein zero causes Charcot-Marie-Tooth disease. *Ann. Neurol.* (in press)
12. Seki N, Takahashi Y, Tomiyama H, Rogaeva E, Murayama S, Mizuno Y, Hattori N, Marras C, Lang AE, St George-Hyslop P, Goto J and Tsuji S. Comprehensive mutational analysis of LRRK2 reveals variants supporting association with autosomal dominant Parkinson's disease. *J. Hum. Genet.* (in press)
13. Mitsunaga S, Suzuki Y, Kuwana M, Sato S, Kaneko Y, Homma Y, Narita A, Kashiwase K, Okudaira Y, Inoue I, Kulski JK, Inoko H. Association between six classical HLA loci and rheumatoid arthritis: a comprehensive analysis. *Tissue Antigen* (in press)

14. Yasuno K, Bakircioglu M, Low SK, Bilguvar K, Gaal E, Ruigrok YM, Niemela M, Hata A, Bijlenga P, Kasuya H, Jaaskelainen JE, Krex D, Auburger G, Simon M, Krschek B, Ozturk AK, Mane S, Rinkel GJ, Steinmetz H, Hernessniemi J, Schaller K, Zenbutu H, Inoue I, Palotie A, Cambien F, Nakamura Y, Lifton RP, Gunel M. Common variant near the endothelin receptor type A (EDNRA) gene is associated with intracranial aneurysm risk. *Proc. Natl. Acad. Sci. USA* 108, 19707-19712, 2011. (doi: 10.1073/pnas.1117137108)
15. Roder C, Kasuya H, Tatagiba M, Inoue I, Krschek B. Meta-analysis of microarray gene expression studies on intracranial aneurysms. *Neuroscience* 201, 105-113, 2012 (doi:10.1016/j.neuroscience.2011.10.033).
16. Nakaoka H, Cui T, Oka A, Mitsunaga S, Kashiwase K, Homma Y, Sato S, Suzuki Y, Inoko H, Inoue I. A systems genetic approach provides a bridge from discovered genetic variants to biological pathways in rheumatoid arthritis. *PLoS One* 6, e25389, 2011 (doi:10.1371/journal.pone.0025389).
17. Mitsunaga S, Okudaira Y, Kunii N, Cui T, Hosomichi K, Oka A, Suzuki Y, Homma Y, Inoue I, Inoko H. Exact break point of a 50 kb deletion 8 kb centromeric of the HLA-A locus with HLA-A*24:02: the same deletion observed in other A*24 alleles and A*23:01 allele. *Immunogenet.* 63, 467-474, 2011 (doi: 10.1007/s00251-011-0521-0).
18. Mitsunaga S, Homma Y, Narita A, Kashiwase K, Okudaira Y, Shiina Y, Inoue I, Inoko H. Particular human leukocyte antigen alleles are associated with biochemical traits in the Japanese population. *Hum. Immunol.* 72, 566-570, 2011 (doi:10.1016/j.humimm.2011.03.011).
19. Yamaguchi T, Hosomichi K, Narita A, Shirota T, Tomoyasu Y, Maki K, Inoue I. Exome resequencing combined with linkage analysis identifies novel PTH1R variants in a primary failure of tooth eruption in Japanese. *J. Bone Miner. Res.* 26, 1655-1661, 2011. Digital Object Identifier (doi:10.1002/jbmr.385).

和文

20. 中岡博史、井ノ上逸朗:ゲノム情報を含む医学情報処理の将来。(2011) *医学の歩み*. 236, 667-674.