

平成 26 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」研究代表者

[徳永 勝士]

[東京大学 大学院医学研究科・教授]

[個別化医療に向けたヒトゲノムバリエーションデータベース]

§1. 研究実施体制

(1) 徳永グループ

- ① 研究代表者: 徳永 勝士 (東京大学医学系研究科、教授)
- ② 研究項目: ヒト多型・変異 DB の機能拡張とデータ収集
 - 1) ゲノムワイド関連解析の DB の機能拡張とデータの収集
 - 2) Human genome variation DB, HLA DB の拡張構築

(2) 辻グループ

- 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: ヒト多型・変異データの体系化とデータ登録
 - 1) 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」で産出した全ゲノム配列解析による神経変性疾患の新規変異の整理と登録
 - 2) 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) 井ノ上グループ (研究機関別)

- 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: 遺伝統計学的計算手法の開発とデータ登録
 - 1) 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コントロールを行うとともに登録を行う。
 - 2) 新学術以外で産出された変異データの収集と整理・登録
 - 3) HLA ハプロタイプの配列決定およびDB登録
 - 4) GWAS メタ解析手法の開発と適用
 - 5) エクソームデータからの新規Alu挿入の探索手法の開発および登録

(4) 小池グループ

- ① 主たる共同研究者: 小池 麻子 (日立製作所 研究開発グループ、ストラテジースタッフ)
- ② 研究項目: 多型・変異関連の計算手法の開発
 - 1) 前プロジェクトから引き継ぎ、GWAS データの解析と運用受け入れ・再配布と解析・登録を行う。
 - 2) Human genome variation DB を他の DB と連携させ、pathway 情報、メチル化情報、種間での保存性を鑑みた形で変異を解釈できるような知識型データベースとしての機能追加を行う。
 - 3) 文献から、変異 (構造多型を含む) と疾患感受性、薬剤応答性、ウィルス耐性などの表現型情報との関係性を抽出し、Human genome variation DB へ登録する。その際、HGVS が推奨しているフォーマットに準拠し、データの互換性を高める。
 - 4) 必要に応じて、Human genome variation DB の参照ゲノムデータを登録データ及び公開データから計算して登録する。

§ 2. 研究実施内容

目的:

比較的低価格での高速大容量の SNP タイピング技術の確立により、SNP による疾患関連解析(GWAS)が可能となり、ここ 10 年程は世界レベルでの大規模なプロジェクトも計画・実施され、多くの疾患関連候補 SNP が発見された。その一部は機能的な疾患関連性も明らかになりつつある。一方、ここ数年は、次世代シーケンサ(NGS)によるゲノムワイドな疾患関連変異の検出も可能となりつつあり、家族性の単一遺伝子疾患の難治性・希少疾患のみならず、徐々に多因子疾患にも適用され、多くの疾患関連候補新規変異が発見されつつある。また、癌に関しては ICGC(International cancer genome consortium)や TCGA(The cancer of genome atlas) などの大型プロジェクトの実施により、多様な癌細胞がシーケンシングされ、新規の体細胞変異が多く発見されるとともに、癌細胞の性質・悪性度に関係する変異も明らかになってきた。

これらのデータは膨大であることから散逸を防ぎ、研究者間で共有仕組みを作ることは重要である。この問題を解決するために NCBI では dbGAP が、EBI では EGA を構築し、phenotype と genotype の関係性を探索するための GWAS データ、及び NGS データをアーカイブする仕組みを作成している。しかし、集団によっても疾患関連変異が異なることもあり、アジア人の変異・表現型の体系化が必要である。

概要:

我々の研究グループは 2006-2010 年の統合データベースプロジェクト、2011-2013 年度の統合化推進プログラムにおいて、GWAS DB、SNP control DB、CNV DB、CNV association DB を構築し、GWAS データの登録・共有化を実施した。更に、GWAS-DB を拡張した human genome variation DB を構築し、GWAS に限らず変異の登録を可能とし、また、文献からの抽出した生殖細胞の変異の体系化を行ってきた。本 PJ においては、変異データの充実化及び体系化、変異と薬剤応答性との関係の充実化と、既存の体細胞変異データベースとの連携を図る。

進捗状況:2014 年度は、主に human genome variation DB への文献から抽出する疾患・薬剤・変異データの登録を充実させると共に、体細胞変異のデータベースとの連携機能を入れ、生殖細胞の変異のうち、がんを特徴づける体細胞変異が明示できるようにした。また、データの一部 RDF 化に取り組んだ。

研究成果:

1) Human genome variation DB の充実化

マニュアルキュレーションを行い文献から疾患情報 4000 レコード以上を追加し、データを充実させた。また、日本 PGx データサイエンスコンソーシアム、東北メディカルメガバンク(東北メディカル・メガバンク機構の地域住民コホート調査に参加した 20 歳以上の 1070 人の成人の集団内頻度が 5%を超える既知 SNP)、京大コホートデータなども日本人健常者データとして取り込み、各デ

ータごとに参照ゲノムとして表示できるようにした。(図 1)これにより、疾患関連変異の患者及びその研究の中での健常者頻度情報と上記データの健常者の頻度との差がわかるようになっている。

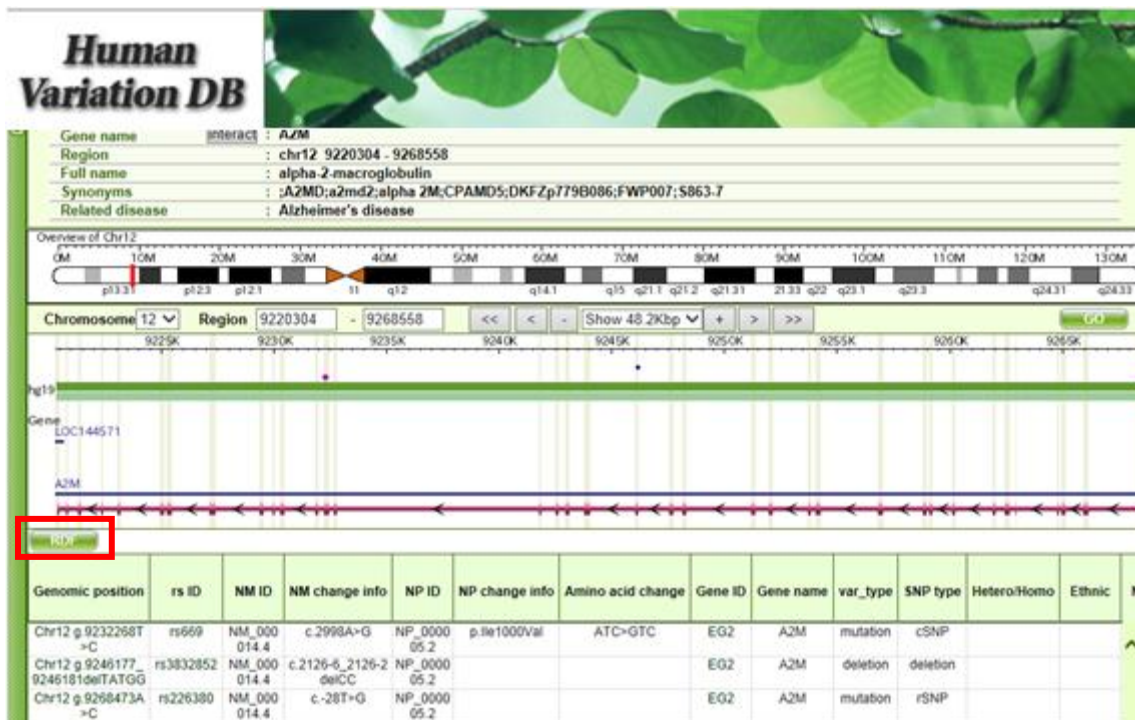
また、偏頭痛および四肢痛(Nagata E. PLoS One 2014)、脾類上皮嚢腫(Omer WH. BMC Med Genet 2014)、および高グレード漿液性卵巣がん(Hayano T. PLoS One 2014)で産出された変異データについて、データの収集と整理を行った。

がん変異データとして COSMIC のデータと連携し、がんのバー変異を生殖細胞変異と共に表示可能とし、生殖細胞変異のうちドライバー変異かどうかを明示できるようにした。(図1)

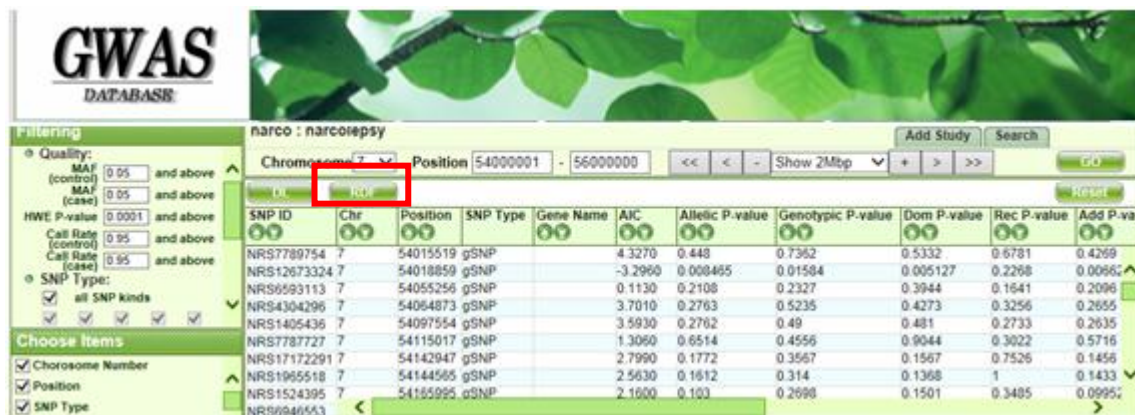
DBの一部データは、Ensemble の定義に従い RDF 化を実施した。SPARQL エンドポイントとしては Virtuoso を利用している。(図 2)



図 1. 特定遺伝子の表示例 (COSMIC データも併せて表示)



(a)



(b)

図 2. RDF の出力

その他、セキュリティー対策として、内部用データベースなどもユーザ情報を暗号化するなどの対策を実施している。

2) GWAS DB

内部用のデータ登録から外部用のデータ登録へのデータ移行の簡易化などを行うと共に、公開しているデータ(理研のオーダーメイド医療実現プロジェクトなど)のデータ登録を実施した。また、痛風患者について、GWAS メタ解析手法を用い関連遺伝子領域を同定した(Matsuo H. Ann Rheum Dis. 2015)。

3) HLA-DB

より簡便で効率の良いハイスループットな HLA ハプロタイプの配列決定を行うために、磁気ビーズ

法を改変することでライブラリーサイズとモル濃度を均一化し、サンプル間で均一なシーケンスカバレッジを得る手法 (BeNUS)を開発した (Hosomichi K. BMC Genomics 2014)。この手法を用いて新たに取得したデータを用い配列解析を実施した。

4) NGS データからの新規トランスポゾン挿入の探索手法の開発および登録

新規トランスポゾン挿入探索手法 (iHoanys)の開発を進めるとともに、既存の手法 (Mobster)を用いて卵巣がんのトランスポゾン探索を行い、エクソン領域への新規体細胞トランスポゾン挿入変異を検出した (論文投稿中)。

今後の見通し:

更なる変異データの充実化を図る。

§3. 成果発表等

(3-1) 原著論文発表

- ① 発行済論文数 (国内 (和文) 0 件、国際 (欧文) 9 件):
- ② 未発行論文数 (“accepted”、“in press”等) (国内 (和文) X 件、国際 (欧文) XX 件)
- ③ 論文詳細情報
 1. Yamasaki M, Miyagawa T, Toyoda H, Khor SS, Koike A, Nitta A, Akiyama K, Sasaki T, Honda M, and Tokunaga K. “Genome-wide analysis of CNV (copy number variation) and their associations with narcolepsy in a Japanese population.”, J. Hum. Genet., 59(5): 235-240, 2014 (DOI: 10.1038/jhg.2014.13.)
 2. Hosomichi K, Mitsunaga S, Nagasaki H, Inoue I. "A Bead-based Normalization for Uniform Sequencing depth (BeNUS) protocol for multi-samples sequencing exemplified by HLA-B." BMC Genomics. 15: 645, 2014 (DOI: 10.1186/1471-2164-15-645)
 3. Awata T, Yamashita H, Kurihara S, Morita-Ohkubo T, Miyashita Y, Katayama S, Mori K, Yoneya S, Kohda M, Okazaki Y, Maruyama T, Shimada A, Yasuda K, Nishida N, Tokunaga K, and Koike A. A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PLoS One 9(11): e111715, 2014. (DOI: 10.1371/journal.pone.0111715.)
 4. Iio E, Matsuura K, Nishida N, Maekawa S, Enomoto N, Nakagawa M, Sakamoto N, Yatsunami H, Kurosaki M, Izumi N, Hiasa Y, Masaki N, Ide T, Hino K, Tamori A, Honda M, Kaneko S, Mochida S, Nomura H, Nishiguchi S, Okuse C, Itoh Y, Yoshiji H, Sakaida I, Yamamoto K, Watanabe H, Hige S, Matsumoto A, Tanaka E, Tokunaga K, and Tanaka Y. Genome-wide association study identifies a PSMD3 variant associated with neutropenia in interferon-based therapy for chronic hepatitis C. Hum. Genet. 134(3): 279-289, 2015. (DOI: 10.1007/s00439-014-1520-7.)

5. Ueta M, Sawai H, Sotozono C, Hitomi Y, Kaniwa N, Kim MK, Seo KY, Yoon KC, Joo CK, Kannabiran C, Wakamatsu TH, Sangwan V, Rathi V, Basu S, Ozeki T, Mushiroda T, Sugiyama E, Maekawa K, Nakamura R, Aihara M, Matsunaga K, Sekine A, Gomes JÁP, Hamuro J, Saito Y, Kubo M, Kinoshita S, and Tokunaga K. *IKZF1*, a new susceptibility gene for cold medicine-related Stevens-Johnson syndrome/toxic epidermal necrolysis with severe mucosal involvements. . *J. Allergy Clin. Immunol.* [Jan 2015 Epub ahead of print] (DOI: 10.1016/j.jaci.2014.12.1916.)
6. Nagata E, Fujii N, Hosomichi K, Mitsunaga S, Suzuki Y, Mashimo Y, Tsukamoto H, Satoh T, Osawa M, Inoue I, Hata A, Takizawa S. "Possible association between dysfunction of vitamin D binding protein (GC Globulin) and migraine attacks." *PLoS One.* 9 (8): e105319, 2014 (DOI: 10.1371/journal.pone.0105319)
7. Omer WH, Narita A, Hosomichi K, Mitsunaga S, Hayashi Y, Yamashita A, Krasniqi A, Iwasaki Y, Kimura M, Inoue I. "Genome-wide linkage and exome analyses identify variants of HMCN1 for splenic epidermoid cyst." *BMC Med Genet.* 15: 115, 2014, (DOI: 10.1186/s12881-014-0115-4)
8. Hayano T, Yokota Y, Hosomichi K, Nakaoka H, Yoshihara K, Adachi S, Kashima K, Tsuda H, Moriya T, Tanaka K, Enomoto T, Inoue I. "Molecular characterization of an intact p53 pathway subtype in high-grade serous ovarian cancer." *PLoS One.* 9 (12): e114491, 2014, (DOI:10.1371/journal.pone.0114491)
9. Matsuo H, Yamamoto K, Nakaoka H, Nakayama A, Sakiyama M, Chiba T, Takahashi A, Nakamura T, Nakashima H, Takada Y, Danjoh I, Shimizu S, Abe J, Kawamura Y, Terashige S, Ogata H, Tatsukawa S, Yin G, Okada R, Morita E, Naito M, Tokumasu A, Onoue H, Iwaya K, Ito T, Takada T, Inoue K, Kato Y, Nakamura Y, Sakurai Y, Suzuki H, Kanai Y, Hosoya T, Hamajima N, Inoue I, Kubo M, Ichida K, Ooyama H, Shimizu T, Shinomiya N. "Genome-wide association study of clinically defined gout identifies multiple risk loci and its association with clinical subtypes." *Ann Rheum Dis.* pii: annrheumdis-2014-206191, 2015 (DOI: 10.1136/annrheumdis-2014-206191)
10. Mitsui J, Matsukawa T, Sasaki H, Yabe I, Matsushima M, Dürr A, Brice A, Takashima H, Kikuchi A, Aoki M, Ishiura H, Yasuda T, Date H, Ahsan B, Iwata A, Goto J, Ichikawa Y, Nakahara Y, Momose Y, Takahashi Y, Hara K, Kakita A, Yamada M, Takahashi H, Onodera O, Nishizawa M, Watanabe H, Ito M, Sobue G, Ishikawa K, Mizusawa H, Kanai K, Hattori T, Kuwabara S, Arai K, Koyano S, Kuroiwa Y, Hasegawa K, Yuasa T, Yasui K, Nakashima K, Ito H, Izumi Y, Kaji R, Kato T, Kusunoki S, Osaki Y, Horiuchi M, Kondo T, Murayama S, Hattori N, Yamamoto M, Murata M, Satake W, Toda T, Filla A, Klockgether T, Wüllner U, Nicholson G, Gilman S, Tanner CM, Kukull WA, Stern MB, Lee VM-Y, Trojanowski JQ, Masliah E, Low PA, Sandroni P, Ozelius LJ, Foroud T, and Tsuji S. Variants associated with Gaucher disease in multiple system atrophy. *Ann Clin Transl*

Neurol. 2: 417–426, 2015 DOI: 10.1002/acn3.185

11. Ishiura H, Takahashi Y, Hayashi T, Saito K, Furuya H, Watanabe M, Murata M, Suzuki M, Sugiura A, Sawai S, Shibuya K, Ueda N, Ichikawa Y, Kanazawa I, Goto J and Tsuji S. Molecular epidemiology and clinical spectrum of hereditary spastic paraplegia in the Japanese population based on comprehensive mutational analyses. *J Hum Genet* 59:163-72, 2014 doi: 10.1038/jhg.2013.139

(3-2) データベースおよびウェブツール等の構築と公開

- ① 公開中のデータベース・ウェブツール等
別紙1参照。

(3-3) その他の著作物(総説、書籍など)

●詳細情報

1. 小池麻子、徳永勝士：ヒトゲノムの多様性の体系化とバリエーションデータベース、遺伝子医療の現状とゲノム医療の近未来、医学のあゆみ 250(5)：326-330, 2014.
2. 宮川卓、徳永勝士：遺伝疫学-多因子疾患感受性遺伝子の探索法、臨床遺伝学の最新知識、産婦人科の実際、64(3)：309-313, 2015.
3. Hirofumi Nakaoka, Kazuyoshi Hosomichi, Shigeki Mitsunaga, Hidetoshi Inoko, Ituro Inoue. "The Admixed Origin of Japanese Population from HLA Alleles." *Major Histocompatibility Complex* 21 (1): 37, 2014, (DOI:10.12667/mhc.21.37)
4. 細道一善、井ノ上逸朗 次世代シーケンサー解析スタンダード 実験医学別冊、2014
5. 細道一善、井ノ上逸朗 次世代シーケンサーによるゲノム解析技術の進歩 医学のあゆみ、250 巻、2014

(3-4) 国際学会および国内学会発表

- ① 招待講演 (国内 4 件、国際 2 件)

〈国内〉

- 1) 徳永勝士：ゲノム解析から医療へ貢献する：薬剤応答遺伝子を中心に、医療薬学フォーラム 2014・第 22 回クリニカルファーマシーシンポジウム、東京、2014. 6. 28.
- 2) 徳永勝士：Complex disease の遺伝研究アップデート、第 18 回小児分子内分泌研究会、札幌北広島、2014. 7. 6.
- 3) 徳永勝士：Complex disease の遺伝子探索：現状と課題、生命医薬情報学連合大会 2014 セッション：個別化医療に向けたゲノム解析、仙台、2014. 10. 3.
- 4) 徳永勝士：Complex disease の遺伝研究：現状と課題、第 15 回眼科 DNA チップ研究会、神戸、2014. 11. 13.

〈国際〉

1. 発表者、タイトル、学会名、場所、月日

- 1) Katsushi Tokunaga: Lessons from genome-wide search for disease susceptibility genes and drug response genes, National Cancer Center, Korea, 2014.7.18.
- 2) Katsushi Tokunaga: HLA and non-HLA susceptibility genes to human immune-mediated diseases, Siriraj International Conference in Medicine and Public Health 2014, Bangkok, 2014.7.24.
- 3) Ishiura H, Koh K, Shimazaki H, Mitsui J, Takahashi Y, Goto J, Yoshimura K, Doi K, Morishita S, Sasaki H, Takiyama Y, Tsuji S, JASPAC (Japan Spastic Paraplegia Research Consortium). Exome analysis of 116 patients supposed to be autosomal recessive hereditary spastic paraplegia established molecular diagnoses of 49 patients with broad genetic heterogeneities. The 64th Annual Meeting of the American Society of Human Genetics

② 口頭講演 (国内 4 件、国際 0 件)

〈国内〉

- 1) 細道一善、井ノ上逸朗 et al. HLA-omics : HLA 領域におけるゲノム多様性、メチル化および遺伝子発現の統合的解析 第 23 回日本組織適合性学会大会、長崎、2014 年、9 月 14 日
- 2) 徳永勝士：個別化医療に向けたヒトゲノムバリエーションデータベース、トーゴの日シンポジウム 2014、東京、2014. 10. 5.
- 3) 早野崇英 p53 パスウェイ正常型高グレード漿液性卵巣癌の分子機序の解析 日本人類遺伝学会第 59 回大会、東京、2014 年 11 月 21 日
- 4) 細道一善、井ノ上逸朗 et al. HLA-omics : HLA 領域におけるゲノム多様性、メチル化および遺伝子発現の統合的解析 日本人類遺伝学会第 59 回大会、東京、2014 年 11 月 22 日

③ ポスター発表 (国内 3 件、国際 2 件)

〈国内〉

1. 早野崇英 NGS データを用いたトランスポゾン *Alu* 検出プログラムの開発 日本分子生物学会第 37 回大会、神奈川、2014 年 11 月 25 日
2. 小池麻子、澤井裕美、川嶋実苗、井ノ上逸朗、辻省次、徳永勝士、統合化推進プログラムにおけるヒトゲノムバリエーションデータベース 統合データベースシンポジウム、2014 年 10 月 10 日
3. 小池麻子、澤井裕美、川嶋実苗、井ノ上逸朗、辻省次、徳永勝士、統合化推進プログラムにおけるヒトゲノムバリエーションデータベース 日本分子生物学会 特別企画 2014 年 11 月 25 日-27 日

〈国際〉

1. A. Koike, H. Sawai, I. Inoue, S. Tsuji, K. Tokunaga Update and expansion of Human Variation Database in Japanese Database Integration Program. ASHG 2014

Annual Meeting、USA、2014年10月19日

2. 細道一善、井ノ上逸朗 et al. PCR-based method for complete HLA gene sequencing and capture-based method for entire HLA region sequencing. ASHG 2014 Annual Meeting、USA、2014年10月19日

(3-5) 知的財産権の出願

特になし。

(3-6) 受賞・報道等

① 新聞報道

1. 日刊工業新聞・京都新聞など、スティーブンス・ジョンソン症候群発症関連遺伝子を特定、平成27年2月9日～11日

§4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2015/2/20	ゲノム医科学とバイオインフォマティクスの接点と集学研究	国立遺伝学研究所	30人程度	ゲノム医学研究者とバイオインフォマティクス研究者が集学的な議論をおこない、お互いが補完しあう形であらたな医学研究の進展に寄与することを目的とする

以上

別紙1 既公開のデータベース・ウェブツール等

No.	研究開発課題名	データベース、ウェブツール等の名称【公開】	概要【公開】 (150字程度)	URL【公開】	公開開始日	関連論文 (論文リストに記載があれば、その番号でも可)	備考
01	個別化医療に向けたヒトゲノムバリエーションデータベース	ヒトゲノムバリエーションデータベース Human genome variation database	各種実験で検出された関連変異を登録するレポジトリ型のデータベースです。文献から抽出した疾患関連変異及び特徴的な臨床情報、及び、日本人健常者のSNV頻度データも登録・閲覧できるようにしています。	https://gwas.biosciencedbc.jp/cgi-bin/hvdb/hv_top.cgi			
02	個別化医療に向けたヒトゲノムバリエーションデータベース	copy number variation association database (CNV association DB)	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ます。	http://gwas.biosciencedbc.jp/cgi-bin/cccdB/ccv_top.cgi			
03	個別化医療に向けたヒトゲノムバリエーションデータベース	CNV control DB	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ます。	http://gwas.biosciencedbc.jp/cgi-bin/cnvdb/cnv_top.cgi			
04	個別化医療に向けたヒトゲノムバリエーションデータベース	Genome wide association database (GWAS DB)	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ます。	http://gwas.biosciencedbc.jp/cgi-bin/gwasdb/gwas_top.cgi		9	
05	個別化医療に向けたヒトゲノムバリエーションデータベース	HLA database (HLA DB)	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ます。	http://gwas.biosciencedbc.jp/cgi-bin/hlab/hla_top.cgi		2	
06	個別化医療に向けたヒトゲノムバリエーションデータベース	SNP Control DB (SNP DB)	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リル・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ます。	http://gwas.biosciencedbc.jp/snpdb/snp_top.php			