

平成 24 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 (研究機関別)

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 の拡張
- ・HLA 領域における多型・変異の DB の拡張

(2) 「ヒト多型・変異データの体系化」グループ (研究機関別)

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 (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

- ・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。
- ・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 (研究機関別)

①主たる共同研究者:小池 麻子 (日立製作所 中央研究所 部員、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

目的:

近年の高速大容量の SNP タイピング技術の確立によるゲノムワイド関連解析による疾患関連解析が世界レベルで大規模に進行し、それに伴う大量な遺伝子型データが産出されつつある。我々の研究グループは 2006-2010 年度の統合 DB プロジェクトにおいて、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。2012 年度には、これらの DB を引き続き運用して新たな成果を受け入れるとともに、メタ解析や imputation 機能などを追加し高機能化させる。

一方、次世代シーケンサーの性能向上に伴いリシーケンスによる新規疾患関連変異の発見が次々と報告されている。一検体あたりゲノムの 0.1%にあたる約 300 万の SNV(single nucleotide variant)が検出されるが、現在では、その 10%程度はまだ dbSNP に未登録な状態であることから、変異情報は今後も増加すると予想される。これらの問題を解決するためにも、本プロジェクトでは、次世代シーケンサーによる新規発見の変異を登録し、データの散逸を防ぐとともに、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。2011 年度に、Human Variation DB を構築し、データの受け入れ開始した。また、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を利用し、次世代シーケンサーによる新規発見の変異解析に利用できるように日本人健常者の標準的なゲノム配列を整備・提供するためのプロトコルを検討した。2012 年度には、Human Variation DB の完成度を上げるとともに、GWAS-DB との横断検索機能などを追加していくことを実施する。また、変異情報を網羅的に登録できる仕組みを作成するために、Evidence が不十分な変異に関しては、簡易型 DB を新たに構築し、そちらへの登録を呼びかける。

概要:

2011 年度に引き続き 2012 年度においても、統合 DB で開発した GWAS 関連の DB の機能追加とデータ収集を行うと共に、次世代シーケンサーから産出される変異データを登録するための Human Variation DB の拡張を行う。また、Human Variation DB とは別に、疾患原因・関連変

異かの evidence が十分取られていないデータで臨床の現場で埋もれているデータを登録できるよう簡易型の変異登録 DB (report DB) を新たに構築する。

進捗状況:

Human Variation DB 及び、HLA DB を拡張し、データを追加すると共に、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。簡易型の変異登録 DB (report DB) を構築した。

研究成果:

4-1) Human variation DB の構築

Human variation DB 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 の基本構造を 2011 年度に構築しており、2012 年度は、以下記すように、次世代シーケンサー用 variation database を充実させるとともに、GWAS-DB、簡易型変異データベース report DB への横断検索も実装した。

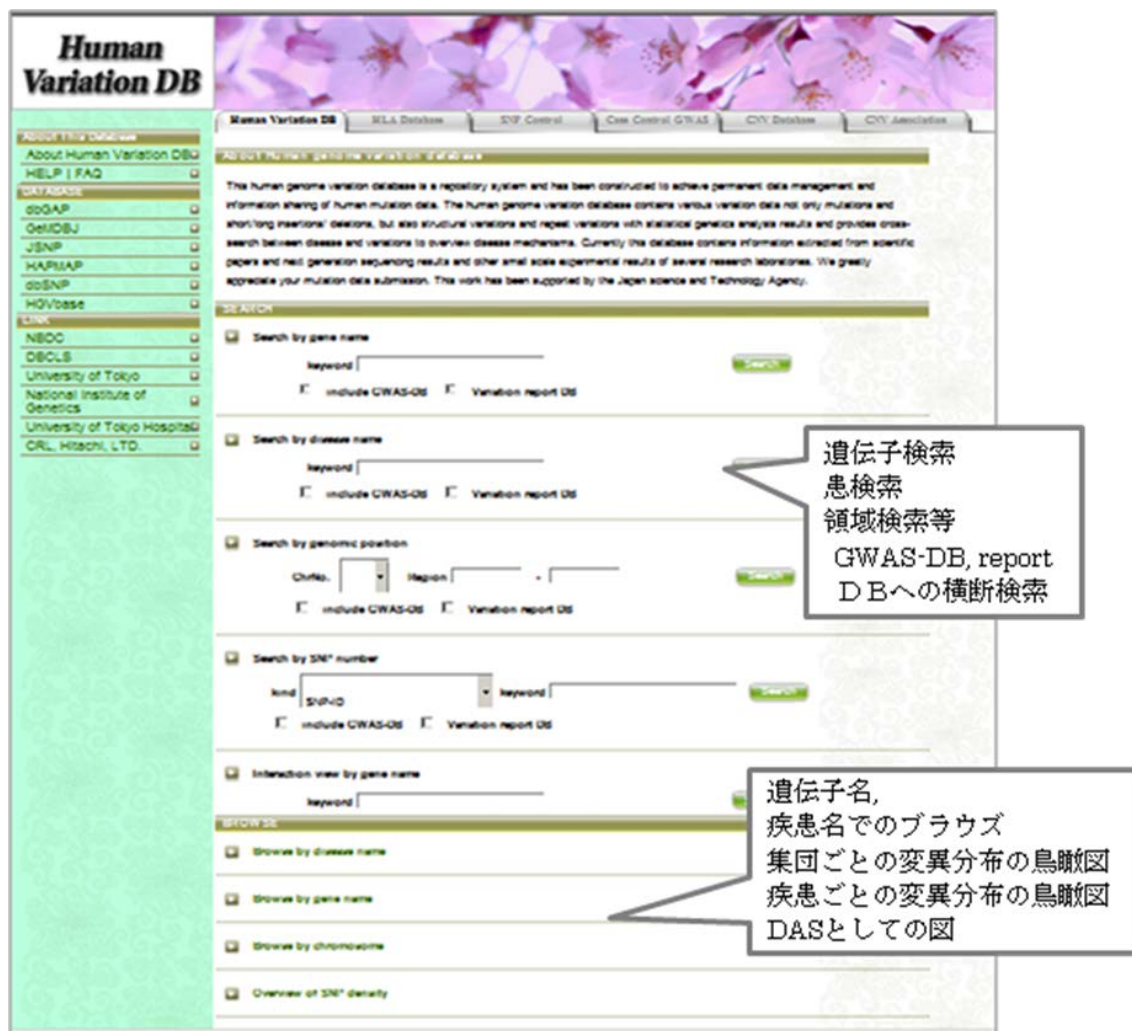


図 1. Human Variation DB から GWAS-DB, reportDB への横断検索機能追加

変異の機能的影響に関する解釈の補助として、Human variation DB に蛋白質のドメイン情報、snoRNA, miRNA, 生物種間での配列保存情報表示などの機能追加を行った。(図 2) また、参照配列ゲノムも複数登録・表示可能とし、同一変異に関する集団ごとの疾患リスクが存在することを鑑みて、集団ごとの疾患リスク変異のアリル頻度を表示可能とした。(図 3) 集団ごとにゲノム上の変異密度分布を計算し表示可能とし、集団間の違いの鳥瞰図を得られるようにした。(図 4)

疾患に関わる変異の全体像を把握するための一環として、疾患ごと、及び、疾患全体での疾患感受性変異/候補変異の分布もゲノム全体で表示可能とした。(図 5) 更に、変異の機能(phenotype)への影響を理解しやすくするための第一歩として、蛋白質相互作用(BIND, HPRD の情報)で描画した蛋白質ネットワーク上に変異情報(新規収集変異+OMIM)を投影できるようにした。(図 6)

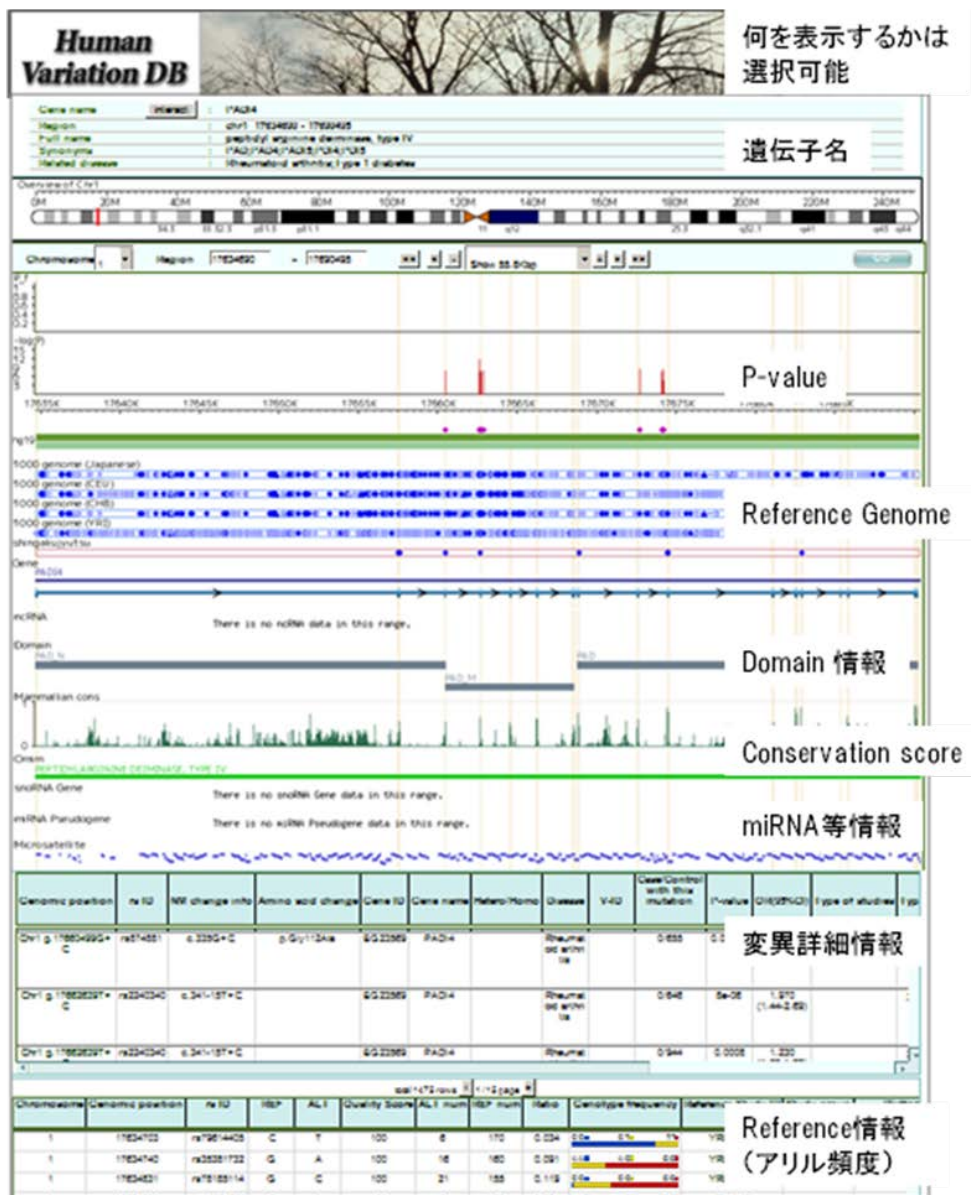


図 2 Human Variation DB の遺伝子ごとの表示例

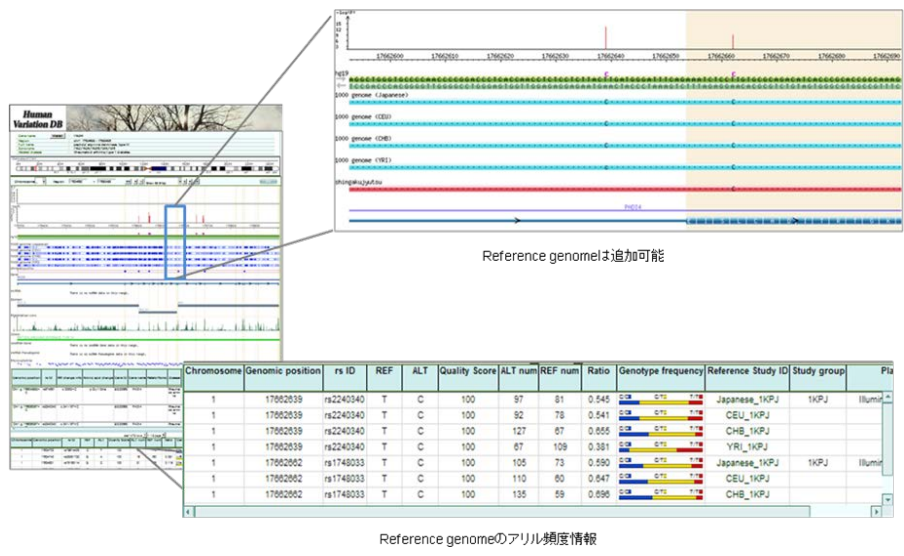


図 3 Human Variation DB の参照配列の表示例

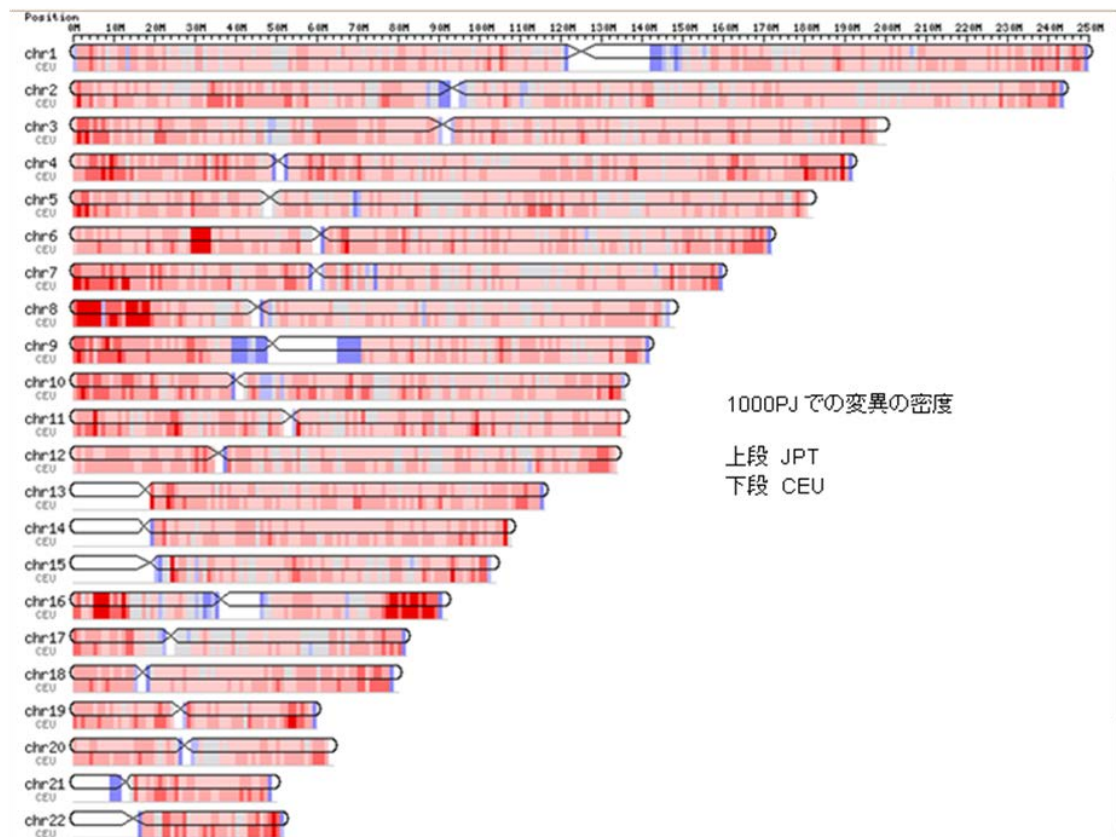


図 4 集団ごとの変異分布の比較例

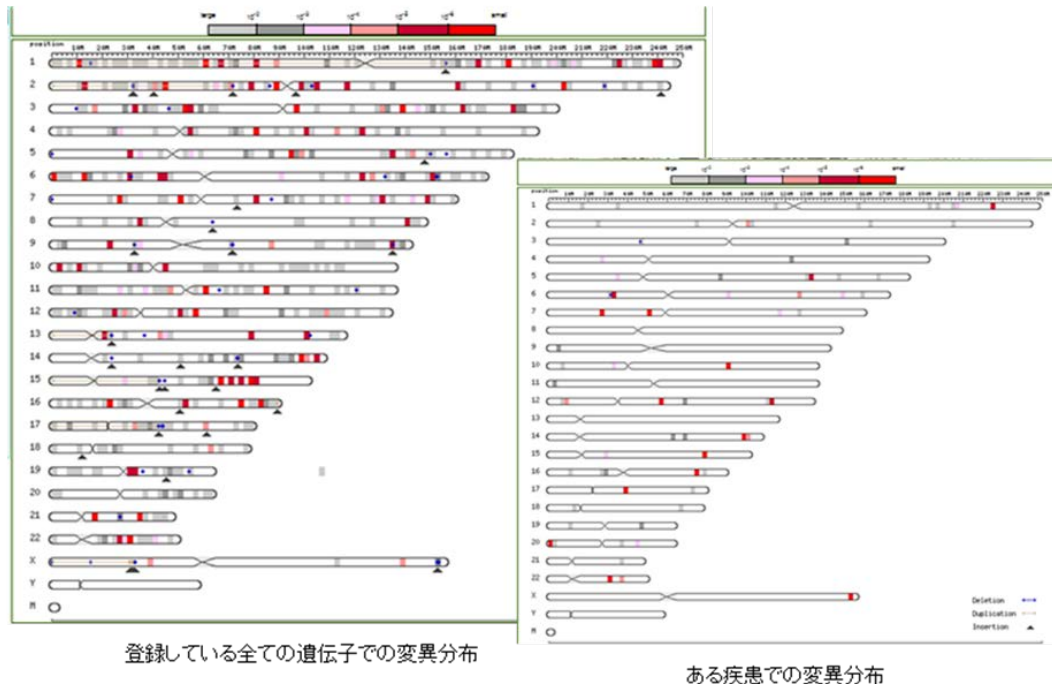


図 5 ゲノム上の疾患感受性変異/疾患感受性候補変異の分布の表示例

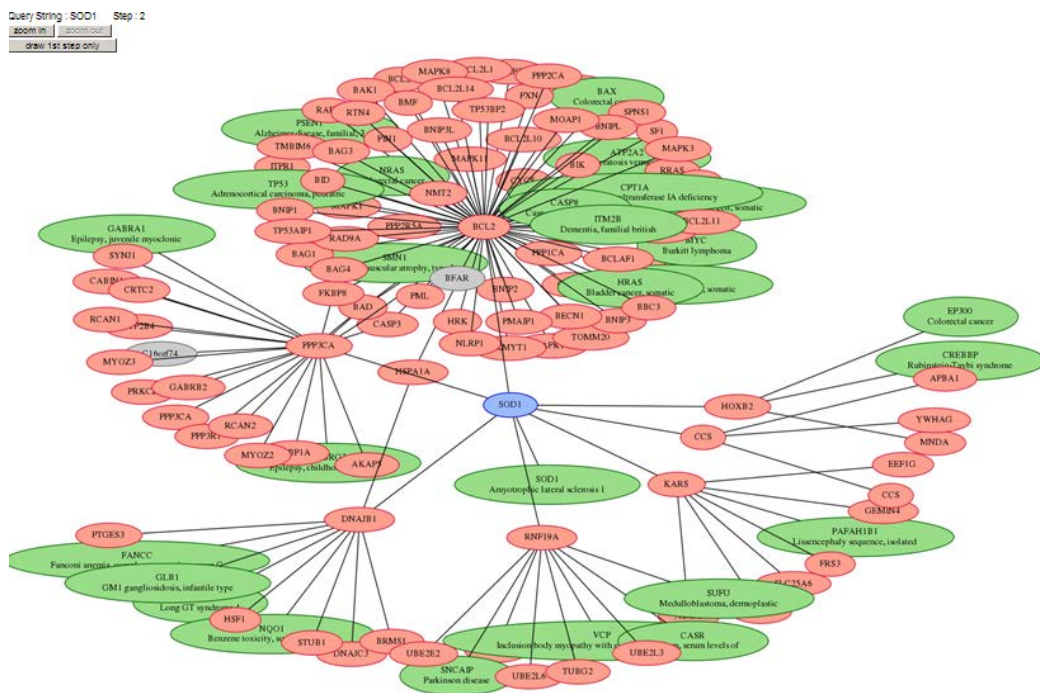


図 6 蛋白質相互作用のネットワーク上に疾患変異情報を投影した例

Human variation 内部用 DB には、新学術領域研究「ゲノム支援」で産出された健常者 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 2 遺伝子変異(新規)+5 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複数公開しており、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 1.5 万エントリーの変異と P-value、オッズ比などの遺伝統計情報、実験手法などの付随情報とともに登録・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、また、公開されている trio data を用いて計算手法の最適化を行っている。

現在までに、SNV の検出には GATK の Unified Genotyper と Samtools の mpileup を試しているが、両者からコールされる変異数は異なる。Samtools および GATK それぞれにおいて実際に変異が確認された割合はそれぞれ 84.9% および 79.1% と約 8 割について正確に変異を検出できていることが確認された。さらに、Samtools で検出された SNV のうち、偽陽性率は 15.1% であったのに対し、GATK では 5.8% と低い偽陽性率での SNV 検出には GATK が良好な成績を示した。一方 Samtools で検出され、キャピラリーシーケンサーにより確認された SNV のうち 15.1% は GATK で検出されなかった。すなわち、GATK は偽陽性率が低い一方で偽陰性率が Samtools に比べて高く、今後データに応じたパラメーターの条件検討の必要性が示唆された。GATK は偽陰性を減らす為、データ量に応じて SNV 検出のパラメーターを再検討する必要があり、Samtools は偽陽性を減らす為に Mapping Quality や Segmental Duplication の情報を使った再評価が必要である。

	Samtools pileup	GATK Unifiedgenotyper
偽陽性率	15.1%	5.8%
偽陰性率	0%*	15.1%
正解率	84.9%	79.1%

*samtools でコールされた変異リストを用いたため

図 Samtools および GATK で検出された SNV における偽陽性および偽陰性

4-2) HLA DB の構築

ヒトの主要組織適合遺伝子複合体 (Human leukocyte antigen; HLA) 領域はヒト第6染色体短腕部 6p21.31 に位置し、免疫応答をつかさどる HLA 遺伝子を含む 3.8 Mb から構成される。ヒトゲノムの中でも群を抜いて多型性に富むゲノム領域であり、HLA 遺伝子 6 座 (HLA-A, -B, -C, -DR, -DQ, -DP) に計 7,300 種類を超える膨大な HLA 多型 (アレル) がこれまでに同定されている。またリウマチや糖尿病など 100 種近くの生活習慣病、自己免疫疾患、がん、造血幹細胞移植に伴う GVH 病、ウイルス感染症における防御と重症化など数多くの医学的興味を有する。特に、ステイブンス・ジョンソン症候群、中毒性表皮壊死融解症および薬剤性過敏症症候群などの薬剤副作用


用と HLA アレルは極めて強い関連が複数報告されており、HLA アレルが薬剤副作用に大きく寄与していることが確認されている。さらには各遺伝子座における特定の HLA アレルが強い連鎖不平衡により、HLA-A24-B52-DR15 など特有の HLA ハプロタイプを形成している。以上の理由から HLA DB は HLA 領域の多型・変異情報をハプロタイプとして登録する必要があるため、上記の DB とは別に HLA 用の多型・変異の DB を 2011 年に構築している。2012 年度は、上述の疾患および薬剤副作用と関連するアレルを含む、33セルラインにおける *HLA-A*、*-C*、*-B*、*-DRB1*、*-DQB1* および *-DPB1* の 6 遺伝子の全長のシーケンス(計 198 アレル)を登録し、データの拡充を行った。

4-3) 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 20 疾患 (27 studies) , 17 形質を登録しており、公開用 DB には 12 疾患を登録している。SNP-DB は 1,100 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。GWAS-DB に関しては、affected sib pair の解析とができるように、パイプラインの構築と描画を行った。

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。

4-4) 簡易型変異データベース(report DB)の構築

疾患原因・関連変異かの evidence が十分取られていないデータで臨床の現場で埋もれているデータを登録できるよう簡易型の変異登録 DB を構築した。Human variation DB と異なり、登録者への負担を軽減するため、変異の位置情報と変異情報のみが必須の登録情報としている。Evidence が十分とれて論文が publish されたデータに関しては、Human Variation DB に登録を行う。

Variation report DB 

Search Registration

Accession number:	<input type="text"/>	* New registration <input type="radio"/> Search number <input type="text"/>	<input type="button" value="Search"/>
Variation kinds:	<input type="text"/>		
Gene symbol:	<input type="text"/>		
Mutations (codon):	<input type="text"/>	Ex. gAAG-AAT	
Mutation:	<input type="text"/>		
Amino Acid changes:	<input type="text"/>	Ex. Lys-Glu	
Splicing mutations:	<input type="text"/>	Ex. S00001 c.sdelG	
Sequence:	<input type="text"/>		
Genomic position:	<input type="text"/>	Ex. Chr21 3000A>G	
Phenotype:	<input type="text"/>	Ex. ALS	
References:	<input type="text"/>		
Submitter:	<input type="text"/>		
Comments:	<input type="text"/>		

File Upload

Load File

図 6 Variation report DB の登録画面例

4-5) 倫理審査方法について

NBDC での倫理委員会でのデータの預け入れと再配布についての規約に則り、運用を始める準備を行った。この際、データの統一的な扱いを容易にするために、実験条件や解析条件の記述を標準化に関しては、EBI, dbGAP と同一のメタデータを用いる方向を議論している。

4-6) 国際連携等について

EBI, dbGAP, Gene2phen など、学会の際には連携に関する会合を実施するとともにメールによる話合い続けている。

今後の見通し:

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) 原著論文発表

- ① 発行済論文数(国内(和文) 0件、国際(欧文) 5件):
 - ② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文)0件)
 - ③ 論文詳細情報
1. Miyashita A, Koike A, Jun G, Wang LS, Takahashi S, Matsubara E, Kawarabayashi T, Shoji M, Tomita N, Arai H, Asada T, Harigaya Y, Ikeda M, Amari M, Hanyu H, Higuchi S, Ikeuchi T, Nishizawa M, Suga M, Kawase Y, Akatsu H, Kosaka K, Yamamoto T, Imagawa M, Hamaguchi T, Yamada M, Moriaha T, Takeda M, Takao T, Nakata K, Fujisawa Y, Sasaki K, Watanabe K, Nakashima K, Urakami K, Ooya T, Takahashi M, Yuzuriha T, Serikawa K, Yoshimoto S, Nakagawa R, Kim JW, Ki CS, Won HH, Na DL, Seo SW, Mook-Jung I; Alzheimer Disease Genetics Consortium, St George-Hyslop P, Mayeux R, Haines JL, Pericak-Vance MA, Yoshida M, Nishida N, Tokunaga K, Yamamoto K, Tsuji S, Kanazawa I, Ihara Y, Schellenberg GD, Farrer LA, and Kuwano R: SORL1 Is Genetically Associated with Late-Onset Alzheimer's Disease in Japanese, Koreans and Caucasians. Plos ONE. 8(4): e58168, 2013 (DOI:101371)
 2. Nishida N, Sawai H, Matsuura K, Sugiyama M, Ahn SH, Park JY, Hige S, Kang JH, Suzuki K, Kurosaki M, Asahina Y, Mochida S, Watanabe M, Tanaka E, Honda M, Kaneko S, Orito E, Itoh Y, Mita E, Tamori A, Murawaki Y, Hiasa Y, Sakaida I, Korenaga M, Hino K, Ide T, Kawashima M, Mawatari Y, Sageshima M, Ogasawara Y, Koike A, Izumi N, Han KH, Tanaka Y, Tokunaga K, Mizokami M. Genome-wide association study confirming association of HLA-DP with protection against chronic hepatitis B and viral clearance in Japanese and Korean. PLoS One 7(6):e39175. 2012 (DOI: 10.1371)
 3. Otowa T, Kawamura Y, Nishida N, Sugaya N, Koike A, Yoshida E, Inoue K, Yasuda S, Nishimura Y, Liu X, Konishi Y, Nishimura F, Shimada T, Kuwabara H, Tochigi M, Kakiuchi C, Umekage T, Miyagawa T, Miyashita A, Shimizu E, Akiyoshi J, Someya T, Kato T, Yoshikawa T, Kuwano R, Kasai K, Kato N, Kaiya H, Tokunaga K, Okazaki Y, Tanii H, Sasaki T. Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. Transl Psychiatry 2:e186.2012. (DOI: 10.1038)

4. Nakaoka H, Mitsunaga S, Hosomichi K, Shyh-Yuh L, Sawamoto T, Fujiwara T, Tsutsui N, Suematsu K, Shinagawa A, Inoko H, Inoue I. Detection of ancestral informative HLA alleles confirm the admixed origin of Japanese population. *PLoS One* 8(4):e60793. 2013 (DOI: 10.1371)

5. Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, Nakamura H, Komori A, Nakamuta M, Zeniya M, Hashimoto E, Ohira H, Yamamoto K, Onji M, Kaneko S, Honda M, Yamagiwa S, Nakao K, Ichida T, Takikawa H, Seike M, Umemura T, Ueno Y, Sakisaka S, Kikuchi K, Ebinuma H, Yamashiki N, Tamura S, Sugawara Y, Mori A, Yagi S, Shirabe K, Taketomi A, Arai K, Monoe K, Ichikawa T, Taniai M, Miyake Y, Kumagi T, Abe M, Yoshizawa K, Joshita S, Shimoda S, Honda K, Takahashi H, Hirano K, Takeyama Y, Harada K, Migita K, Ito M, Yatsushashi H, Fukushima N, Ota H, Komatsu T, Saoshiro T, Ishida J, Kouno H, Kouno H, Yagura M, Kobayashi M, Muro T, Masaki N, Hirata K, Watanabe Y, Nakamura Y, Shimada M, Hirashima N, Komeda T, Sugi K, Koga M, Ario K, Takesaki E, Maehara Y, Uemoto S, Kokudo N, Tsubouchi H, Mizokami M, Nakanuma Y, Tokunaga K, and Ishibashi H: Genome-wide association study identifies *TNFSF15* and *POU2AF1* as susceptibility loci for primary biliary cirrhosis in Japanese. *Am. J. Hum. Genet.* 91(4): 721-728, 2012. (doi: 10.1016)

(3-2) データベースおよびウェブツール等の構築と公開

公開中のデータベース・ウェブツール等
別紙を参照

(3-3) その他の著作物(総説、書籍など)

川嶋実苗、徳永勝士:よくわかるゲノムワイド関連解析②:GWASを読む - 質の管理、医学のあゆみ 243(6): 555-560, 2012.

徳永勝士:よくわかるゲノムワイド関連解析③: GWAS の今後の課題 医学のあゆみ 243(10): 934-936, 2012.

徳永勝士:GWAS(ゲノムワイド関連解析)による疾患リスク遺伝子の同定と人類学、特集【ゲノム人類学入門】「出アフリカ」以後の人類拡散と疾患リスク遺伝子 - もう一つの進化医学、生物の科学 遺伝 67(3):317-321, 2013.



(3-4) 国際学会および国内学会発表

① 招待講演 (国内 3 件、国際 4 件)

〈国内〉

1. 発表者、タイトル、学会名、場所、月日

徳永勝士: 基調講演: 疾患関連遺伝子のゲノム全域探索の現状と展望、「ワークショップ: 最近の遺伝子研究からみた肝臓病の現状と個別化医療への展望」第 48 回日本肝臓学会総会、金沢、2012.6.7.

徳永勝士: 疾患感受性と薬剤応答性: ゲノムからのアプローチ、シンポジウム「組織適合性の将来」、第 11 回日本組織適合性学会近畿地方会、大阪、2013.2.2.

徳永勝士: 疾患感受性と薬剤応答性: ゲノム解析研究の貢献、平成 24 年度 次世代医学セミナー・シンポジウム「ダイナミックなゲノムー遺伝子解析の最前線」、福島、2013.2.28.

〈国際〉

Tokunaga K: Genome-wide approaches to complex disease: advances and perspectives, Genomic Analysis of Diseases Workshop, Nature Conference, Hangzhou, China, 2012.5.17-19.

Tokunaga K: Genome-wide search for disease genes and drug response genes: Implications and perspectives, The 39th International Symposium on Nucleic Acids Chemistry ISNAC2012, 名古屋, 2012.11.17.

Tokunaga K: Genomic approaches to the etiology of autoimmune diseases, Centennial of Hashimoto Disease International Symposium II - Autoimmune Diseases - Etiology and Therapeutics, 福岡, 2012.12.2.

Tokunaga K, Koike A, Inoue I, and Tsuji S: Human Variation Database in Japanese Integrated Database Project, International Symposium on Genome Science Expanding Frontiers of Genome Science, 東京, 2012.1.9.

② 口頭講演 (国内 0 件、国際 0 件)

③ ポスター発表 (国内 3 件、国際 2 件)

〈国内〉

1. 小池麻子、西田奈央、吉田真希子、川嶋実苗、井ノ上逸朗、辻省次、徳永勝士、統合化推進プロジェクトにおけるヒトゲノムバリエーションデータベースの拡張 人類遺伝学会年会 2012年10月25日
2. 小池麻子、西田奈央、吉田真希子、川嶋実苗、井ノ上逸朗、辻省次、徳永勝士、データ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、トーゴーの日シンポジウム 時事通信ホール 2012年10月05日

特別展示のデモ

1. 小池麻子、西田奈央、川嶋実苗、吉田真希子、井ノ上逸朗、辻省次、徳永勝士、分子生物学会 データ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、福岡国際会議場 2012年12月11-14日

〈国際〉

1. 発表者、タイトル、学会名、場所、月日

1) A. Koike, N. Nishida, M. Yoshida, M. Kawashima, I. Inoue, S. Tsuji, K. Tokunaga, Extensions of Human Variation Database in Integrated Database Project, ASHG, Nov. 11th, 2011.

- 2) A. Koike, M. Yoshida, K. Tokunaga

Extensions of Human Variation Database in Database Integration Program, AGBT, Florida, Feb, 21th, 2012.

(3-5) 知財出願

- ① 平成24年度特許出願件数(国内 0件、海外 0件)
- ② 平成23、24年度の累積件数(国内 0件、海外 0件)

- ③ その他の知的財産権

なし

(3-6) 受賞・報道等

なし

§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)

なし

