

令和元年5月31日現在

機関番号：14401

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K07222

研究課題名（和文）疾患に関連するゲノム解析のための統合化された情報処理基盤の構築

研究課題名（英文）Development of an integrated data-processing environment for disease-related genome analysis

研究代表者

中谷 明弘（NAKAYA, Akihiro）

大阪大学・医学系研究科・特任教授（常勤）

研究者番号：60301149

交付決定額（研究期間全体）：（直接経費） 3,800,000円

研究成果の概要（和文）：疾患ゲノムの解析に必要となる情報処理システムの開発を進めた。1000ゲノムプロジェクトの日本人DNAサンプルの10検体分（男女各5検体ずつ）の全エクソン配列の解読を行った。同プロジェクトから公開されている日本人DNAサンプルの約100検体分の全エクソン配列の解読の結果と併せてデータベース化した。集団内でゲノム配列中の変異領域がどのように分布しているかを評価するアプリケーションプログラムを開発した。染色体に沿ったゲノムスキャンによって特定の条件を満たす変異が連続する領域をブロック構造として抽出する手法も開発した。これらの開発物を活用して実データの解析を臨床系研究者との共同研究によって実施した。

研究成果の学術的意義や社会的意義

独自に開発したソフトウェアとデータベースを用いて解析を行うことができる環境を整備した。コントロール検体群のゲノム情報の整備を行った。ヒト疾患に関する実データの解析を行った。

研究成果の概要（英文）：We developed a data processing system for disease-related genome analysis. We carried out whole-exome sequencing (WES) of ten Japanese individuals using the DNA samples provided by the 1000 Genomes Project. We processed the sequencing results using the analysis pipeline and stored the obtained variants in a database developed for this study. The database also includes the variants in the approximately one hundred Japanese individuals obtained by processing the results of WES provided by the 1000 Genomes Project. We developed an application program that evaluates how the chromosomal regions with variants distribute in a population. We also developed a genome scanning method that extracts the regions where variants satisfying a given condition are found with high frequency and then generates block diagrams representing the regions and their mutual relationships. Using the database and programs developed, we carried out analysis of clinical data in collaboration with clinical researchers.

研究分野：ゲノム情報学

キーワード：ゲノム解析 次世代シーケンサ マイクロアレイ ヒト疾患関連変異 解析ソフトウェア データベース

様式 C-19、F-19-1、Z-19、CK-19(共通)

1. 研究開始当初の背景

ヒト検体の次世代シーケンサー (NGS: Next-Generation Sequencer) による全エクソン配列解析 (WES: Whole Exome Sequencing) や高密度マイクロアレイが SNV (一塩基変異) や InDel (短い挿入/欠損) あるいは CNV (コピー数変異) の検出に用いられている。典型的な例では、WES 解析パイプライン (Illumina HiSeq) によってヒト 1 検体当たり約 8 万箇所の変異 (SNV および InDel) が検出され、アレイ出力データからは数十万~百数十万箇所での SNV/CNV の変異情報が得られている (Affymetrix Axiom/SNP6.0)。しかしながら、それらの膨大な情報は、変異と疾患の関連性、あるいは、変異同士の関連性を直接的かつ直観的に示すものではない。さらに、多因子的形質や量的形質を制御する変異群が構成する交互作用ネットワークの存在が示すように、変異を「点」ではなく「領域」、あるいは、それらの「相互関係」に基づいて扱うことが必要になる (手作業や表計算ソフトでは難しい情報処理である)。データの加工や集計に止まらない探索機能や可視化機能をもつ高度なソフトウェアが欠かせない状況が続いている。

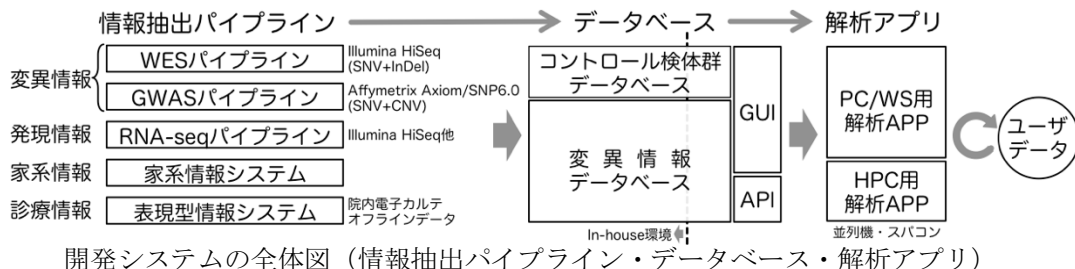
2. 研究の目的

次世代シーケンサーや高密度マイクロアレイが出力するゲノム配列の情報を用いて、ヒトの疾患に関連する変異の探索を行う情報処理システムの開発を行う。そのために必要となる情報抽出パイプライン、変異情報データベース、解析アプリケーションソフトウェアを実際の大規模ヒト検体群データの解析を通して設計および構築する。疾患関連の変異領域を探索する新しい解析手法の開発、および、解析で必要となるコントロール検体群のゲノム配列の解読とデータベース化を進める。これらの処理機能やデータが対話的かつ視覚的に統合化された誰でも使えるソフトウェアとデータベースを開発する。

3. 研究の方法

疾患ゲノムの解析に必要な情報処理システムをパイプライン・データベース・アプリの 3 項目から構成して開発を進める。パイプラインの整備を進めると共に、コントロール検体群のゲノム配列を解読してデータベース化を進める。それと並行して、疾患家系内や集団内でゲノム配列中の変異領域がどのように分布しているかを抽出するアプリを開発する。解析結果は随時データベース化を進める。現場の研究者が直観的に活用するためのユーザインタフェースや可視化システムの開発も進める。

開発するソフトウェアとデータベースの互いの関係は下図の通りである。(1) 情報抽出パイプラインによって、変異・発現・家系・診療に関する情報を解析用データとして整備する。それらのデータを (2) 変異情報データベースに格納する。また、1000 人ゲノムプロジェクトの日本人検体の WES 解析を行った結果をコントロール検体群データベースに格納する。それらを用いた実際のデータ解析を行う (3) 解析アプリケーションソフトウェアを開発する。



4. 研究成果

全エクソン配列解析 (WES 解析) パイプラインは、独自の分散ジョブ管理システムを構築し、クラスター型計算機の計算ノード (20CPU コア・256GB メモリ) 毎に 1 時間当たり 1 検体弱の高スループットを達成している。データベースは、Linux サーバ上の web サーバ (Apache) と連携する関係データベース (MySQL 互換) へのデータの蓄積と PHP による web インタフェースの構築を行える環境を整備済みである。GUI と可視化部分に関しては、web アプリは JavaScript (HTML5/Canvas) によって構築、PC/WS 用解析アプリは OS に非依存に JavaVM 上の Java/JavaFX アプリケーションとして構築した (正確には Scala/ScalaFX アプリケーション)。

(1) 情報抽出パイプラインの整備

次世代シーケンサー (Illumina HiSeq) による全エクソン配列解析 (WES 解析) パイプラインを構築した。全ゲノム配列解析 (WGS 解析) にも対応できるように構成されている。RNA-seq 解析パイプラインおよび DNA メチル化解析パイプラインを構築した。網羅的染色体高次構造解析のための環境の整備も進めた。また、高密度アレイ (Affymetrix Axiom/GeneChip) によるコピー数変異解析 (CNV 解析) を対象とした GWAS 解析パイプラインを構築した。また、家系情報システムとして、検体間の血縁情報の記述と蓄積のための環境を整備を進めた (特許出願済の関連技術に基づく: 特開 2016-207078)。表現型情報システムとして、疾患の有無や各種検査値等

の表現型情報の蓄積を行う仕組みを整備した。これらによって解析に用いる初期データの取得と蓄積を行う環境を整備した。

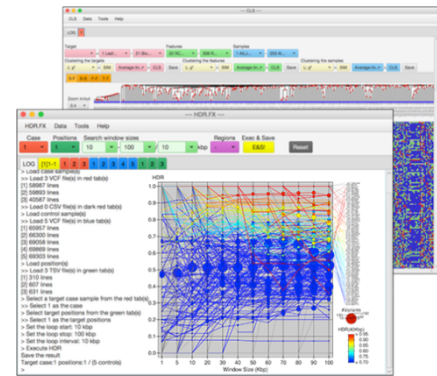
(2) 変異情報データベースの構築

米国Coriell Instituteから購入した1000ゲノムプロジェクトの日本人DNAサンプル(MGP00009, PANEL OF 120 JAPANESE IN TOKYO, JAPAN)から選抜した10検体(男女各5検体ずつ)のIllumina HiSeq 2500による全エクソン配列の解読(WES解析)を行った。キャプチャキットとしてAgilent SureSelect All Human Exon V6を用いた。その一部(男女2検体ずつ)は同キットV5も用いて全エクソン配列の解読を行い、さらにその一部(男女1検体ずつ)は全ゲノム配列の解読(WGS解析)を行った。検体の選抜に当たっては、文献化されている既存研究の結果や1000ゲノムプロジェクトによって公開されている全エクソン配列解析のデータ(Agilent SureSelect All Human Exon V2)の主成分分析を行った結果に基づいた。配列の解読は外部委託(タカラバイオ(株))にて実施して、未加工のシーケンサ出力(FASTQ形式)を入手した。

上記コントロール検体群の次世代シーケンサ出力(FASTQ形式)を構築した解析パイプラインで処理し、注釈情報の付与された変異情報(VCF形式)を獲得した。各検体の変異情報の集計を行ってデータベース化を進めた。また、1000ゲノムプロジェクトから公開されている日本人DNAサンプルのシーケンサ出力(FASTQ形式, Agilent SureSelect All Human Exon V2)を同一の解析パイプラインで再処理した結果も併せてデータベース化した(約100検体)。また、B型肝炎に関連した検体群(疾患群および対照群を含む約2,000検体)のAffymetrix Axiom/GeneChipによるコピー数変異(CNV)をはじめとして、臨床系研究者との共同研究を通して解析を行ったさまざまな疾患(下記(4)項参照)に関する変異情報をデータベースないしデータストレージに蓄積した。

(3) 解析アプリケーションソフトウェアの開発

常染色体劣性遺伝の疾患ハプロタイプ(遺伝子型の並びパターン)を探索するソフトウェア(HDR:Hamming Distance Ratio法)の改良を行い、ホモ接合変異のみだけでなくヘテロ接合変異等も探索の対象に扱えるようにした(HDR2)。右図は実行時の画面のスナップショットである。また、染色体に沿ったゲノムスキャンによって特定の条件を満たす変異が連続する領域をブロック構造として抽出する手法を開発した。ブロック構造として、検体内で特定の接合状態(ホモまたはヘテロ)やコピー数変異が連続する領域や、検体群内で連鎖不平衡状態にある領域を対象として扱うことが可能になっている。本手法は令和元年度より開始の後続の課題において実装方法や可視化方法を含めて研究開発を継続する予定である。さらに、表現型と遺伝型の関係をベイズ理論(ギブスサンプリング)を用いてモデル化(数式で表現)する手法に関する理論的背景についても書籍の一部として出版した。



解析アプリケーションソフトウェア

HDR2 <http://www.gi.med.osaka-u.ac.jp/software/hdr2/>

(4) 臨床系研究者との共同研究による実臨床データの解析

遺伝性疾患を中心として、眼科疾患(円錐角膜)、精神疾患(自閉症や統合失調症)、歯学系疾患(侵襲性歯周炎や口蓋形成不全)、腎臓疾患(腎尿細管間質性疾患や家族性高血圧および代謝性アルカローシス)、循環器疾患(遺伝性心血管疾患や心筋症)、ミトコンドリア病、小児科疾患(ムコ多糖症や先天性難聴および嚥下障害)、神経性疾患(アルツハイマー病)、消化器疾患(B型肝炎や非アルコール性脂肪肝)、免疫疾患(スギ花粉症)などを対象として、全エクソン配列(WES)解析やコピー数変異(CNV)解析およびトランスクリプトーム解析(RNA-seq)などによる実臨床データの解析を実施した。

本研究は大阪大学・大学院医学系研究科・ゲノム情報学共同研究講座にて実施した。研究代表者として中谷明弘(大阪大学特任教授)がシステム開発と研究統括を担当する他、岡崎敦子(大阪大学特任助教)が遺伝学的解析と臨床情報を担当し、菊地正隆(大阪大学特任講師)が発現量解析と疾患情報を担当した。研究協力者として小林香織(大阪大学研究員/NEC医療ソリューション事業部)が参画してシステム開発を担当した。

5. 主な発表論文等

[雑誌論文] (計12件)

- ① [Imai-Okazaki A](#), Kishita Y, Kohda M, Mizuno Y, Fushimi T, Matunaga A, Yatsuka Y, Hirata T, Harashima H, Takeda A, [Nakaya A](#), Sakata Y, Kogaki S, Ohtake A, Murayama K, and Okazaki Y (2019). Cardiomyopathy in children with mitochondrial disease: prognosis

- and genetic background. *International Journal of Cardiology*, 279:115-121. (査読有)
doi:10.1016/j.ijcard.2019.01.017
- ② Imai-Okazaki A, Kishita Y, Kohda M, Yatsuka Y, Hirata T, Mizuno Y, Harashima H, Hirono K, Ichida F, Noguchi A, Yoshida M, Tokorodani C, Nishiuchi R, Takeda A, Nakaya A, Sakata Y, Murayama K, Ohtake A, and Okazaki Y (2018). Barth syndrome: Different approaches to diagnosis. *Journal of Pediatrics*, 193:256-60. (査読有)
doi:10.1016/j.jpeds.2017.09.075
- ③ Kikuchi M, Miura K, Morita K, Yamamori H, Fujimoto M, Ikeda M, Yasuda Y, Nakaya A, and Hashimoto R (2018). Genome-wide association analysis of eye movement dysfunction in schizophrenia. *Scientific Reports*, 8:12347. (査読有)
doi:10.1038/s41598-018-30646-9
- ④ Masumoto R, Kitagaki J, Fujihara C, Matsumoto M, Miyauchi S, Asano Y, Imai A, Kobayashi K, Nakaya A, Yamashita M, Yamada S, Kitamura M, and Murakami S (2018). Identification of genetic risk factors of aggressive periodontitis using genome-wide association studies in association with those of chronic periodontitis. *Journal of Periodontal Research*. (査読有)
doi:10.1111/jre.12620
- ⑤ Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, and Nakaya A (2018). Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. *bioRxiv*. 25 September 2018. (査読無)
doi:10.1101/426312
- ⑥ Imai-Okazaki A, Kohda M, Kobayashi K, Hirata T, Sakata Y, Murayama K, Ohtake A, Okazaki Y, Nakaya A, and Ott J (2017). HDR-del: A tool based on Hamming distance for prioritizing pathogenic chromosomal deletions in exome sequencing. *Human Mutation*, 38:1796-800. (査読有)
doi:10.1002/humu.23298
- ⑦ Nakazawa T*, Kikuchi M*, Ishikawa M, Yamamori H, Nagayasu K, Matsumoto T, Fujimoto M, Yasuda Y, Fujiwara M, Okada S, Matsumura K, Kasai K, Hayata-Takano A, Shintani N, Numata S, Takuma K, Akamatsu W, Okano H, Nakaya A, Hashimoto H, and Hashimoto R (2017). Differential gene expression profiles in neurons generated from lymphoblastoid B-cell line-derived iPS cells from monozygotic twin cases with treatment-resistant schizophrenia and discordant responses to clozapine. *Schizophrenia Research*, 181:75-82. (査読有)
doi:10.1016/j.schres.2016.10.012
- ⑧ Kinoshita M, Numata S, Tajima A, Yamamori H, Yasuda Y, Fujimoto M, Watanabe S, Umehara H, Shimodera S, Nakazawa T, Kikuchi M, Nakaya A, Hashimoto H, Imoto I, Hashimoto R, and Ohmori T (2017). Effect of Clozapine on DNA methylation in peripheral leukocytes from patients with treatment-resistant schizophrenia. *International Journal of Molecular Sciences*, 18:E632. (査読有)
doi:10.3390/ijms18030632
- ⑨ Yamamoto S, Kaimori J, Yoshimura T, Namba T, Imai A, Kobayashi K, Imamura R, Ichimaru N, Kato K, Nakaya A, Takahara S, and Isaka Y (2017). Analysis of an ADTKD family with a novel frameshift mutation in MUC1 reveals characteristic features of mutant MUC1 protein. *Nephrology Dialysis Transplantation*, 32:2010-7. (査読有)
doi:10.1093/ndt/gfx083
- ⑩ Hotta K*, Kikuchi M*, Kitamoto T, Kitamoto A, Ogawa Y, Honda Y, Kessoku T, Kobayashi K, Yoneda M, Imajo K, Tomeno W, Nakaya A, Suzuki Y, Saito S, and Nakajima A (2017). Identification of core gene networks and hub genes associated with progression of non-alcoholic fatty liver disease by RNA sequencing. *Hepatology Research*, 47:1445-8. (*equally contributed) (査読有)
doi:10.1111/hepr.12877
- ⑪ Imai A, Kohda M, Nakaya A, Sakata Y, Murayama K, Ohtake A, Lathrop M, Okazaki Y, and Ott J (2016). HDR: a statistical two-step approach successfully identifies disease genes in autosomal recessive families. *Journal of Human Genetics*, 61:959-63. (査読有)
doi:10.1038/jhg.2016.85
- ⑫ Imai A, Kishita Y, Nakayama Y, Fujita S, Futatani T, Kohda M, Yatsuka Y, Nakaya A, Sakata Y, Murayama K, Ohtake A, and Okazaki Y (2016). Dried blood spots for newborn screening allows easy determination of a high heteroplasmy rate in severe infantile cardiomyopathy. *International Journal of Cardiology*, 221:446-9. (査読有)
doi:10.1016/j.ijcard.2016.06.287

〔学会発表〕（計9件）

- ① 菊地正隆, 原範和, 長谷川舞衣, 宮下哲典, 桑野良三, 池内健, 中谷明弘, 染色体高次構造を介したアルツハイマー病リスクバリエーションの影響, 第41回日本分子生物学会年会, パシフィコ横浜, 2018年11月30日.
- ② 菊地正隆, 宮下哲典, 池内健, 中谷明弘, ポリジェニックハザードスコアを用いたアルツハイマー病発症年齢解析, 第37回認知症学会, ロイトン札幌・札幌市教育文化会館, 2018年10月12日.
- ③ 小林香織, 今井敦子, 山崎悟, 菊地正隆, 足尾勉, 上條憲一, 松村泰志, 中谷明弘, エクソーム解析基盤を用いた日本人変異解析における解析手法の検証と可視化, 第40回日本分子生物学会年会, 神戸国際展示場, 2017年12月8日.
- ④ 菊地正隆, 小林香織, 澤井裕美, 西田奈央, 杉山真也, 溝上雅史, 徳永勝士, 中谷明弘, 日本人集団におけるB型肝炎の全ゲノムコピー数多型解析, 日本人類遺伝学会 第62回大会, 神戸国際会議場, 2017年11月15日~18日.
- ⑤ Kohda M, Kishita Y, Mizuno Y, Imai A, Nakaya A, Hirata T, Yatsuka Y, Bornha NN, Harashima H, Murayama K, Ohtake A, and Okazaki Y. Genetic analysis of mitochondrial disorder, 第5回生命医薬情報学連合大会 (IIBMP2016), 2017年9月29日~10月1日, 東京国際交流館プラザ平成.
- ⑥ 神田将和, 木下善人, 水野洋介, 今井敦子, 中谷明弘, 平田智子, 八塚由紀子, Nurun N. Bornha, 原嶋宏子, 村山圭, 大竹明, 岡崎康司, ミトコンドリア病を疑う集団におけるゲノム解析, NGS現場の会第五回研究会, 仙台国際センター, 2017年5月22日~24日.
- ⑦ Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, and Nakaya A, The prediction method of deleterious variants for Alzheimer's disease using chromatin higher-order structure, The 7th BRI International Symposium 2017, Brain Research Institute, Niigata University, March 10-11, 2017.
- ⑧ 菊地正隆, 原範和, 長谷川舞衣, 宮下哲典, 桑野良三, 池内健, 中谷明弘, アルツハイマー病感受性領域が近接する染色体領域の同定, 第39回日本分子生物学会年会, パシフィコ横浜, 2016年12月1日.
- ⑨ Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, and Nakaya A, Identification of chromosomal regions interacting with susceptibility loci for Alzheimer's disease, Alzheimer's Association International Conference (AAIC) 2016, Metro Toronto Convention Center, Toronto, Canada, July 22-28, 2016.

〔図書〕（計1件）

- ① Nakaya A and Isobe S (2017). Derivation of linear models for quantitative traits by Bayesian estimation with Gibbs sampling. In: Varshney R, Roorkiwal M, Sorrells M (eds) Genomic selection for crop improvement. Springer, Cham, ISBN:978-3-319-63168-4, Chapter 3, 23-53.

6. 研究組織

(1) 研究分担者

研究分担者氏名：岡崎 敦子(今井 敦子)

ローマ字氏名：(OKAZAKI, atsuko)

所属研究機関名：大阪大学

部局名：大学院医学系研究科

職名：特任助教

研究者番号(8桁)：70761691

研究分担者氏名：菊地 正隆

ローマ字氏名：(KIKUCHI, masataka)

所属研究機関名：大阪大学

部局名：大学院医学系研究科

職名：特任講師(常勤)

研究者番号(8桁)：90722538

(2) 研究協力者

研究協力者氏名：小林 香織

ローマ字氏名：(KOBAYASHI, kaori)

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。