

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 20 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25430181

研究課題名(和文) アルツハイマー病のゲノム解析に向けた日本人のコピー数多型データベース

研究課題名(英文) Copy number variation database of Alzheimer's disease for 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) 4,200,000円

研究成果の概要(和文)：アルツハイマー病(AD)患者を含む2,000名以上の日本人検体群のコピー数変異(CNV)のデータベース(AD/GD CNVdb)を構築した。非発症検体群(約1,000名)に加えて、2つの発症群である早期発症群(EOAD, 約300名)と晩期発症群(LOAD, 約1,000名)が含まれる。アフィメトリクス社製マイクロアレイ(SNP 6.0)の約180万箇所のマーカを用いて3つの検体群でのCNV頻度を評価し、検体群内で閾値以上のCNV頻度のマーカが連続している領域を抽出した。得られたCNV領域を発症群と非発症群との間で比較を行ったところ、発症群においてCNVが広がっていることを示唆する結果を得た。

研究成果の概要(英文)：We developed a comprehensive database (AD/GD CNVdb) that comprises genome-wide copy number variations (CNVs) in more than 2,000 unrelated Japanese subjects including patients of Alzheimer's disease (AD). Besides the healthy controls (N=1,000), it contains diseased subjects categorized into two types, EOAD (early-onset AD, N=300) and LOAD (late-onset AD, N=1,000). All the subjects were analyzed by the Affymetrix SNP 6.0 Array to identify CNVs, and the raw intensity data at more than 1.8 million markers were converted into the actual copy number states. In each subject group, we evaluated rates of CNVs (gain and loss) at the markers, and extracted the CNV regions where the continuous markers indicate rates of CNVs higher than a threshold along the genome sequence. We then conducted case-control studies among the subject groups using the CNV regions, and visualized the results to elucidate the relationships between the structural variations and the disease.

研究分野：ゲノム情報学

キーワード：アルツハイマー病 コピー数変異 マイクロアレイ

1. 研究開始当初の背景

(1) アルツハイマー病のゲノム解析

アルツハイマー病 (AD) をはじめとする多くの疾患は、多因子性という用語が示す通り、遺伝的要因や環境的要因が複雑に影響し合っていると考えられている。数十年という期間にわたって多様に個人差のある症状を示しながら緩徐 (insidious) に進行する AD では、各個人に複雑な遺伝的背景が存在していることが予想される。

(2) 一塩基変異とコピー数変異

疾患への感受性をはじめとするさまざまな表現型には、一塩基変異 (SNV) による配列の変異に加えて、ゲノム配列中のコピー数変異 (CNV) が関連していることが明らかにされ (Sebat et al., 2004, *Science*; Iafrate et al., 2004, *Nat. Genet.*)、例えば、HapMap プロジェクトの 270 検体 (一部に日本人も含む) を用いて CNV マップが作成されている (Redon et al., 2006, *Nature*)。また、実際に、AD 等の疾患に関連した検体群での CNV 情報が蓄積されて解析可能な状態になってきている。

(3) AD の網羅的な CNV 解析

AD に関しては、孤発性 AD を標的として、SNV を用いた全ゲノム関連解析 (GWAS) が行われ、*APOE* 遺伝子が人種を超えた有力な感受性遺伝子の一つであると確認されてはいるものの、疾患のリスクを十分には説明できていない。一方、家族性 AD に関しては、*APP* 遺伝子を含むゲノム DNA 配列領域の重複が複数の家系で共通して見い出され、2010 年代に入って AD の網羅的な CNV 解析が欧米人種を中心に発表されるに至っている (表 1)。しかしながら、AD に関連した網羅的な CNV 解析の結果は日本人の検体群では未だ発表されていない。

表 1: アルツハイマー病の CNV 解析

文献	発症型 表現型	発症 群	対照 群
Heinzen 他 2010 年	孤発性晩期 (Caucasian)	331	368 531
Shae 他 2011 年	発症年齢 (連続値) (Caucasian)	40 507	0
Swaminathan 他 2011 年	孤発性晩期 軽度認知 機能障害 (Caucasian)	222 136	143
Rovelet-Lecrux 他 2012 年	孤発性晩期 家族性早期 孤発性早期 (Caucasian)	912 21 12	1,078
Ghani 他 2012 年	孤発性晩期 (Hispanic)	559	554

数字は検体数

2. 研究の目的

アルツハイマー病 (AD) のゲノム解析に向けた日本人のコピー数変異 (CNV) のデータベースを構築することを目的とする。そのために、日本人 AD 検体群でゲノム配列中の CNV の情報を抽出し、その領域サイズや検体群内頻度などのパラメータを柔軟に設定してウェブブラウザ上で検索する機能を開発する。構築したデータベースを活用して AD と健常者の CNV の情報を比較することによって、AD に特異的なゲノム構造変異の全体像の解明を進める。情報処理と同時に実験による再検証も行っており、データの信頼性を高める。同一プラットフォームで作成された日本人の AD および健常者の CNV の情報が相互リンクされたリファレンスデータベースを構築し、AD の解析を軸としつつ日本人のゲノム構造変異情報のハブとなる基礎的なリソースの整備を目指す。また、検体群間の比較による大域的な評価に加えて、網羅的な CNV の情報を個人ごとに評価することによって、検体群内での頻度や平均値では見い出せない疾患に関連する表現型とパーソナルゲノムとの関係評価を行う。稀少変異 (rare variant) が大きな役割をもつことが予想される AD の解析に向けて、関連解析や検体群の層別化 (クラスタリング) 等の手法を開発する。

3. 研究の方法

(1) 日本人 CNV 情報のデータベース化

研究開始時点までに、表 2 に示す日本人 3 検体群 (合計約 2,200 検体) および罹患同胞対検体群 (ASP) (約 150 検体) のマイクロアレイ (Affymetrix SNP 6.0) による CNV 情報 (各検体ごとに約 180 万マーカー) を得ている。これら日本人アルツハイマー病 (AD) に関連する検体のコピー数変異 (CNV) の情報をデータベース化することを最初の目的とする。本研究での CNV 情報はマイクロアレイを用いて取得したものであるが、生データは 100GB 単位のデータであって必ずしも規模が小さくはないため、マーカーごと、あるいは、検体ごとに全ての CNV 遺伝子型を整理したマスターデータベースを構築する。そこに蓄積された情報に基づき、検体群内に存在する CNV 領域のゲノム配列中の位置や領域のサイズ、検体群内での変異頻度を明らかにする。

表 2: 対象検体群内検体数 (検査/発症年齢)

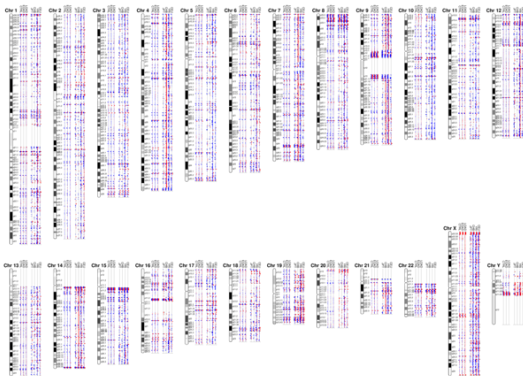
	CTRL	EOAD	LOAD	合計
女性	563 (77.1)	169 (54.7)	699 (72.9)	1,431
男性	409 (76.8)	95 (54.3)	267 (73.0)	771
合計	972	264	966	2,202

CTRL:健常群, EOAD:早期発症型, LOAD:晩期発症型

(2) AD 関連 CNV 情報のデータベース化

解析検体群は、研究分担者の桑野を中心に 10 年以上に渡って蓄積されているものであり、健常検体群を含む異なった発症種別の AD 検体群から構成されていることが特徴である。それらは、表 1 に示した既存の研究実施例を凌駕する検体数を含んでいる。最初に、既に CNV との関連が知られている早期発症型検体群 (EOAD) に特異的な CNV 情報を抽出する。その際には、健常検体群 (CTRL) および既に公開されている日本人健常者 CNV 情報との比較を行うことによって、疾患の有無に関係なく日本人に共有される CNV を除外する。引き続き、約 80 の同胞対グループを含む罹患同胞対検体群 (ASP) (約 150 検体) において、血縁関係にある検体間で共有されている CNV を抽出する。さらに、晩期発症型検体群 (LOAD) において、上記 2 検体群で得られた CNV の特性を確認すると共に、晩期発症型検体群 (LOAD) に特異的な CNV を抽出する。これらの異なるタイプの AD の検体群間の差異を探索することによって、日本人 AD に関連する CNV を包括的に探索して DB 化する。図 1 は解析検体群および HapMap 検体群他における CNV 情報の一部を図示したものである。各縦軸方向が各染色体上でのマーカーの位置を示しており、横軸方向に各位置での集団内での CNV 頻度をヒストグラム表示してある。右向き (赤) が重複、左向き (青) が欠損の頻度である。

図 1: 検体群内での CNV の分布 (詳細略)



赤:gain, 青:loss

(3) 日本人共通 CNV 情報のデータベース化

AD に特有の CNV の抽出に並行して、日本人 AD および健常検体に共通する CNV を抽出してデータベース化する。図 1 でも示した通り、同一の検体群内でも均一なゲノム構造となっている訳ではなく、逆に、異なった検体群間でも共通した変異構造となっている領域も存在しているため、AD に関する CNV 解析を行う際には、日本人に共通のゲノムの構造変異情報を明らかにして差し引く必要がある。

4. 研究成果

平成 25 年度は、DNA アレイによる生データの整理を行ってマスターデータベースを構築

した後、各検体群での CNV 抽出処理を行ってその結果をデータベース化した。既知の遺伝子や CNV との関係を精査すると共に、一部のマーカーは定量的リアルタイム PCR 法によってタイピング結果の確認を行った。平成 26 年度は、データベース検索のためのインタフェースを開発した。平成 27 年度は、AD 検体群および健常検体群のゲノム構造変異の全体像のダイアグラムを生成する手法を開発した。また、随時、CNV の領域サイズや頻度情報を集計して、AD 特異的な CNV に関する情報を整理した。また、DNA シーケンサを用いた TCC 法 (tethered conformation capture 法) による染色体高次構造情報や AD に関する文献情報を収集した情報に基づいた比較や検証を行った。

(1) 日本人初の AD 関連の CNV 情報データベースの構築

AD に関する日本人の網羅的な CNV 解析は初めて行われたものである。十分な検体数で複数の AD 疾患種別をカバーするため、日本人 AD 関連 CNV のリファレンスデータベースとしてさまざまな関連疾患の解析を含めた用途に用いることが可能である。また、統計処理等によって選別された限られた数の CNV のみではなく、全ゲノムに渡って日本人 CNV 情報を参照できるようにすることの意義は大きい。また、情報処理と同時に実験による再検証 (TaqMan アッセイによる定量的リアルタイム PCR 法) も一部で行っている。アルツハイマー病 (AD) の検体群ごとに CNV 情報を抽出してデータベースに格納し、本研究の健常検体群 (CTRL) および既に公開されている健常者 CNV 情報とも比較して、AD に特異的な CNV を抽出してある。

(2) マイクロアレイによる CNV 情報の整理

Affymetrix SNP 6.0 を用いて検体ごとの CNV 情報 (180 万マーカー) を評価した解析検体群全てのデータを本研究予算にて導入してイントラネット内に設置したサーバー上のマスターデータベースに格納した。その際には、健常検体群での CNV マップとして、本研究の健常検体群 (CTRL)、および、九州大学のグループによる CNV 領域データ

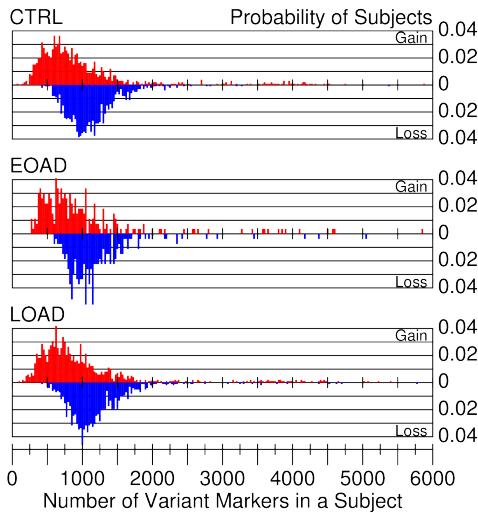
(Kukita et al., 2010, *Am. J. Hum. Genet.*) や HapMap プロジェクトによる既存の日本人健常者の CNV データを用いた (これらの健常者の CNV 情報を統合したデータベースも作成した)。ゲノムワイドな関連解析を行うとともに、既知の候補遺伝子や CNV との関係を精査し、一部のマーカーは、TaqMan アッセイによる定量的リアルタイム PCR 法 (ABI 7900HT を使用) によりタイピング結果の確認を行った。

(3) マーカー単独での評価

図 2 は、健常検体群 (CTRL)、早期発症型検体群 (EOAD)、および、晩期発症型検体群

(LOAD)において、各検体の CNV に関してコピー数の増加(赤)と減少(青)を検出したマーカー数の分布を示している。各検体でおおよそ 1,000 箇所前後の増加と減少を検出していることが分かる。また、全体的な傾向として、減少を検出したマーカー数の方が増加を検出したマーカー数よりも多くなっていることが分かる。しかしながら、このようなマーカーレベルの CNV 情報と疾患との関連性は必ずしも明らかではない。

図 2：検体毎の CNV 検出マーカー数の分布



(4) マーカー領域での評価

検体群ごとの CNV の分布の全体像は前出の図 1 に示す通りである。一方、図 2 は、使用したマイクロアレイが備える隣接するマーカー間の距離の頻度(赤)とゲノム被覆率(青)の関係を示したものである。被覆率は、距離が閾値以下のマーカーに挟まれる区間長の合計の全ゲノム配列長に対する割合で定義されている。閾値 10kbp の場合に 85%程度の被覆率であり、ゲノムワイドな CNV 解析が実現できることを示している。この結果に基づいて、閾値以下の距離だけ離れたマーカーに検体群内で増加ないし減少を検出しているマーカーがある場合に CNV 領域として検出して、その長さと同頻度を示したものが図 4 である。健常検体群 (CTRL) と早期発症型検体群 (EOAD) に比べて、晩期発症型検体群での CNV 領域が伸長していることが分かる。これらの CNV 領域の情報がデータベース化されている。

(5) AD 関連の CNV 領域の例

検体群内での CNV 領域の情報に基づいて AD に関連した CNV 領域を検索することができる。例えば、図 5 は、構築した CNV データベースと既存のアルツハイマー病に関する遺伝子のデータベースとを連携させて検索した結果を可視化したものである。既に疾患との関連が知られているある遺伝子 (Gene X) の領域 (最左列の青矢印) での CNV 情報が、左から健常検体群 (CTRL)、早期発症型検体

図 3：マーカー区間長とゲノム被覆率の関係

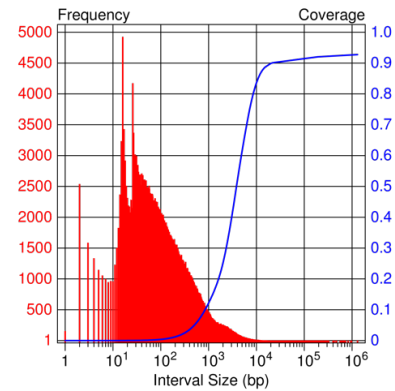
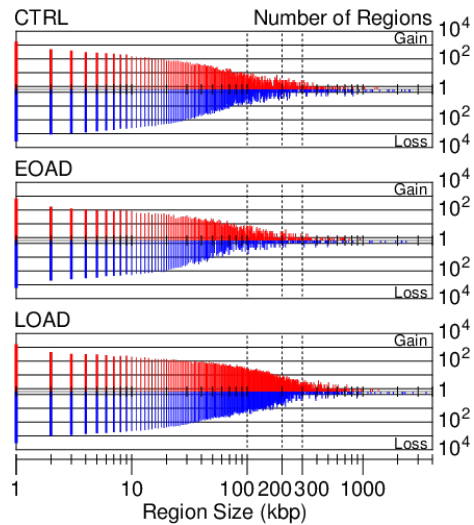
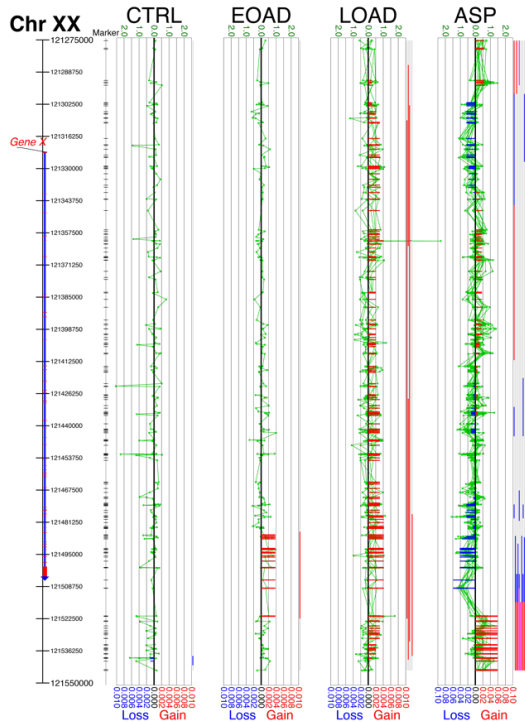


図 4：検体群毎の CNV 領域数の分布



群 (EOAD)、晩期発症型検体群 (LOAD)、および、罹患同胞対検体群 (ASP) の順にプロットされている。健常検体群 (CTRL) と早期発症型検体群 (EOAD) では、変異が検出されないかごく限られた領域のみにコピー数の増加が検出されている(赤横棒)のに対し、晩期発症型検体群 (LOAD) では遺伝子領域全体でコピー数の増加が検出されている(赤横棒)。赤横棒の長さは検体群内での変異頻度を示しており、晩期発症型検体群 (LOAD) での変異頻度は約 0.4%程度である(稀少変異)。また、最右列の罹患同胞対検体群 (ASP) においても晩期発症型検体群 (LOAD) と同様に CNV が検出されているが、CNV の種類としてはコピー数の減少も含んでいる(青横棒)。各検体群の右側には、変異を検出した検体でのコピー数の情報を付記してある。上下方向の線分 1 本が 1 検体に対応しており、染色体上の各位置で当該の検体のコピー数が増加あるいは減少していれば、それぞれ赤と青で着色してある(隣接するマーカーの CNV 情報によって両マーカー間を補完してある)。また、遺伝子領域を示す青矢印にはエクソン領域が赤色で示されている。その右側の“Marker”の列にはマーカーの位置が示されている。各検体群での緑色の折れ線はマイクロアレイによるシグナル値を示している(詳細省略)。

図 5 : AD 関連の CNV 領域の例



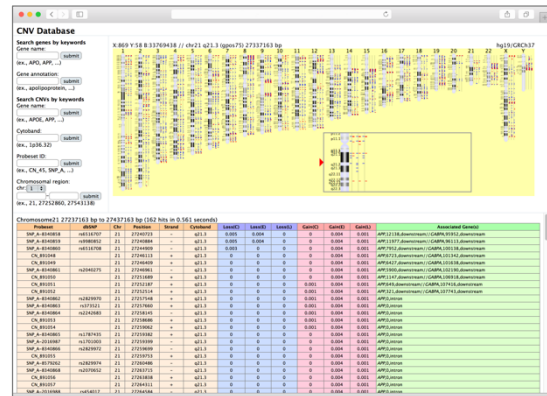
(6) データベース検索インターフェース

CNV 領域は、遺伝子領域と対応しているとは限らず、数マーカー（データの単一マーカー）のみを含む短いものから複数の遺伝子を含むような長いものまで存在している。また、マイクロアレイのデータは、遺伝子間の領域の CNV も検出する。このため、遺伝子名のみでの情報の検索や管理は十分ではない。そこで、検体群内での CNV 情報に加えて、各マーカーに関するゲノム配列中の位置（染色体と塩基位置）、近隣の遺伝子との位置関係、および、遺伝子内での領域情報（エクソン領域かイントロン領域かなど）の情報を含めた情報のデータベースを構築した。また、機能注釈情報を含めた遺伝子のデータベースも構築した。これらを連携することによって、マーカーの ID に加えてゲノム配列中の位置や遺伝子の機能情報に関するキーワードで CNV 情報が検索できるようになっている。これらのデータベースは関係データベース (MySQL) を用いて一元化して管理されている。

さらに図 6 に示すようなウェブブラウザを介した検索システムを構築した。上述のゲノム配列中の位置や遺伝子名および機能注釈情報内の文字列をクエリーとして、CNV 情報の検索を実施することができる（画面上部左側）。また、図 1 に相当する検体群毎の CNV 情報の分布が描画された部分（画面上部右側の黄色背景部分）をマウスポインタが通過すると当該領域の拡大図がリアルタイムで表示され、また、染色体上の位置をクリックするとその位置をクエリーとしたデータベース検索が実行されてその結果（各検体群内での CNV の頻度と近在の遺伝子の情報）が画面下部に表

示されるようになっている。全ての検索結果にはユニークな URL が付与される仕組みになっており、他のデータベースとの連携や統合が容易に実現できるようになっている。これらの機能は、新規に開発した PHP スクリプトと JavaScript プログラムが MySQL で構築されたデータベースと通信して実現している。また、本報告書で使用している各図は、新規に開発したデータ可視化用の Perl スクリプトによるものである（結果を PostScript 形式で出力する）。これらのシステムは、疾患や変異の種類に特化していないので、ゲノム情報に関する解析やデータベース構築で広く活用できるものになっている。

図 6 : データベース検索インターフェース



5. 主な発表論文等

(研究代表者及び研究分担者には下線)

[雑誌論文] (計 1 件)

1. Miyashita A, Hatsuta H, Kikuchi M, Nakaya A, Saito Y, Tsukie T, Hara N, Ogishima S, Kitamura N, Akazawa K, Kakita A, Takahashi H, Murayama S, Ihara Y, Ikeuchi T, and Kuwano R, Genes associated with the progression of neurofibrillary tangles in Alzheimer's disease, *Translational Psychiatry*, 4:e396, 2014.

[学会発表] (計 9 件)

1. Nakaya A, Kikuchi M, Miyashita A, Hara N, Kasuga K, Nishida N, Tokunaga K, Kuwano R, and Ikeuchi T, Genomic structure and distribution of copy number variations in relation to Japanese Alzheimer's disease, 第 38 回日本分子生物学会年会, 神戸国際会議場, 2015 年 12 月 1 日~4 日.
2. 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 38 回日本分子生物

- 学会年会, 神戸ポートピアホテル, 2015年12月1日~4日.
3. Nakaya A, Kikuchi M, Miyashita A, Hara N, Kasuga K, Nishida N, Tokunaga K, Kuwano R, and Ikeuchi T, AD/GD CNVdb (Alzheimer's Disease Genomic Databases, Copy Number Variation Database), 日本バイオインフォマティクス学会2015年年会, 京都大学宇治キャンパスおうばくプラザ, 2015年10月29日~31日.
 4. 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, 日本人類遺伝学会第60回大会, 京王プラザホテル, 2015年10月14日~17日.
 5. 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, American Society of Human Genetics Annual Meeting 2015 (ASHG2015), Baltimore, MD, USA, October 6-10, 2015.
 6. 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34回日本認知症学会学術集会, リンクステーション青森・ホテル青森, 2015年10月2日~4日.
 7. Kikuchi M, Hara N, Hasegawa M, Miyashita A, Kuwano R, Ikeuchi T, and Nakaya A, Identifying chromosomal regions interacting with Alzheimer's disease GWAS SNPs, QBiC Symposium 2015, 理化学研究所生命システム研究センター, 大阪, 2015年8月24日~26日.
 8. 中谷明弘, 宮下哲典, 菊地正隆, 長谷川舞衣, 原範和, 西田奈央, 徳永勝士, 井原康夫, 池内健, 桑野良三, AD/GD CNVdb: アルツハイマー病のコピー数変異データベース, 第36回日本分子生物学会年会, 神戸国際会議場, 2013年12月3日.
 9. 中谷明弘, 疾患関連ゲノム多型情報の解析手法, NGS現場の会第三回研究会, 神戸国際会議場, 2013年9月4日~5日.

6. 研究組織

(1) 研究代表者

中谷明弘 (NAKAYA, Akihiro)
新潟大学・研究推進機構・准教授
(平成 25 年度)

大阪大学・大学院医学系研究科・特任教授(常勤)
(平成 26 年度~平成 27 年度)

研究者番号: 60301149

(2) 研究分担者

桑野良三 (KUWANO, Ryozo)

新潟大学・脳研究所・フェロー

研究者番号: 20111734

宮下哲典 (MIYASHITA, Akinori)

新潟大学・脳研究所・助教

研究者番号: 60323995

(平成 25 年度まで研究分担者)