

令和元年6月3日現在

機関番号：12602
 研究種目：基盤研究(C)（一般）
 研究期間：2016～2018
 課題番号：16K07211
 研究課題名（和文）エクソーム解析における疾患原因変異同定率の低迷を打破するための多面的アプローチ

研究課題名（英文）Development and assessment of a pathogenic mutations search method for individuals who are difficult to identify the mutations in standard-exome analysis

研究代表者
 宮 冬樹（Miya, Fuyuki）
 東京医科歯科大学・難治疾患研究所・講師

研究者番号：50415311
 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）：次世代シーケンサーを用いたエクソーム解析は疾患原因変異同定に有用な手法であるが、その同定率はこれまで30%程度で未同定例が多いという課題があった。そこで我々は、原因変異同定率を更に向上させるための複数手法を統合させた解析手法を開発した。通常法に加え、独自開発した挿入・欠失（indel）検出ソフトの開発、コピー数変異の検出における最良法の検証と組み込み、ミトコンドリアDNAの解析手法の追加、タンパク質ドメイン解析による候補変異の絞り込み手法、等を組み込むことで新たな疾患変異を同定する手法を確立した。実際に多数の新規疾患原因変異を同定し、研究期間内に20報を超えるエクソーム関連論文を発表した。

研究成果の学術的意義や社会的意義
 我々の研究成果で疾患の遺伝子変異を同定する新たな手法論を構築し発表することで、その同定率を10%以上増やすことができた。解析手法を改良することによって、未同定疾患原因変異の新たな発見も可能であることを証明したことで、将来的な解析への応用は当然であるが、過去データの再検証で新規原因変異が同定されることも期待される。病気の原因変異が分かると、病気のメカニズム解明や薬剤ターゲット候補の発見、将来の発症予測、等に役立たせることができ、本成果は今後その一助になると考えられる。

研究成果の概要（英文）：Whole-exome sequencing (WES) using next generation sequencing is a useful method to identify disease-causing mutations. However, often no candidate mutations are identified using commonly available methods. The reported success rate of WES for Mendelian diseases is around 30% worldwide. We have developed an integrative analysis methodology to identify pathogenic mutations. The analysis methodology consists of the following 6 methods: 1) standard variants calling method, 2) detection of intermediate-size insertions and deletions (indels) using our own developed method, 3) copy number variants analysis for WES data, 4) analysis of mitochondrial DNA for WES data, and 5) domain enrichment analysis for candidate mutations. Through this combinatorial method, it becomes possible to identify variants and structural abnormalities that had been difficult to search and identify previously. Also, the method makes it possible to expand the limits of mutation detection of standard WES analysis.

研究分野：メディカルバイオインフォマティクス

キーワード：エクソーム解析 疾患遺伝子変異 遺伝性疾患 次世代シーケンサー CNV解析 ミトコンドリア 全ゲノムシーケンズ

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

次世代シーケンサーを用いた疾患研究では、既知の疾患原因変異の90%以上がタンパク質をコードするエクソン領域に存在することを鑑み、エクソン領域 DNA のみを対象としてキャプチャーしてシーケンスする「エクソーム解析」が2010年頃から世界的に広く実施されてきた。これまで申請者らも先天性神経疾患（小頭症、大頭症、大脳皮質形成異常症、等）を対象にしたエクソーム解析により多数の新規疾患原因変異を同定してきた（Okamoto, Miya et al. J. Hum. Genet. 2014、Harada, Miya et al. Child Nerv. Syst. 2015、等）。

しかしながら、エクソーム解析により多数の疾患原因変異が発見されているものの、実際の疾患原因変異同定率は世界的に約30%程度と見積もられている。残り70%が未同定の原因は様々考えられるが、以前、申請者らは、原因変異がエクソームではカバーし切れていないエクソン領域に存在する可能性を考え、その領域を補完する手法を開発した（Miya et al. Sci. Rep. 2015）。申請者らは、一般的な手法では約7~10%のエクソン領域が読み逃されていることを確認し、広く使われているハイブリダイゼーション法に加え、読み逃し領域に対して選択的環状化法という別の手法を組み合わせることで、エクソン領域のカバー率を97%以上まで上昇させ、新規の疾患原因変異を同定することにも成功した。これにより、エクソーム解析では候補変異が何も残らなかった検体の14%程度で新たな変異を同定できる可能性を示唆したが、未だに未同定検体が多数存在するのが実情であった。

2. 研究の目的

本研究ではエクソーム解析の同定率をさらに上げ、多数の疾患原因変異を明らかとするための探索法の確立を第一義の目的とし、実際に疾患原因変異を探索することとした。当初からの目的として、エクソームからのコピー数変異（CNV）探索の最適条件の決定、エクソームデータからのミトコンドリアDNA（mtDNA）解析、を挙げておりその手法の確立を行った。さらに、当初予定以上の手法開発も行い、中間サイズ（50塩基から1万塩基程度）の挿入・欠失（indel）の同定法の開発、候補変異のタンパク質ドメイン解析による絞り込みの検証、も目的とした。また、当初はエクソーム補完リシーケンス法の改良も予定していたが、想定以上に次世代シーケンスの技術開発と低コスト化が進んだため、予定を変更し、変異解析において全ゲノムシーケンスがエクソームシーケンスに変わりうるかの比較検証、へ拡張することにした。

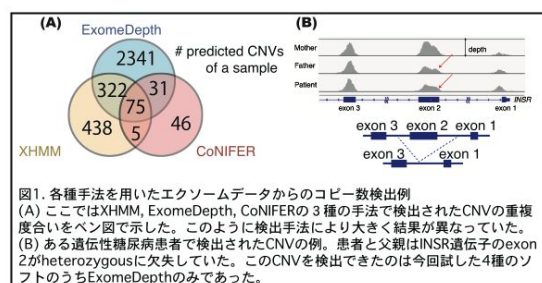
3. 研究の方法

全ての方法について、検体は申請者がこれまで実施してきた主に先天性神経疾患および先天性糖尿病のエクソーム検体データを用いた。

- (1) エクソームからのCNV探索の最適条件の決定
既存のエクソームデータに対して、CNV解析ソフトウェアを複数試行し、統合解析手法の最適条件を検証し、実際に疾患原因CNVが同定可能かを検証した。
- (2) エクソームデータからのmtDNA解析
通常のエクソーム解析ではエクソンキャプチャープロブのターゲット領域のみの解析を行うが、元々の存在量の豊富さからmtDNAが非特異的に混入しているという報告が知られていたことから、解析対象領域をmtDNAにも拡張し、十分なシーケンス量が得られているかと、実際のmtDNA上の疾患原因変異を検出できるかどうかを検証した。
- (3) 中間サイズ（50塩基から1万塩基程度）のindelの同定法の開発
次世代シーケンスデータのマッピングの際にリードの両端部分がマッピングされないリードを収集し、再アライメント等を行うことで、通常のGATK等のcallerでは見つけることができない長鎖のindelを同定するプログラムを構築し、他ソフトと比較検証、サンガーシーケンスで確認実験を行った。
- (4) 候補変異のタンパク質ドメイン解析による絞り込みの検証
遺伝性糖尿病のうち、インスリン受容体遺伝子（*INSR*）に疾患原因変異が同定された検体およびデータベース上の世界的な*INSR*の変異情報と表現型を比較解析し、特定のドメインにエンリッチメントしていないかどうかを統計解析し、立体構造解析も行った。これにより、疾患候補変異の絞り込みへの応用につなげられることが期待される。
- (5) 変異解析において全ゲノムシーケンスがエクソームシーケンスに変わりうるかの比較検証
エクソーム解析を実施済のヒト健康人由来リンパ芽球細胞株NA18943に対し、全ゲノムシーケンスもを行い、エクソーム解析が対象としているタンパク質コーディング領域および既知の疾患原因変異領域のカバー率を解析し、費用対効果の面から手法としてどちらが優れているのかを検証した。

4. 研究成果

- (1) エクソームからのコピー数変異探索の最適条件の探索結果
多数のCNV解析ソフトウェアの中から、



call 手法が異なる ExomeDepth、XHMM、CoNIFER、CODEX2、EXCAVATOR2 を用いて、既存の申請者らが実施済の約 800 検体のエクソームデータについて CNV 解析を実施した。結果として、ソフトごとに検出された CNV が大きく異なるデータが得られた (図 1A)。一例として、糖尿病の原因変異 CNV が検出された例では、各種ソフトのうち ExomeDepth のみで原因変異が検出された (図 1B)。多数ソフトの積集合が高精度と考えられるが、糖尿病例のように真の疾患原因変異 CNV が特定のソフトでのみ検出されることもあり、多数ソフトを統合して和集合で得られた CNV から遺伝子アノテーション情報によって絞り込んでいくことが最も有用な手法と考えられた。

(2) エクソームデータからの mtDNA 解析結果

既存の申請者らが実施済の約 800 検体のエクソームデータについて mtDNA が call できているかどうかを調査した結果、エクソームキャプチャーのターゲット領域外の mtDNA も全長で call されていることが示された (図 2A)。元々の mtDNA はゲノム DNA に比べ量が豊富であり、非特異的な混入により良い意味で想定外の検出が可能であったと考えられた。実際に疾患原因変異も同定できることを証明した (図 2B)。

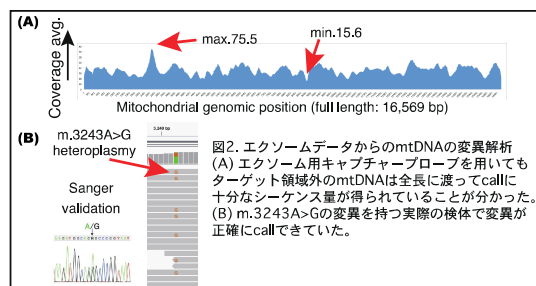


図2. エクソームデータからのmtDNAの変異解析 (A) エクソーム用キャプチャープロブを用いてもターゲット領域外mtDNAは全長に渡ってcallに十分なシーケンス量が得られていることが分かった。(B) m.3243A>Gの変異を持つ実際の検体で変異が正確にcallできていた。

(3) 中間サイズ (50 塩基から 1 万塩基程度) の indel の同定法の開発

通常の caller (GATK 等) では検出が難しい中間サイズの indel の call 手法を開発した (図 3A)。既存の他の長鎖 indel caller のソフトに比べても我々の IMSindel と名付けた手法の方がより長く、多数の indel が検出されていた (図 3B に NA18943 細胞の検出結果を示す)。478 検体のエクソームデータについて検証した結果、大部分は singleton (一人だけが保因する indel) であり (図 3C)、これは通常の SNV や短い indel と同様の傾向であった。ソフトウェアを全世界に公開し、論文化した (発表論文 3)。

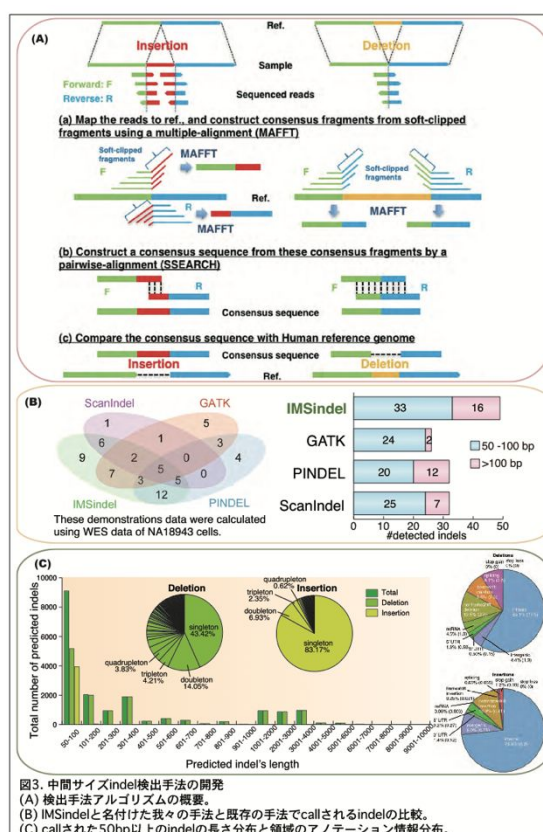


図3. 中間サイズindel検出手法の開発

(A) 検出手法アルゴリズムの概要。(B) IMSindelと名付けた我々の手法と既存の手法でcallされるindelの比較。(C) callされた50bp以上のindelの長さ分布と領域のアノテーション情報分布。

(4) 候補変異のタンパク質ドメイン解析による絞り込みの検証

遺伝性の糖尿病で *INSR* 遺伝子に変異を有する我々の検体およびデータベースから収集した情報を元に、タンパク質の特定のドメイン上に変異が集積していないかを調査した結果、重症タイプは Fibronectin type III ドメインに集積している等の結果が得られ、タンパク質立体構造解析でも大きな構造変化が生じることが明らかとなった (図 4)。本

結果は発表論文 4 にて発表した。本結果のように、別疾患でも重症化予測や予後予測等に使用できる可能性があり、候補変異からの絞り込みにも応用可能と考えられた。

(5) 変異解析において全ゲノムシーケンスがエクソームシーケンスに変わりうるかの比較検証

ヒト健康人由来リンパ芽球細胞株

NA18943 に対し、エクソーム解析と全ゲノムシーケンスの双方を行い比較検証した。エクソーム解析が対象としているタンパク質コーディング領域は、エクソームシーケンス (WES) の方が全ゲノムシーケンス (WGS) よりも 3 倍以上平均読み深度 (depth) が深いデータを用いたが、その場合でも実際の call が得られているコーディング領域の塩基数は WGS の方が多いという結果であった (図 5)。この要因は、エクソーム解析ではキャプチャーが難しい領域 (キャプチャープロブの設計が難しい領域とほぼ同義と考えられる) についても WGS ではシーケンスが得られていること、エクソームキャプチャープロブの設計を行った時から現在までの間にデー

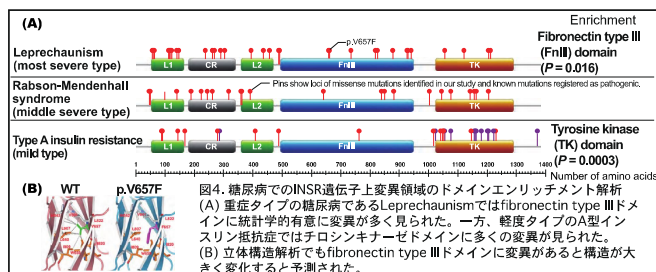


図4. 糖尿病でのINSR遺伝子変異領域のドメインエンリッチメント解析 (A) 重症タイプの糖尿病であるLeprechaunismではFibronectin type IIIドメインに統計学的有意に変異が多く見られた。一方、軽度タイプのA型インスリン抵抗症ではチロシンキナーゼドメインに多くの変異が見られた。(B) 立体構造解析でもfibronectin type IIIドメインに変異があると構造が大きく変化すると予測された。

データベース上のコード領域の情報が更新されてプローブの未設計領域があること、の2点が主な要因と考えられた。さらに既知の疾患遺伝子変異のカバー率もWGSの方が高かった(図5)。両者でcallが得られている塩基の一致率は99.994%と非常に高く、WGSもcall精度が高いと考えられた。不一致領域のサンガーシーケンスでは、エクソームとWGSの双方の正解率がほぼ同程度でWGSの方がやや高いほどであった。

We sequenced the NA18943 sample with both WES and WGS.

	WES with HiSeq4000 (SureSelect V6)	WGS with BGISEQ500
Sequenced bases	9.09 Gb	100.2 Gb
Mean depth in CDS regions	98.37	35.20
Coverage at least 5x in CDS regions	95.96%	99.80%
Coverage of HGMD disease causative mutations	99.76%	99.93%
Coverage of NCBI ClinVar pathogenic mutations	99.80%	99.89%
#Mismatch call in CDS regions called bases by both	211 (/ 35,956,922)	
Concordance rate	99.9994%	

図5. 同一検体のエクソーム解析と全ゲノム解析データの比較

結果の総論として、上記のような新規手法を用いて実際に先天性神経疾患や糖尿病の解析を行い、多数の新たな疾患原因変異を同定することができ、発表論文に記載した20報以上の論文を出版した。疾患原因変異同定率は、10%以上向上させることができたが、未だに疾患原因変異が未同定の検体が約半数残っている要因の解明は今後の課題でもある。最後に記載した全ゲノム解析は、ゲノムの大きな構造異常や、intronや非遺伝子領域のCNVのブレークポイントの同定に優位であり、エクソーム解析とWGSの費用差が2倍以内程度まで迫ってきていることから、基礎研究の分野ではWGSの活用が現実味を帯びてきたと考えられる。また、近年はAI(人工知能・ディープラーニング)をゲノム解析分野でも活用する動きが広がり始めており、実際に申請者も解析を開始したところである。WGSとAIの組み合わせにより、今後更なる疾患原因変異の同定が期待され、引き続き疾患原因変異の更なる同定のための研究を継続していく予定である。

5. 主な発表論文等

[雑誌論文](計24件)

1. Kato K, **Miya F**, (他15名、2番目). MYCN de novo gain-of-function mutation in a patient with a novel megalencephaly syndrome. *J. Med. Genet.* (2018, in press). doi: 10.1136/jmedgenet-2018-105487.
2. Horii I*, **Miya F*** (*equal contribution), Negishi Y, Hattori A, Ando N, Boroevich KA, Okamoto N, Kato M, Tsunoda T, Mami M, Kanemura Y, Kosaki K, Saitoh S. A novel homozygous missense mutation in the SH3-binding motif of STAMBP causing microcephaly-capillary malformation syndrome. *J. Hum. Genet.* **63**, 957-963 (2018). doi: 10.1038/s10038-018-0482-3.
3. Shigemizu D*, **Miya F*** (*equal contribution), (他16名、2番目). IMSindel: An accurate intermediate-size indel detection tool incorporating de novo assembly and gapped global-local alignment with split read analysis. *Sci. Rep.* **8**, 5608 (2018). doi: 10.1038/s41598-018-23978-z.
4. Hosoe J*, Kadowaki H*, **Miya F*** (*equal contribution), (他16名、3番目). Structural basis and genotype-phenotype correlations of INSR mutations causing severe insulin resistance. *Diabetes* **66**, 2713-2723 (2017). doi: 10.2337/db17-0301.
5. Okamoto N*, Tsuchiya Y*, **Miya F*** (*equal contribution), Tsunoda T, Yamashita K, Boroevich KA, Kato M, Saitoh S, Yamasaki M, Kanemura Y, Kosaki K, Kitagawa D. A novel genetic syndrome with STARD9 mutation and abnormal spindle morphology. *Am. J. Med. Genet. A* **173**, 2690-2696 (2017). doi: 10.1002/ajmg.a.38391.
6. Okamoto N, **Miya F**, Hatuskawa Y, Suzuki Y, Kawato K, Yamamoto Y, Tsunoda T, Kato M, Saitoh S, Yamasaki M, Kanemura Y, Kosaki K. Siblings with optic neuropathy and RTN4IP1 mutation. *J. Hum. Genet.* **62**, 927-929 (2017). doi: 10.1038/jhg.2017.68.
7. Kato K, **Miya F*** (*equal contribution), Horii I, Ieda D, Ohashi K, Negishi Y, Hattori A, Okamoto N, Kato M, Tsunoda T, Yamasaki M, Kanemura Y, Kosaki K, Saitoh S. A novel

- missense mutation in the HECT domain of NEDD4L identified in a girl with periventricular nodular heterotopia, polymicrogyria, and cleft palate. *J. Hum. Genet.* **62**, 861-863 (2017). doi: 10.1038/jhg.2017.53.
8. Hori I, Otomo T, Nakashima M, **Miya F**, (他 24 名、4 番目). Defects in autophagosome-lysosome fusion underlie Vici syndrome, a neurodevelopmental disorder with multisystem involvement. *Sci. Rep.* **7**, 3552 (2017). doi: 10.1038/s41598-017-02840-8.
 9. Okamoto N, **Miya F**, Tsunoda T, Kato M, Saitoh S, Yamasaki M, Kanemura Y, Kosaki K. Novel MCA/ID syndrome with ASH1L mutation. *Am. J. Med. Genet. A*, **173**, 1644-1648 (2017). doi: 10.1002/ajmg.a.38193.
 10. Negishi Y, **Miya F**, (他 20 名、2 番目). A combination of genetic and biochemical analyses for the diagnosis of PI3K-AKT-mTOR pathway-associated megalencephaly. *BMC Med. Genet.* **18**, 4 (2017). doi: 10.1186/s12881-016-0363-6.
 11. Hamada N, Negishi Y, Mizuno M, **Miya F**, (他 10 名、4 番目). Role of a heterotrimeric G-protein, Gi2, in the corticogenesis: Possible involvement in periventricular nodular heterotopia and intellectual disability. *J. Neurochem.* **140**, 82-95 (2017). doi: 10.1111/jnc.13878.
 12. Hori I*, **Miya F*** (*equal contribution), (他 11 名、2 番目). Novel Splicing Mutation in the *ASXL3* gene causing Bainbridge-Ropers Syndrome. *Am. J. Med. Genet. A* **170**, 1863-1867 (2016). doi: 10.1002/ajmg.a.37653.
 13. Tsutsumi M, Yokoi S, **Miya F**, Miyata M, Kato M, Okamoto N, Tsunoda T, Yamasaki M, Kanemura Y, Kosaki K, Saitoh S, Kurahashi H. Novel compound heterozygous variants in *PLK4* identified in a patient with autosomal recessive microcephaly and chorioretinopathy. *Eur. J. Hum. Genet.* **24**, 1702-1706 (2016). doi: 10.1038/ejhg.2016.119.

他 11 件

〔学会発表〕(計 36 件)

1. **Miya F**. Beyond the limitations of exome analysis for genetic disorders. 13th International Symposium of Institute Network for Biomedical Sciences, Fukuoka, 2018 Oct. 19, 口頭発表.
2. **Miya F**. Beyond the limitations of exome analysis for genetic disorders. 13th International Symposium of Institute Network for Biomedical Sciences, Fukuoka, 2018 Oct. 18, ポスター発表.
3. **宮 冬樹**, 重水 大智, 金村 米博, 齋藤 伸治, 岡本 伸彦, 加藤 光広, 松永 達雄, 務台 英樹, 小崎 健次郎, 角田 達彦. Beyond the limitations of exome analysis for genetic disorders (exome 解析による疾患原因変異探索の限界を突破するための手法の検証), 日本人類遺伝学会第 63 回大会, 横浜, 2018 年 10 月, 口頭発表.
4. **宮 冬樹**, 重水 大智, 金村 米博, 齋藤 伸治, 岡本 伸彦, 加藤 光広, 山崎 麻美, 松永 達雄, 務台 英樹, 小崎 健次郎, 角田 達彦. 既存の exome 解析では疾患原因変異同定が困難な検体の原因変異探索手法の開発 (Development and assessment of a pathogenic mutations search method for subjects where exome has previously failed), 日本人類

遺伝学会 第 62 回大会, 神戸, 2017 年 11 月 18 日, 口頭発表.

5. **宮 冬樹**, 重水 大智, 齋藤 伸治, 須藤 章, 中川 英刀, 奥田 修二郎, 岡本 伸彦, 加藤 光広, 山崎 麻美, Keith A. Boroevich, 金村 米博, 小崎 健次郎, 角田 達彦. 既存の exome 解析では疾患原因変異同定が困難な検体の原因変異探索手法の開発. 次世代現場の会 第 5 回研究会, 仙台, 2017 年 5 月, ポスター発表.
 6. **Fuyuki Miya**, Mitsuhiro Kato, Tadashi Shiohama, Nobuhiko Okamoto, Shinji Saitoh, Mami Yamasaki, Daichi Shigemizu, Tetsuo Abe, Takashi Morizono, Keith A. Boroevich, Kenjiro Kosaki, Yonehiro Kanemura, Tatsuhiko Tsunoda. A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations. ICHG (The 13th International Congress of Human Genetics), Kyoto, 2016 年 4 月, ポスター発表.
- 他 30 件

〔図書〕(計 2 件)

1. 山形 誠也, 服部 文子, **宮 冬樹**, 久保田 裕子, 遠藤 剛, 根岸 豊, 中村 勇治, 角田 達彦, 小崎 健次郎, 齋藤 伸治. 難治性てんかんと退行を示した POGZ 変異症の 1 例. 脳と発達 51 (1) 29 - 32 2019 年 1 月.
2. 緒方 怜奈, 安永 由紀恵, 渡辺 恭子, **宮 冬樹**, 加藤 光広. PNKP 遺伝子に複合ヘテロ接合変異を認めた難治性てんかんを伴う小頭症の一例. てんかん研究 35 (2) 542 - 542 2017 年 9 月.

〔産業財産権〕

特になし。

〔その他〕

研究室 WEB サイト : <http://www.tmd.ac.jp/mesm/index.html>

6 . 研究組織

(1)研究分担者

なし。

(2)研究協力者

なし。

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。