

平成 30 年 6 月 21 日現在

機関番号：34204

研究種目：若手研究(B)

研究期間：2015～2017

課題番号：15K21487

研究課題名(和文) ミスセンス変異の分子機能及び表現型への影響を予測する手法の開発

研究課題名(英文) Development of a new method to predict disease-causing mechanisms of missense mutations

研究代表者

土方 敦司 (HIJIKATA, Atsushi)

長浜バイオ大学・バイオサイエンス学部・プロジェクト特任講師

研究者番号：80415273

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：遺伝性疾患の原因となるミスセンス変異の分子機能及び疾患メカニズムを予測することは困難な場合が多く、遺伝子診断などの臨床の現場においても問題となるケースがある。本研究では、疾患原因ミスセンス変異とタンパク質立体構造、特に分子間相互作用に着目した解析を行った。その結果、変異による遺伝様式及び疾患メカニズムとタンパク質立体構造との間に明確な関係があることを新たに見出した。本結果に基づき疾患メカニズムを予測するまったく新しい手法を開発した。

研究成果の概要(英文)：Estimating molecular mechanisms of missense mutations on a particular disease is still challenging in the research field. In this study, I analyzed disease-causing missense mutations with its positions on protein structure, especially on macromolecular structures and found a clear relationship between the disease-causing mechanisms and the mutation residue positions on the 3D structures. According to the relationship, I developed a novel method to predict the disease-causing mechanisms caused by given missense mutations.

研究分野：構造バイオインフォマティクス

キーワード：ミスセンス変異 遺伝性疾患 タンパク質立体構造 データベース ハプロ不全 ドミナントネガティブ
機能獲得変異

1. 研究開始当初の背景

次世代シーケンサーによる塩基配列解析技術の進展により、これまで原因不明であった遺伝性疾患の責任遺伝子及び変異候補の特定が可能になってきた。しかし原因変異がミスセンス変異である場合、原因変異を特定することが困難であることが多い。

これまで、生命情報学的アプローチにより、ミスセンス変異の影響の有無について予測する手法が数多く開発されており、ある程度の成果は上げているものの、既存の手法は、ミスセンス変異が分子機能及び表現型への影響の中身については考慮しておらず、どのような分子機能に影響を与えるかを予測することはできず、今後急速に増加すると想定される個人ゲノムデータを解釈する上で重要な課題となっている。

2. 研究の目的

本研究は、ヒト遺伝性疾患の原因となるミスセンス変異に着目し、従来法では予測が困難とされる、ミスセンス変異が与える分子機能及び表現型への影響の予測を可能とする手法の開発を行うことを目的とする。

この目的のために、これまで開発してきた、ヒト疾患変異解析システム (Mutation@A Glance) をさらに発展させ、ミスセンス変異部位のタンパク質立体構造・分子機能に関するデータと、インターネット等に散在している、変異タンパク質の機能解析および疾患表現型に関する文献情報とを整理・統合した新しい情報基盤を構築する。このデータを解析することにより、個々のミスセンス変異と分子機能および疾患表現型への影響との関連性を明らかにする。

3. 研究の方法

(1) Mutation@A Glance への公的データの統合と機能拡張：疾患変異タンパク質の機能解析データについて、公表されている文献情報などから可能な限り収集し、リレーショナルデータベースに格納した上で、遺伝子及びタンパク質立体構造情報との紐付けを行った。

(2) ヒト遺伝性疾患の遺伝様式及び疾患発現メカニズムの分類

遺伝性疾患には大きく分けて2種類の遺伝様式が知られる(常染色体潜性(AR)、常染色体顕性(AD))。またADには、さらに3種類の疾患発現メカニズム(ハプロ不全(HI)、ドミナントネガティブ(DN)、機能獲得(GF))が知られているが、これまでまとめたデータベースは存在していない。そこで文献情報及び公的データベース(OMIM)からこれらの情報を抽出し、整理統合した。

(3) 統合データに基づく変異部位と分子機能及び疾患表現型の相関解析：(1)で統合化したデータベースを用いて、各疾患変異タンパク質について、個々のミスセンス変異がもたらす分子機能及び疾患表現型への影響を分類し、タンパク質立体構造との関連性の解析

を行った。

(4) 解析結果に基づく予測手法の開発と性能評価：(3)の解析結果に基づいた特徴量を用いて、分子機能及び疾患表現型への影響を予測するための手法の開発を行い、その性能の評価をした。

4. 研究成果

(1) Mutation@A Glance へのヒトバリエーションデータの整理・統合

インターネット上で公開されている、ヒト配列バリエーションデータベース dbSNP、疾患関連変異データベース (ClinVar)、がん体細胞変異データベース (COSMIC)、健康人エクソームバリエーションデータベース (ExAC) のデータをダウンロードし、リレーショナルデータベースに格納した。その結果、約8億件のヒトバリエーションについて、Mutation@A Glance 上で参照することができるようになった。

(2) ヒト遺伝性疾患の遺伝様式及び疾患表現型の分類

文献情報及び公的データベース(OMIM)の情報を参考にして、ミスセンス変異が疾患原因となっている1951遺伝子について、その遺伝様式及び顕性遺伝型疾患の疾患表現型についてマニュアルキュレーションによる分類を行った(表1)。その結果、1372のARの疾患及び1140のADの疾患について遺伝子の情報を整理した。さらに、ADの疾患のうち401(35.1%)については、HI、DN、GFのいずれかに分類することができた。

カテゴリ	遺伝子	疾患
常染色体潜性(AR)	1233	1372
常染色体顕性(AD)	845	1140
ハプロ不全(HI)	115	122
ドミナントネガティブ(DN)	132	143
機能獲得(GF)	118	136
不明・情報なし	578	738
合計	1951	2,366

表1 遺伝様式と疾患表現型の内訳

(3) 疾患遺伝様式及び疾患発現メカニズムとタンパク質立体構造との関連性

上記(2)で得た情報に基づき、各疾患の原因となるミスセンス変異がタンパク質立体構造上の関連性について調べた。その結果、ARとADの変異が起きやすい構造上の特徴に違いが見られた。すなわち、ARの変異はタンパク質立体構造の内部に多く、一方でADの変異は分子間相互作用面に多く存在していることがわかった。さらに解析を進めて、ADの3つの疾患メカニズムとの関連性を調べたところ、HIに比べて、DN及びGFは分子間相互作用面に有意に多く局在する傾向があることが明らかとなった。

次に、相互作用する相手分子に特徴があるかどうかを調べたところ、HI は DNA との相互作用部位に多く局在すること、DN は DNA との相互作用部位に加えてホモオリゴマーのインターフェースに多く局在することがわかった。一方で、GF は同一サブユニット内のドメイン間のインターフェースに有意に多いことを初めて明らかにした(図1)。このことは、変異による疾患発症メカニズムとタンパク質の高次構造とが密接に関係していることを示している。以上の結果を論文としてまとめて発表した(主な発表論文)。

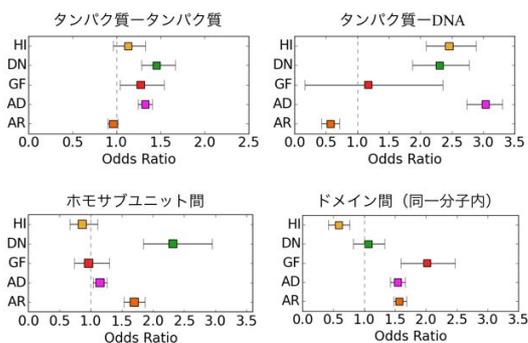


図1 遺伝様式 (AR, AD) および AD 疾患メカニズム (HI, DN, GF) と分子間 (内) 相互作用との関係 Odds Ratio が 1 よりも大きいと、変異部位とその相互作用との関連性が強いことを示す。

(4) 分子間相互作用情報を取り入れた疾患変異のメカニズム予測法とその評価

上記(3)で得られた結果は、ミスセンス変異部位における分子相互作用の情報が疾患メカニズムの予測に利用できることを示唆している。そこで、変異アミノ酸残基の立体構造的な特徴情報(モノマーでの溶媒接触度、複合体構造での溶媒接触度、リガンド分子までの距離など)を、従来の手法でも取り入れられているアミノ酸残基の配列保存性などの進化的情報、置換アミノ酸の物理化学スコアやヒト集団内アレル頻度などと組み合わせた特徴量を用いた機械学習による予測法を開発した。機械学習アルゴリズムには、比較的好く用いられ精度が高いランダムフォレストアルゴリズムを選択し、学習セットに対して、以下の3種類の分類器を構築した。

- 疾患変異と中立変異の2クラス分類
- 遺伝様式 AR, AD と中立変異の3クラス分類
- AR 及び AD の3つの疾患発症メカニズム (HI, DN, GF) の4クラス分類

それぞれの分類手法について評価を行なった。

従来の予測法と同様に、疾患変異かそうでないかを区別する。テストセットに対して本手法を適用してその精度を従来法と比較した。本手法の AUC (Area Under Curve) は 0.88 となった。この値は従来

法の中でも最も精度の高い、REVEL (Ioannidis et al., 2016) (AUC=0.90) と比べて遜色のない結果であることがわかった。

研究期間中に新しい手法が提案された (MAPPIN, Gosalia et al., 2017)。テストセットに対して本手法の結果と比較したところ、3クラス分類の精度において、MAPPIN の 0.56 に対して、本手法の Accuracy は 0.88 であり、従来法に比べて高い精度を持っていることがわかった。AR 及び HI, DN, GF の4クラスを分類する。本手法をテストセットに適用したところ、本手法の Accuracy は 0.71 となった。同様の手法はこれまで存在していないため本手法が世界初となる。ランダムに4クラスを分類した場合の精度は0.25であるのでそれに比べれば十分に高いが誤って分類される数も多いため、今後改良していく必要があると考えられる。

(5) Mutation@A Glance インターフェースの改良

本研究によって取り入れた情報について、より多くの研究者、特に日本の臨床系の研究者に利用しやすくするため、Mutation@A Glance のウェブインターフェースを刷新した(図2)。

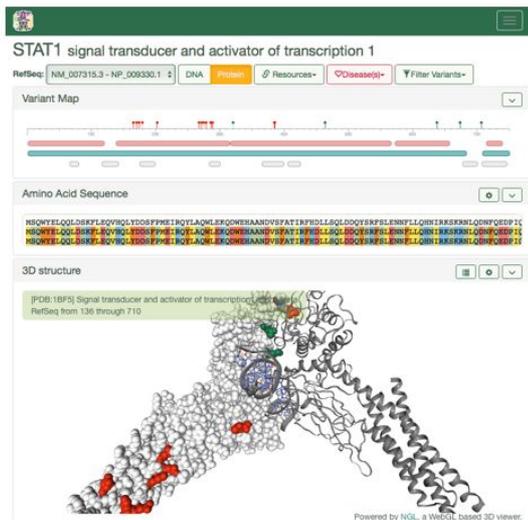


図2 新しい Mutation@A Glance のウェブインターフェース

STAT1 の変異を複合体立体構造上にマップ。疾患メカニズムごとにバリエーションの色を変えて表示させている (DN 変異 (緑)、GF 変異 (赤))。両変異は空間的な局在が異なっている。

主な改良点は以下のとおり:

- レイアウトをレスポンスにし、ユーザーのディスプレイの大きさに依存しない表示に変更、
- 同一遺伝子上の異なる疾患ごとにバリエーションを色分けして表示、

より大きな複合体が表示できる立体構造ビューワ(NGL viewer)を実装、マルチプルアラインメントを表示させ、アミノ酸サイトごとの保存性を視覚化した。

なお本ウェブサイトは、以下の URL で公開し誰でもアクセス可能である。

<http://mutation.nagahama-i-bio.ac.jp/>
公開と同時に、日本語のユーザーマニュアルを作成した(冊子版及びダウンロード版)。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 6 件)

Hijikata A., Tsuji T., Shionyu M., Shirai T., Decoding disease-causing mechanisms of missense mutations from supramolecular structures, Sci Rep, 査読有、7 巻、8541 頁、2017 年、DOI:10.1038/s41598-017-08902-1.

Miyaonori Y., Hijikata A., Nishino Y., Gohara M., Onoue Y., Kojima S., Kojima C., Shirai T., Kainosho M., Homma M., Structural and functional analysis of the C-terminal region of FliG, an essential motor component of Vibrio Na⁺-driven flagella, Structure, 査読有、25 巻、1540-1548.e3 頁、2017 年、DOI: 10.1016/j.str.2017.08.010.

Fujiki R., Hijikata A., Shirai T., Okada S., Kobayashi M., Ohara O. Molecular mechanism and structural basis of gain-of-function of STAT1 caused by pathogenic R274Q mutation, J Biol Chem, 査読有、292 巻、6240-6254 頁、2017 年、DOI: 10.1074/jbc.M116.753848.

Kawasaki Y., Oda H., Ito J., Niwa A., Tanaka T., Hijikata A., (他 13 名), Identification of a high-frequency somatic NLRC4 mutation as a cause of autoinflammation by pluripotent cell-based phenotype dissection, Arthritis Rheumatol, 査読有、69 巻、447-459 頁、2017 年、DOI: 10.1002/art.39960.

Tsujii-Hosokawa A., Takasawa K., Nomura R., Miyakawa Y., Numakura C., Hijikata A., Shirai T., Ogawa Y., Kashimada K., Morio T. Molecular mechanisms of insulin resistance in 2 cases of primary insulin receptor defect-associated diseases, Pediatr Diabetes, 査読有、18 巻、917-924 頁、2017 年、DOI: 10.1111/pedi.12508.

Hijikata A., Yura K., Ohara O., Go M. Structural and functional analysis of Barth syndrome-causing mutations and

alternative splicing in the tafazzin acyltransferase domain, Meta Gene, 査読有、4 巻、92-106 頁、2015 年、DOI: 10.1016/j.mgene.2015.04.001.

[学会発表](計 13 件、抜粋 7 件)

土方敦司、白井 剛、Mutation@A Glance: ヒトにおける意義不明バリエーションの解明を目指した統合解析ツール、生命科学合同年次大会 (ConBio2017)、2017 年

土方敦司、辻 敏之、塩生真史、白井 剛、Decoding disease-causing mechanisms of missense mutations from supramolecular structures, 第 6 回生命医薬情報連合大会、2017 年

土方敦司、辻 敏之、塩生真史、白井 剛、Towards predicting functional consequences of genetic variants in humans through supramolecular complex structures, 第 55 回日本生物物理学会年会、2017 年

土方敦司、辻 敏之、塩生真史、白井 剛、タンパク質超分子複合体構造から読み解くミスセンス変異と疾患表現型との関係、第 5 回 NGS 現場の会、2017 年

Hijikata A., Shionyu M., Shirai T., A new approach for protein-ligand binding predictions based on matching of vector-presented amino acid residues, 第 5 回生命医薬情報学連合大会、2016 年
土方敦司、塩生真史、白井 剛、アミノ酸残基相互作用ベクトルマッチに基づくタンパク質-低分子ドッキング法、第 16 回日本蛋白質科学会年会、2016 年

土方敦司、由良敬、小原 収、郷通子、Tafazzin トランスアシラーゼドメインにおける Barth 症候群関連変異と選択的スプライシングの構造及び機能に関する影響、第 38 回日本分子生物学会、2015 年

[その他]

ホームページ等

Mutation@A Glance:

<http://mutation.nagahama-i-bio.ac.jp/>

研究成果プレスリリース

「ヒト疾患を引き起こすミスセンス変異の発症メカニズムの一端を解明」
<http://www.nagahama-i-bio.ac.jp/research/>
ヒト疾患を引き起こすミスセンス変異の発症メカ/

6. 研究組織

(1) 研究代表者

土方 敦司 (HIJIKATA, Atsushi)

長浜バイオ大学・バイオサイエンス学部・プロジェクト特任講師

研究者番号: 80415273