

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 11 日現在

機関番号：13802

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24790336

研究課題名(和文) イオンチャネル関連疾患における遺伝型-表現型予測法の開発に関する研究

研究課題名(英文) Development of genotype-phenotype prediction method for channelopathy

研究代表者

吉田 秀一 (YOSHIDA, SHUICHI)

浜松医科大学・医学部・助教

研究者番号：10580574

交付決定額(研究期間全体)：(直接経費) 1,500,000円、(間接経費) 450,000円

研究成果の概要(和文)：本研究課題では、ともに電位依存性Na⁺チャンネル(Nav1.1チャンネル)をコードするSCN1A遺伝子変異に起因する二疾患、SCN1A関連てんかん性脳症と全般てんかん熱性けいれんプラスにおける遺伝子型-表現型予測モデルの構築を通し、遺伝性疾患の遺伝子型-表現型予測へ情報学的手法(機械学習法・データマイニング)を応用することの有用性を検討した。その結果、変異に伴って生じたアミノ酸の物理化学的特性変化を予測因子とした表現型予測モデル、特に疎水性、等電荷点、極性を予測因子に含むモデルで高い分類精度が得られ、情報学的手法を活用した遺伝子型-表現型予測構築手法の有用性が示された。

研究成果の概要(英文)：We described a computational model for phenotype prediction of SCN1A-related epilepsies involving the Machine-learning approaches (such as Support vector machines, Random Forest) that were trained some predicting factors obtained by data-mining analysis among the databases. Our models show high accuracy, especially, the prediction models include IE, P and HP of physicochemical property as predicting factor. Our findings indicate the possibility of phenotype prediction for entirely new missense mutations by an application of the physico-chemical properties of amino acid residues.

研究分野：医歯薬学

科研費の分科・細目：基礎医学・人類遺伝学

キーワード：遺伝子型-表現型 機械学習法 SCN1A

1. 研究開始当初の背景

電位依存性 Na^+ チャネル ($\text{Na}_v1.1$) をコードする *SCN1A* (OMIM *182389) は、熱性けいれん (FS)、全般てんかん熱性けいれんプラス (GEFS+, OMIM#604223)、乳児重症ミオクロニーてんかん (OMIM#609208) に代表される *SCN1A* 関連乳児てんかん性脳症 (SRIIE) の責任遺伝子である。同一の遺伝子 *SCN1A* の変異に起因するにも関わらず、予後良好な FS から進行的で重篤な SRIIE まで幅広い臨床スペクトラムを呈す。加えて、初発時期の臨床症状は極めて酷似しており、臨床症状からの早期診断は困難を極める。したがって、遺伝子診断が期待される一方、遺伝子型-表現型の相関は不明確な部分も多いことや新たに見つかる変異の多くは *de novo* であることなどから未だ多くの課題を有している。

様々な分野の種々の問題において、データマイニングや機械学習法に代表される情報学的手法の応用は、それまで見つけることができなかつた新たな知見の発見に大きく貢献している。バイオ・ゲノム研究においても、個々の研究により生み出され、データベースとして蓄積されている知見をデータマイニングにより包括的に解析する取り組みがなされ、成果を上げつつある。

2. 研究の目的

上述した背景より、本研究課題では、データマイニングと機械学習法に基づく *SCN1A* 関連てんかん症候群の遺伝子型-表現型の相関の解明と表現型予測モデルの構築を通し、遺伝性疾患における遺伝子型-表現型予測モデル構築への情報学的手法の有用性を検証した。

3. 研究の方法

SCN1A データセットは、*SCN1A* Variant Database (Claes et al. *Hum Mutat.* 2009) に登録されている *SCN1A* 変異のうち、アミノ酸の置換を伴うミスセンス変異を抽出・取得した。また、予測因子は、変異があった箇所のアミノ酸の進化的配列保存性を評価する指標として (1) PSIC (position- specific independent counts) (Sunyaev et al. *Protein Eng.* 1999; Adzhubei et al *Nat Med.* 2010) を用いるとともに、AA index database (Kawashima et al. *Nucleic Acids Res.* 2008) に登録されているアミノ酸の物理化学的特性指標を対象として、(2) 変異型と野生型

$\text{Na}_v1.1$ チャネルでの統計的に有意な変化を示す物理化学的特性指標をデータマイニングにより選定・使用した。

次いで、*SCN1A* ミスセンス変異データセットに対し、(1)(2)の2系統を予測因子とした機械学習法により、*SCN1A* 関連てんかん症候群における遺伝子型-表現型相関解析と予測モデルの構築、表現型との関連が強い予測モデルの絞り込みを行った。このとき、構築した表現型予測モデルの分類精度は、sensitivity, specificity, ROC (receiver operating characteristic curve) 解析などにより多角的に評価し、最も分類精度が高くなるよう各種パラメータを調整した。また、機械学習データセットを用いたクロスバリデーションによる評価と *SCN1A* と相同な *SCN2A* ($\text{Na}_v1.2$ 、てんかん責任遺伝子の1つ)のミスセンス変異から構築したデータセットによる評価を行った。(2)の変異によるアミノ酸の物理化学的特性変化を基にした遺伝子型-表現型予測モデル構築の流れを図1に示した。(1)を予測因子とした遺伝子型-表現型予測モデル構築の流れも予測因子以外の部分は、同様であるため、流れ図は省略する。

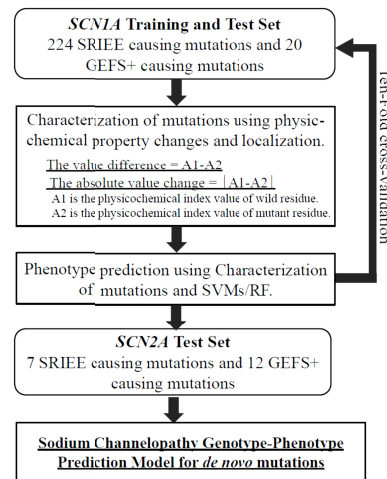


図1. 変異によるアミノ酸の物理化学的特性変化を基にした遺伝子型-表現型予測モデル構築の流れ図

4. 研究成果

(1) PSIC スコアの変化に基づく遺伝子型-表現型予測

SCN1A 関連てんかん症候群における2つの表現型 (SRIIE と GEFS+) において、変異型と野生型 $\text{Na}_v1.1$ チャネルとのアミノ酸の進化的配列保存性の差異 (PSIC score difference) の分布を図2に示す。

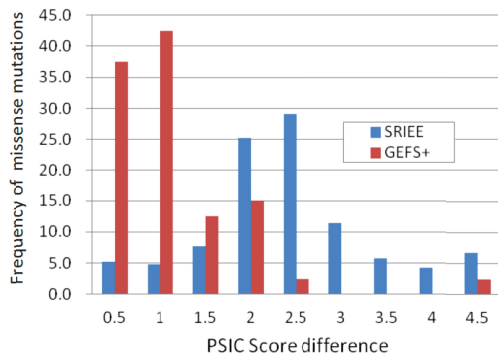


図 2. *SCN1A* 関連てんかん症候群における 2 つの表現型 (SRIIEE と GEFS+) と変異型-野生型 $Na_v1.1$ チャンネル間の PSIC score difference 分布

GEFS+群と SRIIEE 群のスコア分布において、明確な閾値は観察されなかったが、2~2.5 近傍にあると仮定し、ROC 解析により最適解を求めた。この時、Accuracy が最も高くなる PSIC score difference をカットオフ値とした選択すると、sensitivity 0.74、specificity 0.42 の分類精度が得られた (図 3)。

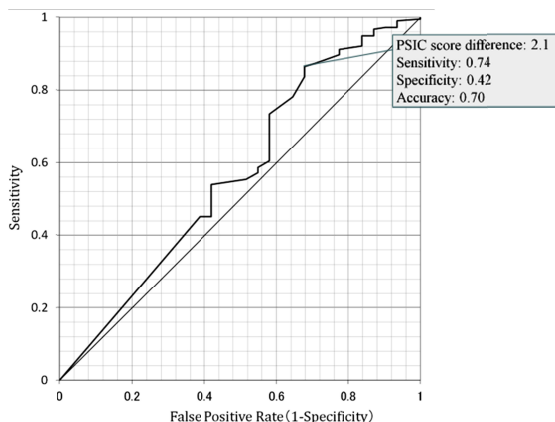


図 3. PSIC スコアをカットオフ値として可変した際の ROC 解析

(2) 物理化学的特性変化に基づく遺伝子型-表現型予測

変異型と野生型 $Na_v1.1$ チャンネルでのアミノ酸の物理化学的特性指標の変化の大きさを基に構築した表現型予測モデルの各種分類精度を図 4 にヒートマップ表示した (Accuracy 上位 5) サポートベクタマシン (SVM)、ランダムフォレスト (RF)、決定木などの機械学習法のうち、本研究課題には SVM に基づく方法が最適であった。また、野生型と変異型 $Nav1.1$ チャンネル間でのアミノ酸の物理化学

的特性変化からデータマイニングにより得た予測因子のうち、特に疎水性、等電荷点、極性を予測因子に含むモデルで高い表現型分類精度が得られた。一方で、機会学習法の宿命ではあるが、クロスバリデーションによる評価で高い分類精度を示すモデルが得られても、異なるデータセットで評価した際に、必ずしも再現しないことが、本課題の結果でも示された (図 4 . B)。

今後は、上記課題の解決に向けて、(1) PSIC スコアによるアミノ酸の進化的配列保存性因子および(2) アミノ酸の物理化学的特性因子、あるいは新たな予測因子を組み合わせる論理和予測モデルへの拡張と汎用性を高める工夫を施すことで遺伝性疾患に全般に対して適用可能な遺伝子型-表現型予測手法へと展開させていく予定である。

A)	Model	Acc	Sen	Spe	PPV	NPV
	P+PR+HP+IE+Localization(SVM)	0.88	0.94	0.74	0.91	0.80
	P+HP+IE(SVM)	0.85	0.96	0.71	0.86	0.80
	P+PR+HP+IE(SVM)	0.84	0.95	0.57	0.86	0.78
	P+HP+PR(SVM)	0.84	0.95	0.57	0.86	0.78
	P+PR+HP+IE+Localization(RF)	0.89	0.95	0.64	0.92	0.76

B)	Model	Acc	Sen	Spe	PPV	NPV
	P+PR+HP+IE+Localization(SVM)	0.75	0.74	0.54	0.71	0.60
	P+HP+IE(SVM)	0.72	0.76	0.51	0.66	0.60
	P+PR+HP+IE(SVM)	0.71	0.75	0.37	0.66	0.58
	P+HP+PR(SVM)	0.71	0.75	0.37	0.66	0.58
	P+PR+HP+IE+Localization(RF)	0.76	0.75	0.44	0.72	0.56

図 4. 各表現型予測モデルの分類精度ヒートマップ表示。A) ブートストラップ法による評価、B) *SCN2A* テストセットに対する評価。

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 5 件)

吉田秀一, 兼子 直: てんかんとイオンチャンネル, 医学のあゆみ (特集: イオンチャンネル病のすべて, 245:740-744, 2013.

Yamada J, Zhu G, Okada M, Hirose S, Yoshida S, Shiba Y, Migita K, Mori F, Sugawara T, Chen L, Liu F, Yoshida S, Ueno S, Kaneko S: A novel prophylactic effect of furosemide treatment on autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE). *Epilepsy Res*, 107:127-137, 2013, 査読有.

doi: 10.1016/j.eplepsyres.2013.08.018.
Saruwatari J, Deguchi M, Yoshimori Y, Noai M, Yoshida S, Ogusu N, Oniki K, Yoshida S, Yasui-Furukori N, Kaneko S, Ishitsu T, Nakagawa K: Superoxide dismutase 2 Val16Ala polymorphism is a risk factor for the valproic acid-related elevation of serum aminotransferases. *Epilepsy Res*, 99:183-186, 2012, 査読有.

doi: 10.1016/j.eplepsyres.2011.10.033.
Tsuchimine S, Yasui-Furukori N, Sasaki K, Kaneda A, Sugawara N, Yoshida S, Kaneko S: Association between the dopamine D2 receptor (DRD2) polymorphism and the personality traits of healthy Japanese participants. *Prog Neuropsychopharmacol Biol Psychiatry*, 38:190-193, 2012, 査読有.
doi: 10.1016/j.pnpbp.2012.03.008.

〔学会発表〕(計4件)

Yoshida S, Kaneko S, Nishio T: A Computational Approach for Predicting *SCN1A*-related Epilepsy Phenotypes based on Physico-chemical Property Changes, The 24th International Conference on Genome Informatics, Singapore (December 2013).

吉田秀一: てんかんの遺伝子診断 up to date (シンポジウム 個別化治療 up to date: 遺伝子診断の現状と治療への応用), 第46回日本てんかん学会, 東京(2012年10月11日)

菅原貴征, 吉田秀一, 和田一丸, 廣瀬伸一, 兼子直: DNAチップを用いた様々なてんかん類型における遺伝子変異の同定, 第46回日本てんかん学会, 東京(2012年10月11日)

Yoshida S, Shimizu T, Kaneko S, Nishio T: Predicting *SCN1A*-related Epilepsy Phenotypes Based on the Functional Effect in *SCN1A* with the Amino-Acid Substitution, Joint Conference on Informatics in Biology, Medicine and Pharmacology, Tokyo (October 2012)

〔図書〕(計3件)

吉田秀一, 兼子直: てんかんとイオンチャンネル, 別冊 医学のあゆみ (特集: イオンチャンネル病のすべて), p46-50, 医歯薬出版, 2014.

兼子直, 吉田秀一: 第1章 総論 C. てんかんとチャンネル・遺伝子異常/第4章 治療 4. 薬剤感受性と遺伝子 polymorphism, シリーズ アクチュアル脳・神経疾患の臨床「てんかんテキスト New Version」, 中山書店, p19-24/ p206-212, 2012.

吉田秀一, 兼子直: I. てんかんとは/ . てんかんと遺伝, てんかん教室第3版(兼子直編), 新興医学出版, p1-13/ p217-221, 2012.

6. 研究組織

(1) 研究代表者

吉田 秀一 (YOSHIDA, Shuichi)
浜松医科大学・医学部・助教
研究者番号: 10580574

(2) 研究分担者

該当なし

(3) 連携研究者

西尾 卓広 (NISHIO, Takuhiro)
浜松医科大学・医学部・准教授
研究者番号: 90172626

兼子 直 (KANEKO, Sunao)
湊病院・北東北てんかんセンター・センター長
研究者番号: 40106852