A Computational Approach for Predicting *SCN1A*-related Epilepsy Phenotypes based on Physico-chemical Property Changes

Shuichi Yoshida [1,3] yoshida@hama-med.ac.jp
Sunao Kaneko [2,3] sk@epigene.org
Takuhiro Nishio[1] nishio@hama-med.ac.jp

[1] Dept. of Integrated Human Science, Hamamatsu University School of Medicine, Hamamatsu 4313192, Japan
[2] North Tohoku Epilepsy Center, Minato Hospital, Hachinohe 0310813, Japan
[3] The Epilepsy Genetic Study Group Japan

Keywords: *SCN1A*, epilepsy, missense mutation, genotype-phenotype correlation, phenotype prediction

1 Introduction

Mutations in *SCN1A* encoding the neuronal voltage-gated sodium channel α1-subunit are most common genetic cause of inherited and idiopathic epilepsy. *SCN1A* mutations were initially identified in two clearly distinct epilepsy syndromes; the generalized epilepsy febrile seizures plus (GEFS+) which is essentially a benign epileptic syndrome and the *SCN1A*-related infantile epileptic encephalopathy (SRIEE) which is a severe epileptic syndrome with delayed psychomotor development and refractoriness to drug therapy. More than 650 mutations of *SCN1A* have been identified in patients with SRIEE phenotypes, and 20 mutations in patients with GEFS+. Approximately 50% of these mutations are truncation mutations and result in the SRIEE phenotypes, clearly demonstrating haploinsufficiency of *SCN1A*. However, the reason why similar missense mutations in *SCN1A* resulting in different phenotypes has not been fully clarified yet. The genetic test for the *SCN1A*-related epilepsy phenotypes has extremely important implication for research and clinical practice. The majority of SRIEE patients, however, carry *de novo SCN1A* mutations. Thus, it is significant to predict the *SCN1A*-related epilepsy phenotypes resulting from *de novo* missense mutations. Previously, we reported the effect of sites of *SCN1A* missense mutation on the epilepsy phenotype severity and also showed a significant association between *SCN1A*-related epilepsy phenotypes and the change of the physico-chemical property by amino acid substitutions [1,2]. In this study, we describe a computational model for phenotype prediction of *SCN1A*-related epilepsies involving the Support vector machines (SVMs) or Random Forest (RF) algorithms that were trained the significant value difference of physico-chemical properties for *SCN1A* missence mutations.

2 Method

2.1 Data sets

We obtained a total of 249 *SCN1A* missense mutations from the *SCN1A* Variant Database [3]. Then, we classified these *SCN1A* missense mutations into the *SCN1A*-related epilepsy groups (SRIEE phenotype, GEFS+ phenotype), then we extracted 5 mutations affected non-epileptic syndorome.

Table 1. A list of the predictors for *SCN1A*-related epilepsy phenotype prediction

Predictors	Effect site	Genotype-phenotype correlation	Ref.
Localization	Whole	The missense mutations in the pore regions were associated with SIREE phenotype than mutations in other regions.	[1]
Absolute value changes of IE	Pore	The absolute value change of IE were significant difference each epilepsy phenotypic groups	[2]
Value difference of HP	S1–S4	Compare with SIREE phenotype, the GEFS+ phenotype has significant low value difference of P.	[2]
Value difference of PR	S1–S4	Compare with SIREE phenotype, the GEFS+ phenotype has significant low value difference of PR.	[2]
Value difference of P	Whole	Compare with SIREE phenotype, the GEFS+ phenotype has significant high value difference of P.	[2]
	S1-S4	Compare with IM and SIREE phenotypes, the GEFS+ phenotype has significant high value difference of P.	

2.2 Predictors for SCN1A-related epilepsy phenotypes

The predictors were listed in Table 1. Previously, we reported that missense mutations in the pore regions of *SCN1A* were associated with more severe phenotype than mutations in other regions [2]. And we also described that several significant associations were shown between the clinical properties and the value changes in the physicochemical property of amino acids with substitution in *SCN1A*, especially the absolute value changes of isoelectric point (IE) in the pore region and the value difference of hydrophobicity indices (hydropathy (HP), polarity (P) and polar requirement (PR)) in the S1–S4 region of *SCN1A* [3]. We used these physic-chemical property changes and mutation localization (pore/non-pore regions) as the predictors for *SCN1A* -related epilepsy phenotype prediction.

2.3 Support Vector Machine and Random Forest Classifiers

SVMs are universal classifiers that learn a variety of data distributions from training samples, and as such, are applicable to the classification and regression tasks. SVMs use kernel functions to map original data to future space of higher dimensions and locate an optimal separating hyperplane there. RF is a classifier consisting of an ensemble of tree-structured classifier. We used R language (R 3.0.1) implementation of the randomForest package for RF and e1076 package for SVMs. We estimated the performance of phenotype prediction using RF and SVMs that were trained by the predictors. Ten-fold cross-validation has been used to test the performance of phenotype prediction, and we compared the performance of prediction models that trained different combination of the predictors and classifiers. Various performance measures are used including accuracy, sensitivity, specificity, Positive predictive value and Negative predictive value.

3 Results and Discussions

The best performance was “P+PR+HP+IE+Localization” model using radial basic function with gamma= 0.12 and cost=1 among the tested SVM classifer (accuracy 88.0 %, sensitivity 94.1 %, specificity 74.0 %). Other models also show high accuracy, especially, the prediction models include IE, P and HP of physicochemical property as predicting factor. P and HP are the indices of hydrophobicity of the residue. The larger value in P is the higher the hydrophilicity of the side chains, and there is opposite association between the value of HP and hydrophilicity. Hence positive value difference in HP and negative value difference in P all meant the decrease of hydrophobicity. Therefore those changes showed that amino acid replacements decreasing hydrophobicity were most associated between benign and severer phenotype. IE is considered to constitute important physico-chemical properties of amino acid residues, and was presumed to have significant effects on the amino acid compositions of proteins in the evolutional stage. Large changes in them may result in severe state change of the surface of the pore, and may markedly affect the function of the ion channel. This hypothesis also suggests the importance of IE in the functions of some proteins, and explains why many proteins have the tendency to minimize changes in IE during their evolutionary stage. Our findings indicate the possibility of phenotype prediction for entirely new missense mutations by an application of the physico-chemical properties of AA residues. Although this retrospective study has some limitations, we think that this research could provide a new strategy for genetic diagnosis.

Acknowledgements

This research was supported by JSPS KAKENHI Grant numbers 24790336, 22890079.

References

[1] Kanai K, *et al.*, Effect of localization of missense mutations in SCN1A on epilepsy phenotype severity. ***Neurology***, 63:329-334, 2004.
[2] Kanai K, Yoshida S, *et al.*, Physico-chemical property changes of amino acid residue accompanied with missense mutations in SCN1A affect the epilepsy phenotype severity. ***J Med Genet***, 46:671-679, 2009.
[3] Claes LR, *et al.*,The SCN1A variant database: a novel research and diagnostic tool. ***Hum Mutat,*** 30:904-920, 2010.