Comprehensive Detection of TM Protein Genes which might be Associated with Febrile Seizure

Yasuhito Inoue

Kazuaki Kanai

Satoshi Mizuta

Motohiro Okada

Sunao Kaneko

Toshio Shimizu

gs04904@si.hirosaki-u.ac.jp

VZR03355@nifty.ne.jp

slmizu@si.hirosaki-u.ac.jp

okadamot@cc.hirosaki-u.ac.jp

ks@cc.hirosaki-u.ac.jp

slsimi@si.hirosaki-u.ac.jp

1 Department of Electronic and Information System Engineering, Faculty of Science and Technology, Hirosaki University, Hirosaki 036-8561, Japan

2 Department of Neurology, Graduate School of Medicine, Chiba University, Chiba 260-8677, Japan

3 Department of Neuropsychiatry, School of Medicine, Hirosaki University, Hirosaki 036-8562, Japan

Keywords: febrile seizure, transmembrane protein, transmembrane topology, ion channel

1 Introduction

Febrile seizure (FS) is a common neurological disorder in childhood [4]. Although the loci on the human chromosomes where the responsible genes are located, 2q23-q24, 5q14-q21, 6q22-q24, 8q13-q21, and 19p13.3, are acquainted, the responsible genes have not been identified yet. It is expected that we can get the candidates of the responsible genes of FS by extracting TM protein genes such as voltage gated ion channels and ligand gated ion channels on the above mentioned chromosome regions, because it has been reported that the responsible genes of generalized epilepsy with FS plus (GEFS+), which is closely related to FS, are some ion channels.

In this study, at first we selected the predicted TM protein sequences from ORFs located on the target chromosome regions, and next we extracted the sequences which were related to ion channel by homology search against SWISS-PROT [2]. Besides, based on the fact that TM proteins can be functionally classified and identified by their TM topology [5,7], we retrieved the candidate TM proteins according to the predicted number of TM segments from the remaining sequences which were not identified by homology search.

2 Materials and Methods

2.1 Dataset

We used 28,104 human open reading frames (ORFs; build 35 [8]) registered in GenBank for this study. We focused on five chromosomal regions, 2q23-q24, 5q14-q21, 6q22-q24, 8q13-q21 and 19p13.3, where the responsible genes of FS have been reported to exist. There were 63, 63, 126, 100 and 234 ORFs in the chromosomal regions, respectively (Table 1). Out of their protein sequences translated from their ORFs, we segregated TM protein sequences and predicted their TM topologies according to the following procedure: (1) predict and remove signal peptide region using SignalP 2.0 [3]; (2) predict TM protein sequence candidates using SOSUI 2.0; and (3) predict TM topology using HMMTOP 2.0, TMHMM 2.0, and ConPred II [1]. In order to increase the capture rate of 7-tms protein sequences, we employed a combination of HMMTOP 2.0 and TMHMM 2.0 (‘HMMTOP+TMHMM’) [6] at the step 3. In this procedure, the sequences predicted as other TM proteins than 7-tms ones by HMMTOP 2.0 were processed by TMHMM 2.0, and the sequences predicted as 7-tms proteins by TMHMM 2.0 were also treated as GPCR candidates for further uses. When TM protein sequences were not predicted as 7-tms proteins by both of the two methods, TM topology information predicted by ConPred II was utilized.
2.2 Detection of TM protein genes associated with febrile seizure

The predicted TM protein sequences were annotated with SWISS-PROT (Release 48.0) by BLAST homology search. The BLAST search was carried out with the default settings (first gap penalty, -11; additional gap penalty, -1; substitution matrix, BLOSUM 62) and the threshold E-value of 1e-3. We calculate the global sequence identities between a query sequence and the matched SWISS-PROT sequence using ALIGN with the default settings, except for the substitution matrix (BLOSUM 62 used). When TM protein sequences were not matched by BLAST, we classified them according to the number of TM segments as “sodium channel (1-tms and 24-tms),” “acetylcholine and GABA\_ receptors (4-tms),” “potassium channel (2-tms and 6-tms),” “extracellular calcium-sensing receptor (7-tms) and syntaxin (1-tms)” and “metabotropic glutamate receptor (7-tms),” in which we permitted the margin of ±1-tms, taking into account a possible error in the TM topology prediction. Moreover, when TM protein sequences were predicted as ≥16-tms by ConPred II, they were classified as “sodium channel”.

3 Results and Discussion

Table 1 lists the number of detected TM protein genes and their predicted number of TM segments in the chromosomal regions, 2q23-q24, 5q14-q21, 6q22-q24, 8q13-q21, and 19p13.3. There were 63, 63, 126, 100, and 234 ORFs in the regions respectively, and 27, 11, 40, 13, and 59 sequences, respectively, were predicted as TM proteins. Out of them, 24, 8, 38, 11, and 53 sequences were functionally identified by BLAST search, respectively, and the functions of the remaining sequences were guessed from the predicted number of TM segments according to the classification procedure. For example, 14, 4, and 18 sequences were classified as candidates of “sodium channel”, “acetylcholine and GABA\_ receptors”, and “potassium channel” genes, respectively. We also detected the “calcium sensing receptor and syntaxin” and “metabotropic glutamate receptors” genes in the chromosomal regions.

Table 1: Numbers of ORFs and TM protein genes in the human loci focused.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Size (bp)</th>
<th>ORFs</th>
<th>TM protein genes</th>
<th>BLAST hit(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2q23-q24</td>
<td>16,981,699</td>
<td>63</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>5q14-q21</td>
<td>22,741,715</td>
<td>63</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>6q22-q24</td>
<td>35,980,095</td>
<td>126</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>8q13-q21</td>
<td>12,062,687</td>
<td>100</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>19p13.3</td>
<td>7,467,154</td>
<td>234</td>
<td>59</td>
<td>53</td>
</tr>
</tbody>
</table>

\(^a\)TM protein genes of which function was annotated by the BLAST homology search against SWISS-PROT 48 with E-value < 1e-3.

References