IDENTIFICATION OF FOUR GENES ON HUMAN CHROMOSOME 3 HOMOLOGOUS TO THE KNOWN GENES ON OTHER CHROMOSOMES BY IN SILICO ANALYSIS

* Rakhmanaliev E.R., Klimov E.A.

Vavilov Institute of General Genetics, RAS, Moscow, Russia, e-mail: elian-rr@newmail.ru

Key words: NotI-clones, NotI-STSs, gene-markers, gene homologs

Resume

Motivation: In this work, we analyzed the structure of genomic sequences of human chromosome 3, marked by NotI-STSs. The NotI-STSs had homology with genes localized on other chromosomes that allowed us to suppose the presence of homologous sequences for these genes on chromosome 3.

Results: Four nucleotide sequences of human chromosome 3, marked by NotI-STSs (NB1-100, NL3-004, NLM-246 and NRL-402) which have high homology with genes, earlier localized in other genomic regions, were characterized in silico. We have shown that RINZF gene earlier localized on 8q13-q21.1 has the full-length copy on human chromosome 3. For three NotI-STSs (NL3-004, NLM-246 and NRL-402), which were markers of genes LOC132160, ATP11B and ITGA9 on chromosome 3, genes KIAA1157 (12q14.1), HSA9947 (1p36) and SCYA5 (17q11.2-q12) were determined as having homology to the NotI-STS sequences, respectively. Similarity of regulatory regions for three pairs of genes, (LOC132160 / KIAA1157, ATP11B / HSA9947 and ITGA9 / SCYA5), marked by the NotI-STSs, was shown.

Introduction

The endonuclease NotI restriction sites (5'-GCGGCCGC-3') are located in CpG-island, which are associated with 5'-UTR of genes. Therefore, STSs (sequenced tagged site), created on base NotI-sites might be considered as the universal markers of genes. Library of NotI-clones of human chromosomes 3 was created earlier (Zabarovsky et al., 1990; 1996). In our laboratory, 113 NotI-STSs for 84 NotI-clones were created (Sulimova et al., 1999). We have determined the physical localization of 30 NotI-STSs by radiation hybrid mapping method and constructed NotI-map of human chromosome 3, including 60 NotI-STSs (data in press). The search of homologies for the localized NotI-clone sequences with corresponding nucleotide sequences, presented in public databases (GenBank, EMBL and TIGR) by the program BLAST has revealed the high level of associations (91,7%) of NotI-STSs with human genes or ESTs. The localization of the majority NotI-STSs and corresponding homologous genes and ESTs completely coincided. However, four genes significant homology with NotI-STSs, were earlier localized on other human chromosomes. These results allowed us to suppose the presence of homologs (or pseudogenes) of these genes on human chromosome 3. To confirm these suggestions, we performed in silico analysis of genomic sequences of human chromosome 3, adjacent to sites of NotI-STSs localization.

Methods

The homologies were searched by the BLAST-program provided by NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). Exon-intron structure of novel genes was created by BLASTN (NCBI) and GENSCAN (http://genes.mit.edu/GENSCAN.html) programs. Promoter regions were identified using PromoterInspector (http://genomatix.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl) and Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html) programs. We also used programs at GeneBee server (http://genebee.msu.ru/) for the search of amino acid homologies and construction of the full local similarity maps for hypothetical proteins. Information concerning proteins, encoded by novel genes, was obtained from OMIM database (http://www.ncbi.nlm.nih.gov/OMIM/).

Results and Discussion

Screening of human genomic sequences for homologous sequences to earlier RH-mapped NotI-STSs was revealed. The data allowed us to identify four nucleotide sequences on human chromosome 3, homologous to the genes previously localized in other genome regions (Table). It allowed us to suggest presence of four earlier non-described gene-homologs (or pseudogenes) on chromosome 3.

The NotI-STS NB1-100 has 99% homology with mRNA RINZF gene, encoding protein with yet unknown function, containing “zinc fingers” domain. The RINZF gene was earlier localized on chromosome 8. However, marker NB1-100 also has 99% homology with fragment of clone AC009812, localized on chromosome 3. The comparative analysis of RINZF gene and clone AC009812 has revealed that clone AC009812 includes nucleotide sequence, identical to the
sequence of RINZF gene. The exon-intron structure of a copy of the RINZF gene constructed by BLAST and GENSCAN programs was identical (in the number, length and nucleotide sequence of exons and introns) the exon-intron RINZF gene structure, predicted by computer analysis performed with the BLAST program and presented in database MapViewer. The gene identical to RINZF gene has all necessary regulatory elements present in any gene: TATA-box, promoter region, poly(A)-sites and splicing sites. Therefore we can suggest that full-length copies of RINZF gene are present on both human chromosome 3 and 8. This is not an artifact, since localization of NotI-STS NB1-100 and localization of homologous clone AC005812 on chromosome 3 completely coincided.

<table>
<thead>
<tr>
<th>NotI-STS</th>
<th>NotI-STS localization</th>
<th>Gene on chromosome 3 marked by NotI-STS</th>
<th>Detected gene-homolog</th>
<th>Localization of gene-homolog</th>
</tr>
</thead>
<tbody>
<tr>
<td>NB1-100</td>
<td>133.5 3p21.33</td>
<td>Human hypothetical gene LOC132160</td>
<td>Human zinc finger protein RINZF (RINZF)</td>
<td>8q13-q21.1</td>
</tr>
<tr>
<td>NL3-004</td>
<td>189.0 3p21.1</td>
<td>Human gene LOC132160</td>
<td>Human gene for KIAA1157 protein (KIAA1157)</td>
<td>12q14.1</td>
</tr>
<tr>
<td>NLM-246</td>
<td>687.0 3q27.2</td>
<td>ATPase, Class VI type 11B (ATP11B)</td>
<td>Human putative ATPase gene (HS9947)</td>
<td>1p36</td>
</tr>
<tr>
<td>NRL-402</td>
<td>127.1 3p21.3</td>
<td>Human integrin, alpha 9 gene (ITG49)</td>
<td>Human small inducible cytokine A5 (RANTES) gene (SCYA5)</td>
<td>17q11.2-q12</td>
</tr>
</tbody>
</table>

Footnote: *We detected full-length copy of RINZF gene from chromosome 8, which yet untitled.

The hypothetical gene LOC132160 (4.5 kb in length) is present on chromosome 3, near the site of localization of NotI-STS NL3-004. Marker NL3-004 has homology with a fragment of the hypothetical gene KIAA1157 (290 kb) (chromosome 12). These genes (LOC132160 and KIAA1157) have different exon-intron structure and cDNA sequences. In spite of these differences, genes LOC132160 and KIAA1157 encoded similar proteins (identity 55%). Using the protein-protein BLAST, we revealed the nearest homolog for these proteins – protein phosphatase 2C (identity 85% for LOC132160 and 54% for KIAA1157). Probably, two considered genes encoded proteins, which referred to the class of proteins PP2C, from serin/threonin phosphatases family (Marley et al., 1998). The products of these genes can be involved in the same metabolic way. Therefore similarity of promoter regions of these genes, might be due to probable similar regulation at the stages of initiations of transcriptions.

Analogous results were received for NotI-STS NLM-246, localized in 5’UTR of ATPase gene, Class VI, Type 11B (ATP11B) (chromosome 3). The gene has also homology with the hypothetical gene HSA9947 (chromosome 1). Between the genes, no reliable homology was detected on nucleotide level. However, both genes encoded proteins (ATPases), from one protein family.

The NotI-STS NRL-402 is a marker of the gene integrin (ITG49), localized on chromosome 3, and has homology with the gene chemokine (SCYA5) localized on chromosome 17. Between these genes, no homology was observed on nucleotide or protein levels. Homology of NRL-402 with gene SCYA5 is explained by the presence of conservative regions in 5’UTR of the genes. Therefore, we characterized in silico four nucleotide sequences of human chromosome 3, marked by NotI-STSs (NB1-100, NL3-004, NLM-246 and NRL-402) which have high homology with genes, earlier localized in other genomic regions. We have shown, that earlier localized on 8q13-q21.1 gene RINZF has the full-length copy on human chromosome 3. For three NotI-STSs (NL3-004, NLM-246 and NRL-402), which were marked of genes, no reliable homology was detected on nucleotide or protein levels. However, both genes encoded proteins, which referred to the class of proteins PP2C, from serin/threonin phosphatases family (Marley et al., 1998). The products of these genes can be involved in the same metabolic way. Therefore similarity of promoter regions of these genes, might be due to probable similar regulation at the stages of initiations of transcriptions.

Analogous results were received for NotI-STS NLM-246, localized in 5’UTR of ATPase gene, Class VI, Type 11B (ATP11B) (chromosome 3). The gene has also homology with the hypothetical gene HSA9947 (chromosome 1). Between the genes, no reliable homology was detected on nucleotide level. However, both genes encoded proteins (ATPases), from one protein family.

The NotI-STS NRL-402 is a marker of the gene integrin (ITG49), localized on chromosome 3, and has homology with the gene chemokine (SCYA5) localized on chromosome 17. Between these genes, no homology was observed on nucleotide or protein levels. Homology of NRL-402 with gene SCYA5 is explained by the presence of conservative regions in 5’UTR of the genes. Therefore, we characterized in silico four nucleotide sequences of human chromosome 3, marked by NotI-STSs (NB1-100, NL3-004, NLM-246 and NRL-402) which have high homology with genes, earlier localized in other genomic regions. We have shown, that earlier localized on 8q13-q21.1 gene RINZF has the full-length copy on human chromosome 3. For three NotI-STSs (NL3-004, NLM-246 and NRL-402), which were marked of genes, no reliable homology was detected on nucleotide or protein levels. However, both genes encoded proteins, which referred to the class of proteins PP2C, from serin/threonin phosphatases family (Marley et al., 1998). The products of these genes can be involved in the same metabolic way. Therefore similarity of promoter regions of these genes, might be due to probable similar regulation at the stages of initiations of transcriptions.

Acknowledgements

This work was supported by “Human Genome Project” (project № 89’99) and RFBR (project № 00-15-9777). The authors are grateful to prof. Sulimova G.E. for helpful discussions.

References