Detection of tissue-specific genes and computational analysis of testis-specific gene expression regulatory regions

Akifumi Yamashita¹
uhmin@gen-info.osaka-u.ac.jp

Naohisa Goto¹
ngoto@gen-info.osaka-u.ac.jp

Seiji Nishiguchi²
nishigu@hirakataryoiku-med.or.jp

Kazuori Shimada²
kazshim@hirakataryoiku-med.or.jp

Hiromichi Yamanishi²
hirochan@hirakataryoiku-med.or.jp

Teruo Yasunaga¹
yasunaga@gen-info.osaka-u.ac.jp

¹ Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan
² Hirakata Ryoikuen, 2-1-1 Tsudahigashi, Hirakata, Osaka 573-0122, Japan

Keywords: cAMP response element (CRE), Tissue-specific expression, GNF gene expression database

1 Introduction

Accumulation of microarray data enabled us to analyze the expression level of many genes in many different tissues. This technology can be used to find genes with unique expression pattern, such as tissue-specific genes. For most genes, the major cause of the difference of gene expression comes from the transcriptional control. The transcriptional control of each gene is done by the interactions between gene regulatory proteins and short stretches of DNA of defined sequence in a regulatory region relatively near the site where transcription begins [1]. In the previous study, we searched for the regulatory sequences of testis-specific genes, because the genes showing testis-specific expression are the most abundant among the genes exhibiting tissue-specific expression [5]. In this study we further investigated the conservation of these motifs in the regulatory regions of testis-specific genes and non-testis-specific genes.

2 Methods

Tissue-specific genes were selected using the gene expression database of the Genomics Institute of the Novartis Research Foundation (GNF) Mouse GNF1M (MAS5-condensed) expression database, which was downloaded from the website [4, 7]. The gene expression intensities were averaged for each tissue, and thereafter log-transformed and normalized. Bono et al. (2003) [2] denoted the gene as tissue-specific if the normalized value exceeded mean +3 standard deviation (S.D.) for their cDNA microarray data and mean +2 S.D. for Affy chips. However, we denoted the gene as “tissue-specific” if the normalized value in the tissue exceeded mean +4 S.D., for more stringent evaluation. We will refer to this value as the tissue specificity of the gene. Because testis-specific genes are abundant, and spermatogenesis is an excellent model for studying the regulation of gene expression during differentiation [6], we used testis-specific genes for further analysis. We extracted 5′-regulatory regions of these testis-specific genes that are arbitrarily defined as [-300 to +50] relative to the transcription start sites for further analysis. Among these regulatory regions, we searched for over-represented 8-nucleotide sequences (8-mers) compared with the randomized regulatory regions. Appearance frequency of over-represented 8-mers is also compared with those within regulatory regions of non-testis-specific genes which are denoted as those having absolute value of testis-specificity less than 1.

3 Results and Discussions

Using the expression data, we found 1,012 testis-specific genes. They were the most abundant among the tissue-specific genes. From these testis-specific genes, we obtained 634 complete regulatory regions. We also
obtained 8,466 regulatory regions of non-testis-specific genes. Among these testis-specific regulatory regions, we found 117 significantly over-represented 8-mers that appeared 2,648 times within the 634 testis-specific regulatory regions. Of these, 64 over-represented 8-mers were significantly more frequent within the regulatory regions of testis-specific genes than within those of non-testis-specific genes. This group contained cAMP response element (CRE)-like sequence, sequences corresponding to parts of ZII element, CSS-A, and X-box. In this group, CRE-like 8-mers contained 4 types that differed from the canonical CRE 8-mer by 1 nucleotide, but appearance frequency of the canonical CRE in the regulatory regions of testis-specific genes is not significantly differ from those of non-testis-specific genes (Table 1). Further, flanking region of CRE-like sequence with testis-specific regulatory regions are more conserved than those within non-testis-specific genes (Fig. 1). We consider that these CRE-like 8-mers participate in the regulatory expression of testis-specific genes to a greater extent than the canonical CRE 8-mer.

Table 1: Appearance of CRE(-like) 8-mers within testis-specific regulatory regions and non-testis-specific regulatory regions

<table>
<thead>
<tr>
<th>8-mer</th>
<th>Testis-specific (634)</th>
<th>Non-testis-specific (8,466)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGATGTCA</td>
<td>8.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>TGACATCA</td>
<td>5.4%</td>
<td>1.2%</td>
</tr>
<tr>
<td>TGACCTCA</td>
<td>3.3%</td>
<td>0.8%</td>
</tr>
<tr>
<td>TGAGGTCA</td>
<td>3.3%</td>
<td>0.9%</td>
</tr>
<tr>
<td>TGACGTCA*</td>
<td>3.8%</td>
<td>2.8%</td>
</tr>
</tbody>
</table>

* canonical CRE

Fig. 1: Sequence logo [3] analysis of the flanking region of CRE(-like) sequences within the regulatory regions of testis-specific (left) and non-testis-specific (right) genes. CRE(-like) sequence within testis-specific regulatory regions conserved TG and CA at the 5' and 3' flanking regions of the core sequence, respectively, but they were not conserved within those of non-testis-specific genes. a-d corresponding to CRE-like sequences, e corresponds to canonical CRE sequence.

References