Bioinformatic characterization of non-annotated transcription-units related to Celiac Disease

Silvia Fernandez Portela

Methods

Table 1. Patients and biopsies for RNAseq and for qPCR analyses. Active CD: children at diagnosis (on a gluten-containing diet, with CD-associated antibodies, atrophy of intestinal villi and crypt hyperplasia). Treated CD: same patients in remission after being treated with GFD for 2 years (asymptomatic, antibody negative and normalized intestinal epithelium). Control: tissue samples from non-celiac individuals not suffering from inflammation at the time of endoscopy, used as controls.

<table>
<thead>
<tr>
<th></th>
<th>Active CD</th>
<th>Treated CD</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Seq</td>
<td>276</td>
<td>276</td>
<td>276</td>
</tr>
<tr>
<td>Validation</td>
<td>16</td>
<td>16</td>
<td>15</td>
</tr>
</tbody>
</table>

RNAseq

For RNAseq analysis, first Sickle [4] was used to remove low quality reads. Then the Tuxedo protocol [5] was followed. Briefly, sequenced reads were mapped against human reference genome (hg38) using TopHat [6] and providing GENCODE 24 [7] as the reference transcriptome. Cufflinks [3] was used to find new transcripts.

Bioinformatic characterization

The bioinformatic characterization was performed using public databases (Figure 2).

Results and Discussion

From the RNAseq experiment 276 unannotated transcription units are extracted and referred to four different regions: 135 differentially expressed, 33 which are completely off in celiac patients; 13 completely off in healthy patients; 95 relation changes, genes that are correlated with each other.

Bioinformatic classification was very successful, with only 5% of unannotated regions that were not similar to anything (Figure 4). Nevertheless, bioinformatic analysis deserves a further study because it may happen that unannotated transcript is really annotated; this can happen due to a misreading of the bioinformatic analysis after RNAseq.

After the bioinformatics analysis, the choice of three regions for validation was based on the number of isometrics and exon (both equal to 1). Furthermore, is supposed that transcripts that switch on and off are more interesting than others. Thus three transcripts are selected (Table 2). One that is turned off in CD (XLOC_022314) and two that are turned off in CD patients (XLOC_010878, XLOC_012919).

In qPCR results, two of the three transcripts selected (turned off in CD) are highly expressed in controls compared to Treated and even Active CD (Figures 5a and 5b). So they are expressed in healthy subjects and begins activate when they avoid gluten. They are activated also in treated patients.

It has been demonstrated that unannotated regions (Table 2) change their expression in healthy subjects, treated patients and CD patients. Therefore these results confirm RNAseq data information.

Consequently, the detection of new transcripts by the RNAseq could open a new line of study for celiac disease research, and can be extrapolated to other areas in Genetics.

Table 2. Three chosen transcripts.

<table>
<thead>
<tr>
<th>XLOC_002314</th>
<th>XLOC_012919</th>
<th>XLOC_010878</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr20:50,278,410-50,279,321</td>
<td>chr16:5,017,237-5,018,210</td>
<td>chr14:104,082,856-104,083,931</td>
</tr>
<tr>
<td>Search USCS Genome Browser</td>
<td>Search USCS Genome Browser</td>
<td>Blast</td>
</tr>
<tr>
<td>Similar to LincRNA: LINC01272</td>
<td>mRNA: SEC14L5</td>
<td>Predicted LOC105370691</td>
</tr>
</tbody>
</table>

qPCR and statistical analysis

qPCR and Statistical analysis followed the next workflow (Figure 3).

Figure 4. Bioinformatic characterization results. 

Figure 5. qPCR results of three transcripts. Fold change to control average vs. Control (CNT) Treated (T) and Active (AC) patients.

References & Acknowledgements

This work is part of projects ISCIII-P13/1201 GVSAN- 2011111034 to JRB, and are approved by the Cruces University Hospital and Basque Clinical Trials and Ethics Committees (Codes CEIC- E09/10 and P20/1307) Biopsies of distal duodenum were obtained by endoscopy after informed consent was obtained from all subjects or their parents. qPCR are made by SGIKER by Fluidigm platform.

I would like to thank Dr. Jose Ramon Bilbao for his support and encouragement and Dr. Koldo Garcia-Etxebarria for his invaluable technical assistance.


[Image 1]