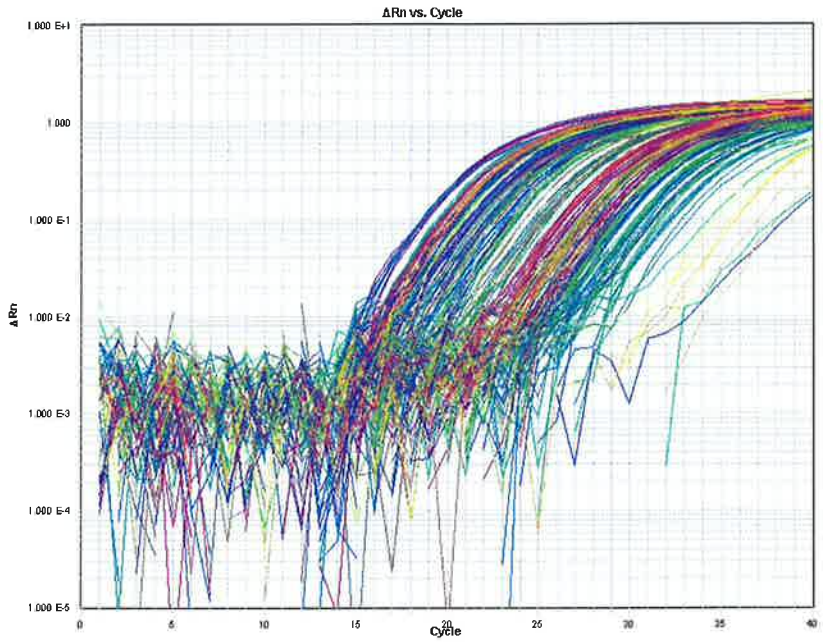
CHALMERS



Exploratory study of gene expression in small intestine from genes associated with celiac disease

LISA OLSSON

Master of Science Thesis in Biotechnology

Supervisors: Åsa Torinsson Naluai, Department of Clinical Genetics & Genomics Core Facility
Christofer Flood & Hannah Hansen Erlandsson Genomics Core Facility
UNIVERSITY OF GOTHENBURG

Examiner: Marie Alminger, Department of Chemical and Biological Engineering, Food Science,
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2011

THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Exploratory study of gene expression in small intestine from genes associated with celiac disease

LISA OLSSON

Department of Clinical Genetics & Genomics core facility
UNIVERSITY OF GOTHENBURG

Department of Chemical and Biological Engineering
Food Science
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2011

Exploratory study of gene expression in small intestine from genes associated with celiac disease

MASTER OF SCIENCE THESIS IN THE MASTER DEGREE PROGRAMME BIOTECHNOLOGY

Supervisors:
Åsa Torinsson Naluai
Department of Clinical Genetics & Genomics Core Facility
University of Gothenburg

Christofer Flood
Hannah Hansen Erlandsson
Genomics Core Facility
University of Gothenburg

Examiner:
Marie Alminger
Department of Chemical and Biological Engineering
Food Science
Chalmers University of Technology

Department of Chemical and Biological Engineering
Food Science
Chalmers University of Technology
Göteborg, Sweden, 2011

Exploratory study of gene expression in small intestine from genes associated with celiac disease

LISA OLSSON
Chalmers University of Technology

Abstract

Celiac disease is a complex disorder characterized by chronic inflammation in the small intestine caused by an auto immune reaction when gluten is digested. Parts of its genetic background are known, but there are still large parts that are unknown. Genetic studies have confirmed strong linkage and association to the HLA region on chromosome 6 and polymorphism in genes in HLA can explain the pathogenesis of the disease. However, these haplotypes are common variants and cannot alone answer as the trigging factor of the disease. Genome-wide association studies (GWAS), another approach to identify other regions, which can contribute to the genetic background of the disease. In this Master thesis, mRNA expression of genes linked to loci found to have strong association to celiac disease in a GWAS by Östensson et.al. (in manuscript) was examined. The differences in expressions between individuals with celiac disease and control individuals are investigated and the effects of insulin stimulation during nutrient deprivation. In untreated samples a difference in expression could be seen between individuals with celiac disease and controls for the gene encoding the insulin receptor (INSR). Lower expression of the neurotensin (NTS) gene could also be seen for the celiac group. When stimulated with insulin under nutrient deprivation differences in expressions between the celiac group and controls could be seen for the genes encoding LPP, α-actinin -1 (ACTN1), insulin receptor and TIPRL, which could not be seen for unstimulated samples.

Exploratory study of gene expression in small intestine from genes associated with celiac disease

LISA OLSSON
Chalmers University of Technology

Sammanfattning

Celiaki eller glutenintolerans är en komplex sjukdom som uppkommer av en autoimmun reaktion när födoämnen innehållande gluten bryts ner. Reaktionen orsakar en kronisk inflammation i tunntarmen som påverkar slemhinnans struktur och dess förmåga att absorbera näringsämnen. Studier har visat att det finns en stor genetisk komponent bakom uppkomsten av sjukdomen och en del av den genetiska bakgrunden är känd, men en del saknas. Genom studier av variationer i hela det humana genomet, genome-wide association studies (GWAS), kan associationer mellan regioner och sjukdomen upptäckas. I följande studie undersöks genuttrycket av intressanta gener från regioner associerade till celiaki enligt resultat från en GWAS av Östensson et.al. (opublicerat). mRNA-nivåer i obehandlade tunntarmsbiopsier från individer med celiaki jämförs med mRNA-nivåer i kontroller. mRNA-nivåer i biopsier stimulerade med insulin under svält undersöks också. I de obehandlade proven sågs ett lägre uttryck för generna som kodar insulinreceptorn (INSR) och neurotensin (NTS). I de behandlade proven sågs skillnader i genuttryck för generna LPP, TIPRL, α-actinin-1 (ACTN1) och insulinreceptorn. Vid insulinstimulering ökade genuttrycket i kontrollerna medan individerna med celiaki inte hade någon förändring eller hade ett lägre genuttryck.

Contents

1 Introduction

1.1 Pathogenesis of celiac disease

Celiac disease is an autoimmune disease, characterized by chronic inflammation in the small intestine. The disorder is caused by both genetic factors and by environmental factors. The unique feature of celiac disease is that the environmental factor, which also serves as the antigen trigger, that causes the condition, is known. This environmental factor is ingestion of gluten. Gluten is a complex mixture of around 100 related proteins rich in wheat. Wheat gluten is a complex mixture of polypeptides and the most common components are gliadin and glutenin. Other related protein mixtures that also induce celiac disease are present in rye and barley (Koning et al. 2005).

The condition is life-long but can be treated if gluten is excluded from the diet. If gluten is not excluded from the diet an immune response to gliadin causes an inflammatory reaction. The inflammation causes damage to the mucosa of primarily the duodenum, the upper part of the small intestine. A chronic inflammation causes the villi of the epithelium in the intestine to degrade. The mucosa is flattened which lead to a reduced absorption area and reduced ability to absorb nutrition (Green & Cellier 2007).

It was first recognized as a malabsorption syndrome in small children but is now seen as a condition that can be diagnosed at any age. Today the rate of children and adults with celiac disease is approaching one percent. It is found in populations all over the world and the numbers of diagnoses are increasing (Green & Cellier 2007).

Classical symptoms in childhood are diarrhea, growth retardation and abdominal distension. But other common symptoms are vomiting, irritability and constipation. In adults the symptoms are more diverse and the disease resembles more a multisystemic disorder with the primary site in the intestine. The fraction of patients with classical symptoms is decreasing and other ways of presenting the disease are discovered. The classical symptom diarrhea is only present in 50 percent of cases. Other celiac disease related effects are iron-deficiency anemia and osteoporosis. Many have previously been diagnosed with irritable bowel syndrome. Other unusual symptoms are neurological problems as peripheral neuroparthy and severe ataxi or abnormalities in the blood chemistry. Some patients have no symptoms at all or are diagnosed after endoscopy for symptoms not associated with celiac disease or through screening. Sensitive and specific serological testing has made it possible to screen relatives of diagnosed patients and other high risk groups, for example, patients with insulin dependent diabetes. Diagnosis of the condition requires serological antibody detection and a duodenal biopsy with characteristic findings and a positive response to a gluten-free diet (Green 2005).

Among adults, as in many other autoimmune diseases, the prevalence is higher in women. There are three times as many women as men that have the disease. The prevalence of other diseases is higher among patients with celiac disease compared to the general population. These diseases are generally other autoimmune diseases such as type 1 diabetes, psoriasis, thyroid diseases,

neurological problems, autoimmune liver diseases and autoimmune cardiomyopathy. Celiac patients also have an increased rate of malignancies (Green 2005).

1.2 Genetics

That celiac disease has a genetic influence is seen since it clusters genetically in families. It has a sibling relative risk of ~20% (Sollid, 2002) and a high concordance between monozygotic twins, around 70-75 % (Greco et.al., 2002). This suggests that there is a large genetic component linked to the disease.

1.2.1 The HLA factor in celiac disease

Linkage and association studies have shown strong evidence of linkage to the genes in the HLA-region on chromosome 6 (6p21). Especially the HLA-DQ2 and the HLA-DQ8 genes which encode MHC class II molecules involved in antigen presentation in the adaptive immune system (Abadie et.al., 2011). These proteins are present on the surfaces of B-cells, T-cells and macrophages and function as a holder for antigen and present the antigen to, for example, a T-cell receptor on a DC4+ helper cell (Louka & Sollid, 2003).

Gluten is rich in glutamine and proline residuals and reaches the duodenum fairly undigested. The enzyme transglutaminase 2 (TG2) is naturally expressed in the duodenum and glutamine rich gluten is a good substrate for TG2. The TG2 enzyme deamidates the glutamine residuals creating peptides with negatively charged residuals. The risk haplotypes, of HLA-DQ2 and HLA-DQ8 genes, linked to celiac disease have properties which improve binding to deamidated gluten peptides and increases the risk to promote pathogenic T cell respond (Louka & Sollid, 2003, Abadie et.al., 2011). Gluten specific CD4+ T-cells can be isolated from patients with celiac disease and in these patients there is also often an increase in TG2 expression (Abadie et.al., 2011).

However it has been questioned that the CD4+ T-cell response alone can cause the cell degradation of mucosa and villous atrophy. First the response must reach a certain threshold to induce tissue damage. This can be reached, for example, by expressing the version of the HLA-molecule which form the most stable MHC-peptide complex or by increasing the amount of antigen, for example through an increased TG2 expression. Secondly it is suggested that the patients must have an intestinal stress response together with the antigluten- immune response to activate cell degradation. As suggested by Ferguson et. al. (1993), this stress response can be caused by factors such as hyper permeability, nutrient deficiency, increased intake of gluten, effects of intestinal infections and non HLA-associated genes. The celiac disease has a wide biological, histological and clinical spectrum and the patients can have a histological picture from normal villous architecture to total villous atrophy. A correlation between the amount of TG2 antibodies presence and the degree of villious atrophy has also been seen (Green, 2005, Abadie et.al 2011).

Not only is the expression of TG2 in persons with celiac disease increased, they also develop auto antibodies against TG2. This suggests that celiac disease have an autoimmune component. Trough epidemiological studies it has been shown that there are links between celiac disease and other autoimmune disorders, particular type 1 diabetes and autoimmune thyroiditis (Abadie et.al., 2011).

Most patients with celiac disease have one or two copies of the DR3-DQ2 haplotype or are DR5-DQ7 and DR7-DQ2 heterozygote. The susceptibility is not influenced by whether the haplotypes are carried *in cis* or *in trans* instead it is the interaction between the two genes that influence the susceptibility to celiac disease. However, 20-30% of the Caucasian population express these haplotypes and the incidence of the disease is only approximately 0.5-1% (Sollid, 2002). Although the association between celiac disease and certain genotypes of HLA the genes give some answers to the diseases pathogenesis. It cannot alone answer for the entire genetic background of celiac disease.

1.2.2 Searching for non-HLA genes through Genome wide association studies

To find the remaining heredity, genome wide association studies (GWAS) is used. In these studies single nucleotide polymorphisms (SNP) represents variation in most of the genes in the human genome. With this method small differences in the genome between individuals can be identified. The method is not depended on previous knowledge of the biochemical, molecular function or pathogenesis. By comparing which variation of a SNP is more often present in a group of individuals with celiac disease compared to a control group, associations to celiac disease can be identified (Hirschhorn & Daly, 2005). Most associations found by GWAS are typically common variants in the human genome and probably have a low, individual, impact on the susceptibility of celiac disease. It is more likely that the non-HLA genes influence the susceptibility of celiac disease trough control or other mechanisms altering gene-expression, rather than a change on the protein coding level (Abadie et.al, 2011).

In the GWAS published so far association has been reported to 39 non-HLA loci, which together explain ~5% of the risk for the disease. From these regions 115 genes have been identified as possible candidate genes (Trynka et.al, 2010). Among the genes reported a large part are involved in the T cell-mediated immunological response. They can be associated with functions such as triggering and activation of T-cell response, immune cell signaling, T-cell maturation but also functions in the non-specific immune system for example the response to exogenous factors such as infections. Among the genes reported there are also genes involved in stress response pathways (Abadie et.al, 2011, Dubois et.al, 2010, Trynka et.al, 2010).

According to a GWAS by Östensson et.al. (in manuscript), using 100 celiac families from Sweden and Norway, regions containing genes involved in nutrient sensing, macromolecular processing and growth factor signaling were found to have strong association to celiac disease. Many genes were involved in the insulin pathway and pathways involved in stress reactions triggered by nutrient deprivation. Present study is based on data from this GWAS.

Among the genes of interest is the insulin receptor-gene (INSR), which together with insulin stimulates the uptake of glucose and the NTS-gene, the encoder of neurotensin, which stimulate insulin secretion at low glucose levels (Béraud-Dufour et.al, 2010). Another gene in a loci with strong association is the TIPRL-gene, which negatively regulate the TOR pathway. The TOR pathway is involved in cell growth regulating response to nutrition and tightly connected with the insulin pathway. At low nitrogen levels or during starvation conditions the TOR pathway is inhibited, resulting in further signaling and transcription of genes. TIPRL act as an inhibiting factor on the TOR pathway (Jacinto et.al, 2001). Other loci with strong association in the same

GWAS include genes involved in intracellular processes like cytoplasmic vesicles, protein and macromolecular transport and actin binding. These processes are important in for example nutrient sensing and the response to insulin. Among the genes with the strongest association were alpha-actinin 1 (ACTN1) and LPP. LPP was also a gene with strong association in the GWAS by Dubois et. al (2010). ACTN1 is an actin binding protein and has been shown to be involved in remodeling the cell cytoskeleton and cell junctions in response to, for example, Insulin like growth factor (IGF-1) (Guvakova et. al, 2002). LPP is a focal adhesion protein that has been shown to bind α-actinin. It is involved in intra cell signaling and cell-cell contact (Li et.al, 2002).

One way of investigating the result from the GWAS further is to study the correlation between genotype and the expression of a gene or genes nearby the SNP in the associated loci. About 50% of the disease associated SNP affects expression of closely located genes. By investigating the expression of a gene or genes nearby (expression quantitative trait loci, eQTLs) SNPs with functional effects on a nearby gene can be identified (Hrdlickova, 2011). These studies must be made in every type of tissue relevant for the disease. In the case of celiac disease some tissues of interest are blood and intestinal epithelial cells (Trynka el. al, 2010).

The objective of this Master thesis is to make an exploratory study of mRNA expression from genes linked to loci found to have strong association to celiac disease in the GWAS by Östensson et.al. (in manuscript). Differences in mRNA expressions between individuals with celiac disease and control individuals are investigated. Since many of the genes of interest are involved in insulin signaling and nutrient sensing, it is also investigated how insulin stimulation during nutrient deprivation influences the mRNA expression.

2 Materials and Methods

Real-time PCR or quantitative PCR (qPCR) is a technique that can detect and measure the amount of product that is generated in each cycle in a PCR reaction, this amount is in direct relation to the original amount within a sample. The technique can be used to analyze and quantify, for example, gene expression.

2.1 PCR

During the polymerase chain reaction (PCR) a DNA fragment is amplified between two, in beforehand decided, places on the DNA molecule. By designing two experiment specific primers, compatible to the two ends of the fragment, the chosen fragment is copied in each PCR step. The DNA sample is mixed in buffer with dNTPs, building bricks in the elongation, a heat stable polymerase, which elongates the fragments, and the designed primers. During one PCR step, illustrated in figure 1, the temperature is increased and DNA denaturize, the temperature is then lowered to let the primers anneal, the temperature is increased again and the polymerase, which bind to the primers, elongate the fragment (Kubista, 2006). Since one copy is made from each strand available the number of copies are increasing exponentially;

(1)

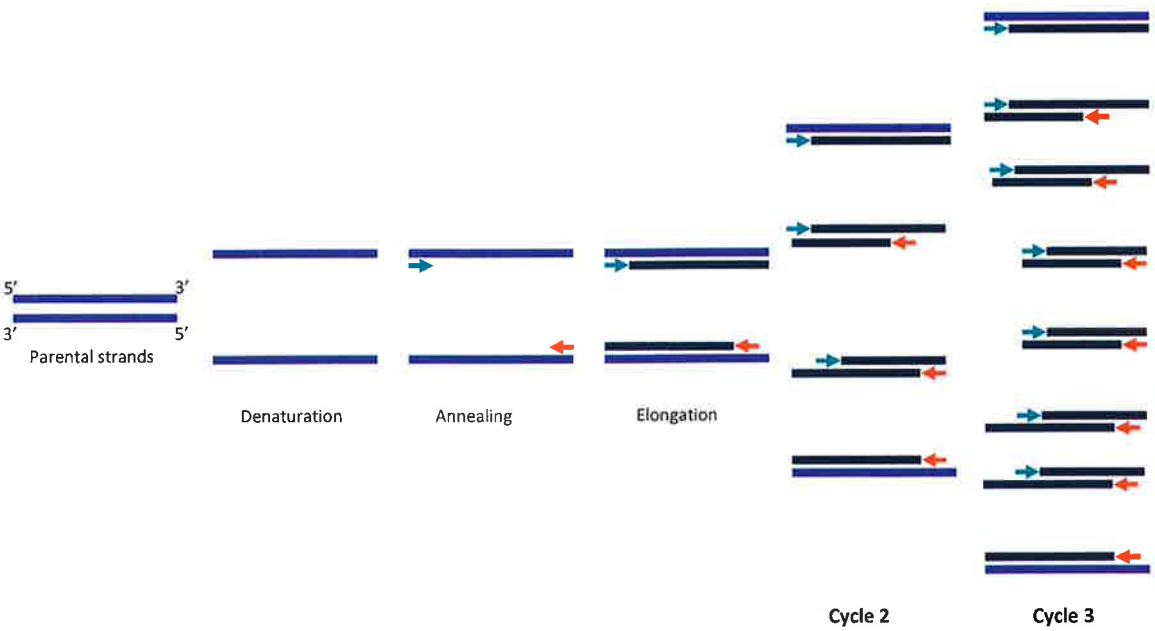


Figure 1. The polymerase chain reaction.

2.2 Quantitative PCR

When running a quantitative PCR a fluorescent reporter probe can be used to report the accumulation of product during the PCR (Kubista, 2006). In this study the TaqMan chemistry from Applied Biosystem is used (Applied Biosystems by Life Technologies Carlsbad Carlifonia, USA). This chemistry contains a fluorescent, sequence specific, probe. The probe consists of a reporter in one end and a molecule, to prevent fluorescent signaling, to the other. The probe binds to the fragment, which is being amplified. When the complex is intact the fluorescence from the probe is highly reduced by the proximity of the quencher molecule by fluorescence resonance energy transfer (FRET) between the two molecules. After the primer has annealed and during amplification the probe is cleaved by the Taq polymerase. The reporter, when no longer close to the quencher, increases its fluorescence signal, as seen in Figure 2. In each cycle a probe is cleaved resulting in an accumulation of fluorescent signal proportional to the number of amplicons generated

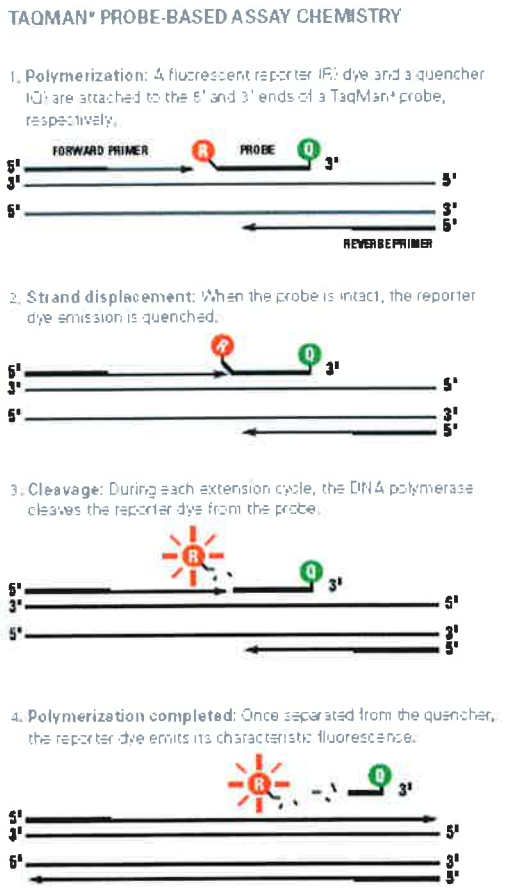


Figure 2. Overview of TaqMan®-Probe-Based Assay Chemistry for qPCR

The amount of fluorescent signal detected reflects the initial amount of sample. When the PCR products reach a certain amount the exponential growth can be detected. At this point the signal from the product is larger than the signal from background; this phase is called growth phase. The number of cycles until this occurs reflects the starting amount in the sample. The different amounts of starting material in different samples is compared by evaluating how many cycles that are required for the different samples to reach a threshold level in the growth phase. The cycle for this amplicon concentration or threshold value is called the Ct-value and reflects the individual sample's initial concentration (Kubista 2006, Ginzinger 2002).

The quantification can either be made by absolute quantification, where a standard curve of known sample concentrations is used as reference or by relative quantification where a relative amount of product is calculated through normalization with product from a reference gene. The normalization is necessary to compensate for the differences in the amount of biological starting material. When using a reference gene for normalization the most challenging problem is to find a gene with a constant gene expression independent of individual or situation for the tissue used (Ginzinger 2002).

2.3 Data analysis

If the PCR is ideal, as in equation 1, the number of replicates is growing exponentially with the base 2. However, depending on the primer's properties the PCR is rarely ideal and varies depending on the assays used. The PCR's efficiency is noted E and is a value between 0 and 1, see equation 2:

$$[N]_n = [N_0](1+E)^n \tag{2}$$

According to the definition of threshold cycle (Ct), the concentration of product from a target gene of interest and the concentration of the product from the reference gene are constant at their respective Ct values:

$$[N]_T = [N_0]_T(1+E_T)^{Ct_T} = k_T \tag{3}$$

$$[N]_R = [N_0]_R(1+E_R)^{Ct_R} = k_R \tag{4}$$

When a target sample $[N]_T$ is normalized to the reference gene $[N]_R$:

$$\frac{[N]_T}{[N]_R} = \frac{[N_0]_T(1+E_T)^{Ct_T}}{[N_0]_R(1+E_R)^{Ct_T}} = \frac{k_T}{k_R} = K \tag{5}$$

Rearranged:

$$\frac{[N_0]_T}{[N_0]_R} = K\left[\frac{(1+E_R)}{(1+E_T)}\right]^{-\Delta Ct} \tag{6}$$

$$\Delta Ct = Ct_T - Ct_R \tag{7}$$

The constant K includes variables as RT-yield and properties of the detection chemistry for each assay; these are assumed to be constant for all samples but not necessary equal to 1. They are however difficult to determine experimentally and instead the normalized target concentration is compared to another normalized calibrator sample. The choice of calibrator sample only affect how the data is shown and can for example, as in this case, be an average of the control samples. The calibrated expression ratio is given by:

$$RQ = \frac{\left[\frac{[N_0]_T}{[N_0]_R}\right]_{sample}}{\left[\frac{[N_0]_T}{[N_0]_R}\right]_{cal}} = \frac{\left[\frac{(1+E_R)}{(1+E_T)}\right]^{-\Delta Ct_{sample}}}{\left[\frac{(1+E_R)}{(1+E_T)}\right]^{-\Delta Ct_{cal}}} = \left[\frac{(1+E_R)}{(1+E_T)}\right]^{-\Delta\Delta Ct} \tag{8}$$

$$\Delta\Delta Ct = \Delta Ct_{sample} - \Delta Ct_{cal} \tag{9}$$

If also the assumption is made that the PCR efficiency is 100%:

$$RQ = 2^{-\Delta\Delta Ct} \tag{10}$$

This method to calculate the relative quantification (RQ) is known as the ΔΔCt method (Livak 1007, Livak et. al. 2001). To be able to analyze the result with parametric tests, such as student t-

tests, the result must be normally distributed. Gene expression data is most often not normally distributed when represented as relative quantification but become normally distributed when logarithmically transformed to fold difference or fold change. An increased fold change corresponding to one Ct value represents double the expression for that sample and a decrease of one represent half of the expression. (Bergkvist et.al, 2010);

$$Fold\ change = \ln RQ \tag{11}$$

2.3.1 PCR efficiency

The PCR efficiency can be calculated from a standard curve. The Ct values from a run with a dilution series are plotted against the natural logarithm of the initial concentrations. From equation 2 at n= Ct the following relation is received:

$$\log[N_0] = -\log[E+1]Ct + \log[N_{Ct}] \tag{11}$$

If E and N_{Ct} are constant Ct and $\log[N_0]$ have a linear relation. The slope k is equal to $-\log[E+1]$. From the slope the efficiency, E can be calculated as:

$$E = 10^{-k} + 1 \tag{12}$$

(Rutledge & Cote, 2003).

2.3.2 Reference gene analysis

When determining the gene expression two types of variation are measured. The true biological variation of interest but also non-specific variation caused by sample quality and quantity, the extraction yield and the enzymatic reactions in the reverse transcriptase PCR. To avoid introducing variation in the experimental procedure it is of importance to standardize each step and use similar sample sizes in the extraction, reverse transcriptase PCR and in the qPCR (Ståhlberg et.al., 2004).

At the 3rd London qPCR symposium (April 2005) the most appropriate method for normalization of real-time PCR data for biological sample quantity was considered to be the use of reference genes. A reference gene is a gene with a stable expression, constant mRNA level, in all samples and under all conditions. This method offers an internal control that is affected by all kinds of variance during the experimental process. However, a major problem with the method is to find a constantly expressed gene for all purposes. There are no universal reference genes with the same expression, at all conditions, in every type of tissue. For every system studied the reference gene expression used must be carefully validated under the conditions used (Vandesompele et. al, 2009).

In gene expression studies often only a single reference gene is used. In these cases the effects of an inappropriate reference gene are then unnoticed. In a study by Vandesompele et. al. (2002) the errors associated with the use of a single reference gene was quantified. The result demonstrate that when using a single, non validated, reference gene for normalization the errors in expression was 3 fold in 25% of the cases and 6 fold in 10% of the cases.

There are several methods and software that can be used for validation and selection of reference genes. In this study the software qBase^{PLUS} is used for quality control analysis and reference gene validation of the qPCR-data. The software includes a tool to evaluate and select the most stably expressed reference genes using the method geNorm (Hellemans et.al, 2007). In geNorm the stability of multiple reference genes are tested by investigating the pairwise variation with all other reference genes in the experiment. The stability value (M) is the average pairwise variation of one reference gene with all other tested candidate reference genes. Genes with the lowest M-value are the most stably expressed. Normalization can then be made with a normalization factor baseed on the geometric mean of the best reference genes (Vandesompele, 2002). The coefficient of variance (CV) is a marker for the stability of the expression of the reference genes, which should be constant after normalization. Helleman et.al (2007) has experimentally determined that the M value and the CV value should be lower than 25% and 0,5 respectively for stably expressed reference genes.

2.4 Sample preparation, RNA isolation and cDNA synthesis

Human small intestine biopsies from three clinics in Swedenwere gathered from 30 individuals. One sample from each individual was instantly placed in RNAlater(untreated sample), to protect from degradation. For each individual two biopsies were taken and placed in DMEM low glucose medium with 5% Fetal calf serum (FCS). One biopsy was incubated in low glucose medium alone (treated sample) and one in low glucose medium including 10mM insulin for 24 hours (stimulated, treated sample). After incubation the biopsies were put in RNAlater. Total RNA was extracted from treated and untreated samples using Qiagen TissueLyser and Qiagen mRNeasy kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Total RNA was eluted into 40µl RNAase free water. RNA concentration was measured with a NanoDrop™ ND-1000 Spectrophotometer (Nanodrop Technologies, USA). Samples were stored in -20° until used.

mRNA was transferred to complimentary DNA by reverse transcriptase using SuperScript® VILO™ cDNA synthesis kit (Invitrogen, USA) according to manufacturer's instructions. 1 µg of RNA was used for the untreated samples, from which generally larger concentrations of RNA were received. For the treated samples RNA amounts between 300ng and 1000 ng were used. The tubes were incubated for 10 min at 25°C, 60 min at 42°C and the reaction was terminated at 85°C for 5 min. The cDNA was stored at -20°C until used.

2.5 Pilot study

To evaluate the efficiency of the qPCR assays, evaluate gene expression in the treated samples and evaluate appropriate cDNA concentration to use in upcoming experiments, a pilot study was made. A new experimental setup design was also tested. In this experimental setup the primermix was evaporated in the reactionplate and cDNA mixed with master mix and water was then added in the total reaction volume, to reduce the critical technical variation in cDNA volume. Twelve TaqMan gene expression assays (Applied Biosystems, USA) were evaluated, three reference genes (EPCAM, GUSB;ACTB) and nine candidate target genes (DCC, EPCAM, ACTN1, NTS, STK32A, STK32B, TIPRL, INSR) in two different pooled cDNA mixtures (untreated samples and treated samples) at five concentrations (50, 25, 5, 0.5, 0.05 ng/µl). The plate was set up using the Biomek FX (Beckman Coulter, USA) and qPCR was run in 5µl reactions and duplicates on a ABI PRISM 7900 (Applied Biosystems, USA) according to the manufacturer's instructions.

2.6 qPCR and data analysis

Five target genes, selected from the pilot study, and three reference genes were analyzed using the three samples from each individual (one untreated and two treated). A volume corresponding to 5-10ng of undiluted cDNA was used for each reaction together with 1x TaqMan Gene Expression Assay and 2x TaqMan Genotyping Master Mix in a total reaction volume of 5 µl. Each plate was set up in triplicates using the Biomek FX (Beckman Coulter, USA) with inter plate triplicates The qPCR was made on a ABI PRISM 7900 (Applied Biosystems, USA) using the SAMI core system (Beckman Coulter, USA).
The raw data was analyzed using SDS version 2.4 and RQ manager 1.2.1. Ct values were extracted using a threshold value of 0,02, illustrated in Figure 3.

The Ct replicates were evaluated and for differences in Ct values larger than 0,5 outliers were removed. An average Ct value for each sample was calculated using the arithmetic mean. Average Ct value for samples with undetectable expression was set to 40. The reference genes were validated by plotting the ct-values of the reference genes against each other. After validation the target genes were normalized against an arithmetic mean of the three reference genes, according to equation 7. Relative quantities and Fold change were calculated using an average of the untreated control as a calibrator sample according to equations 9, 10 and 11. The efficiencies of the assays where assumed to be close to 100% (±10%) according to manufacturers guarantees (Applied Biosystems Application Note, 2006). Calculations were implemented using R.

Reference gene stability was evaluated using the geNorm module in the software qBase^{PLUS} v2.1 (Biogazelle, Belgium) on the untreated samples.

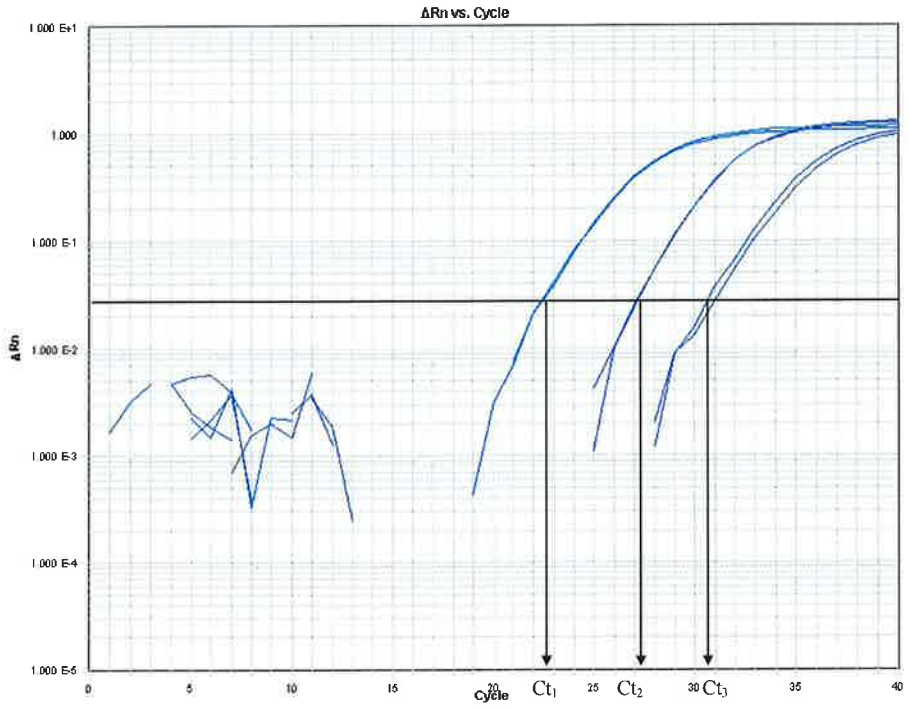


Figure 3. Example of Ct-value extraction from logarithmic transformed raw data.

3 Results

3.1 Pilot study

Reliable results, which could be used as data points in standard curves, were only received for the genes INSR and EPCAM. The efficiencies (E) were calculated as 0,91 and 0,99 respectively, see Figure 4.

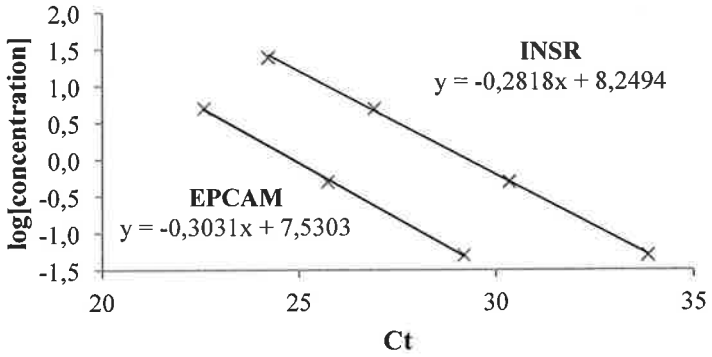


Figure 4. The real-time PCR standard curve. The Ct values are plot against the logarithmic concentrations in the series of dilutions. From the slope (k) the efficiency (E) can be calculated, according to eq. 12.

The five genes with best data quality and expression detected were chosen for the main experiment. The five chosen genes were the insulin receptor (INSR), neurotensin (NTS), α-actinin 1 (ACTN1), LPP and TIPRL. The untreated samples generally had earlier amplification and better data quality. The pilot study showed that the preferable concentration of cDNA for the majority of the assays were 5 ng/µl.

3.2 Reference gene validation and quality evaluation

To validate the stability of the expression for the three reference genes the correlation of the reference genes was first visualized by plotting the Ct values for each reference gene against the other two, see Figure 5. A correlation can be seen in all individual comparisons. The correlation is best in the EPCAM/GUSB comparison. An arithmetic mean of the three reference genes was used for normalization in the data analysis.

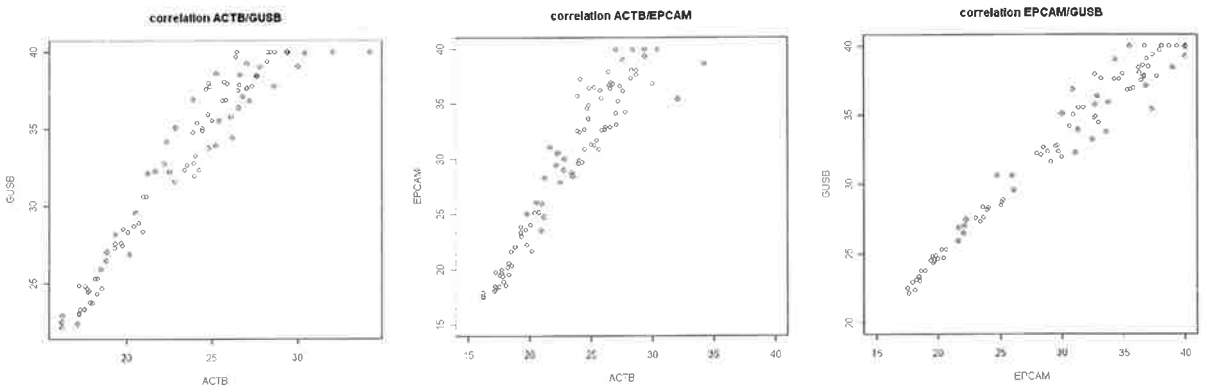


Figure 5. Correlations plots of reference genes.

The reference gene stability was also evaluated using geNorm. According to the geNorm algorithm of the reference gene stability the experimental design layout of present experiment was acceptable. The variation among replicates were evaluated and the replicates which fell outside of the set value of 0,5 were removed. The coefficient of variation CV and expression stability value M, for the three reference genes were calculated to 0,33 and 0,81. This result is higher than the recommended values of 0,2 and 0,5. Out of the three reference genes analyzed; GUSB was ranked as the most stably expressed reference gene and ACTB the least. The software could not recommend any optimal reference gene and suggested that the experiment had low reference gene stability.

3.3 Celiac and controls in untreated samples

When comparing the expression from individuals with celiac disease and control individuals in untreated samples, se Figure 6, a lower level of expression from the INSR gene can be seen in the individuals with celiac disease (p=0,046). Unpublished results show that the expression of NTS in individuals with celiac disease is lower than the expression in controls. These results also support these findings (p=0,042).

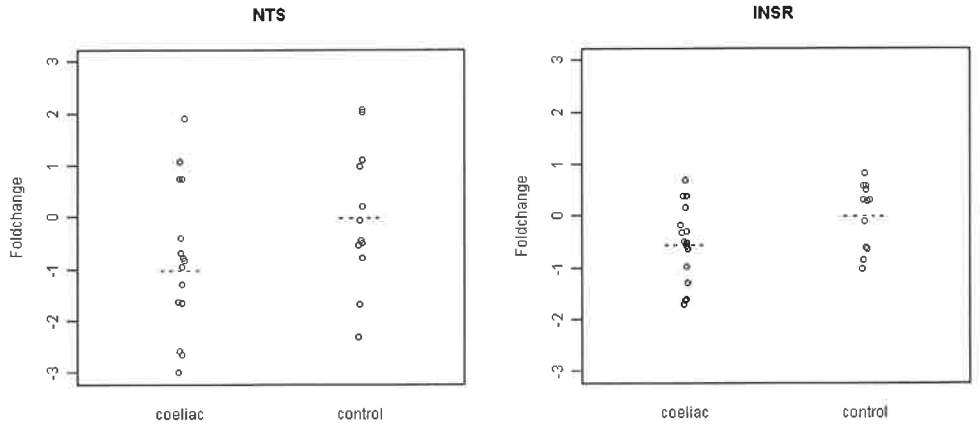


Figure 6. Fold change plotted in the two compared groups, celiac and control, for NTS and INSR.

3.4 Evaluation of expression after insulin stimulation

In the samples stimulated with insulin under nutrient deficiency a trend can be observed indicating differences in mRNA levels between the two groups, celiac and controls, for the LPP, ACTN1, INSR and TIPRL genes. For LPP, the expression increases slightly for both patients and controls during treatment compared to untreated samples. When stimulated with insulin the expression decreases in the celiac group, to the untreated levels, and increases significantly in the control group (p=0,02 for the stimulated samples), illustrated in Figure 7. For ACTN1 the expression increases in the treated samples compared to the untreated, but there are no

differences between the two groups. When stimulated with insulin the expression increases only in the control group (p=0,03), see Figure 7.

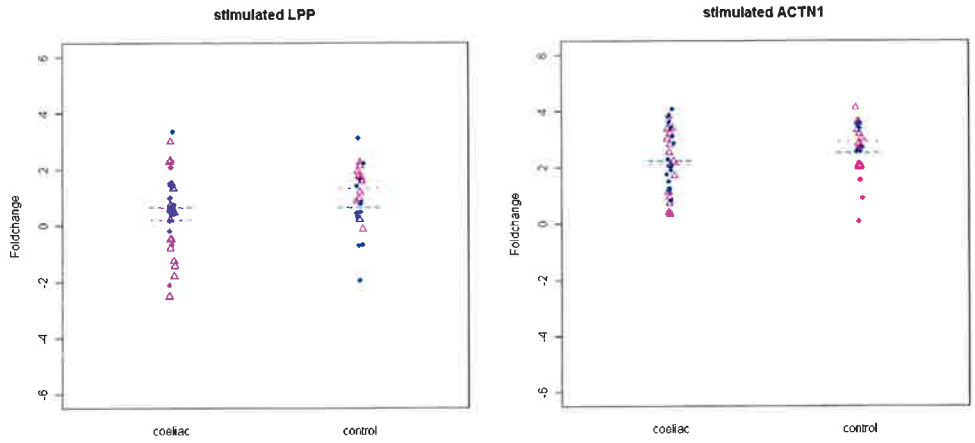


Figure 7. Fold change of LPP and ACTN1 expression plotted for the two groups, celiac and control, for samples ex vivo stimulated.
(• - unstimulated samples (~~ = mean), Δ -samples stimulated with insulin (- - = mean),)

For INSR and TIPRL the mRNA levels also increase when treated. A trend could be observed that the expression in the celiac group is lower than the control group in the samples stimulated with insulin, see Figure 8. For NTS no reliable expression data for the treated samples was received.

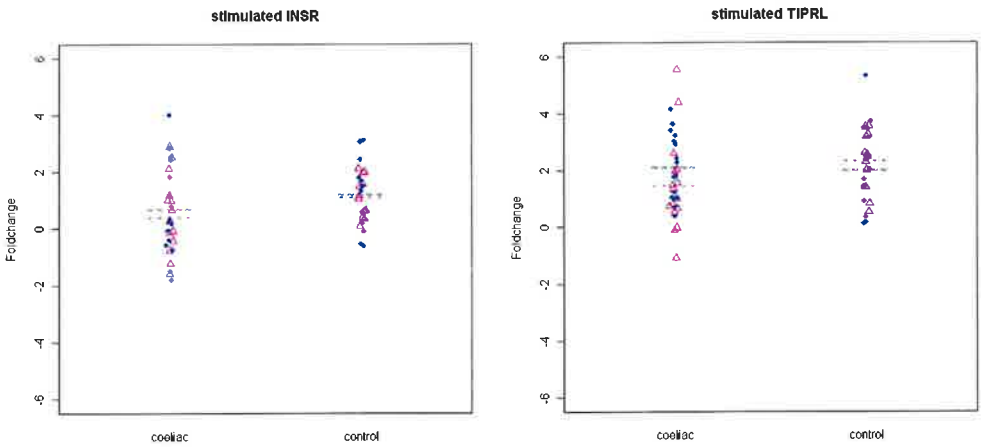


Figure 8. Fold change of INSR and TIPRL expression plotted for the two groups, celiac and control, for samples ex vivo stimulated. P-values 0,08 and 0,09 respectively for INSR and TIPRL in stimulated samples.
(• - unstimulated samples (~~ = mean), Δ -samples stimulated with insulin (- - = mean))

4 Discussion

The genetic background to complex diseases, as celiac disease, is expected to be influenced by genes in tens to hundreds of different loci. Many causal risk variants are probably common variants with low individual effect on the susceptibility of the disease and, because of that, hard to identify. A first step, before detailed mapping of the candidate loci, is to investigate eventual functional effects on nearby genes to associated SNP (eQTL mapping). eQTL mapping in combination with genotyping and detailed mapping through sequencing can, in the future, hopefully confirm causal risk variants behind celiac disease. In the present study, expression from genes near SNPs with strong association to celiac disease, are investigated.

When comparing expression from the untreated samples a difference could be seen between the celiac group and the control group for the INSR and the NTS genes. The celiac group has lower expression than the control group for both genes. Both genes are involved in glucose homeostasis (Lee & Gorospe, 2010, Béraud-Dufour et.al, 2010). The INSR gene encodes for the insulin receptor and is involved in glucose metabolism as a mediator for the metabolic function of insulin. Changes in glucose levels trigger changes in insulin concentrations and changes of proteins involved in insulin signaling, for example the insulin receptor. Insulin sensitivity, the efficiency of the uptake of glucose, is also strongly dependent on the number of insulin receptors present on the plasma membrane of cells in most tissues (Lee & Gorospe, 2010). Neurotensin encoded by the NTS gene is one of the neurotransmitters, in the central nervous system, that reduces food intake and increases energy expenditure (Wilding, 2002). In the peripheral tissues neurotensin act as a paracrine and endocrine modulator and is secreted into the blood stream after food intake and stimulates insulin secretion in beta cells (Béraud-Dufour, 2010). Neurotensin has also been shown to accelerate intestinal transit by causing an increase of intestinal contractions (Bueno & Fioramonti, 1994).

In the treated samples stimulated with insulin we could see changes in mRNA levels that could not be seen in the treated, unstimulated samples. The mRNA levels of LPP and ACTN1 increased in the control samples but not in the celiac samples, when treated with insulin under starvation. ACTN1 encodes the α-actinin 1 protein and is an actin binding protein. LPP is a focal adhesion protein, involved in intra cell signaling and cell-cell contact, and have been shown to bind α-actinin (Li et.al, 2002). Control individuals seem to have an ability to respond to the higher insulin levels by increasing the expression of proteins involved in cell signaling, in this case LPP and ACTN1, that individuals with celiac disease lack. Foster et.al (2006) could show that cytoskeleton proteins and related proteins, including α-actinin 1, are involved in the GLUT 4 signaling and regulation in response to insulin. GLUT 4 is an insulin regulated glucose transporter, which is closely involved with the insulin receptor in the glucose metabolic signaling pathway. This shows that there could be a connection between the altered expression levels of LPP and ACTN1 and the trend of altered expression levels of the insulin receptor (INSR).

The celiac group also had decreased TIPRL expression when stimulated with insulin. Both unstimulated and insulin stimulated samples are under starvation but the expressions in the celiac group drop when treated with insulin. The TOR pathway is inhibited when the cell is under starvation to allow transcription of stress response proteins. TIPRL encodes a protein that is involved in inhibition of the TOR pathway (Jacinto et.al, 2001). When stimulated with insulin this part in the down regulation of the TOR pathway decreases in the individuals with celiac

disease but not in the control individuals. The TOR pathway is also closely linked to the insulin signaling pathway and the changes in TIPRL expression can also be a part of an altered response to nutrition levels or glucose homeostasis signaling.

These results are from experiments made on relatively few individuals and the studied parameters probably have small individual effects on the precipitation of the disease. The statistical analyzes made are meant as guidance for supplementary, validating, studies with further individuals. No statistical adjustments for multiple testing have been done.

In these experiments a new plate setup was tested. This plate setup lets cDNA, mastermix and water to be mixed instead of assay, mastermix and water. This increases the critical pipetting volume for the robot, 5 µl instead of 1 µl, which reduces the effects of possible technical variation. Each sample for each gene is also distributed with the same tip on the 96 pod, which also reduces the possible variation within the analysis for each sample. However, the replicates and gene detection are, with this method, on different plates, resulting in possible larger differences within replicates. However, the majority of the replicates in this study was, for well-expressed genes, with in the 0,5 Ct-value mark. Having the replicates on different plates also make discovery of inter plate technical variation possible. This method uses additional 96 plates, which could be an extra cost and includes an extra step of drying, which requires an extra day to be included in the plans for the project.

Both the lower RNA yield and the result from the pilot study indicate that the treated samples are on the limit of surviving the transport and treatment. The lower RNA yields of the treated samples could be caused by the treatment resulting in that part of the cells in the biopsies has died or that the biopsies of the treated samples simply are smaller. Because of the lower RNA yield for the treated samples a lower concentration of these samples was used in the cDNA synthesis, hence a higher concentration of cDNA for the treated samples would be necessary to reach the same amount in the qPCR. However, the cDNA synthesis product cannot be too concentrated, because of the high concentrations of reagents in the undiluted product. The highest concentration possible of cDNA, according to the results from the pilot study, was used in the qPCR. From the results of the pilot study the highest concentration of cDNA possible to use in the qPCR was between 5-10 ng/µl.

The effects of using too large concentration of undiluted cDNA could be seen in the undiluted, 50ng/µl reactions and some effects in the 25ng/µl reactions. Those concentrations did not give reliable result and could not be used as data points in the standard curve calculations. The two genes, for which a standard curve was received, were well expressed genes with reliable results for the lowest concentrations. Manufacturer guarantees that the assays efficiencies are close to 100% (±10), which the result for the two assays, with reliable result, show is true. Since the manufacturer guarantees 100% efficiency for all their assays (Applied Biosystems Application Note, 2006) 100% efficiency were assumed for all assays used in the data analysis.

In the correlation plots in Figure 5 the three reference genes are compared to each other to control the stability of their expression. A trend can be seen that low expression in one gene correlate to low expression in another and high expression in one correlate to high in another. These results indicate that the expressions of the reference genes correlate and could be used for

normalization of biological variation. There is however a shift in the expression for ACTB compared to the other two reference genes. ACTB seem generally to be more expressed than the other two, however, the correlation with the other two still remains. From the analysis of the correlation plots the data analysis of the qPCR result was made on a combination of the three reference genes for normalization. The use of a combination reduces the effects of individual samples reference gene expression. A combination of these three reference genes also is expected to represent an average for all tissue types in the biopsies analyzed. The biopsies consist of epithelial cells but also muscle cells and immunological cells. EPCAM was chosen especially as a reference gene stably expressed in epithelial cells.

In the analysis of reference target stability in the qBase^{PLUS} software the parameters coefficient of variation (CV) and expression stability value (M) were just above the set values. In a larger validating study other reference genes might be considered and validated. Especially ACTB would probably be more carefully validated since it was the least stably expressed reference gene in this experiment. Investigated target genes of interest are also genes in close connection with actin and the expression of the target genes may not be independent of the expression of ACTB, which could create a bias.

5 Conclusions

Of the five genes investigated in this study (INSR, NTS, TIPRL, ACTN1 and LPP) all seem to be somewhat connected to celiac disease, as the GWAS previously pointed toward. The insulin receptor and neurotensin had a lower mRNA expression in the individuals with celiac disease than the controls. This result suggests that these individuals seem to have some sort of defect or change in their signaling response to nutrition levels or glucose homeostasis, since both NTS and INSR are involved in these signaling pathways. It would be interesting to compare expression levels with blood insulin and glucose levels in individuals with celiac disease and control individuals, to investigate possible correlated differences.

A difference in mRNA levels could also be seen for ACTN1, LPP, INSR and TIPRL when ex vivo stimulation with insulin on nutrient deprivation was performed, changes could not be seen without the stimulation. All genes investigated are involved in glucose homeostasis or can be connected to pathways involved in glucose homeostasis. Indicating that a different expression pattern of genes connected to glucose homeostasis or cellular responses to nutrient levels could be connected to the disease. Further studies are needed, on a larger sample material and additional reference genes, to confirm these differences in expression levels and more studies on these signaling pathways is warranted to fully understand what triggers the disease. Studies like these are one piece when unwinding the mystery of why certain individuals suffer from celiac disease.

Acknowledgements

I would especially like to thank my main supervisor Åsa Torinsson Naluai for a very rewarding time, for constantly teaching me new things and for many interesting conversations, which always formed new thoughts and questions. Big hugs and thanks to my other two supervisors. Christofer, for letting me figure things out myself, trusting me in order for me to make my own mistakes and for all the talks about anything and everything. Hannah, thanks for everything, if only everyone could have their best friend at work every day.

I would also like to thank Frida Abel, manager at Genomics, for including me in the daily work at Genomics, for letting me take part in trainings and courses and learn more than I ever expected. Thanks also to Maria Nethander och Arvid Sondén for their time and support during my struggles with learning R. Last but not least thanks to all the staff at Genomics for making me look forward to go to work, for always taking time to answers questions and making be part of the group.

Bibliography

Abadie V, Sollid L M, Barreiro L B, Jabri B (2011). Integration of Genetic and Immunological Insights into a Model of Celiac Disease Pathogenesis. *Annu. Rev. Immunol*, 29: 493-525.

Applied Biosystems Application Note (2006). Amplification Efficiency of TaqMan® Gene Expression Assays", publication 127AP05.

Béraud-Dufour S, Abderrahmani A, Noel J, Brau F, Waeber G, Mazella J, Coppola T (2010). Neurotensin is a regulator of insulin secretion in pancreatic beta-cells. *The International Journal of Biochemistry & Cell Biology*, 42(10): 1681-1688.

Bergkvist A, Rusnakova V, Sindelka R, Andrade Garda J M, Björn Sjögreen (2010). Gene expression profiling – Clusters of possibilities. *Methods*, 50:323–335

Ferguson A, Arranz E, O'Mahony S, (1993). Clinical and pathological spectrum of coeliac disease-active, silent, latent, potential. *Gut*, 34(2):150-151

Foster L J, Rudich A, Talior I, Patel N, Huang X, Furtado M L, Bilan PJ, Mann M, Klip A (2006). Insulin-dependent Interactions of Proteins with GLUT4 Revealed through Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)*. *Journal of Proteome Research*, 5(1):64-75.

Ginzinger D G (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Experimental Hematology*, 30:503-512.

Greco L, Romino R, Coto I, Di Cosmo N, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R, D'Agate C, Tinto N, Sacchetti L, Tosi R, Stazi M A (2002). The first large population based twin study of coeliac disease. *Gut*, 50(5):624-628.

Green P H R, C Cellier (2007). Celiac Disease. *The New England Journal of Medicine* , 357:17 1731-1743.

Green P H R (2005). The Many Faces of Celiac Disease: Clinical Presentation of Celiac Disease in the Adult Population. *Gastroenterology*, 128:74-78.

Guvakova M A, Adams J C, Boettiger D (2002). Functional role of alpha-actinin, PI 3-kinase and MEK1/2 in insulin-like growth factor I receptor kinase regulated motility of human breast carcinoma cells. *J Cell Sci,* 115(Pt 21):4149-4165.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*, 8:R19.

Hirschhorn J N, Daly M J (2005). Genome-wide association studies for common diseases and complex traits. *Nature Review Genetics*, 6:95-108

Jacinto E, Guo B, Arndt KT, Schmelzle T, Hall MN (2001). TIP41 interacts with TAP42 and negatively regulates the TOR signaling pathway. *Mol Cell,* 8:1017–1026.

Koning F, Schuppan D, Cerf-Bensussan, Sollid L M (2005). Pathomechanisms in celiac disease. *Best Practice & Clinical Gastroenterology*, 19:373-387.

Kubista M, Andrade, J M, Bengtsson M, Forootan A, Jonák J, Lind, K et al. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine* , 27:95-125.

Lee E K & Gorospe M (2010). Minireview: posttranscriptional regulation of the insulin and insulin-like growth factor systems. *Endocrinology*, 151(4):1403-1408.

Li B, Zhuang L, Reinhard M, Trueb B (2003). The lipoma preferred partner LPP interacts with alpha-actinin. *Journal of Cell Science*, 116:1359-1366.

Livak K (1997). Relative Quantitation of Gene Expression: ABI PRISM® 7700 Sequence Detection System, Applied Biosystems User Bulletin #2 (part number 4303859)

Livak K J, Schmittgen T D (2001). Analysis of Relative Gene Expression Data Unsing Real-Time Quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method. *Methods*, 25:402-408.

Louka A S, Sollid L M (2003). HLA in coeliac diseas: Unravelling the complew genetics of complex disorder. *Tissue Antigens* , 61:105-117.

Rutledge R G, Cote C (2003). Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research*, 31, e93.

Sollid, L. M. (2002). Coeliac Disease: Dissecting a complex inflammatory disorder. *Nature Reviews Immunology* , 2:647-655. (A)

Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M (2004). Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry.* 50:509-515.

Trynka G, Wijmenga C, van Heel D A (2010). A genetic perspective on coeliac disease. *Trends in Molecular Medicine*, 16: 537-550. (E) +ORGINALARTIKELN

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology 3*, RESEARCH0034.

Vandesompele J, Kubista M, Pfaffl M W (2009). Reference gene validation software for improved normalization In: Real-time PCR: current technology and applications, Logan J, Edwards K, Saunders N A (red). Caister Academic Press: Norfolk UK.

Appendix A

AssayID:

Hs00218324_m1	DOK5
Hs00180437_m1	DCC
Hs00944352_m1	LPP
Hs00901885_m1	EPCAM
Hs00939627_m1	GUSB
Hs00998100_m1	ACTN1
Hs00357333_g1	ACTB
Hs00175048_m1	NTS
Hs00402102_m1	STK32A
Hs00179683_m1	STK32B
Hs00295580_m1	TIPRL
Hs00961554_m1	INSR