

Characterisation of gene and pathway expression in stabilised blood from children with coeliac disease

Hanna Gustafsson Bragde ,^{1,2} Ulf Jansson,³ Mats Fredrikson,² Ewa Grodzinsky,² Jan Söderman^{1,2}

To cite: Bragde HG, Jansson U, Fredrikson M, *et al*. Characterisation of gene and pathway expression in stabilised blood from children with coeliac disease. *BMJ Open Gastro* 2020;**7**:e000536. doi:10.1136/bmjgast-2020-000536

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjgast-2020-000536>).

Received 4 September 2020
Revised 5 November 2020
Accepted 21 November 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Laboratory Medicine, Region Jönköping County, Jönköping, Sweden

²Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

³Department of Paediatrics, Region Jönköping County, Jönköping, Sweden

Correspondence to

Dr Hanna Gustafsson Bragde; hanna.gustafsson.bragde@rjl.se

ABSTRACT

Introduction A coeliac disease (CD) diagnosis is likely in children with levels of tissue transglutaminase autoantibodies (anti-TG2) >10 times the upper reference value, whereas children with lower anti-TG2 levels need an intestinal biopsy to confirm or rule out CD. A blood sample is easier to obtain than an intestinal biopsy sample, and stabilised blood is suitable for routine diagnostics because transcript levels are preserved at sampling. Therefore, we investigated gene expression in stabilised whole blood to explore the possibility of gene expression-based diagnostics for the diagnosis and follow-up of CD.

Design We performed RNA sequencing of stabilised whole blood from active CD cases (n=10), non-CD cases (n=10), and treated CD cases on a gluten-free diet (n=10) to identify diagnostic CD biomarkers and pathways involved in CD pathogenesis.

Results No single gene was differentially expressed between the sample groups. However, by using gene set enrichment analysis (GSEA), significantly differentially expressed pathways were identified in active CD, and these pathways involved the inflammatory response, negative regulation of viral replication, translation, as well as cell proliferation, differentiation, migration, and survival. The results indicate that there are differences in pathway regulation in CD, which could be used for diagnostic purposes. Comparison between GSEA results based on stabilised blood with GSEA results based on small intestinal biopsies revealed that type I interferon response, defence response to virus, and negative regulation of viral replication were identified as pathways common to both tissues.

Conclusions Stabilised whole blood is not a suitable sample for clinical diagnostics of CD based on single genes. However, diagnostics based on a pathway-focused gene expression panel may be feasible, but requires further investigation.

INTRODUCTION

Small intestinal biopsies have long been an essential part of coeliac disease (CD) diagnostics, but based on the recommendations from the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition in 2012, the intestinal biopsy may be excluded from the diagnostic flow for

Summary box

What is already known about this subject?

- Paediatric coeliac disease (CD) diagnosis still requires sampling and evaluation of small intestinal biopsies in many cases.
- Gene expression in small intestinal biopsies differs between patients with active CD and patients with non-CD.
- Blood is generally a more accessible sampling material compared with small intestinal biopsies.
- Stabilised blood is suitable for routine gene expression-based diagnostics because transcript levels are preserved at sampling.

What are the new findings?

- Stabilised whole blood is not a suitable sample for clinical diagnostics of CD based on single genes.
- Pathways of potential interest in CD have been identified.

How might it impact on clinical practice in the foreseeable future?

- The results provide a potential starting point for the development of pathway-focused gene expression panels for CD.
- These findings add to the knowledge of CD pathogenesis.

symptomatic children with high levels (>10 times the upper reference value, URV) of tissue transglutaminase autoantibodies (anti-TG2) if additional requirements are met.¹ In symptomatic children with anti-TG2 levels that are 1–10 times the URV, an intestinal biopsy is needed to confirm or rule out CD as a diagnosis. If the biopsy sample is Marsh grade 0 or 1, the diagnosis is unclear and further investigations are recommended.¹ In symptomatic IgA-competent patients with anti-TG2 <1 times the URV, a CD diagnosis is less likely, but not ruled out.¹ Additional biomarkers would aid in diagnosing CD; we previously investigated gene expression in small intestinal biopsies by RNA sequencing and found differences in gene expression

in active CD cases versus non-CD cases that were detectable also in CD cases with no or low-grade intestinal injuries (Marsh 0–1).² However, blood is a more accessible sampling material than small intestinal biopsies that can be sampled repeatedly to both diagnose and follow-up with patients with CD. RNA sequencing of CD4+ T cells and whole genome microarray transcriptome analysis of peripheral blood mononuclear cells from mainly adult CD cases have identified genes and pathways of interest in CD.^{3–5} Because stabilised whole blood preserves transcript levels during collection, transport, and storage,⁶ the use of stabilised whole blood would facilitate the incorporation of gene expression-based diagnostics in clinical practice. We performed RNA sequencing of stabilised whole blood from paediatric patients with active CD, non-CD, and treated CD on a gluten-free diet (GFD) to identify potential CD diagnostic biomarkers and pathways involved in CD pathogenesis.

METHODS

Study subjects

Children and adolescents (<18 years of age) were referred to Ryhov County Hospital in Jönköping, Sweden with suspected CD or were followed-up after a period on a GFD to verify mucosal recovery. The patients included in this study were enrolled during the years 2009–2014, where biopsy samplings on GFD were still performed as a part of the standard diagnostic process for CD at the hospital. Most patients were referred for small intestinal biopsy due to elevated anti-TG2 regardless of symptoms. Some children with anti-TG2 levels below the URV on a gluten-containing diet (with or without selective IgA deficiency) were referred for biopsy to exclude CD. The patients provided written consent before blood and duodenal biopsy specimens were collected.

Study groups

Two study groups (table 1) were defined based on subjects on a gluten-containing diet: (1) group CDM3 with a CD diagnosis (histopathologic assessment Marsh 3B–3C and anti-TG2 >7U/mL), (2) group M0 cleared from CD (Marsh 0 and anti-TG2 ≤7U/mL). A third study group (table 1) was defined based on CD subjects at follow-up on a GFD: (3) Group CDM0 (Marsh 0 and anti-TG2 ≤7U/mL). Whole blood RNA samples were sequenced from all study subjects in all groups. Additionally, RNA samples from small intestinal biopsies from subjects in groups M0 and CDM3 were sequenced. These subjects were included as subgroups in a previous study involving RNA sequencing of small intestinal biopsies.³ Based on the previous analysis, one subject in group M0 was more similar to Marsh 3 subjects, and therefore, that subject was excluded from all analyses.

Sample collection and processing

Sera were collected using blood tubes containing polymer gel and clot activator (Becton, Dickinson and Company, Franklin Lakes, New Jersey, USA). Serum levels of anti-TG2 and IgG antibodies against deamidated gliadin were determined using EliA-kits from Thermo Fisher Scientific (Waltham, Massachusetts, USA) according to Bragde *et al.*⁷ A cut-off of 7U/mL was used.

Blood for RNA isolation was collected in Tempus Blood RNA tubes (Life Technologies, Carlsbad, California, USA), kept overnight at 4°C, and then stored at –20°C until RNA isolation. Total RNA from stabilised blood was purified using the Tempus Spin RNA Isolation Reagent kit (Life Technologies) according to the manufacturer's instructions. RNA concentrations were determined using the Qubit 2.0 Fluorometer and the Qubit RNA BR Assay Kit (Thermo Fisher Scientific). RNA integrity was assessed using the Agilent 2100 Bioanalyzer with the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa

Table 1 Descriptive statistics of the study groups

	No. of cases	Age at biopsy (years)*	Gender M/F	Marsh grade (span)	Anti-TG2*† (U/mL)	Anti-DG*‡ (U/mL)	HLA-DQ2.5cis§
M0	9	8.4 (2.2–17)	4/5	M0	0.29 (<0.10–1.3)	0.57 (<0.40–1.6)	67%, 22%, 11%
CDM0	10	8.8 (3.0–17)	2/8	M0	2.3 (0.10–7.0)	2.5 (0.50–5.8)	10%, 70%, 20%
CDM3	10	8.6 (2.3–16)	4/6	M3B–3C	600 (36–2858)	151 (9.0–781)	10%, 80%, 10%

Subjects in the M0 and CDM3 groups consumed a gluten-containing diet. Study subjects in group M0 did not receive a coeliac disease (CD) diagnosis, whereas a CD diagnosis was confirmed in study subjects in group CDM3. Group CDM0 contained subjects with a CD diagnosis who were followed-up after a period on a gluten-free diet.

*Mean (min–max)

†Levels of IgA autoantibodies against tissue transglutaminase (anti-TG2) in sera. For two subjects in group M0 serum outcomes were not available, but plasma outcomes were within range of the serum outcomes. Two subjects with IgA deficiency were included, one in group M0 and one in group CDM3. IgG anti-TG2 levels were measured instead, and the outcomes were within the accepted ranges for the groups (anti-TG2 ≤7 U/mL for group M0 and anti-TG2 >7 U/mL for group CDM3).

‡Levels of IgG antibodies against deamidated gliadin (anti-DG) in sera. For three subjects in group M0 and three subjects in group CDM3, no serum outcomes were available, but plasma outcomes were within range of the serum outcomes.

§The percentages of study subjects with 0, 1, or 2 human leucocyte antigen alpha chain DQA1*05 and beta chain DQB1*02 alleles (HLA-DQ2.5) in cis are accounted for in each group. HLA, human leucocyte antigen.

Clara, California, USA), according to the manufacturer's instructions. Small intestinal biopsies were collected and RNA was isolated as described in Bragde *et al.*² Blood was also collected in EDTA tubes (Becton, Dickinson and Company), and DNA was extracted from 350 µL of blood using the Biorobot EZ1 and the EZ1 DNA Blood 350 µL Kit according to the manufacturer's instructions (Qiagen, Hilden, Germany).

The intestinal biopsies from the subjects were histopathologically assessed using the modified Marsh scale^{8,9} according to the method described in Bragde *et al.*⁷ Additional assessments were available for 10 subjects that were included in a previous study.¹⁰

RNA sequencing

Libraries for RNA sequencing of stabilised whole blood were prepared using TruSeq Stranded Total RNA with Ribo-Zero Globin (Illumina, San Diego, California, USA). RNA was sequenced as described previously in Bragde *et al.*²

Genotyping

The single nucleotide polymorphism rs2187668 was used to type the human leucocyte antigen (HLA) alpha chain DQA1*05 and beta chain DQB1*02 alleles (HLA-DQ2.5) in cis.^{2,11}

Statistical analysis

RNA sequencing data were aligned and analysed using RStudio V.1.0.143.¹² The analysis flow up until differential expression calculation for individual genes was based on Law *et al.*¹³ and used functionalities from both limma (V.3.36.2)¹⁴ and edgeR (V.3.22.3).¹⁵ Sequencing reads were mapped and features (official gene symbols) were counted using RSubread V.1.30.4¹⁶ and genome build hg38 and annotations (V.85) from Ensembl.¹⁷ Entrez ID information and the name of each gene were added using the Bioconductor package AnnotationDbi.¹⁸ Genes with missing information were excluded from further analyses. Counts were converted to counts per million (cpm). Genes with expression >2cpm in at least 10 samples were included in subsequent analyses. Multidimensional scaling with distances between samples based on the top 500 genes with the largest SD between all samples was carried out using Glimma¹⁹ to visualise sample differences according to the primary condition of interest (Marsh grade) or possible confounding factors. Scaling factors were calculated based on the trimmed mean of the M values. Heteroscedasticity was removed from count data using the voom function in limma. An empirical Bayes moderated t-test, with and without adjustment for multiple testing (5% false discovery rate, FDR²⁰), was used to assess significant differences from zero for each individual contrast; thus, there was no fold-change (FC) cut-off needed. P values <0.05 after adjustment for multiple testing were considered significant.

To identify potential biological processes of interest in CD, gene set enrichment analysis (GSEA) was performed

using GSEA software V.3.0 from the Broad Institute,^{21,22} 10 000 gene set permutations, and gene set versions from 1 June 2018 ('Human_GO_bp_no_GO_ica_symbol.gmt' and 'Human_Reactome_June_01_2018_symbol.gmt'). The absolute value of the logarithm of the p value for each gene was multiplied by the direction (sign) of the FC, thus resulting in a ranked list of genes with upregulated genes in the top and downregulated genes in the bottom. The list was used as the input for the GSEA (GSEAPre-ranked), and analysis output was Gene Ontology (GO) Biological Process terms^{23,24} or Reactome pathways²⁵ containing genes overrepresented (FDR-adjusted p value <0.075) at the top or bottom of the ranked list. Each GO term or Reactome pathway was accompanied by an enrichment score, reflecting the degree to which the gene set was overrepresented at the top or bottom of the ranked list. The enrichment score was adjusted to account for differences in gene set size and in correlations between gene sets and the expression data set, and the adjusted value was termed normalised enrichment score. The results were visualised using EnrichmentMap V.3.1.0²⁶ in Cytoscape V.3.6.1²⁷ with an edge similarity cut-off (combined Jaccard and Overlap, 50/50) of 0.375. Clusters were identified and annotated using AutoAnnotate V.1.2, and the Markov clustering algorithm (clusterMaker2 V.1.3.1²⁸), and clusters were weighted based on similarity coefficients that were calculated based on overlaps between data sets. The word clouds (WordCloud V.3.1.1²⁹) that associated with the clusters were adjusted to clarify context.

Deconvolution was performed using immunoStates³⁰ within MetaIntegrator V.2.0.0 with RNA sequencing data formatted as log₂-transformed transcripts per million. The resulting cell ratios were investigated for differences between groups using the Kruskal-Wallis one-way analysis of variance by ranks in Statistica V.13.3 (Statsoft).

To compare the GSEA results based on blood samples with the results based on duodenal biopsies, small intestinal biopsy RNA sequencing data from Bragde *et al.*² were reanalysed using the same analysis pipeline as the blood data.

Power calculations were performed using ssizeRNA V.1.2.9.³¹

RESULTS

All results are based on RNA sequencing of stabilised whole blood unless otherwise specified. RNA sequencing resulted in a mean of 24.5 million reads per sample (13.1–29.5 million reads), and a mean of 10.7 million reads per sample (6.4–13.2 million reads) were mapped to exons and summarised to gene counts. A total of 56 303 genes were initially included in the analysis, but after filtering and annotation, 12 290 genes were further analysed.

Data set assessment

Unsupervised clustering suggested that gender might influence gene expression. Twenty-two genes were

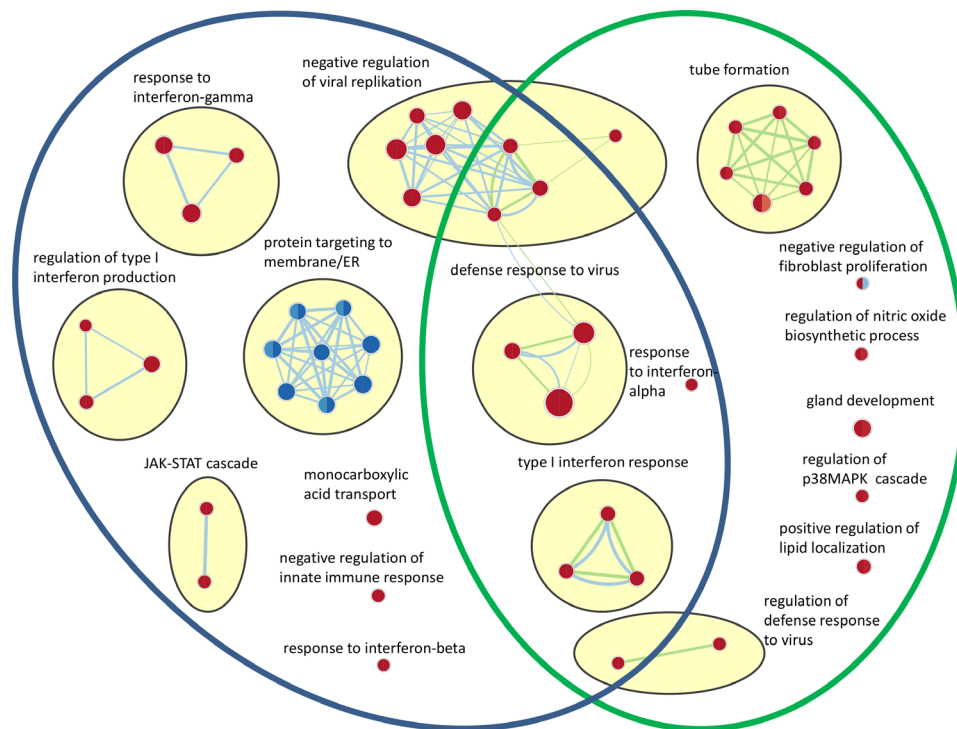


Figure 1 Visualisation of significant (FDR-adjusted p value <0.075) GO terms identified by GSEA when comparing CDM3 versus M0 and CDM3 versus CDM0. GO terms are represented by nodes coloured based on normalised enrichment scores. Darker shades represent more positive (red) or negative (blue) scores, and the node size reflects the gene set size. The nodes are divided into two halves in which the left half represents CDM3 versus M0 and the right half represents CDM3 versus CDM0. Nodes are connected by edges, and the thickness of the edges represents the overlap between gene sets (green edges for the CDM3 versus M0 comparison and blue edges for the CDM3 versus CDM0 comparison). Nodes associated with CDM3 versus M0 and CDM3 versus CDM0 comparisons have been encircled in green and blue, respectively. ER, endoplasmic reticulum; FDR, false discovery rate; GO, Gene Ontology; GSEA, gene set enrichment analysis.

identified as differentially expressed genes (DEGs) based on gender not considering CD status, and 21 of these associated with the X and Y chromosomes (online supplemental file 1). Five of the genes were identified as pseudogenes, including the autosomal gene. Differential expression and pathway analyses of CD produced similar results with and without controlling for gender. Therefore, to avoid unnecessary stratification, the results were not adjusted for gender.

Differential expression on a gene level

Even without an FC cut-off, no significant DEGs were identified in the CDM3 versus M0, CDM3 versus CDM0, and CDM0 versus M0 comparisons. Results for genes with an absolute FC difference ≥ 1.5 are presented without FDR correction (p value <0.05 ; online supplemental file 2).

Differential expression on a gene set level

Differences in gene expression in stabilised whole blood from children with active CD (CDM3), treated CD (CDM0), and non-CD children (M0) were further investigated using GSEA with GO terms or Reactome pathways. Significant differential expression differences of genes associated with GO terms (figure 1, online supplemental file 3) and Reactome pathways (figure 2, online supplemental file 4) were identified in the CDM3 versus M0 and

CDM3 versus CDM0 comparisons, but not in the CDM0 versus M0 comparison.

Eleven GO terms, related to type I interferon response and defence response to virus, were associated with higher RNA levels in the CDM3 group based on comparisons of CDM3 versus M0 and CDM3 versus CDM0. Thirteen GO terms related to ‘tube formation’, proliferation, nitric oxide biosynthesis, the p38 mitogen-activated protein kinase cascade, and lipid localisation were specific for the CDM3 versus M0 comparison and associated with higher RNA levels in CDM3. Twenty-four GO terms were specific for the CDM3 versus CDM0 comparison, and eight of these GO terms were associated with lower RNA levels in CDM3 and related to ‘protein targeting to membrane/endoplasmic reticulum (ER)’. The remaining 16 GO terms were associated with higher RNA levels in CDM3 and were related to immune response processes and monocarboxylic acid transport.

Two Reactome pathways were identified in both the CDM3 versus M0 and the CDM3 versus CDM0 comparison. Higher RNA levels in CDM3 were associated with ‘interferon alpha beta signalling’ and lower RNA levels in CDM3 associated with ‘signalling by ERBB4’. Eight Reactome pathways were identified for CDM3 versus M0, with higher RNA levels in CDM3 associated with pathway ‘formation of the beta-catenin:T cell factor (TCF)

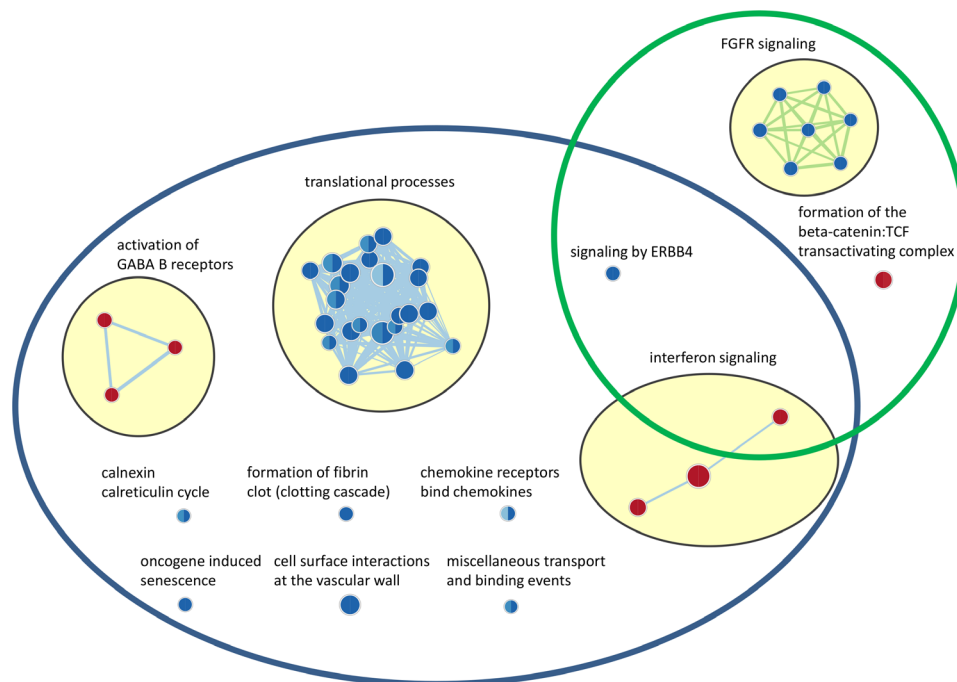


Figure 2 Visualisation of significant (FDR-adjusted p value <0.075) Reactome pathways identified by GSEA when comparing CDM3 versus M0 and CDM3 versus CDM0. Reactome pathways are represented by nodes coloured based on normalised enrichment scores. Darker shades represent more positive (red) or negative (blue) scores, and the node size reflects the gene set size. The nodes are divided into two halves in which the left half represents CDM3 versus M0 and the right half represents CDM3 versus CDM0. Nodes are connected by edges, and the thickness of the edges represents the overlap between gene sets (green edges for the CDM3 versus M0 comparison and blue edges for the CDM3 versus CDM0 comparison). Nodes associated with CDM3 versus M0 and CDM3 versus CDM0 comparisons have been encircled in green and blue, respectively. FDR, false discovery rate; FGFR, fibroblast growth factor receptors; GSEA, gene set enrichment analysis; TCF, T cell factor.

transactivating complex', and lower RNA levels in CDM3 associated with seven pathways in the cluster 'fibroblast growth factor receptors (FGFR) signalling'. Three Reactome pathway clusters were identified for CDM3 versus CDM0, and higher RNA levels in CDM3 were associated with 'activation of GABA B receptors' and 'interferon signalling', and lower RNA levels in CDM3 were associated with the cluster 'translational processes' along with several other pathways.

Gene expression in blood compared with small intestinal biopsies

GO terms identified in both blood and small intestinal GSEA (CDM3 versus M0) were associated with type I interferon response, defence response to virus, and negative regulation of viral replication (online supplemental file 3), and Reactome pathways were associated with interferon signalling and formation of the beta-catenin:TCF transactivating complex (online supplemental file 4).

Cell composition

Abundances of blood cell types inferred from deconvolution of RNA sequencing data indicated no significant differences between the groups (online supplemental file 5; p values 0.17–0.99).

DISCUSSION

On a per gene level, no significant DEGs in stabilised blood were identified in this study. To identify genes suitable for use as biomarkers, we aimed for an FC >3 or FC <-3. In a previous study of gene expression in duodenal biopsies (CDM3 versus M0) using RNA sequencing, 2.5% of the analysed genes were differentially expressed at this FC level.² Assuming 10 times fewer DEGs in blood, the analysed number of genes (12 290) should render approximately 25 DEGs at this FC level with group sizes corresponding to a power of 80%. However, no DEGs that withstood adjustment for multiple testing were found, indicating that even fewer DEGs are detectable at this FC level in whole blood using RNA sequencing. DEGs have previously been successfully identified in studies of gene expression based on purified cells from CD blood samples.^{3–5 32} This discrepancy indicates that RNA stabilised whole blood is not a suitable sampling material to identify differential expression in CD, and thus unsuitable for CD diagnostics based on a single or a few genes.

Based on GSEA, our study groups (table 1) differed significantly with respect to several Reactome pathways and GO terms. In this discussion, we chose to focus on the GO terms and Reactome pathways with the most positive/negative normalised enrichment scores because these would have the most diagnostic potential. The results suggest concomitant regulation by both



pro-inflammatory and anti-inflammatory processes. The GO term 'negative regulation of innate immune response' was enhanced in active CD when compared with treated CD, and a non-significant increase was also seen in active CD when compared with non-CD. Additionally, expression of genes associated with the Reactome pathway 'cell surface interactions at the vascular wall' was lower in active CD when compared with treated CD, and a non-significant decrease was also seen in active CD when compared with non-CD. This pathway describes key interactions between platelets and leukocytes and the endothelium in response to injury. Furthermore, clusters of GO terms/Reactome pathways related to translational processes and targeting of proteins to the membrane/ER were associated more with treated CD than active CD, possibly reflecting post-transcriptional processes in the initiation and regulation of inflammation.³³ GO terms and Reactome pathways associated with interferon type I (alpha and beta) were enriched in active CD when compared with both non-CD and treated CD, and significant differences in Reactome pathways/GO terms related to type II interferon (interferon gamma) and the JAK-STAT cascade were observed when active CD was compared with treated CD. The JAK-STAT cascade is highly involved in immune responses, many of the JAK-STAT receptors respond to cytokines,³⁴ and JAK-STAT is an important signal transduction pathway for interferon gamma.³⁵ In CD lesions, there is constitutive activation of the STAT pathway,³⁵ and the JAK-STAT pathway was identified in a transcriptome analysis of CD4+ T cells from CD cases by Quinn *et al.*³ In this study, Reactome pathways related to the activation of GABA B receptors were identified in active CD when compared with treated CD, and GABA B receptors could be involved in the active immune response because they have been shown to be expressed in neutrophils and seem to be involved in neutrophil migration by acting as chemoattractant receptors.³⁶ Additionally, our results indicated that negative regulation of viral replication and defence response to virus are associated with active CD. This is in agreement with the suggestion that the response to viral infection is a contributing factor to the onset of CD.³⁷ The GO term 'monocarboxylic acid transport' was enriched in active CD when compared with treated CD. It is of interest to note that levels of faecal short chain fatty acids are altered in children with CD,³⁸ and that short chain fatty acids have been proposed to have anti-inflammatory effects by influencing regulatory T cells.³⁹ Reactome pathway cluster 'FGFR signalling', and GO term cluster 'tube formation' and the GO term 'negative regulation of fibroblast proliferation' were associated with active CD when compared with non-CD, indicating that processes involving cell proliferation, differentiation, migration, and survival could be altered in CD.

Biological processes identified by overlapping GSEA results from blood and small intestinal biopsies included response to type I interferon, defence response to virus, and formation of the beta-catenin:TCF transactivating

complex. This indicates that, in both tissues, an immune response similar to the response found against virus infections and activation of transcription through the WNT pathway could be involved in CD.

Even if no single gene biomarker was identified, and the identified pathways were general and associated primarily with inflammation, further studies could still reveal gene expression profiles unique for CD within this biological context. In a study by Tuller *et al.*, where gene expression and protein-protein interactions in six autoimmune diseases were analysed, pathways in common for most or all diseases were identified.⁴⁰ However, there were also evidence of differential contribution of inflammation-related and more general pathways (eg, interferon and apoptosis), and also that in different diseases the same pathway (eg, apoptosis) was regulated by different mechanisms.⁴⁰

Any conclusions as to whether the results from this study can be of use for diagnostic purposes cannot be drawn due to the small groups. For this purpose larger studies are needed, including patients at different stages of the disease, and should probably be based on a different sampling material.

In conclusion, our study indicates that the use of stabilised whole blood for identification of single gene expression biomarkers in patients with CD is not optimal. However, by analysing gene expression on a pathway level, it is possible to identify significant differences in expression between active CD, treated CD, and non-CD cases. Our results indicate concomitant regulation of both pro-inflammatory and anti-inflammatory processes in CD, and highlight processes such as transport, and cell proliferation, differentiation, migration, and survival. Parallel analysis of RNA sequencing results from small intestinal biopsies and stabilised blood indicate that factors involved in the defence against viral infections are activated in both tissue types in CD. With regard to CD diagnostics, results from GO term and Reactome pathway analyses could be used to develop pathway-focused gene expression panels that could be used to predict disease status in single samples, and should be investigated further. A good starting point would be to study genes involved in pathways relating to defence against viral infections.

Acknowledgements The authors acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The authors also wish to thank all participating patients, Research Nurse Inga-Lena Hultman at the Department of Paediatrics, and the staff at the Endoscopy Department and the Surgical Department at Ryhov County Hospital, Jönköping, Sweden.

Contributors HGB and JS conceived of the study with support from UJ. HGB, JS and UJ planned the study with support from MF and EG. HGB, UJ and JS collected material for the study. HGB carried out the laboratory work, except for antibody testing, histopathologic assessment, and RNA sequencing. HGB performed the data analysis and interpretation in discussion with JS and with support from UJ, MF and EG. HGB drafted the manuscript, with help from all other authors, especially JS. All authors read and approved the final manuscript.

Funding Financial support was provided by Futurum—the Academy for Health and Care, Region Jönköping County, and by the Medical Research Council of Southeast Sweden.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The protocol was approved by the Regional Ethical Review Board in Linköping, Sweden (2011/239-31)

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Count data and relevant metadata are available from the corresponding author upon request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Hanna Gustafsson Bragde <http://orcid.org/0000-0001-9104-3863>

REFERENCES

- Husby S, Koletzko S, Korponay-Szabó IR, *et al.* European Society for pediatric gastroenterology, hepatology, and nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* 2012;54:136–60.
- Bragde H, Jansson U, Fredrikson M, *et al.* Celiac disease biomarkers identified by transcriptome analysis of small intestinal biopsies. *Cell Mol Life Sci* 2018;75:4385–401.
- Quinn EM, Coleman C, Molloy B, *et al.* Transcriptome analysis of CD4+ T cells in coeliac disease reveals imprint of Bach2 and IFN γ regulation. *PLoS One* 2015;10:e0140049.
- Sanginetto M, Graziano G, D'Amore S, *et al.* Identification of peculiar gene expression profile in peripheral blood mononuclear cells (PBMC) of celiac patients on gluten free diet. *PLoS One* 2018;13:e0197915.
- Ciccocioppo R, Panelli S, Conti Bellocchi MC, *et al.* The transcriptomic analysis of circulating immune cells in a celiac family unveils further insights into disease pathogenesis. *Front Med* 2018;5:182.
- Duale N, Brunborg G, Rønningen KS, *et al.* Human blood RNA stabilization in samples collected and transported for a large Biobank. *BMC Res Notes* 2012;5:510.
- Bragde H, Jansson U, Fredrikson M, *et al.* Potential blood-based markers of celiac disease. *BMC Gastroenterol* 2014;14:176.
- Rostami K, Kerckhaert J, Tiemessen R, *et al.* Sensitivity of antiendomysium and anti gliadin antibodies in untreated celiac disease: disappointing in clinical practice. *Am J Gastroenterol* 1999;94:888–94.
- Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol* 1999;11:1185–94.
- Bragde H, Jansson U, Jarlsfelt I, *et al.* Gene expression profiling of duodenal biopsies discriminates celiac disease mucosa from normal mucosa. *Pediatr Res* 2011;69:530–7.
- van Heel DA, Franke L, Hunt KA, *et al.* A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007;39:827–9.
- RStudio. Integrated Development for R [program]. Boston, MA URL: <http://www.rstudio.com/>
- Law CW, Alhamdoosh M, Su S, *et al.* Rna-Seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* 2016;5:1408.
- Ritchie ME, Phipson B, Wu D, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 2013;41:e108.
- Zerbino DR, Achuthan P, Akanni W, *et al.* Ensembl 2018. *Nucleic Acids Res* 2018;46:D754–61.
- AnnotationDbi: Annotation Database Interface [program]. R package version 1.42.1. 2018.
- Su S, Law CW, Ah-Cann C, *et al.* Glimma: interactive graphics for gene expression analysis. *Bioinformatics* 2017;33:2050–2.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 1995;57:289–300.
- Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Mootha VK, Lindgren CM, Eriksson K-F, *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 2003;34:267–73.
- Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. the gene ontology Consortium. *Nat Genet* 2000;25:25–9.
- The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res* 2017;45:D331–8.
- Fabregat A, Jupe S, Matthews L, *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res* 2018;46:D649–55.
- Merico D, Isserlin R, Stueker O, *et al.* Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010;5:e13984.
- Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
- Morris JH, Apeltsin L, Newman AM, *et al.* clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* 2011;12:436.
- Oesper L, Merico D, Isserlin R, *et al.* WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol Med* 2011;6:7.
- Vallania F, Tam A, Lofgren S, *et al.* Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nat Commun* 2018;9:4735.
- ssizeRNA: Sample Size Calculation for RNA-Seq Experimental Design [program]. R package version 1.2.9, 2017. Available: <https://CRAN.R-project.org/package=ssizeRNA>
- Garber ME, Saldanha A, Parker JS, *et al.* A B-cell gene signature correlates with the extent of Gluten-Induced intestinal injury in celiac disease. *Cell Mol Gastroenterol Hepatol* 2017;4:1–17.
- Mazumder B, Li X, Barik S. Translation control: a multifaceted regulator of inflammatory response. *J Immunol* 2010;184:3311–9.
- Morris R, Kershaw NJ, Babon JJ. The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci* 2018;27:1984–2009.
- Mazzarella G, MacDonald TT, Salvati VM, *et al.* Constitutive activation of the signal transducer and activator of transcription pathway in celiac disease lesions. *Am J Pathol* 2003;162:1845–55.
- Rane MJ, Gozal D, Butt W, *et al.* Gamma-Amino butyric acid type B receptors stimulate neutrophil chemotaxis during ischemia-reperfusion. *J Immunol* 2005;174:7242–9.
- Brown JJ, Jabri B, Dermody TS. A viral trigger for celiac disease. *PLoS Pathog* 2018;14:e1007181.
- Tjellström B, Stenhammar L, Högborg L, *et al.* Gut microflora associated characteristics in children with celiac disease. *Am J Gastroenterol* 2005;100:2784–8.
- Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014;12:661–72.
- Tuller T, Atar S, Ruppin E, *et al.* Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. *Genes Immun* 2013;14:67–82.