Review protocol: Evaluating the Diagnostic Accuracy of Gene Expression Profiles in Peripheral Blood Mononuclear Cells (PBMCs) for Celiac Disease: A Systematic Review

Use this template to help you plan your systematic review. See the PROSPERO database for examples of protocols for current reviews www.crd.york.ac.uk/prospero/.

Background

Describe the population and condition or phenomenon of interest and contextualize it. In other words, describe what this review is about.

Celiac (coeliac) disease, also known as gluten-sensitive enteropathy, is an autoimmune disorder caused by ingesting a specific type of protein called gluten found in wheat, barley, and rye. The gliadin molecule, which is a substructure of gluten, induces an autoimmune response by CD4+ T-cells that damages the lining of the small intestine and causes what is known as villous atrophy. This results in malabsorption and maldigestion.

Celiac disease has multiple symptoms that vary widely between patients of different ages and their exposure duration to gluten. Systemic symptoms include weight loss, persistent anemia, chronic fatigue, osteopenia, osteoporosis, and fractures.

The gold standard of diagnosing celiac disease is using an intestinal (duodenal) biopsy. The patient must undergo a gluten challenge (GC) before the procedure to see the effect of gluten on the small intestine and avoid a false diagnosis. *HLA-DQ2* or *HLA-DQ8* haplotypes are considered mandatory for someone to develop celiac disease. However, these haplotypes are found in 30-40% of the general population, and someone having these haplotypes does not necessarily mean that the disease will express. It means that they have a high chance of developing it. This makes the diagnostic process difficult. There are also regional differences in the prevalence of these genes due to the exposure to different environmental factors.

The problem with the gold standard is that it is invasive, and serological tests can give false positives due to the similarity of celiac disease with other gastrointestinal diseases, such as Crohn's disease. Moreover, subtypes of celiac disease such as asymptomatic (silent), latent, potential or refractory celiac disease are harder to diagnose due to the variation in their presentation and symptoms compared to classical or active celiac disease. It is important to note that these subtypes develop from the same autoimmune response triggered by gliadin, but their progression and presentation can vary. Silent celiac disease is difficult to diagnose, as the patient has little to no symptoms but severe intestinal damage. Latent celiac disease occurs when the patient has a previous history of celiac disease and has positive serology but no villous atrophy presentation or other tissue abnormalities is recognized. Potential celiac disease refers to someone who has positive serological tests but still hasn't had any intestinal damage; they might develop celiac disease, and this often occurs with children. Refractory celiac disease is the most severe form of celiac disease as malabsorptive symptoms and villous atrophy still persists even after going on a strict gluten free diet.

Multiple studies have analyzed gene expression differences of specific CeD-associated genes, potentially serving as a diagnostic biomarker for celiac disease. However, there is still not a definitive statement on which genes are the most promising to diagnose celiac disease reliably and what limitations need to be addressed.

This systematic review aims to assess studies examining the diagnostic accuracy of using differential gene expression profiles in PBMCs of CeD patients.

Objective

Describe the justification for this review. In other words, describe why this review/the information it collects is important.

More than 80% of celiac disease patients remain undiagnosed due to the diseases' various symptoms and its overlap with other inflammatory diseases. Till now, the gold standard of diagnosing celiac disease is invasive and is not able to diagnose subtypes of celiac disease such as latent and potential celiac disease. Moreover, Traditional serologic tests and biopsies require a gluten challenge, making the process of diagnosis stressful and potentially harmful for patients.

Over the past few years, multiple studies examined differences in gene expression profiles in PBMCs in celiac patients compared to healthy individuals and have shown great promise to be used as biomarkers in clinical tests and replacing the invasive gold standard.

The aim of this systematic review is to evaluate the diagnostic potential of leveraging differential gene expression profiles in PBMCs of Celiac disease patients by:

1. Comparing diagnostic metrics, such as sensitivity and specificity values, of the differential gene expression profiles.
2. Discussing the genes' roles in Celiac disease pathogenesis.
3. Addressing gaps in the literature regarding the current understanding of the genes' roles in the pathogenesis of the disease.
4. Discussing the limitations and future directions of using differential gene expression profiles in clinical diagnosis.

Review question

Full review question Provide the full review question in sentence format, and then break up the question according to the PICO framework (or other frameworks as appropriate).	
What is the accuracy of diagnosing celiac disease using gene expression differences in? Peripheral blood mononuclear cells (PBMCs) compared to the gold standard and serological tests?	
Population	Diagnosed celiac disease patients (with a classic, non-classic, or silent subtype) using the gold standard or high serological values
Intervention/ Exposure	Diagnostic testing based on gene expression differences in PBMCs or its subpopulations.
Comparison	Gold standard and serological tests.
Outcome	Evaluating the diagnostic accuracy of differential gene expression profile of a specific gene or combination of genes in PBMCs of celiac disease patients.

Search strategy

Databases

List the bibliographic databases to be searched.

If you are unsure look at the list of databases on the E-resources tab of your subject guide www.reading.ac.uk/library/subjects and try scoping searches with your key terms to decide which ones are relevant.

- Medline
- Cochrane
- Embase

Search terms

What are your key search terms? Pick out the key words and think about alternatives you will need to include in your search.

For guidance on constructing a search see: libguides.reading.ac.uk/database-searching

- Celiac disease
- celiac sprue
- sprue
- gene expression
- genetic expression testing
- genetic association study
- genetic association studies
- gene burden test, quantitative trait locus
- quantitative trait loci
- eQTL colocalization
- leukocyte
- leukocytes
- peripheral blood mononuclear cell
- peripheral blood mononuclear cells
- transcriptome-wide association studies

Identifying other useful sources

Outline any other ways you will find useful sources, such as hand searching contents lists of key journals, looking at relevant websites, looking at references and citations for relevant articles.

- Looking at references and citations for relevant articles

Eligibility criteria

Give more detail about your PICO (or other framework) concepts by explicitly stating what would and would not meet inclusion. For more information about this and other frameworks see libguides.reading.ac.uk/systematic-review/protocol

PICO	Inclusion Criteria	Exclusion Criteria
Population	• Celiac disease patients of any age diagnosed with the gold standard or high serological values, either have the classic, non-classic or silent subtype of celiac disease.	• Latent, potential and refractory celiac disease are excluded.
Intervention	• Differences in gene expression profiles in PBMCs or its subpopulations in celiac patients compared to healthy controls	• Gene expression differences in intestinal cells or in animals or in saliva are excluded.
Comparison	• Gold standard and serological tests such as antigliadin test, anti-endomysial test, anti-tissue transglutaminase (tTG) test, IgA-DGP test, and IgG-DGP test	• Absence of a healthy control group or solely including controls with other conditions that may affect celiac disease gene expression profile.
Outcomes	• Studies reporting the diagnostic accuracy of PBMCs gene expression for diagnosis, either through formal diagnostic metrics, (sensitivity/specificity/AUC) or classification/discriminatory modelling approaches. This also includes data permitting to calculate diagnostic accuracy metrics.	• Studies lacking quantitative diagnostic accuracy results or insufficient data to calculate diagnostic accuracy metrics

Additional limits

Outline any other limits you impose e.g. language, publication type, study design (e.g. randomised controlled trials)?

The studies that will be chosen have to be published in English but not necessarily from an English-speaking country.

Study designs include but not limited to randomized controlled trails (RCTs), genome wide association studies, mendelian randomization, observational studies, case-control studies, quasi- experimental studies, longitudinal studies, and mendelian randomization studies. The rationale for including these study designs is the presence of a diagnostic accuracy metrics, whether formal sensitivity/specificity values, or discriminatory/classification ones.

Both peer-reviewed journal articles and relevant grey literature will be included in the search strategy. Editorials and animal-only studies will be excluded.

Study quality assessment

What criteria will you use to assess methodological quality? How will quality assessment be performed?

The QUADAS-2 quality assessment tool will be applied to the included studies to assess the risk of bias after applying the inclusion and exclusion criteria.

The tool evaluates diagnostic accuracy according to 4 domains:

1. Patient selection
2. Index test
3. Reference standard
4. Flow and timing

Two reviewers will independently assess each included study using QUADAS-2. Any disagreements will be discussed and resolved by a third reviewer when needed.

Data extraction & synthesis

Outline the procedures you intend to use to analyse and summarise the study results, including whether or not you intend to carry out meta-analyses. How will you extract the data and how will it be analysed and summarised (statistical or narrative)?

The data that will be extracted include study design, diagnostic accuracy outcomes (if provided in the study or to be calculated from extractable data), author/year, sample size, control group characteristics and diagnosis, diet of celiac disease patients, technique used, genes with diagnostic power, differences in gene expression in PBMCs, *HLA* haplotype of patients, *HLA* haplotype of controls, and subtype of celiac disease patients. Narrative synthesis and thematic analysis will be performed and included studies will be compared in summary tables.

After data extraction, If studies are sufficiently similar in their metrics, meta-analysis will be conducted to determine the most statistically significant candidates for diagnosis.