Author’s response to reviews

Title: Comparison of DNA methylation profiles from saliva in Coeliac disease and non-Coeliac disease individuals

Authors:

Nerissa Hearn (nerissa.hearn@westernsydney.edu.au)
Christine Chiu (christine.chiu@mq.edu.au)
Joanne Lind (joanne.lind@mq.edu.au)

Version: 1 Date: 26 Nov 2019

Author’s response to reviews:

Shiwei Duan (Reviewer 1)
The gender is not matched in the validation cohort. The authors could perform a subgroup analysis by gender in both pilot study and the validation study.

Subgroup analysis by sex may be important in diseases where differences may exist between the sexes in terms of pathology or response to treatment. However, in the case of CD no difference in age at diagnosis or in reason for presentation and diagnosis (gastrointestinal symptoms, screening for family history, bone disease etc.) has been observed between the sexes (PMID:15764149). For this reason, the authors do not agree that stratifying the methylation HM450 array analysis by sex would be beneficial. Subgroup analysis would also require correction for multiple testing to decrease the probability of false positive claims, thereby reducing the statistical power.

For the validation cohort, the disparity in the sexes is noted in the Discussion pg. 16, line 256 “The disproportionate number of female participants in the validation cohort is another factor. While CD is more frequent in females to males (1.4 to 1) this does not account for the 75% of female participants in the validation cohort. Recruitment setting may account for this as most attendees at the Gluten Free Expos were female due to the events being marketed as food and cooking demonstrations. Recruitment in a more gender neutral setting could help to correct this imbalance. However, as no differences in age at diagnosis, presenting symptoms or response to treatment has been reported between males and females with CD [22], it is unlikely sex impacted the results.”

The methylome analysis could be redone using gender as a covariate to provide a new list for validation in the large cohort.

IlluminaBeadChip 450K array methylation analysis in the pilot cohort included sex as a covariate in the model. This has been reported in the Results and Methods (pg. 6 line 104, and pg. 19 line 338). The authors do not agree that stratifying analysis by sex is beneficial for the reasons given above.

Stephen Brown (Reviewer 2)
The cases are carefully and thoughtfully characterized, as are the controls. The methylation studies themselves have been competently done and appropriately analyzed. In the initial study group, the authors report some statistically significant methylation differences between cases and controls;
however, what is most remarkable is how few and how modest the differences were. Interestingly, the validation cohort, in which just 3 loci were analyzed with pyrosequencing, the initial results were not confirmed. Overall, I see this as an interesting and well done study, with essentially negative results. Even though results are negative, it will be of interest. I have just a couple of specific comments:
1) When speaking of underlying SNPs that affect methylation results, the authors could be a bit clearer. To clarify and highlight how underlying DNA variants that may affect methylation were addressed, the Results section on page 5, line 96 has been amended to the below.

“Methylation patterns within SMYD3 and HLADPB1 were indicative of the presence of an underlying DNA variant (Figure 2). Known DNA polymorphisms at these sites were identified using the UCSC genome browser. The rs201044038 polymorphism in SMYD3 results in the insertion of a thymine (T) between the cytosine and guanine disrupting the CpG site at cg4798314. The population frequency of rs201044038 is unknown and was not excluded during the initial methylation pre-processing when probes within known variants with a minor allele frequency greater than 1% were removed. Similarly, the rs9276 variant in HLADPB1 has an unknown population frequency and results in a substitution of the guanine leading to a loss of the cg14373797 site and methylation. No DNA variants were reported at the CpG sites within HLADQB1, SLC17A3 and ZFYVE19.
DNA sanger sequencing to genotype the rs201044038 and rs9276 polymorphisms within the pilot cohort found the frequency of rs201044038 was significantly lower in CD individuals when compared to non-affected controls (57% vs 96%, p=0.002), while the frequency of rs9276 was significantly higher in CD individuals when compared to non-affected controls (52 % vs 4%, p<0.001). Individuals who were heterozygous or homozygous for the rs246668601 or the rs9276 variants had reduced or no methylation at these sites.”

2. Along the same lines, figure 3 raises some questions. It appears that each of the 3 cg loci has a bimodal (or trimodal) distribution of methylation. It would be helpful if the authors would comment on this. If all three loci are affected by the presence of an underlying SNP, it would be very important to discuss.

There were no underlying DNA variants at the three CpG sites near SLC17A3 when searching the UCSC Genome Browser. While the methylation patterns at these sites are suggestive of two groups, the methylation levels (~20% vs ~55%) are not indicative of an underlying DNA variant. The patterns observed in Figure 2 for SMYD3 and HLADPB1 are what is expected when there are underlying DNA variants that affect the CpG sites (~0% vs ~50% vs ~100%). The following sentence has been added to the Results when discussing whether DNA variants were identified at the CpG sites of interest. We hope that this makes things clearer (pg. 7, line 141)

“No DNA variants were reported at the CpG sites within HLADQB1, SLC17A3 and ZFYVE19.”

Marco Lucarelli, BS (Reviewer 3)
The manuscript compares DNA methylation from saliva in subjects affected by coeliac disease and non-coeliac controls by using the Illumina Infinium Methylation 450 Beadchip array and pyrosequencing. The aim of the work is to identify useful epigenetic biomarkers from easily accessible saliva samples, in respect to more difficult to obtain intestinal biopsies. The work appears promising and original, since no previous methylation data obtained from saliva samples of coeliac subjects have been reported. This makes the manuscripts worth of publication although the data are still preliminary. However, some minor questions need to be addressed.

1) The Abstract suffers from some unclear wording and punctuation in the Results section.
The abstract has been amended and reads:

“Coeliac disease (CD) is an autoimmune disease characterised by mucosal inflammation in the small intestine in response to dietary gluten. Genetic factors play a key role with CD individuals carrying either the HLA-DQ2 or HLA-DQ8 haplotype, however these haplotypes are present in half the general population making them necessary but insufficient to cause CD. Epigenetic modifications, including DNA methylation that can change in response to environmental exposure could help to explain how interactions between genes and environmental factors combine to trigger disease development. Identifying changes in DNA methylation profiles in individuals with CD could help discover novel genomic regions involved in the onset and development of CD.

The Illumina Infinium Methylation450 Beadchip array (HM450) was used to compare DNA methylation profiles in saliva, in CD and non-CD affected individuals. CD individuals who had been diagnosed at least 2 years previously; were on a GFD; and who were currently asymptomatic, were compared to age and sex-matched non-CD affected healthy controls. Methylation differences within the HLA region at HLA-DQB1 were identified on HM450 but could not be confirmed with pyrosequencing. Significant methylation differences near the SLC17A3 gene were confirmed on pyrosequencing in the initial pilot cohort. Interestingly pyrosequencing sequencing of these same sites within a second cohort of CD and non-CD affected controls produced significant methylation differences in the opposite direction. Altered DNA methylation profiles appear to be present in saliva in CD individuals. Further work to confirm whether these differences are truly associated with CD is needed.”

2) The first two paragraphs of the Results (lines 85-99) and the paragraph from line 143 to line 150 could be moved to the Methods section, as description of case series.

The first part of the results section has been condensed and amended to mainly include the main findings with respect to the demographics between the two groups. Reference to the selection criteria for CD and non-affected controls has been removed and re-iterated in the Methods section. The first paragraph of the Results section now reads:

“The pilot cohort comprised of saliva samples obtained from 31 CD individuals (12 female) and 28 controls (13 female), matched for sex and age. The mean age of individuals with CD was 42 ± 15 years old and healthy controls was 37 ± 14 years old. For CD individuals, a family history of CD was reported in 39% of the group, and the mean length of time since diagnosis was 8.7 ± 6.5 years and ranged from 2.1 to 26.2 years. All individuals were Caucasian and no significant differences in BMI, smoking status, and alcohol consumption between the two groups was observed. The frequency of another autoimmune condition was higher in individuals with CD compared to controls (45.2% vs 17.9%). Demographic information is summarised in Table 1.”

3) The Authors should better describe what criteria they used to choose genes for additional experiments and, among these genes, how they choose CpG sites to be deeper analyzed. In synthesis, I think a better description of experimental planning is needed.

The following paragraph has been included in the Methods section which outlines how CpG sites were selected for pyrosequencing.

“Pyrosequencing assays were performed on 6 CpG sites within the HLADQB1 (1 CpG), SLC17A3 (3 CpG) and ZFYVE19 (2 CpG) genes to confirm the methylation status of CpG sites in the pilot cohort. These sites were selected as they did not contain underlying DNA variations, had |Δβ| >5% at the CpG site; and primers to enable amplification for pyrosequencing could be designed. All pyrosequencing assays were designed, optimised, performed, and analysed by AGRF. Percentage methylation at the select CpG sites for each sample were provided to us by AGRF. CpG sites that were confirmed as being
differentially methylated in the pilot cohort, were then quantified in the second larger validation cohort using the same pyrosequencing assays.”

4) More attention is needed to explain the differences in methylation observed in the SLC17A3 gene between the array and the sequence in the two cohorts. Have the Authors tried to sequence the samples from the first cohort? Do discrepancies depend on differences between the two cohorts or between the two methods?

Methylation differences identified from the HM450 array were first confirmed via pyrosequencing in the pilot cohort. If the differences were still present, the same pyrosequencing assays were used to measure those CpG sites in the second validation cohort. Therefore, the difference in the direction of the methylation change is unlikely to be due to discrepancies between methods (HM450 array vs Pyrosequencing). Furthermore, no underlying DNA variants at the three CpG sites were reported in UCSC Genome Browser, and no DNA variants were observed in the pyrograms (like sequence data) at and adjacent to these CpG sites after pyrosequencing. Discrepancies in the findings may be a result of the differences in the makeup of the pilot and validation cohorts. Further information has been included in the Discussion, page 16, line 250.

“The differing results observed indicates that the smaller pilot cohort was not reflective of the larger validation cohort, or vice versa. In the pilot cohort recruitment of CD and controls were similar, whereas the validation cohort had a higher proportion of CD (69%) compared to controls (31%). Furthermore, individuals with a family history of CD are at a 10x higher risk of developing CD. The exclusion criteria for non-affected controls in the pilot cohort included a family history of disease. Whereas in the validation cohort, 38.4% of non-affected controls reported a family history of CD. This may have impacted on the results and led to the discrepancies observed.”

5) Cohorts. It is not clear what is the difference between the "illumina" cohort and the "pyrosequencing" cohort. Which is the cohort described in Methods? Is the illumina cohort a subgroup selected from the larger cohort described in table 4? If so, table 4 should be described first. Moreover, if so, Authors should explain whether the "pyrosequencing" cohort also include the individuals analyzed by the Illumina or not and discuss their rationale in this experimental design. In synthesis, I think a more ordinated presentation of case series in Methods is needed, with a link to the two tables, as well as a more detailed description of what has been done on one cohort in respect to the other and its rationale.

The Methods have been amended to include the follow subheadings and descriptions of the two cohorts. We hope that this helps to clarify the differences between the two cohorts. Subsequent reference to the cohorts is differentiated by the terms “pilot” or “secondary”.

“Pilot cohort: Illumina Infinium HumanMethylation450 Beadchip array
The pilot cohort consisted of n=59 individuals consisting of CD individuals and non-affected controls. CD status was confirmed via endoscopy reports. For inclusion, these individuals had a CD diagnosis that was more than 2 years old, adhered to a strict gluten free diet since diagnosis; were free of any associated symptoms; and carried at least one HLA-DQ2 or DQ8 haplotype. For non-affected controls, these individuals reported no family history of CD, and were age (± 3 years) and sex matched to the CD group. Non-affected controls were free of any associated symptoms; carried at least one HLA-DQ2 or DQ8 haplotype; and were negative for CD using the Simtomax® CD assay, a point-of-care test that detects antibodies against deamidated gliadin peptides (HealthScreen Solutions, AUS). This is a commercially available screening tool that has a negative predictive value of 99.1% [24].

Secondary cohort: Validation group to confirm any methylation differences
An additional n=221 CD cases and non-affected controls were recruited to enable validation of any
differentially methylated CpG sites identified in the pilot cohort. These individuals were defined as having CD if they fulfilled the following criteria: had been diagnosed with CD via duodenal biopsy by a gastrointestinal specialist; were currently on a gluten free diet; and carried at least one HLA-DQ2 or HLA DQ8 haplotype. Individuals were classified as non-affected controls if they reported not having CD or CD associated symptoms and were not on a gluten-free diet (GFD).”

6) The Authors should discuss the hypothetical correlation between the methylation data and the expression of the differentially methylated genes: does the theoretical up-regulation (hypomethylation) or down-regulation (hypermethylation) have a possible role in the coeliac disease?

The following has been included into the Discussion:
“The three sites are located upstream of the SLC17A3 gene which would suggest that alterations in methylation could affect chromatin interactions and/or transcription factor binding, influencing the expression of SLC17A3 or downstream genes. Loss of function mutations in SLC17A3 have been shown to result in hyperuricemia [21] therefore, hypermethylation at these regions may result in elevated blood uric acid levels, which is consistent with elevated levels of uric acid levels seen in individuals with CD [19]…”

7) The Authors only report difference in CpG methylation, although evidence indicate that significant and functional CpN methylation may be masked by the use of PCR primers that do not allow to evidence it, with the indirect consequence of underestimating also differences in CpG methylation:


What kind of primers for illumina assay and pyrosequencing have the Authors used? Are these primers biased versus CpN methylation? These aspects should be discussed in relation to the obtained data (the indicated papers could be quoted as support).

The Illuminia Infinium HumanMethylation450 Beadchip (HM450) array provides single base resolution genotyping of targeted CpG sites using probes on a microarray. The array uses two probe types with a different chemistry which each have a different dynamic range, which has the potential to affect DNA methylation measurements. Despite this, the epigenetic research community has widely used this array, and specific bioinformatics methods (embedded in the packages used to analyse this dataset) have enable us to address the technical impact of the two probe designs.

The Qiagen Pyromark platform is utilised by the Australian Genome Research Facility in their Targeted Sites and Methylation Validation service. We used this service to validate methylation findings identified on the HM450 array. This technology is a type of NGS technique. All pyrosequencing assays were designed, fully optimised, and performed by the Australian Genome Research Facility, a commercial fee for service provider. The references supplied by the reviewer refer to issues observed in older methodologies that were either not used in this study or have been addressed via specific primer design software utilised by AGRF.

8) Discussion, line 191: Cpg -> CpG.
This error has been corrected.
9) Line 266. The term "general population" is ambiguous in this context. Does the selection concern the cohort of affected patients or control subjects? If both, the term cannot be used.

Individuals were recruited at events that were open to the general population. Attendees at these Gluten Free Expos were individuals with a CD diagnosis, their families, or members of the public with an interest in the area. The authors understand how reference to "general public" may be confusing and have removed reference to it in the first line of the Methods. This now reads: “Individuals were recruited at the annual Gluten-Free Expos in Sydney and Melbourne, Australia.”

10) Bisulfite modification. The bisulfite modification is poorly described. Authors should describe the kit used and the positive and negative controls they adopted for the bisulfite modification and subsequent sequencing protocol.

Methylation arrays, bisulfite pyrosequencing and DNA sanger sequencing were all performed by the Australian Genome Research Facility, a commercial fee for service provider. All assays were designed, optimised, performed, and analysed by the staff within the facility, therefore the exact details of the kits and processes used were not available to the authors. The Methods section: Sanger sequencing and Bisulfite pyrosequencing (page 22, Lines 407) has been amended to clarify the rationale and steps taken. This now reads

“CpG sites within genes that mapped to DMRs with a mean $|\Delta \beta| >5\%$ were further investigated. The UCSC genome browser was used to investigate whether known DNA variants were present in and around the CpG sites of interest. In cases where known DNA variants (single nucleotide polymorphisms SNPs) were reported that may alter the CpG site, and thus methylation status, DNA sequencing was used to determine whether underlying DNA variation was responsible for differences in methylation. All sequencing assays were designed, optimised, and performed by AGRF. Samples from across the methylation profile (high vs low) were sequenced. Forward and reverse sequences were provided to us by AGRF, and sequences were analysed using Sequencer v.5.4 (Genecodes, USA). Pyrosequencing assays were performed on 6 CpG sites within the HLADQB1 (1 CpG), SLC17A3 (3 CpG) and ZFYVE19 (2 CpG) genes to confirm the methylation status of CpG sites in the pilot cohort. These sites were selected as they did not contain underlying DNA variations, had $|\Delta \beta| >5\%$ at the specific site; and primers to enable amplification for pyrosequencing could be designed. All pyrosequencing assays were designed, optimised, performed, and analysed by AGRF. Percentage methylation at the select CpG sites for each sample were provided to us by AGRF. CpG sites that were confirmed as being differentially methylated in the pilot cohort, were then quantified in the second larger validation cohort using the same pyrosequencing assays.”

11) Line 365. The Authors say that Sanger sequencing was performed, with no other details, within a paragraph called "Bisulfite sequencing" including pyrosequencing. In this paragraph each sequencing method used, also if it was performed by a facility, should be described in greater detail. In addition, when in the text the Authors refer to results obtained by sequencing, which method was used should made clear. For example, in the "Results" section of the present manuscript is some Sanger sequencing results reported?

Please see response for above comment for changes to Methods section.