Title: Multi-Omics Assessment of Genetic Risk for Celiac Disease in Down Syndrome

Authors: Marisa Gallant Stahl¹,² MD MSCS, Jessica R Shaw², Neetha Paul Eduthan², Angela L Rachubinski², Keith P Smith², Belinda Enriquez Estrada², Matthew D Galbraith²,³, Ronald J Sokol¹ MD, Sameer Chavan⁴, Laura Ann Leaton⁴,⁵, Katherine M Kichula⁴,⁵, Paul J Norman⁴,⁵ PhD, Jill M Norris⁶ PhD MPH, Edwin Liu¹ MD & Joaquin M Espinosa²,³ PhD

Affiliations:

1. Digestive Health Institute, Children’s Hospital Colorado, Department of Pediatrics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States.
2. Linda Crnic Institute for Down Syndrome, University of Colorado, Aurora, CO
3. Department of Pharmacology, University of Colorado Anschutz Medical Campus.
4. Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
5. Department of Immunology and Microbiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
6. Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Correspondence:

Joaquin M. Espinosa, PhD
Linda Crnic Institute for Down Syndrome & Department of Pharmacology
University of Colorado School of Medicine
12700 East 19th Avenue, Box 4011, Aurora, CO 80045
Tel: 303-724-7389; Fax 303-724-5741
Email: joaquin.espinosa@cuanschutz.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Objectives:
Individuals with Down syndrome (DS) display high risk of celiac disease (CD), but the mechanisms underlying this increased susceptibility await elucidation. Here, we examined the prevalence of HLA genotypes associated with CD risk in the general population and tested a previously developed genetic risk score (GRS) for CD in people with DS.

Methods:
HLA genotypes were obtained for 204 individuals with DS in the Human Trisome Project cohort study, of whom 9% had CD. We compared HLA genotype frequencies in those with and without CD against frequencies observed in the general population. CD permissive HLA haplotypes explored were DQ2.5, DQ2.2, DQ8.1, and DQ7.5. We also analyzed 38 non-HLA-DQ alleles used to generate the CD GRS.

Results:
Frequencies of risk genotypes were different for CD in DS versus CD in the general population. For example, we observed lower frequency of DQ2.5/DQ2.5 and higher prevalence of DQ7.5/X and X/X in CD in DS. Although GRS values were significantly increased in those with CD and DS, their predictive power was decreased relative to the general population. Transcriptome analysis revealed dysregulated expression of many genes composing the GRS in DS. Proteomics analysis showed that GRS values correlate with elevation of specific immune factors in DS.

Conclusions:
The genetic risk profile of CD in DS is different relative to the general population, which is likely due to dysregulation of immune pathways in DS. Larger studies are needed to elucidate pathogenic mechanisms and to develop a validated GRS for CD in DS.
Keywords:
Human leukocyte antigen, trisomy 21, gluten.

Funding:
This work was supported by the NIH Office of the Director via the NIH INCLUDE Project through NIAID grant R01AI150305, and NCATS grants 5UL1TR002535 and KL2-TR002534. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional funding was provided by the Linda Crnic Institute for Down Syndrome, the GLOBAL Down Syndrome Foundation, the Anna and John J. Sie Foundation, the GI & Liver Innate Immune Program, the Human Immunology and Immunotherapy Initiative, and the University of Colorado School of Medicine.

Disclosures:
R.J.S. serves on the advisory committee for Mirum Pharma and Abireo Pharma and is a consultant for Astellas.

Abbreviations:
Celiac Disease (CD); Disomy 21 (D21); Down Syndrome (DS); Genetic Risk Score (GRS); Human Leukocyte Antigen (HLA); Interferon (IFN); Trisomy 21 (T21)
Author contributions:

Marisa Gallant Stahl: conception of the work, data interpretation, funding acquisition, drafting the manuscript, final approval of the version to be published.

Jessica R. Shaw: data analysis and interpretation, critical revision of the manuscript, final approval of the version to be published.

Neetha Paul Eduthan: data analysis and interpretation, critical revision of the manuscript, final approval of the version to be published.

Angela L. Rachubinski: IRB protocol writing and oversight, participant recruitment and consent, biospecimen acquisition, data acquisition, critical revision of the manuscript, final approval of the version to be published.

Keith P. Smith: biospecimen acquisition and processing, critical revision of the manuscript, final approval of the version to be published.

Belinda Enriquez Estrada: IRB protocol oversight, participant recruitment and consent, biospecimen acquisition, data acquisition, critical revision of the manuscript, final approval of the version to be published.

Matthew D. Galbraith: data analysis and interpretation, critical revision of the manuscript, final approval of the version to be published.

Ronald J. Sokol: conception of the work, data interpretation, funding acquisition, critical revision of the manuscript, final approval of the version to be published.

Sameer Chavan: data acquisition and analysis, critical revision of the manuscript, final approval of the version to be published.

Laura A. Leaton: Data acquisition and analysis, critical revision of the manuscript, final approval of the version to be published.

Katherine M. Kichula: Data acquisition and analysis, critical revision of the manuscript, final approval of the version to be published.

Paul J. Norman: data acquisition and analysis, critical revision of the manuscript, final approval of the version to be published.
Jill M. Norris: Conception of the work, data interpretation, critical revision of the manuscript, final approval of the version to be published.

Edwin Liu: Conception of the work, data interpretation, critical revision of the manuscript, final approval of the version to be published.

Joaquin M. Espinosa: Conception of the work, IRB protocol writing and oversight, participant recruitment and consent, data generation, analysis and interpretation, critical revision of the manuscript, final approval of the version to be published.

All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements: We would like to thank Monica Campbell, Tonya Brunetti, and the Colorado Anschutz Research Genetics Organization (CARGO) team for their assistance with preparation and analysis of the MEGA chip samples. We thank Dr. Kim Jordan at her team at the Human Immune Monitoring Shared Resource for outstanding service in generation of the immune marker dataset. We are also grateful to the Colorado Translational and Sciences Institute and the Colorado Multiple Institutional Review Board for assistance in all clinical research projects involving the Crnic Institute. Special thanks to Michelle Sie Whitten, the team at the Global Down Syndrome Foundation and Dr. John Reilly for logistical support at multiple stages of the project. We are grateful to all participants in the Human Trisome Project as well as many team members who assisted in participant recruitment, consent, biospecimen collection and processing, as well as clinical data annotation.
What is Known:

- Celiac disease is more common in individuals with Down syndrome, but the impact of HLA risk genotypes in this population is unclear.
- A celiac disease genetic risk score incorporating HLA-DQ and non-HLA SNPs has been developed with good predictive accuracy in the general population.

What is New:

- Individuals with DS may still develop CD even without the traditional HLA-DQ risk factors.
- A modified CD genetic risk score may be applied to individuals with DS with good accuracy and specificity.
- The immune dysregulation characteristic of DS involves dysregulated expression of many genes involved in CD etiology.
INTRODUCTION

Down syndrome (DS) is caused by trisomy 21 (T21), the most common chromosomal anomaly, occurring in 1 in ~700 live births \(^1\). Celiac disease (CD) is more common in DS, affecting up to 10-18% of individuals with T21, representing a ~10-fold increase relative to the general population \(^2-5\). CD may be asymptomatic in DS, and even symptomatic presentation can be difficult to identify, leading to diagnostic delays \(^6,7\). Despite the observation of increased prevalence, the underlying mechanisms driving this phenomenon remain unknown.

While human leukocyte antigen (HLA) predisposition represents the largest known contribution to the genetic risk of developing CD in the general population, whether or not the same HLA predispositions drive increased CD risk in DS is controversial. Some studies suggest that individuals with DS have the same HLA predispositions for CD as the general population \(^3,8-11\), while others suggest that individuals with DS and CD may have unique HLA genetic risk profiles \(^12-15\). Furthermore, even in the general population, HLA typing alone has poor positive predictive value \(^16-18\). Many individuals with permissive HLA genotypes will not ever develop CD. Accordingly, a genetic risk score (GRS) for CD has been generated based on genetic data including both HLA and non-HLA single nucleotide polymorphisms (SNPs) from genome-wide association studies. This GRS is more discriminative of CD risk than HLA genotyping alone \(^17\).

The best screening practices for CD in individuals with DS have yet to be defined, and methods for risk discrimination have yet to be examined in individuals with DS. In order to address the current gaps in knowledge, we investigated the HLA risk profile for individuals with DS and CD and explored if the GRS developed for the general population could be applied to individuals with DS.
METHODS

Study Participants.

All participants involved in this study (n=581) were enrolled in the Crnic Institute Human Trisome Project Biobank™ (HTP). All study activities were approved by the Colorado Multiple Institutional Review Board (COMIRB) under protocol numbers 15-2170 and 20-1977. See also clinicaltrials.gov entry NCT02864108 and www.trisome.org. Samples analyzed in this study were collected between 2016 and 2020. Written informed consent was obtained from all study participants or their legal guardians. Biospecimen collection in the HTP biobank includes a blood draw, a tongue swab, and optional urine and/or stool samples. A clinical history for each participant was curated from both medical record review and a participant/family report, with medical record taking precedence in cases of discordance between the two sources. The biological datasets described in this manuscript were generated from deidentified blood-derived biospecimens and linked demographics and clinical metadata for analysis (See Figure - Supplemental Digital Content 1).

Blood sample collection and processing.

Peripheral blood samples were collected into PAXgene RNA and DNA Tubes (Qiagen, PAXgene RNA tubes, Cat # 762165, PAXgene DNA tubes, Cat # 761115) and BD Vacutainer K2 EDTA tubes (BD, Cat # 366643). Samples in PAXgene tubes were processed for RNA-sequencing or GWAS genotyping as described below. For the plasma-based measurements of immune markers, EDTA tubes were processed within 2 hrs of blood draw by centrifugation at 700 x g for 15 min to separate plasma, buffy coat (white blood cells, WBCs), and red blood cells (RBCs), which in turn were aliquoted, flash-frozen, and stored at -80°C. Subsequent processing was carried out as described below, with aliquots selected to minimize freeze/thaw cycles.

Genotyping via Illumina Multi-Ethnic Global Array (MEGA) MEGA arrays.
Whole genome SNP genotyping was performed using the Infinium Expanded Multi-Ethnic Genotyping Array (MEGA-EX) according to manufacturer’s instructions (Illumina LLC, San Diego, CA) as previously described.

**HLA allele imputation.**

We performed HLA allele imputation using the HIBAG algorithm. SNPs specific to the pre-built MEGA imputation model of HIBAG (InfiniumMEGA-Broad-HLA4-hg19) were first extracted from the whole-genome SNP data using PLINK, and duplicate SNPs removed. Imputed HLA-DQ alleles were used to infer the presence or absence of each CD-associated DQ heterodimer (DQ2.5, DQ8, DQ2.2, DQ7.5) in each participant. “X” was used to denote any heterodimer that was not one of the four listed. From the combination of heterodimers carried, we then derived the HLA-DQ genotype carried by each participant. Imputed HLA-DQ genotypes for the 19 CD cases were subsequently validated independently. *HLA* genes were targeted for DNA sequencing using a biotinylated DNA probe-based capture method, with modifications as described. Paired ends of 250 bp each were sequenced using a NovaSeq™ 6000 instrument and SP Reagent Kit (Illumina Inc, San Diego, CA). The alleles were determined from the sequence data using the consensus from three algorithms: NGSengine® 2.10.0 (GenDX, Utrecht, the Netherlands), HLA Twin™ (Omixon Biocomputing Ltd. Budapest, Hungary) and HLA*LA. There was 100% concordance between the imputed and observed genotypes.

**Genetic Risk Score (GRS).**

Of 38 non-HLA-DQ SNPs in the GRS, 34 were directly genotyped on the MEGA chip. Genotypes for the remaining four SNPs were obtained through whole genome imputation via the Michigan imputation server (https://imputationserver.readthedocs.io/en/latest/prepare-your-data/). MEGA genotype data was prepared for whole genome imputation according to best practices, as recommended in documentation provided by the Michigan Imputation Server and using resources curated by Will Rayner at the Wellcome Trust Institute. Whole genome imputation was performed by the Michigan Imputation Server using the 1000 Genome Reference Panel (Phase 3, version 5) with reference population “Other/Mixed”.

It is made available under a CC-BY-NC-ND 4.0 International license.

[Note: The preprint is not certified by peer review]
All non-HLA-DQ SNPs in the genetic risk score were imputed with $R^2 > 0.5$ and were therefore considered eligible for downstream analyses. While the published GRS used haplotype tag SNPs to infer HLA-DQ genotypes, we used the imputed HLA-DQ genotypes as described above. For one SNP located on chromosome 21 (UBASH3A rs$^{1893592}$), we performed manual genotyping from raw intensity data from the MEGA chip. The Illuminaio R package was used to read in raw intensity data for this SNP. From a biplot of the allelic intensity ratio vs. the combined SNP intensity, we inferred four genotype clusters corresponding to 0, 1, 2, or 3 copies of the measured allele. We then inferred the dosage of the GRS allele, allowing up to three copies of the allele to be included in the GRS calculation. Following genomic imputation, SNPs required for the GRS calculation were extracted from imputed VCF files. Dosage data was then cleaned in R to reflect the dosage of GRS allele for each SNP. The GRS for each participant was then calculated using a log-additive model in which dosage of 15 HLA-DQ haplotype (0 or 1) and the dosages of 38 non-HLA-DQ SNPs (0, 1, 2 or 3) are multiplied by the corresponding log odds ratios as published previously $^{17}$ using the following formula:

$$\text{GRS score} = \log\text{OR}_{\text{HLADQ}2.5/2.5} \times \text{dosage}_{\text{HLADQ}2.5/2.5} + \ldots \log\text{OR}_{\text{HLADQn}} \times \text{dosage}_{\text{HLADQn}} + \log\text{OR}_{\text{SNP1}} \times \text{dosage}_{\text{SNP1}} + \ldots \log\text{OR}_{\text{SNPn}} \times \text{dosage}_{\text{SNPn}}.$$  

See Table - Supplemental Digital Content 2 for HLA genotypes and Table - Supplemental Digital Content 3 for non-HLA SNP data.

Whole blood transcriptome analysis.

Whole-blood RNA was extracted from PAXgene RNA Tubes and purified using a PAXgene Blood RNA Kit (Qiagen, Cat # 762164). RNA quality was assessed using a 2200 TapeStation system (Agilent) and quantified on a Qubit fluorometer (ThermoFisher Scientific). Globin RNA depletion, poly-A(+) RNA enrichment, and strand-specific library preparation were carried out using a GlobinClear kit (ThermoFisher Scientific, Cat # AM1980), NEBNext Poly(A) mRNA Magnetic Isolation Module, and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, Cat #s E7490 and E7760). Paired-end 150 bp sequencing was carried out on an Illumina NovaSeq 6000 instrument by Novogene Co., Ltd and delivered in FASTQ format.
**Profiling of plasma inflammatory markers using Meso Scale Discovery (MSD) assays.**

For each EDTA plasma sample, two technical replicates (12-25 µL depending on required dilution) were measured using the MSD multiplex immunoassay platform V-PLEX Human Biomarker 54-Plex Kit (Cat # K15248D) on a MESO QuickPlex SQ 120 instrument. Assays were performed according to manufacturer instructions and concentration values were calculated against a standard curve using provided calibrators. MSD data are reported as concentration values in picograms per milliliter of plasma and log transformed for improved visualization.

**Statistical analysis.**

Unless stated otherwise, all statistical analyses were performed in R (version 4.0.4).

The frequency of each HLA-DQ genotype was calculated among individuals with DS with and without CD. We qualitatively compared HLA genotype frequencies among CD cases and controls with DS to published frequencies observed in the general population.  

Principal components for genetic ancestry were estimated from directly observed genotypes using PLINK 1.9. Input to the principal components analysis calculation included only the samples eligible for inclusion in the GRS analysis, namely unrelated individuals with DS, known CD status, and complete genotypes for all genetic predictors in the GRS. The detailed methods as well as PCA plots colored by self-reported race and ethnicity are shown in Supplemental Methods - Supplemental Digital Content 4. The first five principal components were used as covariates in logistic regression modeling of CD cases versus controls in DS to calculate odds ratios for each SNP used in the GRS. Having derived the GRS summary metric, we evaluated its association with CD in DS and quantified its ability to predict CD in these individuals. To evaluate whether the GRS was significantly associated with CD in DS, we utilized a logistic regression model for which the outcome of CD was predicted by the GRS. From the fitted logistic regression model, we estimated a receiver operating characteristic (ROC) curve using the R package pROC (version 1.17.0.1). The Area Under the ROC (AUC) was used to
quantify the predictive capacity of the GRS in individuals with DS. This was compared to the published predictive capacity of the CD GRS in the general population \(^1\). To generate the DS-GRS, the observed odds ratios in DS were used in place of the published weights in the formula described above. See Supplementary Digital Content 2-3 for the odds ratios in DS. Differential gene expression of the included non-HLA-DQ SNPs among those in the HTP with and without DS and those with DS with and without CD was performed using R package DESeq2 (version 1.28.1) \(^2\) with age, sex and sample source as covariates. Spearman correlation analysis was used to define associations between plasma immune factors and the calculated DS-GRS using the rcorr function from the Hmisc package (v 4.5-0). Multiple comparisons were accounted for using the Benjamini-Hochberg (false discovery rate) method.
RESULTS

Individuals with Down syndrome and celiac disease display a distinct genetic risk profile.

To investigate the genetic risk profile of CD in DS, we analyzed multi-omics datasets generated by the HTP. At the time of this analysis, within the larger HTP cohort, 204 participants had genomics data available generated by Illumina MEGA arrays (see Figure - Supplemental Digital Content 1). Nineteen of these participants had a self-reported and/or medical record diagnosis of CD (see Table - Supplemental Digital Content 5).

After deriving HLA haplotypes for all participants with DS, we performed a comparison of HLA frequencies in those with and without CD relative to the frequencies observed in the general population. For those with DS without CD, the distribution of HLA-DQ frequencies did not differ significantly with respect to the unaffected general population (Figure 1A-B). However, among those with CD and DS, we observed a higher frequency of traditionally low risk HLA-DQ genotypes relative to those in the general population affected by CD. In fact, 11% of individuals with DS harbor traditionally low risk genotypes (DQ-7.5/7.5, DQ-7.5/X, and DQ X/X) compared to 3% of individuals in the general population (see Table - Supplemental Digital Content 2). Nevertheless, most (17/19) individuals with CD and DS still present with traditionally permissive genotypes. Interestingly, none of the participants with CD and DS presented DQ-2.5 homozygosity, which has been observed to be the highest risk genotype in the general population. It should be noted that the study may have been underpowered to detect DQ-2.5 homozygosity. Participants with CD and DS were most likely to have the DQ-2.2/X and DQ-2.5/X genotypes (21% and 26%, respectively). Participants with CD and DS also did not have the DQ-8/X genotype, which is why an odds ratio for this genotype is not included in Figure 1B.

Next, we examined the odds of CD among the different non-HLA risk alleles included in the GRS developed with data from the general population (see Table - Supplemental Digital Content 3). Sixteen of the included risk alleles were associated with increased odds of CD in DS, while twenty-two were associated with decreased odds of CD in DS (Figure 1C). After accounting for multiple comparisons, no individual risk variant was significantly associated with CD risk in DS. On average, individuals with CD and DS display higher GRS values relative to those with DS and no CD, albeit with
strong inter-individual variation (Figure 1D). In fact, one of the participants with CD and DS presented the lowest GRS value among the entire cohort, thus revealing the limitations of the GRS in calling CD in DS (Figure 1D). Overall, the published GRS yielded an AUC of 0.735 in DS (Figure 1E), a lesser performance relative to that observed in the general population, where it provided an AUC of 0.835 as reported by Sharp et al. With this in mind, we incorporated the observed odds ratios for each of these SNPs in DS and reweighted the GRS creating a DS-specific CD GRS (Figure 1D). Expectedly, the DS-GRS performed better than the published GRS with an AUC of 0.888 (Figure 1E, see Figure - Supplemental Digital Content 6).

Overall, these results indicated that increased risk of CD in DS is associated with a differential genetic risk profile, whereby risk alleles identified in the general population may have either lower or higher predictive value in DS.

Trisomy 21 dysregulates expression of genes involved in the celiac disease risk profile.

Next, we hypothesized that the differential genetic risk profile of CD in DS could be explained, in part, by effects of T21 on the expression of genes composing the GRS. We reasoned that if T21 either downregulates or upregulates mRNA expression for one or more of these genes, especially non-HLA risk genes, this would impact the behavior of the GRS in terms of predicting CD in DS. To test this, we analyzed whole blood transcriptome data generated through the HTP for individuals with and without DS (T21, n=304 versus D21, n=96), which enabled us to define the differential impacts of karyotype and CD diagnosis. Indeed, we observed that twenty of the non-HLA genes in the GRS were differentially expressed in individuals with DS, with 8 being significantly downregulated and 12 significantly upregulated (Figure 2A-C, see Figure - Supplemental Digital Content 7). Additionally, three of these non-HLA risk genes were also differentially expressed based on CD diagnosis in DS (i.e., PTPRK, ICOS, and ARHGAP31) (Figure 2D).

Prominent examples of CD risk genes downregulated in DS include PTPRK, which encodes a protein tyrosine phosphatase previously associated with CD risk in the general population, and which was found to be downregulated in the mucosa of active CD compared to treated patients and unaffected
controls, POU2AF1, a transcriptional cofactor expressed primarily in B cells, and CIITA, a transcriptional cofactor essential for major histocompatibility complex class II gene transcription. Other examples of downregulated genes in DS are ETS1, REL, TAGAP, LPP and DDX6. Among CD risk genes upregulated in DS, salient examples include the FAS ligand (FASLG), a key inducer of apoptosis; UBASH3A, a gene encoded on chromosome 21 involved in regulation of T cell receptor activity; and IL18RAP, which encodes an accessory subunit of the IL18 receptor, with mutations being associated with inflammatory bowel disease. Other upregulated genes in DS include STAT4, PLEK, RGS1, ARHGAP31, PTPN2, SH2B3, CSK, ZFP36L1, and UBE2L3.

Altogether, these results demonstrate that T21 affects the expression of CD risk genes, which in turn could lead to differential contributions of these genes, as well as HLA-DQ genotypes, to the genetic risk profile in DS.

**Genetic risk score associates with specific immune markers in Down syndrome.**

Next, we hypothesized that the genetic risk for CD in DS would correlate with differences in the inflammatory profile of individuals with DS. To test this, we defined Spearman correlations between the re-weighted DS-GRS and 54 inflammatory markers measured in matching plasma samples by MSD assays, using T21 samples only. This revealed that four immune markers were significantly positively correlated with the DS-GRS among individuals with T21: eotaxin (CCL11), eotaxin-3 (CCL26), MCP-4 (CCL13), and IL-27 (Figure 3A-B). Eotaxin and eotaxin-3 are chemokines that stimulate the migration of eosinophils and their role in atopic disease has been well described, but there are no previous reports of an association with CD. MCP-4 is a chemokine that has also been implicated in atopic disease and autoimmune diseases like rheumatoid arthritis, but also has not previously been associated with CD. IL-27 is a cytokine for which expression has previously been tied to the differentiation of gliadin-specific T cells. These plasma chemokine and cytokine levels were also compared for those with and without DS and those with DS with and without CD with several significant differences being observed in each of these comparisons (Figure 3C-D). Whereas eotaxin and MCP4...
are significantly elevated in the cohort with DS versus euploid controls, eotaxin 3 and IL-27 are not
\textbf{(Figure 3C)}. Furthermore, eotaxin, eotaxin 3, and MCP-4 are significantly elevated in individuals with
DS and CD versus individuals with DS and no CD \textbf{(Figure 3D)}.

Overall, these results suggest that the genetic risk of CD in DS is accompanied by a distinctive
inflammatory milieu among those with T21.
DISCUSSION

It is well established that individuals with DS are highly predisposed to multi-organ autoimmunity, as evidenced by high rates of autoimmune thyroid disease, autoimmune skin disorders, and CD, but the underlying mechanisms by which T21 causes these effects are not well understood. In the general population, CD risk stratification based on HLA-DQ genotype has been well characterized and is used in the clinic. Here, we demonstrate that individuals with CD and DS may not carry the traditionally high-risk HLA-DQ genotypes. Furthermore, applying a CD GRS developed for the general population that includes information on many non-HLA-DQ alleles, we found that while GRS values are indeed higher in those with CD and DS, the predictive accuracy of the GRS is decreased in DS. Clearly, a better understanding of such differential genetic risk profile for CD in DS may shed light on the increased prevalence of autoimmunity in this at-risk population more broadly, but also enhance our understanding of CD etiology and development of autoimmunity in general.

Previous studies have yielded mixed results with respect to HLA-DQ risk for CD in DS. Some studies demonstrated that individuals with DS carry the same typical HLA DQ-2.5, DQ-2.2, and DQ-8 predispositions for CD as the general population, while others demonstrated that individuals with CD and DS do not always carry these typical HLA-DQ predispositions. Reasons for disparate results include small sample sizes of individuals with CD and DS and different technologies for HLA-DQ typing. The largest previous study included 12 individuals with suspected CD and DS. Several studies only evaluated for DQ-2.5 and/or DQ-8 missing many of the traditionally permissive HLA-DQ genotypes such as DQ-2.2 or DQ-2.5 in the trans conformation. Our study used a well-described HLA imputation method and confirmed these imputed HLA-DQ genotypes with Next Generation Sequencing to accurately evaluate for the presence of traditionally permissive HLA-DQ haplotypes in 19 individuals with CD and DS versus 182 controls with DS and no CD, thus becoming the largest study of CD genetic risk in DS to date. In individuals without DS, HLA-DQ genotyping is often used to rule out CD in individuals on a gluten-free diet, in individuals with inconclusive biopsy results, and in at-risk populations such as those with a family history for risk stratification and screening guidance. In fact, the negative predictive value of HLA-DQ genotyping in individuals without DS approaches 100%
However, our results indicate that HLA-DQ genotyping may not have the same clinical application in individuals with DS and even those with traditionally low-risk HLA-DQ genotypes may still be at risk of developing CD. Therefore, these observations reveal the need for a customized approach for CD screening in DS.

To further explore the genetic profile of CD in DS, we tested a CD genetic risk score developed and validated in the general population composed of four HLA-DQ, five non-DQ HLA, and 33 non-HLA SNPs to our DS cohort. This GRS did not perform as well in DS, which can be explained in part by the fact that the risk associated with these SNPs is different in CD in DS versus published risk for CD in the general population (Figure 1C). In turn, such differential risk could be linked to the fact that most of these genes were dysregulated at the mRNA level in the peripheral immune compartment of individuals with DS (Figure 2). Therefore, we hypothesize that the widespread dysregulation of genes involved in CD pathogenesis in DS modulates the contributions of risk SNPs identified for these genes in the general population. Given that SNPs can affect gene transcription, mRNA processing and stability, and/or protein activity, the effects of these SNPs can be attenuated or enhanced by convergent mechanisms affecting the expression and/or activities of these same genes and their products. For example, the PTPRK gene, located on chromosome 6, has been demonstrated to play important roles in maintenance of cell-cell junctions and has subsequently been implicated in intestinal barrier function. In the general population, individuals with active CD have been shown to have reduced expression of PTPRK in intestinal biopsies in line with increased intestinal barrier permeability. Our study demonstrated that PTPRK expression was decreased in individuals with DS compared to the general population and that expression was further reduced in individuals with DS with versus without CD, all of which may contribute to different frequencies of PTPRK SNPs in DS. Given that all these risk SNPs affect genes in a small number of interconnected pathways (Figure 4), it is difficult to predict how differences in mRNA expression may impact on SNP frequency. Furthermore, we should note that our transcriptome analysis employed whole blood RNA, which may not reflect gene expression changes in the gut epithelium or other relevant tissues.
Noteworthy, individuals with DS display chronic activation of the interferon (IFN) response, a key aspect of the innate immune system, which could contribute to their increased risk of autoimmunity, including CD. Indeed, many of the pathways involved in CD pathogenesis (Figure 4) are affected by IFN signaling. Such IFN hyperactivity can be explained by the fact that four of the six IFN receptors are encoded on chr21 and widely overexpressed throughout the immune cell lineages of individuals with DS \cite{45-48}. Given that the IFN response employs JAK1-2 kinases for signal transduction, these findings justify the ongoing testing of JAK inhibition to decrease the burden of autoimmunity in this population (NCT04246372) \cite{49-51}.

While these findings are interesting, several limitations must be accounted for when interpreting these results. Our study included 19 individuals with CD and DS, which is larger than previous studies, but an even larger sample size would be needed to define statistical significance for many of our observations. Another limitation is the broad clinical inclusion criteria for CD diagnosis. Four individuals only had self-reported diagnoses of CD that were unable to be confirmed in the medical record. Of the ones that were confirmed in the medical record, several individuals were diagnosed based on positive antibody results and did not receive a confirmatory intestinal biopsy. This could lead to misclassification error and future studies should include individuals with deeply annotated clinical data as well as multi-omics data to improve the generalizability of the results. Similarly, gluten-free diet status was not available at the time of the blood draw for most participants. It is known that gene expression can vary based on whether individuals are on a gluten-free diet and whether they have active CD \cite{28}.

Despite these limitations, our study has the largest sample size to date of individuals with CD and DS and is the first to have extensive multi-omics data available for this population. In conclusion, we have demonstrated that individuals with CD and DS have a differential genetic risk profile relative to CD in the general population that also corresponds with differences in gene expression.
REFERENCES


**FIGURES and FIGURE LEGENDS**

**Figure 1.** (A) Stacked bar plot showing the HLA-DQ frequency distribution in those with and without celiac disease and with and without Down syndrome (DS). The high-risk genotypes are highlighted in...
shades of red and the traditionally low-risk genotypes are highlighted in shades of blue. (B) Bar graphs displaying the odds of celiac disease in individuals with Down syndrome enrolled in the Human Trisome Project based on the presence of specific HLA-DQ genotypes. The gray line represents the odds ratio published in Sharp, et al. for the general population. (C) Bar graphs displaying the odds of celiac disease in individuals with Down syndrome enrolled in the Human Trisome Project based on the presence of specific non-HLA-DQ SNPs. The gray line represents the odds ratio published in Sharp, et al. for individuals without Down syndrome. (D) Sina plots displaying genetic risk score (GRS) values based on celiac disease status. The plot on the left uses the published GRS developed by Sharp et al. and the plot on the right uses the GRS reweighted with the observed odd ratios in the Human Trisome Project (DS-GRS). (E) Receiver operator curves of the published celiac disease GRS (red) and the DS-GRS (blue) for individuals with Down syndrome enrolled in the Human Trisome Project.
Figure 2. (A) Heat maps showing odds ratios (OR) of celiac disease (heatmaps A and B) and fold change in gene expression (heatmaps C and D). Heat map A shows ORs (log scale) of celiac disease for indicated SNPs from published data in Sharp et al for the general euploid population (controls). Heat map B shows OR of celiac disease for the same SNPs in individuals with Down syndrome in the Human Trisome Project. Heat map C shows fold change (log2 scale) in mRNA gene expression in Down syndrome versus controls. Heat map D shows fold change in mRNA expression in Down syndrome with versus without celiac disease.
whole blood transcriptome of individuals with T21 versus euploid controls in the Human Trisome Project. Heat map D shows fold change in gene expression in individuals with Down syndrome with versus without celiac disease in the Human Trisome Project. *Denotes a q-value <0.1 (false discovery rate 10%) Gray boxes represent SNPs for which expression data was not available. (B-C) Sina plots showing differences in gene expression in the whole blood transcriptome of individuals with Down syndrome versus euploid controls. (D) Sina plots showing differences in gene expression in the whole blood transcriptome of individuals with Down syndrome with versus without celiac disease.
Figure 3. (A) Volcano plot displaying the correlation of inflammatory markers measured in plasma samples from individuals with Down syndrome versus their genetic risk score for celiac disease weighted based on SNP frequencies observed in Down syndrome (DS-GRS). (B) Scatter plots
displaying correlation between plasma levels of indicated inflammatory markers and DS-GRS values in individuals with Down syndrome. Rho and q values are derived from Spearman correlation analysis. Lines indicate linear regression fit with shaded gray area displaying the 95% confidence interval. Dots are colored by density. (C-D) Sina plots comparing levels of indicated immune factors in individuals with Down syndrome versus euploid controls (C) and in individuals with Down syndrome with versus without celiac disease (D). p values in Cand D calculated by Wilcoxon rank sum test.
**Figure 4.** Graphical abstract displaying key pathways of celiac disease pathogenesis and the impact of trisomy 21 on the expression of genes involved in each pathway. Composed with graphic elements from BioRender.

### Pathways and genes involved in celiac disease etiology

<table>
<thead>
<tr>
<th>Barrier function</th>
<th>Tissue damage</th>
<th>Antigen presentation</th>
<th>T cell activation and differentiation</th>
<th>B cell activation and maturation</th>
<th>Cytokine release</th>
<th>Cell recruitment and migration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPRK</td>
<td>FASLG</td>
<td>CIITA</td>
<td>TAGAP</td>
<td>POU2AF1</td>
<td>DDX6</td>
<td>LPP</td>
</tr>
<tr>
<td>LPP</td>
<td>HLA - DQA1</td>
<td>ETS1</td>
<td>REL</td>
<td>SH2B3</td>
<td>STAT4</td>
<td>SH2B3</td>
</tr>
<tr>
<td>ARHGAP31</td>
<td>HLA - DQB1</td>
<td>SH2B3</td>
<td>CSK</td>
<td>ZFP36L1</td>
<td>UBE2L3</td>
<td>UBE2E3</td>
</tr>
<tr>
<td>C10RF106</td>
<td>HLA-F</td>
<td>UBASH3A</td>
<td>PLEK</td>
<td>SH2B3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>HLA-DP</td>
<td>ARHGAP31</td>
<td>RGS1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICOS</td>
<td>IL18RAP</td>
<td>IL12A</td>
<td>ICOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gene expression downregulated in the peripheral immune compartment in Down syndrome**

**Gene expression upregulated in the peripheral immune compartment in Down syndrome**
Supplemental Digital Content 1 – Figure

Diagram of experimental design. Blood samples obtained from research participants in the Human Trisome Project were analyzed through a multi-omics pipeline including GWAS analysis with the Illumina MEGA array, whole blood transcriptome analysis via RNAseq, and targeted proteomics using Meso Scale Discovery assays to measure levels of inflammatory markers.
### Supplemental Digital Content 2 – Table: HLA genotype frequency by celiac status and karyotype

<table>
<thead>
<tr>
<th>HLA-DQ Genotype</th>
<th>Observed frequencies* in those with Down syndrome</th>
<th>Published frequencies** in those without Down syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq Controls (DS without celiac)</td>
<td>Freq Cases (DS with celiac)</td>
</tr>
<tr>
<td>DQ2.5/DQ2.5</td>
<td>0 (n=0)</td>
<td>0 (n=0)</td>
</tr>
<tr>
<td>DQ2.5/DQ2.2</td>
<td>0.022 (n=4)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>DQ7.5/DQ2.5</td>
<td>0.088 (n=16)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>DQ2.5/DQ8</td>
<td>0.011 (n=2)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>DQ7.5/DQ2.2</td>
<td>0.043 (n=8)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>DQ2.5/X</td>
<td>0.114 (n=21)</td>
<td>0.263 (n=5)</td>
</tr>
<tr>
<td>DQ8/DQ8</td>
<td>0 (n=0)</td>
<td>0 (n=0)</td>
</tr>
<tr>
<td>DQ2.2/DQ8</td>
<td>0.032 (n=6)</td>
<td>0.158 (n=3)</td>
</tr>
<tr>
<td>DQ7.5/DQ8</td>
<td>0.038 (n=7)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>DQ8/X</td>
<td>0.108 (n=20)</td>
<td>0 (n=0)</td>
</tr>
<tr>
<td>DQ2.2/X</td>
<td>0.108 (n=20)</td>
<td>0.211 (n=4)</td>
</tr>
<tr>
<td>DQ2.2/DQ2.2</td>
<td>0 (n=0)</td>
<td>0 (n=0)</td>
</tr>
<tr>
<td>DQ7.5/DQ7.5</td>
<td>0 (n=0)</td>
<td>0 (n=0)</td>
</tr>
<tr>
<td>DQ7.5/X</td>
<td>0.151 (n=28)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>X/X</td>
<td>0.286 (n=53)</td>
<td>0.053 (n=1)</td>
</tr>
<tr>
<td>Total (n=204)</td>
<td><strong>0.907 (n=185)</strong></td>
<td><strong>0.093 (n=19)</strong></td>
</tr>
</tbody>
</table>

**DQ2.5** = DQA1*05:01–DQB1*02:01  
**DQ2.2** = DQA1*02:01–DQB1*02:02  
**DQ8** = DQA1*03–DQB1*03:02  
**DQ7.5** = DQA1*05–DQB1*03  
**X** = None of the above annotated HLA haplotypes  
* The Crnic Institute Human Trisome Project (HTP).  
<table>
<thead>
<tr>
<th>HLA-DQ Genotype</th>
<th>Published OR* in those without Down syndrome</th>
<th>Observed OR** in those with Down syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ2.5/DQ2.5</td>
<td>144.03</td>
<td>NA</td>
</tr>
<tr>
<td>DQ2.5/DQ2.2</td>
<td>142.59</td>
<td>2.95</td>
</tr>
<tr>
<td>DQ7.5/DQ2.2</td>
<td>51.94</td>
<td>1.58</td>
</tr>
<tr>
<td>DQ2.5/DQ8</td>
<td>42.1</td>
<td>4.19</td>
</tr>
<tr>
<td>DQ2.5/X</td>
<td>36.23</td>
<td>2.79</td>
</tr>
<tr>
<td>DQ7.5/DQ2.5</td>
<td>26.05</td>
<td>0.61</td>
</tr>
<tr>
<td>DQ8/DQ8</td>
<td>18.92</td>
<td>NA</td>
</tr>
<tr>
<td>DQ2.2/DQ8</td>
<td>14.3</td>
<td>5.84</td>
</tr>
<tr>
<td>DQ2.2/DQ2.2</td>
<td>7.32</td>
<td>NA</td>
</tr>
<tr>
<td>DQ7.5/DQ8</td>
<td>4.81</td>
<td>1.21</td>
</tr>
<tr>
<td>DQ8/X</td>
<td>3.67</td>
<td>0</td>
</tr>
<tr>
<td>DQ2.2X</td>
<td>2.77</td>
<td>2.3</td>
</tr>
<tr>
<td>DQ7.5/X</td>
<td>1.6</td>
<td>0.28</td>
</tr>
<tr>
<td>DQ7.5/DQ7.5</td>
<td>1.48</td>
<td>NA</td>
</tr>
<tr>
<td>X/X</td>
<td>1</td>
<td>0.13</td>
</tr>
</tbody>
</table>
### Supplemental Digital Content 3 - Table: Non-HLA-DQ SNP risk of celiac disease by karyotype

<table>
<thead>
<tr>
<th>Putative Non-HLA-DQ Gene</th>
<th>SNP</th>
<th>Chromosome: Position</th>
<th>Allele</th>
<th>Published OR* in those without Down syndrome</th>
<th>Observed OR** in those with Down syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-B*0801</td>
<td>rs4143332</td>
<td>6:31348365</td>
<td>G</td>
<td>1.53</td>
<td>0.93</td>
</tr>
<tr>
<td>HLA-DP Intergenic</td>
<td>rs3128927</td>
<td>6:33074288</td>
<td>T</td>
<td>1.45</td>
<td>1.05</td>
</tr>
<tr>
<td>HLA-F</td>
<td>rs1611710</td>
<td>6:29828916</td>
<td>T</td>
<td>1.42</td>
<td>1.76</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>rs17264332</td>
<td>6:138005515</td>
<td>G</td>
<td>1.29</td>
<td>1.3</td>
</tr>
<tr>
<td>PTPRK</td>
<td>rs55743914</td>
<td>6:128293562</td>
<td>T</td>
<td>1.21</td>
<td>1.99</td>
</tr>
<tr>
<td>IL18RAP</td>
<td>rs990171</td>
<td>2:103086770</td>
<td>A</td>
<td>1.2</td>
<td>1.36</td>
</tr>
<tr>
<td>HLA-DP Intergenic</td>
<td>rs6937061</td>
<td>6:33079812</td>
<td>G</td>
<td>1.19</td>
<td>0.78</td>
</tr>
<tr>
<td>ICOS</td>
<td>rs1980422</td>
<td>2:204610396</td>
<td>C</td>
<td>1.19</td>
<td>1.38</td>
</tr>
<tr>
<td>SH2B3</td>
<td>rs3184504</td>
<td>12:111884608</td>
<td>T</td>
<td>1.19</td>
<td>2.17</td>
</tr>
<tr>
<td>ETS1</td>
<td>rs61907765</td>
<td>11:128391937</td>
<td>T</td>
<td>1.18</td>
<td>0.8</td>
</tr>
<tr>
<td>REL</td>
<td>rs13003464</td>
<td>2:61186829</td>
<td>G</td>
<td>1.17</td>
<td>0.94</td>
</tr>
<tr>
<td>PTPN2</td>
<td>rs11875687</td>
<td>18:12843137</td>
<td>C</td>
<td>1.17</td>
<td>1.26</td>
</tr>
<tr>
<td>UBE2E3</td>
<td>rs1018326</td>
<td>2:182007800</td>
<td>C</td>
<td>1.16</td>
<td>0.96</td>
</tr>
<tr>
<td>TAGAP</td>
<td>rs182429</td>
<td>6:159469574</td>
<td>A</td>
<td>1.16</td>
<td>1.11</td>
</tr>
<tr>
<td>POU2AF1</td>
<td>rs7104791</td>
<td>11:111196858</td>
<td>T</td>
<td>1.16</td>
<td>1.25</td>
</tr>
<tr>
<td>UBE2L3</td>
<td>rs4821124</td>
<td>22:21979289</td>
<td>C</td>
<td>1.16</td>
<td>0.96</td>
</tr>
<tr>
<td>CIITA</td>
<td>rs6498114</td>
<td>16:10964118</td>
<td>G</td>
<td>1.14</td>
<td>0.76</td>
</tr>
<tr>
<td>ZFP36L1</td>
<td>rs11851414</td>
<td>14:69259502</td>
<td>C</td>
<td>1.13</td>
<td>0.68</td>
</tr>
<tr>
<td>CSK</td>
<td>rs1378938</td>
<td>15:75096443</td>
<td>T</td>
<td>1.13</td>
<td>0.84</td>
</tr>
<tr>
<td>CCR3</td>
<td>rs7616215</td>
<td>3:46205686</td>
<td>C</td>
<td>1.12</td>
<td>0.91</td>
</tr>
<tr>
<td>IL12A</td>
<td>rs2561288</td>
<td>3:159674928</td>
<td>T</td>
<td>1.12</td>
<td>1.59</td>
</tr>
<tr>
<td>ARHGAP31</td>
<td>rs61579022</td>
<td>3:119123278</td>
<td>A</td>
<td>1.11</td>
<td>0.72</td>
</tr>
<tr>
<td>PLEK</td>
<td>rs10167650</td>
<td>2:68645560</td>
<td>G</td>
<td>0.92</td>
<td>0.72</td>
</tr>
<tr>
<td>C1orf106</td>
<td>rs10800746</td>
<td>1:200881392</td>
<td>T</td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>IRF4</td>
<td>rs1050976</td>
<td>6:408079</td>
<td>T</td>
<td>0.89</td>
<td>0.62</td>
</tr>
<tr>
<td>IL12A</td>
<td>rs1353248</td>
<td>3:159623559</td>
<td>T</td>
<td>0.88</td>
<td>0.96</td>
</tr>
<tr>
<td>PRKCKQ</td>
<td>rs947474</td>
<td>10:6390450</td>
<td>G</td>
<td>0.88</td>
<td>0.5</td>
</tr>
<tr>
<td>UBASH3A</td>
<td>rs1893592</td>
<td>21:43855067</td>
<td>C</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>MMEL1</td>
<td>rs4445406</td>
<td>1:2539400</td>
<td>C</td>
<td>0.87</td>
<td>0.92</td>
</tr>
<tr>
<td>FASLG</td>
<td>rs11801183</td>
<td>1:172862984</td>
<td>T</td>
<td>0.87</td>
<td>1.16</td>
</tr>
<tr>
<td>FASLG</td>
<td>rs12068671</td>
<td>1:172681031</td>
<td>C</td>
<td>0.86</td>
<td>1.21</td>
</tr>
<tr>
<td>ZMIZ1</td>
<td>rs1250552</td>
<td>10:81058027</td>
<td>G</td>
<td>0.86</td>
<td>1.23</td>
</tr>
<tr>
<td>DDX6</td>
<td>rs10892258</td>
<td>11:118579865</td>
<td>A</td>
<td>0.86</td>
<td>0.44</td>
</tr>
<tr>
<td>STAT4</td>
<td>rs6715106</td>
<td>2:191913034</td>
<td>G</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>RGS1</td>
<td>rs1323292</td>
<td>1:192541021</td>
<td>G</td>
<td>0.77</td>
<td>1.31</td>
</tr>
<tr>
<td>LPP</td>
<td>rs2030519</td>
<td>3:188119901</td>
<td>G</td>
<td>0.76</td>
<td>0.69</td>
</tr>
<tr>
<td>IL2</td>
<td>rs13132308</td>
<td>4:123551114</td>
<td>G</td>
<td>0.71</td>
<td>0.7</td>
</tr>
<tr>
<td>HLA-DP Intergenic</td>
<td>rs2301226</td>
<td>6:33034596</td>
<td>G</td>
<td>0.69</td>
<td>1.53</td>
</tr>
</tbody>
</table>


**The Crnic Institute Human Trisome Project (HTP).

The OR for FASLG from the Sharp publication represents the OR included rs12124480, which was the SNP included in the original publication.
Supplemental Digital Content 4 – Supplemental Methods.

Principal components for genetic ancestry.

Excluded from input to principal components analysis calculation were chromosome 21, sex chromosomes, mitochondrial variants, variants with unknown chromosomal mapping, and the specific non-HLA-DQ SNPs used in the GRS. Also excluded from the principal components analysis calculation were ancestry-specific regions which may confound the analysis \(^1,^2\) including chr8:8135000-1200000, chr17: 40900000-45000000, and the HLA region plus or minus 200 kb (chr6:28277797-33648354).

Because HLA alleles have been demonstrated to strongly associate with variants up to 200 kb outside of the formal boundaries of the HLA region \(^3,^4\), we excluded variants within the HLA region as well as variants located within 200 kb of the HLA region boundaries. We further filtered input variants to common variants with a minor allele frequency >5% (Plink --maf 0.05), sample call rate > 95% (--mind 0.05), and variant call rate > 98% (Plink --geno 0.02). Lastly, we performed SNP pruning to identify a subset of pairwise independent variants with a maximum pairwise correlation of 0.20 (Plink --indeppairwise 50 5 0.2). The first 5 principal components (PCs) for ancestry were used as covariates in single-variant logistic regression models to account for population stratification (Supplemental Methods Figure 1).

**Supplemental Methods Figure 1.** Principal component analysis plots displaying the population stratification by race and ethnicity among participants in the GWAS study of the Human Trisome Project.
SNP calling for UBASH3A.

In order to identify the probe identifier corresponding to the UBASH3A SNP of interest (rs1893592), we downloaded the MEGA product manifest from the Illumina website:

https://webdata.illumina.com/downloads/productfiles/multiethnicglobal-8/v1-0/infinium-multi-ethnic-global-8-d1-csv.zip. Using the illuminaio R package, we read the raw intensity data into R and extracted a matrix of data from each idat file. We then mapped the row names of the raw intensity data to the AddressA_ID column of the Illumina product manifest to identify that AddressA_ID of ‘29800833’ corresponded to the chr21 SNP of interest, UBASH3A\textsuperscript{rs1893592}. We identified the TOP/BOT (A/B) alleles for the probe using the strand report for the MEGA product provided on the Illumina website:

https://webdata.illumina.com/downloads/productfiles/multiethnicglobal-8/v1-0/multi-ethnic-global-8-d1-strand-report.zip. Having identified the probe specific to the SNP of interest and the A and B alleles for that probe, we next derived the Combined SNP intensity (R) and the Allelic Intensity Ratio (θ) using the equations:

\[ R = \text{Mean intensity}_{\text{Red}} + \frac{\text{Mean intensity}_{\text{Grn}}}{\pi} \]

\[ \theta = \left(\frac{2}{\pi}\right) \arctan\left(\frac{\text{Mean intensity}_{\text{Grn}}}{\text{Mean intensity}_{\text{Red}}}\right) \]

We then produced scatterplots of θ versus R and visually identified thresholds of θ that clearly separated the samples into four genotype clusters (Supplemental Methods Figure 2). We then used the identified theta thresholds to manually assign a genotype for each sample.

Supplemental Methods Figure 2. Scatter plot demonstrating the genotype assignment of the SNP UBASH3A\textsuperscript{rs1893592}, which was the only SNP located on chromosome 21 in the genetic risk score. Given its chromosomal location, this plot demonstrates how the trisomy was accounted for in the genetic risk score.
Supplemental Methods References


Supplemental Digital Content 6 – Figure

Comparison of GRS algorithms to discriminate celiac disease in Down syndrome. GRS values were calculated as described in the Methods using a published algorithm developed with data from the general population (Sharp et al, 2020) or with a re-weighted algorithm (DS-GRS) using data from individuals with Down syndrome in the Human Trisome Project. Plot displays the comparative predictive power of the two versions of the GRS. The vertical black line indicates the value at which the DS-GRS correctly identifies 50% of celiac disease cases.
Supplemental Digital Content 7 – Figure

A. mRNA expression downregulated in Down syndrome versus controls

- ETS1
- REL
- TAGAP
- LPP
- DDX6

B. mRNA expression upregulated in Down syndrome versus controls

- STAT4
- PLEK
- RGS1
- ARHGAP31
- PTPN2
- SH2B3
- CSK
- ZFP36L1
- UBE2L3

Whole blood transcriptome analysis reveals dysregulation of multiple genes involved in the celiac disease genetic risk score. Sina plots displaying the expression of mRNAs that are either significantly downregulated (A) or upregulated (B) in individuals with Down syndrome versus euploid controls in the Human Trisome Project. Differential expression was calculated using DeSeq2 software.