

## Methods – Figure 1

Genomic data and celiac status were available for 201 individuals with T21 from the Human Trisome Project (HTP), including 19 individuals with CD and 182 without CD. Celiac diagnosis was ascertained from participant surveys and medical records.

Of 38 non-HLA-DQ SNPs in the GRS, 34 were directly genotyped on the Illumina Multi-Ethnic Global Array (MEGA). Genotypes for the remaining four SNPs were obtained through whole genome imputation via the Michigan imputation server (ref).

While the published GRS used haplotype tag SNPs to infer HLA-DQ genotypes, we performed HLA allele imputation using the HIBAG algorithm (ref), which predicts HLA alleles from whole genome genotyping data. 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. From the combination of heterodimers carried, we then derived the HLA-DQ genotype carried by each participant. Imputed HLA-DQ genotypes for Celiac cases were subsequently validated through molecular HLA genotyping, with 100% concordance between the imputed and observed genotypes.

The frequency of each HLA-DQ genotype was calculated among DS individuals with and without CD. We qualitatively compared HLA genotype frequencies among CD cases and controls with DS to published frequencies observed in the typical population.<sup>2</sup>

Principal components for genetic ancestry were estimated from directly observed genotypes using Plink1.9. Input to the PCA calculation included only the samples eligible for inclusion in the GRS analysis, namely unrelated individuals with Down syndrome, known Celiac status, and complete genotypes for all genetic predictors in the GRS. Excluded from input to PCA 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 PCA calculation were ancestry-specific regions which can confound PCA (Krebs et al., 2020, Lam et al., 2015), including chr8:8135000-1200000, chr17: 40900000-45000000, and the HLA region plus or minus 200kb (chr6:28277797-33648354). Because HLA alleles have been demonstrated to strongly associate with variants up to 200kb outside of the formal boundaries of the HLA region (Lam et al., 2013, de Bakker et al., 2006), we excluded variants within the HLA region as well as variants located within 200kb 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 --indep-pairwise 50 5 0.2). The first 5 PCs for ancestry were used as covariates in single-variant logistic regression models to account for population stratification. PCA plots colored by self-reported race and ethnicity are shown in Supplemental Figure **TBD**.

For one GRS SNP located on chromosome 21, we performed manual genotyping from raw intensity data. Study identifiers were mapped to experiment barcodes using the SatrixBarcode\_A column of the HTP sample manifest. In order to identify the probe identifier corresponding to the SNP of interest, we downloaded the MEGA product manifest from the Illumina website (<https://webdata.illumina.com/downloads/productfiles/multiethnic-global-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 rownames 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, 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/multiethnic-global-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 ( $\theta$ ) using the equations:

$$R = \text{Mean intensity}_{\text{Red}} + \text{Mean intensity}_{\text{Grn}}, \text{ and}$$

$$\theta = \left(\frac{2}{\pi}\right) * \arctan\left(\frac{\text{Mean intensity}_{\text{Grn}}}{\text{Mean intensity}_{\text{Red}}}\right).$$

We then produced scatterplots of  $\theta$  versus R and visually identified thresholds of  $\theta$  that clearly separated the samples into four genotype clusters (**Supplemental Figure TBD**). We then used the identified theta thresholds to manually assign a genotype for each sample. From the assigned genotype clusters, we then defined the Dosage variable as the count of the score allele used in the GRS for the SNP.

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 (Michigan Imputation Server, Rayner et al.). 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". All non-HLA-DQ SNPs in the GRS were imputed with  $R^2 > 0.5$  and were therefore considered eligible for downstream analyses.

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 score allele for each SNP. The GRS for each participant was then calculated using the dosage of each HLA-DQ genotype, the dosage of each non-HLA-DQ SNP, and the weight of each genetic predictor as published in Sharp et al., 2019.

Having derived the GRS summary metric, we evaluated its association with Celiac disease in trisomy 21 and quantified its ability to predict Celiac in these individuals. To evaluate whether the GRS was significantly associated with CD in DS, we fit a logistic regression model for Celiac status versus the GRS. From the fitted logistic regression we estimated a receiver operating characteristic (ROC) curve. The Area Under the Curve (AUC) was used to quantify the predictive accuracy of the GRS in individuals with

DS and facilitate comparison of the GRS's performance in DS compared to in the typical population for which it was developed.

In addition to analyses of the GRS, we conducted variant-level analyses to identify whether any genetic predictor in the aggregate GRS was a significant independent predictor of CD in DS. For each HLA-DQ genotype or non-HLA-DQ SNP, we fit a separate logistic regression model for CD status versus the dosage of the GRS predictor, with adjustment for the top 5 PCs for ancestry. Possible dosage values for HLA-DQ genotypes were 0 or 1, while possible values of dosage for non-HLA-DQ SNPs were 0, 1, or 2. Following model fitting, we calculated adjusted p-values for variant dosage using the Benjamini and Hochberg method.

Lastly, we performed analyses of HLA-DQ allele group dosages (2-digit HLA alleles). Whereas 4-digit HLA allele codes denote the specific protein encoded by the allele, 2-digit HLA-DQ alleles groups typically correspond to a specific antigen (REF – Kate Waugh?). For each HLA-DQA1 and HLA-DQB1 allele observed in the HTP, we fit a separate logistic regression for Celiac status versus the dosage of the two-digit allele group, again with adjustment for the top 5 PCs for ancestry. Given that some HLA-DQ alleles have been reported to have non-additive genetic associations, we also performed analyses of HLA-DQ allele group dosage as a categorical predictor with reference level of 1. Variants with a significant associations of dosage of 0 compared to 1, but no significant association of dosage 2 compared to 1, were considered suggestive of a possible dominant mode of inheritance. Meanwhile, variants with a significant association of dosage 2 compared to 1, but no significant association of 0 compared to 1, were defined as suggestive of a recessive mode of inheritance. Further research with larger sample sizes will be needed for conclusive findings regarding non-additive genetic associations of DQ allele groups in DS.

## References

- MHC Region of the Human Genome - Genome Reference Consortium*.  
<https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC?asm=GRCh37>. Accessed 24 Feb. 2022.
- Krebs, Kristi, et al. "Genome-Wide Study Identifies Association between HLA-B\*55:01 and Self-Reported Penicillin Allergy." *American Journal of Human Genetics*, vol. 107, no. 4, Oct. 2020, pp. 612–21. *PubMed*, <https://doi.org/10.1016/j.ajhg.2020.08.008>.
- Lam, Tze Hau, et al. "Intrahaplotypic Variants Differentiate Complex Linkage Disequilibrium within Human MHC Haplotypes." *Scientific Reports*, vol. 5, no. 1, Nov. 2015, p. 16972. *www.nature.com*, <https://doi.org/10.1038/srep16972>.
- de Bakker, P. I. W. et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38, 1166–1172, doi: 10.1038/ng1885 (2006).
- Lam, T. H., Shen, M., Chia, J. M., Chan, S. H. & Ren, E. C. Population-specific recombination sites within the human MHC region. *Heredity*, doi: 10.1038/hdy.2013.27 (2013).
- Data Preparation - Michigan Imputation Server.  
<https://imputationserver.readthedocs.io/en/latest/prepare-your-data/>. Accessed 29 Apr. 2022.
- Rayner, William et al. Strand Home. <https://www.well.ox.ac.uk/~wrayner/strand/>. Accessed 29 Apr. 2022.
- Bajaj, Deepak, et al. "A Combinatorial Approach of Comprehensive QTL-Based Comparative Genome Mapping and Transcript Profiling Identified a Seed Weight-Regulating Candidate Gene in Chickpea." *Scientific Reports*, vol. 5, no. 1, Mar. 2015, p. 9264. *www.nature.com*, <https://doi.org/10.1038/srep09264>.
- <https://webdata.illumina.com/downloads/productfiles/multiethnic-global-8/v1-0/infinium-multi-ethnic-global-8-d1-csv.zip>
- <https://webdata.illumina.com/downloads/productfiles/multiethnic-global-8/v1-0/multi-ethnic-global-8-d1-strand-report.zip>