UNIVERSITY OF CALIFORNIA

Los Angeles

Association of Imputed Human Leukocyte Antigen Genotypes in Frontotemporal Dementia and Cognitive Impairment

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Human Genetics

by

Misty Knight

2024

ABSTRACT OF THE THESIS

Association of Imputed Human Leukocyte Antigen Genotypes in Frontotemporal Dementia and Cognitive Impairment

by

Misty Knight

Master of Science in Human Genetics

University of California, Los Angeles, 2024

Professor Jessica Rexach, Co-Chair

Professor Bogdan Pasaniuc, Co-Chair

Frontotemporal Dementia (FTD) is a progressive, terminal neurodegenerative disorder with a strong genetic component in individual risk. Prior FTD genetic-wide association studies (GWAS) have identified several disease-associated single nucleotide polymorphisms (SNPs) including lead SNPs adjacent to HLA loci, one of which has also been implicated in immunosenescence. As HLA is a major component of immune activation, understanding the role of HLA genetic variation in FTD pathology and senescence may lead to uncovering causal roles

of immune pathways in neurodegeneration. We hypothesize that genetic variation in HLA gene or gene-regulatory regions is driving HLA expression and immune response in disease. To support this hypothesis, we seek to define HLA haplotypes and alleles associated with clinical outcomes and disease pathologies. We use a new but proven methodology that can overcome the challenges of studying the most polymorphic gene region in the human genome that has yet to be applied to FTD cohorts.

The thesis of Misty Knight is approved.

Manish J. Butte

Noah A. Zaitlen

Bogdan Pasaniuc, Committee Co-Chair

Jessica Rexach, Committee Co-Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

In this project, we set out to understand the role of human leukocyte antigen (HLA) gene variation in Frontotemporal Dementia (FTD) and cognitive outcomes in senescence by determining disease-associated genetic variation. To do this, we leverage genetic data of FTD patients from the ARTFL LEFFTDS Longitudinal Frontotemporal Lobar Degeneration (ALLFTD) research study [1] as well as genetic data from the Hillblom longitudinal aging study [2] to impute HLA alleles and haplotypes, then identify associations with clinical outcomes. We sought out to define disease-associated variants as well as those that may be associated with neuro-protective outcomes.

FTD refers to a group of Alzheimer's Disease (AD) related dementias that primarily affect the frontal and temporal lobes. The disease is progressive, terminal and the most common dementia in people under 60 and familial in up to 50% of patients, with men and women almost equally affected [3-5]. The main clinical variants of FTD are behavioral (bvFTD) and language (PPA) variants that cause changes in personality and behavior and language impairment, respectively. Additional variants include Progressive Supranuclear Palsy (PSP), Corticobasal Syndrome (CBS) and concomitant Amyotrophic Lateral Sclerosis (ALS, also known as Lou Gehrig's disease). These clinical variants are further confounded by different molecular pathologies and clinical

symptoms [6]. The heterogeneity of this disease contributes to the many challenges in studying FTD. Currently, FTD disease etiology is not well understood and there is no curative treatment.

Prior genetic-wide association studies of FTD have identified several potential risk loci, including HLA-DRA/DRB5 [5]. HLA-DR has also been found to have decreased expression in $CD14^{+}CD16^{+}$ monocyctes with age, indicating a potential role of immunosenescence in disease progression [7,8]. Conversely, HLA-DRB1*13:02 has been found to be protective against some deleterious effects of APOE4 in dementia, further supporting a potential role of HLA function in disease [9].

Additionally, HLA associations have been identified in Alzheimer's disease and other neurodegenerative disorders [7,8]. In Figure 1 (Misra 2018) we see that many risk loci for neurological diseases have been mapped to or near HLA loci.

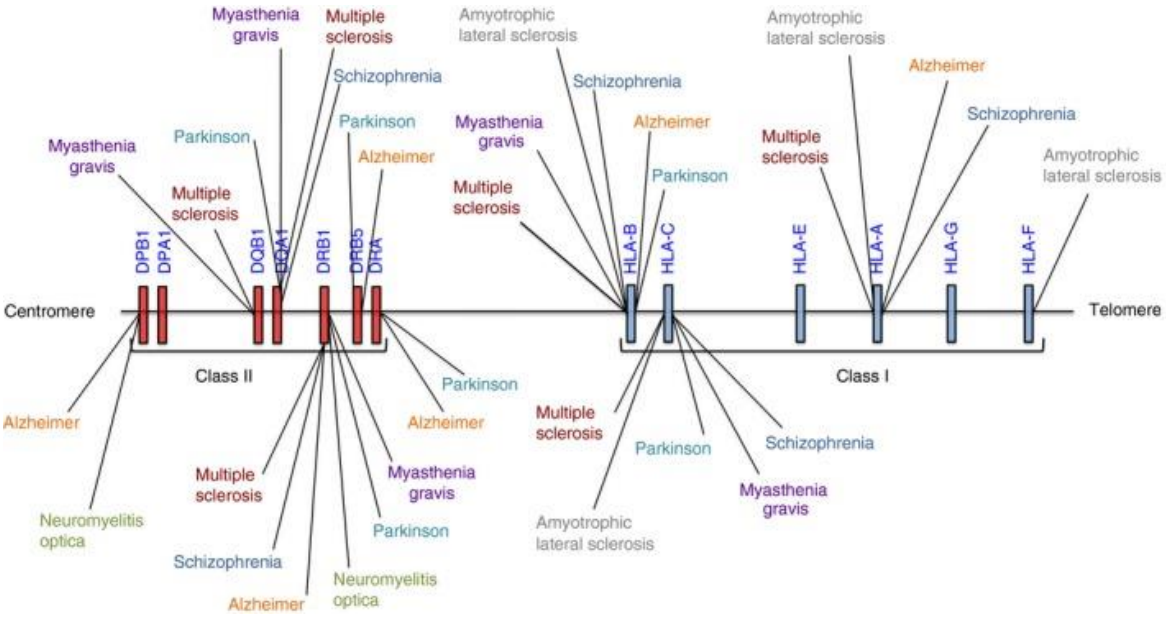

Figure 1. Genomic discovery in neurological disease mapped to extended MHC region on chromosome 6. (Misra 2018)

Further support of this potential HLA – neurodegeneration link has been found in our lab's previous work to identify differential gene expression across multiple neurodegenerative diseases. In this cross-disorder RNA expression data, we see a differential pattern of expression between Alzheimer's Disease and FTD for a subset of Class I HLA genes in key brain regions in Figure 2 (Rexach). Genes within the MHC gene region that encode for antigen presenting molecules such as Class I HLA genes represent the highest genetic risk for many neurological disorders [10]. Thus, determining HLA genotypes and haplotypes associated with increased risk of disease-specific clinical outcomes may uncover potential pathways of the disease.

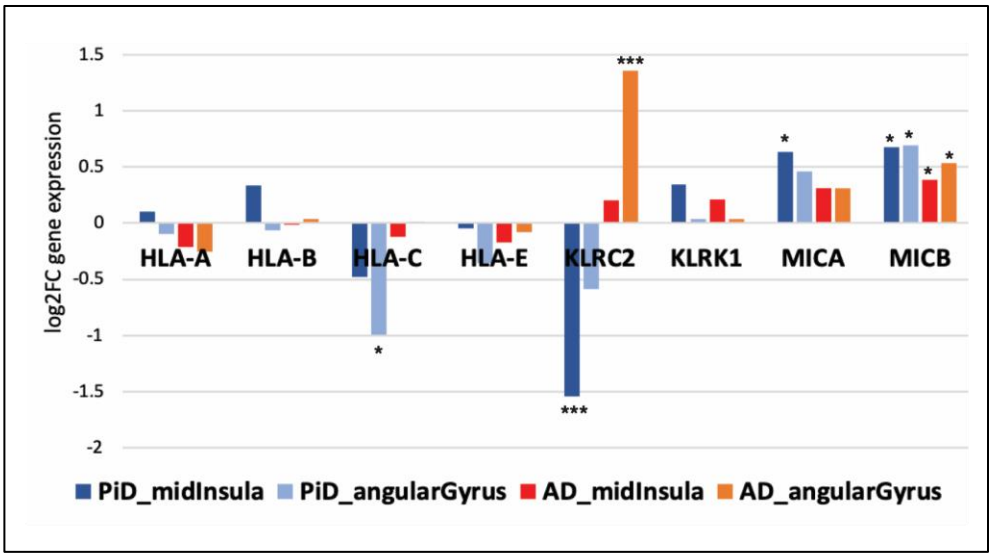

Figure 2. A subset of HLA genes and natural killer cell receptors are differentially expressed in FTD (PiD) brain, showing the insular cortex and angular gyrus, compared to Alzheimer's disease. (Rexach)

CHAPTER 2

Data Sets

2.1 Study Cohorts

In this study, to identify HLA alleles and haplotypes associated with clinical outcomes, we used the ARTFL LEFFTDS Longitudinal Frontotemporal Lobar Degeneration (ALLFTD) [1] cohort for our disease case study, and the Hillblom longitudinal aging study [2] was utilized as a control cohort.

THE ARTFL LEFFTDS Longitudinal Frontotemporal Lobar Degeneration cohort includes individual and familial participants with known familial or sporadic FTD that have completed two or more annual visits. Study data includes genetic data and a wide range of clinical data and biospecimens for hundreds of participants [1].

The Hillblom longitudinal data includes genotyping data and clinical demographics from community-enrolled study participants that completed at least two annual study visits at the University of California, San Francisco Memory and Aging Center. There they underwent evaluation at every visit to determine they met study criteria of no neurological condition or functional decline [2].

The genetic data we used from these cohosts consist of SNP-based genotyping data generated from the Illumina Omni2.5Exome and Illumina Global Screening Array (GSA) platforms.

	ALLFTD (case)	Hillblom (control)
n	694 total 370 Omni2.5Exome 324 Global Screening Array	325 total Omni2.5Exome
Median Age (range)	65 (25 – 88)*	71 (32 – 98) **
Sex	383 Male, 311 Female	158 Male, 167 Female
Phenotypes	FTD (six subtypes)	Normal / Impaired

Table 1. Demographics for study cohorts. * Age at time of first record, ** Age at time of collection.

CHAPTER 3

Methods

3.1 Quality Control of Genomic Data

To begin, the SNP data from each cohort was run through a thorough, standard quality-control pipeline (Figure 3.). The SNP and sample quality control checks included standard filtering in PLINK for sex discrepancies, relatedness, SNP and sample missingness, minor allele frequency (MAF), and Hardy-Weinberg equilibrium (HWE) deviation.

The datasets were then processed through standard ancestry clumping and filtering through PLINK using a multidimensional scaling approach (MDS) anchored by 1kG data. This process was also performed with the R package bigsnpr [12] to confirm proper labeling of ancestry groups and to conform to ALLFTD data processing. The parameters for these methods were defined by prior ALLFTD processing performed by the data managing group. As we were delivered pre-processed chromosome 6 SNP data for this cohort, we were unable to perform these steps on our own, which necessitated strict adherence to the established pipeline with the Hillblom cohort to avoid introduction of batch effects.

Additional filtering steps were completed before HLA imputation and analysis to include only phenotypes in "bvFTD, corticobasalsyndrome, FTD/ALS, PPA-nonfluent, PPA-semantic, PSP". Additional filtering by family gene = "none" was performed for additional analyses for sporadic cases.

Later analysis prompted more stringent filtering for the Hillblom cohort as well to remove any individuals without a normal clinical status. We also took the opportunity to remove individuals that self-reported as non-European ancestry in additional metadata that was obtained.

The total number of samples post filter for the ALLFTD was 577 (295 Omni, 282 GSA). This was the final number used for all analyses. The initial cohort for Hillblom post QC was 325. After additional filtering in step 4, this number was reduced to 140 "normal" controls and 129 with abnormal clinical diagnosis.

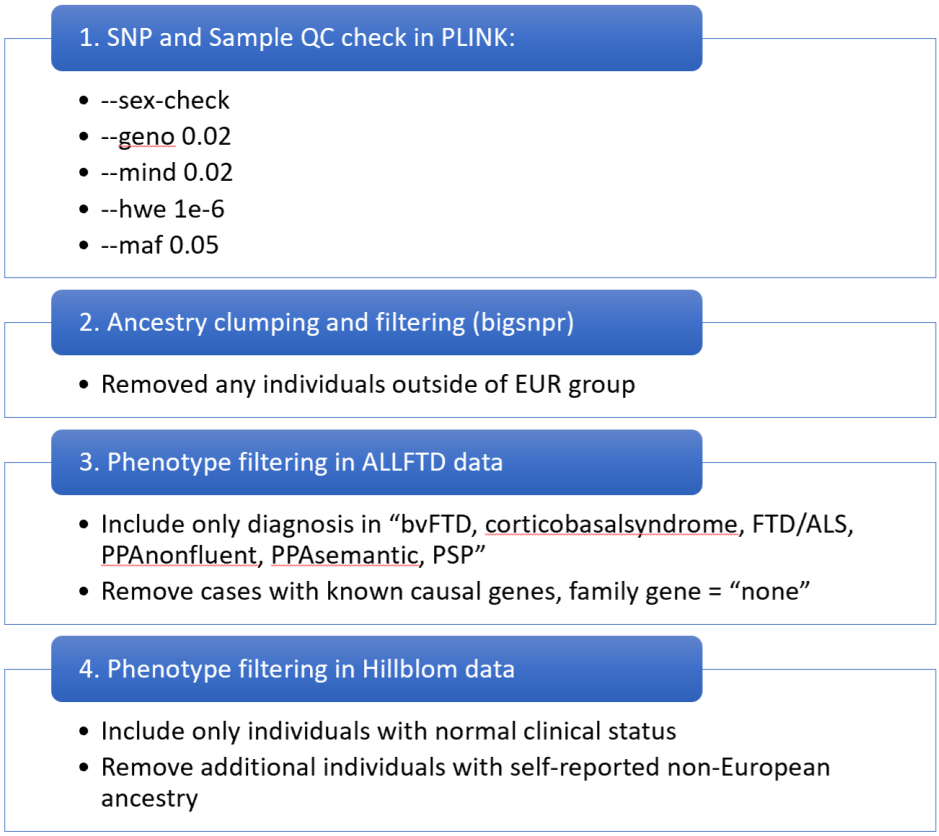

Figure 3. QC Pipeline and data filtering applied to both cohorts. Steps 1-3 were applied pre-analysis. Step 4 was applied for additional analyses.

3.2 Imputation of HLA Genotypes

After quality control and filtering performed strand flips for proper data phasing using a platform specific tool [16]. The data was chromosome sorted and chromosome 6 VCF files were then gzipped and uploaded for imputation. Imputation was performed on the Michigan server "Genotype Imputation HLA (Minimac4) 1.5.8 with [13,14]. This tool utilizes the HLA-TAPAS pipeline that implements its own quality controls steps before phasing the genotype data and performing the imputations with SNP2HLA with the latest reference panel, "Four-digit Multi-ethnic HLA v2 (2022)" [13,14]. Output binary markers for classical HLA alleles were utilized to determine HLA alleles and corresponding haplotypes.

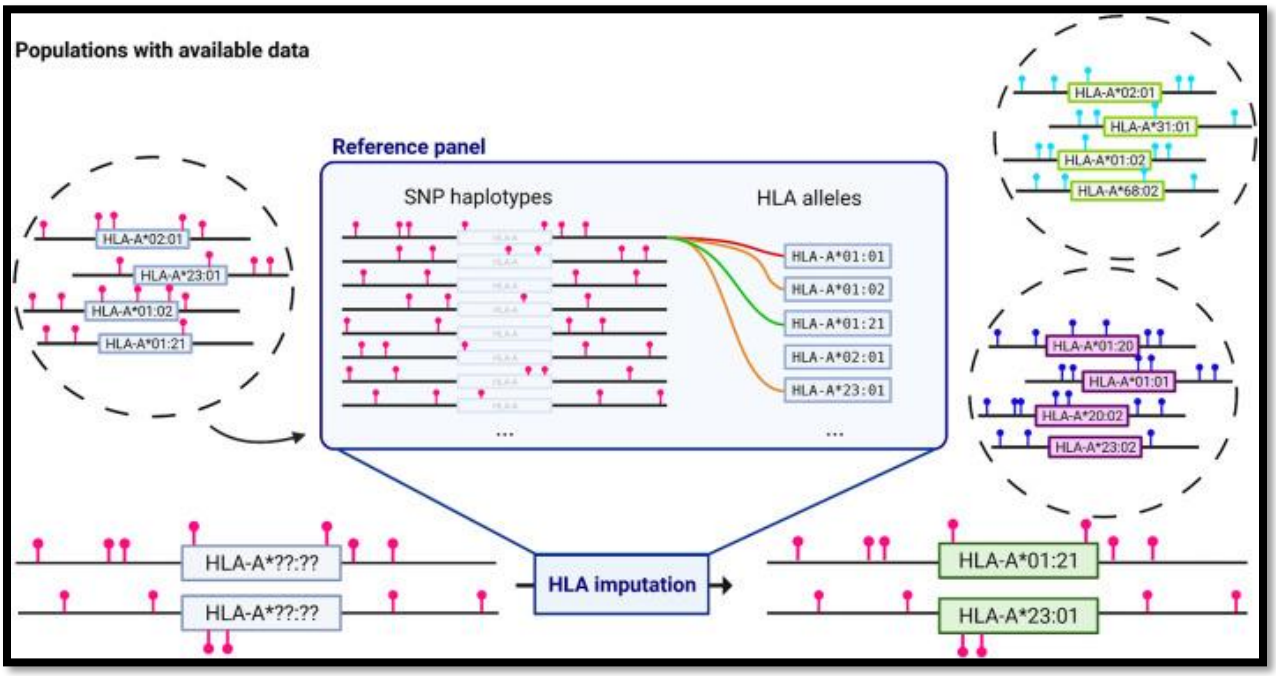

Figure 4. A reference-panel based approach to imputing HLA genotypes from SNP genotype data. (Douillard 2021)

3.3 Analysis of Imputed HLA Genotypes

With the imputed HLA allele data from each cohort, we utilized the R package BIGDAWG [17] to confirm HLA haplotypes and determine associations with disease phenotypes and clinical outcomes.

A custom script was used to format the PLINK binary HLA marker data output from the Michigan Imputation Server to the HLA allele table input required for BIGDAWG. Data needed to be converted from PLINK formatting to tabular data form with subject IDs, phenotypes, and genotype pairs for each HLA allele. HLA alleles needed to be in recognizable format for proper analysis.

We then ran the analysis in BIGDAWG using default method parameters on our combined and coded case/control dataset. This analysis operates in several steps to test for significance at the locus level, allele-level, haplotype level, and finally protein sequence level using data from IMGT/HLA database [17].

To interpret these results, we ran a series of associations based on disease phenotype for a comprehensive discovery phase utilizing each of our genetic datasets. These results were then compared to established population allele frequencies as described by the National Marrow Donor Program [18].

CHAPTER 4

Results

4.1 Imputation Performance

Our first step post-imputation was to determine the imputation completed without any noticeable error and provided consistent results for both of our cohorts.

	Total (2n = 2038)		Hillblom (2n = 650)	ALLFTD Omni (2n = 740)	ALLFTD GSA (2n = 648)
Locus	Number of Alleles	Number Missing	# Missing	# Missing	# Missing
A	38	6	2	1	3
B	57	15	4	4	7
C	27	0	0	0	0
DRB1	45	18	3	3	12
DQA1	10	0	0	0	0
DQB1	15	0	0	0	0
DPA1	6	0	0	0	0
DPB1	23	3	0	1	2

Figure 5. Performance results of the Michigan Imputation HLA Pipeline on our datasets with values represented as allele counts.

4.2 HLA Allele Frequency Variation

After confirming the success of our HLA imputations, we proceeded to our case/control analysis. The ALLFTD cohort was filtered to remove cases with known causal family genes to provide an analysis of sporadic cases. After multiple testing corrections two alleles were found to be significantly different between cohorts. The European ancestry population frequencies [18]

were included as a sanity check for our frequency results. This analysis included the sporadic FTD cases (n=577) and full Hillblom cohort after filtering for quality as previously described (n=325).

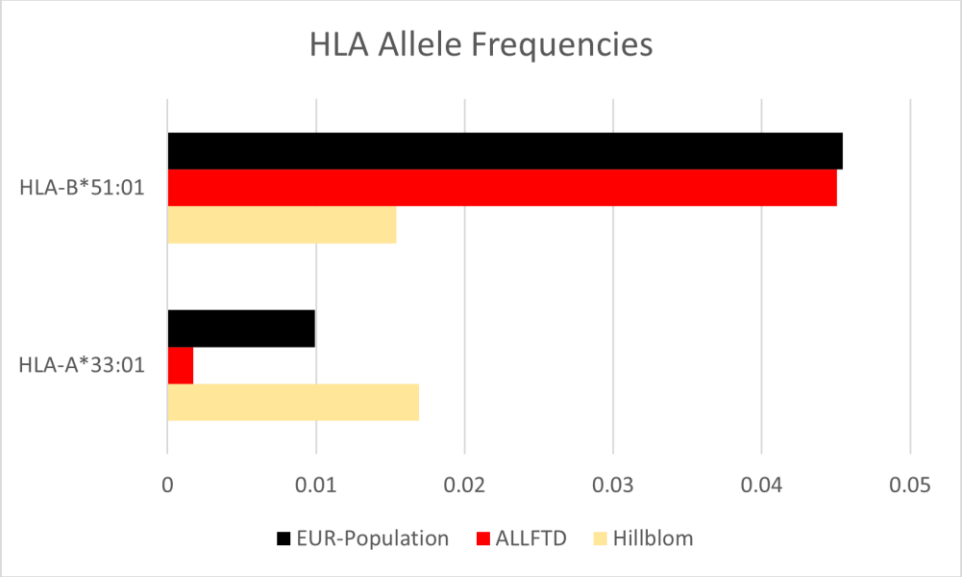

Figure 6. Allele frequencies found in ALLFTD sporadic cases and Hillblom cohort alongside European population frequencies.

Allele	HILL Freq	FTD Freq	OR	p value	EUR Freq
A*33:01	0.016923	0.001733	0.1	0.009039	0.00991
B*51:01	0.015385	0.045061	3.03	0.049368	0.04544

Table 2. Allele frequencies as seen in figure 6.

4.3 Deviations from Population Frequencies

When we examined the frequencies from our FTD case/control analysis in relation to known population frequencies, it became clear that the Hillblom cohort that we were using as a control may be driving our results. By selecting a group without any form of dementia in a longitudinal aging study, we realized we inadvertently selected a group that deviated from the general population. Since this may provide a greater allele frequency for alleles with neuro-protective effects, we decided to look further into the make-up of this cohort. Accordingly, we found diagnostic codes that indicated current cognitive status. To determine if a particular subgroup may have influenced our results, we split the Hillblom cohort according to cognitive diagnosis (162 not impaired, 163 impaired) and ran the analysis again.

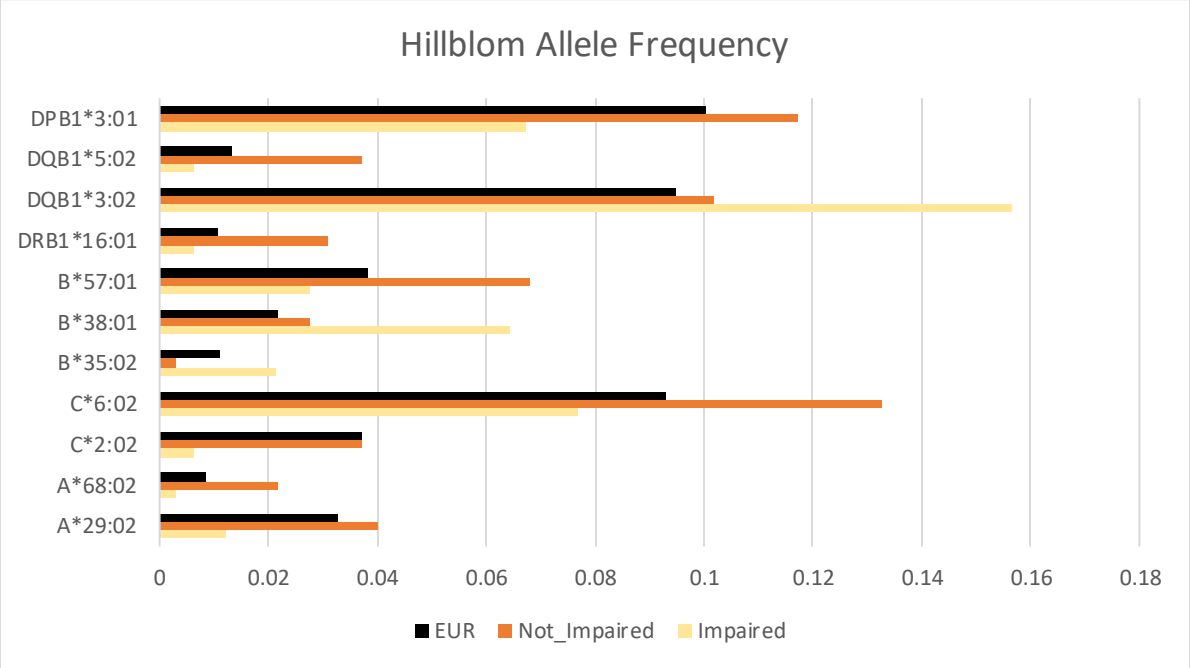

Figure 7. Significantly different allele frequencies between cognitive impairment and non-impaired individuals in Hillblom cohort with corresponding EUR population frequencies

Allele	Impaired Freq	Not Impaired Freq	OR	p value	EUR Freq
A*29:02	0.012269939	0.040123457	3.39	0.02523	0.03279
A*68:02	0.003067485	0.021604938	7.22	0.031386	0.00845
C*2:02	0.006134969	0.037037037	6.23	0.006657	0.03729
C*6:02	0.076687117	0.132716049	1.84	0.019614	0.09301
B*35:02	0.021472393	0.00308642	0.14	0.03426	0.01099
B*38:01	0.064417178	0.027777778	0.42	0.02718	0.0218
B*57:01	0.027607362	0.067901235	2.58	0.015167	0.03832
DRB1*16:01	0.006134969	0.030864198	5.18	0.018907	0.01061
DQB1*3:02	0.156441718	0.101851852	0.61	0.038038	0.09504
DQB1*5:02	0.006134969	0.037037037	6.23	0.006657	0.01315
DPB1*3:01	0.067484663	0.117283951	1.84	0.0283	0.1005

Table 3. Allele frequencies as seen in figure 7.

4.4 Cognition diagnosis-based Analysis

Next, we reached out to the clinical study team to acquire additional and revised clinical metadata for the Hillblom cohort to confirm our findings. We found additional individuals with

reported cognitive issues and the inclusion of self-reported ancestry, so we performed additional filtering to remove individuals with any diagnosis suggestive of cognitive issues and any self-reported non-European ancestry. The total number of this cognitively resilient subset of the Hillblom cohort included in the analysis was 140 individuals (FTD case n = 577, Hillblom control n =140).

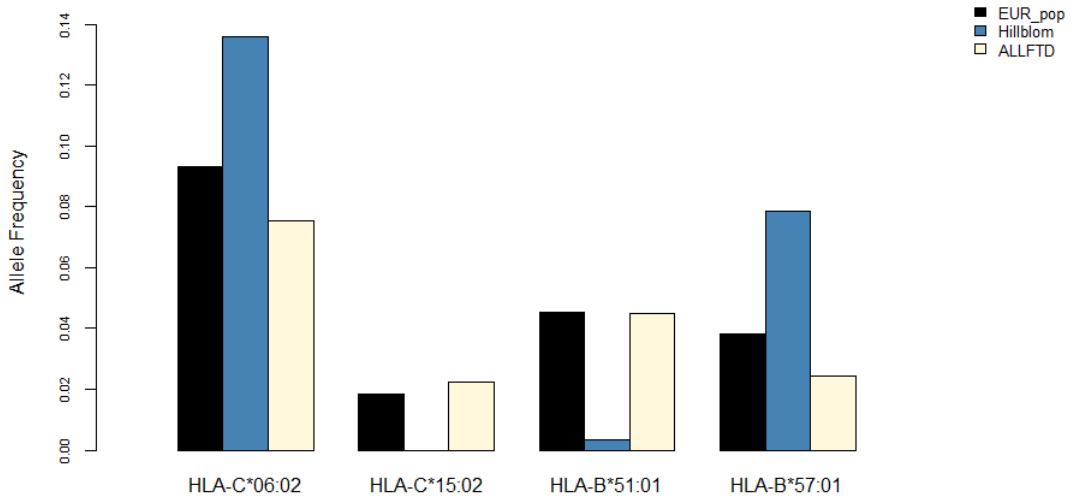

Figure 8. Allele frequencies for HLA alleles were found to be significantly different between Hillblom "normal" diagnostic control and ALLFTD cases with European Population frequencies.

Allele	HILL Freq	FTD Freq	OR	p val	EUR Freq
C*06:02	0.135714	0.075389948	0.52	0.0318408	0.09301
C*15:02	0	0.022530329	Inf	NS	0.01861
B*51:01	0.003571	0.045060659	13.19	0.047557	0.04544

B*57:01	0.078571	0.024263432	0.29	0.000453675	0.03832

Table 4. Allele frequencies as seen in figure 9.

Additionally, significant A~B haplotype frequencies were identified with this analysis (case n = 577, control n = 140).

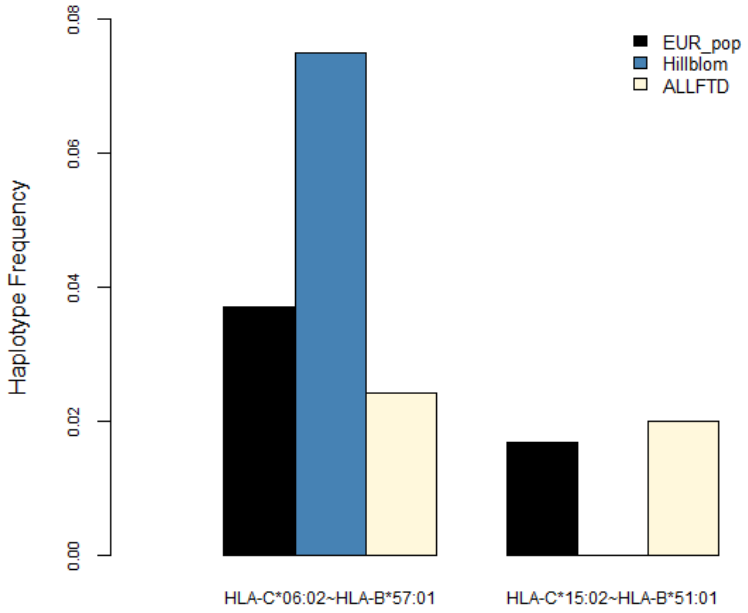

Figure 9. Haplotype frequencies for HLA-B~HLA-C were found to be significantly different between Hillblom "normal" diagnostic control and ALLFTD cases with European Population frequencies.

HLA-C~HLA-B	**HILL Freq**	**FTD Freq**	**OR**	**p val**	**EUR Freq**
06:02~57:01	0.075	0.024263	0.31	2.76E-05	0.03701
15:02~51:01	0	0.019931	Inf	0.017242	0.01677

Table 5. Haplotype frequency values as seen in figure 9.

CHAPTER 5

Discussion

5.1 Conclusion

In our findings, we initially found two HLA allele frequencies to be significantly different between our FTD case cohort and control: HLA-A*33:01 (OR=0.1, p.val=0.009) and HLA-B*51:01 (OR=3.03, p.val=0.049). However, when comparing these frequencies with corresponding population frequencies of the US population with European ancestry, we found our Hillblom study control cohort frequency to significantly deviate from the population values and ultimately influence our results. Therefore, we re-queried the clinical phenotypes and self-reported ancestry of this "control" cohort to assess what differences within the cognitive aging cohort contributed to differences in observed HLA allele frequencies compared to the FTD cohort. We observed that individuals with cognitive resilience (clinical status = "normal") had HLA genotype frequencies that deviated from our case and population frequencies while individuals in the Hillblom aging cohort with reported cognitive decline had HLA frequencies more similar to the FTD cohort.

Among the greatest differences among cognitively resilient cases was the frequency of HLA-B*51:01, which showed a four-fold reduction compared to Hillblom study participants with cognitive decline (0.015385, 0.00357). The frequency of this allele in the FTD cohort was strikingly similar to the population (0.045, 0.047). HLA-B*51:01 is a known risk factor for determining clinical phenotypes of Behcet's syndrome [19], which is an autoimmune disorder.

These findings suggest that HLA-B*51:01 carriers may have increased risk for cognitive decline compared to non-carriers, which has not previously been reported.

Additionally, we saw a total depletion of HLA-C*15:02 and significant increases in HLA-B*57:01 (OR=0.29, p.val=0.000454) and HLA-C*06:02 (OR=0.52, p.val=0.031841) in the cognitively resilient Hillblom cohort compared to ALLFTD study subjects (figure 8). HLA-B*57:01 is known to have protective effects in HIV, possibly alluding to an allele specific viral response pathway of interest. These findings suggest that HLA- B*57:01 carriers may have decreased risk for cognitive decline compared to non-carriers, which has also not previously been reported.

Of additional significance is that these alleles are frequently a part of the same haplotypes, and we can see that the A~B haplotype frequencies of the Hillblom cohort deviate from the ALLFTD cohort as well as the population frequencies (Figure 9). This is not unexpected considering the high linkage disequilibrium in this genomic region and the previously reported allele frequencies. Near the SNP used for imputing HLA-B*57:01 and also in high linkage disequilibrium are the genes MICA, which was found to be differentially expressed in FTD in the cross-disorder study (figure 2) and HCP5, which has also been implicated in HIV outcomes [21]. The genomic region including the SNPs used to impute the haplotype HLA-C*06:02~HLA-B*57:01 is shown in Figure 10.

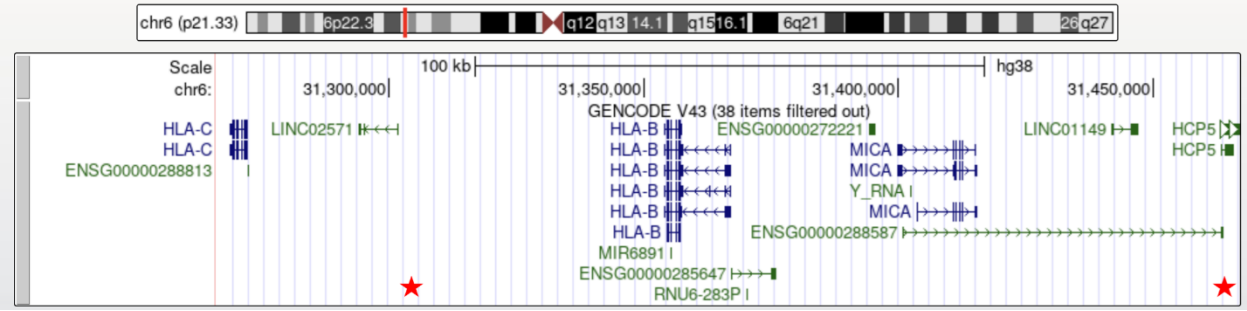

Figure 10. Genomic region chr6 (p21.33) shows genes near SNPs used to impute HLA-C*06:02 and HLA-B*57:01.

Allele	SNP for Imputation	SNP Position GRCH38/hg38
HLA-B*57:01	rs2395029	6:31464003
HLA-C*06:02	rs10484554	6:31306778

Table 6. SNP names and positions used for HLA imputation.

5.2 Future Directions

By performing HLA analysis in the ALLFTD and Hillblom longitudinal aging cohorts, we identified candidate HLA genotypes associated with greater cognitive resilience in aging. The next major important step is to assess the reproducibility of these findings in additional cases. Toward that, the Rexach lab is currently sequencing an additional 160 cases from the Hillblom study to utilize as a replication cohort. In addition, we devised a plan to further test the HLAs identified in our study against existing aging cohorts with cognitive data, including the Alzheimer's Disease Neuroimaging Initiative (ADNI). These datasets might provide additional control data we need to power analysis of less frequent HLA alleles and haplotypes.

Additionally, as further validation, we established a plan to sequence the full HLA and Killer cell Immunoglobulin-like Receptors (KIR) gene and promoter regions at high density for

validation. Utilizing this data along with the ALLFTD and Hillblom cohorts will allow us to expand upon our imputation data while validating the imputation accuracy.

Finally, based on the preliminary observation of differential expression of HLA genes in brain tissue of cases with different neurodegenerative diseases, we are interested in the question of how HLA variants related to differences in disease risk and cognitive aging might affect the expression of HLA genes that are differentially regulated in diseased tissues where they might contribute to disease progression. To explore this, we are genotyping the cross-disorder dataset and will perform the same HLA analysis there to identify cases representing alleles either enriched or depleted in disease cohorts. Identification of such associations would provide a foundation for experimental studies into the role of HLA variation in disease phenotypes of the aging brain.

References

1. Boeve, B.F., Boxer, A.L., Rosen, H.J., Forsberg, L.K., Heuer, H.W., Brushaber, D., Appleby, B., Biernacka, J.M., Bordelon, Y.M., Botha, H., Brannelly, P., Dickerson, B.C., Dickson, D.W., Kimiko, D.-R., Faber, K., Fagan, A., Fields, J.A., Fishman, A., Foroud, T.M., Galasko, D.R., Gavrilova, R.H., Gendron, T.F., Geschwind, D.H., Ghoshal, N., Goldman, J., Graff-Radford, J., Graff-Radford, N.R., Grant, I., Grossman, M., Hsiung, G.-Y.R., Huang, E.J., Huey, E., Irwin, D.J., Jones, D.T., Kantarci, K., Karydas, A.M., Kaufer, D., Knopman, D.S., Kramer, J.H., Kremers, W.K., Kornak, J., Kukull, W.A., Lagone, E., Leger, G.C., Litvan, I., Ljubenkov, P.A., Lucente, D.E., Mackenzie, I.R., Manoochehri, M., Masdeu, J.C., McGinnis, S., Mendez, M.F., Miller, B.L., Miyagawa, T., Nelson, K.M., Onyike, C.U., Pantelyat, A., Pascual, B., Pearlman, R., Petrucelli, L., Rademakers, R., Ramos, E.M., Rankin, K., Rascovsky, K., Rexach, J.E., Ritter, A., Roberson, E.D., Rojas, J.C., Sabbagh, M.N., Salmon, D.P., Savica, R., Seeley, W.W., Staffaroni, A.M., Syrjanen, J., Tartaglia, C., Tatton, N., Taylor, J., Toga, A.W., Weintraub, S., Wheaton, D., Wong, B. and Wszolek, Z. (2020), *Studying the natural history of frontotemporal lobar degeneration (FTLD): The ARTFL LEFFTDS longitudinal FTLD (ALLFTD) protocol. Alzheimer's Dement.*, 16: e045482. https://doi.org/10.1002/alz.045482

2. Casaletto KB, Elahi FM, Staffaroni AM, Walters S, Contreras WR, Wolf A, Dubal D, Miller B, Yaffe K, Kramer JH. *Cognitive aging is not created equally: differentiating unique cognitive phenotypes in "normal" adults.* Neurobiol Aging. 2019 May;77:13-19. doi: 10.1016/j.neurobiolaging.2019.01.007. Epub 2019 Jan 24. PMID: 30772736; PMCID: PMC6486874.

3. Ghosh S, Lippa CF. *Clinical Subtypes of Frontotemporal Dementia*. American Journal of Alzheimer's Disease & Other Dementias®. 2015;30(7):653-661. doi:10.1177/1533317513494442

4. Bang J, Spina S, Miller BL. *Frontotemporal dementia*. Lancet. 2015 Oct 24;386(10004):1672-82. doi: 10.1016/S0140-6736(15)00461-4.

5. Ferrari, R., et al., *Frontotemporal dementia and its subtypes: a genome-wide association study.* Lancet Neurol, 2014. **13**(7): p. 686-99.

6. FTD talk, Rohrer J. *What is Frontotemporal Dementia*, UCL Queen Square Institute of Neurology, https://www.ftdtalk.org/what-is-frontotemporal-dementia/pathology/. 2015.

7. Lutshumba J, Nikolajczyk BS, Bachstetter AD. *Dysregulation of Systemic Immunity in Aging and Dementia*. Front Cell Neurosci. 2021 Jun 22;15:652111. doi: 10.3389/fncel.2021.652111. PMID: 34239415; PMCID: PMC8258160.

8. Jian B, Hu M, Cai W, Zhang B, Lu Z. Update of Immunosenescence in Cerebral Small Vessel Disease. Front Immunol. 2020 Nov 18;11:585655. doi: 10.3389/fimmu.2020.585655.

9. James LM, Georgopoulos AP. *Human Leukocyte Antigen as a Key Factor in Preventing Dementia and Associated Apolipoprotein E4 Risk*. Front Aging Neurosci. 2019 Apr 12;11:82. doi: 10.3389/fnagi.2019.00082. PMID: 31031617; PMCID: PMC6473084.

10. Misra MK, Damotte V, Hollenbach JA. *The immunogenetics of neurological disease.* Immunology. 2018 Apr;153(4):399-414. doi: 10.1111/imm.12869. Epub 2017 Dec 11.

11. Rexach, Jessica. *Cross-Disorder Analysis of Genetic Expression in Dementia*. 2023. Unpublished manuscript.

12. Florian Privé, Hugues Aschard, Andrey Ziyatdinov, Michael G B Blum, *Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr*, Bioinformatics, Volume 34, Issue 16, August 2018, Pages 2781–2787, https://doi.org/10.1093/bioinformatics/bty185

13. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze S, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C. *Next-generation genotype imputation service and methods.* Nature Genetics 48, 1284–1287 (2016).

14. Luo, Y., Kanai, M., Choi, W. et al. *A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response.* Nat Genet 53, 1504–1516 (2021). https://doi.org/10.1038/s41588-021-00935-7

15. Douillard V, Castelli EC, Mack SJ, Hollenbach JA, Gourraud PA, Vince N, Limou S. *Approaching Genetics Through the MHC Lens: Tools and Methods for HLA Research*. Front Genet. 2021 Dec 2;12:774916. doi: 10.3389/fgene.2021.774916. PMID: 34925459; PMCID: PMC8677840.

16. Rayner, William. *Genotyping Chip Strand and Build Files*, University Oxford, https://www.chg.ox.ac.uk/~wrayner/strand/index.html

17. Derek J. Pappas, Wesley Marin, Jill A. Hollenbach, Steven J. Mack, *Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline*, Human Immunology, Volume 77, Issue 3, 2016, Pages 283-287, ISSN 0198-8859, https://doi.org/10.1016/j.humimm.2015.12.006.

18. Maiers M, Gragert L, Klitz W. *High-resolution HLA alleles and haplotypes in the United States population*. Hum Immunol. 2007 Sep;68(9):779-88. doi: 10.1016/j.humimm.2007.04.005. Epub 2007 May 24. Erratum in: Hum Immunol. 2008 Feb;69(2):141. PMID: 17869653.

19. Takeno M. *The association of Behçet's syndrome with HLA-B51 as understood in 2021*. Curr Opin Rheumatol. 2022 Jan 1;34(1):4-9. doi: 10.1097/BOR.0000000000000846.

20. Lobos CA, Downing J, D'Orsogna LJ, Chatzileontiadou DSM, Gras S. *Protective HLA-B57: T cell and natural killer cell recognition in HIV infection*. Biochem Soc Trans. 2022 Oct 31;50(5):1329-1339. doi: 10.1042/BST20220244. PMID: 36111814; PMCID: PMC9704518.

21. van Manen D, Kootstra NA, Boeser-Nunnink B, Handulle MA, van't Wout AB, Schuitemaker H. *Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection*. AIDS. 2009 Jan 2;23(1):19-28. doi: 10.1097/QAD.0b013e32831db247. Erratum in: AIDS. 2010 Mar 27;24(6):939. PMID: 19050382.