Theses and Dissertations--Physiology

Physiology

2022

Immunoregulatory Receptor Genetics, Expression, and Splicing Studies in Alzheimer's Disease

Benjamin C. Shaw

University of Kentucky, bcshaw@outlook.com
Author ORCID Identifier:
https://orcid.org/0000-0001-8862-3840
Digital Object Identifier: https://doi.org/10.13023/etd.2022.036

Right click to open a feedback form in a new tab to let us know how this document benefits you.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

<div align="right">

Benjamin C. Shaw, Student

Dr. Steven Estus, Major Professor

Dr. Kenneth S. Campbell, Director of Graduate Studies

</div>

Immunoregulatory Receptor Genetics, Expression, and Splicing Studies in
Alzheimer's Disease

_____

DISSERTATION
_____

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Medicine
at the University of Kentucky

By
Benjamin Charles Shaw
Lexington, Kentucky
Director: Dr. Steven Estus, Professor of Physiology
Lexington, Kentucky
2022

ABSTRACT OF DISSERTATION

Immunoregulatory receptor genetics, expression, and splicing studies in
Alzheimer's Disease

Microglia are the resident immune cells of the brain, undertaking many critical tissue maintenance functions such as immune surveillance and phagocytosis. Microglial dysfunction has recently been identified as a multi-stage signature of many neurodegenerative diseases, including late-onset Alzheimer's Disease (LOAD). Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in over thirty genes that modulate risk of developing LOAD. In the central nervous system, roughly half of these LOAD-associated genes are primarily expressed in microglia. The proteins encoded by these genes include cell surface receptors that contain either immunomodulatory tyrosine-phosphorylated activating motifs (ITAMs) or inhibitory motifs (ITIMs), including *TREM2*, *CD33*, and *SIGLEC14*. Here, I studied the molecular genetics underlying these three genes and their respective contributions to LOAD risk.

First, I found that *TREM2* undergoes extensive alternative splicing in multiple tissues, including brain. Total *TREM2* expression is not different as a function of LOAD diagnosis ($p = 0.1268$), but *TREM2* expression is increased by 34% in tissues with higher National Institute on Aging/Reagan Institute (NIARI) scores ($p = 0.0033$). I also found that a novel *TREM2* isoform lacking exon 2, *D2-TREM2*, accounts for 11% of the total *TREM2* mRNA in human brain, and that this splicing efficiency is not altered as a function of AD status ($p = 0.4909$) or brain pathology ($p = 0.9502$). I also found that the D2-TREM2 protein has similar subcellular localization to its parent TREM2 protein, as both are primarily retained in the Golgi apparatus.

Next, I studied the exon 2-lacking *CD33* isoform, *D2-CD33*. I developed an *in vitro* model to study the function of the *D2-CD33* using a CRISPR-Cas9 approach in the U937 human monocyte cell line. After validating this model with sequencing, qPCR, and flow cytometry, I found that a nearby pseudogene, *SIGLEC22P*, was used as a repair template in approximately 10% of edited cells. This finding also provided the highest resolution to date of the clinically relevant anti-CD33 P67.6 antibody clone, gemtuzumab.

Finally, I combined a recent LOAD GWAS with a protein quantitative trait loci (pQTL) study to uncover *SIGLEC14* as a potentially overlooked LOAD risk factor. I found that a previously described *SIGLEC14* genetic deletion occurs within a 692 bp crossover region. I also found additional copy number variation not previously described using both qPCR-based and *in silico* assays, with copy numbers identified ranging from zero to four. While *SIGLEC14* deletion does correlate well with a proxy single nucleotide polymorphism (SNP), rs1106476, additional *SIGLEC14* genomic copies do not correlate with this SNP. Further, the *SIGLEC14* genomic deletions correlate stepwise with decreased *SIGLEC14* expression ($p = 0.0002$), and also correlate significantly with decreased *SIGLEC5* expression ($p = 0.0389$).

In conclusion, microglial cell surface receptors are heavily implicated in the risk of developing LOAD, and these studies advance the field by adding to the molecular mechanisms which underlie their risk contribution. Further studies will be needed to address whether these findings can be translated clinically to either potential druggable targets or biomarkers.

KEYWORDS: Molecular genetics, Neurodegenerative disease, Genome-wide association studies, functional genetics

Benjamin Charles Shaw

02/28/2022

Immunoregulatory Receptor Genetics, Expression, and Splicing Studies in
Alzheimer's Disease

By
Benjamin Charles Shaw

Dr. Steven Estus	
Director of Dissertation	
Dr. Kenneth S. Campbell	
Director of Graduate Studies	
02/28/2022	
Date	

DEDICATION

To my father, Craig Shaw, the reason I pursued science and research.

ACKNOWLEDGMENTS

Graduate school is not without its challenges, and these challenges require the support of a broad range of individuals. Mentors, fellow students, friends, and family have each sustained me throughout this process, and I could not be more grateful.

Dr. Steve Estus picked me up in the middle of my training, providing me a home to finish my PhD when I was unable to move to Pittsburgh. Throughout this time, he has been profoundly supportive, mentoring me through four grant applications, four publications, and a behind-the-scenes view into how to run a lab from personnel to grant management. I'm sure I could have been an easier student to mentor, but he has been nothing but patient with me. The skills he has passed on to me will serve me well for decades to come. Dr. Brad Taylor also had an integral role in my training as my first mentor. While I was unable to continue with him after moving his laboratory to Pittsburgh, he provided formative training in experimental design, techniques, and grantsmanship. My time in his laboratory provided the foundation needed to successfully transfer to the Estus laboratory.

My committee members Dr. Bret Smith, Dr. John Gensel, and Dr. Chris Norris have also guided me along this journey. You have all written more letters on my behalf than you signed up for, sat through committee meetings that went far too long, and shifted to a completely different project with me. Bret, your counsel during the time I tried making a long-distance mentorship work, my transition between laboratories, and the way you've watched over me and acted as a second

mentor is especially appreciated. Dr. Beth Garvy, while not acting as a mentor or committee member in any official capacity, has also been enormously supportive. From training me in experimental techniques to helping me navigate the postdoctoral market, I was lucky to find such a caring, knowledgeable, and candid guide.

None of this would have been possible without the seasoned experimentalists and bench scientists from whom I owe everything I've learned at the bench. Renee Donahue, Dr. Suzanne Doolen, Dr. Lilian Custodio-Patsey, Melissa Hollifield have helped me master techniques which I never thought possible. Jim Simpson, you have entertained every qPCR assay I've concocted since joining you and Steve, teaching me that as much experience as I have in a technique there's always more to learn. You've been a great lab manager, and even better friend. I'll certainly miss our frequent golf outings, constantly chasing a better score and being berated about that one shot that went "under the bridge."

Nobody can relate to the unique struggles of this journey like those going through with you. Dr. Bethi Oates, Dr. Brooke Ahern, Dr. Brandon Farmer, and Dr. Ryan Cloyd have walked this path with me, and we have pulled each other up countless times. From office venting sessions to long lunches to our Bourbon Club, we've been there for each other for years. While we all seem to be scattering across the country as this phase ends, I know we will continue to support each other from afar.

Finally, I have to thank my family. My father, Craig Shaw, has been a shining example of discipline, integrity, and perseverance. Through every bump in the road

these past six years, he has been a sounding board and reminded me how far I've come. Knowing I've made him proud is one of my greatest accomplishments. My in-laws, Susan & Walter Hirth and Rachel Mossop, have been a constant source of encouragement. My amazing wife, Heather Shaw, could not have been more supportive. I have worked 90-hour weeks, gone in early and come home late, left for weeks at a time to Pittsburgh, traveled for conferences, and far too often left my fair share of responsibilities at home on the back burner only for you to pick them up. You have shouldered that without complaint while maintaining your own career and keeping up with our wonderful daughter more than I should have let you on your own. Our beautiful daughter, Hayley, has grown up in the lab. I started this PhD when she was only two years old, and during this time she has been to the lab with me so many times she now looks forward to looking under microscopes and seeing what kind of new things I can show her. I can only hope that she maintains this level of curiosity in all that she does.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1   Alzheimer's Disease: A brief overview of pathophysiology and epidemiology

1.1.1   Discovery of Alzheimer's Disease

Alzheimer's Disease (AD) was first described in a 1906 lecture, subsequently published in 1907 by Dr. Alois Alzheimer (Alzheimer, 1907; Strassnig & Ganguli, 2005). The disease was characterized in a 51-year-old patient, Auguste Deter, presenting with a "rapidly worsening memory," an "inability to navigate her way around her dwelling," and "intermittently she is completely delirious." Notably, Frau Deter was considerably younger than the typical dementia patient with an aggressive course of disease. Upon her death, less than five years after entering an asylum, her brain was sectioned and stained. Alzheimer noted "an evenly atrophic brain," with "changes of neurofibrils," and "miliar foci" (Alzheimer, 1907; Strassnig & Ganguli, 2005). These correspond to the loss of brain volume, neurofibrillary tangles, and amyloid plaques as we now know them.

1.1.2   Pathophysiology of Alzheimer's Disease

Histologically, AD is characterized by the accumulation of neurofibrillary tangles (NFTs) and amyloid-β (Aβ) plaques (Figure 1.1) (Nelson *et al.*, 2009). The NFTs typically include hyperphosphorylated tau, a microtubule associated protein, but occasionally also contain TAR DNA-binding protein 43 (TDP-43) in advanced AD (Smith *et al.*, 2018). These NFTs are primarily intracellular, only released once the neuron degenerates. Aβ plaques are primarily extracellular and are initiated through cleavage of the amyloid precursor protein, first by β-secretase then γ-

secretase, to release either $A\beta_{40}$ or $A\beta_{42}$ (Y.-W. Zhang *et al.*, 2011). Of these two species, $A\beta_{42}$ is more likely to oligomerize and seed plaques (Burdick *et al.*, 1992). These plaques also contain other proteins and lipids, many of which are glycosylated (Kida *et al.*, 1995; Martin-Rehrmann *et al.*, 2005; McGeer *et al.*, 1994). This glycosylation may be an important factor in clearance of $A\beta$ plaques (Salminen & Kaarniranta, 2009).

Antemortem, these pathological hallmarks can be characterized using positron emission tomography (PET) imaging. Hyperphosphorylated tau aggregates can be detected using T807/AV1451, while $A\beta$ plaques can be detected using Pittsburgh compound B (PiB) (Dickstein *et al.*, 2016; Klunk *et al.*, 2004). These biomarkers have been incorporated into the A/T/N classification scheme to aid in diagnosis of AD when coupled with neuropsychiatric examinations (Jack *et al.*, 2016). Briefly, the A/T/N scheme includes amyloid (A) as either amyloid PET imaging or CSF $A\beta_{42}$ concentration, tau (T) as either PET or CSF phospho-tau concentration, and general neurodegeneration (N) as either [^{18}F]-fluorodeoxyglucose PET imaging (FDG-PET) or CSF total tau concentration. FDG-PET is a useful metric for neurodegeneration as diseased or dead neurons will take up less of the FDG tracer. These three biomarkers are assigned either a positive or negative binary response based on established cutoff values. While these methods are considered quantitative and useful in diagnosis, postmortem methods are considered more definitive as there can be variable agreement between PiB labeling and histological measures (Bacskai *et al.*, 2007; Sojkova *et al.*, 2011).

There are several postmortem autopsy methods for evaluating AD. The first method to characterize NFTs was developed by Heiko and Eva Braak, with six stages increasing in pathological burden (Braak & Braak, 1991). These stages are independent of amyloid pathology, as the "size and shape of the deposits and their distribution vary from one individual to another," while NFTs "exhibit a well-defined pattern" (Braak & Braak, 1991). Consortium to Establish a Registry on Alzheimer's Disease (CERAD) criteria also incorporate Aβ plaques and age at autopsy along with NFTs (Mirra *et al.*, 1991). Another commonly used postmortem characterization method is the National Institute on Aging/Reagan Institute (NIARI) Criteria (Hyman & Trojanowski, 1997; Khachaturian, 1985). The NIARI criteria combine the Braak scale with CERAD to generate a score which reflects the probability that the individual had AD.

1.1.3   Epidemiology of Alzheimer's Disease

The prevalence of AD is estimated at 6.2 million Americans, with the incidence expected to increase year-over-year for at least the next four decades (Alzheimer's Association, 2021; Rajan *et al.*, 2021). By 2030, this number is expected to reach over 8.3 million individuals. Overall, women have a greater risk of developing AD, with over two-thirds of all current AD patients being women, partially due to increased longevity over men (Alzheimer's Association, 2021). African-American individuals are also disproportionately more affected by AD than white individuals; 19% of African-Americans over age 65 have AD, compared to 10% of white Americans (Alzheimer's Association, 2021). In fact, this disparity may actually be an underestimate, as a missed diagnosis of AD is more common

among elderly African-Americans than whites (Clark *et al.*, 2005; Gianattasio *et al.*, 2019).

There are several risk factors outside of genetics which contribute to AD. By far, age is the best predictor for the development of AD, with risk doubling every five years after age 65 (Gollan *et al.*, 2011; Henderson, 1988; Herrup, 2010). Higher education level and cognitive activity are protective against AD (Jonaitis *et al.*, 2013; Solas *et al.*, 2013). Lifestyle factors including alcohol use, lack of exercise, poor diet, and smoking all contribute to AD risk (Grant, 2014; Henderson, 1988; Kirk-Sanchez & McGough, 2014; Rantanen, 2013; Scarmeas *et al.*, 2006; Soininen & Heinonen, 1982). Medically, common comorbidities include obesity, high cholesterol, Type II Diabetes, and history of traumatic brain injury (Anstey *et al.*, 2011; Leduc *et al.*, 2010; Luchsinger, 2010a, 2010b; Matsuzaki *et al.*, 2010; McKee *et al.*, 2014; McKee *et al.*, 2013; Mortimer *et al.*, 1985; Tolppanen *et al.*, 2014).

1.2   Familial Alzheimer's Disease and Late-Onset Alzheimer's Disease

Familial AD (FAD) is the result of at least one genetic variation in the *PSEN1*, *PSEN2*, or *APP* genes (Bird, 1993). These are rare, dominant variants, where heterozygous individuals have strong penetrance for AD, but only account for approximately 1% of all AD cases (R. J. Guerreiro *et al.*, 2012). *APP* encodes the amyloid precursor protein, which is cleaved by β- and γ-secretases to generate the fibrillogenic $A\beta_{42}$. Mutations in *APP* increase either the $A\beta_{42}{:}A\beta_{40}$ ratio or total $A\beta$ production, leading to increased $A\beta_{42}$ deposition and plaque formation

4

(Holtzman *et al.*, 2011). *PSEN1* and *PSEN2* encode presenilin proteins which are integral components of γ-secretase. Mutations in *PSEN1* are the most common identified cause of FAD (Holtzman *et al.*, 2011). Individuals with pathogenic mutations in one of these three genes typically develop AD by age 50, though variants in *PSEN2* seem to be less penetrant (Jayadev *et al.*, 2008).

Late-onset AD (LOAD), by contrast, typically develops after age 65. The single greatest genetic risk factor for development of LOAD is homozygosity for *APOE4*, increasing risk by approximately 12-fold (Karch & Goate, 2015). *APOE4* was discovered as a risk factor for AD in 1993 (Corder *et al.*, 1993). *APOE2*, conversely, is a protective haplotype (Corder *et al.*, 1994). The *APOE* haplotype is a combination of two single nucleotide polymorphisms (SNPs) rs429358 and rs7412. These are both missense SNPs, encoding arginine to cysteine substitutions, and occur at residues 112 and 158, respectively (Huebbe & Rimbach, 2017). While *APOE* was discovered as an AD risk factor as early as 1993, more recent studies have uncovered myriad genes which impact AD risk.

1.3 Genome-wide Association Studies in Alzheimer's Disease reveal microglia as a major contributing cell type to Late Onset Alzheimer's Disease

Genome-wide Association Studies (GWAS) represent a significant advance in the hunt for low penetrance risk variants. Since 2008, there have been at least nine GWAS for AD (Beecham *et al.*, 2014; Bertram *et al.*, 2008; Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; Karch *et al.*, 2012; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Naj *et al.*, 2011; Wightman *et al.*, 2021). These studies aggregate cohorts of subjects from around the world, with standardized inclusion criteria, and millions

of single nucleotide polymorphism (SNP) genotypes to uncover which SNPs are overrepresented in AD vs. control. Typically, hundreds of thousands of SNPs are genotyped directly—496763 in the case of Hollingworth, et al., 2011—then millions more are imputed based on patterns of coinheritance observed in genome sequencing studies. Jansen et al., 2019 analyzed over 13,000,000 SNPs in their study. The large number of samples in GWAS allow detection of low penetrance variants, such as that in *CD33*, rs3865444, with odds ratios reported between 0.84 to 0.99, indicating the minor allele of rs3865444 is protective from AD (Beecham *et al.*, 2014; Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; Naj *et al.*, 2011).

These GWAS have identified variants in 38 genes, many with multiple independent signals, which are associated with AD risk (Wightman *et al.*, 2021). The most well-validated variants are shown in Figure 1.2 (Karch & Goate, 2015). The majority of genes identified as associated with AD are expressed primarily, or exclusively, in microglia (Hansen *et al.*, 2017; Srinivasan *et al.*, 2016). Among these are the cell surface receptors *CD33*, *TREM2*, *FCER1G*, *HAVCR2*, and *CR1*; signaling mediator *PLCG2*; and actin-interacting protein *ABI3* (Beecham *et al.*, 2014; Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Naj *et al.*, 2011; Wightman *et al.*, 2021).

1.4   *TREM2* in Alzheimer's Disease

1.4.1   Discovery of *TREM2* as an Alzheimer's Disease risk factor

TREM2 was first identified as an AD risk factor in 2013 (R. Guerreiro *et al.*, 2013; Jonsson *et al.*, 2013). Both studies found a strong association (p < 10^{-8} in

both studies) of rs75932628, predicted to result in a p.R47H coding variation, with LOAD (R. Guerreiro *et al.*, 2013; Jonsson *et al.*, 2013). This finding was later replicated at the genome-wide level in multiple studies (Jansen *et al.*, 2019; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Wightman *et al.*, 2021). GWAS have also found that the *TREM2* locus has multiple independent signals; specifically of note, rs143332484 which is predicted to result in a p.R62H coding variation. Thus, *TREM2* variants have been repeatedly validated as genome-wide significant risk factors for AD.

1.4.2 TREM2 structure and function

TREM2 is a Type I transmembrane protein with an extracellular immunoglobulin V-set (IgV) domain which acts as the ligand binding domain. TREM2 is a receptor for both Apolipoprotein E (ApoE) and amyloid beta (Aβ) (Y. Zhao *et al.*, 2018), and regulates Aβ phagocytosis (McQuade *et al.*, 2020; Piers *et al.*, 2020; Y. Zhao *et al.*, 2018), transcriptional changes (Holtman *et al.*, 2015), and microglial transition to a full disease-associated phenotype (Atagi *et al.*, 2015; Keren-Shaul *et al.*, 2017). TREM2 also contains an intramembrane lysine critical for salt-bridge formation to DNAX-activating protein of 12 kDa (DAP12). DAP12 contains a cytosolic immunomodulatory tyrosine-based activation motif (ITAM) and is an obligate adaptor molecule for signaling (Gratuze *et al.*, 2018; Peng *et al.*, 2010). Upon TREM2 stimulation, DAP12 is phosphorylated by Src family kinases. Phospho-DAP12 then recruits spleen tyrosine kinase (Syk) which auto-phosphorylates and begins a signaling cascade ultimately leading to phospholipase C (PLC) activation, intracellular calcium flux, and mitogen-activated

protein kinase (MAPK) activity (Peng *et al.*, 2010). TREM2 is also apparently necessary for a disease-associated microglia (DAM) phenotype based on transcriptome profile as murine *Trem2*-knockout microglia stall in an intermediate disease-associated phenotype when crossed with mouse models of AD (Keren-Shaul *et al.*, 2017). Interestingly, loss of function variants in DAP12 and *TREM2* are causative for Nasu-Hakola disease, characterized by bone defects and early-onset dementia with death by age 60, further highlighting the role of TREM2 in brain function (Paloneva *et al.*, 2002).

TREM2 also undergoes proteolytic cleavage at the cell surface through ADAM10 or ADAM17 between p.H157 and p.S158 (Schlepckow *et al.*, 2017; Thornton *et al.*, 2017). This cleavage of the TREM2 ectodomain results in a soluble form of TREM2 (sTREM2) which may be beneficial in AD. Alternatively, sTREM2 is also predicted to be generated from an alternative splice isoform (ENST00000338469) wherein exon 4 is skipped resulting in a frameshift mutation which abrogates the transmembrane domain (Del-Aguila *et al.*, 2019). Notably, sTREM2 has been found in the plaques of humanized TREM2 mice crossed with the 5xFAD model, indicating sTREM2 retains its affinity for Aβ fragments (Song *et al.*, 2018). TREM2 also has a high avidity for Aβ monomers and oligomers, and sTREM2 likely retains this avidity as well (Lessard *et al.*, 2018). Higher cerebrospinal fluid (CSF) concentrations of sTREM2 inversely correlate with the rate of Aβ accumulation by PET imaging (Ewers *et al.*, 2020). This study also found sTREM2 inversely correlated with tau accumulation by PET imaging, especially at lower Braak stages. Interestingly, sTREM2 in CSF correlated with total and

phosphorylated tau, but not Aβ$_{42}$ (Piccio *et al.*, 2016). The discrepancy between correlation of sTREM2 with Aβ deposition but not Aβ$_{42}$ concentration may be due to an excess of sTREM2 binding Aβ plaques or normalization methods. Also, Aβ deposition rate in the Ewers *et al.* (2020) study may not be directly comparable to the Aβ$_{42}$ concentration in the Piccio *et al.* (2016) study. Taken together, these studies suggest a protective role for sTREM2 in AD given the inverse correlation of CSF sTREM2 and Aβ deposition by PET or tau burden by PET or in CSF, though further validation of this is warranted. This also presents splicing modulation as a potential therapeutic avenue if sTREM2 is functional and protective.

1.4.3   Molecular genetics of *TREM2* in Alzheimer's Disease

This section will focus on three SNPs: rs75932628, rs143332484, and rs2234255. While a multitude of *TREM2* variants are independently associated with AD, these are the most relevant to this work. These are missense mutations and are not known to affect expression nor splicing of the *TREM2* mRNA. The rs75932628 and rs143332484 encode the p.R47H and p.R62H mutations, respectively. The p.R47H mutation reduces ligand affinity, resulting in a partial loss of function in TREM2 associated with increased amyloid burden (Atagi *et al.*, 2015; Cheng *et al.*, 2018; Cosker *et al.*, 2021; Gratuze *et al.*, 2020; Parhizkar *et al.*, 2019). Both the p.R47H and p.R62H mutations reduce Aβ uptake and NFAT signaling, with the p.R62H mutation having a lesser effect compared to p.R47H (Lessard *et al.*, 2018). Consistent with this, both the p.R47H and p.R62H mutations appear to reduce microglial reactive signatures in human AD tissue, with p.R47H

having a greater effect than p.R62H (Zhou *et al.*, 2020). Thus, both the p.R47H and p.R62H mutations result in at least a partial loss of function of TREM2.

The rs2234255 is also a missense coding variant, resulting in a p.H157Y mutation, and increases TREM2 ectodomain cleavage (Thornton *et al.*, 2017). This SNP has a much weaker association with AD; the Jansen *et al.* (2019) GWAS did not find a genome-wide significant association (p = 0.0095, uncorrected) while smaller studies focused on individual populations such as the Han Chinese found weak evidence of an increase in AD risk (p = 0.02) (Jiang *et al.*, 2016). Since this SNP increases cleavage from the cell surface, it likely decreases functional, membrane-bound TREM2 and thus the apparent increase in risk is logical given what is known with regards to other *TREM2* loss-of-function variants. The fact that this SNP has a profoundly weaker effect is perplexing then. Two possible explanations exist: either the ectodomain cleavage rate is not high enough to overcome replacement by newly translated TREM2; or, less likely but more interesting, that the sTREM2 generated is functional and protective as suggested above and this protective effect is counteracting the detrimental effect of losing signaling-competent, cell surface TREM2.

1.5   *CD33* in Alzheimer's Disease

[This section contains material adapted from a published manuscript: Estus, S., Shaw, B. C., Devanney, N., Katsumata, Y., Press, E. E., & Fardo, D. W. (2019). Evaluation of CD33 as a genetic risk factor for Alzheimer's disease. *Acta Neuropathol*. 2019 Aug;138(2):187-199. PMID: 30949760.]

1.5.1 *CD33* in genome-wide association studies

CD33 has appeared in AD GWAS with mixed results over the years. In 2011, two genome wide association studies (GWAS) identified a single nucleotide polymorphism (SNP) located upstream of *CD33*, rs3865444, as associated with AD risk (Hollingworth *et al.*, 2011; Naj *et al.*, 2011). Naj *et al*, using AD Genetic Consortium samples, found a robust association between reduced AD risk and the minor allele of rs3865444 that reached genome wide significance when meta-analyzed with subjects from the three other consortia that would later form the International Genomics of Alzheimer's Project (IGAP) (Naj *et al.*, 2011). The complementary study by Hollingworth *et al.* was supportive of an association between rs3865444 and AD risk although the finding did not reach genome wide significance ($p = 2 \times 10^{-4}$) (Hollingworth *et al.*, 2011). Quite recently, Jansen *et al.* reported on a very large-scale meta-analysis which incorporated an unconventional AD-by-proxy phenotype to gain statistical power. In this analysis, individuals that reported a parent with AD were considered to be AD risk carriers and hence scored as AD. This study totaled 71,880 AD samples and 383,378 non-AD samples. Although subjects overlapped with IGAP, this study reconfirmed genome wide significance for the rs3865444 association with AD ($p = 6.3 \times 10^{-9}$) (Jansen *et al.*, 2019; Lambert *et al.*, 2013). In summary, currently available, very large studies support an association between rs3865444 and AD risk. The reason for the inconsistent association in some cohorts could reflect statistical power, variation in AD diagnostic accuracy or cohort risk, or, conceivably, inconsistent linkage disequilibrium between rs3865444 and the functional *CD33* SNP, or cohort

variations in the allele frequency of SNPs in CD33-related genes that modulate the rs3865444 association with AD risk.

1.5.2 CD33 structure and function

CD33 is a member of the sialic acid-binding immunoglobulin-type lectin (SIGLEC) family. CD33 contains an amino-terminal, extracellular IgV domain which acts as the ligand binding domain, along with an immunoglobulin C_2-set (IgC_2) domain (Estus *et al.*, 2019). The primary ligands are α2-3 and α2-6 linked sialic acids (Rodrigues *et al.*, 2020). CD33 also contains a cytosolic immunomodulatory tyrosine-based inhibitory motif (ITIM) proximal to the membrane and an ITIM-like sequence near their carboxyl terminus (Paul *et al.*, 2000). The ITIM is phosphorylated by Src family kinases and, once phosphorylated, recruits phosphatases such as SHP1 and SHP2 which counteract ITAM activity. A prevailing concept is that SIGLECs, including *CD33*, are activated by "self-associated molecular patterns," or SAMPs, to signal through their ITIM to induce immunosuppression (Varki, 2011), reviewed in Angata (2018); (Lubbers *et al.*, 2018). Briefly, SAMPs including specific glycosylation patterns, i.e. sialic acid linkages, provide a self-recognition signal to suppress innate immune cells, much in the way that pathogen-associated molecular patterns (PAMPs) or "disease-associated molecular patterns" (DAMPs) activate innate immune cells. This contact-dependent inhibition is thought to provide self-nonself discrimination in innate immune cells which do not undergo a selection process. These sialic acids are also commonly found in the ganglioside subclass of glycosphingolipids and as a common post-translational modification. Sialylated glycoproteins and

gangliosides are oftentimes especially abundant in pathophysiologic conditions such as cancer and inflammation, including AD amyloid plaques; their binding to CD33 has been suggested to inhibit plaque clearance (Ana Griciuc *et al.*, 2013; Salminen & Kaarniranta, 2009).

1.5.3 Molecular genetics of *CD33* in Alzheimer's Disease

The primary AD GWAS SNP for *CD33* is rs3865444, which is 372 bp upstream of the *CD33* transcription start site. The minor allele of this SNP is associated with reduced AD risk (Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; Naj *et al.*, 2011). Griciuc *et al*. did not detect an association between rs3865444 and *CD33* mRNA expression and yet found reduced CD33 protein in rs3865444 minor allele carriers (Ana Griciuc *et al.*, 2013). Subsequently, the Estus laboratory previously found total *CD33* mRNA was increased about 25% in AD and decreased modestly with the rs3865444 minor allele (Malik *et al.*, 2013). Moreover, previous work found that a surprisingly common CD33 isoform in human brain was lacking exon 2 (*D2-CD33*, also known as *CD33ΔV-Ig* (Ana Griciuc *et al.*, 2013) and *CD33m* (Perez-Oliva *et al.*, 2011)). Skipping of exon 2 deletes the CD33 IgV domain, leading to an in-frame fusion of the CD33 signal peptide directly to the IgC2 extracellular domain (Perez-Oliva *et al.*, 2011). When *D2-CD33* is quantified in comparison to *CD33* by using isoform specific qPCR primers, a striking correlation between rs3865444 and the proportion of *CD33* mRNA expressed as *D2-CD33* is observed (Malik *et al.*, 2013). The working hypothesis was that rs3865444 was a proxy for a functional SNP because rs3865444 is in the promoter region of *CD33*. Indeed, sequencing and subsequent genotyping studies found an

13

exon 2 SNP, rs12459419, that is in perfect linkage disequilibrium with rs3865444. Minigene studies confirmed that rs12459419 is a functional SNP with the minor allele increasing the proportion of *D2-CD33* (Malik *et al.*, 2013). The finding that rs3865444 and its functional proxy, rs12459419, are associated with *CD33* exon 2 splicing efficiency was subsequently confirmed in several reports (Malik *et al.*, 2015a; Raj *et al.*, 2014; Walker *et al.*, 2015). The association between rs12459419 and *CD33* exon 2 splicing has also been documented by acute myeloid leukemia (AML) researchers, who are interested in CD33 as a target for antibody-based AML therapeutics (Lamba *et al.*, 2017; Mortland *et al.*, 2013). In fact, several, although not all studies, found a correlation between the rs12459419 major allele and increased efficacy of gemtuzumab-ozogamicin, which recognizes an epitope in the CD33 IgV domain (Lamba *et al.*, 2017; Laszlo *et al.*, 2018; Mortland *et al.*, 2013) reviewed in (M. Gbadamosi *et al.*, 2018). The function(s) of D2-CD33, if any, have not been determined. Since loss of exon 2 does not alter the codon reading frame, *D2-CD33* encodes a protein essentially identical to CD33 except for the loss of the IgV domain. Therefore, we and others originally hypothesized that D2-CD33 is non-functional because loss of the IgV domain would result in an inability to bind sialic acid-based ligands (reviewed in (Malik *et al.*, 2015b; L. Zhao, 2018)).

We recently investigated the impact of a four bp insertion/deletion (indel), rs201074739, in exon 3 of *CD33*. The deletion is moderately rare with a ~2.4% minor allele frequency in European populations. The rs201074739 deletion causes a frameshift in the *CD33* codon reading frame at amino acid 155. This results in aberrant amino acids at positions 156-159 and then a premature stop codon.

Hence, instead of CD33 as a 364 amino acid type-1 transmembrane protein, CD33 containing this 4bp deletion is predicted to be a secreted protein consisting of the IgV domain and approximately 16 amino acids of the IgC2 domain. As such, this deletion would preclude both CD33 and D2-CD33. This was demonstrated in a very recent report by Papageorgiou *et al.* describing an individual that was homozygous for this deletion and had no detectable cell surface CD33 on their monocytes (Papageorgiou *et al.*, 2019). Since the rs3865444 minor allele is associated with decreased full-length CD33 and decreased AD risk, we hypothesized that this indel would also be associated with reduced AD risk. To evaluate this possibility, we used results from the most recent IGAP AD meta-analysis (Kunkle *et al.*, 2019). As a positive control to assess whether *CD33* genetics are associated with AD, we examined the *CD33* exon 2 splicing SNP, rs12459419, and found a significant association with AD risk (OR = 0.92 (95% CI: 0.90-0.95), $p = 4.5 \times 10^{-7}$). This was a robust sample set of 21,982 AD cases and 41,944 non-AD cases. However, the 4 bp indel was not significantly associated with AD risk in these same data (p = 0.1337, OR = 0.90 (95% CI: 0.79-1.03). While post-hoc power calculations have well-known limitations (Hoenig & Heisey, 2001), it is somewhat revealing that a dataset of this size (>60k individuals) confers 80% statistical power to detect an odds ratio as small as 0.90—comparable to the effect of rs12459419—for a variant with a minor allele frequency similar to this indel (2.4%).

1.5.4 Current Model for *CD33* and Alzheimer's Disease

Based on our current understanding of *CD33* genetics, AD risk and SNP actions, an interesting paradigm has emerged. The AD-associated SNP, rs3865444, acts through the linked SNP rs12459419 to primarily increase aberrant exon 2 splicing. This results in a D2-CD33 increase at the expense of CD33. Reduced functional CD33 was hypothesized to mediate reduced AD risk. This loss of function hypothesis is derived from the demonstrated ability of ligand binding to the CD33 IgV domain to cause CD33 clustering and ITIM phosphorylation and overall lead to an inhibition in phagocytic-type activity. A decrease in CD33 inhibition of microglial function was thought to increase microglial function, e.g., clearance of amyloid and cell debris and, over time, lead to reduced AD pathology. The conclusion in this loss of function hypothesis is that further reducing CD33 protein via pharmacologic means (antibodies, small molecules, anti-sense approaches) would reduce AD incidence and severity (Malik *et al.*, 2015a; L. Zhao, 2018). This theory is now called into question because of the data regarding the rs201074739 indel variant. Although the indel is expected to cause a 50% reduction in CD33 in heterozygous individuals and to produce a complete loss of CD33 cell surface expression in homozygous individuals (Papageorgiou *et al.*, 2019), this indel does not appear to modulate AD risk. Considering *CD33* genetic findings overall, we conclude that (i) an AD-protective splicing SNP, rs12459419, increases D2-CD33 and decreases CD33 but (ii) an indel that robustly decreases CD33 has no effect on AD risk. While we recognize that future genetic studies may yield new findings, a parsimonious interpretation of available data is that AD protection may

well be mediated by the increase in D2-CD33, i.e., D2-CD33 represents a gain of function variant to protect from AD. Microglial function appears to be regulated by a homeostatic balance between ITIM and ITAM signaling, reviewed in detail elsewhere (Gratuze *et al.*, 2018; Jay *et al.*, 2017; Linnartz & Neumann, 2013; Shi & Holtzman, 2018). That both *TREM2* and *CD33* have genome-wide significant variants associated with AD is strong evidence that this pathway is critical in AD pathogenesis and thus a promising pathway to target for therapeutics. A full overview of the interaction between CD33 and TREM2 is presented as Figure 1.3.

1.6   Involvement of other microglial receptors in Alzheimer's Disease

While GWAS clearly have substantial power, there are also limitations. Among these limitations is the power to detect signals from low penetrance variants especially from moderately rare (1-5% minor allele frequency (MAF)) and the focus on SNPs rather than other variants such as insertion-deletion variants (indels), complex repeated regions such as variable number of tandem repeat (VNTR) and short tandem repeat (STR) regions, and structural variations such as gene deletions (Ebbert *et al.*, 2019; Tam *et al.*, 2019). The TREM2-CD33 pathway has been repeatedly identified as important in AD, both in GWAS and preclinical animal and *in vitro* models (Bhattacherjee *et al.*, 2021; Bhattacherjee *et al.*, 2019; Bradshaw *et al.*, 2013; A. Griciuc *et al.*, 2019; Ana Griciuc *et al.*, 2013; Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; McQuade *et al.*, 2020; Raj *et al.*, 2014; Wightman *et al.*, 2021). With this evidence, other targets, including receptors, in this pathway may be proposed as potential targets as well despite lacking genetic evidence of involvement. These proposed targets include *SIGLEC14*, *SIGLEC5*, *LYN*, *BLNK*,

and *SYK* (B. C. Shaw *et al.*, 2021; Sierksma *et al.*, 2020). SIGLEC14 couples to DAP12 for signaling, while SIGLEC5 is its paired receptor with an ITIM. LYN is a Src family kinase which regulates microglial activity, specifically through phosphorylation of ITIMs and ITAMs. BLNK is a scaffolding protein which stabilizes receptor-kinase interactions. Syk is one of the primary signaling proteins in the ITAM pathway, alongside PLC—itself a genome-wide significant risk factor for AD (Jansen *et al.*, 2019; Kunkle *et al.*, 2019). Thus, while GWAS have provided approximately 40 genes of interest in AD risk, we should expand our preclinical and clinical studies beyond GWAS to include targets within the pathway as well.

1.7   Summary

GWAS have drastically expanded our understanding the pathogenesis of AD yet. Our challenge now is translating these large-scale genetic studies into translational therapies to impact disease course. GWAS risk data combined with bioinformatics and molecular biology allow us to predict the direction of desired effect, i.e. activation or inhibition of a specific protein or pathway. Better understanding of the molecular genetics of each risk variant provides further context on directionality and possible alternative strategies such as splicing modulation. The TREM2-CD33 pathway has been studied extensively and monoclonal antibodies targeting each are currently in clinical trials. Overall, there remains substantial gaps in knowledge between the molecular genetics and protein effects to generate risk or protection.
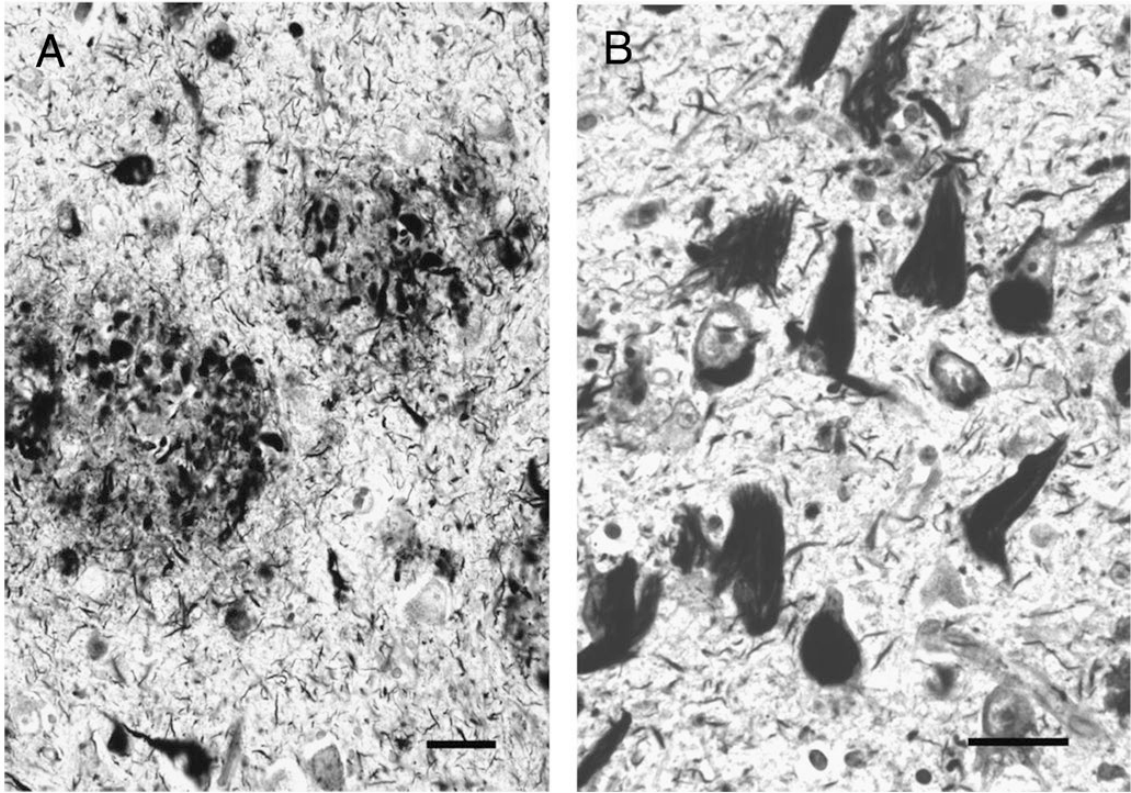
Figure 1.1: Basic pathological features of AD.

(A) Aβ plaques surrounded by degenerating neurons and (B) NFTs composed of intracellular, insoluble hyperphosphorylated tau protein are the hallmark histological features of AD. Adapted from Nelson, et al. 2009. Reproduced with permission.
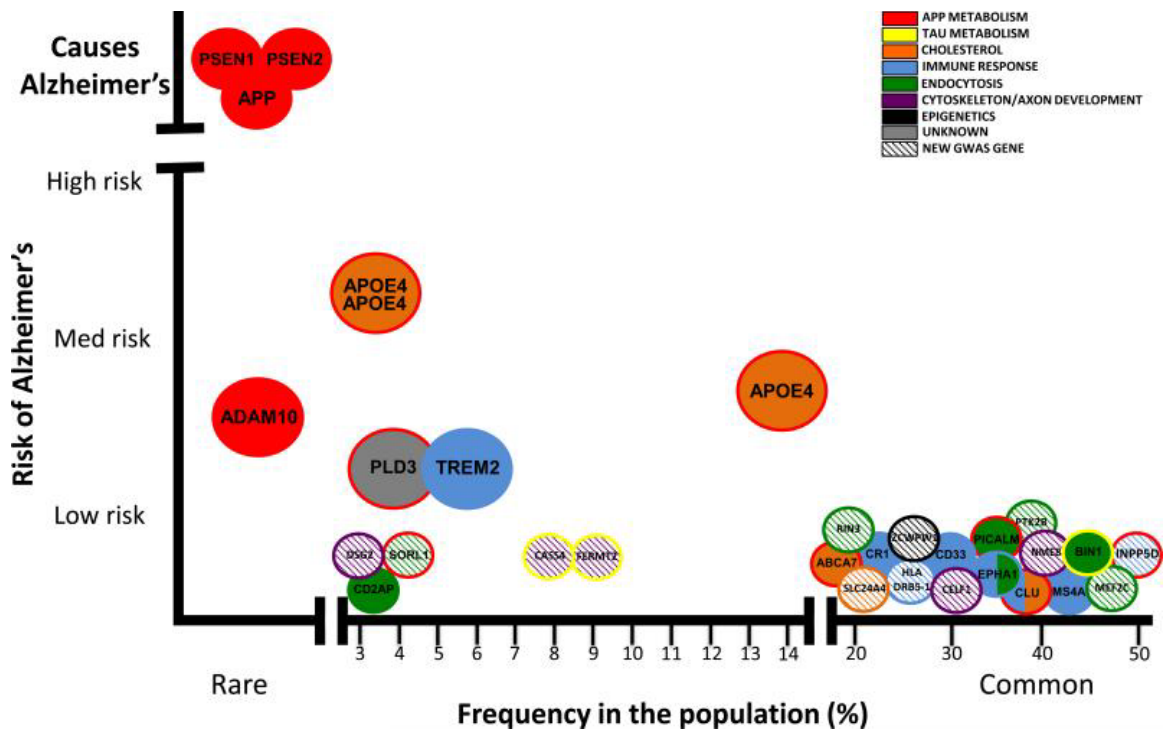
Figure 1.2: Alzheimer's Disease risk variants as a function of frequency in population.

The frequency of the most impactful genetic variant for a given gene is plotted on the X-axis, while its impact on risk of AD is plotted on the Y-axis. Note that in the upper left corner, variants in *APP*, *PSEN1*, and *PSEN2* are shown as "Causes Alzheimer's" as these are FAD-associated genes. Other variants, such as those in *TREM2* and *CD33* are in the medium- to low-risk of AD.
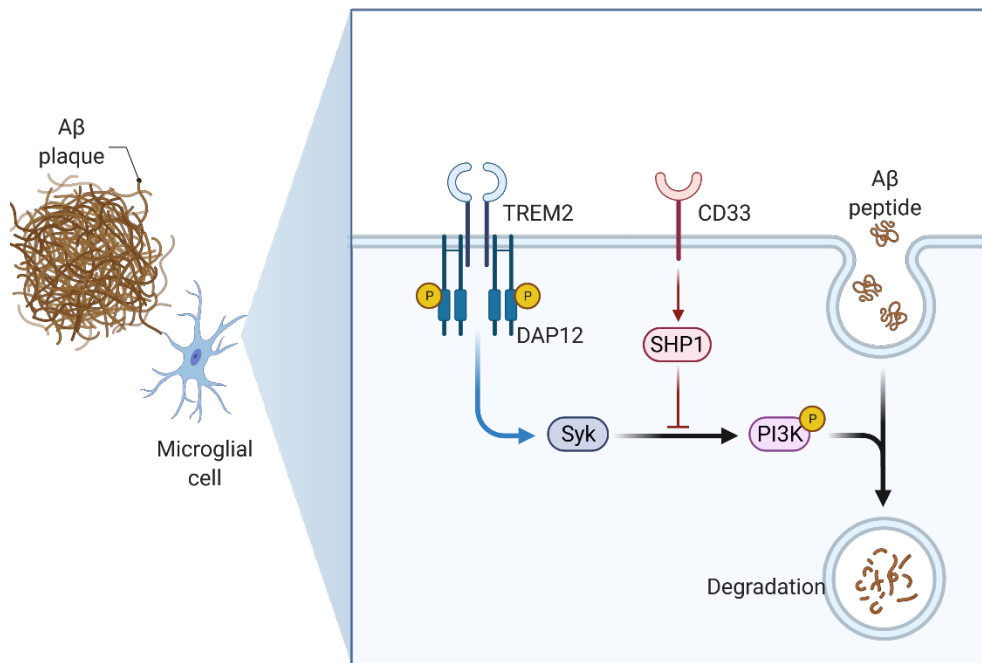
Figure 1.3: TREM2 and CD33 act in opposing fashion.

As noted above, signaling from TREM2 to DAP12 is mediated by a positively charged amino acid in the TREM2 transmembrane domain. Clustering of TREM2-DAP12 after ligand binding or antibody-based ligation leads to Src family kinase recruitment and phosphorylation of the tyrosines within the DAP12 ITAM. This leads to downstream signaling via Syk, PLC, and phosphoinositde-3-kinase, ultimately facilitating phagocytosis of Aβ for degradation. CD33 acts through its ITIM domain to recruit SHP1, and to a lesser extent SHP2, which dephosphorylates Syk and interrupts this signaling pathway, thus decreasing the phagocytic capacity of microglia.

CHAPTER 2. A NOVEL TREM2 SPLICE ISOFORM LACKING THE LIGAND BINDING DOMAIN IS EXPRESSED IN BRAIN AND SHARES LOCALIZATION

[This section contains material adapted from a submitted manuscript: Shaw, B. C., Snider, H. C., Turner, A. K., Zajac, D. J., Simpson, J. F., & Estus, S. (2021). A novel TREM2 splice isoform lacking the ligand binding is expressed in brain and shares localization. *bioRxiv*. Submitted to *J. Alz. Dis.* Dec. 2021.]

2.1 Introduction

TREM2 is an activating receptor expressed on innate immune cells, and genetic variants in the *TREM2* gene are associated with both Nasu-Hakola disease and Alzheimer's Disease (AD) (R. Guerreiro *et al.*, 2013; Jonsson *et al.*, 2013; Paloneva *et al.*, 2002). Genome-wide association studies (GWAS) have confirmed that variants rs75932628 and rs143332484, encoding the p.R47H and p.R62H variants, respectively, in *TREM2* are strong risk factors for developing late-onset AD (LOAD) (Jansen *et al.*, 2019; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Wightman *et al.*, 2021). These variants reduce TREM2 function (Cosker *et al.*, 2021; Dean *et al.*, 2019; Gratuze *et al.*, 2020; Piers *et al.*, 2020). TREM2 is a receptor for both Apolipoprotein E (ApoE) (Atagi *et al.*, 2015) and amyloid beta (Aβ) (Y. Zhao *et al.*, 2018), and regulates Aβ phagocytosis (McQuade *et al.*, 2020; Piers *et al.*, 2020; Y. Zhao *et al.*, 2018), transcriptional changes (Holtman *et al.*, 2015), and microglial transition to a full disease-associated phenotype (Keren-Shaul *et al.*, 2017). Murine models of AD suggest TREM2 may be beneficial early in the disease but detrimental later; *Trem2^{KO}* or *Trem2^{R47H}* mice crossed with amyloid-based AD models (APP/PS1 or 5xFAD mice) develop greater Aβ

pathology (A. Griciuc *et al.*, 2019; Wang *et al.*, 2015), but when crossed with the PS19 humantau-model, reduced tau pathology (Gratuze *et al.*, 2020).

TREM2 is a five-exon gene that has been reported to undergo alternative splicing (Del-Aguila *et al.*, 2019; Numasawa *et al.*, 2011; Yanaizu *et al.*, 2018), wherein exon 3 (*D3-TREM2*) or exon 4 (*D4-TREM2*) are skipped, or exon 4 is extended to a include a 3' portion of intron 3. Each of these three alternative splicing isoforms results in a frameshift mutation; exon 3 skipped has been associated with Nasu-Hakola disease as well (Numasawa *et al.*, 2011; Yanaizu *et al.*, 2018). Interestingly, though both isoforms encode proteins which lack a transmembrane domain and are expected to be secreted, the exon 4 variants have not been reported as associated with Nasu-Hakola disease possibly due to the lack of an identified causal genetic variant for this isoform. Modulation of *TREM2* splicing has been proposed as a potential therapeutic for Nasu-Hakola disease previously (Yanaizu *et al.*, 2018), similar to the recent successes with spinal muscular atrophy (Finkel *et al.*, 2017). Further TREM2 splice variants had not been reported until very recently (Kiianitsa *et al.*, 2021), when an isoform lacking exon 2 (*D2-TREM2*), which encodes the ligand binding domain, was identified. This *D2-TREM2* isoform maintains the reading frame and transmembrane domain but lacks the ligand binding domain.

In this study, we sought to fully characterize *TREM2* alternative splicing in brain. We identify many more alternative splice isoforms than previously reported, and report that this alternative splicing is not brain-specific as it is conserved across multiple tissues. Further, we show that the *D2-TREM2* splice isoform is

23

translated into protein using overexpression paradigms, and that this D2-TREM2 protein has similar compartment localization as the full-length (FL-TREM2) protein. We propose that modulating *D2-TREM2* could be exploited to enhance TREM2 function—by decreasing *D2-TREM2* early, one could increase the then-beneficial functional TREM2; by increasing *D2-TREM2* late in disease, one could inhibit the then-detrimental functional TREM2.

2.2   Methods

2.2.1   Preparation of DNA, RNA, and cDNA from human samples

Human blood and anterior cingulate autopsy tissue from 61 donors were generously provided by the Sanders-Brown Alzheimer's Disease Center neuropathology core and their characteristics and cDNA synthesis have been described elsewhere (Zou *et al.*, 2007). All human subjects research was carried out in accordance with the University of Kentucky Institutional Review Board under protocol number 48095. The matched brain and blood samples were from deceased individuals with an average age at death of 82.4 ± 8.7 (mean $\pm$ SD) years for non-AD and 81.7 ± 6.2 years for AD subjects. The average postmortem interval (PMI) for non-AD and AD subjects was 2.8 ± 0.8 and 3.4 ± 0.6 hrs, respectively. Non-AD and AD samples were comprised of 48% and 55% female subjects, respectively. MMSE scores were, on average, 28.4 ± 1.6 for non-AD subjects and 11.9 ± 8.0 for AD subjects. Total RNA was prepared using a Qiagen RNeasy Lipid Tissue Mini kit (Qiagen #74804) according to manufacturer's instructions. Reverse transcription was carried out using SuperScript IV (Invitrogen

#18091050) according to manufacturer's instructions. For cross-tissue splicing comparison, fetal RNA libraries from various tissues were obtained from a commercial vendor (Stratagene) and their cDNA preparation has been described elsewhere (Burchett *et al.*, 2011).

Cell Culture

The HMC3 human microglial cell line was obtained from American Type Culture Collection (ATCC CRL-3304). Cells were cultured in Eagle's Modified Minimum Essential Medium (EMEM), ATCC modification (ATCC 30-2003) supplemented with 10% fetal bovine serum, defined (HyClone, GE Healthcare SH30070.03); 50 U/mL penicillin, 50 µg/mL streptomycin (Gibco 22400-089). Cells were maintained at 37°C in a 5% CO_2 in air atmosphere.

2.2.2 TREM2 splice isoform identification by PCR and sequencing

The cDNA samples from the anterior cingulate samples were amplified using primers corresponding to *TREM2* exon 1 (5'-CCTGACATGCCTGATCCTCT-3') and exon 5 (5'-GTGTTCTTACCACCTCCCC-3') with Q5 high-fidelity hot-start polymerase (NEB # M0493L). Thermocylcing parameters were as follows: 98°C 30 s; 98°C 5 s, 67°C 5 s, 72°C 45 s, 30 cycles; 72°C 2 min, 25°C hold. PCR products were separated on a 8% acrylamide gel and imaged using a BioRad ChemiDoc XR. Bands were extracted for subsequent amplification as above and purification using a Monarch PCR Cleanup Kit (NEB T1030L). Purified products were sequenced commercially (ACGT; Wheeling, IL) and compared to the reference transcript NM_018965.4 to determine splicing.

2.2.3 Quantification of TREM2 transcripts

Quantitative PCR (qPCR) was used to quantify expression of *TREM2* transcripts. Primers corresponding to sequences within exons 1 and 2 were used to quantify *TREM2* exon 2 expression (forward, 5'-CCTTGGCTGGGGAAGGG-3'; reverse, 5'-TCATAGGGGCAAGACACCTG-3'), as well as primers corresponding to sequences at the exon 1–3 junction and within exon 3 to quantify the *D2-TREM2* isoform (forward, 5'-TTACTCTTTGTCACAGACCCC-3'; reverse, 5'-GGGCATCCTCGAAGCTCT-3'). PCR was conducted using an initial 2 min incubation at 95°, followed by 40 cycles of 10 s at 95°C, 20 s at 60°C, and 20 s at 72°C. The 20 µl reactions contained 1 µM each primer, 1X PerfeCTa SYBR Green Super Mix (Quanta Biosciences), and 20 ng of cDNA. Experimental samples were amplified in parallel with serially diluted standards that were generated by PCR of cDNA using the indicated primers, followed by purification and quantitation by UV absorbance. Results from samples were compared relative to the standard curve to calculate copy number in each sample. Total TREM2 expression was the sum of the copy numbers for TREM2 exon 2 present and exon 2 skipped. Assays were performed in duplicate and normalized to expression of Iba1 (*AIF1*) as the housekeeping gene, as *TREM2* in the CNS is exclusively expressed in microglia. For cross-tissue comparison, the percent exon 2 skipping was calculated by dividing the exon 2 skipped copies by the sum of the exon 2 skipped and mean of exon 2 present copies without normalization. Data in the cross-tissue comparison reflect six technical replicates.

2.2.4 TREM2 transcript cloning

The full-length *TREM2* and *D2-TREM2* transcripts were amplified from gel extracts as above and genomic DNA was amplified using the same primers as above. Amplification was performed with Platinum Taq (Invitrogen 10966034) with the following cycling parameters: 2 min at 94°C; 30 s at 94°C, 30 s at 60°C, 2 min at 72°C, 30 cycles; 7 min at 72°C, 25°C hold. All cloning was performed using a pcDNA 3.1-V5/His TOPO-TA cloning kit (Invitrogen K480001) per manufacturer's instructions. Clones were verified by sequencing (ACGT; Wheeling, IL) and grown for midi-scale production and purification using a Qiagen Plasmid Plus Midiprep kit (Qiagen 12943).

2.2.5 HMC3 Transfection

HMC3 human microglia were transfected with Lipofectamine 3000 with Plus reagent (Invitrogen L3000001) per manufacturer instructions with 0.8 µL of Lipofectamine 3000, 1 µL Plus reagent, and 250 ng plasmid per well in 8 well glass chamber slides (MatTek CCS-8). Cells were incubated for 24 hours prior to processing for microscopy.

2.2.6 Confocal Immunofluorescence Microscopy

Transfected HMC3 cells were fixed with 10% neutral buffered formalin (Fisher Scientific SF100-4) for 30 minutes then blocked and permeabilized for 30 minutes with 10% goat serum (Sigma S26-LITER), 0.1% Triton X-100 (Fisher Scientific BP151-500) in PBS (Fisher BioReagents BP665-1). Primary and secondary antibodies were diluted in the same blocking and permeabilization

buffer and incubated at room temperature for 90 minutes. Cells were washed three times in blocking and permeabilization buffer between primary and secondary antibodies, and three times in PBS prior to coverslip mounting with Prolong Glass with NucBlue mounting media (Invitrogen P36981) and high-tolerance No. 1.5 coverglass (ThorLabs CG15KH1). Images were acquired using a Nikon A1R HD inverted confocal microscope with a 60X oil objective and NIS Elements AR software.

2.2.7 Statistical Analyses

Analyses were performed using GraphPad Prism 8.4.2. Quantitative data were first checked for normality by the D'Agostino & Pearson test. Normally distributed data were analyzed using a two-tailed t-test, while data not normally distributed were analyzed with a two-tailed Mann Whitney test and are noted along with p values.

2.3 Results

2.3.1 TREM2 undergoes extensive alternative splicing in human adult brain tissue

To fully characterize *TREM2* alternative splicing, we PCR-amplified *TREM2* cDNA from anterior cingulate cortex by using primers corresponding to sequences within exons 1 and 5. A gene map, including introns and encoded protein domains, is shown with forward and reverse primers (Figure 2.1, top). We observed substantial alternative splicing in both AD and non-AD individuals (Figure 2.1, left). The identity of each isoform was confirmed by direct sequencing (Figure 2.1, right).

This effort identified multiple novel TREM2 isoforms that have not been reported previously, including isoforms with multiple exons skipped.

Since this assay suggested that the the isoform lacking exon 2 (*D2-TREM2*) is the most abundant variant, we next quantified *TREM2* and this alternate *D2-TREM2* isoform in a series of brain samples. Since *TREM2* is almost exclusively expressed by microglia in the brain, we normalized the *TREM2* copy number to that of *AIF1* (Figure 2.2A). While we do observe a trend toward higher *TREM2* in the AD group, this difference is not significant ($p = 0.1268$, two-tailed t- test). There was an association between total *TREM2* expression and neuropathology when parsed by National Institute on Aging/Reagan Institute (NIARI) scores, where low pathology was defined as NIARI < 3 and high pathology was defined as NIARI ≥ 3 (Figure 2.2B; $p = 0.0033$, Student's t-test). We next investigated whether exon 2 skipping correlated with exon 2 inclusion and found a strong correlation irrespective of AD diagnosis (Figure 2.2C). We then investigated whether *D2-TREM2* occurred more frequently in AD vs. non-AD individuals. We found that exon 2 skipping correlates well with total TREM2 expression (Figure 2.2D), but observed no difference in exon 2 skipping frequency when parsed by AD diagnosis (Figure 2.2D; $p = 0.4909$, Mann Whitney test) or NIARI scores (data not shown; $p = 0.9443$, Welch's t-test).

2.3.2 TREM2 alternative splicing is conserved across tissues

To test whether the observed alternative splicing is specific to brain, we subjected cDNA libraries from aorta, lung, kidney, heart, skeletal muscle, brain, and liver to PCR amplification with primers in TREM2 exons 1 and 5, as described

above (Figure 2.1). We observed each of the previous splice isoforms across multiple tissues, though each isoform was not present in all tissues (Figure 2.3). We also observed a high relative abundance of the isoform lacking both exon 2 and 3 along with the isoform lacking exons 2 and 4. This may reflect PCR bias, where shorter fragments are amplified more efficiently than longer fragments and are overrepresented in relative abundance. Nonetheless, the *D2-TREM2* isoform was an abundant alternative isoform across all tissues surveyed. We then quantified *TREM2* exon 2 skipping frequency in these samples and found similar frequency of exon 2 skipping across tissues (Figure 2.4).

2.3.3   D2-TREM2 protein localizes similar to full-length TREM2

To test whether the D2-TREM2 protein isoform is trafficked similarly to the full-length isoform, we cloned each isoform into expression vectors and transfected HMC3 human microglial cells for confocal microscopy (Figure 2.5). We observed similar staining patterns from both the *FL-TREM2* (Figure 2.5A) and *D2-TREM2* (Figure 2.5B), and this staining pattern is consistent with intracellular retention prevoiusly reported (Prada *et al.*, 2006; Sessa *et al.*, 2004). This implies the D2-TREM2 protein maintains the localization pattern of its parent full-length TREM2. We confirmed the D2-TREM2 isoform is predominantly retained in the Golgi complex (Figure 2.5C) as has been previously reported for the FL-TREM2 and shown here (Figure 2.5D).

2.4 Discussion

The primary findings of this report are that *TREM2* undergoes far more extensive alternative splicing than has been previously reported (Del-Aguila *et al.*, 2019; Kiianitsa *et al.*, 2021; Numasawa *et al.*, 2011; Yanaizu *et al.*, 2018), that D2-TREM2 is a common variant that is not influenced by AD neuropathology, and that D2-TREM2 co-localizes with full-length TREM2. As such, this comprehensive analysis of *TREM2* alternative splicing extends prior knowledge, noting that Ensembl lists only full-length TREM2 (ENST00000373113.8), exon 4 skipped (*D4-TREM2;* ENST00000338469.3), and intron 3 retained isoforms (ENST00000373122.8) as known transcripts; NCBI only lists the full length (NM_018965.4) and *D4-TREM2* (NM_001271821.2). Exon 3 skipping has been previously described in two small cohorts (Numasawa *et al.*, 2011; Paloneva *et al.*, 2002). The *D2-TREM2* isoform is not annotated in either the NCBI nor Ensembl database. The *D2-TREM2* isoform in brain was recently reported (Han *et al.*, 2021; Kiianitsa *et al.*, 2021), and our brain quantitation data closely replicate their results (Figure 2.2C). In addition, we demonstrate that this splice isoform is expressed across multiple tissues (Figure 2.3) with roughly equal relative abundance (Figure 2.4). We further extend these recent findings by demonstrating that isoforms exist with multiple exons skipped, such that all permutations of exons 2, 3, and 4 can be skipped individually or in combination (Figure 2.1), and that exon 2 skipping frequency is not different as a function of AD neuropathology.

While TREM2 has well-described cell surface localization including receptor activity (Atagi *et al.*, 2015; A. Griciuc *et al.*, 2019; McQuade *et al.*, 2020; Y. Zhao

et al., 2018), interactions with DAP12 (Peng *et al.*, 2010), and proteolytic cleavage by ADAM10 (Schlepckow *et al.*, 2017; Zhong *et al.*, 2015), there also exists a considerable intracellular TREM2 pool (Prada *et al.*, 2006; Sessa *et al.*, 2004). Our localization studies support the hypothesis that D2-TREM2 localizes to the same compartment as FL-TREM2. Both appear to be predominantly localized to the Golgi complex (Figure 2.5C-D), replicating previous work with FL-TREM2 (Prada *et al.*, 2006; Sessa *et al.*, 2004). The recent *D2-TREM2* report (Kiianitsa *et al.*, 2021) provided evidence of membrane-bound D2-TREM2 and FL-TREM2 by Western blots on subcellular fractions. Taken together, this suggests that TREM2—and likely D2-TREM2—resides in a transmembrane Golgi pool.

Functionally, TREM2 seems to operate under a feed-forward mechanism which D2-TREM2 may act to inhibit. Experiments *in vitro* suggest that the transmembrane Golgi pool rapidly moves to the cell membrane in response to intracellular calcium flux induced by ionomycin (Prada *et al.*, 2006; Sessa *et al.*, 2004). Ligation of TREM2 using monoclonal antibodies also elicits an intracellular calcium spike. Whether D2-TREM2 inhibits this feed-forward mechanism, where TREM2 signaling induces calcium flux and increases cell surface mobilization, is still unknown.

TREM2 variants were some of the first genetic risk factors outside of *APOE4* identified for LOAD (R. Guerreiro *et al.*, 2013; Jonsson *et al.*, 2013), have been replicated in multiple GWAS (Jansen *et al.*, 2019; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Wightman *et al.*, 2021), and have the highest odds ratios for LOAD after *APOE4* (Karch & Goate, 2015). Variants encoding the p.R47H

(rs75932628) and p.R62H (rs143332484) disrupt ligand binding and are predicted to be partial loss-of-function mutations (Cosker *et al.*, 2021; Dean *et al.*, 2019; Gratuze *et al.*, 2020; Piers *et al.*, 2020). We hypothesize that the D2-TREM2 protein is a functionally null receptor, as the ligand binding IgV domain is missing (Figure 2.1), though this has yet to be confirmed. In a similar case, *CD33* has an isoform with relatively high abundance that is also lacking its ligand binding IgV domain (*D2-CD33*); however genetics (Estus *et al.*, 2019), mouse models (Bhattacherjee *et al.*, 2021), and *in vitro* studies with human pluripotent stem cell-derived microglia (Jannis Wißfeld *et al.*, 2021; J. Wißfeld *et al.*, 2021) suggest the D2-CD33 protein may have a gain of function which acts independent of its loss of receptor activity.

TREM2 remains a promising target in potential AD therapeutics, evidenced by the strong interest in preclinical TREM2 monoclonal antibody treatments including a Phase 2 trial of AL002a (Alector, NCT04592874). Modulation of alternative splice isoforms of *TREM2* may represent a novel therapeutic pathway. The previously reported *D4-TREM2* isoform lacks a transmembrane domain (Figure 2.1) and is likely secreted as soluble TREM2 (sTREM2) which may be AD-protective (Ewers *et al.*, 2020; Zhong *et al.*, 2019). We posit that the *D2-TREM2* isoform acts similarly to the p.R47H and p.R62H variants by reducing TREM2 function, and balanced with *D4-TREM2*, this forms a potentially druggable rheostat. Early in the disease, increased TREM2 function and sTREM2 secretion are apparently beneficial; increasing full-length *TREM2* would promote functional TREM2 signaling and increased *D4-TREM2* would promote increased sTREM2

secretion. Later in disease when inflammation may be detrimental, increasing *D2-TREM2* could provide a switch to decrease TREM2 activity and inflammation (Figure 2.6).

We note one outlying point in our quantitation data (Figure 2.2C, AD subject, top) with substantially higher *D2-TREM2* than the rest of our subjects. This subject may have an underlying genetic variant which regulates exon 2 skipping in *cis*, and current work seeks to better understand this outlier. Future studies will examine in greater depth using RNA sequencing methods whether this splice isoform is associated with any genetic variants, such as the p.R47H and p.R62H variants or other rare mutations.
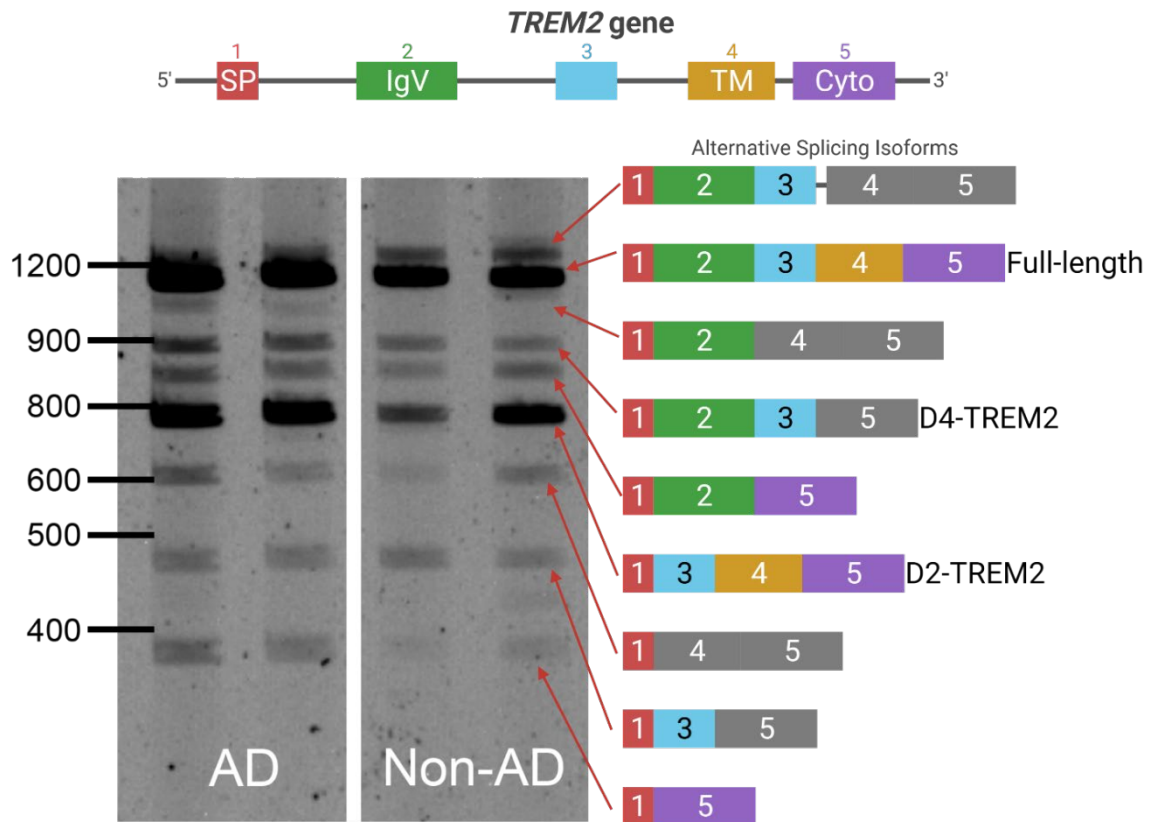
Figure 2.1: *TREM2* undergoes extensive alternative splicing

Top: a schematic of the *TREM2* gene is shown with introns and exons. Approximate locations of the forward (green) and reverse (red) primers are shown. Left: a representative image of the PCR amplification and gel electrophoresis of *TREM2* cDNA from AD and non-AD brains. No differences in splicing patterns were noticed between AD and non-AD. Right: schematic of the splice isoforms identified after sequencing. Colors correspond to exons in the gene model, while frameshifts are shown in grey. A doublet appears in the lower transcripts on the gel in which the lower band corresponds to *CES3* in addition to the identified *TREM2* isoforms. All bands identified in Figure 2.1 have been confirmed by Sanger sequencing.
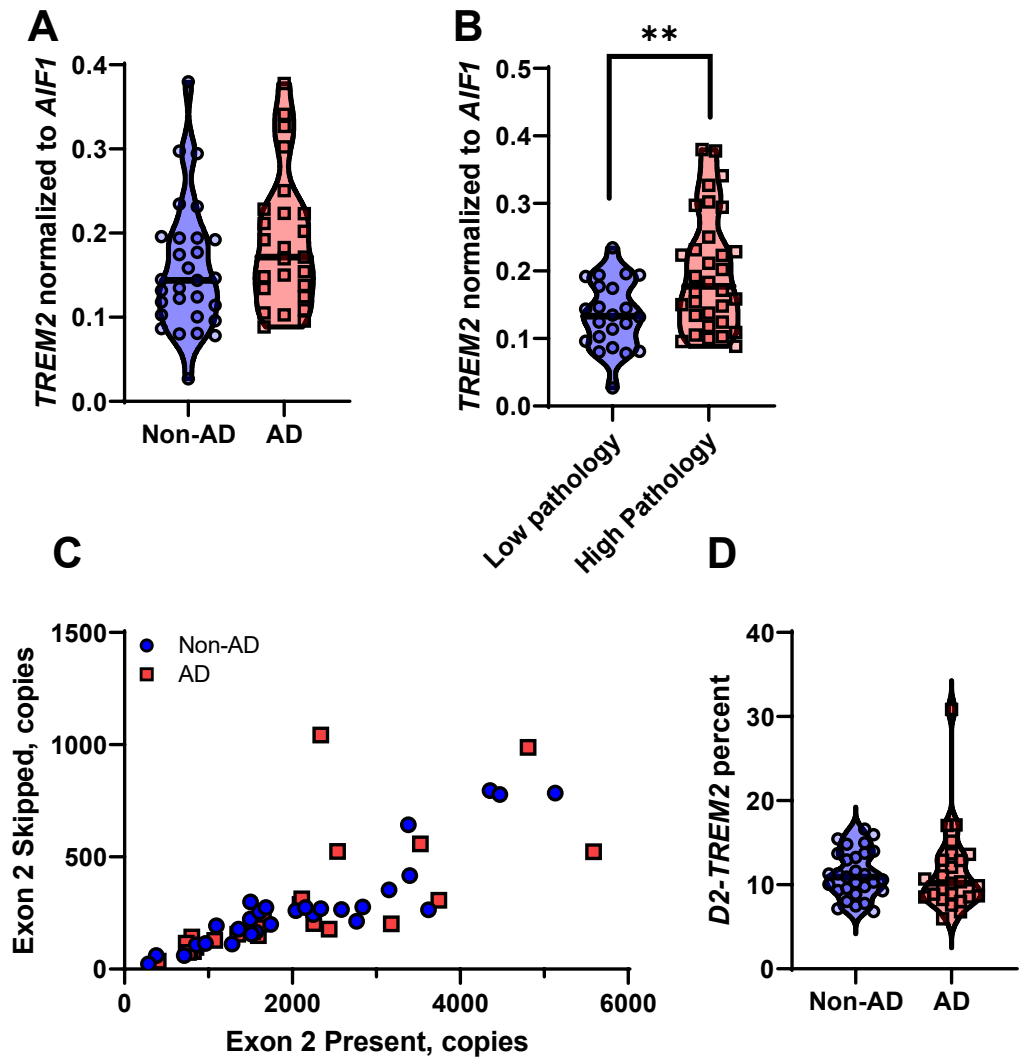
Figure 2.2: Quantification of *TREM2* and exon 2 skipping in human brain tissue

(A): Total *TREM2* expression normalized to the microglial marker *AIF1* expression. *TREM2* expression is not significantly different between AD and non-AD samples (p = 0.1268, t-test). (B) Total *TREM2* expression normalized to microglial marker *AIF1* expression, parsed by NIARI pathology state, where low pathology is defined as a NIARI score < 3 and high pathology is defined as a NIARI score ≥ 3 (p = 0.0033, Student's t-test). (C) Expression of the isoform lacking exon 2 correlates well with expression of the isoform containing exon 2. (D) Exon 2 is

skipped at an approximate frequency of 11%, with no significant difference between AD vs. non-AD (p = 0.4909, Mann-Whitney test) or NIARI pathology (p = 0.9443, Welch's t-test, data not shown).
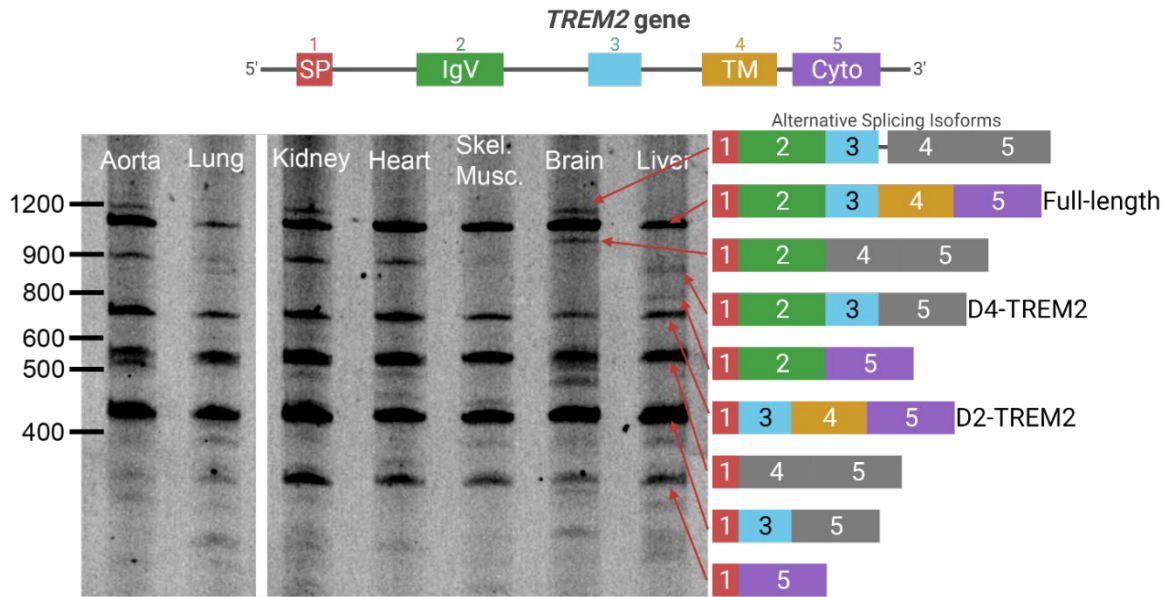
Figure 2.3: Complex patterns of *TREM2* alternative splicing are present in many tissues

Human fetal cDNA libraries from multiple tissues were amplified using the same primers from Figure 2.1. The splice variants from Figure 2.1 are replicated across these six additional tissues. Differences in the relative abundance of each splice variant may be due to differences in developmental stage and/or splicing factor differences between tissues.
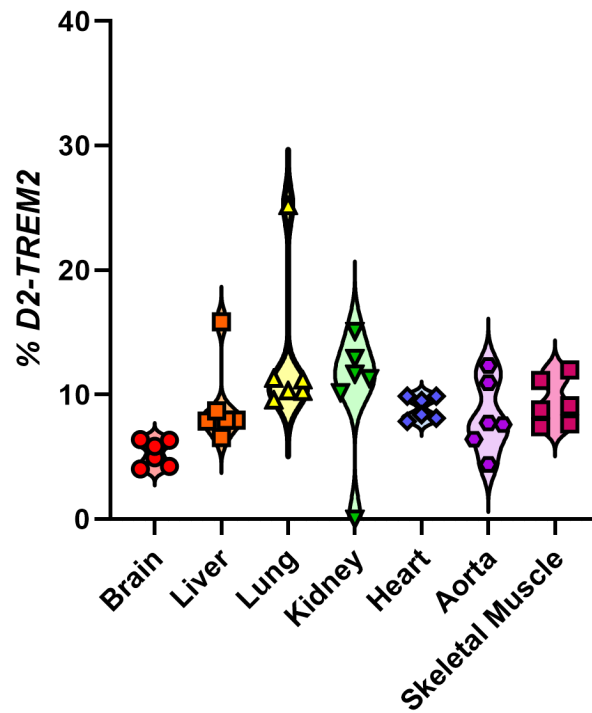
Figure 2.4: Quantification of *TREM2* and exon 2 skipping across tissues

Exon 2 is skipped at a frequency between 5.30 – 13.0%. Data points reflect technical, not biological, replicates from pooled cDNA libraries.
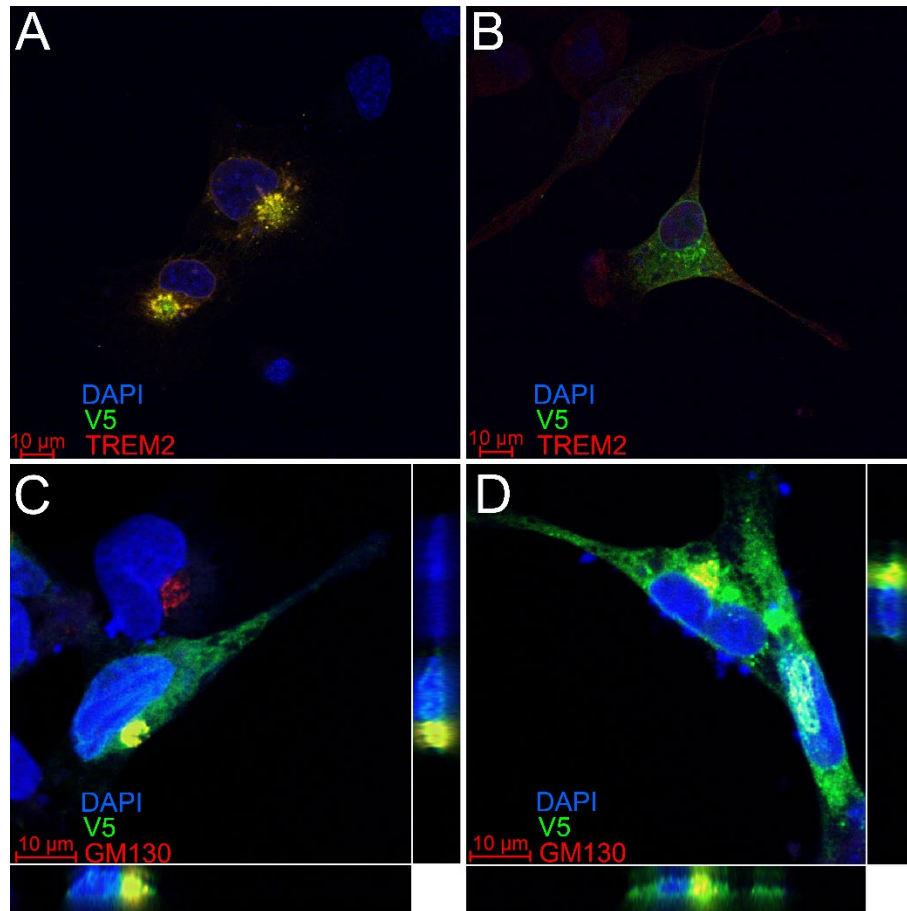
Figure 2.5: TREM2 and D2-TREM2 have a similar subcellular localization

The HMC3 human microglial cell line was transfected with vectors encoding either *full-length TREM2* (A,C) or *D2-TREM2* (B,D). Both vectors have an in-frame V5 epitope tag. In A and B, cells were subsequently labeled with antibodies against a TREM2 epitope encoded by exon 2 (red) or V5 (green). The limited red fluorescent labeling and absence of yellow overlap in the D2-TREM2 transfected cells (B) is due to low endogenous TREM2 expression. In C and D, transfected cells were labeled with antibodies against the Golgi complex marker GM130 (red) and V5 (green). The intracellular pools of both TREM2 (C) and D2-TREM2 (D)

largely colocalize with GM130. Views represent the XY (main), XZ (bottom) and YZ (right) views in each panel.
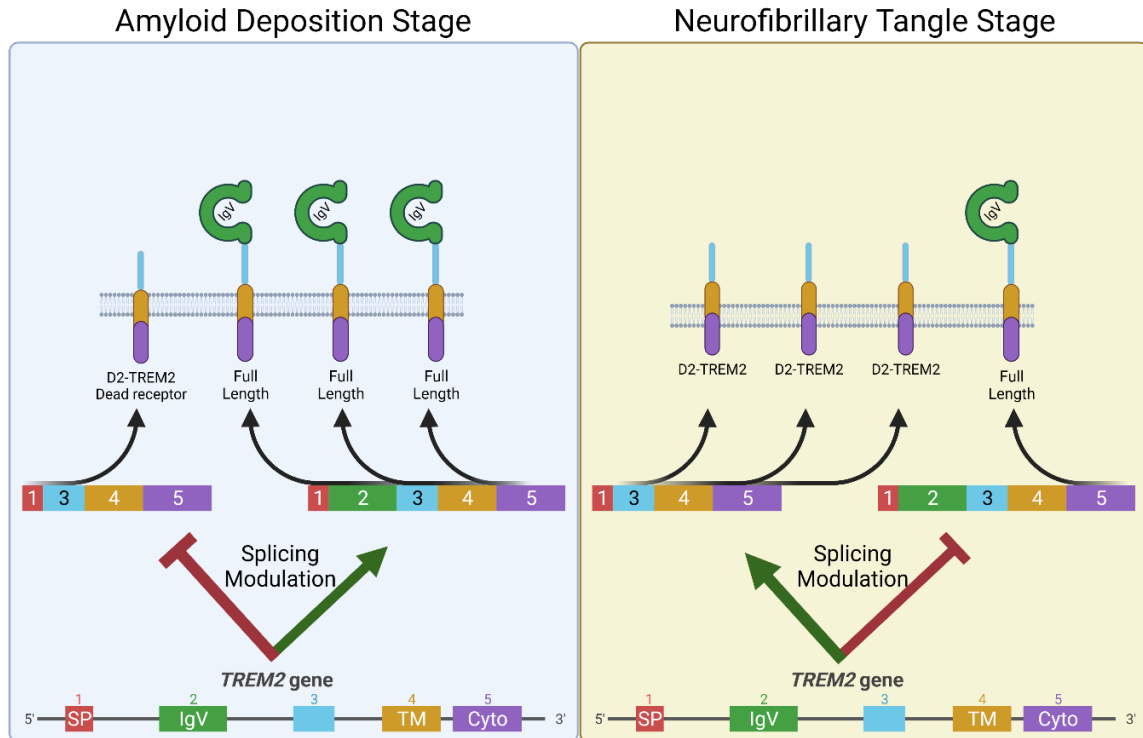
Figure 2.6: Model to exploit alternative splicing in *TREM2* as a potential AD therapeutic.

Early in disease when Aβ pathology is developing, animal models suggest TREM2 signaling is protective. Hence, decreasing *D2-TREM2* and increasing full-length *TREM2* may be helpful in this stage of the disease (Left). Animal models also indicate TREM2 deficiency is protective from tau-related pathology suggesting a detrimental role for TREM2 signaling in this later stage of the disease when hyperphosphorylated tau accumulates. At this point, increased skipping of exon 2 to promote the "dead receptor" D2-TREM2 and thereby inhibit TREM2 signaling may be helpful (Right).

CHAPTER 3. ECTOPIC GENE CONVERSION OF CD33 AFTER CRISPR-CAS9 EDITING LEADS TO AN IN-FRAME, CHIMERIC SIGLEC22P-CD33 PROTEIN

[This section contains material adapted from a published manuscript: Shaw, B. C., & Estus, S. (2021). Pseudogene-Mediated Gene Conversion After CRISPR-Cas9 Editing Demonstrated by Partial CD33 Conversion with SIGLEC22P. *CRISPR J*. 2021 Oct;4(5):699-709. PMID: 34558988.]

3.1   Introduction

The CRISPR–Cas9 system has revolutionized gene editing (Jinek *et al.*, 2012). In this process, a single-stranded guide RNA (sgRNA) directs Cas9 endonuclease to cleave DNA at a sequence-specific site. The DNA cleavage results in a double stranded DNA break (DSB) that is repaired by either homology-directed repair (HDR) or nonhomologous end-joining (NHEJ). The former results in targeted integration of DNA sequence while the latter typically results in gene disruption through the introduction of insertions or deletions (indels). To generate a targeted knock-in or knock-out, an HDR template of exogenous DNA is often supplied as part of the process (Renaud *et al.*, 2016). Alternatively, endogenous HDR templates have been described, including *HBD* sequence being incorporated into *HBB* and sequence from one allele of *HPRT* being incorporated into the other allele (Javidi-Parsijani *et al.*, 2020; Susani *et al.*, 2018). However, to our knowledge, HDR directed by a pseudogene has not been previously reported.

CD33 genetic variants, including rs124594919, have been associated with reduced risk of Alzheimer's Disease (AD) in genome-wide studies (Hollingworth *et*

al., 2011; Jansen *et al.*, 2019; Naj *et al.*, 2011). We and others subsequently identified rs12459419 as a functional SNP that increases the proportion of *CD33* lacking exon 2 (*D2-CD33*) (Bradshaw *et al.*, 2013; Ana Griciuc *et al.*, 2013; Malik *et al.*, 2015a; Raj *et al.*, 2014; Schwarz *et al.*, 2016). This exon encodes the ligand-binding IgV domain of this member of the sialic acid-binding immunoglobulin-type lectin (SIGLEC) family (Estus *et al.*, 2019). Hence, while the extracellular portion of CD33 normally includes an IgV and IgC_2 domain, *D2-CD33* encodes a protein with only the IgC_2 domain (Malik *et al.*, 2013). CD33 inhibits microglial activity through its immunomodulatory tyrosine inhibitory motif (ITIM) and ITIM-like domains, which recruit protein tyrosine phosphatases, SHP1 and SHP2, to impact intracellular calcium flux, phagocytosis, and microglial migration (Balaian *et al.*, 2003; Bhattacherjee *et al.*, 2019; Bradshaw *et al.*, 2013; Ana Griciuc *et al.*, 2013; Hernández-Caselles *et al.*, 2006; Paul *et al.*, 2000; Perez-Oliva *et al.*, 2011; Raj *et al.*, 2014; Walter *et al.*, 2008).

Given that the AD-protective rs12459419 increases *D2-CD33* at the expense of *CD33*, the prevailing theoretical mechanism has been that rs12459419 reduces AD risk through decreased CD33 function. However, recent findings that a *bona fide* loss of function indel, rs201074739, is not associated with AD risk, has led to this hypothesis being revised to suggest that rs12459419 and its related *D2-CD33* isoform represent a gain of function (Bhattacherjee *et al.*, 2021; Estus *et al.*, 2019; J. Wißfeld *et al.*, 2021). The gain-of-function mechanism and localization of D2-CD33 protein remain heavily debated (Bhattacherjee *et al.*, 2021; Bhattacherjee *et al.*, 2019; Estus *et al.*, 2019; Godwin *et al.*, 2020; Ana Griciuc *et*

al., 2013; Humbert *et al.*, 2019; Malik *et al.*, 2015a; Perez-Oliva *et al.*, 2011; Raj *et al.*, 2014; Siddiqui *et al.*, 2017; J. Wißfeld *et al.*, 2021).

Here, we sought to generate a model of physiologic *D2-CD33* expression by using CRISPR-Cas9 to excise *CD33* exon 2 in the U937 human monocyte cell line. During these experiments, we identified a subset of cells which apparently underwent HDR directed by the *SIGLEC22P* pseudogene, located 13.5 kb away from *CD33*. Although the *SIGLEC22P* pseudogene shares approximately 87% identity over 1800 bp with *CD33*, this gene conversion was detected because three nucleotides in *SIGLEC22P* differ from those within the targeted *CD33* exon 2 and result in three missense amino acids in CD33, including p.N20K, p.F21I, and p.W22R. Hence, we report pseudogene directed gene conversion as a mechanism for unanticipated CRISPR mutations.

3.2 Methods

3.2.1 Cell Lines and Antibodies

U937 and HEK293 cell lines were obtained from American Type Culture Collection (ATCC). U937 cells were cultured in RPMI 1640 with HEPES (Gibco 22400-089) supplemented with 10% fetal bovine serum, defined (HyClone, GE Healthcare SH30070.03); 50 U/mL penicillin, 50 µg/mL streptomycin (Gibco 15070-063); and 2 µM L-glutamine (Gibco A2916801). HEK293 cells were cultured in EMEM, ATCC formulation (ATCC 30-2003) supplemented with 10% fetal bovine serum, defined (HyClone, GE Healthcare SH30070.03); 50 U/mL penicillin, 50 µg/mL streptomycin (Gibco 22400-089). Cells were maintained at 37°C in a 5% CO_2 in air atmosphere. The U937 cell line has been reported as either diploid or

triploid at chromosome 19 which contains *CD33* (Lee *et al.*, 2002; Shipley *et al.*, 1988). Antibodies, concentrations, and CD33 domains targeted are shown in Table 3.1.

3.2.2 CRISPR-Cas9 Gene Editing

All CRISPR reagents were purchased from Integrated DNA Technologies (IDT). Single-guide RNAs (sgRNAs) and Cas9 protein (IDT 1081059) were incubated at a 1:1 molar ratio (0.5 nmol each) at room temperature for 10 minutes to form ribonucleotide-protein complexes (RNPs). The sgRNA sequences targeting CD33 exon 2 were 5'-TCCATAGCCAGGGCCCCTGT-3' and 5'-GCATGTGACAGGTGAGGCAC-3' (Humbert *et al.*, 2019). U937 cells were washed three times in PBS (Gibco 10010-023) and resuspended in complete Nucleofector Kit C (Lonza Biosciences VCA-1004) media (10^6 cells per transfection) with 5 µL electroporation enhancer (IDT 1075916) and RNPs. Cells were electroporated using a Nucleofector IIb device (Lonza Biosciences) under protocol V-001 and immediately added to a 12 well plate with 1.5 mL complete media and cultured for two weeks.

3.2.3 Cell Sorting and Flow Cytometry

Edited U937 cells were washed in PBS with 5% heat-inactivated fetal bovine serum (Gibco 10082-147), resuspended at 10^6 cells/mL and then treated with Human TruStain FcX blocker (BioLegend 422302). Cell sorting was carried out in azide-free buffers; for flow cytometry, 0.02% sodium azide was included in all buffers. Cells were stained with HIM3-4-FITC and P67.6-BV711 for one hour on

ice, washed twice with HBSS, then stained with Fixable Viability Dye eFluor780 (Invitrogen 65-0865-18). Cells were resuspended in HBSS (Gibco 24020-117) with 5% heat-inactivated fetal bovine serum (Gibco 10082-147) for sorting. Viable cells were gated using scatter and viability exclusion stain, sorted as either HIM3-4$^+$ P67.6$^+$, HIM3-4$^+$ P67.6$^-$, or HIM3-4$^-$ P67.6$^-$ and collected in complete media. At 48 hours post-sort, cells were split using limiting dilution into a 96 well plate at an average density of 0.5 cells/well and expanded until sufficient cell numbers for analysis were achieved, approximately 8 weeks. Clones were screened by flow cytometry again prior to PCR and sequence analysis.

3.2.4 PCR Screening and Cloning

Genomic DNA from CRISPR-edited U937 clones was isolated with a DNeasy kit (Qiagen 69506) per manufacturer instructions and amplified with Q5 High-Fidelity DNA Polymerase (New England BioLabs M0439L) using forward primer 5'-CACAGGAAGCCCTGGAAGCT and reverse primer 5'-GAG CAGGTCAGGTTTTTGGA (Invitrogen). Thermocycling parameters were as follows: 98°C 1 min; 98°C 15 s, 66°C 15 s, 72°C 45 s, 32 cycles; 72°C 2 min, 25°C hold. PCR products were separated on a 0.8% agarose-TBE gel, purified using a Monarch gel extraction kit (New England BioLabs T1020L), and sequenced by a commercial company (ACGT, Wheeling, IL). The three missense mutations identified were introduced into a previously described pcDNA3.1-CD33-V5/HIS vector using a QuikChange XL kit (Agilent 200517) with forward primer 5'-GCACTT GCAGCCGGATTTTTGGATCCATAGCCAGGGCC-3' and reverse primer 5'-GGCCCTGGCTATGGATCCAAAAATCCGGCTGCAAGTGC-3' (Invitrogen) to

47

generate the pcDNA3.1-KIRCD33-V5/HIS vector, transformed into TOP10 E. coli (Invitrogen C404003), isolated using a Plasmid Plus Midiprep Kit (Qiagen 12945) and verified by sequencing (ACGT) (Malik *et al.*, 2015a).

3.2.5   Gene expression by qPCR

Quantitative PCR (qPCR) was used to quantify expression of total CD33 and D2-CD33 as previously described (Malik *et al.*, 2013). Briefly, primers corresponding to sequences within exons 4 and 5 were used to quantify total *CD33* expression (forward, 5'-TGTTCCACAGAACCCAACAA-3'; reverse, 5'-GGCTGTAACACCAGCTCCTC-3'), as well as primers corresponding to sequences at the exon 1–3 junction and exon 3 to quantify the D2-CD33 isoform (forward, 5'-CCCTGCTGTGGGCAGACTTG-3'; reverse, 5'-GCACCGAGGAGTGAGTAG TCC-3'). PCR was conducted using an initial 2 min incubation at 95°, followed by cycles of 10 s at 95°C, 20 s at 60°C, and 20 s at 72°C. The 20 μl reactions contained 1 μM each primer, 1X PerfeCTa SYBR Green Super Mix (Quanta Biosciences), and 20 ng of cDNA. Experimental samples were amplified in parallel with serially diluted standards that were generated by PCR of cDNA using the indicated primers, followed by purification and quantitation by UV absorbance. Results from samples were compared relative to the standard curve to calculate copy number in each sample. Assays were performed in triplicate and normalized to expression of ribosomal protein L32 (*RPL32*) as the housekeeping gene.

3.2.6 HEK293 Transfection

HEK293 cells were seeded at approximately 70% confluency 24 hours before transfection. Cells were then transfected with Lipofectamine 3000 with Plus Reagent (Invitrogen L3000001) per manufacturer instructions, 250 ng plasmid per well in 8 well glass chamber slides (MatTek CCS-8) for immunofluorescence or 1000 ng per well in 12 well plates (Corning 3513) for flow cytometry. Cells were transfected with either the previously described wild-type CD33 vector (pcDNA3.1-CD33-V5/HIS), pcDNA3.1-KIRCD33-V5/HIS, or no vector control. Cells were incubated for 24 hours before analysis by flow cytometry or immunofluorescence and confocal microscopy.

3.2.7 Confocal Immunofluorescence Microscopy

Transfected HEK293 cells were fixed with 10% neutral buffered formalin (Fisher Scientific SF100-4) for 30 minutes then blocked and permeabilized for 30 minutes with 10% goat serum (Sigma S26-LITER), 0.1% Triton X-100 (Fisher Scientific BP151-500) in PBS (Fisher BioReagents BP665-1). Primary and secondary antibodies were diluted in the same blocking and permeabilization buffer and incubated at room temperature for 90 minutes. Cells were washed three times in blocking and permeabilization buffer between primary and secondary antibodies, and three times in PBS prior to coverslip mounting with Prolong Glass with NucBlue mounting media (Invitrogen P36981) and high-tolerance No. 1.5 coverglass (ThorLabs CG15KH1). Images were acquired using a Nikon A1R HD

inverted confocal microscope with a 60X oil objective and NIS Elements AR software.

3.2.8 Statistical analyses

Analyses were performed using GraphPad Prism 8.4.2. Gene expression data were analyzed by one-way ANOVA followed by Dunnett's multiple comparisons to control (WT cells).

3.3 Results

3.3.1 CRISPR-Cas9-mediated CD33 Exon 2 deletion leads to loss of P67.6 epitope

To generate an *in vitro* model of D2-CD33, we targeted exon 2 for deletion by using guide RNAs corresponding to sequences in the flanking introns as previously described (Humbert *et al.*, 2019). Cells were transfected, maintained for two weeks, and sorted according to Figure 3.1. Live cells were gated by light scatter (Figure 3.1A), singlet events identified (Figure 3.1B), and sorted into separate tubes based on CD33 phenotypes (Figure 3.1C) with unstained cells shown for reference (Figure 3.1D). CD33 immunophenotype was determined with antibodies P67.6 and HIM3-4, which target epitopes in IgV and IgC_2 that are encoded by exon 2 and exon 3, respectively. CD33 domains targeted by each antibody used in this study are shown in Table 3.1. We found that, of the 396,789 sorted cells, 91.7% fell outside of the unedited cell gate, and we presume these cells contain a CRISPR-mediated change in *CD33* sequence. Clonal cell lines were established from bulk collection of the gates drawn in Figure 3.1C. These cell lines were subsequently re-examined by flow cytometry with the same P67.6 and

HIM3-4 antibodies. While unedited cells showed robust labeling by both HIM3-4 and P67.6 (Figure 3.2A), edited cell lines showed strong labeling by HIM3-4 but not P67.6 (Figure 3.2B), or no labeling by either HIM3-4 or P67.6 (Figure 3.2C). Data are representative of three independently established cell lines for each phenotype. Since D2-CD33 protein is not readily apparent on the cell surface (Bhattacherjee *et al.*, 2021; M. O. Gbadamosi *et al.*, 2021; Godwin *et al.*, 2020; Humbert *et al.*, 2019; Malik *et al.*, 2015a; Siddiqui *et al.*, 2017), we expected that the latter cells (Figure 3.2C) were candidates for exon 2 excision, which was confirmed by a PCR product of the appropriate size (Figure 3.2D, right) and by sequencing. However, cell lines with robust cell surface HIM3-4 but not P67.6 labeling were unexpected. Screening by the size of the PCR amplicon with primers corresponding to exon 1 and exon 3 suggested that exon 2 was still present (Figure 3.2D, middle). Sequencing of this PCR fragment revealed that the HIM3-4[+] P67.6[-] clones contained three apparent single nucleotide polymorphisms (SNPs) in exon 2 compared to the unedited, wild-type (WT-CD33) U937 cell line. These SNPs have an identical minor allele frequency (MAF), 9.86×10^{-5} and are indexed as rs3987761, rs3987760, and rs35814802 (Karczewski *et al.*, 2020). Introduction of these SNPs results in changes in three consecutive amino acids (p.N20K, p.F21I, p.W22R) which we refer to as KIR-CD33. The nonsynonymous amino acids are the 4-6th amino-terminal residues of the mature protein. EMBOSS and PSORT II predict cell surface localization for KIR-CD33 and no change in the signal peptide cleavage site (Nakai & Horton, 1999; Rice *et al.*, 2000). Consistent with typical cell-surface localization, HIM3-4 labeled both CD33 and KIR-CD33 in a similar fashion

(Figure 3.2A-B). Notably, total *CD33* gene expression and exon 2 splicing in KIR-CD33 cells does not differ from that of unedited cells (Figure 3.2E-F). CD33 expression was increased in D2-CD33 cells (Figure 3.2E) which exclusively express the *D2-CD33* isoform (Figure 3.2F).

3.3.2 Validation of the CRISPR-induced in-frame disruption of the P67.6 epitope

To rigorously demonstrate that the lack of P67.6 labeling was solely due to the KIR mutations, we introduced the KIR mutations into a previously described CD33 expression vector (Malik *et al.*, 2015a). CD33 and KIR-CD33 vectors were transfected into HEK293 cells, which do not naturally express CD33, and the cells processed for flow cytometry (Figure 3.3) and immunofluorescent confocal microscopy (Figure 3.4). For flow cytometry, cells were labeled with P67.6 or WM53, each of which was conjugated to the same fluorochrome to facilitate direct comparisons (Figure 3.3). Importantly, these antibodies both target exon 2, are both mouse IgG1κ and have comparable degrees of labeling. In the CD33 HEK293 cells, labeling with both WM53 and P67.6 correlated well with HIM3-4 (Figure 3.3A-B). However, in the KIR-CD33 HEK293 cells, labeling with WM53 but not P67.6 correlated with HIM3-4 labeling as P67.6 labeling was not apparent (Figure 3.3C-D). Further gating on the HIM3-4$^+$ cells shows that WM53 labeling is not affected by the KIR-CD33 mutation (Figure 3.3E), while P67.6 labeling in KIR-CD33 cells is comparable to non-transfected control cells (Figure 3.3F). These results were confirmed with immunofluorescent confocal microscopy using an array of anti-CD33 antibodies (Figure 3.4). The *CD33*-transfected HEK293 cells showed consistent double-labeling between an antibody against a cytoplasmic epitope

(H110) and either WM53, P67.6, or PWS44 (Figure 3.4A-C). For the KIR-CD33 cells, robust co-labeling was observed between H110 and WM53 or PWS44 (Figure 3.4D, F). However, P67.6 labeling of KIR-CD33 cells was not detected (Figure 3.4E). We thus conclude that the residues identified—p.N20, p.F21, and p.W22—are necessary for P67.6 binding and that these changes are the reason that this antibody failed to label the CRISPR-edited cells. Although P67.6 has been humanized and used clinically, prior studies have not mapped its epitope at this resolution (Chauhan *et al.*, 2019; Mortland *et al.*, 2013; Perez-Oliva *et al.*, 2011).

3.3.3 Identification of SIGLEC22P as a homology-directed repair template for CD33

The three nucleotide changes (Figure 3.5, red) are 23-27 bp away from the putative Cas9 cleavage site (Figure 3.5, scissors), were not consecutive, and were present and identical in each of the clones with the KIR-CD33 phenotype. Since this was unlikely due to chance, we hypothesized that this was due to HDR from elsewhere in the genome. A search of a 50 bp sequence centered on the KIR mutations revealed that this sequence occurs in the *SIGLEC22P* pseudogene which is located 13.5 kb away from *CD33*. Further investigation found an extended region of homology between *CD33* and *SIGLEC22P* that flanked the Cas9 cleavage site; indeed, the first 500 bp of the genes share 97% identity, including a 143 bp region of otherwise complete identity centered on the KIR mutations (Figure 3.5, underline). The *SIGLEC22P* pseudogene exon 2 contains 11 additional mutations as well as two intronic mutations relative to CD33. None of these mutations were detected in the CRISPR-edited cell lines, indicating the region

used for repair was limited to, at most, the 143 bp region surrounding the KIR mutations near the Cas9-induced double-stranded break. We interpret these results as indicating that, upon double-strand breakage at the beginning of exon 2 in *CD33* (Figure 3.5, blue), the *SIGLEC22P* pseudogene was used as a repair template because of the strong sequence homology to *CD33* (Figure 3.5, underlined). In this process, three *SIGLEC22P*-specific mutations were introduced into *CD33*, resulting in missense mutations in three adjacent codons and thus KIR-CD33. This indicates an in-frame, ectopic gene conversion using pseudogene sequence. Two lines of evidence indicate that the KIR-CD33 cell lines were homozygous for the KIR mutation. First, the DNA sequence chromatogram showed only clear single peaks through the KIR sequence (Figure 3.5, chromatogram). Second, cell populations with an intermediate level of P67.6$^+$ labeling were not detected (Figure 3.1C). Further, sequencing of the *SIGLEC22P* region in the KIR-CD33 clones revealed that the *CD33* sequence was not present, which we interpret to mean that *SIGLEC22P* was used as a repair template, rather than a crossover event.

3.4   Discussion

We show here that the DSB repair pathways initiated after *S. pyogenes* Cas9 cleavage can lead to ectopic gene conversion from a pseudogene in a mitotic human cell line. This gene conversion resulted in an in-frame chimeric protein, wherein less than 150 bp of sequence from a nearby pseudogene replaced the targeted gene sequence (Figure 3.5). Indeed, this distance is consistent with meiotic gene conversion events observed by Jeffreys et al., who showed that gene

conversion occurs through relatively short tracts with a mean length between 55-290 bp (Jeffreys & May, 2004). Given that the *SIGLEC22P* locus in these KIR-CD33 cells lacks any detectable *CD33* sequence, we interpret this as further evidence of a gene conversion, rather than a mitotic crossover event. We speculate that this gene-conversion occurs in *trans*, i.e. from the intact chromosome, as a *cis*-mediated repair would more likely result in a gene fusion with intergenic deletion as has been previously reported (Borot *et al.*, 2019; Kim *et al.*, 2018). This type of deletion event also occurs naturally, in the absence of Cas9 DSBs, as in the case of *SIGLEC14* deletions (Yamanaka *et al.*, 2009). Gene conversion in *trans* after CRISPR-induced DSBs has been demonstrated previously (Susani *et al.*, 2018). We were surprised by the unexpectedly high frequency of conversion observed here; this pseudogene conversion occurred at approximately 1 in 10 edited cells. Exogenous double-stranded regions of homology as short as 58 bp have been used *in vitro* to introduce mutations through CRISPR-Cas9/HDR mechanisms (Renaud *et al.*, 2016). Coincidentally, this ectopic gene conversion disrupted the epitope of a well-validated antibody, P67.6 (Figure 3.6), known clinically as gemtuzumab. Using transiently transfected HEK293 cells expressing the chimeric protein, we demonstrated that these mutations are sufficient to abrogate P67.6 binding, providing the most precise epitope mapping to date of this clinically relevant antibody.

The KIR-CD33 mutations—rs3987761, rs3987760, rs35814802—are indexed in dbSNP and gnomAD with MAF $< 10^{-4}$, and roughly equivalent across populations (Karczewski *et al.*, 2020). We considered the possibility that these

mutations, while rare, occur naturally and are clinically relevant. The SIGLEC family of genes are undergoing rapid evolution in many species, including humans (Padler-Karavani *et al.*, 2014). This rapid evolution has resulted in the pseudogenization of many SIGLECs, including the *CD33* pseudogene *SIGLEC22P*. These same bases are indexed as SNPs in *SIGLEC22P* (rs997169007, rs1049597792, and rs1005338799), also have identical frequencies (MAF = 2.1 x 10^{-4}) in the gnomAD database, and the major and minor alleles are the inverse of the KIR-CD33 SNPs (Karczewski *et al.*, 2020). The region of homology identified here between *CD33* and *SIGLEC22P* is 143 bp, substantially shorter than the 50 bp reads upon which the gnomAD and dbSNP databases are built (Lonsdale *et al.*, 2013). While it is possible these are *bona fide* variants in *CD33*, the most likely explanation is that the *SIGLEC22* pseudogene sequences have been mapped incorrectly to *CD33* by algorithms. These missense SNPs are also not recorded in the BeatAML variant database, which records known AML-associated functional variants such as missense SNPs, further underscoring the low probability that these are true *CD33* variants (Tyner *et al.*, 2018).

Off-target editing is a frequent concern in gene editing workflows, including CRISPR-Cas9. However, gene conversion is often overlooked as a potential confound. For instance, within *CD33*, off-target editing of *SIGLEC22P* resulting in a 14 kb deletion and subsequent *SIGLEC22P-CD33* gene fusion has been previously reported, which may have been the result of a mitotic crossover (Borot *et al.*, 2019; Kim *et al.*, 2018). Neither report noted a gene conversion between

SIGLEC22P and *CD33*. Mitotic crossover initiated by CRISPR-Cas9 cleavage has also been reported in the *HPRT* locus, resulting in a 36 kb crossover (Susani *et al.*, 2018). Gene conversion via CRISPR-Cas9 between a 101 bp homologous region in *HBD* and *HBB* has also been reported; notably, these genes were also the first reported gene conversion event in humans (Javidi-Parsijani *et al.*, 2020; Slightom *et al.*, 1980). To our knowledge, this is the first report of pseudogene-mediated gene conversion during the CRISPR-Cas9 editing process. Our results demonstrate the need for rigorous screening in studies which rely on gene editing, and analysis of the region flanking the expected cut site for homology and possible gene conversion. Since the human genome contains 8000-12000 pseudogenes and approximately 3400 genes in the human genome have known pseudogenes, pseudogene directed homology repair is a potentially considerable confound (Pei *et al.*, 2012; Z. Zhang *et al.*, 2006). Approximately 84% of putative pseudogenes are estimated to be located on a different chromosome than their parent gene, while the remaining 16% have a mean intergenic distance of 1.8 Mb (median of 1.3 Mb) (Sisu *et al.*, 2014). Whether the gene conversion described here occurs in *cis* or *trans*, or if this event is impacted by intergenic distance, is unclear. As pseudogenes are by definition nonfunctional, they may be overlooked as irrelevant during the design process. Given the pseudogene conversion described here, there is a clear need to analyze the sequence of the edit, rather than relying on gene or protein expression. This is especially relevant given that CRISPR-mediated *CD33* knockout allografts and autografts have been proposed as a potential AML treatment strategy in conjunction with chimeric antigen receptor

(CAR) T cells (Borot *et al.*, 2019; Kim *et al.*, 2018). In this strategy, autologous transplantation of *CD33*^{null} hematopoietic stem and progenitor cells (HSPCs) would reconstitute the myeloid system, while engraftment of CAR-T cells would provide long-term surveillance against any surviving CD33$^+$ AML cells. Mitotic gene conversion, and presumably the herein described pseudogene-mediated gene conversion, has been demonstrated after CRISPR-Cas9 editing in primary human cells between *HBD* and *HBB* (Javidi-Parsijani *et al.*, 2020). Screening these HSPCs with the standard gemtuzumab antibody, however, could lead to engraftment of some KIR-CD33 cells as well.

While off-target editing of a nonfunctional pseudogene may have limited impact on downstream results, incorporating elements from a nonfunctional pseudogene into a target gene may lead to deleterious mutations that abrogate function of the target gene entirely, for instance the introduction of a premature stop codon. This is especially important to consider for designs which incorporate unedited cells which have undergone sorting and single cell cloning as a control. We also speculate that this pseudogene-mediated repair will reduce the efficiency of creating a gene disruption if the disruption site has homology in a pseudogene. The breakage site may be repaired by a pseudogene with complete identity at the DSB, requiring more clones to be screened to find a gene disruption. By coincidence, our initial screen for editing included an antibody which overlapped the KIR sequence. This gene conversion, at the protein level, is masked when using an alternative antibody (WM53) targeting the same domain and is not apparent by PCR alone. Combining the data from Figures 3.2D-F and 3.3C, one

could reason that these cells were unedited as there are no apparent differences in PCR fragment size, gene expression, splicing, or cell surface protein expression, and thus incorrectly assume that sequencing is unnecessary. We conclude that, in addition to off-target Cas9-editing confounds, researchers should be aware of the potential for pseudogene-directed homology repair.

Table 3.1: Antibodies used in Chapter 3.

Clone	Conjugate	Vendor	Catalog #	Lot	Dilution/Use	CD33 domain
P67.6	BV711	BioLegend	366624	B302694	2.5 µg/mL FC	IgV
P67.6	Unconjugated	BioLegend	825601	B258818	5 µg/mL IF	IgV
WM53	BV711	BioLegend	303424	B253729	5 µg/mL FC	IgV
WM53	Unconjugated	BioLegend	96281	B274701	5 µg/mL IF	IgV
HIM3-4	FITC	BioLegend	303304	B284834	20 µg/mL FC	IgC_2
PWS44	Unconjugated	Leica	NCL-L-CD33	6024275	2 µg/mL IF	IgC_2
H110	Unconjugated	Santa Cruz	sc-28811	K2211	1 µg/mL IF	Intracellular
Goat anti-Mouse Polyclonal	AlexaFluorPlus 488	Invitrogen	A32723	TF266577	10 µg/mL IF	
Goat anti-Rabbit $F(ab)_2$ Polyclonal	AlexaFluor 568	Invitrogen	A21069	2087701	10 µg/mL IF	

IF: immunofluorescence, FC: flow cytometry
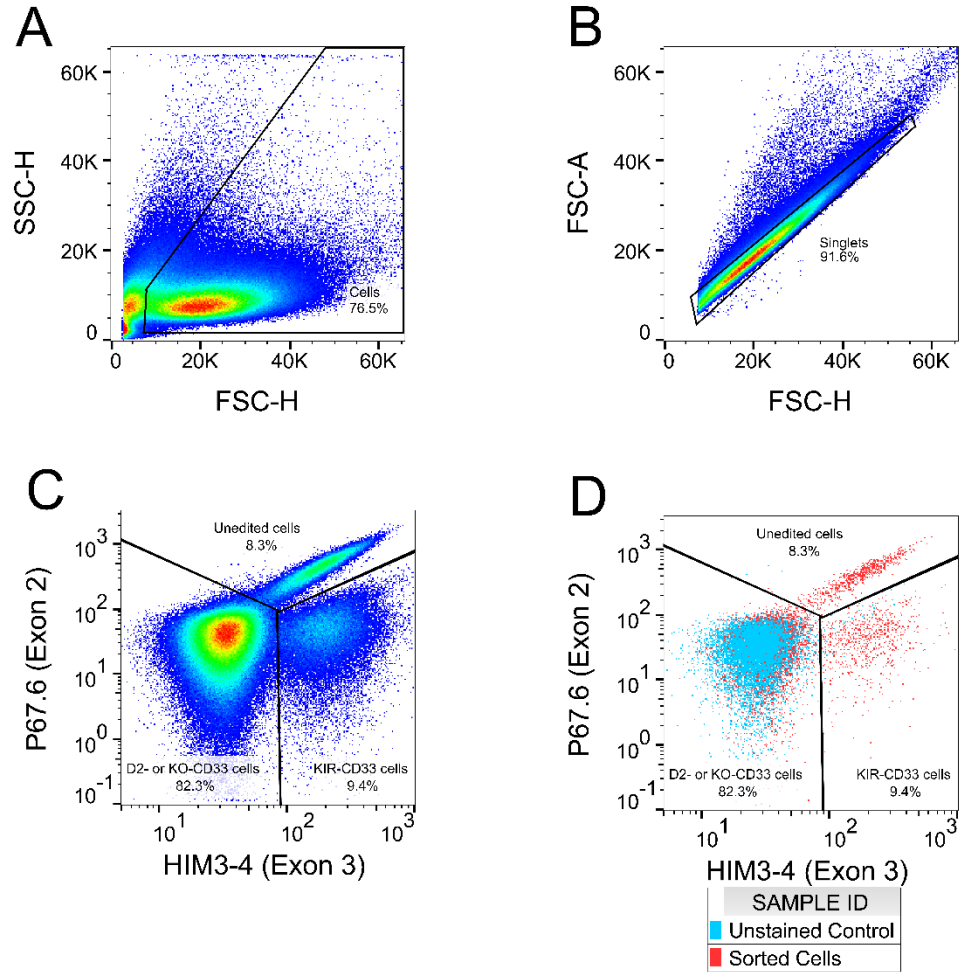
Figure 3.1: Sorting CRISPR-Cas9 CD33 exon 2-edited cells reveals to two cell surface phenotypes

Gates are labeled in each panel and correspond to the following panel, with percent of parent gate shown. (A) Forward scatter (FSC) by side scatter (SSC) gating was used to gate out dead cells (FSC-low SSC-high). (B) Singlet events were selected along the FSC-Area by FSC-Height diagonal. (C) Populations were identified as either unedited (HIM3-4+ P67.6+), potential D2-CD33 or KO-CD33 (HIM3-4- P67.6-), or KIR-CD33 (HIM3-4+ P67.6-). (D) Unstained control (blue)

61

shown on top of the sorted cells (red) for reference. This unexpected KIR-CD33 population occurred at approximately a 1-to-9 frequency with respect to the D2- or KO-CD33 cells, or 10% of all edited cells.
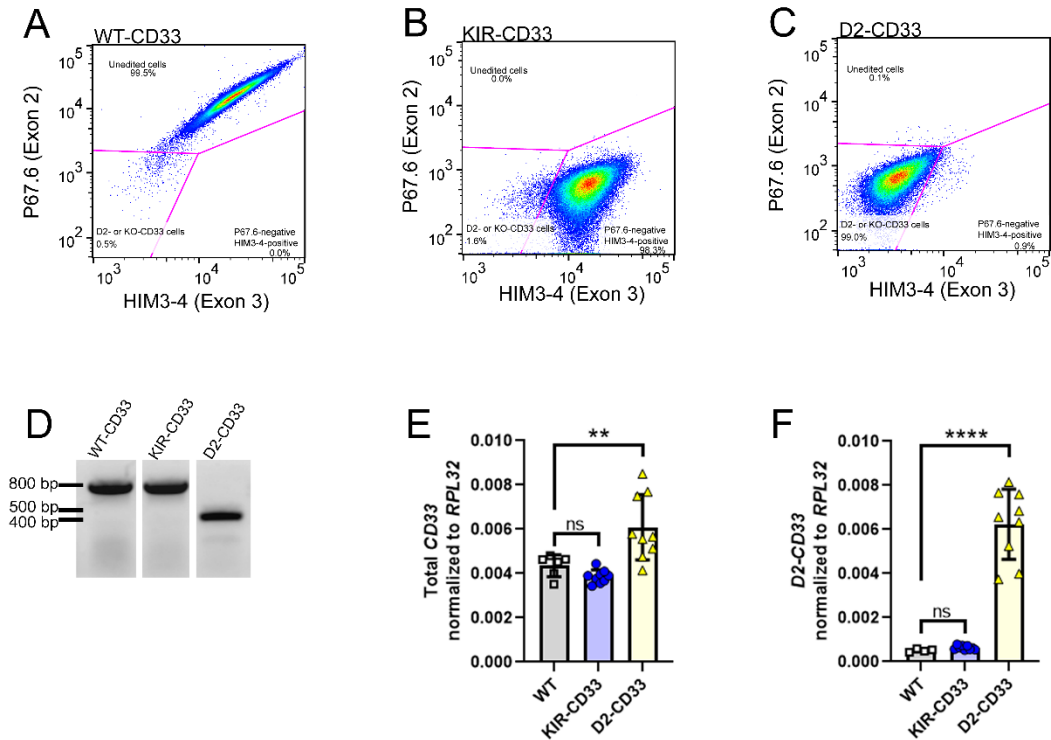
Figure 3.2. CRISPR-Cas9 editing of CD33 exon 2 leads to loss of P67.6 epitope

(A) Unedited U937 cells display robust P67.6 and HIM3-4 labeling. (B) Edited U937 clone that is robustly labeled by HIM3-4 but not P67.6. (C) Edited U937 clone that is not labeled by either HIM3-4 or P67.6. The depicted results in B-C are representative of at least three clonal cell populations established for each phenotype. (D) Genomic DNA PCR of CD33 exon 1 to exon 3 of the above cell lines. PCR products at 789 bp and 428 bp correspond to the expected sizes for the presence and absence of exon 2, respectively. (E) KIR-CD33 mutations do not affect total CD33 gene expression as determined by qPCR, but removal of exon 2 increases total CD33 gene expression by 39.8%. (F) KIR-CD33 mutations do not affect splicing efficiency of exon 2 as determined by qPCR, but removal of exon 2 at the genomic level increases exon 1-exon 3 junction to 100% of total CD33

expression. Data from E-F analyzed by one-way ANOVA followed by Dunnett's multiple comparisons test to unedited control. ** p < 0.01; **** p < 0.0001.
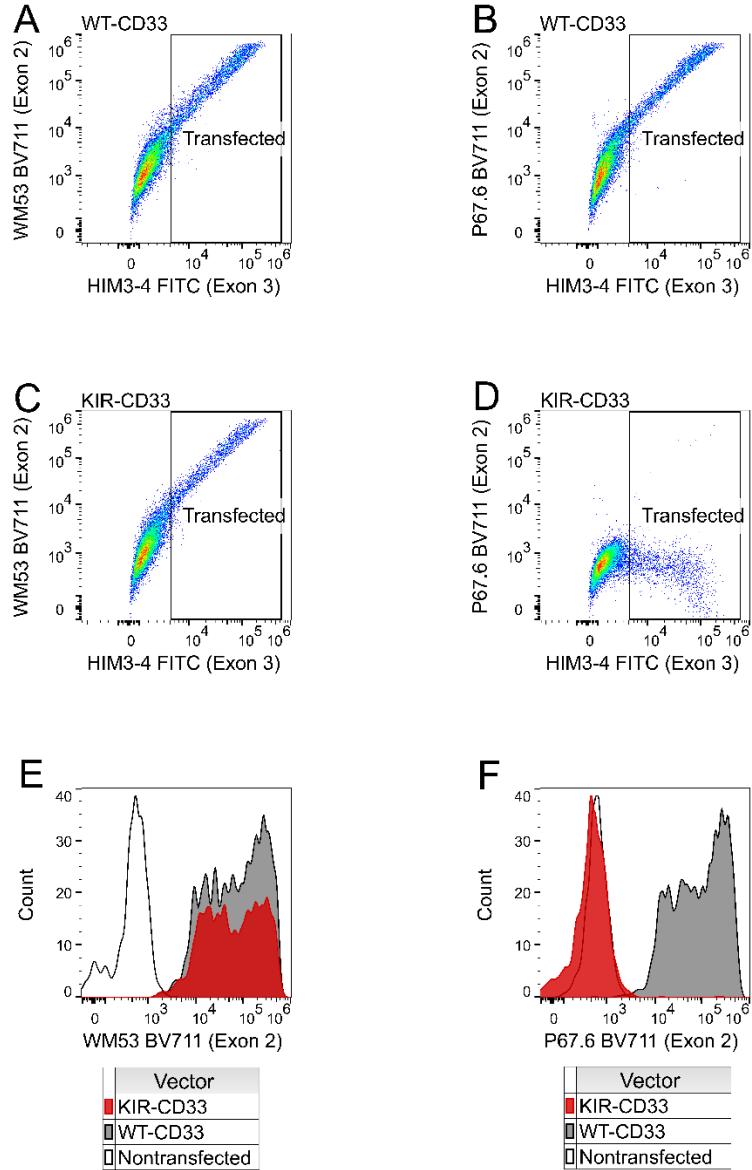
Figure 3.3. Loss of P67.6 epitope is in-frame and preserves cell surface expression

WT-CD33 (A, B) or KIR-CD33 (C, D) expressing HEK293 cells were labeled with HIM3-4 and either WM53 (A, C) or P67.6 (B, D). HIM3-4+ cells identified in the "Transfected" gate (A-D) were gated to show WM53 (E) or P67.6 (F) binding in transfected cells.
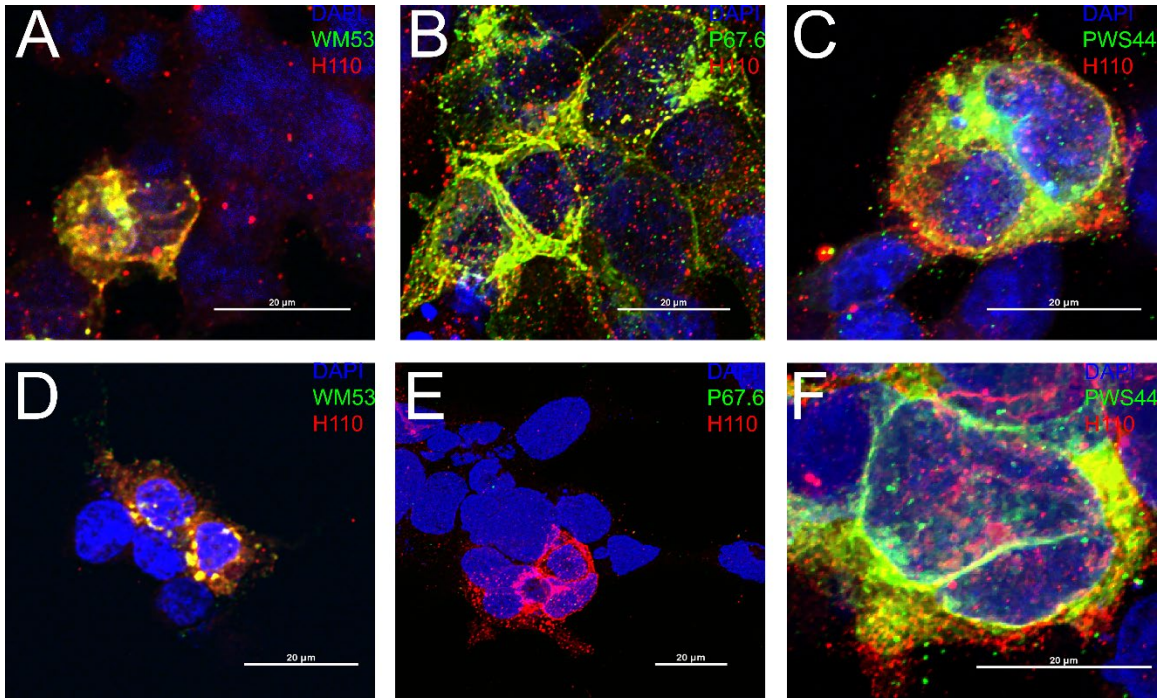
Figure 3.4. Loss of P67.6 epitope does not alter other common CD33 epitopes

WT-CD33 (A-C) or KIR-CD33 (D-F) HEK293 cells were labeled with H110 (red) and either WM53 (A, D; green), P67.6 (B, E; green), or PWS44 (C, F; green) and DAPI (blue).

```
WT-CD33              TGAGTGGCTG TGGGGAGAGG GGTTGTCGGG CTGGGCCGAG CTGACCCTCG TTTCCCCACA
KIR-CD33             .......... .......... .......... .......... .......... ..........
D2-CD33              .......... .......... .......... .......... .......... ..........
SIGLEC22P pseudogene ...A...... C......... .......... .......... .......... ..........


WT-CD33              GGGGCCCTGG CTATGGATCC AAATTTCTGG CTGCAAGTGC AGGAGTCAGT GACGGTACAG
KIR-CD33             .......... .......... ...AA..C.. .......... .......... ..........
D2-CD33              .--------- ---------- ---------- ---------- ---------- ----------
SIGLEC22P pseudogene .......... .......... ...AA..C.. .......... .......... ..........


KIR-CD33 sequencing


WT-CD33              GAGGGTTTGT GCGTCCTCGT GCCCTGCACT TTCTTCCATC CCATACCCTA CTACGACAAG
KIR-CD33             .......... .......... .......... .......... .......... ..........
D2-CD33              ---------- ---------- ---------- ---------- ---------- ----------
SIGLEC22P pseudogene .......... .......... .......... ....C..... .......... ....A...G.
```
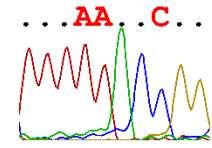
Figure 3.5. Alignment of cell line sequencing data

Figure 3.5. Alignment of cell line sequencing data

Unedited U937 cells at top, followed by KIR-CD33 sequence, D2-CD33 sequence, and reference *SIGLEC22P* sequence. Cas9 cleavage site marked by scissor icons. Mismatches from the unedited *CD33* sequence are denoted by either the differing base (KIR-CD33) or dash in the case of a gap (D2-CD33). Intron 1 in black, exon 2 in blue. The *SIGLEC22P* region used as a repair template is underlined. Mutations introduced into *CD33* in red. A representative post-PCR sequencing chromatogram from one clonal KIR-CD33 is shown at the mutation site with clear, single peaks demonstrating homozygosity.
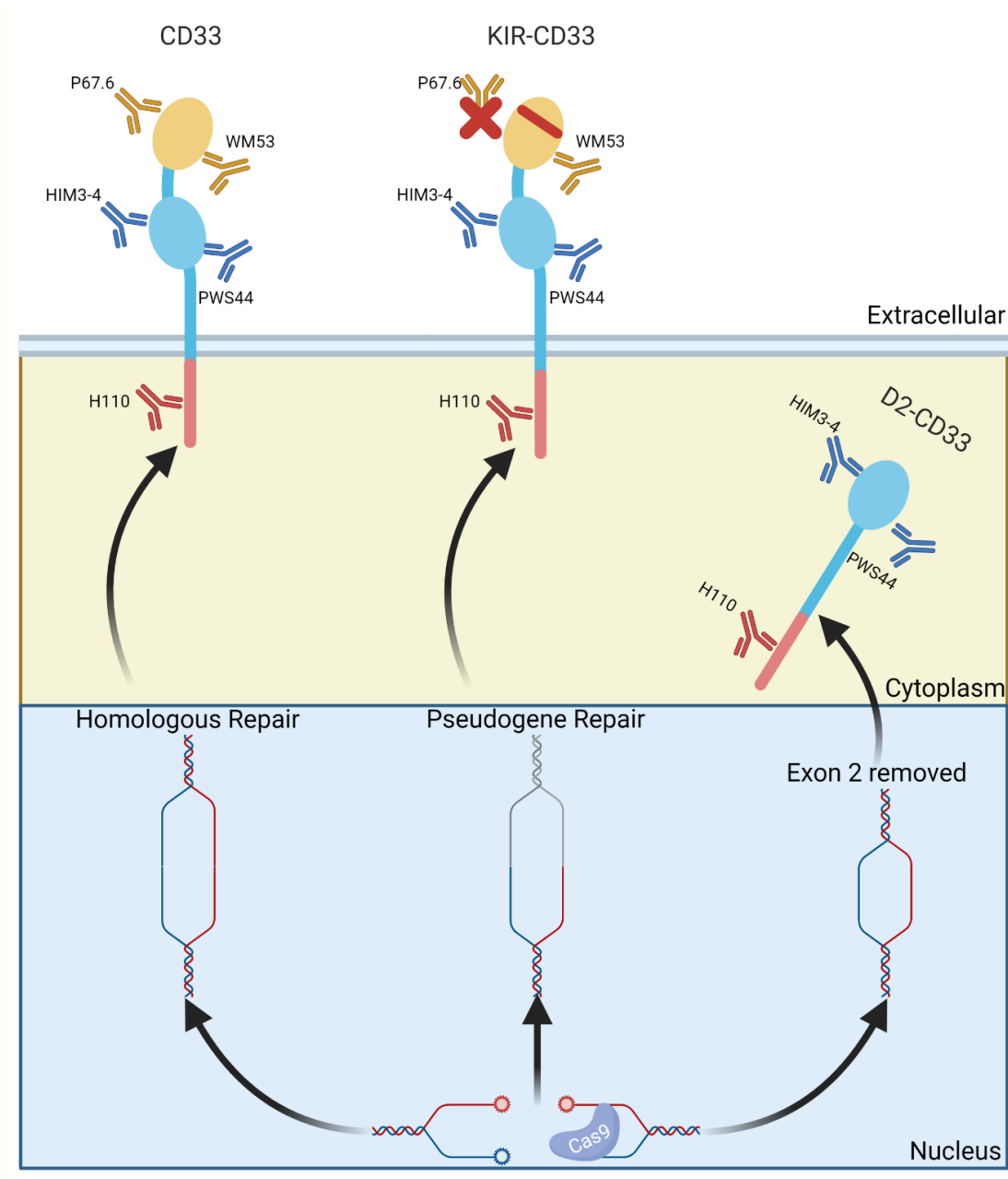
Figure 3.6. Model of pseudogene repair mechanism and anti-CD33 antibody binding sites

CD33 is normally a transmembrane, cell surface receptor with one IgV domain and one IgC$_2$ domain. The KIR-CD33 mutation introduced by pseudogene

directed-repair abrogates P67.6, but not WM53, binding. IgC$_2$ domain antibodies HIM3-4 and PWS44, and the intracellular domain H110 antibody bind both CD33 and KIR-CD33. D2-CD33 is not readily apparent on the cell surface in CRISPR-Cas9 edited U937 cells, implying that under physiologic expression, the D2-CD33 protein is retained in an intracellular vesicle. Overall, pseudogene repair of a CRISPR-Cas9-targeted gene can disrupt the binding of well-validated antibodies without introducing a frameshift, and protein-level expression alone is not sufficient for knockout confirmation. Created with Biorender.

CHAPTER 4. ANALYSIS OF GENETIC VARIANTS ASSOCIATED WITH LEVELS OF IMMUNE MODULATING PROTEINS FOR IMPACT ON ALZHEIMER'S DISEASE RISK REVEAL A POTENTIAL ROLE FOR *SIGLEC14*

[This section contains material adapted from a published manuscript: Shaw, B. C., Katsumata, Y., Simpson, J. F., Fardo, D. W., & Estus, S. (2021). Analysis of Genetic Variants Associated with Levels of Immune Modulating Proteins for Impact on Alzheimer's Disease Risk Reveal a Potential Role for SIGLEC14. *Genes (Basel)*, *12*(7).]

4.1 Introduction

Genome-wide association studies (GWAS) have identified a set of polymorphisms that modulate the risk of Alzheimer's disease (AD) (Hollingworth *et al.*, 2011; Jansen *et al.*, 2019; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Naj *et al.*, 2011; Novikova *et al.*, 2021). The pathways implicated in this process include innate immunity, cholesterol homeostasis, and protein trafficking (Jones *et al.*, 2010; Karch & Goate, 2015; Malik *et al.*, 2015b). Four of these genes, *TREM2*, *CD33*, *PILRA*, and *FCER1G*, are members of the family of non-catalytic tyrosine-phosphorylated receptors (NTRs), which function through immunomodulatory tyrosine-phosphorylated activating motifs (ITAMs) or inhibitory motifs (ITIMs). The underlying immunomodulatory pathway is further implicated by AD-associated variants in phospholipase C (*PLCG2*) and *INPP5D* which encode proteins acting downstream of these ITAM- and ITIM-containing proteins. Functional studies have informed the current hypothesis that the variants associated with AD in the ITAM/ITIM family modulate inflammation and phagocytosis (Bhattacherjee *et al.*, 2021; Bhattacherjee *et al.*, 2019; Chan *et al.*, 2015; Ana Griciuc *et al.*, 2013; Malik

et al., 2015a; Malik *et al.*, 2013; McQuade *et al.*, 2020; Raj *et al.*, 2014; Siddiqui *et al.*, 2017).

The ITAM family, including TREM2, recruit kinases such as spleen tyrosine kinase (Syk) and phosphoinositide 3-kinase (PI3K) to induce downstream signaling, while the ITIM family, including CD33, recruit phosphatases such as SHP-1 to dephosphorylate Syk and ITAMs, thereby counteracting ITAM activity (Estus *et al.*, 2019). These ITAM and ITIM proteins are predominantly expressed in immune cells such as microglia. Overall, these and other studies have shown that microglia contribute to AD pathogenesis, a concept that has been reviewed recently (Efthymiou & Goate, 2017; Gandy & Heppner, 2013; A. Griciuc & Tanzi, 2021).

The critical barrier to progress in translating GWAS candidate genes to treatments is elucidating the actions of the functional variant at the molecular level, i.e., splicing (sQTL), gene expression (eQTL), or protein level (pQTL), to understand whether the pathway affected is detrimental or beneficial to disease risk. GWAS single nucleotide polymorphisms (SNPs) in AD are frequently identified as eQTLs in the brain (Allen *et al.*, 2012). Sun et al. have used GWAS to identify pQTLs for the plasma proteome, including ITIM and ITAM-containing proteins (Sun *et al.*, 2018). To investigate the hypothesis that these pQTLs may uncover additional AD-related genes that may have been overlooked in AD GWAS because of their stringent false-discovery rate controls, we examined the Sun et al. *cis*-pQTL data together with the Jansen et al. AD GWAS results. Parsing the proteins from the genome-wide significant *cis*-pQTL dataset by whether or not an

ITIM/ITAM domain was present, and then examining whether the associated SNP is nominally significant (p < 0.05) for AD association, found a significant overrepresentation of ITIM/ITAM encoding genes with nominal AD associations. Since one of these genes, *SIGLEC14*, has been reported to be deleted in some individuals, we investigated further and found that the pQTL and AD SNP, rs1106476, is a proxy for the previously identified deletion polymorphism (Yamanaka *et al.,* 2009). We defined this deletion further by identifying additional *SIGLEC14* copy number variants and by determining the effect of *SIGLEC14* copy number on the expression of *SIGLEC14* and the neighboring *SIGLEC5*. We conclude that variants in ITIM/ITAM family members, including *SIGLEC14*, represent underappreciated potential genetic risk factors for AD.

4.2   Materials and Methods

4.2.1   Preparation of gDNA, RNA, and cDNA from Human Tissue

Human blood and anterior cingulate autopsy tissue from 61 donors were generously provided by the Sanders-Brown Alzheimer's disease center neuropathology core and have been described elsewhere (Zou *et al.*, 2007). The matched brain and blood samples were from deceased individuals with an average age at death of 82.4 ± 8.7 (mean ± SD) years for non-AD and 81.7 ± 6.2 years for AD subjects. The average postmortem interval (PMI) for non-AD and AD subjects was 2.8 ± 0.8 and 3.4 ± 0.6 h, respectively. Non-AD and AD samples were comprised of 48% and 55% female subjects. MMSE scores were, on average, 28.4 ± 1.6 for non-AD subjects and 11.9 ± 8.0 for AD subjects. These samples were

used for genotyping and gene expression studies. Three additional blood samples matched to whole-genome sequencing (WGS) data were obtained to confirm WGS observations of additional *SIGLEC14* copies. DNA from these patients was prepared using a QIA-amp DNA Blood Mini kit (Qiagen, Germantown, MD, USA) per the manufacturer's instructions.

4.2.2 Genotyping and Copy Number Variant Assays

Copy number variation in *SIGLEC14* was determined using a TaqMan-based copy number variant (CNV) assay (Invitrogen, Waltham, MA, USA; Catalog number 4400291, Assay number Hs03319513_cn) compared to RNAse P (Invitrogen, 4403326). Amplification and quantitation were performed per manufacturer instructions. Genotyping the rs1106476 SNP was performed with a custom TaqMan assay (Invitrogen). This assay discriminates rs1106476 and rs872629, which are in perfect LD. As coinherited SNPs, this variant is also known as rs35495434.

4.2.3 Gene Expression by qPCR

Gene expression was quantified by qPCR with PerfeCTa SYBR Green master mix as previously described (Malik *et al.*, 2015a). *SIGLEC14* was quantified with primers corresponding to a sequence in exons 3 and 5: 5′—CAGGTGAAACGCCAAGGAG—3′ and 5′—GCGAGGAACAGGGACTGG—3′. *SIGLEC5* was quantified with primers corresponding to sequences in exons 4 and 5: 5′—ACCATCTTCAGGAACGGCAT—3′ and 5′—GGGAGCATCACAGAGCAGC—3′. Cycling conditions for all qPCRs were as

follows: 95 °C, 2 min; 95 °C, 15 s, 60 °C, 15 s, 72 °C, 30 s, 40 cycles. Copy numbers present in the cDNA were determined relative to standard curves that were executed in parallel (Estus *et al.*, 2019).

4.2.4   WGS Data Analysis

To investigate the frequency and range of *SIGLEC14* CNV, we performed a read-depth analysis for WGS data. We obtained compressed sequence alignment map (CRAM) files from the AD sequencing project (ADSP) and AD Neuroimaging (ADNI). We extracted paired-end reads mapped to the *SIGLEC14-SIGLEC5* locus under Genome Reference Consortium Human Build 38 (GRCh38/hg38), and then computed the depth at each position using the samtools depth function (Li *et al.*, 2009).

4.2.5   Statistical Analyses

The association of *cis*-pQTL proteins containing ITIM/ITAM domains and AD-associated SNPs was calculated using a simple chi-square test. Gene expression was analyzed by using JMP14 Pro using one-way analysis of variance (ANOVA) followed by Tukey's post-hoc multiple testing correction and graphed in GraphPad Prism 8.

4.3   Results
4.3.1   ITIM/ITAM pQTLs Are Overrepresented in AD GWAS Results

To evaluate whether pQTLs for ITIM or ITAM-containing proteins were associated with AD, we compiled a list of ITIM and ITAM-containing proteins from

prior reviews (Barrow & Trowsdale, 2006; Dushek *et al.*, 2012; Isakov, 1997; Ravetch, 2000). The resulting list contained 187 genes and is provided in the Appendix. The *cis*-acting pQTLs from Sun et al. and AD associations from Jansen et al. were then matched by chromosomal coordinates (Jansen *et al.*, 2019; Sun *et al.*, 2018). Both datasets were provided under Genome Reference Consortium Human Build 37 (GRCh37/hg19). Genes were then subset as either coding for an ITIM/ITAM gene or not and nominally significant ($p < 0.05$) for AD association or not. The SNPs which are associated with both ITIM/ITAM protein levels in plasma and AD risk are shown in Table 4.1. We found that pQTLs that affect ITIM or ITAM genes were significantly overrepresented in nominally significant AD associations ($p = 6.51 \times 10^{-5}$, $\chi^2 = 15.95$, Table 4.2).

4.3.2 SIGLEC14 pQTL Is a Proxy for the Deletion Polymorphism

Previous reports have identified a *SIGLEC14* deletion (Yamanaka *et al.*, 2009). Given the strong pQTL signal from rs1106476 on *SIGLEC14* reported by Sun et al., and the fact that rs1106476 is within the neighboring *SIGLEC5* gene, yet has a *cis*-pQTL effect on *SIGLEC14*, we hypothesized that rs1106476 is a proxy for the *SIGLEC14* deletion polymorphism (Sun *et al.*, 2018). To test this hypothesis, we genotyped a set of DNA samples for rs1106476 and quantified genomic copy number variation (CNV). We found that the proxy SNP correlates with *SIGLEC14* deletion well but not perfectly ($p < 0.0001$, $\chi^2 = 38.40$) (Table 4.3). To better understand this deletion, we then sequenced the region containing the *SIGLEC14-SIGLEC5* fusion in five minor allele carriers (two homozygous for SIGLEC14 deletion and three heterozygous) (Yamanaka *et al.*, 2009). Based on

these sequencing data, relative to reference sequences, we found a 692 bp region of complete identity between *SIGLEC14* and *SIGLEC5*. Within this region, the deletion polymorphism sequence corresponds to *SIGLEC14* at the 5′ end, but *SIGLEC5* on the 3′ end, with respect to reference sequence data (Figure 4.1). Overall, this represents a 17 kb deletion.

4.3.3 SIGLEC14 CNV Is Not Fully Captured by rs1106476

As noted in Table 3, we found some individuals that had three copies of SIGLEC14 as detected by the CNV assay. To validate these findings, we leveraged the ADNI and ADSP WGS datasets and compared read depth in the SIGLEC14 locus with surrounding sequences (Figure 4.2). Both datasets contained individuals with *SIGLEC14* copy numbers ranging from 0–3. The presence of three copies of *SIGLEC14* was cross-validated between WGS data and CNV assay in three individuals. Further, the frequencies across populations are equivalent (Table 4.4; $p = 6.76 \times 10^{-12}$, $\chi^2 = 69.30$). Read depths for Caucasian, African American, and other populations are shown as Figure 4.3, Figure 4.4, and Figure 4.5, respectively.

4.3.4 SIGLEC14 Is Expressed in Human Brain, and CNV Correlates with Gene Expression

To test whether gene expression compensation may neutralize the effect of genomic *SIGLEC14* deletion, we quantified *SIGLEC14* expression relative to *SIGLEC14* gene copy number in cDNA prepared from human brain samples. Consistent with RNAseq studies that show *SIGLEC14* is expressed in microglia, *SIGLEC14* expression strongly correlated with expression of the microglial gene

AIF1 (p < 0.0001, r^2 = 0.409, Figure 4.6A) (Estus *et al.*, 2019; Y. Zhang *et al.*, 2016). When *SIGLEC14* expression is normalized to *AIF1* expression, *SIGLEC14* expression was dependent in a step-wise manner with *SIGLEC14* CNV (p = 0.0002, $F_{2,47}$ = 10.679, Figure 4.6B). Strikingly, individuals with one copy of *SIGLEC14* have a mean *SIGLEC14* expression of 54.6% compared to individuals with two copies. We interpret this to mean that there is no compensatory increase in *SIGLEC14* expression in individuals heterozygous for *SIGLEC14* deletion.

4.3.5   *SIGLEC14* Deletion Leads to Increased *SIGLEC5* Expression

To test whether *SIGLEC5* expression changed with respect to *SIGLEC14* deletion, we quantified *SIGLEC5* expression relative to *SIGLEC14* CNV in these same brain samples. Since *SIGLEC5* does not have its own promoter and there are no H3K27 acetylation peaks between *SIGLEC14* and *SIGLEC5*, we hypothesized that an inverse relationship exists between *SIGLEC14* CNV and *SIGLEC5* expression, where a *SIGLEC14* deletion brings *SIGLEC5* closer to the promoter leading to increased transcription (Figure 4.7) (Fishilevich *et al.*, 2017; Kent *et al.*, 2002; The ENCODE Project Consortium, 2012). We found that *SIGLEC5* expression significantly increases with respect to *SIGLEC14* genomic deletions (Figure 4.8; p = 0.0220, $F_{2,46}$ = 4.151).

4.4   Discussion

The primary finding of this paper is that pQTLs for ITIM and ITAM-containing proteins are overrepresented as being nominally significant for AD risk, suggesting that the ITIM and ITAM family of proteins may contribute to AD pathogenesis. This

adds to the current body of work which supports the hypothesis that AD is mediated, at least in part, by immune cell dysfunction (Hollingworth *et al.*, 2011; Lambert *et al.*, 2013; Mawuenyega *et al.*, 2010; Naj *et al.*, 2011). Indeed, transcriptomics and genomics studies have frequently identified genes predominantly expressed in microglia within the CNS as associated with AD risk (Holtman *et al.*, 2015; Miller *et al.*, 2013; Orre *et al.*, 2014; Wes *et al.*, 2014; B. Zhang *et al.*, 2013). Within a pQTL study, variants that affect the expression of the ITIM/ITAM family of genes—which govern immune cell activation state—are more commonly associated with AD risk than variants for genes, not in this family (Table 4.2). Although we hypothesized that variants that enhanced ITAM levels or decreased ITIM levels would be associated with reduced AD risk, this was not observed. This likely indicates that while some of these pQTLs may reflect increased functional signaling, others may involve alterations in splicing to generate soluble isoforms or may increase susceptibility to cleavage from the cell surface. Hence, an SNP that associates with increased plasma protein levels does not necessarily correlate with increased cell surface expression and signaling.

SIGLEC14 was selected for further investigation based on its previously reported deletion polymorphism and close relationship to another AD-associated gene, CD33 (Jansen *et al.*, 2019; Yamanaka *et al.*, 2009). Since SNPs have previously been recognized as proxies for deletion of other genes (Abdollahi *et al.*, 2008; Hinds *et al.*, 2006; McCarroll *et al.*, 2008), and *SIGLEC14* deletion has been previously reported (Yamanaka *et al.*, 2009), we hypothesized that the strong pQTL signal from rs1106476 reported in Sun et al. (Sun *et al.*, 2018) correlated

with *SIGLEC14* deletion. Indeed, we found that rs1106476 is a proxy for *SIGLEC14* deletion and the minor allele count corresponds to the number of *SIGLEC14* deletions in 89% of cases in our dataset (Table 4.3).

This proxy variant does not, however, predict copy numbers greater than two. For instance, we observed four individuals with three copies of *SIGLEC14*; two of these individuals were homozygous minor for rs1106476 and two were heterozygous for rs1106476 (Table 4.3). Additional copy number variation is also present in the ADSP and ADNI sequencing projects (Figure 4.2). These CNVs are equivalent across populations in these datasets (Table 4.4, Figure 4.3, Figure 4.4, and Figure 4.5). Based on these data and the recombination peak which spans from upstream of *SIGLEC14* through exon 8 of *SIGLEC5* (Figure 4.9), we hypothesize that the additional copies integrate from a deletion event, though far less frequently than the deletion itself (Machiela & Chanock, 2015). Across the 3095 individual WGS dataset in ADSP, we found *SIGLEC14* deletion has a minor allele frequency (MAF) of 0.2023, while insertion occurs at a MAF of only 0.0195, suggesting a 10-times lower rate of integration than deletion (Table 4.4).

In the brain, *SIGLEC14* is predominantly expressed in microglia, in keeping with its putative role as an immune receptor (Figure 4.6A). The *SIGLEC14* deletion polymorphism also strongly correlates with *SIGLEC14* gene expression (Figure 4.6B). Due to the low frequency of the additional copy integration, we do not have sufficient samples with which to correlate *SIGLEC14* expression to additional copy numbers, nor can we conclude whether additional *SIGLEC14* genomic copies are transcribed in frame and subsequently produce protein.

We also find that *SIGLEC14* deletion increases the expression of *SIGLEC5* (Figure 4.8). For individuals with at least one copy of *SIGLEC14*, the expression of *SIGLEC14* is substantially higher than *SIGLEC5*. Coupled with the lack of an independent promoter or H3K27 acetylation peaks between the two genes in GeneHancer or Encode, respectively, we infer that expression of both genes is governed by a common promoter proximal to *SIGLEC14*, that the integrity of this promoter is preserved after *SIGLEC14* deletion, and that *SIGLEC14* deletion results in an increase in *SIGLEC5* expression due to its closer proximity to this common element. The *SIGLEC* family of receptors bind sialic acids as ligands to initiate their signaling cascades, and sialylated proteins, as well as gangliosides, are abundant in amyloid plaques (Ariga *et al.*, 2008; Salminen & Kaarniranta, 2009; Yanagisawa, 2007). This decrease in expression of *SIGLEC14*, an ITAM-coupling protein, and concomitant increase in expression of *SIGLEC5*, an ITIM-containing protein, may lead to a dampened microglial activation state or proportion of activated microglia in deletion carriers. We speculate that decreased *SIGLEC14* ex-pression and increased *SIGLEC5* expression may decrease the phagocytic capacity in AD. This is similar to the inverse relationship between *TREM2* and *CD33*, two well-known AD risk factors. Loss of the ITAM-containing *TREM2* decreases phagocytic capacity, while loss of *CD33* increases phagocytic capacity (Bhattacherjee *et al.*, 2019; A. Griciuc *et al.*, 2019; Malik *et al.*, 2015a). Since TREM2, which couples with DAP12, is critical for the transition of microglia into a full disease-associated phenotype, *SIGLEC14* may also contribute to this transition (Keren-Shaul *et al.*, 2017). Future studies could investigate whether at

the single-cell level *SIGLEC14* CNV affects disease-associated microglial induction.

Copy number variation may represent a relatively unexplored source of genetic variation in AD (Ebbert *et al.*, 2019). GWAS such as Jansen et al. rely on SNPs, which do not always capture the full range of variation (Jansen *et al.*, 2019). Additionally, "camouflaged" genes such as *SIGLEC5* and *SIGLEC14* with high sequence identity due to gene duplication are challenging for WGS and WES technologies which rely on small fragments of DNA sequence, typically under 250 bp reads (Ebbert *et al.*, 2019). As such, variants which may have disease relevance and association may be overlooked with current methods. *SIGLEC14* is an example of one such possibly overlooked risk contributor in AD. *SIGLEC14* encodes an ITAM protein and signals through DAP12 similar to TREM2, and deletion of *SIGLEC14* is associated with increased AD risk, also similar to SNPs that reduce TREM2 function (Hollingworth *et al.*, 2011; Kunkle *et al.*, 2019; Lambert *et al.*, 2013; Naj *et al.*, 2011). Ligands for SIGLEC14, which include sialylated proteins, are commonly found within amyloid plaques similar to ligands for TREM2. We propose that the effect size and significance of association are masked through copy number variation not accounted for using the proxy SNP alone, i.e., loss of SIGLEC14 function likely increases risk, but the proxy SNP rs1106476 occasionally also marks the individuals with an extra *SIGLEC14* copy, thus reducing the power of rs1106476 association with AD. We thus conclude that *SIGLEC14* represents a potentially overlooked AD genetic risk factor due to complex genetics.

Table 4.1. Genes that are nominally significant for AD association with strong pQTL signal

Gene	SNP	P (pQTL)	β (pQTL)	P (AD)	β (AD)	N (AD)	ITIM/ITAM
CD33	rs12459419	0 †	−0.94	7.13×10^{-9}	−0.01330	458,744	ITIM
FCGR3B	rs10919543	3.20×10^{-67}	0.44	0.000317	0.00806	445,293	ITAM
LILRA5	rs759819	2.50×10^{-111}	−0.54	0.00186	0.00717	454,216	ITAM
LILRB2	rs373032	7.60×10^{-146}	−0.72	0.00227	0.00763	463,880	ITIM
SIGLEC9	rs2075803	0†	−1.23	0.00703	0.00576	466,252	ITIM
SIRPB1	rs3848788	1.20×10^{-213}	0.75	0.00942	0.00582	458,092	ITAM
COLEC12	rs2846667	9.30×10^{-12}	0.20	0.0177	0.00586	449,987	ITAM
FCRL1	rs4971155	6.30×10^{-26}	−0.26	0.0197	−0.00520	403,829	ITAM
NCR1	rs2278428	1.10×10^{-15}	−0.36	0.0249	0.00815	466,252	ITAM
SIGLEC14	rs1106476	0 †	−1.19	0.0284	0.00736	458,063	ITAM
FCRL3	rs7528684	1.40×10^{-112}	0.53	0.04	−0.00434	458,744	Both
MRC2	rs146385050	1.30×10^{-11}	−0.22	0.041	−0.00612	396,686	ITAM
SLAMF6	rs11291564	2.60×10^{-12}	0.20	0.042	−0.02450	17,477	ITAM

† The p-value in the analyzed summary statistics was reported as exactly 0. This does not impact our analysis, as our threshold was any cis-pQTL at p < 0.05.

Table 4.2. Overlap of pQTL and AD signals

pQTLs	ITIM/ITAM (%)	Not ITIM/ITAM (%)	Total
AD $p < 0.05$	13 (28)	54 (10)	67
AD $p > 0.05$	34 (72)	488 (90)	522
Total	47 (100)	542 (100)	589

Table 4.3. Evaluation of rs1106476 as a proxy for SIGLEC14 deletion

SIGLEC14 Copies	rs1106476 T/T	rs1106476 A/T	rs1106476 A/A	Total
0	0	1	1	2
1	6	13	0	19
2	39	0	0	39
3	2	2	0	4
Total	47	16	1	64

Blue = predicted correlation of SIGLEC14 deletion vs. rs1106476. Each cell represents the number of DNA samples with the indicated SIGLEC14 copy number and rs1106476 genotype.

Table 4.4 Summary of the *SIGLEC14* CNV in the 3095 sample ADSP WGS dataset

SIGLEC14 Copy Number	Caucasian	African American	Other	Total
0	24	74	44	142
1	304	348	316	968
2	692	522	652	1866
3	21	53	43	117
4	0	1	1	2
Total	1041	998	1056	3095
Deletion MAF	0.1691	0.2485	0.1913	0.2023
Addition MAF	0.0101	0.0276	0.0213	0.0195

MAF: Minor allele frequency
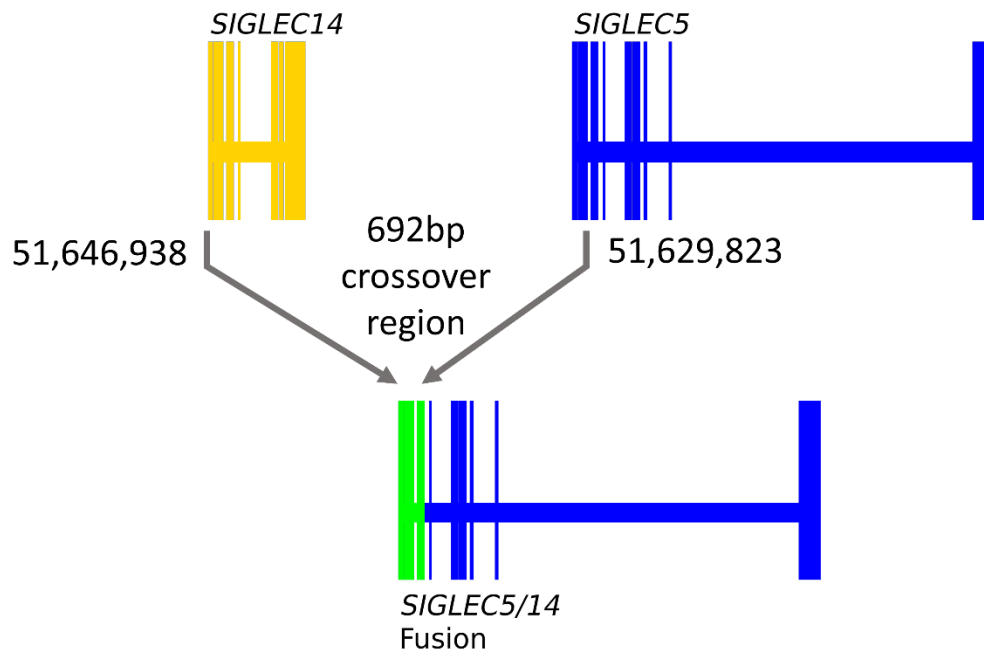
Figure 4.1. Identification of the SIGLEC14 deletion site

Coordinates in both are for reference genome. Exons 1-3 of *SIGLEC14* and *SIGLEC5* are identical which confounds exact determination of the crossover event. The yellow region depicts *SIGLEC14*, the blue region depicts *SIGLEC5*, while the green region depicts the 692 bp region of complete identity where the crossover deletion occurs.
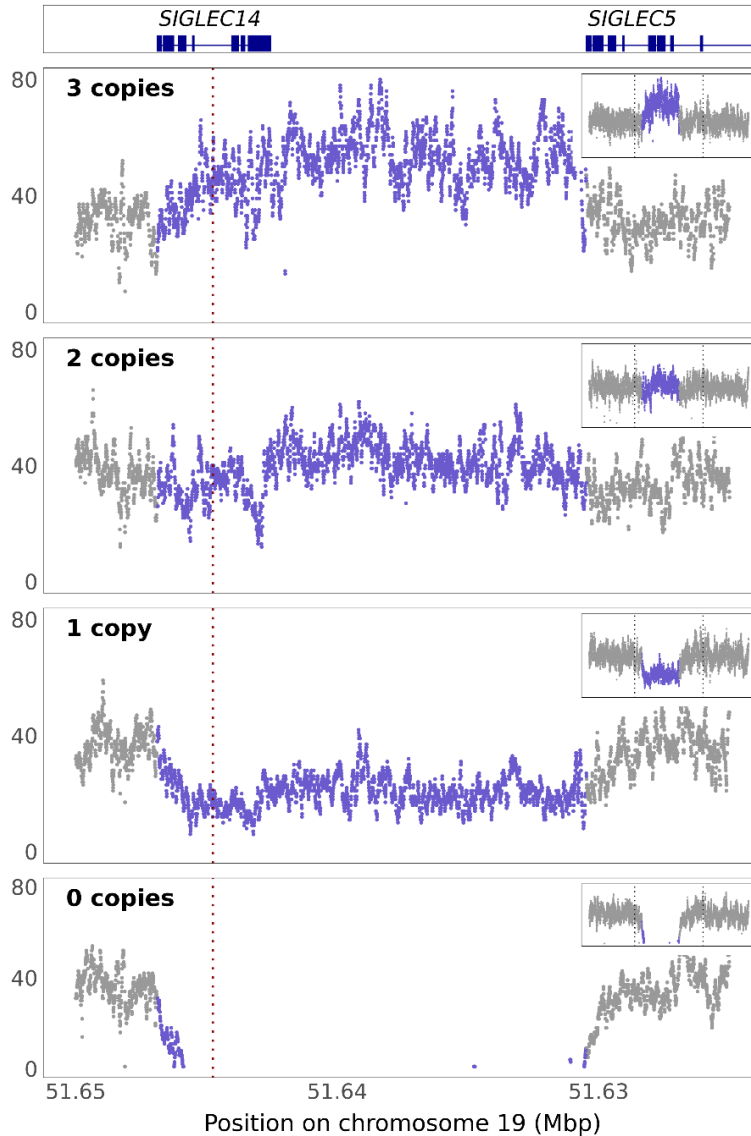
Figure 4.2. *SIGLEC14* CNVs detected in ADNI and ADSP cohorts

Read depth shown by chromosomal position of whole-genome sequencing in a representative example of each CNV detected. Exon/intron maps for *SIGLEC14* and *SIGLEC5* at figure top for reference. Purple: copy number variation. Inset: expanded view of locus. Red dotted line: location of copy number variation assay. The dotted line in the insets shows the boundaries of the full-size image.
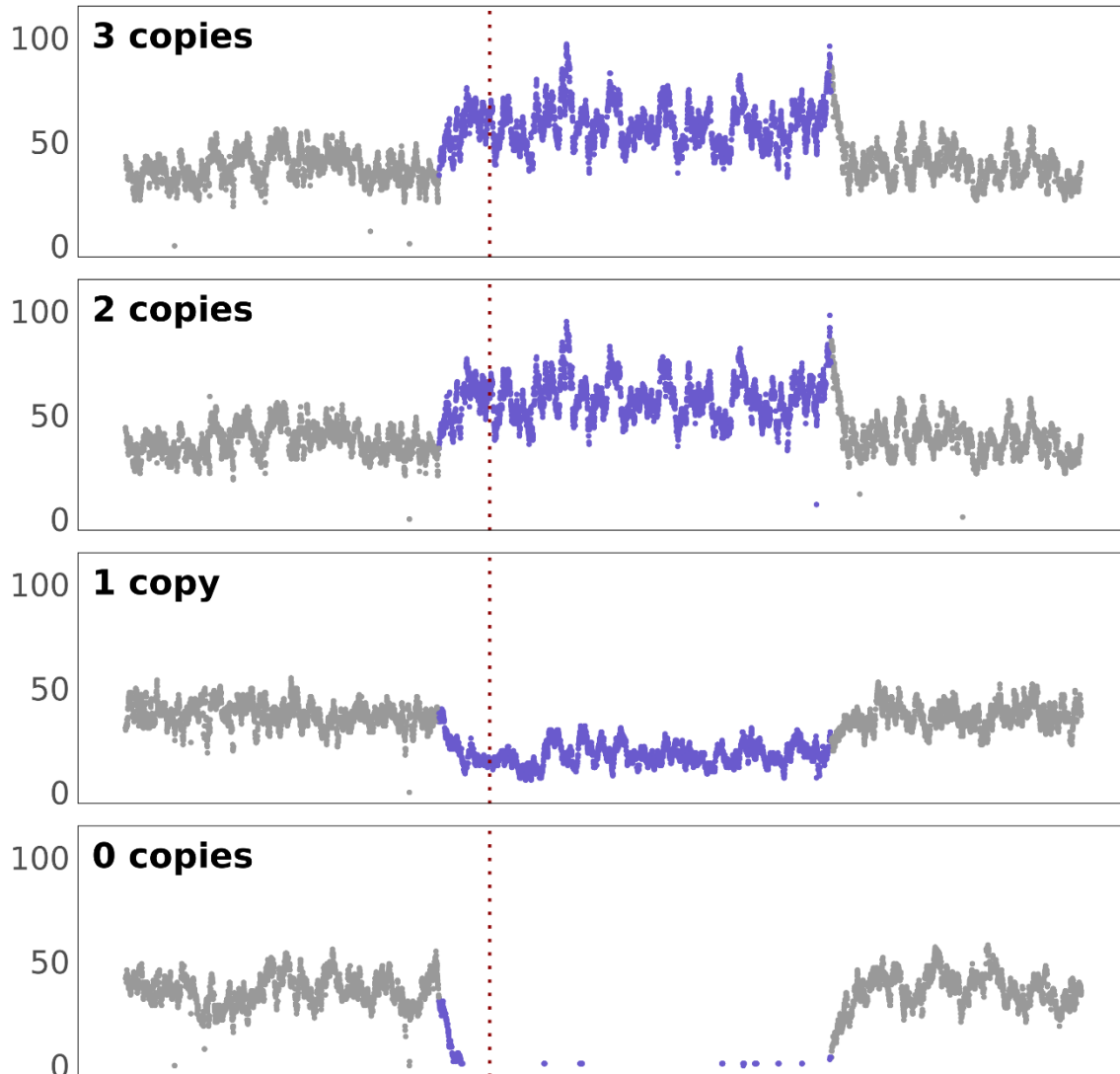
Figure 4.3: WGS read depth data from the ASDP in Caucasian population

Whole genome sequencing (WGS) read depth data from the Alzheimer's Disease Sequencing Project (ASDP) in Caucasian population (n = 1041) reveals copy number variation (CNV) in *SIGLEC14*, including deletions and duplications.
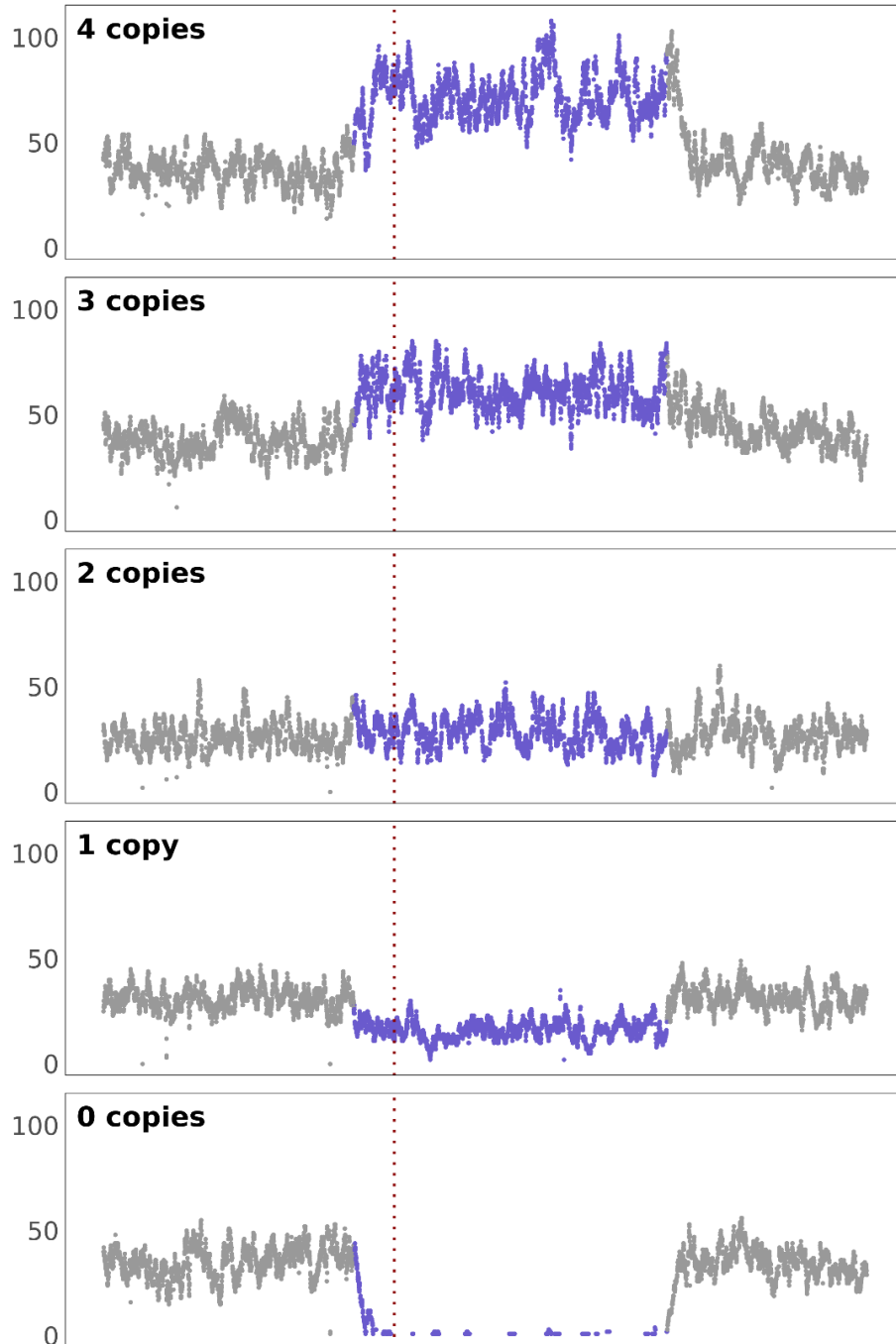
Figure 4.4: WGS read depth data from the ASDP in African American population

WGS read depth data from the ASDP in African American population (n = 998) reveals CNV in *SIGLEC14*, including deletions and duplications. Especially notable is the 4-allele carrier. Frequencies described in Table 4.4
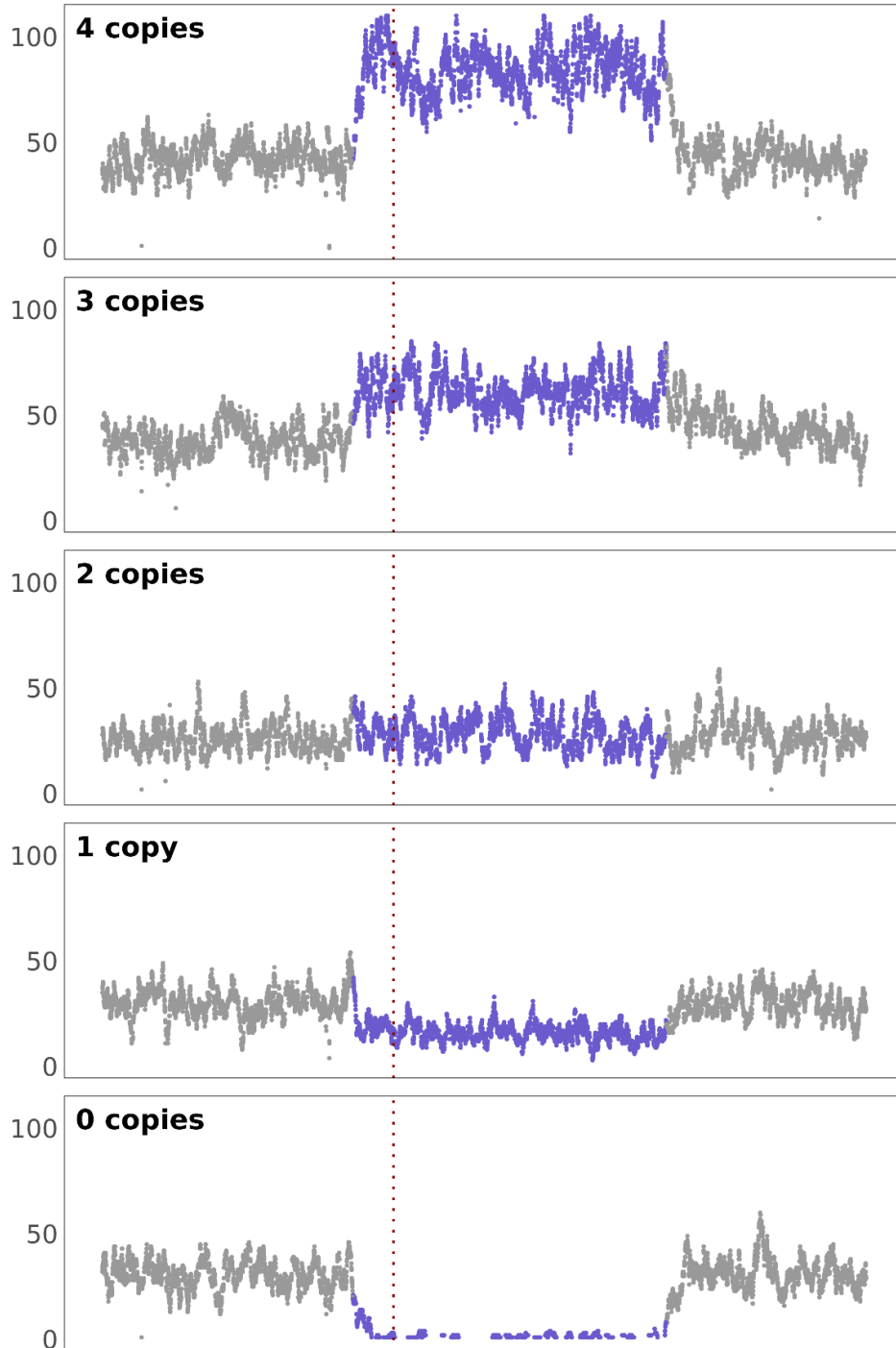
Figure 4.5: WGS read depth data from the ASDP in all other populations

WGS read depth data from the ASDP in all other populations (n = 1056) reveals CNV in *SIGLEC14*, including deletions and duplications. Especially notable is the 4-allele carrier. Frequencies described in Table 4.
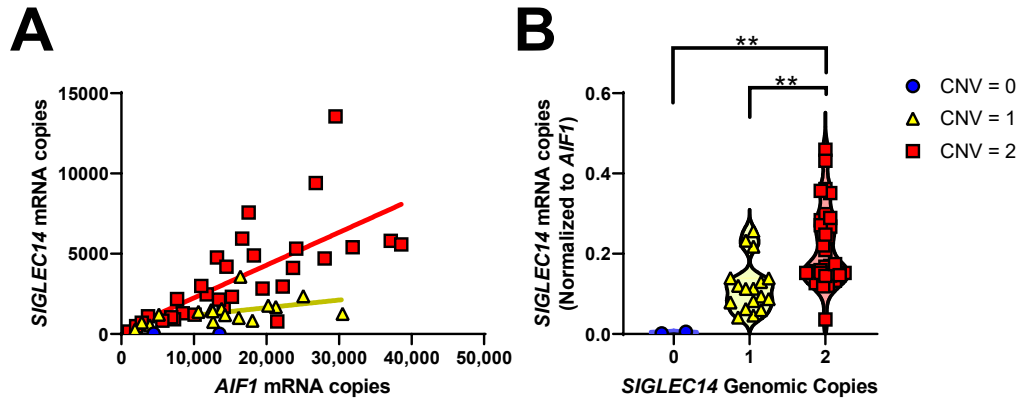
Figure 4.6. SIGLEC14 expression correlates with microglial gene AIF1 and SIGLEC14 CNV

(A) SIGLEC14 is expressed in microglia ($p < 0.0001$, $F_{1,48} = 33.19$, $r2 = 0.409$). (B) SIGLEC14 CNV strongly correlates with SIGLEC14 gene expression ($p = 0.0002$, $F_{2,47} = 10.679$), Tukey's post-hoc multiple comparisons test. ** $p < 0.01$. We do not have statistical power to compare expression with CNV > 2, given its low MAF.
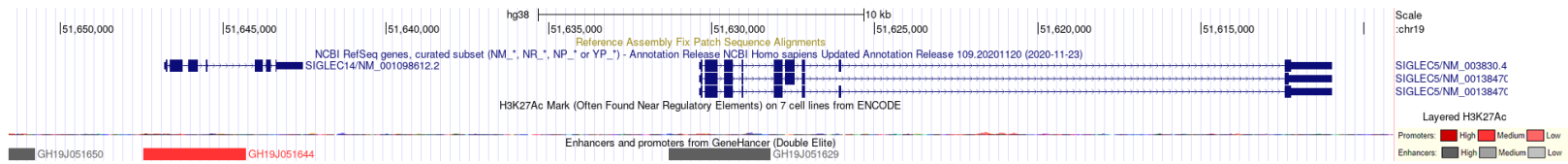
Figure 4.7. The *SIGLEC14* locus contains no H3K27Ac peaks nor regulatory elements between *SIGLEC14* and *SIGLEC5*

Expression of *SIGLEC14* is approximately ten times higher than *SIGLEC5* in individuals with both copies of *SIGLEC14*, while *SIGLEC5* expression is higher in individuals lacking *SIGLEC14* copies, in keeping with a common promoter or enhancer governing the single locus.
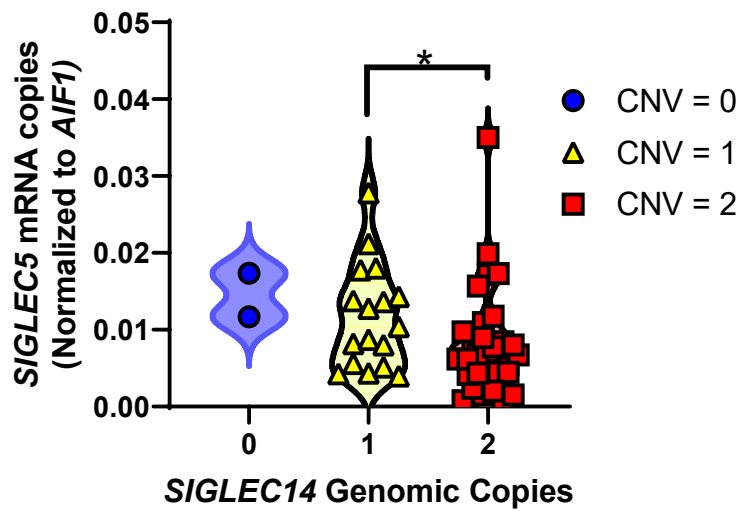
Figure 4.8. *SIGLEC5* expression inversely correlates with *SIGLEC14* CNV

    SIGLEC5 expression increases with fewer copies of *SIGLEC14*, presumably due to proximity to regulatory elements (p = 0.0220, $F_{2,46}$ = 4.151), Tukey's post-hoc multiple comparisons test. * p = 0.0389.
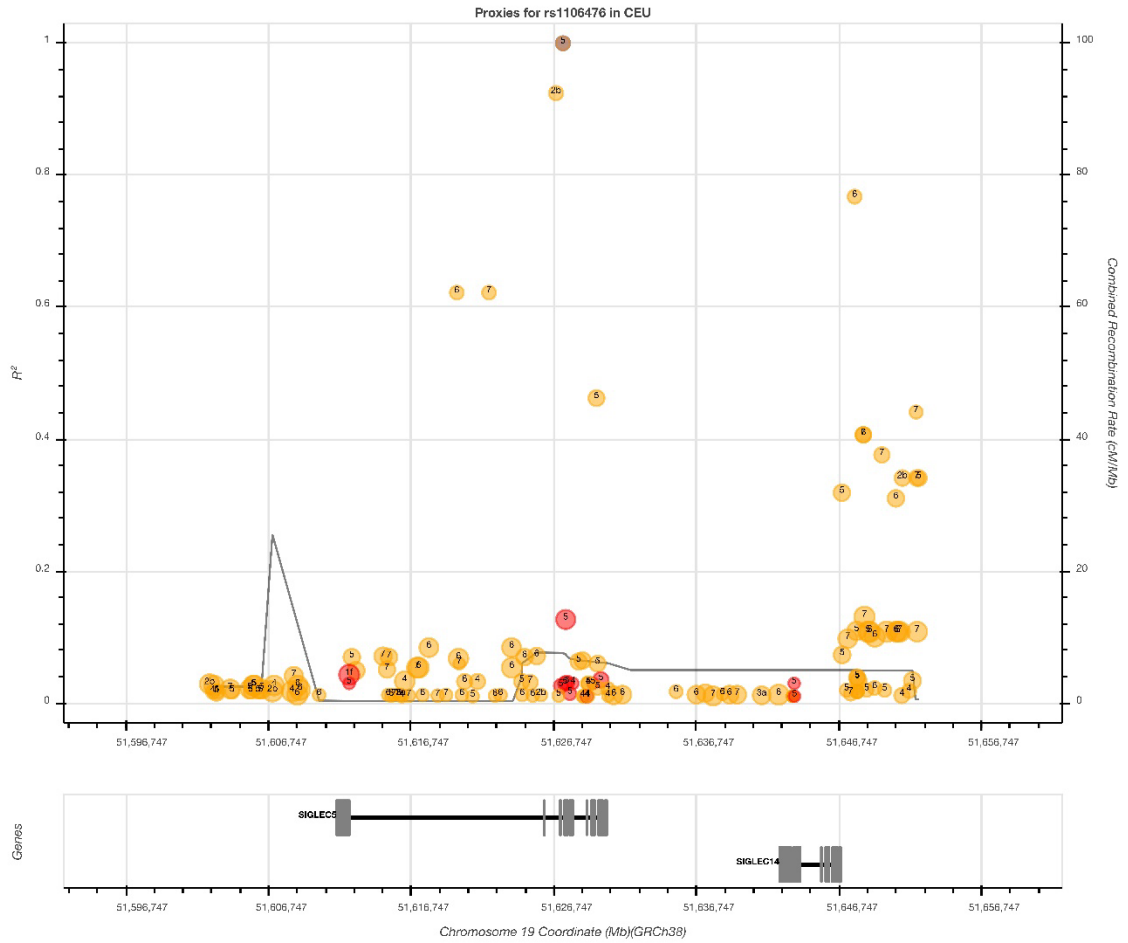
Figure 4.9. *SIGLEC5* and *SIGLEC14* share a broad recombination peak

Recombination peak is indicated by the grey line. Note that, since *SIGLEC14* and *SIGLEC5* are on the minus strand, these genes appear inverted in this figure and read right-to-left.

CHAPTER 5. DISCUSSION AND FUTURE DIRECTIONS

5.1 Primary findings and summary of dissertation

The primary findings of this dissertation are wide-ranging, given the breadth of topics covered. First, I found that *TREM2* undergoes substantial alternative splicing, much more than previously reported. The *D2-TREM2* alternative splice isoform, a novel isoform, did not correlate with any AD-associated SNPs in our dataset nor AD status or pathology but was detected across multiple tissues indicating this is not a brain-specific—and therefore not a microglia-specific—isoform. This *D2-TREM2* isoform is translated to protein and colocalizes primarily with the Golgi marker GM-130, similar to the full-length TREM2 protein, when overexpressed in the HMC3 human microglial cell line.

I also found that CRISPR-Cas9 editing can lead to not only off-target edits, but gene conversions using pseudogenes as repair templates. This finding advanced the fields of molecular biology and acute myeloid leukemia therapeutics simultaneously, as this was a previously unreported phenomenon and provided the highest resolution of the gemtuzumab epitope to date. These studies were initially undertaken to generate a model of endogenous-level expression of *D2-CD33*. Overexpression studies have shown previously that *D2-CD33* protein is expressed at the cell surface, however this finding has not been replicated in physiologically relevant expression levels (Bhattacherjee *et al.*, 2021; Malik *et al.*, 2015a; Siddiqui *et al.*, 2017). Whether the D2-CD33 protein is made at all is still a matter of controversy. Humbert *et al.* (2019) employed a similar CRISPR-Cas9 editing strategy and was unable to detect D2-CD33 protein at the cell surface by

flow cytometry but did not investigate intracellular expression. My flow cytometry results in multiple clones of well-validated edited cells replicated the Humbert *et al.* (2019) paper; I did not detect D2-CD33 protein at the cell surface. Bhattacherjee *et al.* (2021) found by intracellular flow cytometry and a custom-generated antibody that the D2-CD33 protein is expressed in transgenic mouse microglia, though personal communications with the group indicate they are struggling to see this protein using alternative methods. Siddiqui *et al.* (2017) reported that the D2-CD33 isoform is particularly susceptible to degradation using Western blot techniques but was able to detect the isoform by immunofluorescence and found it colocalizes with peroxisomes. Notably, Siddiqui *et al.* (2017) used a custom-generated antibody and clone HIM3-4. Neither our laboratory nor the Macauley laboratory have been able to replicate the custom-generated antibody labeling even in transfected cells. The Macauley laboratory also has not been able to replicate the HIM3-4 labeling in their transgenic microglia expressing *D2-CD33*. Subsequent immunofluorescence and Western blot studies in our laboratory have not conclusively detected any endogenous D2-CD33 protein, implying that under endogenous conditions the mRNA is not translated into protein.

Finally, I found that pQTLs which affect ITIM or ITAM genes were significantly overrepresented in nominally significant AD associations. This is in congruence with previous studies highlighting the importance of microglia in the pathogenesis of AD. During this study, I primarily focused on *SIGLEC14* as there was a known deletion polymorphism with a proxy SNP (Yamanaka *et al.*, 2009). This proxy SNP was nominally associated with AD in the Jansen study ($p = 0.02$)

with a strong pQTL signal ($p < 10^{-10}$) (Jansen *et al.*, 2019; Sun *et al.*, 2018). I found that this proxy SNP, rs1106476, does not fully capture the *SIGLEC14* CNV. Specifically, there was approximately 11% disagreement between the predicted and actual CNV based on the proxy SNP, and this proxy SNP also did not predict the novel additional genomic copies of *SIGLEC14*. Given that the proxy SNP was associated with both a loss of a genomic copy and drastically decreased protein expression, I hypothesized that this would also lead to a corresponding decrease in mRNA expression. We confirmed this using qPCR, showing that individuals with only one genomic copy of *SIGLEC14* have 54.6% of the gene expression when compared with an individual with two genomic copies.

In summary, I have studied AD genetics spanning the breadth of molecular biology—at the DNA, mRNA, and protein levels. I have discovered new splicing isoforms in *TREM2*, helped develop new theories behind the actions of the protective variants in *CD33*, and strengthened the existing evidence implicating microglia in AD pathogenesis. The following sections will describe the implications of my work and the new directions it has opened.

5.2 Future Directions: *TREM2* splicing modulation as a proposed therapeutic target for Alzheimer's Disease

As discussed in Chapter 2, there seems to be a window early in AD during which TREM2 activity is beneficial which later shifts to become detrimental. This is evidenced by AD mouse models crossed with *Trem2*$^{-/-}$ mice. In the 5xFAD model crossed with *Trem2*$^{-/-}$, mice exhibited increased cognitive deficits and amyloid pathology indicating a critical role for TREM2 in limiting early-stage AD (A. Griciuc

et al., 2019; McQuade *et al.*, 2020; Wang *et al.*, 2015). By contrast, the PS19 tau model crossed with *Trem2⁻/⁻* exhibited less tau pathology indicating that TREM2 is actually detrimental during the late-stage disease (Gratuze *et al.*, 2020). The alternative splicing of *TREM2* is particularly relevant here: during the initial phase of AD when amyloid pathology is first detected, a CNS-directed splicing modulator to enhance exon 2 inclusion may increase TREM2 protein in microglia. Presuming the patient is followed in clinic with periodic biomarker imaging, clinicians will notice the initial formation of tau pathology after some years. At this point, the strategy may be shifted to enhance exon 2 skipping to increase *D2-TREM2*. Since D2-TREM2 protein is missing its ligand binding domain, we hypothesize that this is a dead receptor as was the initial hypothesis regarding D2-CD33; increased expression of this dead receptor may decrease TREM2 activity. Strategies to test splicing modulators are currently under development in the Estus laboratory. I am currently working on the development of an expression vector which produces eGFP when exon 2 is included (i.e., typical splicing) but switches to dsRED when exon 2 is skipped (i.e., *D2-TREM2*). This vector can then be ectopically expressed in various cell lines for high-throughput screening of splicing factors and small molecule drugs using a flow cytometer and a ratiometric response between eGFP and dsRED expression to determine the shift in splicing efficiency when compared to a control sample.

Another approach using *TREM2* splicing modulation surrounds the *D4-TREM2* isoform. This may be useful in the earliest stages of AD, or even as a prophylactic in genetically at-risk individuals. As discussed in Chapter 1, sTREM2

retains its affinity and avidity for both Aβ and APOE. It is thus conceptually possible that sTREM2 acts to cap Aβ$_{42}$ monomers and oligomers and preclude growth into large plaques. Further work is needed to determine if sTREM2 is beneficial in limiting plaque growth, however. If sTREM2 does in fact prevent Aβ deposition, very early (e.g., 10 years prior to the onset of neuropsychiatric symptoms or mild cognitive impairment) modulation to promote *D4-TREM2* may prove beneficial in delaying the onset of mild cognitive impairment (MCI) and ultimately AD. This strategy with respect to sTREM2 is not limited to splicing, however. It is also conceivable that recombinant sTREM2 protein could be used similar to the recently approved yet controversial aducanumab, an anti-amyloid monoclonal antibody therapy. Given that sTREM2 is substantially smaller than an antibody, it may more readily cross the blood-brain barrier though other limitations such as hydrophobicity must also be considered. In summary, the alternative splicing isoforms of *TREM2* have generated two potential avenues of translational research which could ultimately lead to therapeutics and a potential high-throughput screening mechanism for *TREM2* exon 2 splicing modulators.

5.3   Future Directions: *CD33* gene editing and gain-of-function studies

In Chapter 3, I detailed my work using CRISPR-Cas9 to generate an endogenous expression model of *D2-CD33*. During this model generation, I came upon the unexpected *SIGLEC22P* pseudogene-mediated gene conversion within *CD33* and showed that this conversion disrupts the gemtuzumab epitope. This has interesting implications for AML research and treatment. Currently, a CRISPR-Cas9 strategy is under consideration to hedge against relapses. The antibody-drug

conjugate gemtuzumab-ozogamicin couples the anti-CD33 antibody gemtuzumab to a calicheamicin, ozogamicin, which targets the minor groove of DNA in the cell and promotes strand scission. Upon gemtuzumab-CD33 ligation, the receptor-antibody-drug complex is internalized and the ozogamicin is released into the cell, thus targeting only CD33-expressing cells. Researchers are currently investigating the use of CRISPR-Cas9 to generate $CD33^{null}$ hematopoietic stem cells for autologous transplantation after remission of AML. This would allow the use of gemutzumab-ozogamicin in the event of a relapse, but only targeted to the pre-existing CD33-expressing cells responsible for the original AML. This approach relies on complete ablation of CD33 which is important for self-recognition, and thus may be deleterious if no longer expressed by myeloid cells. By contrast, one could instead simply target the three bases identified in Chapter 3 to generate hematopoietic cells which are not tagged by gemtuzumab and may still retain CD33 function. To date, we have not confirmed whether the KIR-CD33 is functional, but based on x-ray crystallography these three residues are not within the ligand binding domain and thus not predicted to disrupt signaling. Clearly more studies would be required to better understand the safety profile of such a mutation and its functional effects, but my data show equivalent expression levels as wild-type CD33 on the cell surface. Since this KIR-CD33 retains cell surface localization and potentially function but is masked from gemtuzumab, this approach has a clear advantage over simply abrogating all CD33 expression.

Shifting focus to AD, the generation of an endogenous expression model of *D2-CD33* has provided needed clarity with respect to the protein expression and

localization. In my studies, I did not detect expression of the D2-CD33 protein at the cell surface by flow cytometry, intracellularly by confocal microscopy, or within cell lysates by Western blot using antibody clones recognizing multiple epitopes. I interpret this to mean that the D2-CD33 protein is either not translated or is unstable and degraded relative to full-length CD33. Also of note is the fact that D2-CD33 protein is translated, stable, and trafficked to the cell surface using overexpression cDNA vectors, and readily recognized by numerous antibody clones in all three methods which failed under endogenous expression. This illustrates that the antibodies and methods are capable of detecting D2-CD33 protein if it were present in these edited cells. If we then consider the gain-of-function hypothesis, which I still hold as substantiated by genetic evidence, then we can rule out a change in protein function as the function gained. This presents more possibilities: the protein may be causing a misfolded protein response in the Golgi thus stressing the cell, or the mRNA itself has some activity. In the case of the former, it is possible that the misfolded protein response leads to induction of genes which promote degradation of proteins or generation of additional proteasomes and lysosomes, potentially aiding in the degradation of phagocytosed amyloid as well. In the latter case, it is possible that the *D2-CD33* mRNA acts as some sort of long non-coding RNA (lncRNA) to suppress expression of another unknown gene. These are both highly speculative, but interesting, hypothetical mechanisms by which *D2-CD33* may have a gain-of-function. I will note, however, that it is possible that the *D2-CD33* gain-of-function hypothesis is incorrect and this isoform simply generates a dead receptor as has

102

been the prevailing hypothesis until the identification of the rs201074739 indel. Investigating the alternatives listed above is not without merit, though, as there is always something to be learned. After all, we would not have the clarity we do now regarding the gemtuzumab epitope nor the possibility of using KIR-CD33 edited cells in AML treatments without my current work on the gain-of-function hypothesis.

5.4   Future Directions: Alternative microglial targets for Alzheimer's Disease

Finally, I will turn to the investigation of genes not significantly associated by GWAS as potential AD risk factors, but suggestive based on pathway analyses. I focused on *SIGLEC14* in Chapter 4, identifying additional copy number variation and linking deletion of *SIGLEC14*, a DAP12-coupling protein, to increased expression of *SIGLEC5*, an ITIM protein. These are paired receptors with the same extracellular domain and, thus, the same ligands. As members of the SIGLEC family, they also recognize sialic acids which are abundant in post-translational modifications and gangliosides present in the heterogeneous Aβ plaques. The implication here is that loss of the activating SIGLEC14 with concomitant increase in inhibitory SIGLEC5 expression may lead to a more subdued microglial phenotype, possibly decreasing overall microglial activity. This also brings about a blind-side of GWAS—these studies are typically focused on SNPs, rather than all variations, and thus missing a substantial portion of human genetic variability. Notably, some small indels are captured in GWAS; for instance, the 4 bp rs201074739 indel within *CD33* was captured in the recent Jansen *et al.* (2019) study. However, larger indels and structural variants, such as the 692 bp

SIGLEC14-SIGLEC5 gene fusion or variants within repeated regions such as in *CR1*, are missed in current GWAS. Including the full range of variation provides not only better coverage to identify risk factors but also improves our interpretation. The *D2-CD33* gain-of-function hypothesis only came about due to the presence of the rs201074739 indel in the Jansen *et al.* (2019) summary statistics. This allowed us to reason that if a complete loss-of-function variant, such as a frameshift indel, is not genome-wide significant, then a partial loss-of-function with less of an effect also should not be significantly associated with disease; thus, any associated variants must have a gain-of-function. The full range of variation can add to our understanding of variants with limited information, or at least inform our experiments. Studies using whole-genome sequencing instead of SNP-based GWAS are currently underway.

Looking broader, if we understand the directionality of the effect of a variant on disease, we can then infer whether to target its up- or downstream signaling partners. In the case of the *TREM2-CD33* signaling pathway, the preclinical *in vivo* and *in vitro* models are clear that TREM2 promotes amyloid uptake, while CD33 represses amyloid uptake. The consensus of the field is that, in the early stages of disease, targeting amyloid clearance is paramount. TREM2-activating antibodies are currently in clinical trials to facilitate this amyloid clearance. It thus stands to reason that targeting adjacent proteins may prove equally useful. Increasing the total or active Syk, BLNK, or PLCγ proteins through transcription factor activation, kinases, or bispecific antibodies (i.e., clustering TREM2 with other DAP12-coupling proteins to facilitate DAP12 phosphorylation) may increase microglial

activity and amyloid phagocytosis. This highlights the importance of the big picture of using GWAS as a tool in translational research—this big-data style research is a means, not an end, to understanding the complexities of heritance in human disease and translating to therapies. Once a pathway is well-validated by big-data techniques, we should not limit ourselves only to those few risk factors identified in the dataset. Rather, we should be marrying big-data with basic cell and molecular biology to develop sound preclinical models surrounding a risk factor, targeting surrounding and ancillary proteins as needed for drug development. As a case-in-point, *ABI3* has been identified in multiple GWAS as an AD risk factor. This is an intracellular protein, involved in regulating actin dynamics. Drugs targeting this molecule specifically will have to consider the active site of ABI3, in addition to being cell permeant and crossing the blood-brain barrier. Targeting upstream signaling which induces *ABI3* expression may be a better option. Specifically, pharmaceutical companies are most interested in receptor and enzyme inhibitors, in addition to receptor activating ligands. Pathway analysis allows us to expand into these more easily targeted steps of the pathway. Thus, it is important to see the forest through the trees when considering the implications of GWAS in translational research.

5.5 Closing Remarks

In summary, I have uncovered new *TREM2* alternative splicing isoforms, CRISPR-Cas9 repair mechanisms, and copy number variation in *SIGLEC14*. During the *CD33* model generation, I also concluded that the D2-CD33 protein is not stably expressed at physiologic expression levels, thus calling into question

our current interpretations behind this gain-of-function mechanism. While I still hold this hypothesis, I now believe that the protein itself is not functional, but either the misfolded protein response or the mRNA is causing some change which leads to increased microglial activation. As is often the case with good scientific work, this dissertation presents more questions than it answers. My work has covered a wide breadth of topics which have served to advance the field of AD research and molecular biology as a whole, while also providing a firm foundation on which to begin a scientific career.

APPENDIX

This table contains the list of all ITIM and ITAM containing, or DAP12-coupling, proteins and their gene names used in Chapter 4. The "Name" column refers to a protein or gene name, while the "Alias 1" and "Alias 2" columns refer to alternative, common names for the same protein.

NAME	ALIAS 1	ALIAS 2
CLEC1A		
CLEC1B		
CLEC2A		
CLEC2B		
CD69	CLEC2C	
CLEC2D		
CLEC2L		
CLEC3A		
CLEC3B		
CLEC4A		
CLEC4C		
CLEC4D		
CLEC4E		
CLEC4F		
CLEC4G		
ASGR1	CLEC4H1	

NAME	ALIAS 1	ALIAS 2
ASGR2	CLEC4H2	
FCER2	CLEC4J	CD23
CD207	CLEC4K	
CD209	CLEC4L	
CLEC4M		
CLEC5A		
CLEC6A		
CLEC7A		
OLR1	CLEC8A	
CLEC9A		
CLEC10A		
CLEC11A		
CLEC12A		
CLEC12B		
CD302	CLEC13A	
LY75	CLEC13B	CD205
PLA2R1	CLEC13C	
MRC1	CLEC13D	CD206
MRC2	CLEC13E	CD280
CLEC14A		
CLEC16A		
CLEC17A		

NAME	ALIAS 1	ALIAS 2
KLRA1		
KLRB1	CLEC5B	NK1.1
KLRC1		
KLRC2		
KLRC3		
KLRC4		
KLRD1		
KLRF1	CLEC5C	
KLRG1	CLEC15A	
KLRG2	CLEC15B	
KLRK1		
AGC1		
ATRNL1		
BCAN		
CD248		
CD72		
CD93		
CHODL		
CL-K1-IA		
CL-K1-IB		
CL-K1-IC		
CLECSF5		

NAME	ALIAS 1	ALIAS 2
COLEC10		
COLEC11		
COLEC12		
CSPG3		
FCER2		
FREM1		
HBXBP		
LAYN		
LOC348174		
LOC728276		
MAFA		
MBL2		
MGC34761		
MICL		
MRC1L1		
OLR1		
PAP		
PKD1		
PKD1L2		
PLA2R1		
PRG2		
PRG3		

NAME	ALIAS 1	ALIAS 2
REG1A		
REG1B		
REG3A		
REG3G		
REG4		
SELE		
SELL		
SELP		
SFTPA1		
SFTPA2		
SFTPA2B		
SFTPD		
SRCL		
THBD		
VCAN		
CD64		
CD32		
CD16A		
CD16B		
FCER1		
FCD23		
CD89		

NAME	ALIAS 1	ALIAS 2
FCAMR		
FCRN		
FCRL1		
FCRL2		
FCRL3		
FCRL4		
FCRL5		
FCRL6		
FCRLA		
FCRLB		
KIR2DL1		
KIR2DL2		
KIR2DL3		
KIR2DL4		
KIR2DL5A		
KIR2DL5B		
KIR2DS1		
KIR2DS2		
KIR2DS3		
KIR2DS4		
KIR2DS5		
KIR3DL1		

NAME	ALIAS 1	ALIAS 2
KIR3DL2		
KIR3DL3		
KIR3DS1		
LILRA1		
LILRA2		
LILRA3		
LILRA4		
LILRA5		
LILRA6		
LILRB1		
LILRB2		
LILRB3		
LILRB4		
LILRB5		
LILRB6		
LILRB7		
LILRA6		
LILRA5		
SIGLEC-1		
SIGLEC-2	CD22	
SIGLEC-3	CD33	
SIGLEC-4	MAG	

NAME	ALIAS 1	ALIAS 2
SIGLEC-5		
SIGLEC-5		
SIGLEC-6		
SIGLEC-7		
SIGLEC-8		
SIGLEC-9		
SIGLEC-10		
SIGLEC-11		
SIGLEC-12		
SIGLEC-13		
SIGLEC-14		
SIGLEC-15		
SIGLEC-16		
SIGLEC-17		
SLAMF1	CD150	
SLAMF2	CD48	
SLAMF3	CD229	LY9
SLAMF4	CD244	
SLAMF5	CD84	
SLAMF6	CD352	
SLAMF7	CD319	CRACC
SLAMF8	CD353	

NAME	ALIAS 1	ALIAS 2
SLAMF9		
NKP44		
NKP46		
NKP30		
PILRA		
PILRB		
CD28		
ICOS		
BTLA		
CTLA-4		
PD-1		
CD200R1		
CD5		
CD6		
LAIR1		
GPVI		
OSCAR		

BIBLIOGRAPHY

Abdollahi, M. R., Huang, S., Rodriguez, S., Guthrie, P. A. I., Smith, G. D., Ebrahim, S., Lawlor, D. A., Day, I. N. M., & Gaunt, T. R. (2008). Homogeneous Assay of rs4343, anACEI/D Proxy, and an Analysis in the British Women's Heart and Health Study (BWHHS). *Disease Markers*, *24*(1), 11-17. https://doi.org/10.1155/2008/813679

Allen, M., Zou, F., Chai, H. S., Younkin, C. S., Crook, J., Pankratz, V. S., Carrasquillo, M. M., Rowley, C. N., Nair, A. A., Middha, S., Maharjan, S., Nguyen, T., Ma, L., Malphrus, K. G., Palusak, R., Lincoln, S., Bisceglio, G., Georgescu, C., Schultz, D., Rakhshan, F., Kolbert, C. P., Jen, J., Haines, J. L., Mayeux, R., Pericak-Vance, M. A., Farrer, L. A., Schellenberg, G. D., Petersen, R. C., Graff-Radford, N. R., Dickson, D. W., Younkin, S. G., Ertekin-Taner, N., Alzheimer's Disease Genetics, C., Apostolova, L. G., Arnold, S. E., Baldwin, C. T., Barber, R., Barmada, M. M., Beach, T., Beecham, G. W., Beekly, D., Bennett, D. A., Bigio, E. H., Bird, T. D., Blacker, D., Boeve, B. F., Bowen, J. D., Boxer, A., Burke, J. R., . . . Woltjer, R. L. (2012). Novel late-onset Alzheimer disease loci variants associate with brain gene expression. *Neurology*, *79*(3), 221-228. https://doi.org/10.1212/WNL.0b013e3182605801 PMID:22722634

Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *17*(3), 327-406. https://doi.org/10.1002/alz.12328

Alzheimer, A. (1907). Uber eine eigenartige Erkrankung der Hirnrinde. *Zentralbl. Nervenh. Psych.*, *18*, 177-179.

Angata, T. (2018). Possible Influences of Endogenous and Exogenous Ligands on the Evolution of Human Siglecs. *Front Immunol*, *9*, 2885. https://doi.org/10.3389/fimmu.2018.02885 PMID:30564250

Anstey, K. J., Cherbuin, N., Budge, M., & Young, J. (2011). Body mass index in midlife and late-life as a risk factor for dementia: a meta-analysis of prospective studies. *Obes Rev*, *12*(5), e426-437. https://doi.org/10.1111/j.1467-789X.2010.00825.x PMID:21348917

Ariga, T., McDonald, M. P., & Yu, R. K. (2008). Thematic Review Series: Sphingolipids. Role of ganglioside metabolism in the pathogenesis of Alzheimer's disease—a review. *Journal of Lipid Research*, *49*(6), 1157-1175. https://doi.org/10.1194/jlr.r800007-jlr200

Atagi, Y., Liu, C.-C., Painter, M. M., Chen, X.-F., Verbeeck, C., Zheng, H., Li, X., Rademakers, R., Kang, S. S., Xu, H., Younkin, S., Das, P., Fryer, J. D., & Bu, G. (2015). Apolipoprotein E Is a Ligand for Triggering Receptor Expressed on Myeloid Cells 2 (TREM2). *Journal of Biological Chemistry*, *290*(43), 26043-26050. https://doi.org/10.1074/jbc.m115.679043

Bacskai, B. J., Frosch, M. P., Freeman, S. H., Raymond, S. B., Augustinack, J. C., Johnson, K. A., Irizarry, M. C., Klunk, W. E., Mathis, C. A., Dekosky, S. T., Greenberg, S. M., Hyman, B. T., & Growdon, J. H. (2007). Molecular Imaging With Pittsburgh Compound B Confirmed at Autopsy. *Archives of Neurology*, *64*(3), 431. https://doi.org/10.1001/archneur.64.3.431

Balaian, L., Zhong, R. K., & Ball, E. D. (2003). The inhibitory effect of anti-CD33 monoclonal antibodies on AML cell growth correlates with Syk and/or ZAP-70 expression. *Exp Hematol, 31*(5), 363-371. https://doi.org/10.1016/s0301-472x(03)00044-4 PMID:12763134

Barrow, A. D., & Trowsdale, J. (2006). You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling. *Eur J Immunol, 36*(7), 1646-1653. https://doi.org/10.1002/eji.200636195 PMID:16783855

Beecham, G. W., Hamilton, K., Naj, A. C., Martin, E. R., Huentelman, M., Myers, A. J., Corneveaux, J. J., Hardy, J., Vonsattel, J. P., Younkin, S. G., Bennett, D. A., De Jager, P. L., Larson, E. B., Crane, P. K., Kamboh, M. I., Kofler, J. K., Mash, D. C., Duque, L., Gilbert, J. R., Gwirtsman, H., Buxbaum, J. D., Kramer, P., Dickson, D. W., Farrer, L. A., Frosch, M. P., Ghetti, B., Haines, J. L., Hyman, B. T., Kukull, W. A., Mayeux, R. P., Pericak-Vance, M. A., Schneider, J. A., Trojanowski, J. Q., Reiman, E. M., Alzheimer's Disease Genetics, C., Schellenberg, G. D., & Montine, T. J. (2014). Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet, 10*(9), e1004606. https://doi.org/10.1371/journal.pgen.1004606 PMID:25188341

Bertram, L., Lange, C., Mullin, K., Parkinson, M., Hsiao, M., Hogan, M. F., Schjeide, B. M., Hooli, B., Divito, J., Ionita, I., Jiang, H., Laird, N., Moscarillo, T., Ohlsen, K. L., Elliott, K., Wang, X., Hu-Lince, D., Ryder, M., Murphy, A., Wagner, S. L., Blacker, D., Becker, K. D., & Tanzi, R. E. (2008). Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *Am J Hum Genet, 83*(5), 623-632. https://doi.org/10.1016/j.ajhg.2008.10.008 PMID:18976728

Bhattacherjee, A., Jung, S. J., Ho, M., Eskandari-Sedighi, G., St. Laurent, C. D., McCord, K. A., Bains, A., Gaurav, S., Sarkar, S. S., Plemel, J., & Macauley, M. S. (2021). The CD33 short isoform is a gain-of-function variant that enhances Aβ1-42 phagocytosis in microglia. *Molecular Neurodegeneration*.

Bhattacherjee, A., Rodrigues, E., Jung, J., Luzentales-Simpson, M., Enterina, J. R., Galleguillos, D., St Laurent, C. D., Nakhaei-Nejad, M., Fuchsberger, F. F., Streith, L., Wang, Q., Kawasaki, N., Duan, S., Bains, A., Paulson, J. C., Rademacher, C., Giuliani, F., Sipione, S., & Macauley, M. S. (2019). Repression of phagocytosis by human CD33 is not conserved with mouse CD33. *Commun Biol, 2*, 450. https://doi.org/10.1038/s42003-019-0698-6 PMID:31815204

Bird, T. D. (1993). Alzheimer Disease Overview. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. W. Gripp, G. M. Mirzaa, & A. Amemiya (Eds.), *GeneReviews(®)*. University of Washington, Seattle.

Borot, F., Wang, H., Ma, Y., Jafarov, T., Raza, A., Ali, A. M., & Mukherjee, S. (2019). Gene-edited stem cells enable CD33-directed immune therapy for myeloid malignancies. *Proceedings of the National Academy of Sciences*, 201819992. https://doi.org/10.1073/pnas.1819992116

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica, 82*(4), 239-259. https://doi.org/10.1007/bf00308809

Bradshaw, E. M., Chibnik, L. B., Keenan, B. T., Ottoboni, L., Raj, T., Tang, A., Rosenkrantz, L. L., Imboywa, S., Lee, M., Von Korff, A., Alzheimer Disease Neuroimaging, I., Morris, M. C., Evans, D. A., Johnson, K., Sperling, R. A., Schneider, J. A., Bennett, D. A., & De Jager, P. L. (2013). CD33 Alzheimer's disease locus: altered monocyte function and amyloid biology. *Nat Neurosci*, *16*(7), 848-850. https://doi.org/10.1038/nn.3435 PMID:23708142

Burchett, M. E., Ling, I. F., & Estus, S. (2011). FBN1 isoform expression varies in a tissue and development-specific fashion. *Biochemical and Biophysical Research Communications*, *411*(2), 323-328. https://doi.org/10.1016/j.bbrc.2011.06.140

Burdick, D., Soreghan, B., Kwon, M., Kosmoski, J., Knauer, M., Henschen, A., Yates, J., Cotman, C., & Glabe, C. (1992). Assembly and aggregation properties of synthetic Alzheimer's A4/beta amyloid peptide analogs. *Journal of Biological Chemistry*, *267*(1), 546-554. https://doi.org/https://doi.org/10.1016/S0021-9258(18)48529-8

Chan, G., White, C. C., Winn, P. A., Cimpean, M., Replogle, J. M., Glick, L. R., Cuerdon, N. E., Ryan, K. J., Johnson, K. A., Schneider, J. A., Bennett, D. A., Chibnik, L. B., Sperling, R. A., Bradshaw, E. M., & De Jager, P. L. (2015). CD33 modulates TREM2: convergence of Alzheimer loci. *Nat Neurosci*, *18*(11), 1556-1558. https://doi.org/10.1038/nn.4126 PMID:26414614

Chauhan, L., Shin, M., Wang, Y. C., Loken, M., Pollard, J., Aplenc, R., Hirsch, B. A., Raimondi, S., Ries, R. E., Bernstein, I. D., Gamis, A. S., Alonzo, T. A., Meshinchi, S., & Lamba, J. K. (2019). CD33_PGx6_Score Predicts Gemtuzumab Ozogamicin Response in Childhood Acute Myeloid Leukemia: A Report From the Children's Oncology Group. *JCO Precis Oncol*, *3*. https://doi.org/10.1200/po.18.00387 PMID:32914031

Cheng, Q., Danao, J., Talreja, S., Wen, P., Yin, J., Sun, N., Li, C. M., Chui, D., Tran, D., Koirala, S., Chen, H., Foltz, I. N., Wang, S., & Sambashivan, S. (2018). TREM2-activating antibodies abrogate the negative pleiotropic effects of the Alzheimer's disease variant Trem2(R47H) on murine myeloid cell function. *J Biol Chem*, *293*(32), 12620-12633. https://doi.org/10.1074/jbc.RA118.001848 PMID:29599291

Clark, P. C., Kutner, N. G., Goldstein, F. C., Peterson-Hazen, S., Garner, V., Zhang, R., & Bowles, T. (2005). Impediments to Timely Diagnosis of Alzheimer's Disease in African Americans. *Journal of the American Geriatrics Society*, *53*(11), 2012-2017. https://doi.org/https://doi.org/10.1111/j.1532-5415.2005.53569.x

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Jr., Rimmler, J. B., Locke, P. A., Conneally, P. M., Schmader, K. E., & et al. (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*, *7*(2), 180-184. https://doi.org/10.1038/ng0694-180 PMID:7920638

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's

Disease in Late Onset Families. *Science*, *261*(5123), 921-923. https://doi.org/doi:10.1126/science.8346443

Cosker, K., Mallach, A., Limaye, J., Piers, T. M., Staddon, J., Neame, S. J., Hardy, J., & Pocock, J. M. (2021). Microglial signalling pathway deficits associated with the patient derived R47H TREM2 variants linked to AD indicate inability to activate inflammasome. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-91207-1

Dean, H. B., Roberson, E. D., & Song, Y. (2019). Neurodegenerative Disease–Associated Variants in TREM2 Destabilize the Apical Ligand-Binding Region of the Immunoglobulin Domain [Original Research]. *Frontiers in Neurology*, *10*(1252). https://doi.org/10.3389/fneur.2019.01252

Del-Aguila, J. L., Benitez, B. A., Li, Z., Dube, U., Mihindukulasuriya, K. A., Budde, J. P., Farias, F. H. G., Fernández, M. V., Ibanez, L., Jiang, S., Perrin, R. J., Cairns, N. J., Morris, J. C., Harari, O., & Cruchaga, C. (2019). TREM2 brain transcript-specific studies in AD and TREM2 mutation carriers. *Molecular Neurodegeneration*, *14*(1). https://doi.org/10.1186/s13024-019-0319-3

Dickstein, D. L., Pullman, M. Y., Fernandez, C., Short, J. A., Kostakoglu, L., Knesaurek, K., Soleimani, L., Jordan, B. D., Gordon, W. A., Dams-O'Connor, K., Delman, B. N., Wong, E., Tang, C. Y., DeKosky, S. T., Stone, J. R., Cantu, R. C., Sano, M., Hof, P. R., & Gandy, S. (2016). Cerebral [(18) F]T807/AV1451 retention pattern in clinically probable CTE resembles pathognomonic distribution of CTE tauopathy. *Transl Psychiatry*, *6*(9), e900. https://doi.org/10.1038/tp.2016.175 PMID:27676441

Dushek, O., Goyette, J., & van der Merwe, P. A. (2012). Non-catalytic tyrosine-phosphorylated receptors. *Immunol Rev*, *250*(1), 258-276. https://doi.org/10.1111/imr.12008 PMID:23046135

Ebbert, M. T. W., Jensen, T. D., Jansen-West, K., Sens, J. P., Reddy, J. S., Ridge, P. G., Kauwe, J. S. K., Belzil, V., Pregent, L., Carrasquillo, M. M., Keene, D., Larson, E., Crane, P., Asmann, Y. W., Ertekin-Taner, N., Younkin, S. G., Ross, O. A., Rademakers, R., Petrucelli, L., & Fryer, J. D. (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol*, *20*(1), 97. https://doi.org/10.1186/s13059-019-1707-2 PMID:31104630

Efthymiou, A. G., & Goate, A. M. (2017). Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol Neurodegener*, *12*(1), 43. https://doi.org/10.1186/s13024-017-0184-x PMID:28549481

Estus, S., Shaw, B. C., Devanney, N., Katsumata, Y., Press, E. E., & Fardo, D. W. (2019). Evaluation of CD33 as a genetic risk factor for Alzheimer's disease. *Acta Neuropathol*. https://doi.org/10.1007/s00401-019-02000-4 PMID:30949760

Ewers, M., Biechele, G., Suárez-Calvet, M., Sacher, C., Blume, T., Morenas-Rodriguez, E., Deming, Y., Piccio, L., Cruchaga, C., Kleinberger, G., Shaw, L., Trojanowski, J. Q., Herms, J., Dichgans, M., Initiative, t. A. s. D. N., Brendel, M., Haass, C., & Franzmeier, N. (2020). Higher CSF sTREM2 and microglia activation are associated with slower rates of beta-amyloid

accumulation. *EMBO Molecular Medicine*, *12*(9), e12308. https://doi.org/https://doi.org/10.15252/emmm.202012308

Finkel, R. S., Mercuri, E., Darras, B. T., Connolly, A. M., Kuntz, N. L., Kirschner, J., Chiriboga, C. A., Saito, K., Servais, L., Tizzano, E., Topaloglu, H., Tulinius, M., Montes, J., Glanzman, A. M., Bishop, K., Zhong, Z. J., Gheuens, S., Bennett, C. F., Schneider, E., Farwell, W., & De Vivo, D. C. (2017). Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *New England Journal of Medicine*, *377*(18), 1723-1732. https://doi.org/10.1056/nejmoa1702752

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., & Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, *2017*. https://doi.org/10.1093/database/bax028

Gandy, S., & Heppner, F. L. (2013). Microglia as dynamic and essential components of the amyloid hypothesis. *Neuron*, *78*(4), 575-577. https://doi.org/10.1016/j.neuron.2013.05.007 PMID:23719156

Gbadamosi, M., Meshinchi, S., & Lamba, J. K. (2018). Gemtuzumab ozogamicin for treatment of newly diagnosed CD33-positive acute myeloid leukemia. *Future Oncol*, *14*, 3199-3213. https://doi.org/10.2217/fon-2018-0325 PMID:30039981

Gbadamosi, M. O., Shastri, V. M., Hylkema, T., Papageorgiou, I., Pardo, L., Cogle, C. R., Doty, A., Loken, M. R., Meshinchi, S., & Lamba, J. K. (2021). Novel CD33 antibodies unravel localization, biology and therapeutic implications of CD33 isoforms. *Future Oncol*, *17*(3), 263-277. https://doi.org/10.2217/fon-2020-0746 PMID:33356566

Gianattasio, K. Z., Prather, C., Glymour, M. M., Ciarleglio, A., & Power, M. C. (2019). Racial disparities and temporal trends in dementia misdiagnosis risk in the United States. *Alzheimers Dement (N Y)*, *5*, 891-898. https://doi.org/10.1016/j.trci.2019.11.008 PMID:31890853

Godwin, C. D., Laszlo, G. S., Wood, B. L., Correnti, C. E., Bates, O. M., Garling, E. E., Mao, Z. J., Beddoe, M. E., Lunn, M. C., Humbert, O., Kiem, H.-P., & Walter, R. B. (2020). The CD33 splice isoform lacking exon 2 as therapeutic target in human acute myeloid leukemia. *Leukemia*, *34*(9), 2479-2483. https://doi.org/10.1038/s41375-020-0755-7

Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. *Neuropsychologia*, *49*(14), 3826-3830. https://doi.org/10.1016/j.neuropsychologia.2011.09.041

Grant, W. B. (2014). Trends in diet and Alzheimer's disease during the nutrition transition in Japan and developing countries. *J Alzheimers Dis*, *38*(3), 611-620. https://doi.org/10.3233/JAD-130719 PMID:24037034

Gratuze, M., Leyns, C. E. G., & Holtzman, D. M. (2018). New insights into the role of TREM2 in Alzheimer's disease. *Molecular Neurodegeneration*, *13*(1), 66. https://doi.org/10.1186/s13024-018-0298-9

Gratuze, M., Leyns, C. E. G., Sauerbeck, A. D., St-Pierre, M.-K., Xiong, M., Kim, N., Serrano, J. R., Tremblay, M.-È., Kummer, T. T., Colonna, M., Ulrich, J.

D., & Holtzman, D. M. (2020). Impact of TREM2R47H variant on tau pathology–induced gliosis and neurodegeneration. *Journal of Clinical Investigation*, *130*(9), 4954-4968. https://doi.org/10.1172/jci138179

Griciuc, A., Patel, S., Federico, A. N., Choi, S. H., Innes, B. J., Oram, M. K., Cereghetti, G., McGinty, D., Anselmo, A., Sadreyev, R. I., Hickman, S. E., El Khoury, J., Colonna, M., & Tanzi, R. E. (2019). TREM2 Acts Downstream of CD33 in Modulating Microglial Pathology in Alzheimer's Disease. *Neuron*, *103*(5), 820-835 e827. https://doi.org/10.1016/j.neuron.2019.06.010 PMID:31301936

Griciuc, A., Serrano-Pozo, A., Parrado, Antonio R., Lesinski, Andrea N., Asselin, Caroline N., Mullin, K., Hooli, B., Choi, Se H., Hyman, Bradley T., & Tanzi, Rudolph E. (2013). Alzheimer's Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. *Neuron*, *78*(4), 631-643. https://doi.org/https://doi.org/10.1016/j.neuron.2013.04.014

Griciuc, A., & Tanzi, R. E. (2021). The role of innate immune genes in Alzheimer's disease. *Curr Opin Neurol*, *34*(2), 228-236. https://doi.org/10.1097/WCO.0000000000000911 PMID:33560670

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J. S., Younkin, S., Hazrati, L., Collinge, J., Pocock, J., Lashley, T., Williams, J., Lambert, J. C., Amouyel, P., Goate, A., Rademakers, R., Morgan, K., Powell, J., St George-Hyslop, P., Singleton, A., Hardy, J., & Alzheimer Genetic Analysis, G. (2013). TREM2 variants in Alzheimer's disease. *N Engl J Med*, *368*(2), 117-127. https://doi.org/10.1056/NEJMoa1211851 PMID:23150934

Guerreiro, R. J., Gustafson, D. R., & Hardy, J. (2012). The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiology of Aging*, *33*(3), 437-456. https://doi.org/10.1016/j.neurobiolaging.2010.03.025

Han, S., Na, Y., Koh, I., Nho, K., & Lee, Y. (2021). Alternative Splicing Regulation of Low-Frequency Genetic Variants in Exon 2 of TREM2 in Alzheimer's Disease by Splicing-Based Aggregation. *International Journal of Molecular Sciences*, *22*(18), 9865. https://doi.org/10.3390/ijms22189865

Hansen, D. V., Hanson, J. E., & Sheng, M. (2017). Microglia in Alzheimer's disease. *Journal of Cell Biology*, *217*(2), 459-472. https://doi.org/10.1083/jcb.201709069

Henderson, A. S. (1988). The risk factors for Alzheimer's disease: a review and a hypothesis. *Acta Psychiatr Scand*, *78*(3), 257-275. https://doi.org/10.1111/j.1600-0447.1988.tb06336.x PMID:3057813

Hernández-Caselles, T., Martínez-Esparza, M., Pérez-Oliva, A. B., Quintanilla-Cecconi, A. M., García-Alonso, A., Alvarez-López, D. M. R., & García-Peñarrubia, P. (2006). A study of CD33 (SIGLEC-3) antigen expression and function on activated human T and NK cells: two isoforms of CD33 are generated by alternative splicing. *Journal of Leukocyte Biology*, *79*(1), 46-58. https://doi.org/https://doi.org/10.1189/jlb.0205096

Herrup, K. (2010). Reimagining Alzheimer's disease--an age-based hypothesis. *J Neurosci*, *30*(50), 16755-16762. https://doi.org/10.1523/JNEUROSCI.4521-10.2010 PMID:21159946

Hinds, D. A., Kloek, A. P., Jen, M., Chen, X., & Frazer, K. A. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genetics*, *38*(1), 82-85. https://doi.org/10.1038/ng1695

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am Stat*, *55*(1), 19-24.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., Abraham, R., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., Powell, J., Proitsi, P., Lupton, M. K., Brayne, C., Rubinsztein, D. C., Gill, M., Lawlor, B., Lynch, A., Brown, K. S., Passmore, P. A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A. D., Beaumont, H., Warden, D., Wilcock, G., Love, S., Kehoe, P. G., Hooper, N. M., Vardy, E. R., Hardy, J., Mead, S., Fox, N. C., Rossor, M., Collinge, J., Maier, W., Jessen, F., . . . Williams, J. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*, *43*(5), 429-435. https://doi.org/10.1038/ng.803 PMID:21460840

Holtman, I. R., Raj, D. D., Miller, J. A., Schaafsma, W., Yin, Z., Brouwer, N., Wes, P. D., Möller, T., Orre, M., Kamphuis, W., Hol, E. M., Boddeke, E. W. G. M., & Eggen, B. J. L. (2015). Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis. *Acta Neuropathologica Communications*, *3*(1). https://doi.org/10.1186/s40478-015-0203-5

Holtzman, D. M., Morris, J. C., & Goate, A. M. (2011). Alzheimer's Disease: The Challenge of the Second Century. *Science Translational Medicine*, *3*(77), 77sr71-77sr71. https://doi.org/10.1126/scitranslmed.3002369

Huebbe, P., & Rimbach, G. (2017). Evolution of human apolipoprotein E (APOE) isoforms: Gene structure, protein function and interaction with dietary factors. *Ageing Res Rev*, *37*, 146-161. https://doi.org/10.1016/j.arr.2017.06.002 PMID:28647612

Humbert, O., Laszlo, G. S., Sichel, S., Ironside, C., Haworth, K. G., Bates, O. M., Beddoe, M. E., Carrillo, R. R., Kiem, H.-P., & Walter, R. B. (2019). Engineering resistance to CD33-targeted immunotherapy in normal hematopoiesis by CRISPR/Cas9-deletion of CD33 exon 2. *Leukemia*, *33*(3), 762-808. https://doi.org/10.1038/s41375-018-0277-8

Hyman, B. T., & Trojanowski, J. Q. (1997). Editorial on Consensus Recommendations for the Postmortem Diagnosis of Alzheimer Disease from the National Institute on Aging and the Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer Disease. *Journal of Neuropathology and Experimental Neurology*, *56*(10), 1095-1097. https://doi.org/10.1097/00005072-199710000-00002

Isakov, N. (1997). Immunoreceptor tyrosine-based activation motif (ITAM), a unique module linking antigen and Fc receptors to their signaling cascades. *Journal of Leukocyte Biology*, *61*(1), 6-16. https://doi.org/10.1002/jlb.61.1.6

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., Hampel, H., Jagust, W. J., Johnson, K. A., Knopman, D. S., Petersen, R. C., Scheltens, P., Sperling, R. A., & Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539-547. https://doi.org/10.1212/wnl.0000000000002923

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hagg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Braekhus, A., Brathen, G., de Leeuw, C., Desikan, R. S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T. J., Jonsson, P. V., Kiddle, S. J., Rongve, A., Saltvedt, I., Sando, S. B., Selbaek, G., Shoai, M., Skene, N. G., Snaedal, J., Stordal, E., Ulstein, I. D., Wang, Y., White, L. R., Hardy, J., Hjerling-Leffler, J., Sullivan, P. F., van der Flier, W. M., Dobson, R., Davis, L. K., Stefansson, H., Stefansson, K., . . . Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. https://doi.org/10.1038/s41588-018-0311-9 PMID:30617256

Javidi-Parsijani, P., Lyu, P., Makani, V., Sarhan, W. M., Yoo, K. W., El-Korashi, L., Atala, A., & Lu, B. (2020). CRISPR/Cas9 increases mitotic gene conversion in human cells. *Gene Ther*, *27*(6), 281-296. https://doi.org/10.1038/s41434-020-0126-z PMID:32020049

Jay, T. R., von Saucken, V. E., & Landreth, G. E. (2017). TREM2 in Neurodegenerative Diseases. *Mol Neurodegener*, *12*(1), 56. https://doi.org/10.1186/s13024-017-0197-5 PMID:28768545

Jayadev, S., Steinbart, E. J., Chi, Y. Y., Kukull, W. A., Schellenberg, G. D., & Bird, T. D. (2008). Conjugal Alzheimer disease: risk in children when both parents have Alzheimer disease. *Arch Neurol*, *65*(3), 373-378. https://doi.org/10.1001/archneurol.2007.61 PMID:18332250

Jeffreys, A. J., & May, C. A. (2004). Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nature Genetics*, *36*(2), 151-156. https://doi.org/10.1038/ng1287

Jiang, T., Tan, L., Chen, Q., Tan, M.-S., Zhou, J.-S., Zhu, X.-C., Lu, H., Wang, H.-F., Zhang, Y.-D., & Yu, J.-T. (2016). A rare coding variant in TREM2 increases risk for Alzheimer's disease in Han Chinese. *Neurobiology of Aging*, *42*, 217.e211-217.e213. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2016.02.023

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821. https://doi.org/10.1126/science.1225829 PMID:22745249

Jonaitis, E., La Rue, A., Mueller, K. D., Koscik, R. L., Hermann, B., & Sager, M. A. (2013). Cognitive activities and cognitive performance in middle-aged

adults at risk for Alzheimer's disease. *Psychol Aging*, *28*(4), 1004-1014. https://doi.org/10.1037/a0034838 PMID:24364404

Jones, L., Holmans, P. A., Hamshere, M. L., Harold, D., Moskvina, V., Ivanov, D., Pocklington, A., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Pahwa, J. S., Jones, N., Stretton, A., Morgan, A. R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M. K., Brayne, C., Rubinsztein, D. C., Gill, M., Lawlor, B., Lynch, A., Morgan, K., Brown, K. S., Passmore, P. A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A. D., Love, S., Kehoe, P. G., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Schurmann, B., Heun, R., Kolsch, H., van den Bussche, H., Heuser, I., Peters, O., Kornhuber, J., Wiltfang, J., . . . Williams, J. (2010). Genetic evidence implicates the immune system and cholesterol metabolism in the aetiology of Alzheimer's disease. *PLoS One*, *5*(11), e13950. https://doi.org/10.1371/journal.pone.0013950 PMID:21085570

Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P. V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A. I., Lah, J. J., Rujescu, D., Hampel, H., Giegling, I., Andreassen, O. A., Engedal, K., Ulstein, I., Djurovic, S., Ibrahim-Verbaas, C., Hofman, A., Ikram, M. A., Van Duijn, C. M., Thorsteinsdottir, U., Kong, A., & Stefansson, K. (2013). Variant of TREM2 Associated with the Risk of Alzheimer's Disease. *New England Journal of Medicine*, *368*(2), 107-116. https://doi.org/10.1056/nejmoa1211103

Karch, C. M., & Goate, A. M. (2015). Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry*, *77*(1), 43-51. https://doi.org/10.1016/j.biopsych.2014.05.006 PMID:24951455

Karch, C. M., Jeng, A. T., Nowotny, P., Cady, J., Cruchaga, C., & Goate, A. M. (2012). Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One*, *7*(11), e50976. https://doi.org/10.1371/journal.pone.0050976 PMID:23226438

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434-443. https://doi.org/10.1038/s41586-020-2308-7 PMID:32461654

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, A. D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996-1006. https://doi.org/10.1101/gr.229102

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., Itzkovitz,

S., Colonna, M., Schwartz, M., & Amit, I. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell*, *169*(7), 1276-1290 e1217. https://doi.org/10.1016/j.cell.2017.05.018 PMID:28602351

Khachaturian, Z. S. (1985). Diagnosis of Alzheimer's Disease. *Archives of Neurology*, *42*(11), 1097-1105. https://doi.org/10.1001/archneur.1985.04060100083029

Kida, E., Choi-Miura, N. H., & Wisniewski, K. E. (1995). Deposition of apolipoproteins E and J in senile plaques is topographically determined in both Alzheimer's disease and Down's syndrome brain. *Brain Res*, *685*(1-2), 211-216. https://doi.org/10.1016/0006-8993(95)00482-6 PMID:7583250

Kiianitsa, K., Kurtz, I., Beeman, N., Matsushita, M., Chien, W.-M., Raskind, W. H., & Korvatska, O. (2021). Novel TREM2 splicing isoform that lacks the V-set immunoglobulin domain is abundant in the human brain. *Journal of Leukocyte Biology*, *1*(9). https://doi.org/https://doi.org/10.1002/JLB.2HI0720-463RR

Kim, M. Y., Yu, K.-R., Kenderian, S. S., Ruella, M., Chen, S., Shin, T.-H., Aljanahi, A. A., Schreeder, D., Klichinsky, M., Shestova, O., Kozlowski, M. S., Cummins, K. D., Shan, X., Shestov, M., Bagg, A., Morrissette, J. J. D., Sekhri, P., Lazzarotto, C. R., Calvo, K. R., Kuhns, D. B., Donahue, R. E., Behbehani, G. K., Tsai, S. Q., Dunbar, C. E., & Gill, S. (2018). Genetic Inactivation of CD33 in Hematopoietic Stem Cells to Enable CAR T Cell Immunotherapy for Acute Myeloid Leukemia. *Cell*, *173*(6), 1439-1453.e1419. https://doi.org/10.1016/j.cell.2018.05.013

Kirk-Sanchez, N. J., & McGough, E. L. (2014). Physical exercise and cognitive performance in the elderly: current perspectives. *Clin Interv Aging*, *9*, 51-62. https://doi.org/10.2147/CIA.S39506 PMID:24379659

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G. F., Estrada, S., Ausén, B., Debnath, M. L., Barletta, J., Price, J. C., Sandell, J., Lopresti, B. J., Wall, A., Koivisto, P., Antoni, G., Mathis, C. A., & Långström, B. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol*, *55*(3), 306-319. https://doi.org/10.1002/ana.20009 PMID:14991808

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., Olaso, R., Raybould, R., Chen, Y., Kuzma, A. B., Hiltunen, M., Morgan, T., Ahmad, S., Vardarajan, B. N., Epelbaum, J., Hoffmann, P., Boada, M., Beecham, G. W., Garnier, J. G., Harold, D., Fitzpatrick, A. L., Valladares, O., Moutet, M. L., Gerrish, A., Smith, A. V., Qu, L., Bacq, D., Denning, N., Jian, X., Zhao, Y., Del Zompo, M., Fox, N. C., Choi, S. H., Mateo, I., Hughes, J. T., Adams, H. H., . . . Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*, *51*(3), 414-430. https://doi.org/10.1038/s41588-019-0358-2 PMID:30820047

Lamba, J. K., Chauhan, L., Shin, M., Loken, M. R., Pollard, J. A., Wang, Y. C., Ries, R. E., Aplenc, R., Hirsch, B. A., Raimondi, S. C., Walter, R. B., Bernstein, I. D., Gamis, A. S., Alonzo, T. A., & Meshinchi, S. (2017). CD33 Splicing Polymorphism Determines Gemtuzamab Ozogamicin Response in De Novo Acute Myeloid Leukemia: Report From Randomized Phase III Children's Oncology Group Trial AAML0531. *J Clin Oncol*, *35*(23), 2674-2682. https://doi.org/10.1200/JCO.2016.71.2513 PMID:28644774

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., DeStafano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thorton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., Vardarajan, B. N., Kamatani, Y., Lin, C. F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M. L., Ruiz, A., Bihoreau, M. T., Choi, S. H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O. L., De Jager, P. L., Deramecourt, V., Johnston, J. A., Evans, D., Lovestone, S., Letenneur, L., Moron, F. J., Rubinsztein, D. C., Eiriksdottir, G., Sleegers, K., Goate, A. M., Fievet, N., Huentelman, M. W., Gill, M., . . . Amouyel, P. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*, *45*(12), 1452-1458. https://doi.org/10.1038/ng.2802 PMID:24162737

Laszlo, G. S., Beddoe, M. E., Godwin, C. D., Bates, O. M., Gudgeon, C. J., Harrington, K. H., & Walter, R. B. (2018). Relationship between CD33 expression, splicing polymorphism, and in vitro cytotoxicity of gemtuzamab ozogamicin and the CD33/CD3 BiTErAMG 330. *Haematologica*. https://doi.org/10.3324/haematol.2018.202069 PMID:30115657

Leduc, V., Jasmin-Belanger, S., & Poirier, J. (2010). APOE and cholesterol homeostasis in Alzheimer's disease. *Trends Mol Med*, *16*(10), 469-477. https://doi.org/10.1016/j.molmed.2010.07.008 PMID:20817608

Lee, J.-Y., Lee, C.-H., Shim, S.-H., Seo, H.-K., Kyhm, J.-H., Cho, S., & Cho, Y.-H. (2002). Molecular cytogenetic analysis of the monoblastic cell line U937. *Cancer Genetics and Cytogenetics*, *137*(2), 124-132. https://doi.org/10.1016/s0165-4608(02)00565-4

Lessard, C. B., Malnik, S. L., Zhou, Y., Ladd, T. B., Cruz, P. E., Ran, Y., Mahan, T. E., Chakrabaty, P., Holtzman, D. M., Ulrich, J. D., Colonna, M., & Golde, T. E. (2018). High-affinity interactions and signal transduction between Aβ oligomers and TREM 2. *EMBO Molecular Medicine*, *10*(11). https://doi.org/10.15252/emmm.201809027

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352 PMID:19505943

Linnartz, B., & Neumann, H. (2013). Microglial activatory (immunoreceptor tyrosine-based activation motif)- and inhibitory (immunoreceptor tyrosine-based inhibition motif)-signaling receptors for recognition of the neuronal

glycocalyx. *Glia*, *61*(1), 37-46. https://doi.org/10.1002/glia.22359 PMID:22615186

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., . . . Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580-585. https://doi.org/10.1038/ng.2653

Lubbers, J., Rodriguez, E., & van Kooyk, Y. (2018). Modulation of Immune Tolerance via Siglec-Sialic Acid Interactions. *Front Immunol*, *9*, 2807. https://doi.org/10.3389/fimmu.2018.02807 PMID:30581432

Luchsinger, J. A. (2010a). Diabetes, related conditions, and dementia. *J Neurol Sci*, *299*(1-2), 35-38. https://doi.org/10.1016/j.jns.2010.08.063 PMID:20888602

Luchsinger, J. A. (2010b). Type 2 diabetes, related conditions, in relation and dementia: an opportunity for prevention? *J Alzheimers Dis*, *20*(3), 723-736. https://doi.org/10.3233/JAD-2010-091687 PMID:20413862

Machiela, M. J., & Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, *31*(21), 3555-3557. https://doi.org/10.1093/bioinformatics/btv402 PMID:26139635

Malik, M., Chiles, I. I. I. J., Xi, H. S., Medway, C., Simpson, J., Potluri, S., Howard, D., Liang, Y., Paumi, C. M., Mukherjee, S., Crane, P., Younkin, S., Fardo, D. W., & Estus, S. (2015a). Genetics of CD33 in Alzheimer's disease and acute myeloid leukemia. *Human Molecular Genetics*, *24*(12), 3557-3570. https://doi.org/10.1093/hmg/ddv092

Malik, M., Parikh, I., Vasquez, J. B., Smith, C., Tai, L., Bu, G., LaDu, M. J., Fardo, D. W., Rebeck, G. W., & Estus, S. (2015b). Genetics ignite focus on microglial inflammation in Alzheimer's disease. *Mol Neurodegener*, *10*, 52. https://doi.org/10.1186/s13024-015-0048-1 PMID:26438529

Malik, M., Simpson, J. F., Parikh, I., Wilfred, B. R., Fardo, D. W., Nelson, P. T., & Estus, S. (2013). CD33 Alzheimer's Risk-Altering Polymorphism, CD33 Expression, and Exon 2 Splicing. *J Neurosci*, *33*(33), 13320-13325. http://www.jneurosci.org/content/jneuro/33/33/13320.full.pdf

Martin-Rehrmann, M. D., Hoe, H.-S., Capuani, E. M., & Rebeck, G. W. (2005). Association of apolipoprotein J-positive β-amyloid plaques with dystrophic neurites in alzheimer's disease brain. *Neurotoxicity Research*, *7*(3), 231-241. https://doi.org/10.1007/bf03036452

Matsuzaki, T., Sasaki, K., Tanizaki, Y., Hata, J., Fujimi, K., Matsui, Y., Sekita, A., Suzuki, S. O., Kanba, S., Kiyohara, Y., & Iwaki, T. (2010). Insulin resistance is associated with the pathology of Alzheimer disease: the Hisayama study.

Neurology, *75*(9), 764-770. https://doi.org/10.1212/WNL.0b013e3181eee25f PMID:20739649

Mawuenyega, K. G., Sigurdson, W., Ovod, V., Munsell, L., Kasten, T., Morris, J. C., Yarasheski, K. E., & Bateman, R. J. (2010). Decreased Clearance of CNS -Amyloid in Alzheimer's Disease. *Science*, *330*(6012), 1774-1774. https://doi.org/10.1126/science.1197623

McCarroll, S. A., Huett, A., Kuballa, P., Chilewski, S. D., Landry, A., Goyette, P., Zody, M. C., Hall, J. L., Brant, S. R., Cho, J. H., Duerr, R. H., Silverberg, M. S., Taylor, K. D., Rioux, J. D., Altshuler, D., Daly, M. J., & Xavier, R. J. (2008). Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nature Genetics*, *40*(9), 1107-1112. https://doi.org/10.1038/ng.215

McGeer, P. L., Klegeris, A., Walker, D. G., Yasuhara, O., & McGeer, E. G. (1994). Pathological proteins in senile plaques. *Tohoku J Exp Med*, *174*(3), 269-277. https://doi.org/10.1620/tjem.174.269 PMID:7761992

McKee, A. C., Daneshvar, D. H., Alvarez, V. E., & Stein, T. D. (2014). The neuropathology of sport. *Acta Neuropathol*, *127*(1), 29-51. https://doi.org/10.1007/s00401-013-1230-6 PMID:24366527

McKee, A. C., Stern, R. A., Nowinski, C. J., Stein, T. D., Alvarez, V. E., Daneshvar, D. H., Lee, H. S., Wojtowicz, S. M., Hall, G., Baugh, C. M., Riley, D. O., Kubilus, C. A., Cormier, K. A., Jacobs, M. A., Martin, B. R., Abraham, C. R., Ikezu, T., Reichard, R. R., Wolozin, B. L., Budson, A. E., Goldstein, L. E., Kowall, N. W., & Cantu, R. C. (2013). The spectrum of disease in chronic traumatic encephalopathy. *Brain*, *136*(Pt 1), 43-64. https://doi.org/10.1093/brain/aws307 PMID:23208308

McQuade, A., Kang, Y. J., Hasselmann, J., Jairaman, A., Sotelo, A., Coburn, M., Shabestari, S. K., Chadarevian, J. P., Fote, G., Tu, C. H., Danhash, E., Silva, J., Martinez, E., Cotman, C., Prieto, G. A., Thompson, L. M., Steffan, J. S., Smith, I., Davtyan, H., Cahalan, M., Cho, H., & Blurton-Jones, M. (2020). Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19227-5

Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S., & Geschwind, D. H. (2013). Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Medicine*, *5*(5), 48. https://doi.org/10.1186/gm452

Mirra, S. S., Heyman, A., McKeel, D., Sumi, S. M., Crain, B. J., Brownlee, L. M., Vogel, F. S., Hughes, J. P., van Belle, G., & Berg, L. (1991). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*, *41*(4), 479-486. https://doi.org/10.1212/wnl.41.4.479 PMID:2011243

Mortimer, J. A., French, L. R., Hutton, J. T., & Schuman, L. M. (1985). Head injury as a risk factor for Alzheimer's disease. *Neurology*, *35*(2), 264-267. https://doi.org/10.1212/wnl.35.2.264 PMID:3969219

Mortland, L., Alonzo, T. A., Walter, R. B., Gerbing, R. B., Mitra, A. K., Pollard, J. A., Loken, M. R., Hirsch, B., Raimondi, S., Franklin, J., Pounds, S., Cao, X., Rubnitz, J. E., Ribeiro, R. C., Gamis, A., Meshinchi, S., & Lamba, J. K. (2013). Clinical significance of CD33 nonsynonymous single-nucleotide polymorphisms in pediatric patients with acute myeloid leukemia treated with gemtuzumab-ozogamicin-containing chemotherapy. *Clin Cancer Res*, *19*(6), 1620-1627. https://doi.org/10.1158/1078-0432.CCR-12-3115 PMID:23444229

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., Gallins, P. J., Buxbaum, J. D., Jarvik, G. P., Crane, P. K., Larson, E. B., Bird, T. D., Boeve, B. F., Graff-Radford, N. R., De Jager, P. L., Evans, D., Schneider, J. A., Carrasquillo, M. M., Ertekin-Taner, N., Younkin, S. G., Cruchaga, C., Kauwe, J. S., Nowotny, P., Kramer, P., Hardy, J., Huentelman, M. J., Myers, A. J., Barmada, M. M., Demirci, F. Y., Baldwin, C. T., Green, R. C., Rogaeva, E., St George-Hyslop, P., Arnold, S. E., Barber, R., Beach, T., Bigio, E. H., Bowen, J. D., Boxer, A., Burke, J. R., Cairns, N. J., Carlson, C. S., Carney, R. M., Carroll, S. L., Chui, H. C., Clark, D. G., Corneveaux, J., Cotman, C. W., Cummings, J. L., . . . Schellenberg, G. D. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*, *43*(5), 436-441. https://doi.org/10.1038/ng.801 PMID:21460841

Nakai, K., & Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci*, *24*(1), 34-36. https://doi.org/10.1016/s0968-0004(98)01336-x PMID:10087920

Nelson, P. T., Braak, H., & Markesbery, W. R. (2009). Neuropathology and Cognitive Impairment in Alzheimer Disease: A Complex but Coherent Relationship. *Journal of Neuropathology & Experimental Neurology*, *68*(1), 1-14. https://doi.org/10.1097/NEN.0b013e3181919a48

Novikova, G., Kapoor, M., Tcw, J., Abud, E. M., Efthymiou, A. G., Chen, S. X., Cheng, H., Fullard, J. F., Bendl, J., Liu, Y., Roussos, P., Björkegren, J. L., Liu, Y., Poon, W. W., Hao, K., Marcora, E., & Goate, A. M. (2021). Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21823-y

Numasawa, Y., Yamaura, C., Ishihara, S., Shintani, S., Yamazaki, M., Tabunoki, H., & Satoh, J. I. (2011). Nasu-Hakola disease with a splicing mutation of TREM2 in a Japanese family. *European Journal of Neurology*, *18*(9), 1179-1183. https://doi.org/10.1111/j.1468-1331.2010.03311.x

Orre, M., Kamphuis, W., Osborn, L. M., Melief, J., Kooijman, L., Huitinga, I., Klooster, J., Bossers, K., & Hol, E. M. (2014). Acute isolation and transcriptome characterization of cortical astrocytes and microglia from young and aged mice. *Neurobiology of Aging*, *35*(1), 1-14. https://doi.org/10.1016/j.neurobiolaging.2013.07.008

Padler-Karavani, V., Hurtado-Ziola, N., Chang, Y. C., Sonnenburg, J. L., Ronagh, A., Yu, H., Verhagen, A., Nizet, V., Chen, X., Varki, N., Varki, A., & Angata,

T. (2014). Rapid evolution of binding specificities and expression patterns of inhibitory CD33-related Siglecs in primates. *The FASEB Journal*, *28*(3), 1280-1293. https://doi.org/10.1096/fj.13-241497

Paloneva, J., Manninen, T., Christman, G., Hovanes, K., Mandelin, J., Adolfsson, R., Bianchin, M., Bird, T., Miranda, R., Salmaggi, A., Tranebjærg, L., Konttinen, Y., & Peltonen, L. (2002). Mutations in Two Genes Encoding Different Subunits of a Receptor Signaling Complex Result in an Identical Disease Phenotype. *The American Journal of Human Genetics*, *71*(3), 656-662. https://doi.org/10.1086/342259

Papageorgiou, I., Loken, M. R., Brodersen, L. E., Gbadamosi, M., Uy, G. L., Meshinchi, S., & Lamba, J. K. (2019). CCGG deletion (rs201074739) in CD33 results in premature termination codon and complete loss of CD33 expression: another key variant with potential impact on response to CD33-directed agents. *Leuk Lymphoma*, 1-4. https://doi.org/10.1080/10428194.2019.1569232 PMID:30721105

Parhizkar, S., Arzberger, T., Brendel, M., Kleinberger, G., Deussing, M., Focke, C., Nuscher, B., Xiong, M., Ghasemigharagoz, A., Katzmarski, N., Krasemann, S., Lichtenthaler, S. F., Muller, S. A., Colombo, A., Monasor, L. S., Tahirovic, S., Herms, J., Willem, M., Pettkus, N., Butovsky, O., Bartenstein, P., Edbauer, D., Rominger, A., Erturk, A., Grathwohl, S. A., Neher, J. J., Holtzman, D. M., Meyer-Luehmann, M., & Haass, C. (2019). Loss of TREM2 function increases amyloid seeding but reduces plaque-associated ApoE. *Nat Neurosci*, *22*(2), 191-204. https://doi.org/10.1038/s41593-018-0296-9 PMID:30617257

Paul, S. P., Taylor, L. S., Stansbury, E. K., & McVicar, D. W. (2000). Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2. *Blood*, *96*(2), 483-490. https://www.ncbi.nlm.nih.gov/pubmed/10887109 PMID:10887109

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Reymond, A., Hubbard, T. J., Harrow, J., & Gerstein, M. B. (2012). The GENCODE pseudogene resource. *Genome Biology*, *13*(9), R51. https://doi.org/10.1186/gb-2012-13-9-r51

Peng, Q., Malhotra, S., Torchia, J. A., Kerr, W. G., Coggeshall, K. M., & Humphrey, M. B. (2010). TREM2- and DAP12-dependent activation of PI3K requires DAP10 and is inhibited by SHIP1. *Sci Signal*, *3*(122), ra38. https://doi.org/10.1126/scisignal.2000500 PMID:20484116

Perez-Oliva, A. B., Martinez-Esparza, M., Vicente-Fernandez, J. J., Corral-San Miguel, R., Garcia-Penarrubia, P., & Hernandez-Caselles, T. (2011). Epitope mapping, expression and post-translational modifications of two isoforms of CD33 (CD33M and CD33m) on lymphoid and myeloid human cells. *Glycobiology*, *21*(6), 757-770. https://doi.org/10.1093/glycob/cwq220 PMID:21278227

Piccio, L., Deming, Y., Del-Águila, J. L., Ghezzi, L., Holtzman, D. M., Fagan, A. M., Fenoglio, C., Galimberti, D., Borroni, B., & Cruchaga, C. (2016). Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and

associated with mutation status. *Acta Neuropathologica*, *131*(6), 925-933. https://doi.org/10.1007/s00401-016-1533-5

Piers, T. M., Cosker, K., Mallach, A., Johnson, G. T., Guerreiro, R., Hardy, J., & Pocock, J. M. (2020). A locked immunometabolic switch underlies TREM2 R47H loss of function in human iPSC-derived microglia. *The FASEB Journal*, *34*(2), 2436-2450. https://doi.org/10.1096/fj.201902447r

Prada, I., Ongania, G. N., Buonsanti, C., Panina-Bordignon, P., & Meldolesi, J. (2006). Triggering receptor expressed in myeloid cells 2 (TREM2) trafficking in microglial cells: Continuous shuttling to and from the plasma membrane regulated by cell stimulation. *Neuroscience*, *140*(4), 1139-1148. https://doi.org/https://doi.org/10.1016/j.neuroscience.2006.03.058

Raj, T., Ryan, K. J., Replogle, J. M., Chibnik, L. B., Rosenkrantz, L., Tang, A., Rothamel, K., Stranger, B. E., Bennett, D. A., Evans, D. A., De Jager, P. L., & Bradshaw, E. M. (2014). CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Human Molecular Genetics*, *23*(10), 2729-2736. https://doi.org/10.1093/hmg/ddt666

Rajan, K. B., Weuve, J., Barnes, L. L., McAninch, E. A., Wilson, R. S., & Evans, D. A. (2021). Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). *Alzheimer's & Dementia*, *n/a*(n/a). https://doi.org/https://doi.org/10.1002/alz.12362

Rantanen, T. (2013). Midlife fitness predicts less burden of chronic disease in later life. *Clin J Sport Med*, *23*(6), 499-500. https://doi.org/10.1097/JSM.0000000000000039 PMID:24169299

Ravetch, J. V. (2000). Immune Inhibitory Receptors. *Science*, *290*(5489), 84-89. https://doi.org/10.1126/science.290.5489.84

Renaud, J.-B., Boix, C., Charpentier, M., Anne, Cochennec, J., Duvernois-Berthet, E., Perrouault, L., Tesson, L., Edouard, J., Thinard, R., Cherifi, Y., Menoret, S., Fontanière, S., Noémie, Fraichard, A., Sohm, F., Anegon, I., Concordet, J.-P., & Giovannangeli, C. (2016). Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Reports*, *14*(9), 2263-2272. https://doi.org/10.1016/j.celrep.2016.02.018

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*, *16*(6), 276-277. https://doi.org/10.1016/s0168-9525(00)02024-2 PMID:10827456

Rodrigues, E., Jung, J., Park, H., Loo, C., Soukhtehzari, S., Kitova, E. N., Mozaneh, F., Daskhan, G., Schmidt, E. N., Aghanya, V., Sarkar, S., Streith, L., St. Laurent, C. D., Nguyen, L., Julien, J.-P., West, L. J., Williams, K. C., Klassen, J. S., & Macauley, M. S. (2020). A versatile soluble siglec scaffold for sensitive and quantitative detection of glycan ligands. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18907-6

Salminen, A., & Kaarniranta, K. (2009). Siglec receptors and hiding plaques in Alzheimer's disease. *Journal of Molecular Medicine*, *87*(7), 697-701. https://doi.org/10.1007/s00109-009-0472-1

Scarmeas, N., Stern, Y., Tang, M. X., Mayeux, R., & Luchsinger, J. A. (2006). Mediterranean diet and risk for Alzheimer's disease. *Ann Neurol*, *59*(6), 912-921. https://doi.org/10.1002/ana.20854 PMID:16622828

Schlepckow, K., Kleinberger, G., Fukumori, A., Feederle, R., Lichtenthaler, S. F., Steiner, H., & Haass, C. (2017). An Alzheimer-associated TREM2 variant occurs at the ADAM cleavage site and affects shedding and phagocytic function. *EMBO Molecular Medicine*, *9*(10), 1356-1365. https://doi.org/https://doi.org/10.15252/emmm.201707672

Schwarz, F., Springer, S. A., Altheide, T. K., Varki, N. M., Gagneux, P., & Varki, A. (2016). Human-specific derived alleles of CD33 and other genes protect against postreproductive cognitive decline. *Proc Natl Acad Sci U S A*, *113*(1), 74-79. https://doi.org/10.1073/pnas.1517951112 PMID:26621708

Sessa, G., Podini, P., Mariani, M., Meroni, A., Spreafico, R., Sinigaglia, F., Colonna, M., Panina, P., & Meldolesi, J. (2004). Distribution and signaling of TREM2/DAP12, the receptor system mutated in human polycystic lipomembraneous osteodysplasia with sclerosing leukoencephalopathy dementia. *European Journal of Neuroscience*, *20*(10), 2617-2628. https://doi.org/10.1111/j.1460-9568.2004.03729.x

Shaw, B. C., & Estus, S. (2021). Pseudogene-Mediated Gene Conversion After CRISPR-Cas9 Editing Demonstrated by Partial CD33 Conversion with SIGLEC22P. *Crispr j*, *4*(5), 699-709. https://doi.org/10.1089/crispr.2021.0052 PMID:34558988

Shaw, B. C., Katsumata, Y., Simpson, J. F., Fardo, D. W., & Estus, S. (2021). Analysis of Genetic Variants Associated with Levels of Immune Modulating Proteins for Impact on Alzheimer's Disease Risk Reveal a Potential Role for SIGLEC14. *Genes (Basel)*, *12*(7), 1008. https://doi.org/10.3390/genes12071008 PMID:34208838

Shaw, B. C., Snider, H. C., Turner, A. K., Zajac, D. J., Simpson, J. F., & Estus, S. (2021). A novel TREM2 splice isoform lacking the ligand binding is expressed in brain and shares localization. *bioRxiv*, 2021.2011.2023.469712. https://doi.org/10.1101/2021.11.23.469712

Shi, Y., & Holtzman, D. M. (2018). Interplay between innate immunity and Alzheimer disease: APOE and TREM2 in the spotlight. *Nat Rev Immunol*, *18*(12), 759-772. https://doi.org/10.1038/s41577-018-0051-1 PMID:30140051

Shipley, J. M., Sheppard, D. M., & Sheer, D. (1988). Karyotypic analysis of the human monoblastic cell line U937. *Cancer Genetics and Cytogenetics*, *30*(2), 277-284. https://doi.org/10.1016/0165-4608(88)90195-1

Siddiqui, S. S., Springer, S. A., Verhagen, A., Sundaramurthy, V., Alisson-Silva, F., Jiang, W., Ghosh, P., & Varki, A. (2017). The Alzheimer's disease-protective CD33 splice variant mediates adaptive loss of function via diversion to an intracellular pool. *J Biol Chem*, *292*(37), 15312-15320. https://doi.org/10.1074/jbc.M117.799346 PMID:28747436

Sierksma, A., Lu, A., Mancuso, R., Fattorelli, N., Thrupp, N., Salta, E., Zoco, J., Blum, D., Buée, L., De Strooper, B., & Fiers, M. (2020). Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU

pathology. *EMBO Molecular Medicine*, *12*(3), e10606. https://doi.org/https://doi.org/10.15252/emmm.201910606

Sisu, C., Pei, B., Leng, J., Frankish, A., Zhang, Y., Balasubramanian, S., Harte, R., Wang, D., Rutenberg-Schoenberg, M., Clark, W., Diekhans, M., Rozowsky, J., Hubbard, T., Harrow, J., & Gerstein, M. B. (2014). Comparative analysis of pseudogenes across three phyla. *Proceedings of the National Academy of Sciences*, *111*(37), 13361-13366. https://doi.org/10.1073/pnas.1407293111

Slightom, J. L., Blechl, A. E., & Smithies, O. (1980). Human fetal G gamma- and A gamma-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*, *21*(3), 627-638. https://doi.org/10.1016/0092-8674(80)90426-2 PMID:7438203

Smith, V. D., Bachstetter, A. D., Ighodaro, E., Roberts, K., Abner, E. L., Fardo, D. W., & Nelson, P. T. (2018). Overlapping but distinct TDP-43 and tau pathologic patterns in aged hippocampi. *Brain Pathol*, *28*(2), 264-273. https://doi.org/10.1111/bpa.12505 PMID:28281308

Soininen, H., & Heinonen, O. P. (1982). Clinical and etiological aspects of senile dementia. *Eur Neurol*, *21*(6), 401-410. https://doi.org/10.1159/000115514 PMID:7173282

Sojkova, J., Driscoll, I., Iacono, D., Zhou, Y., Codispoti, K.-E., Kraut, M. A., Ferrucci, L., Pletnikova, O., Mathis, C. A., Klunk, W. E., O'Brien, R. J., Wong, D. F., Troncoso, J. C., & Resnick, S. M. (2011). In Vivo Fibrillar β-Amyloid Detected Using [11C]PiB Positron Emission Tomography and Neuropathologic Assessment in Older Adults. *Archives of Neurology*, *68*(2). https://doi.org/10.1001/archneurol.2010.357

Solas, M., Aisa, B., Tordera, R. M., Mugueta, M. C., & Ramirez, M. J. (2013). Stress contributes to the development of central insulin resistance during aging: implications for Alzheimer's disease. *Biochim Biophys Acta*, *1832*(12), 2332-2339. https://doi.org/10.1016/j.bbadis.2013.09.013 PMID:24090692

Song, W. M., Joshita, S., Zhou, Y., Ulland, T. K., Gilfillan, S., & Colonna, M. (2018). Humanized TREM2 mice reveal microglia-intrinsic and -extrinsic effects of R47H polymorphism. *Journal of Experimental Medicine*, *215*(3), 745-760. https://doi.org/10.1084/jem.20171529

Srinivasan, K., Friedman, B. A., Larson, J. L., Lauffer, B. E., Goldstein, L. D., Appling, L. L., Borneo, J., Poon, C., Ho, T., Cai, F., Steiner, P., van der Brug, M. P., Modrusan, Z., Kaminker, J. S., & Hansen, D. V. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. *Nature Communications*, *7*(1), 11295. https://doi.org/10.1038/ncomms11295

Strassnig, M., & Ganguli, M. (2005). About a peculiar disease of the cerebral cortex: Alzheimer's original case revisited. *Psychiatry (Edgmont (Pa. : Township))*, *2*(9), 30-33. PMID:21120106

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., Oliver-Williams, C., Kamat, M. A., Prins, B. P., Wilcox, S. K., Zimmerman, E. S., Chi, A., Bansal, N.,

Spain, S. L., Wood, A. M., Morrell, N. W., Bradley, J. R., Janjic, N., Roberts, D. J., Ouwehand, W. H., Todd, J. A., Soranzo, N., Suhre, K., Paul, D. S., Fox, C. S., Plenge, R. M., Danesh, J., Runz, H., & Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature*, *558*(7708), 73-79. https://doi.org/10.1038/s41586-018-0175-2 PMID:29875488

Susani, L., Castelli, A., Lizier, M., Lucchini, F., Vezzoni, P., & Paulis, M. (2018). Correction of a Recessive Genetic Defect by CRISPR-Cas9-Mediated Endogenous Repair. *Crispr j*, *1*, 230-238. https://doi.org/10.1089/crispr.2018.0004 PMID:31021260

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467-484. https://doi.org/10.1038/s41576-019-0127-1

The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57-74. https://doi.org/10.1038/nature11247

Thornton, P., Sevalle, J., Deery, M. J., Fraser, G., Zhou, Y., Ståhl, S., Franssen, E. H., Dodd, R. B., Qamar, S., Gomez Perez-Nievas, B., Nicol, L. S., Eketjäll, S., Revell, J., Jones, C., Billinton, A., St George-Hyslop, P. H., Chessell, I., & Crowther, D. C. (2017). TREM2 shedding by cleavage at the H157-S158 bond is accelerated for the Alzheimer's disease-associated H157Y variant. *EMBO Molecular Medicine*, *9*(10), 1366-1378. https://doi.org/https://doi.org/10.15252/emmm.201707673

Tolppanen, A. M., Ngandu, T., Kareholt, I., Laatikainen, T., Rusanen, M., Soininen, H., & Kivipelto, M. (2014). Midlife and late-life body mass index and late-life dementia: results from a prospective population-based cohort. *J Alzheimers Dis*, *38*(1), 201-209. https://doi.org/10.3233/JAD-130698 PMID:23948937

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., Long, N., Schultz, A. R., Traer, E., Abel, M., Agarwal, A., Blucher, A., Borate, U., Bryant, J., Burke, R., Carlos, A., Carpenter, R., Carroll, J., Chang, B. H., Coblentz, C., D'Almeida, A., Cook, R., Danilov, A., Dao, K.-H. T., Degnin, M., Devine, D., Dibb, J., Edwards, D. K., Eide, C. A., English, I., Glover, J., Henson, R., Ho, H., Jemal, A., Johnson, K., Johnson, R., Junio, B., Kaempf, A., Leonard, J., Lin, C., Liu, S. Q., Lo, P., Loriaux, M. M., Luty, S., Macey, T., Macmaniman, J., Martinez, J., Mori, M., Nelson, D., . . . Druker, B. J. (2018). Functional genomic landscape of acute myeloid leukaemia. *Nature*, *562*(7728), 526-531. https://doi.org/10.1038/s41586-018-0623-z

Varki, A. (2011). Since there are PAMPs and DAMPs, there must be SAMPs? Glycan "self-associated molecular patterns" dampen innate immunity, but pathogens can mimic them. *Glycobiology*, *21*(9), 1121-1124. https://www.ncbi.nlm.nih.gov/pubmed/21932452 PMID:21932452

Walker, D. G., Whetzel, A. M., Serrano, G., Sue, L. I., Beach, T. G., & Lue, L. F. (2015). Association of CD33 polymorphism rs3865444 with Alzheimer's disease pathology and CD33 expression in human cerebral cortex. *Neurobiol Aging*, *36*(2), 571-582. https://doi.org/10.1016/j.neurobiolaging.2014.09.023 PMID:25448602

Walter, R. B., Raden, B. W., Zeng, R., Hausermann, P., Bernstein, I. D., & Cooper, J. A. (2008). ITIM-dependent endocytosis of CD33-related Siglecs: role of intracellular domain, tyrosine phosphorylation, and the tyrosine phosphatases, Shp1 and Shp2. *J Leukoc Biol*, *83*(1), 200-211. https://doi.org/10.1189/jlb.0607388 PMID:17947393

Wang, Y., Cella, M., Mallinson, K., Jason, Katherine, Michelle, Gilfillan, S., Gokul, Sudhakar, S., Bernd, David, John, & Colonna, M. (2015). TREM2 Lipid Sensing Sustains the Microglial Response in an Alzheimer's Disease Model. *Cell*, *160*(6), 1061-1071. https://doi.org/10.1016/j.cell.2015.01.049

Wes, P. D., Easton, A., Corradi, J., Barten, D. M., Devidze, N., Decarr, L. B., Truong, A., He, A., Barrezueta, N. X., Polson, C., Bourin, C., Flynn, M. E., Keenan, S., Lidge, R., Meredith, J., Natale, J., Sankaranarayanan, S., Cadelina, G. W., Albright, C. F., & Cacace, A. M. (2014). Tau Overexpression Impacts a Neuroinflammation Gene Expression Network Perturbed in Alzheimer's Disease. *PLoS One*, *9*(8), e106050. https://doi.org/10.1371/journal.pone.0106050

Wightman, D. P., Jansen, I. E., Savage, J. E., Shadrin, A. A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B. S., Drange, O. K., Martinsen, A. E., Skogholt, A. H., Willer, C., Bråthen, G., Bosnes, I., Nielsen, J. B., Fritsche, L. G., Thomas, L. F., Pedersen, L. M., Gabrielsen, M. E., Johnsen, M. B., Meisingset, T. W., Zhou, W., Proitsi, P., Hodges, A., Dobson, R., Velayudhan, L., Agee, M., Aslibekyan, S., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Cameron, B., Coker, D., Partida, G. C., Dhamija, D., Das, S., Elson, S. L., Filshtein, T., Fletez-Brant, K., Fontanillas, P., Freyman, W., Gandhi, P. M., Hicks, B., Hinds, D. A., Huber, K. E., Jewett, E. M., . . . Posthuma, D. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nature Genetics*, *53*(9), 1276-1282. https://doi.org/10.1038/s41588-021-00921-z

Wißfeld, J., Mathews, M., Mossad, O., Picardi, P., Cinti, A., Redaelli, L., Pradier, L., Brüstle, O., & Neumann, H. (2021). Reporter cell assay for human CD33 validated by specific antibodies and human iPSC-derived microglia. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-92434-2

Wißfeld, J., Nozaki, I., Mathews, M., Raschka, T., Ebeling, C., Hornung, V., Brustle, O., & Neumann, H. (2021). Deletion of Alzheimer's disease-associated CD33 results in an inflammatory human microglia phenotype. *Glia*. https://doi.org/10.1002/glia.23968 PMID:33539598

Yamanaka, M., Kato, Y., Angata, T., & Narimatsu, H. (2009). Deletion polymorphism of SIGLEC14 and its functional implications. *Glycobiology*, *19*(8), 841-846. https://doi.org/10.1093/glycob/cwp052 PMID:19369701

Yanagisawa, K. (2007). Role of gangliosides in Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1768*(8), 1943-1951. https://doi.org/10.1016/j.bbamem.2007.01.018

Yanaizu, M., Sakai, K., Tosaki, Y., Kino, Y., & Satoh, J.-i. (2018). Small nuclear RNA-mediated modulation of splicing reveals a therapeutic strategy for a

TREM2 mutation and its post-transcriptional regulation. *Scientific Reports*, *8*(1), 6937. https://doi.org/10.1038/s41598-018-25204-2

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Alexei, Zhang, C., Xie, T., Tran, L., Dobrin, R., Fluder, E., Clurman, B., Melquist, S., Narayanan, M., Suver, C., Shah, H., Mahajan, M., Gillis, T., Mysore, J., Marcy, John, David, Molony, C., David, Gudnason, V., Amanda, Eric, Neumann, H., Zhu, J., & Emilsson, V. (2013). Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. *Cell*, *153*(3), 707-720. https://doi.org/10.1016/j.cell.2013.03.030

Zhang, Y.-W., Thompson, R., Zhang, H., & Xu, H. (2011). APP processing in Alzheimer's disease. *Molecular Brain*, *4*(1), 3. https://doi.org/10.1186/1756-6606-4-3

Zhang, Y., Sloan, Steven A., Clarke, Laura E., Caneda, C., Plaza, Colton A., Blumenthal, Paul D., Vogel, H., Steinberg, Gary K., Edwards, Michael S. B., Li, G., Duncan, John A., III, Cheshier, Samuel H., Shuer, Lawrence M., Chang, Edward F., Grant, Gerald A., Gephart, Melanie G. H., & Barres, Ben A. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron*, *89*(1), 37-53. https://doi.org/10.1016/j.neuron.2015.11.013

Zhang, Z., Carriero, N., Zheng, D., Karro, J., Harrison, P. M., & Gerstein, M. (2006). PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics*, *22*(12), 1437-1439. https://doi.org/10.1093/bioinformatics/btl116 PMID:16574694

Zhao, L. (2018). CD33 in Alzheimer's Disease - Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology*, 1-9. https://doi.org/10.1159/000492596 PMID:30541012

Zhao, Y., Wu, X., Li, X., Jiang, L.-L., Gui, X., Liu, Y., Sun, Y., Zhu, B., Piña-Crespo, J. C., Zhang, M., Zhang, N., Chen, X., Bu, G., An, Z., Huang, T. Y., & Xu, H. (2018). TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function. *Neuron*, *97*(5), 1023-1031.e1027. https://doi.org/10.1016/j.neuron.2018.01.031

Zhong, L., Chen, X.-F., Zhang, Z.-L., Wang, Z., Shi, X.-Z., Xu, K., Zhang, Y.-W., Xu, H., & Bu, G. (2015). DAP12 Stabilizes the C-terminal Fragment of the Triggering Receptor Expressed on Myeloid Cells-2 (TREM2) and Protects against LPS-induced Pro-inflammatory Response. *Journal of Biological Chemistry*, *290*(25), 15866-15877. https://doi.org/10.1074/jbc.m115.645986

Zhong, L., Xu, Y., Zhuo, R., Wang, T., Wang, K., Huang, R., Wang, D., Gao, Y., Zhu, Y., Sheng, X., Chen, K., Wang, N., Zhu, L., Can, D., Marten, Y., Shinohara, M., Liu, C.-C., Du, D., Sun, H., Wen, L., Xu, H., Bu, G., & Chen, X.-F. (2019). Soluble TREM2 ameliorates pathological phenotypes by modulating microglial functions in an Alzheimer's disease model. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09118-9

Zhou, Y., Song, W. M., Andhey, P. S., Swain, A., Levy, T., Miller, K. R., Poliani, P. L., Cominelli, M., Grover, S., Gilfillan, S., Cella, M., Ulland, T. K., Zaitsev, K., Miyashita, A., Ikeuchi, T., Sainouchi, M., Kakita, A., Bennett, D. A.,

Schneider, J. A., Nichols, M. R., Beausoleil, S. A., Ulrich, J. D., Holtzman, D. M., Artyomov, M. N., & Colonna, M. (2020). Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nature Medicine*, *26*(1), 131-142. https://doi.org/10.1038/s41591-019-0695-9

Zou, F., Gopalraj, R. K., Lok, J., Zhu, H., Ling, I. F., Simpson, J. F., Tucker, H. M., Kelly, J. F., Younkin, S. G., Dickson, D. W., Petersen, R. C., Graff-Radford, N. R., Bennett, D. A., Crook, J. E., Younkin, S. G., & Estus, S. (2007). Sex-dependent association of a common low-density lipoprotein receptor polymorphism with RNA splicing efficiency in the brain and Alzheimer's disease. *Human Molecular Genetics*, *17*(7), 929-935. https://doi.org/10.1093/hmg/ddm365

VITA

Education
- *Eastern Kentucky University*   2014 – 2016
  M.S. Chemistry
- *Eastern Kentucky University*   2013 – 2015
  B.S. Chemistry

Research Experience
- *PhD Dissertation Project*
  University of Kentucky   October 2018 – 2022
  Department of Physiology
  Mentor: Steven Estus, PhD
  Dissertation Title: *Immunoregulatory receptor genetics, expression, and splicing studies in Alzheimer's Disease*
- *PhD Laboratory Rotation*
  University of Kentucky   October 2016 – December 2016
  Department of Microbiology, Immunology, and Molecular Genetics
  Mentor: Beth Garvy, PhD
  Central Nervous System Lymphocyte Trafficking in Experimental Autoimmune Encephalomyelitis
- *Previous PhD Laboratory*
  University of Kentucky   May 2016 – October 2018
  Department of Physiology
  Mentor: Bradley K. Taylor, PhD
  Sphingosine-1-Phosphate Receptor 1 Modulation of Central Neuropathic Pain in Experimental Autoimmune Encephalomyelitis
- *Masters Thesis Project*
  Eastern Kentucky University   August 2014 – May 2016
  Department of Chemistry
  Mentor: Jamie Fredericks, PhD
  A Novel Forensic DNA Profiling Method based on Molecular Beacons Without DNA Purification

Publications and Preprints
- Shaw, B.C, Snider, H.S., Turner, A.K., Zajac, D.J., Simpson, J.F., Estus, S. A novel TREM2 splice isoform lacking the ligand binding is expressed in brain and shares localization. Preprint. *bioRxiv*. 2021 Nov.
- Shaw, B.C., Estus, S. Pseudogene-mediated gene conversion after CRISPR-Cas9 editing demonstrated by partial CD33 conversion with SIGLEC22P. *CRISPR J*. 2021 Sept 23;4(5).

- Shaw, B.C., Maglinger, G. B., Ujas, T., Rupareliya, C., Fraser, J.F., Grupke, S., Kesler, M., Gelderblom, M., Pennypacker, K.R., Turchan-Cholewo, J., Stowe, A.M. Isolation and Identification of Leukocyte Populations in Intracranial Blood Collected During Mechanical Thrombectomy. *J Cereb. Blood Flow Metab*. 2021 July 11.
- Shaw, B.C., Katsumata, Y., Simpson, J.F., Fardo, D.W., Estus, S. Analysis of Genetic Variants Associated with Levels of Immune Modulating Proteins for Impact on Alzheimer's Disease Risk Reveal a Potential Role for SIGLEC14. *Genes*. 2021 Jun 30;12(7):1008.
- Katsumata Y, Fardo DW, Bachstetter AD, Artiushin SC, Wang WX, Wei A, Brzezinski LJ, Nelson BG, Huang Q, Abner EL, Anderson S, Patel I, Shaw, B.C., Price DA, Niedowicz DM, Wilcock DW, Jicha GA, Neltner JH, Van Eldik LJ, Estus S, Nelson PT. Alzheimer Disease Pathology-Associated Polymorphism in a Complex Variable Number of Tandem Repeat Region Within the MUC6 Gene, Near the AP2A2 Gene. *J Neuropathol Exp Neurol*. 2020 Jan 1;79(1):3-21.
- Estus, S., Shaw, B.C., Devanney, N., Katsumata, Y., Press, E., Fardo, F. Evaluation of CD33 as a Genetic Risk Factor for Alzheimer's Disease. *Acta Neuropathol*. 2019 Aug;138(2):187-199.
- Doolen, S.D., Iannitti, T., Donahue, R.R., Shaw, B.C., Taylor, B.K. Fingolimod Reduces Neuropathic Pain Behaviors in a Mouse Model of Multiple Sclerosis by a Sphingosine-1 Phosphate Receptor 1-Dependent Inhibition of Central Sensitization in the Dorsal Horn*. Pain*. 2018 Feb;159(2):224-38.
- Shaw, B.C., Reekers, C, Fenske, W, Fugate, D, Brock, M, Fredericks, J. Rapid Identification of Fermentation Spoilage Microbes Using Molecular Beacons and a Two-Step Direct DNA Amplification Protocol. *JASBC*. 2020 Feb;78(3):195-201.
- Shaw, B.C. A novel forensic DNA profiling method based on molecular beacons without DNA purification. M.S. thesis, Eastern Kentucky University 2016

National Presentations
- Shaw, B.C., Snider, H., Zajac, D., Simpson, J.F., Estus, S. *Novel TREM2 isoforms highly expressed in human brain and their potential impact on Alzheimer's disease*. Society for Neuroscience Annual Meeting, 2021. Poster presentation, 11/2021. Chicago, IL.
- Shaw, B.C., Devanney, N., Press, E., Katsumata, Y., Fardo, D.W., Estus, S. CD33 *Molecular Genetics in Alzheimer's Disease Risk*. Alzheimer's Association International Conference 2019. Poster presentation, 07/2019. Los Angeles, CA.
- Shaw, B.C., Doolen, S.D., Iannitti, T., Donahue, R.R., Taylor, B.K. *Fingolimod behaves as an agonist at sphingosine-1-phosphate receptor 1 to inhibit pain in an experimental autoimmune encephalomyelitis model of multiple sclerosis.* American Pain Society, Poster presentation, May 2017, Pittsburgh, PA.

Funding History
- F99NS120365   Benjamin Shaw, PI   7/1/2021 – 6/30/2023
  NINDS Diversity Specialized Predoctoral to Postdoctoral Advancement in Neuroscience (D-SPAN)
  This fellowship provides two years of predoctoral support for dissertation research and training. During this time, the award will support my technical, scientific, and professional development through additional workshops and conference opportunities. This fellowship also includes an additional four years of career development support as a postdoctoral scholar through a K00, which will be awarded upon completion of my doctoral program and non-competing review of the career award application.
- RF1AG059717-01S1   Steven Estus, PI   3/15/2019 – 6/30/2021
  NIA Research Supplements to Promote Diversity in Health-Related Research
  This grant provides support for two years of dissertation research and training. During this time, the award has supported training me in general molecular biology and proteomics studies and developed critical thinking skills through frequent discussions with my mentor and conference attendance.
- T32GM118292-01   Bret Smith, PI   7/1/2017 – 6/30/2022
  NIGMS Graduate training in integrative physiology
  This grant provides support for graduate students during their first year in the department. During this time, the students undertake nine hours of didactic lecture in integrative physiology as well as ethics training, a quantitative methods course, a clinical shadowing experience and sit for their comprehensive and qualifying exams.

Professional Development
- Functional Genetics Boot Camp, Columbia SHARP program, July 2020
- McKusick Short Course in Human and Mammalian Genetics and Genomics, Jackson Laboratories, July 2020

Awards, Honors, and Scholarships
- Outstanding Poster Award, Markesbery Symposium on Aging at University of Kentucky (2021)
- 2nd Place, Poster Competition, 17th Annual World Congress on Pain (2018)
- 3rd Place, Graduate Oral Presentation Competition, Kentucky Academy of Science (2015)
- 3rd Place, Undergraduate Poster Competition, Kentucky Academy of Science (2013)
- Navy Achievement Medal—for exceptional leadership (2011)