Investigating the Genetic Factors influencing Inflammation, Pain and Depression in Parkinson's Disease

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

Hannah Jane Hendry

School of Medicine
Cardiff University

September 2022

Thesis Summary

Parkinson’s disease (PD) is a neurodegenerative movement disorder caused by progressive loss of dopaminergic neurons. Beyond motor symptoms, patients experience a host of non-motor symptoms. Two common symptoms are pain and depression, which have been correlated with each other. PD is also characterised by increased activation of the immune system within the periphery and CNS. A genetic risk factor for PD is at the Human Leukocyte Antigen (HLA) locus, a highly polymorphic region encoding proteins that control the adaptive immune response. This project aimed to develop the understanding of genetic factors influencing inflammation, pain, and depression in PD.

Firstly, a range of bioinformatics techniques including HLA imputation were applied to a PD dataset to determine the HLA loci most associated with PD risk and protection. This resulted in identification of HLA-B, HLA-C, HLA-DRB1, and HLA-DQA1 as loci to further analyse. The Pacific-Biosciences long-read sequencing method was applied to these loci from PD samples. Results from sequencing data indicated the HLA alleles associated with PD protection (HLA-DRB1*04) and risk (HLA-DQA1*01). These results were compared to HLA imputation in a large case-control dataset, which corroborated the top associated alleles.

Secondly, an investigation into the relationship between pain and depression in PD was conducted. Two GWAS of depression in PD were conducted in the UKBB and Proband cohorts, and a GWAS of multisite chronic pain (MCP) in PD was conducted in the UKBB cohort. The results indicated putative genetic associations with these symptoms. Polygenic risk score (PRS) analysis showed no evidence for correlation of genetic influences with MDD, but did for MCP. A Mendelian randomisation analysis was performed, finding no evidence for a causative relationship between these symptoms; this suggests independent causative factors.

Overall, novel data to identify potential genetic influences of these PD characteristics was collected, which can help direct future investigations.

Acknowledgements

Firstly, I would like to say a huge thank you to my main supervisor Nigel Williams. It has definitely been an interesting period whilst working on this thesis, and I have been incredibly lucky to have such a warm and encouraging supervisor throughout the process. Thank you for the kindness and patience you have shown me; it has been a pleasure to learn so much from you. I would also like to thank my second supervisor Simon Jones for your advice and insight during the project, particularly in the development stage. Thank you to the Hodge Foundation for funding this PhD. We would also like to thank Carwyn Griffiths and his family, whose generous donations supported the research presented in this PhD thesis.

I would like to thank everyone in the MRC Centre who has helped me with this work, especially Ellis, Mark, Alex, Ying, Lucinda, Patrick and everyone in the core team – your advice and training has been greatly appreciated. I would particularly like to thank Jo for your help with the PacBio work, and Leon for all your help and advice with the bioinformatics analysis. Thank you also to everyone in the Jones group who were so kind and welcoming, especially Aisling, Ana and Alicia for your training.

Thank you to everyone at the PGR office for helping us through the pandemic and ensuring PGR students were well supported, especially Mandy who has done so much for us all.

Finally, thank you so much to my family for all your support over the past few years. I am so grateful to my parents for your confidence, encouragement, and love throughout all my education, I promise I will get a job now. Thank you to my sister Rachel for all the dinners and emotional support and lockdown walks, it has been a joy to spend this time with you in Cardiff. Thank you Rhidian, for turning up at the right time and being the best friend and support bubble I could ask for. I hope you appreciate all the useful knowledge you have gained about the HLA.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APC	antigen presenting cell
α-syn	alpha-synuclein
ß	beta
BBB	blood brain barrier
BD	broad depression
BP	base position
BPI	brief pain inventory
BRCA1	Breast cancer Gene 1
BRIP1	BRCA1 Interacting Protein 1
CD	cluster of differentiation
CHR	chromosome
CNS	central nervous system
DF	degrees of freedom
DISH	Direct imputing summary association statistics HLA variants
DNA	deoxyribonucleic acid
DRG	dorsal root ganglion
eQTL	expression quantitative trait locus
GM-CSF	granulocyte macrophage colony stimulating factor
GWAS	genome wide association study
hg19	human genome build 19
HIBAG	HLA Imputation using attribute BAGging
HLA	human leukocyte antigen
HWE	Hardy-Wienberg equilibrium
IFN-γ	interferon gamma
IV	instrumental variable
IVW	inverse variance weighted
KPPS	Kings Parkinson's Pain Scale
LADS	Leeds Anxiety Depression Scale
LC	locus corealus

LD	linkage disequilibrium
MAF	minor allele frequency
MCP	multisite chronic pain
MDD	major depressive disorder
MHC	major histocompatibility complex
mQTL	methylation quantitative trait locus
MR	Mendelian randomisation
NSAID	non-steroidal anti-inflammatory drugs
OR	odds ratio
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PD	Parkinson’s disease
PE	preeclampsia
RNA	ribonucleic acid
SD	standard deviation
SE	standard error
SFMPQ	Short Form McGill Pain Questionnaire
SMRT	Single Molecule, Real-Time
SN	substantia nigra
SNP	single nucleotide polymorphism
SSRI	selective serotonin reuptake inhibitor
STR	short tandem repeat
TCR	T cell receptor
TNF-α	tumour necrosis factor alpha
TLR	Toll-like receptor
TSMR	two sample Mendelian randomisation
UKBB	UK BioBank
UPDRS	Unified Parkinson's disease Rating Scale
VAS	Visual Analogue Scale

ZF	zinc finger
ZMW	zero mode waveguide

1 Introduction

1.1 Parkinson's disease

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting over 6 million individuals globally (1). PD is primarily a movement disorder characterised by tremor, rigidity of movement, bradykinesia, and dyskinesia. In 2015 an investigation into the global burden of disease found its incidence to be the fastest growing neurological disease (1), and the number of PD cases is anticipated to exceed 12 million by 2040(2). PD predominantly affects those over 60 years of age, with 1% of this age group suffering from PD and 5% of those over 85(3). The impact of PD extends beyond loss of motor control, with patients experiencing a range of physical and psychological symptoms including sleep disturbance, mood disorders, cognitive decline, systemic inflammation, digestive issues, and chronic pain (4).

Pathological hallmarks of PD include accumulation of aggregated alpha-synuclein (α-syn) in the midbrain, leading to progressive degeneration of dopaminergic neurons in the substantia nigra, which results in symptomatic loss of motor function(5). Braak staging indicates that neurodegeneration begins in medulla oblongata and spreads through the subcortical areas to eventually include extensive cortical regions. This contributes to development of non-motor symptoms including cognitive decline, sleep alterations and pain(5).

There are limited treatments for PD; the dopamine precursor Levadopa is the current gold standard treatment to address the loss of dopamine production, and other avenues such as deep brain stimulation are considered in severe treatment resistant cases (6). Currently there is no way to prevent or reverse neurodegeneration, so new therapeutic targets to treat and alleviate symptoms are an important area of investigation.

1.2 Non-motor aspects of PD

Whilst motor symptoms are characteristic of PD patients, the host of other non-motor PD symptoms can be experienced to varying degrees. Often these symptoms can be experienced as prodromal PD, with onset occurring years before characteristic motor symptoms (7). Whilst the experience of these symptoms differs across patients, they can have a significant impact on quality of life. It is questionable how much traditional PD therapies which focus on dopamine replacement can impact these non-motor symptoms. To have a greater impact on the quality of life of PD patients, an improved understanding of the underlying causes of these non-motor symptoms is required so that they may be targeted appropriately.

1.2.1 Pain in PD

1.2.1.1 Prevalence of pain in PD

Pain is one of the most common non-motor symptoms in PD. Different investigations have attempted to determine the exact prevalence and severity of pain in PD. A recent investigation assessing 1957 PD patients found 85% of PD patients reported pain, with 42% reporting moderate to severe pain (8). Pain was assessed with three different tests: Short Form McGill Pain Questionnaire (SFMPQ), Visual Analogue Scale (VAS) for pain severity over the last month and the Kings Parkinson's Pain Scale (KPPS). Experience of overall pain was not predicted by disease duration or motor impairment, but female gender and younger age were two predictors. Importantly, a multiple regression model found that pain influenced quality of life more significantly than motor impairment (8).

A different investigation into the prevalence of pain in 176 PD patients in Norway found that 146 (83%) reported experiencing pain, with musculoskeletal pain reported by 70%. This investigation used three separate measures of pain: Brief Pain Inventory (BPI), the Bodily Pain (BP) Scale of the SF-36, and a clinical examination. It was also found that the experience of pain was not associated with disease duration or severity, with female gender the only predictor (9).

The DoPaMiP survey similarly set out to characterise pain in PD in French patients, but with a focus on chronic pain (10). Out of 450 PD patients assessed, 278 (62%) experienced chronic pain. This was defined as pain lasting more than 3 months. 26% of patients reported having pain unrelated to PD (mainly caused by osteoarthritis) whilst 39.3% had chronic pain exclusively associated with PD. Of those with PD related pain, this was associated with several factors such as younger age of onset and more severe depressive symptoms than those without PD related pain, but not with disease duration or severity.

These results consistently show that the level of pain experienced in PD patients is greater than the general population. It is not only important to consider the prevalence of pain but what impact this has on sufferers. As mentioned above, in one investigation pain was found to negatively influence quality of life more than motor impairment (8). To better understand what the greatest burden is to PD sufferers, a separate investigation into the patient's perspective asked a cohort of 265 PD patients to report their three worst symptoms (4). In early PD patients (less than six years of disease), pain was ranked as the fourth overall worst symptom behind slowness, tremor and stiffness. Whilst over 60% of these patients reported slowness or tremor as the worst aspect of their disease, 10% reported pain as their worst feature and 10% reported it as their second. In patients with later stage PD, pain was the sixth ranked most troublesome feature with mood disorders and fluctuating responses to medication becoming more burdensome. This indicates the significance of pain to the experience of PD patients, and why it is an important target of consideration.

1.2.1.2 Causes of pain in PD

The most common types of pain experienced in PD are musculoskeletal (pain affecting bones, joints, and muscles), radicular (back pain radiating from spinal nerves), and dystonic (painful muscle movements) (8). It has been suggested that PD pain, in particular musculoskeletal pain, could be attributed to motor impairments such as stiffness and loss of movement. However, studies have observed that severity of motor impairment has not been shown to be correlated with levels of pain in PD (8), suggesting this is not the case.

Alterations in central pain processing pathways could also be an underlying cause of pain in PD. Loss of dopaminergic neurons in PD can affect the mesolimbic dopamine system, which affects the experience of painful stimuli and motivational behaviour in response to it. Within chronic pain patients, reduced D2 receptor binding and responsiveness to dopamine has been observed, affecting motivating behaviour (11). This impairment of the response to mesolimbic dopamine activity could however be a result of experiencing persistent painful stimuli rather than a cause of it. One investigation into the effects of dopamine in PD pain observed that PD patients appear to have greater impairment in emotional-motivational pain processing rather than sensory-discriminative pain processing, which could be improved by L-Dopa administration, suggesting that dopamine depletion is a key factor in the experience of this symptom (12). However, a greater understanding of the pain processing pathways affected is required in order to attribute peripheral and central neuropathic pain experienced in PD to pathological factors.

A GWAS was recently conducted to identify genetic risk factors for developing pain in PD (13). PD patients were divided into those with no/low pain or high pain levels, with two genome wide significant SNPs found to be associated with high pain. These were located at the TRPM8 locus, which is a cold-sensing ion channel also involved in inflammation and analgesia. Cannabinoids have been shown to act as TRPM8 antagonists (14), which could be of significance to PD pain considering that trials of cannabinoids have been shown to relieve a range of symptoms in PD including pain and inflammation (15). Furthermore, other TRPM8 antagonists have been investigated for their analgesic properties (16), such as AMTB which reduces painful bladder syndrome and allodynia in different animal models (17,18). TRPM8 is also highly expressed in the caudate, which forms part of the striatum, implicating it in regulation of pain via the basal ganglia pain processing pathway (19). As this region is subject to PD degeneration, this genetic association could indicate an important role for the caudate in PD pain. TRPM8 is also expressed in dorsal root ganglion (DRG) neurons, where its role in depolarising neurons has been linked to the development of neuropathic pain (20). Despite several hypotheses, the mechanism of how this genetic risk factor could be impacting PD is still yet to be fully established.

1.2.1.3 Treatments for pain in PD

Current treatment options for pain in PD include traditional analgesic drugs and dopaminergic therapies, however there are varying reports as to their success. Due to the potential role of dopamine in PD pain, dopaminergic treatments have been studied. Levodopa has previously not been shown to improve PD pain, but some studies suggest it increases the threshold for painful stimuli in PD patients (21). However, a more recent investigation into musculoskeletal pain in PD found that over 80% of PD patients experiencing musculoskeletal pain were responsive to Levodopa, resulting in a reduction in pain intensity scores (22). Alternatively, while a double-blind study of dopamine agonist rotigotine found an improvement in average pain severity in PD patients, these results did not reach significance (23). Mixed results of dopamine therapies could indicate that addressing dopamine depletion cannot completely eliminate the underlying cause of PD pain. However, PD patients experiencing pain most often report it during ‘off states’ (when motor symptoms are worse and medication does not work effectively to control them) rather than 'on states' (where motor symptoms are regulated), suggesting that dopamine levels do impact pain sensations, and further exploration of these therapies is worthwhile (24).

Of other treatment options, non-steroidal anti-inflammatory drugs (NSAIDs) are most used to treat PD pain, whilst other therapies such as opioid analgesics have shown potential to be effective. Opioid analgesic tapentadol demonstrated ability to lower pain severity as well as anxiety and depression in 21 PD patients in one trial (25), and a study of combined oxycodone and naloxone (OXN) provided some evidence that this opioid-based therapy was able to reduce PD pain (particularly musculoskeletal), although the final endpoint improvement was insignificant and adverse side effects were reported (26). One investigation found that 28% of PD patients experiencing pain reported paracetamol to be effective, 12% found NSAIDs effective, 10% found opioids effective and 3% found drugs targeting central pain (gabapentin, pregabalin etc) effective (4). No pharmacological treatments have proved to be universally effective, with mixed to poor results. Alternative therapies such as deep brain stimulation (DBS) have demonstrated ability to alleviate painful symptoms in PD patients (27), however are not viable options for many patients.

The current lack of effective treatments for pain in PD could be due to insufficient understanding of the correct targets to address PD pain, and the heterogeneity of pain experiences in PD patients. With greater knowledge of factors which result in the experience of pain in PD, more effective interventions for PD pain could be identified.

1.2.2 Depression in PD

1.2.2.1 Prevalence of depression in PD

Mood disorders are similarly a common issue, with depression being the most common psychiatric symptom experienced by PD patients (28). An estimated 35% of patients experience clinically significant depressive symptoms, with experiences of sadness, pessimism, and increased anxiety the most widely reported experiences from PD patients (28). While estimates vary, up to 90% of PD patients experience some symptoms of depression (28). In the assessment of worse disease aspects for PD patients, those with advanced PD report that mood disorders rather than pain were one of the most troublesome symptoms, with 7.5% listing this as the most significant symptom, and it ranked second overall behind fluctuating response to medication (4). This indicates the significant of mood disorders for those suffering from the disease long-term.

1.2.2.2 Causes of depression in PD

Different reasons for the prevalence of mood disorders in PD have been proposed. PD diagnosis can be a risk factor for developing a mood disorder, with a difficult illness bringing new struggles in day-to-day life. As with pain, neurological changes in PD can also directly affect mood, with various studies observing that degeneration of dopaminergic neuron projections, as well as loss of noradrenergic limbic and brainstem structures, are linked with PD depression (29). This potential impact of dopaminergic levels on PD depression is supported by patients with motor fluctuations commonly reporting 'off period' depression (30), and PD patients with depression having lower striatal dopamine transporter (DAT) binding (31). Dopamine depletion can have adverse effects on the reward-motivation system as described earlier, which also leads to experience of anhedonia as well as pain dysregulation (32). Whilst not correlated with severity of motor symptoms, the DoPaMiP study observed that depressive symptoms were associated with PD pain (10), which has also been observed elsewhere (33). This suggests that potential shared central mechanisms such as monoamine depletion caused by PD degeneration could be driving PD pain and depression.

Other biological processes in PD that could be affecting mood disorders include the increased levels of inflammation in PD. It has been observed that levels of inflammatory cytokines are elevated in patients with major depressive disorder (MDD), and that some antidepressants also act to reduce cytokine levels (34). TNF-α levels are increased in PD patients with depression compared to those without (35), indicating that inflammatory factors can also be an important factor in PD depression. Systemic activation of T cells has been shown to affect mood via depletion of serum amino acids tryptophan and tyrosine (36), with the resulting serotonin and dopamine deficiency leading to anxiety-like behaviours and increased fear response in a mouse model. The extent to which these inflammatory factors influence PD mood disorders is uncertain, and greater understanding is required of the specific causes of PD-specific depression.

Whilst no GWAS of depression in PD has been conducted to date, potential genetic associations have been investigated. One study previously identified a CB1 receptor gene polymorphism that could impact the expression of this gene that was associated with depression in PD, however no further genetic associations were explored (37). The endocannabinoid system has been shown to be a target in both pain and depression therapies, so this could be an avenue of further investigation. Furthermore, another study investigating the variable effects of PD associated SNPs on different clinical features of disease identified one SNP near the BRIP1 gene which was associated with depression (38). This encodes the BRCA1 Interacting Protein 1, so is difficult to identify the potential impact of this gene function on PD depression.

Differences between experience of depression in Mendelian and idiopathic form of PD could also provide insight into the genetic correlates of this symptom. One investigation found that there was no association with carrying genetic variants in LRRK2, PRKN, or APOE4 with higher rates of depression (39). Familial PD rates of depression were approximately 37.5%, which is comparative to idiopathic PD rates. A different study did observe depression to be more common in SNCA triplication PD patients (40,41). However as this was a small sample of only three familial PD patients, further exploration of genetic associations with PD depression is necessary to determine if SNCA variation is also a factor influencing idiopathic PD.

1.2.2.2 Treatments for depression in PD

SSRIs are traditionally the go-to therapy for PD patients with depression, although their effectiveness is questionable(28). Some trials indicate they are less effective at treating PD depression than non-PD depression. A study of the SSRI paroxetine found this to significantly improve depressive symptoms compared to placebo (42), yet alternative trials have found that paroxetine was no better than the placebo (43). Overall, meta-analysis indicates that there is no strong evidence for the efficacy of SSRIs with PD depression, and that a lack of well powered research into efficacy of antidepressant medications for PD patients hinders the development of better therapeutic options (44).

Given the potential impact of dopamine on depression in PD, dopamine replacement therapies could also be effective anti-depressants as well as treating motor symptoms. Levodopa has not been shown to have any significant antidepressant effects, but a clinical trial of dopamine agonist Pramipexole showed a reduction in depressive symptoms compared with placebo in PD patients (45,46). Greater research into the pharmacological and nonpharmacological interventions into PD depression is required, which will be aided by a greater understanding of factors influencing PD depression.

1.3 The immune system in PD

A key element of the pathophysiology of PD is the activation of the immune system. As described previously, this can potentially influence non-motor symptoms such as depression, and other motor and non-motor symptoms such as pain through its impact on neurodegeneration. The following sections will outline important areas of the immune system and how they relate to PD pathology.

1.3.1 Structure of the immune system

The immune system has two main approaches of responding to a pathogen: the innate response in which myeloid cells such as macrophages perform an immediate general response, and the adaptive response wherein lymphoid cells present a specialised response for the pathogen. To initiate the adaptive response, various antigen presenting cells (APCs) including macrophages, dendritic cells and B cells present peptides derived from these pathogens to adaptive immune cells, primarily T cells and natural killer (NK) cells. These multiply and produce specific antibodies, allowing direct targeting (47). A key element of the adaptive immune system is the major histocompatibility complex (MHC), which consists of the cell surface proteins that enable activation of T cells.

1.3.2 Structure and function of the Major Histocompatibility Complex

MHC molecules are surface proteins that form complexes with antigen peptides to facilitate presentation. They are heterodimers, forming pockets to tightly anchor the antigen in a precise orientation which can then be recognised by T cell receptors (TCR) (48). This process occurs with foreign peptides that have been digested by APCs but also viral proteins and self-derived pathogenic proteins, including α-syn (Figure 1-1).

A huge variety of MHC proteins are involved in this process, with over 200 genes in the MHC system (49). These are separated into class I, class II and class III. The most studied and well-defined MHC genes include the class I genes HLA-A, HLA-B, and HLA-C, and class II genes HLA-DRA, HLA-DRB, HLA-DQA, HLA-DQB, HLA-DPA, and HLA-DPB.

Class I molecules are expressed in most nucleated cells and present antigens to CD8+ cytotoxic T cells, initiating them to clear all 'non-self' molecules. Class II however are primarily expressed on APCs where they present antigens to CD4+ helper T cells, initiating helper responses such as cytokine production. The types of peptides presented also differ; class I presents endogenous proteins such as those from a viral infection, while class II present extracellular proteins from bacterial infections for example (49) (Figure 1-1).

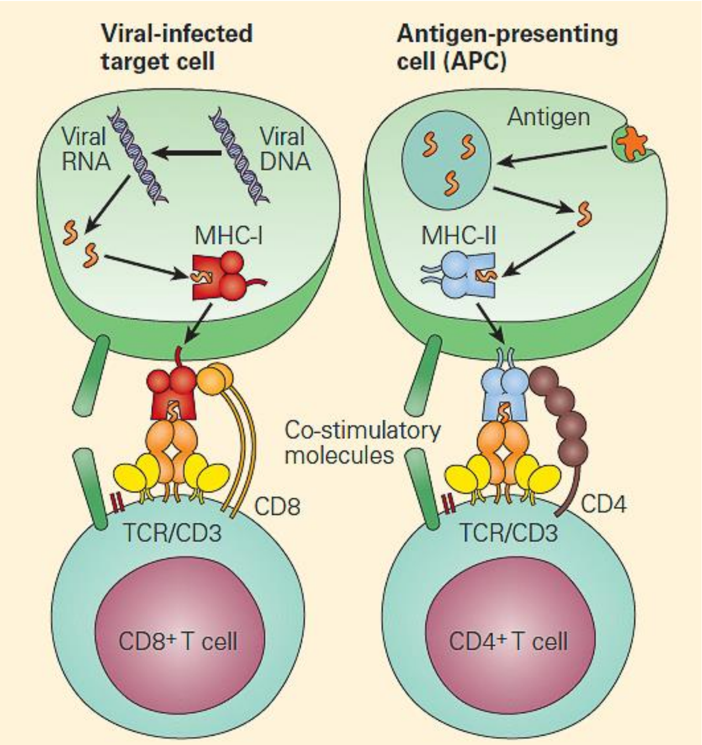


Figure 1-1: Function of the MHC-I and MHC-II molecules.

From Bellanti, JA (Ed), Immunology IV: Clinical Applications in Health and Disease (2012)

To prevent the damaging consequences of unnecessary activation of these processes, this response is tightly controlled. Appropriate activity is ensured by MHC restriction, which requires a specific match up of the antigen and MHC complex with the TCR for activation (50). Extensive MHC diversity ensures that the correct MHC-antigen complex can form for the huge array of potential antigens that could be met. One of the ways in which the MHC is adapted to provide this diversity among MHC proteins and control the immune response is by the unique properties of the genomic region.

1.3.3 Human Leukocyte Antigen genomic region

Located on the short arm of chromosome 6, the HLA (Human Leukocyte Antigen) region covers about 4000 kb and is amongst the most highly polymorphic regions of the human genome (Figure 1-2). The HLA region encodes the MHC proteins, and is so called as it is the human specific version of the MHC. It is characterised by high linkage disequilibrium (LD) between variants, few recombination regions, and high gene density and allele diversity compared to the rest of the genome (51). The IGMT database holds the sequences of all identified alleles of HLA genes, currently listing over 25,000 alleles for 45 genes within the HLA locus (52).

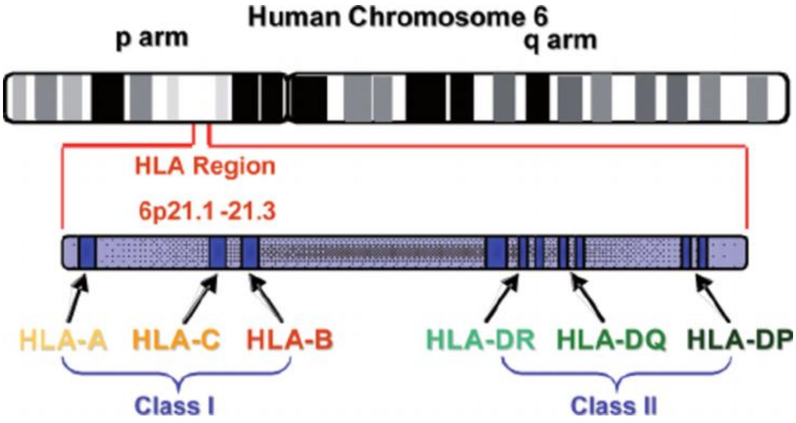


Figure 1-2 HLA genomic region.

From McCarty, Influence of the Human Leukocyte Antigen Complex on the Development of Cutaneous Fibrosis: An Immunogenetic Perspective (2010)

The most polymorphic genes of class I are HLA-A, B and C which have 5,266, 6,537, and 5,140 alleles respectively. The majority of the class II genes are polymorphic; the most polymorphic is HLA-DRB1, which has 2,581 alleles sequenced so far (53). This polymorphism arises from point mutations but also gene conversion, in which class I genes transfer sections of DNA to replace homologous regions (51). Polymorphisms most often occur in peptide binding grooves, with any changes to the amino acid sequence having significant functional effects (54). These new alleles will confer different capacity for binding and presenting peptides, creating a huge array of potential MHC-antigen interactions. Non-binding groove variants can also have significant functional changes, such as cytoplasmic tail alterations affecting MHC transport to the cell surface (53).

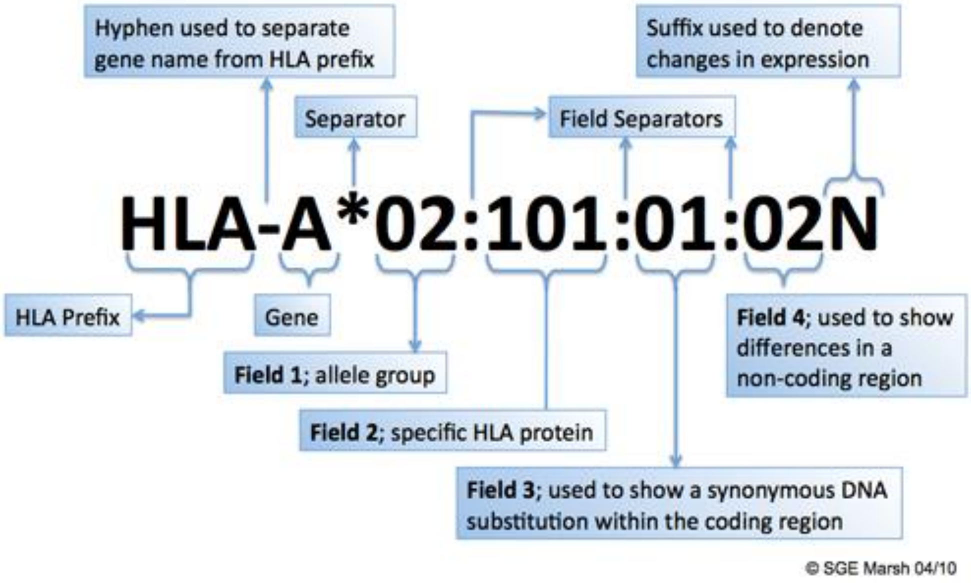

Hyphen used to separate gene name from HLA prefix
Separator
Field Separators
Suffix used to denote changes in expression
HLA-A*02:101:01:02N
HLA Prefix
Gene
Field 1; allele group
Field 2; specific HLA protein
Field 3; used to show a synonymous DNA substitution within the coding region
Field 4; used to show differences in a non-coding region
© SGE Marsh 04/10

***Figure 1-3 HLA Nomenclature**.*
From http://hla.alleles.org/nomenclature/naming.html

Detailed nomenclature is used to categorise each HLA allele, consisting of four sets of numbers divided by colons (Figure 1-3). The first indicates the allele group or allotype, detected by a different serological antigen. The second is the subtype, with a new number for each discovered subtype. Alleles with different numbers for the first two values have different amino acid sequences. Synonymous substitutions in the coding sequence are differed by the third number, and polymorphisms in the non-coding regions are differed by the fourth number.

This variety within the HLA is necessary to support the need for humans to adapt to constantly changing environmental pathogens. With extensive polymorphism, there is a greater chance that individuals will be heterozygous for each HLA gene, increasing chances of survival via a greater capacity for antigen presentation (55). Group survival is also increased with a greater diversity, increasing the likelihood that a fraction of individuals will be able to target a novel antigen. Some alleles will be enriched in certain populations, which have historically given greater chances of survival in a particular environment. It has also been observed that environments with more pathogenic diversity have populations with greater HLA diversity (55). The low recombination frequency of the HLA region results in certain sets of alleles being consistently inherited together. This process could ensure that alleles which work together are collectively inherited, facilitating epistasis (51). This also contributes to the long-range LD patterns of the region.

This adaptation creates a level of complexity when trying to determine HLA risk, with extreme diversity making it difficult to identify associations between particular HLA alleles and disease. Furthermore, many GWAS have a process of excluding regions of high LD or SNP frequency including the HLA locus. If tested, HLA SNPs that are significantly associated with a phenotype will likely be in LD with many other variants over a large region, complicating the interpretation of results. Understanding the HLA variants which carry risk can be important for many conditions, so alternative approaches are often taken to work around this. A growing body of evidence indicates the importance of the adaptive immune response in PD development, so revealing the HLA alleles involved will aid in furthering this understanding.

1.3.4 Adaptive immune response in PD

The activation of the immune response in neurodegenerative diseases has long been studied, with toxic protein accumulation and cell death initiating inflammatory processes. Whether this is a necessary mechanism that helps to counteract the effects on cell degeneration, or a factor which contributes to disease progression, is subject to investigation.

Increased activity of the adaptive immune response in PD has been observed. Circulating levels of pro-inflammatory cytokines, including TNF-α and IFN-γ (56), have consistently been found to be upregulated in PD patients. Differences in T cell subpopulations have also been measured, with a greater proportion of activated T cells compared to naïve, increased CD8+ and decreased CD4+ expression, and a shift to a H1-type immune response observed in patients (57,58). There is some indication that activity of T-regs, which normally act to suppress the activated immune response, may be impaired in PD patients (59) and therefore reduce the ability to control excess inflammation. However, research into the exact nature of the T cell response in PD is ongoing, with inconsistencies across some results. While most of this T cell activity has been monitored in the periphery, the central nervous system is increasingly a focus of study with regards to the PD immune response.

Inflammation within the brain is characteristic of PD, indicated by upregulated DR-positive microglia in the substantia nigra of PD patients (60). This inflammatory response has been suggested to accelerate neurodegeneration and cognitive decline (61). Traditionally the brain has been considered an immune privileged site, with the blood brain barrier (BBB) preventing the entry of peripheral immune cells to exacerbate the response of brain specific microglia. Neurodegeneration however can cause breakdown of this barrier and erode the distinction between these immune systems. Neuroimaging of PD patients has confirmed the breakdown of areas of the barrier including near the midbrain (62), facilitating the entry of peripheral cells. CD8+ and CD4+ T cells, but not B cells, were found to be present in post-mortem brain tissue of PD patients, with the same results observed in MPTP animal models (63,64). A different animal approach involving injection of human α-syn also led to T cell infiltration (65). In this study, infiltration was observed before the development of motor symptoms. This evidence highlights how the activity of these cells of the adaptive immune response can contribute to central as well as peripheral pathology, potentially occurring in early critical stages of the disease.

Different possibilities exist for the pathway by which the adaptive immune response is initiated in PD. α-syn specific T cells have been identified (66), indicating T cells interact with APCs presenting α-syn peptides to generate an immune response. This could occur via breakdown of aggregated α-syn within the brain, or peripheral α-syn. As well as α-syn aggregation within the CNS, there is evidence for early build up in the enteric nervous system, potentially activating α-syn specific T cells within the gut (67). It has been suggested that this early activation of an immune response in the periphery could contribute to the weakening of the BBB and infiltration of T cells (67).

Results from animal studies support this immune response being a significant pathological factor of disease development. The same study that observed T cell infiltration in MPTP mice found that knock down of genes necessary for T cell function resulted in protection against this (64). CD4-/- mice were resistant to MPTP-induced PD symptoms including dopaminergic cell death, whereas CD8a-/- mice were as susceptible as wildtype animals to cell death. This indicated the impact of T cell infiltration could be driven by this specific subset. Transfer of wild type T cells reversed this protection in CD4-/- animals. All mutant animals had similar levels of striatal dopamine and metabolites following MPTP injections, suggesting that T cell infiltration does not seriously effect dopamine production. In a different animal model involving viral overexpression of α-syn, deletion of CCR2 (a receptor expressed on peripheral monocytes) resulted in prevention of peripheral T cell entry, reduced MHCII expression and reduced dopaminergic neuron degeneration (68).

As well as genetic models interfering directly with T cell function, knock outs of MHCII genes have demonstrated protection from different PD models (69,70). One mouse model found expression of human α-syn induced MHCII expression in microglia, with knock-out of MHCII then preventing the antigen presentation of α-syn and subsequent microglial activation, attenuating neurodegeneration (70). A separate model also found that MPTP treatment induced MHCII expression in astrocytes and microglia, while MHCII knock out mice showed significantly reduced MPTP-induced neurodegeneration and cytokine production. Knockout of the class II transactivator, a factor required for MHCII induction, was also sufficient to reduce α-syn induced neurodegeneration and T cell infiltration in mice (71). This demonstrates the central function of MHCII molecules to the immune response and resulting neurodegeneration in models of PD, which could be comparable their role in human disease development.

Collectively this evidence demonstrates the importance of the HLA region to PD pathology, with MHC expression and T cell activity both influencing the extent of neurodegeneration observed. Different cell types are important to consider in this interaction, with expression of MHC proteins observed in SN and LC neurons as well as microglia and astrocytes, facilitating interaction with CD8+ T cells (72). Additional investigation is needed to understand which genes or alleles are significant in this interaction, although current genetic research has suggested some candidates for further study.

1.3.5 HLA in PD genetics

The use of GWAS to investigate the polygenic risk of PD has led to the identification of several of the main risk loci, although most heritable factors are still not understood. Recent advances in data collection have allowed large analysis to be conducted, with GWAS now covering millions of individuals. Among well-established loci, results from the past decade have indicated regions of the HLA locus that confer risk for PD (Table 1). Fully characterising this risk will help to explain the observed impact of MHC proteins on disease development.

One of the earliest observations of a potential HLA risk for PD came in a 2009 GWAS of 1,713 idiopathic PD cases and 3,978 controls from European ancestry. They observed a significant association for rs13192471 within HLA-DQB1, but this did not remain significant when replicated in a wider cohort (73). Subsequently, the UK PD consortium and Welcome Trust case-controlled consortium performed the then largest PD GWAS on 2,190 idiopathic PD cases and 5,667 controls. They found several loci within the short arm of chromosome 6 but outside of the HLA locus, however these too did not remain significant in the replication sample (74).

Following these early indications, more robust results emerged. A GWAS of 2,000 late onset PD cases and 1,986 controls in the NGRC dataset in 2010 found a new risk SNP in the HLA region, rs3129882, which was located within HLA-DRA (75). The association was confirmed and gained significance in a meta-analysis with further datasets. This SNP was proposed to be an eQTL for HLA-DRA, HLA-DQA2 and HLA-DRB5 based on previous expression data. Following this initial study, the NGRC data was further stratified into the subpopulations of sporadic-PD (1,565) and familial-PD (435) to examine if these groups held different associations. They found rs3129882 was more associated with sporadic than familial PD (76), which differed this from other risk variants observed. This result was further tested for interaction with toxin exposure in PD patients, with results suggesting this variant can interact with the environmental impact of a common insecticide to increase PD risk (77).

Whilst these results indicated the HLA risk associated with sporadic PD, results for Mendelian forms of PD have also included HLA loci. The first GWAS of familial PD was conducted in 2009, with 857 cases and 867 controls. A SNP within HLA-DQB1, rs9275184, was found to be significantly associated. The cohort was then used for an additive meta-analysis with a previous GWAS, resulting in a total of 1,124 cases and 1,137 controls. The same SNP gained significance and was in the 20 most significantly associated SNPs in the meta-analysis (78).

Since these results, PD GWAS analyses have increased in power and size. In 2011 a meta-analysis of the then 5 largest GWAS studies was conducted, totalling 5,333 cases and 12,019 controls. Of the 11 loci that passed significance threshold, one was the SNP chr6:32588205 located near HLA-DRB5. This was taken as a confirmation of the earlier indications of HLA being a PD risk locus (79).

An updated meta-analysis was published in 2019, vastly increasingly the scale of analysis (80). A total of 37.7 thousand cases, 18.6 thousand UK Biobank proxies, and 1.4 million controls were included. Consistent with the previous analysis, HLA-DRB5 was identified as a risk locus, with rs11245576 near this gene passing genome wide significance. Also included in this analysis was a summary-based Mendelian randomisation approach to investigate whether QTL properties contributed to any of the results. An meQTL for HLA-DRB5, rs34039593, was identified as the top associated meQTL within the HLA locus.

Study	Top HLA SNP	Nearest gene	P value	Cohort	Effect allele	Cases	Controls
Gasser (2009)	rs13192471	HLA-DQB1	2.65E-04 Did not remain significant when replicated	Idiopathic PD	C	1713	3978
Myers (2009)	rs9275184	HLA-DQB1	Original P = 3.2E-04 Meta-analysis P = 9.5E-05	Familial PD	C	857 (+267)	867 (+270)
Payami (2010)	rs3129882	HLA-DRA	Sporadic P = 5.5 E-10 Familial P = 2.4 E-08	Idiopathic PD (Sporadic and Familial)	G	2000	1986
Nalls (2011)	chr6:32588205	HLA-DRB5	2·58E-08	Idiopathic PD	G	5333	12,019
Ahmed (2012)	rs660895	HLA-DRB1	P < .0001	Idiopathic PD	G	7996	36,455
Nalls (2019)	rs112485576	HLA-DRB5	6.96E-28	Idiopathic PD	A	37.7 thousand cases, 18.6 thousand UK biobank proxies	1.4 million

Table 1: Top HLA risk variants for PD from published GWAS results

Beyond identifying HLA risk SNPs, data from these GWAS have been used for further imputation of HLA allele information. One study used the NGRC dataset (75) to impute HLA alleles using SNP2HLA and HLA*IMP, as well as conducting further HLA sequencing of 196 PD cases and 204 controls (81). The HLA haplotype

B*07:02_C*07:02_DRB5*01_DRB1*15:01_DQA1*01:02_DQB1*06:02 was found to be positively associated with PD, while the group of alleles C*03:04, DRB1*04:04 and DQA1*03:01 were negatively associated. A separate cohort was also used for an imputation investigation using HLA*IMP (82), which identified an association for PD with the HLA-DRB1*04 allele. This study also performed a meta-analysis on 4 GWAS data sets, finding an association with rs660895 within HLA-DRB1.

The growing power of GWAS has aided in refining of the risk locus at the HLA region, although there are still different candidates for the main genes and alleles associated. HLA-DRB5 has been presented as a gene of interest from the largest meta-analysis, due to its proximity to associated SNPs and their ability to act as eQTLs for this gene. However, this overlooks the potential trans and cis acting effects of the risk variants, and how these could be associated with other HLA loci. Furthermore, HLA-DRB5 is only present in approximately 20% of the population (83). Caution must therefore be taken before focusing on this as the main gene of interest. Other class II genes have been repeatedly implicated in PD, and imputation of data have shown a variety of potentially associated alleles. Further analysis of these risk loci is necessary to identify how these genes are involved and the biological pathways affected.

The results of these GWAS investigations have indicated several HLA risk loci of interest, yet how these loci impact HLA expression or function is not always clear. Risk variants are often in intronic regions, affecting expression or function of unknown genes. Whilst some PD risk SNPs have been identified as potential eQTLs, for most the gene or extent of impact is unknown. Animal studies have shown the extreme effects that complete knockdown of HLA genes can have on PD pathology, yet these effects will not likely be comparable to biological consequences of altered expression caused by polygenic risk SNPs. Therefore, it is important to understand the extent of the impact of genetic risk on gene expression.

A recent investigation using public expression data sought to further the understanding of which genes are most impacted (84). An initial study examined the distribution of PD risk SNPs overlapping tissue-specific regulatory elements, finding that most enrichment was seen in non-neuronal tissue including lymphocytes (85). This indicated the altered gene expression outside of neuronal tissue could be more significant than central brain effects. However, this only examined tissue-specific active regions rather than affected gene sets.

Subsequently, a genome-wide screen was carried out to gather 7,607 PD risk SNPs and 23,759 proxy variants (LD > 0.8) from public databases. GTEx was used to identify PD risk eQTLs and examine the correlation between GWAS significance and eQTL expression. In addition, overlap between enhancer regions and risk SNPs was identified, and gene set enrichment analysis performed on the genes affected.

795 genes were found to be affected by risk SNPs, with strong associations seen at chromosome 6. Gene set and pathway enrichment were conducted across different methods, with most methods showing that within the majority of gene sets, antigen presentation pathway dominated. To establish whether this was due to the strength of the HLA locus overwhelming other associations, risk SNPs were further subclassified to those that both overlapped dbSUPER defined superenhancers and which disrupted transcription factor binding. These 95 genes were also functionally enriched for antigen presentation processes. A strong correlation was observed between associated eQTL expression changes and PD GWAS significance, suggesting eQTL associations directly impacted PD risk. The subset of brain tissue-specific eGenes also included 20 at the HLA locus, which aligned with a previous investigation into eQTLs within the prefrontal cortex of PD patients that found HLA gene expression was impacted (86).

Overall, these results indicate that of the known PD risk SNPs that act as eQTLs, they significantly effect antigen related processes involving HLA genes. The HLA genes that were associated with increased expression included HLA-B, HLA-C, HLA-DQA1, HLA-DQB1, HLA-DQB1-AS1, HLA-DRB1, and HLA-DRB5, whilst those that showed decreased expression included HLA-DOB, HLA-DQA2, HLA-DQB2, and HLA-DRB6. It was suggested that the pathways altered by PD eQTLs could contribute to excessive neuroinflammation in PD patients, although the exact mechanism and extent to which this contributes to PD pathology is still unknown.

1.4 Aims and Outline

The main aim of this thesis is to contribute to the understanding of genetic factors influencing inflammation, pain, and depression in PD. As detailed in this chapter, there is at present some understanding of the potential genetic factors influencing these, primarily through GWAS or other association studies. However, two main areas that require further research have been identified.

The first is the understanding of the nature of the HLA association with PD. HLA alleles are long, complex, and highly polymorphic genes, with new alleles being identified regularly. The methods applied to date have primarily focused on identifying associated loci and commonly known alleles that could be driving the association between PD and the HLA. This has been insufficient to capture unambiguous data of the full length of HLA alleles within PD subjects. Therefore, if rare or novel HLA alleles are associated with PD, they have yet to be identified.

The second is the understanding of the factors influencing the development of two of the primary non-motor symptoms in PD: pain and depression. These symptoms have both been correlated with each other, as well as to inflammatory processes that occur in PD. There has been no genome wide association study of depression in PD, and no exploration of how these two symptoms could be influencing each other in PD. Greater understanding of how these symptoms develop is needed to improve treatment options.

To address these issues, the aims of each experimental chapter are as follows:

1. Apply a range of bioinformatic approaches to a PD case control dataset to understand which HLA loci are most associated with PD risk and protection. This will include various HLA imputation methods and QTL analysis. This will be used to select HLA loci most appropriate to target for long-read sequencing.
2. Apply long-read sequencing to the HLA locus in PD samples to determine which alleles are associated with PD risk and protection. This approach will implement the Pacific Biosciences long-read technology to study the complex HLA region, and so will be able to uncover any unknown alleles or structural variants within the HLA locus of PD samples.
3. Study the genetic factors influencing pain and depression in PD, and the relationship between these symptoms. This will include conducting further GWAS of pain and depression in PD, a polygenic risk score analysis to identify any shared genetic factors between PD and non-PD symptoms, and a Mendelian randomisation study to identify any causal relationships.

2 Exploring Parkinson’s Disease Associated HLA Loci

2.1 Introduction

Numerous PD GWAS studies have indicated the HLA is a genetic risk factor for PD. This was first observed in a 2009 GWAS of fewer than 6,000 subjects, in which a significant association for rs13192471 within the HLA-DQB1 locus was identified (73). However, this did not remain significant when replicated in a wider cohort. Over the subsequent decade, the power of PD GWAS grew significantly, enabling a greater ability to detect genetic associations of interest. The most recent PD GWAS meta-analysis in 2019 was able to combine data from 37.7 thousand cases, 18.6 thousand UK biobank proxies, and 1.4 million controls (80). This identified the top HLA risk SNP to be rs11245576, with the minor allele having a protective effect on disease (P = 6.96E-28, OR = 0.85). As the nearest gene was HLA-DRB5, this was proposed as the HLA gene of interest; however, this gene is only present in approximately 20% of individuals who carry HLA-DRB1 alleles within the HLA-DR15 serotype (83). Therefore, this is unlikely to explain the full extent of the genetic association with PD at the HLA locus.

As this proposed risk locus is potentially not the only HLA locus involved in PD risk, other avenues of investigation are required to understand additional HLA loci that confer risk for PD. Furthermore, once specific risk loci are identified, the highly polymorphic nature of the HLA locus means there are potentially thousands of alleles with differing disease associations. The complex pattern of linkage disequilibrium within the HLA means that there can be multiple alleles that are associated, however this also makes it difficult to pinpoint the causative association (51).

Alternative strategies to observing the nearby HLA genes to the top risk SNP have been applied to understand PD HLA risk. One such approach is to impute HLA alleles within a sample, and test these for association with disease. HLA imputation is the method by which common SNP data is used to infer the specific HLA alleles carried by an individual, using a reference panel of samples with known HLA data. Different imputation software can be applied using various computational approaches to give the ‘best guess’ alleles based on SNP data. One of the initial PD HLA imputation studies was conducted in 2013, which applied SNP2HLA and HLA*IMP software's to impute HLA alleles (81). This study identified C*03:04 (P = 8.3x 10^{-6}, OR = 0.72) and DRB1*04:04 (P = 4.3x10^{-5}, OR = 0.65) as independent HLA alleles associated with PD risk. As well as identifying the potential DRB1 allele of interest, this was also the first indication of a class I HLA allele association with PD. Since this result, sample sizes of imputation studies and computational power of imputation programmes have developed. A more recent imputation approach in 2021 was published after completion of the work in this chapter. This applied a different computational approach with use of the HIBAG software (87). The association of HLA-DRB1*04:04 was replicated in this study, with this being the main PD association (P = 8.21x10^{-5}, OR = 0.84). PD association was also identified for HLA-DQA1*03:01, HLA-DQB1*03:02, and HLA-DRB1*04:01 alleles, however no HLA class I alleles were found to be associated with disease.

Other factors beyond imputed alleles associated with PD can aid in understanding of whether PD associated HLA associations are indicating specific HLA loci. For example, QTL (quantitative trait loci) properties of associated SNPs can be informative. This was the case within the 2019 PD GWAS meta-analysis, which identified an meQTL (methylation QTL) for HLA-DRB5, rs34039593, as the top associated meQTL within the HLA locus (80).

One of the main aims of this thesis is to apply a long-read sequencing approach to HLA loci in PD samples. Before pursuing sequencing work, it is important to have a more detailed picture of which HLA loci it is worthwhile to investigate with this approach. To achieve this, a PD case-control sample in which the HLA alleles had previously not been imputed was obtained. This consists of 5,322 cases and 10,018 controls that were previously included within the 2011 PD meta-analysis (79). Multiple different HLA imputation approaches were applied to compare results across classical imputation methods and newer approaches. Additionally, various up to date QTL databases were applied to GWAS and conditional analysis results to gain improved insight into the HLA loci of interest indicated by these results. This data, combined with previously published results, allows an informed selection of the HLA alleles most likely to be associated with PD to target for deeper analysis with long-read sequencing.

2.1.1 Aims

The aims of this current investigation include the following:

1. To conduct a GWAS and conditional analysis in a PD case-control dataset to further analyse the HLA risk variants identified.
2. To apply bioinformatics approaches including a range of different HLA imputation approaches and QTL searches to this dataset to further identify the most likely HLA loci that could be associated with PD.
3. To select HLA risk loci for PD for further investigation for PacBio long-read sequencing.

2.2 Materials and Methods

2.2.1 Study Population

The individuals used for this investigation were obtained from the dataset gathered for the 2011 PD meta-analysis (79). These represent 4 separate European/American populations; USA-NIA, UK, Germany and France (Table 2). Individuals from the USA dbGAP dataset were also used. 15,340 individuals in total were included in this analysis, with 5,322 cases and 10,018 controls. The populations had a similar age and sex divide. Data collection and quality control procedures were similar for each sample (79).

Population group	**Cases**			**Controls**			**Total**
	Sample size	**Women (%)**	**Mean age at onset (years [SD])**	**Sample size**	**Women (%)**	**Mean age at examination (years [SD])**	
US-NIA	971	40.5	55.9 (15.1)	3034	52.8	62 (15.6)	4005
UK	1705	43.3	65.8 (10.8)	5200	49.5	NA	6905
Germany	742	39.8	56 (11.6)	944	48	NA	1686
France	1039	41.2	48.9 (12.8)	0	NA	NA	1039
USA-dbGAP	876	40.4	61.5 (9.2)	857	60.2	NA	1733
Total (after removal of duplicates)	5322			10,018			15,340

Table 2: Characteristics of datasets included in the PD case control sample for this study.

2.2.2 Genotyping and Quality Control

The genotyped SNPs that were present across all datasets were obtained, with a total of 252,356 common SNPs. All samples from each of the populations had similar standardised quality control, including the inclusion criteria of < 95% genotyping success rate per SNP and < 95% call rate per sample. A further 6,468,921 SNPs were then imputed using the HRCv1.1 reference panel, totalling 6,721,277 SNPs. SNPs were excluded according to the following thresholds: imputation quality (INFO) > 0.8, missingness (geno) > 0.02, minor allele frequency (MAF) < 0.01, and HWE P value < $1x10^{-6}$. Individuals were excluded according to missing genotype data (mind) > 0.01.

2.2.3 GWAS and Conditional Analysis

A genome wide association study (GWAS) was conducted with this dataset using Plink version 1.9. Initially, a principal component analysis (PCA) was conducted. By controlling for principal components (PCs), confounding bias by population stratification can be reduced. To generate the PCs, the PD case-control dataset was merged with the FIN, CHB and YRI samples from 1000 genomes reference genotype data, and the resulting dataset LD pruned. This merged dataset was used for a principal components analysis (PCA). The output eigenvec file was imported to R studio version 3.2.0 to calculate which of the PCs were significantly associated with disease state using a logistical regression model. All PCs which passed the significance threshold ($P < 1x10^{-10}$) were then used as covariates for the GWAS. The association analysis was conducted using logistic regression in Plink with 14 PCs included as covariates.

SNPs from the HLA region (hg19 28–33.5 Mb) were extracted from the results of the association analysis to find significant associations in this region. The top SNP from this region was used as a covariate for a further conditional analysis to observe any further independent associations. The same conditions and covariates were used in this conditional analysis.

2.2.4 Imputation Approaches

A variety of imputation statistical programmes were used to impute HLA alleles, amino acids, and SNPs from the genotype data of this population.

2.2.4.1 SNP2HLA

SNP2HLA is an imputation method that uses Beagle to impute HLA alleles from genotyped data using a reference panel (88). This imputation method produces allele dosages for 2- and 4-digit classical HLA alleles for HLA-A, -B, -C, -DQA1, -DQB1 and -DRB1, as well as polymorphic amino acid positions and SNPs.

SNP2HLA version 1.0.3 was used along with Beagle version 3.0.4, linkage2beagle and Plink version 1.9. Recommended parameters were used with 10 iterations and a 1000 marker window. The imputation was run using the HapMap-CEPH (124 samples) reference panel provided.

Following imputation, variants with INFO < 0.5 were excluded, leaving a total of 3,136 alleles for further analysis. A regression analysis for association with disease state was then conducted with Plink using this allele dosage data.

2.2.4.2 HIBAG

HIBAG (HLA Imputation using attribute BAGging) is an imputation method which uses attribute bagging to impute 4-digit classical HLA alleles for HLA-A, -B, -C, -DRB1, -DQA1, -DQB1 and -DPB1 from genotype data (89). HIBAG version 1.20.0 along with R studio version 3.2.0 were used with the European-HLA4-hg19 (2,572 samples) reference dataset provided to impute HLA alleles from the sample data. An association analysis for disease state was run with these imputed alleles in R studio.

2.2.4.3 DISH

DISH (Direct imputing summary association statistics HLA variants) is the most recently developed imputation software, which imputes summary association statistics of HLA alleles from GWAS output summary association statistics (90). This method imputes association statistics for 2- and 4-digit classical HLA alleles for HLA-A, -B, -C, -DQA1, -DQB1 and -DRB1, as well as polymorphic amino acid positions and SNPs. DISH version 1.0 was used with R studio version 3.2.0 for imputation.

2.2.5 Haploview

Haploview version 4.1 was used to observe linkage disequilibrium (LD) between genotyped SNPs and imputed HLA SNPs, alleles and amino acids. Genotype data for the PD dataset and SNP2HLA imputation results were used as input data and converted to Haploview input files using Plink. LDLink was used to find improved LD R^2 and D’ values where applicable.

2.2.6 Public QTL Databases

Public databases were used to search for QTL properties of the top associated variants from the GWAS, conditional analysis and imputation results. The following datasets were included in this search.

2.2.6.1 eQTLGen

eQTLGen is a database which includes eQTL information from 37 expression datasets, with a total of 31,684 individuals tested across all cohorts (91). Gene expression and genotype data was obtained from blood samples, with 19,250 genes expressed in blood tested for association. Every eQTL reported was tested in at least 2 cohorts. P values provided are Bonferroni corrected, and the number of cohorts and samples in which this SNP-gene combination was tested is provided. Effect is provided as a Z-score of the assessed allele.

2.2.6.2 GTEx

The GTEx portal is a source of gene expression data, with samples collected from 54 non-diseased tissue sites across nearly 1,000 individuals (92). 838 donors from the database had whole genome sequencing and RNA-seq data used for eQTL analysis. Results provide the tissue type in which this QTL was observed, as well as the adjusted P values and normalised effect size (slope of linear regression) for the alternative allele.

2.2.6.3 QTLbase

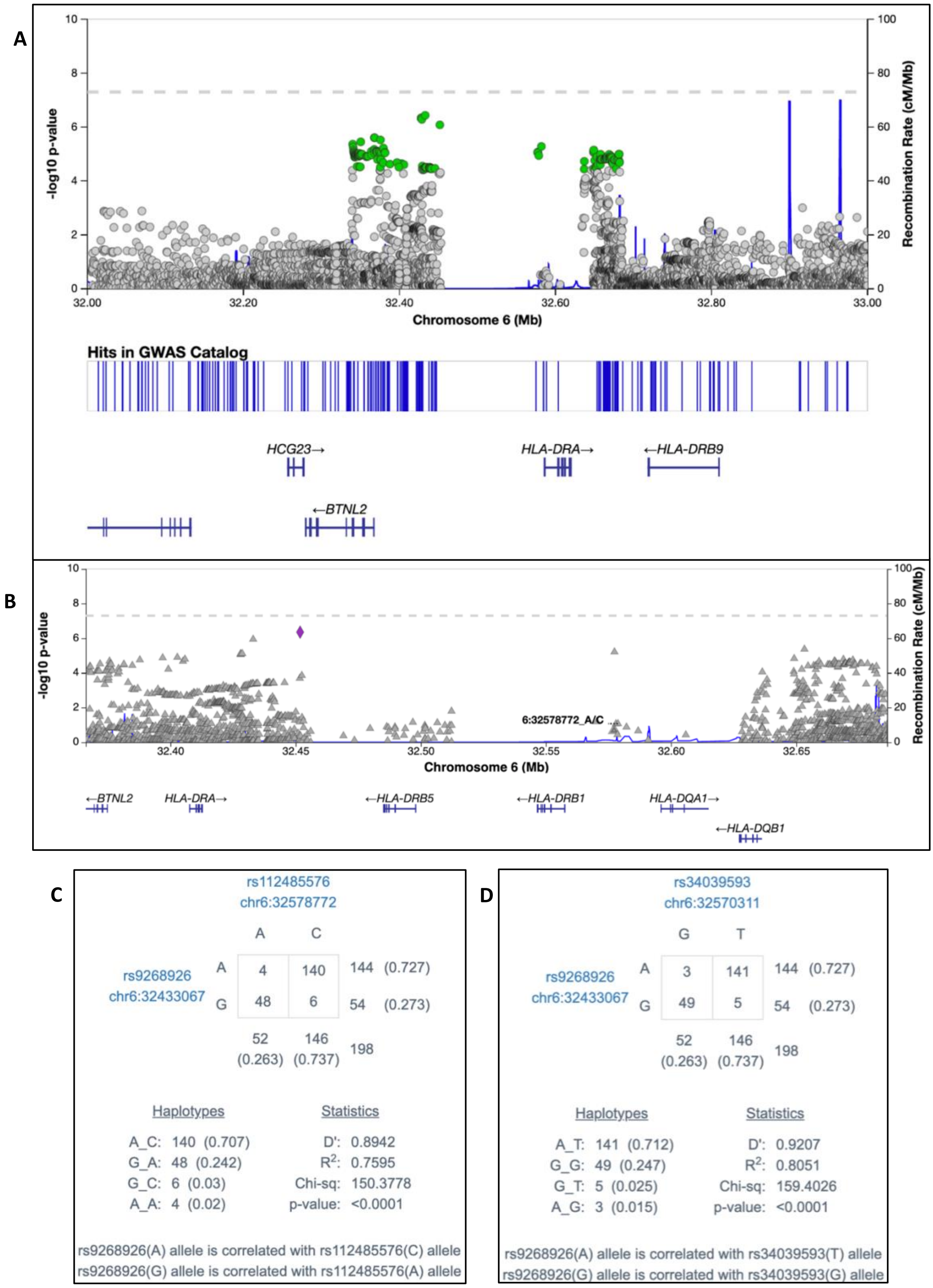
QTLbase is a database which gathers published QTL summary statistics from different studies across more than 70 tissue or cell types (93). Release 1.1 was consulted, including 167 independent studies for eQTL data in the database. Results from QTLbase provide the study in which the QTL was identified, and the sample size of that study. The P value and effect size from the original study are provided.

2.3 Results

2.3.1 GWAS

A GWAS was conducted using the PD case-control dataset gathered (see Appendix 1 for Manhattan and QQ plots). The genomic inflation factor was λ = 1.099. The most associated SNPs within the HLA region ($P < 8.00 \times 10^{-6}$) are listed in Table 3. The most significantly associated SNP is rs9268926 ($P = 3.67 \times 10^{-7}$, OR = 0.84). A plot of this region is displayed in more detail in Figure 2-1A. The minor allele G is associated with disease protection, and the major allele A is associated with disease risk (Table 3). This SNP is in partial linkage disequilibrium (LD) with the most associated HLA SNP from the 2019 meta-analysis results, rs112485576 ($R^2 = 0.76$, D' = 0.89). Of the rs112485576 alleles, the minor A protective allele is correlated with the minor G protective allele (Figure 2-1C). This SNP is also in LD with the top meQTL from the same analysis, rs34039593 ($R^2 = 0.81$, D' = 0.92), with the minor G allele correlated with the minor G allele of rs9268926 (Figure 2-1D). This suggests these results are comparable and could be indicating the same association, although this is inconclusive.

With the quality control criteria set at this standard (INFO > 0.8, geno > 0.02, MAF < 0.01, and HWE P value < 1×10^{-6}), rs112485576 was excluded from this GWAS. The criteria were relaxed so that this variant was included, reducing the exclusion criteria to INFO > 0.5, geno > 0.35, MAF < 0.002 and HWE < 0.00001 (Figure 2-1B). The PD association for this variant was insignificant in this analysis (P = 0.085, OR = 0.62). This indicates that the association at rs112485576 required a larger sample size to be significantly associated; the partial LD between this SNP and rs9268926 suggests the top GWAS result could still be indicating the same HLA association (Figure 2-1C). These relaxed criteria were only used for the purpose of exploring the rs112485576 results; the analysis with the original stringent QC criteria was used for the subsequent imputation analysis.



C

rs112485576
chr6:32578772

rs9268926 chr6:32433067		A	C		
	A	4	140	144	(0.727)
	G	48	6	54	(0.273)
		52 (0.263)	146 (0.737)	198	

Haplotypes	Statistics
A_C: 140 (0.707)	D': 0.8942
G_A: 48 (0.242)	R^2: 0.7595
G_C: 6 (0.03)	Chi-sq: 150.3778
A_A: 4 (0.02)	p-value: <0.0001

rs9268926(A) allele is correlated with rs112485576(C) allele
rs9268926(G) allele is correlated with rs112485576(A) allele

D

rs34039593
chr6:32570311

rs9268926 chr6:32433067		G	T		
	A	3	141	144	(0.727)
	G	49	5	54	(0.273)
		52 (0.263)	146 (0.737)	198	

Haplotypes	Statistics
A_T: 141 (0.712)	D': 0.9207
G_G: 49 (0.247)	R^2: 0.8051
G_T: 5 (0.025)	Chi-sq: 159.4026
A_G: 3 (0.015)	p-value: <0.0001

rs9268926(A) allele is correlated with rs34039593(T) allele
rs9268926(G) allele is correlated with rs34039593(G) allele

Figure 2-1: LocusZoom plot of top GWAS result in HLA region.

(A) The region in the original GWAS with the most associated HLA SNP, rs9268926 (P = 3.67 x 10^{-7}, OR = 0.8428). A section within this region was removed during quality control due to poor imputation score (B) The secondary GWAS conducted with reduced quality control thresholds to include rs112485576, which is labelled as position 6:32578772. Lead SNP in purple, SNPs in LD, 0.6 < R^2 ≤ 0.4 in green. (C-D) LDLink box plot demonstrating the correlated alleles between the Nalls top SNP rs112485576 (C), and the Nalls top meQTL rs34039593 (D) with rs9268926.

SNP	Position (hg19)	Effect allele	Non-effect allele	Odds Ratio	P value
rs9268926	32433067	G	A	0.84	3.67E-07
rs9268833	32428062	T	C	0.86	4.65E-07
rs9268834	32428079	A	C	0.86	4.65E-07
rs9268835	32428115	A	G	0.86	4.65E-07
rs9268838	32428715	A	G	0.86	5.23E-07
rs554742089	32451822	T	C	0.85	8.33E-07
rs3817966	32367847	C	T	0.87	2.49E-06
rs3817963	32368087	C	T	0.87	2.49E-06
rs9268499	32375695	A	G	0.87	3.00E-06
rs9268458	32350384	A	C	0.87	3.55E-06
rs1980496	32340070	T	C	0.88	4.32E-06
rs521539	32581973	A	G	0.86	5.24E-06
rs9268400	32340654	A	G	0.87	5.49E-06
rs9268516	32379489	T	C	0.87	6.01E-06
rs9268514	32378945	A	T	0.87	6.48E-06
rs2294880	32367722	G	A	0.87	6.68E-06
rs9275098	32649161	T	C	0.82	7.09E-06
rs9268401	32341318	G	A	0.87	7.15E-06
rs9275095	32649088	G	C	0.82	7.76E-06
rs9268482	32367777	T	A	0.87	7.85E-06

Table 3: Top 20 variants within the HLA region associated with PD from the GWAS results.

To establish if there is existing eQTL data for rs9268926 and rs112485576, public QTL databases were searched. Results from the GTEx consortium indicates rs9268926 could be associated with increased HLA-DQA2 expression, with the top 9 associations all for HLA-DQA2 (Table 4). HLA-DQA2 is a paralogue of HLA-DQA1. However, QTLbase results indicate that mixed eQTL results have been published for rs9268926, including effects on HLA-DRB1 and HLA-DPB1 expression (Table 5). The QTL properties of rs9268926 therefore remain ambiguous.

Gene	P Value	Effect size	Tissue
HLA-DQA2	3.00E-78	1.1	Muscle - Skeletal
HLA-DQA2	2.40E-73	1	Whole Blood
HLA-DQA2	3.90E-67	1.2	Adipose - Subcutaneous
HLA-DQA2	9.20E-65	0.94	Skin - Sun Exposed (Lower leg)
HLA-DQA2	7.20E-64	1	Artery - Tibial
HLA-DQA2	1.60E-62	1.1	Lung
HLA-DQA2	5.80E-62	1.1	Nerve - Tibial
HLA-DQA2	2.40E-57	1	Thyroid
HLA-DQA2	3.70E-55	1.1	Esophagus - Muscularis
HLA-DRB6	1.80E-52	0.88	Muscle - Skeletal

Table 4: Top 10 results from the GTEx database for rs9268926

Gene	P Value	Effect Size	Tissue	Source	Sample size
HLA-DRB1	3.03E-28	0.76	Brain-Prefrontal Cortex	Fromer (2016)	467
HLA-DRB1	2.08E-15	0.74	Brain-Prefrontal Cortex	Fromer (2016)	467
HLA-DPB1	2.59E-12	0.19	Blood	Jansen (2017)	4896
HLA-DQA1	1.22E-11	0.88	Brain-Prefrontal Cortex	Fromer (2016)	467
HLA-DQA1	2.94E-09	0.87	Brain-Prefrontal Cortex	Fromer (2016)	467
SLC44A4	9.90E-08	0.15	Blood	Jansen (2017)	4896

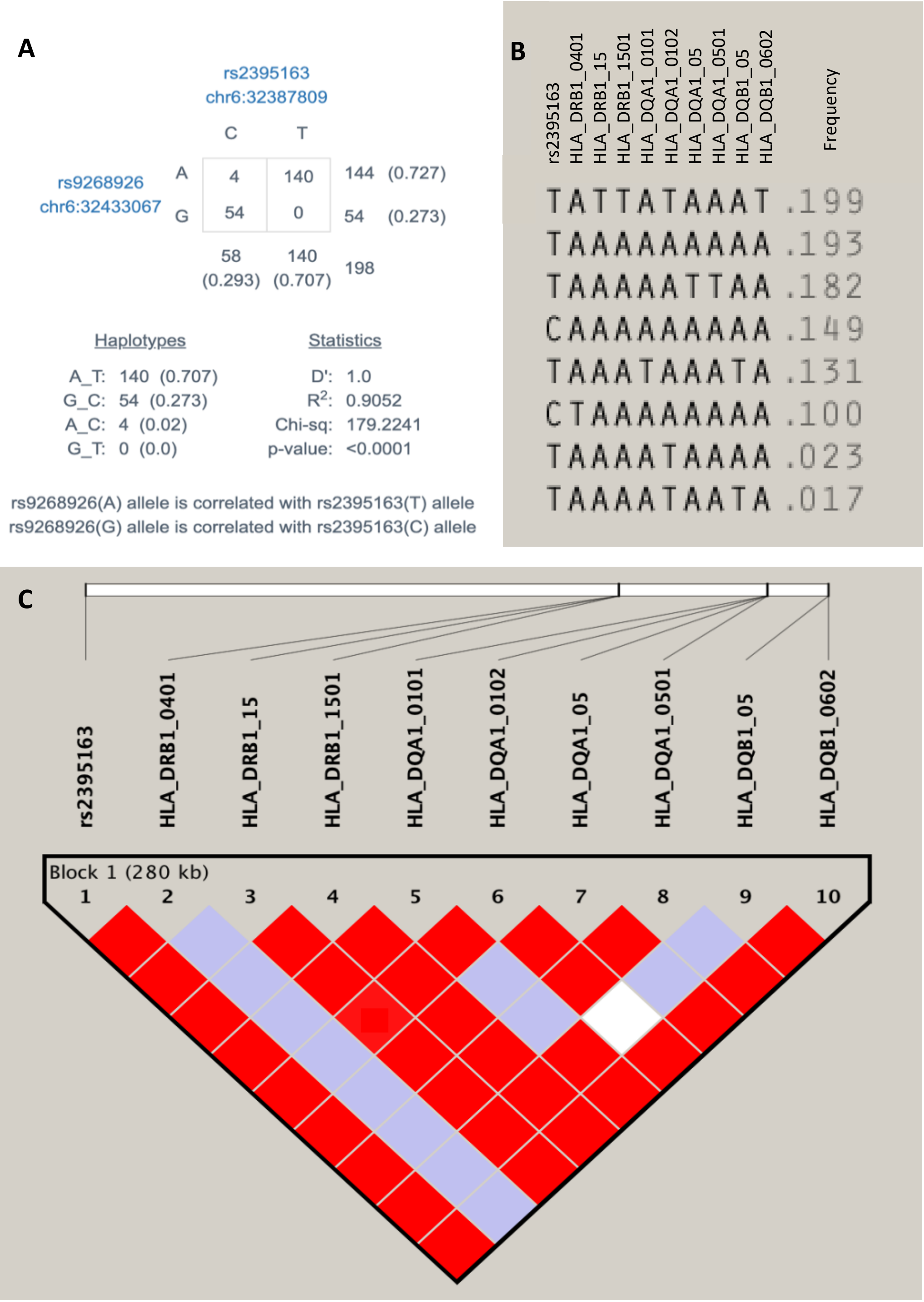
Table 5: Top associations from the QTLbase database for rs9268926

The 2019 meta-analysis proposed their top HLA variant (rs112485576) to be an eQTL for HLA-DRB5, based upon its proximity to the gene and on published eQTL data from eQTLGen. A search of the database revealed that whilst HLA-DRB5 expression was found to be significantly reduced when tested in the largest number of cohorts, there are also mixed eQTL results for this SNP, with an increase in HLA-DQA2 expression included as the most significant association (Table 6). This indicates that HLA-DRB5 is potentially not the only candidate for the affected gene, and other loci could be being impacted.

Gene	P value	Z-score	No. Cohorts	No. Samples
HLA-DQA2	3.27E-310	55.72	13	5500
HLA-DRB1	3.27E-310	-38.01	13	5500
HLA-DRB6	4.69E-100	-21.23	30	26959
HLA-DQB1	6.65E-96	-20.78	28	21135
HLA-DQB2	5.92E-91	20.22	14	10575
HLA-DQB1-AS1	4.58E-34	-12.17	12	4992
HLA-DRB5	7.37E-32	-11.75	33	22312

Table 6: Significant HLA associations for rs112485576 within eQTLGen database

This analysis of the HLA region has shown which SNP allele is correlated with risk and which is correlated with protection, and the potential QTL properties of the effect alleles. Haploview was then used to assess which HLA alleles were in LD of D' = 1 with these alleles. The HapMap CEU HLA reference panel was used for analysis. This panel did not contain the top SNP rs9268926, but the proxy SNP rs2395163 was analysed (R^2=0.91, D'=1) (Figure 2-2A). The minor C allele of rs2395163 is correlated with the minor G allele of rs9268926. The following alleles formed haplotypes with alleles of this SNP; HLA-DRB1*0401, DRB1*15:01, DQA1*01:01, DQA1*01:02, DQA1*05:01 and DQB1*06:02 (2-2B-C). Figure 2-2C shows the Haploview block with these alleles, while Figure 2-2B indicates which alleles form haplotypes with the PD risk associated allele T and protective associated allele C. The number 01-10 in Figure 2-2B indicate the variants in the same order as presented in 2-2C, with 'A' indicated HLA allele absence and 'T' indicating HLA allele presence. This demonstrates that HLA-DRB1*04:01 is correlated with the C allele, whilst HLA-DRB1*15, HLA-DQA1*01, HLA-DQA1*05 and both HLA-DQB1 alleles are correlated with the T allele. Therefore HLA-DRB1*04:01 is correlated with PD protection, and the others with PD risk.



A

		rs2395163 chr6:32387809 C	T	
rs9268926 chr6:32433067	A	4	140	144 (0.727)
	G	54	0	54 (0.273)
		58 (0.293)	140 (0.707)	198

Haplotypes		Statistics	
A_T:	140 (0.707)	D':	1.0
G_C:	54 (0.273)	R^2:	0.9052
A_C:	4 (0.02)	Chi-sq:	179.2241
G_T:	0 (0.0)	p-value:	<0.0001

rs9268926(A) allele is correlated with rs2395163(T) allele
rs9268926(G) allele is correlated with rs2395163(C) allele

B

rs2395163	HLA_DRB1_0401	HLA_DRB1_15	HLA_DRB1_1501	HLA_DQA1_0101	HLA_DQA1_0102	HLA_DQA1_05	HLA_DQA1_0501	HLA_DQB1_05	HLA_DQB1_0602	Frequency
T	A	T	T	A	T	A	A	A	T	.199
T	A	A	A	A	A	A	A	A	A	.193
T	A	A	A	A	A	T	T	A	A	.182
C	A	A	A	A	A	A	A	A	A	.149
T	A	A	A	T	A	A	A	T	A	.131
C	T	A	A	A	A	A	A	A	A	.100
T	A	A	A	A	T	A	A	A	A	.023
T	A	A	A	A	T	A	A	T	A	.017

***Figure 2-2: Haploview plots of HapMap HLA alleles in LD (D'=1) with the proxy for rs9268926, rs2395163*.**

(A) LDLink plot demonstrated the correlated alleles between rs9268926 and rs2395163. rs9268926 was the most significantly associated HLA SNP with PD in the current GWAS (P = 3.67 x 10-7, OR = 0.84). The rs9268926 minor allele G is correlated with the minor allele C of proxy SNP rs2395163 (B) Haploview haplotype plot showing which HLA alleles form haplotypes with alleles of proxy SNP rs2395163. 'T' and 'C' indicate the rs2395163 allele whilst for the HLA alleles 'A' indicates absence of HLA allele while 'T' indicates presence. (C) LD block of the HLA alleles in D' = 1 with the SNP rs2395163. Red and blue squares indicate D'=1, with red indicating LOD (log of the likelihood odds ratio, a measure of confidence in the value of D) ≥ 2 and blue indicating LOD < 2. White squares indicate D' < 1.

2.3.2 Conditional Analysis

A subsequent conditional association analysis was conducting with rs9268926 included as a covariate to test for any independent HLA associations in this dataset. The results from this analysis indicated there was one potential independent association, with one SNP (rs9295987) reaching suggestive significance (P = 9.82 x 10^{-5}, OR = 0.80) (Table 7).

SNP	Position (hg 19)	Effect allele	Non-effect allele	OR	P
rs9295987	31349844	G	A	0.80	9.82E-05
rs9261503	30111863	A	G	0.86	1.29E-04
rs9261505	30112408	G	A	0.86	1.29E-04
rs9261504	30111932	T	C	0.86	1.51E-04
rs9261501	30111526	T	C	0.86	1.63E-04
rs757262	30114955	T	G	0.86	1.63E-04
rs757259	30115542	A	C	0.86	1.63E-04
rs1573299	30115965	T	G	0.86	1.63E-04
rs1573297	30116341	T	G	0.86	1.63E-04
rs9261514	30116537	A	G	0.86	1.63E-04

Table 7: Top 10 associated variants within the HLA region from the conditional analysis results.

This SNP was also searched for within the QTL databases. The top associations did not include HLA alleles, so the HLA specific associations were extracted rather than the top results. Both results from GTEx database and QTLbase indicate that this SNP is potentially correlated with expression of class I genes HLA-B and HLA-C. The two most significant HLA associations within GTEx were for increased expression of HLA-C (Table 8), and the top results from QTLbase were mixed associations with HLA-C expression, with HLA-B also significant (Table 9).

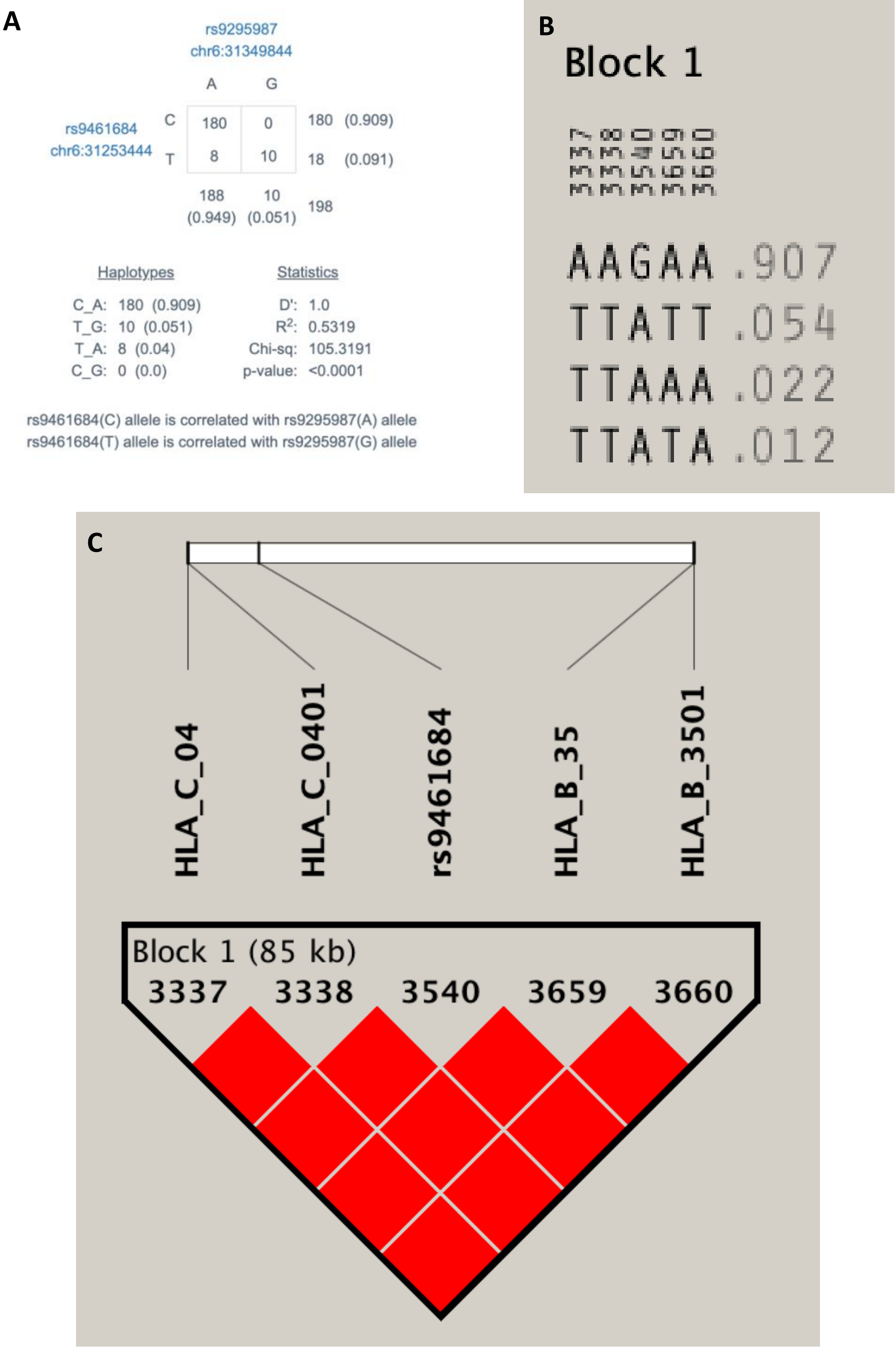
Gene	P Value	Effect size	Tissue
HLA-C	2.5E-13	0.50	Adipose - Subcutaneous
HLA-C	2.2E-10	0.34	Whole Blood
HLA-C	4.30E-06	0.38	Lung
HLA-B	7.80E-06	-0.28	Nerve - Tibial
HLA-C	8.50E-06	0.62	Spleen
HLA-C	9.90E-06	0.37	Adipose - Visceral (Omentum)
HLA-B	1.00E-04	-0.16	Artery - Tibial

Table 8: HLA specific results from the GTEx database for rs9295987

Gene	P-Value	Effect Size	Tissue	Source	Sample size
HLA-C	2.50E-55	-0.81	Blood	Lloyd-Jones (2017)	2765
HLA-B	9.17E-13	0.36	Blood-T cell CD4+ naive	Chen (2016)	197
HLA-C	6.38E-11	0.57	Adipose-Subcutaneous	GTEx Consortium (2015)	385
HLA-A	9.10E-09	-0.29	Blood	Lloyd-Jones (2017)	2765
HLA-H	1.70E-08	-0.29	Blood	Lloyd-Jones (2017)	2765
HLA-A	1.90E-07	-0.28	Blood	Lloyd-Jones (2017)	2765

Table 9: HLA specific results from the QTLbase database for rs9295987

Similarly with the main associated HLA SNP, Haploview analysis of the HapMap dataset was used to assess which HLA alleles were in an LD of D’ = 1 with these alleles. rs9295987 was not present in this reference panel dataset, but the proxy SNP rs9461684 was present (D’ = 1, R^2 = 0.53). The minor T allele of rs9461684 is correlated with the minor G allele of rs9295987 (Figure 2-3A). This SNP formed a haplotype with the HLA alleles HLA-C*04:01 and HLA-B*35:01 (Figure 2-3 B-C). Both of these HLA alleles were only present on the major A allele, indicated they are correlated with disease protection.



		rs9295987 chr6:31349844 A	G	
rs9461684 chr6:31253444	C	180	0	180 (0.909)
	T	8	10	18 (0.091)
		188 (0.949)	10 (0.051)	198

Haplotypes		Statistics	
C_A:	180 (0.909)	D':	1.0
T_G:	10 (0.051)	R^2:	0.5319
T_A:	8 (0.04)	Chi-sq:	105.3191
C_G:	0 (0.0)	p-value:	<0.0001

rs9461684(C) allele is correlated with rs9295987(A) allele
rs9461684(T) allele is correlated with rs9295987(G) allele

Figure 2-3: Haploview plots of HapMap HLA alleles in LD (D'=1) with the proxy for rs9295987, rs9461684.

(A) LDLink plot demonstrated the correlated alleles between rs9461684 and rs9295987. (B) Haploview haplotype plot showing which HLA alleles form haplotypes with rs9461684 alleles. The numbers 3337-3660 correlate with the order of HLA alleles demonstrated in the block shown in (C). 'A' indicates absence of HLA allele while 'T' indicates presence.

2.3.3 Imputation

The genotype data from the PD case-control dataset was used in a variety of imputation methods to predict which HLA alleles are associated with PD from the genotype data.

2.3.3.1 SNP2HLA

The SNP2HLA imputation strategy resulted in a total of 6,700 imputed alleles, including 180 2- and 4- digit classic alleles, 728 amino acids and 5,787 SNPs. The most associated alleles from the SNP2HLA association analysis ($P < 5 \times 10^{-5}$) are listed (Table 10).

The association analysis with imputed alleles differs from that described above for the overall GWAS. When SNPs were tested for association with PD, the minor vs the major allele are tested. This provides one allele that is associated with the risk group, and one allele associated with the protective group. At the HLA loci, there are potentially thousands of alleles to test. For each HLA allele, the presence (P) and absence (A) are tested for association with PD phenotype; in other words, one allele is tested against all other possible alleles. This is the same in alternative imputation association studies. This means the nature of each allele to confer risk or protection for PD can be identified by the odds ratio value (Table 10).

SNP	Effect allele	Non-effect allele	Frequency	Odds Ratio	P Value
rs3763316	T	C	0.33	0.85	6.26E-06
rs9268516	A	G	0.33	0.85	6.30E-06
rs3763311	A	G	0.32	0.86	6.80E-06
rs3793127	A	G	0.28	0.87	6.80E-06
rs3763309	A	C	0.28	0.87	6.82E-06
rs3763312	A	G	0.27	0.87	6.87E-06
rs2076520	T	C	0.28	0.87	7.97E-06
rs2076522	C	G	0.28	0.87	8.35E-06
rs7454108	C	T	0.10	0.83	1.02E-05
DRB1*13:01	P	A	0.30	0.83	1.49E-05
rs3998159	G	T	0.10	0.83	1.96E-05
rs3957148	C	T	0.10	0.83	2.08E-05
rs9275184	G	A	0.10	0.83	2.53E-05
rs3134996	T	A	0.64	0.89	3.65E-05

Table 10: Top associated alleles of the SNP2HLA imputed alleles ($P < 5 \times 10^{-5}$).

These results indicate that HLA-DRB1 is a locus of interest, with the allele HLA-DRB1*13:01 being the only classical HLA allele significantly associated with PD ($P = 1.49 \times 10^{-5}$, OR = 0.83) (Table 10). Presence of this allele is associated with PD protection.

The top associated imputed SNP from the SNP2HLA analysis, rs3763316, was then investigated for QTL properties. Results from the GTEx database indicate that rs3763316 could be an eQTL for HLA-DQA2, with the majority of results being for increased HLA-DQA2 expression (Table 11). However, the QTLbase results also indicate mixed results have been published, including HLA-DQB1 and HLA-DRA associations (Table 12). HLA-DQA2 is still the most significant association in this database.

Gene	P Value	Effect size	Tissue
HLA-DQA2	8.80E-55	0.83	Whole Blood
HLA-DQA2	2.10E-54	0.86	Muscle - Skeletal
HLA-DQA2	3.60E-52	0.96	Adipose - Subcutaneous
HLA-DQA2	4.10E-49	0.77	Skin - Sun Exposed (Lower leg)
HLA-DQA2	4.90E-46	0.9	Nerve - Tibial
HLA-DQA2	6.70E-46	0.84	Thyroid
HLA-DRB6	9.80E-45	0.65	Whole Blood
HLA-DQA2	2.20E-41	0.84	Lung
HLA-DQA2	8.60E-41	0.76	Artery - Tibial
HLA-DRB6	4.90E-40	0.71	Muscle - Skeletal

Table 11: Top 10 results from the GTEx database for rs3763316

Trait	P-Value	Effect Size	Tissue	Source	Samples size
HLA-DQA2	5.91E-201	-0.0864	Blood	Yao (2017)	5257
HLA-DQA2	5.91E-201	-0.0864	Blood	Yao (2017)	5257
HLA-DQA1	3.60E-177	0.8844	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	4.70E-177	0.8844	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	3.70E-103	0.6705	Blood	Lloyd-Jones (2017)	2765
HLA-DQB1	3.40E-74	0.5598	Blood	Lloyd-Jones (2017)	2765
HLA-DRB6	5.10E-68	-0.5432	Blood	Lloyd-Jones (2017)	2765
HLA-DRA	1.00E-55	-0.503	Blood	Lloyd-Jones (2017)	2765
HLA-DQA2	8.79E-55	0.8321	Blood	NA	

Table 12: Top 10 results from the QTLbase database for rs3763316

2.3.3.2 HIBAG

A total of 80 classical HLA alleles were imputed using the HIBAG method. The top 10 significantly associated alleles from the HIBAG association analysis are listed (Table 13). The HIBAG results indicate that the DRB1*04:01 allele is the only HLA allele significantly associated with PD (P = 1.56 x 10^{-6}, OR = -0.03) (Figure 2-4), with no other alleles passing the significance threshold.

Allele	P Value	OR
DRB1*04:01	1.56E-06	-0.03
DQB1*03:02	1.23E-04	-0.03
DQA1*03:01	1.77E-04	-0.03
DQA1*02:01	1.44E-03	0.02
DQA1*03:03	0.01	-0.03
DRB1*15:01	0.01	0.02
C*02:02	0.01	0.04
DQB1*05:01	0.01	0.02
C*14:02	0.02	0.63
DPB1*03:01	0.02	-0.02

Table 13: Top 10 HIBAG alleles associated with PD phenotype

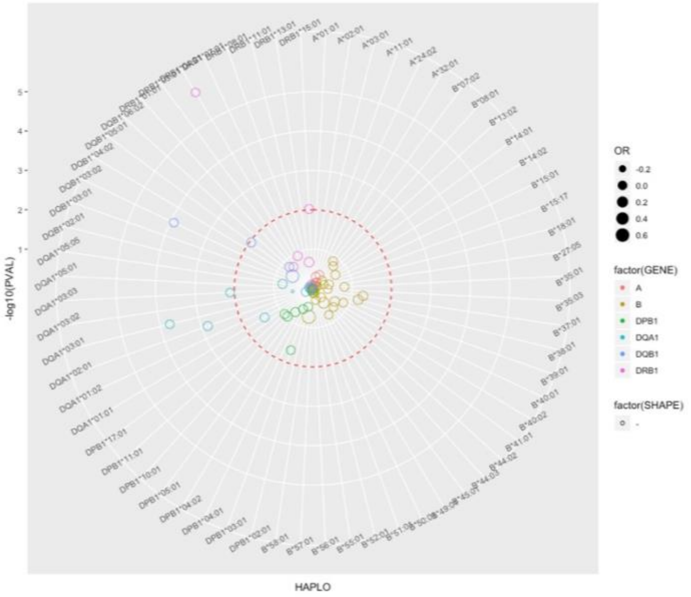


Figure 2-4: Plot of PD association of HLA alleles of HIBAG analysis.

2.3.3.3 DISH

Imputation using the DISH method resulted in a total of 4,441 imputed alleles. The top 20 associated alleles from the DISH analysis are listed (Table 14). The results from this imputation method indicate that HLA-DQA1 alleles and amino acids are most significantly associated with PD, with the top associated HLA allele being HLA-DQA1*03:01 (P = 3.25 x 10^{-5}). Alleles from HLA-B are also significantly associated with PD to a lesser degree by this analysis.

Variant	Position (hg19)	Effect allele	Non-effect allele	Z Score	P Value
SNP_DQA1_32713235	32605257	C	A	4.40	1.07E-05
AA_DQA1_-16_32713236_L	32605258	P	A	4.40	1.07E-05
AA_DQA1_-16_32713236_M	32605258	A	P	4.40	1.09E-05
SNP_DQA1_32717257	32609279	C	T	4.39	1.14E-05
AA_B_12_31432680	31324701	V	M	4.39	1.15E-05
SNP_B_31432681	31324702	C	T	4.39	1.15E-05
rs1048087	32609286	C	T	4.38	1.16E-05
SNP_DQA1_32717256_C	32609278	A	P	4.38	1.17E-05
AA_DQA1_69_32717257_L	32609279	A	P	4.38	1.17E-05
SNP_DQA1_32717264	32609286	C	T	4.38	1.17E-05
AA_DQA1_187_32718380_A	32610402	A	P	-4.16	3.24E-05
HLA_DQA1_03	32608306	P	A	-4.16	3.25E-05
HLA_DQA1_0301	32608306	P	A	-4.16	3.25E-05
SNP_DQA1_32717108	32609130	T	C	-4.16	3.25E-05
SNP_DQA1_32717120	32609142	G	C	-4.16	3.25E-05
AA_DQA1_26_32717128	32609150	S	T	-4.16	3.25E-05
SNP_DQA1_32717128	32609150	G	C	-4.16	3.25E-05
AA_DQA1_47_32717191_Q	32609213	P	A	-4.16	3.25E-05
SNP_DQA1_32717217_A	32609239	P	A	-4.16	3.25E-05
AA_DQA1_56_32717218_R	32609240	P	A	-4.16	3.25E-05

Table 14: Top 20 PD associated alleles from the DISH association analysis

2.3.4 Haploview

In order to investigate the correlation between SNPs that were associated with PD in the GWAS and HLA alleles that were associated following imputation, Haploview was used to measure the linkage disequilibrium (LD) between genotyped and imputed SNPs from the GWAS dataset, and imputed alleles, amino acids and SNPs from the SNP2HLA method. The top associated HLA SNP from the main GWAS (rs9268926), the top associated SNP from the conditional analysis (rs9295987), and the top meQTL (rs34039593), were measured for LD with all 180 2- and 4- digit classic alleles from the SNP2HLA imputation.

None of the HLA alleles passed the LD threshold of $R^2 = 0.7$ to indicate LD with the most associated SNP from the GWAS, rs9268926. However, it was in partial LD with three SNPs that were imputed by SNP2HLA; rs3793127, r s3763309, rs3763312 (Figure 2-5). The LD between rs9268926 and these three SNPs is D' = 0.8974, $R^2 = 0.7853$.

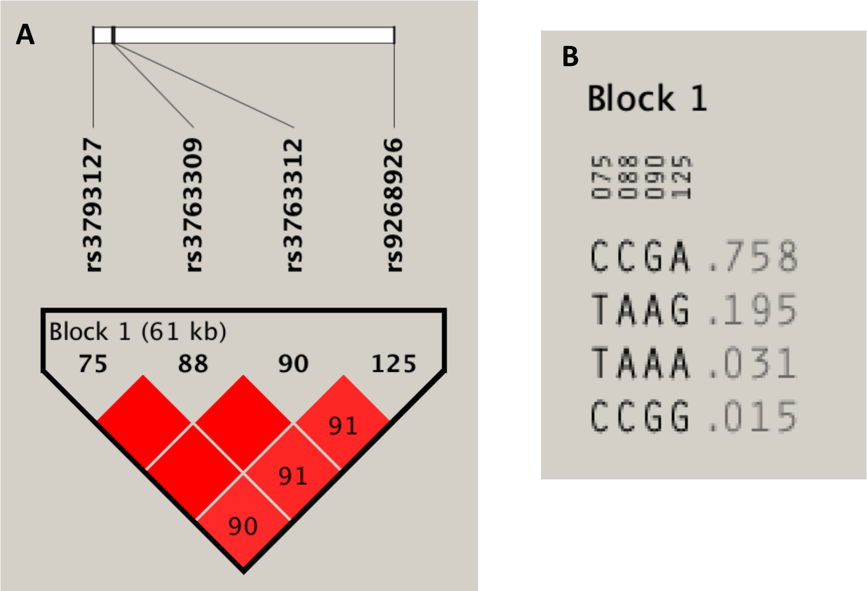


Figure 2-5: Haploview plot indicating LD between rs9268926 and SNP2HLA imputed SNPs rs3793127, rs3763309, rs3763312.

The minor G allele is correlated with the minor T, A and A alleles of rs3793127, rs3763309, and rs3763312 respectively.

The top meQTL from the 2019 meta-analysis was also in partial LD with these three SNPs (D' = 0.8935, R^2 = 0.7393). The top SNP from this meta-analysis, rs112485576 does not have an R^2 above 0.7 with any of the SNPs from SNP2HLA.

The most associated SNP from the conditional analysis, rs9295987, was observed to be in partial LD with four imputed alleles of class I loci (Table 15) (Figure 2-6).

Allele	R^2	D'
HLA_B_3501	0.633	0.934
HLA_B_35	0.633	0.934
HLA_C_0401	0.547	0.822
HLA_C_04	0.547	0.822

Table 15: SNP2HLA Imputed alleles in LD with rs9295987.

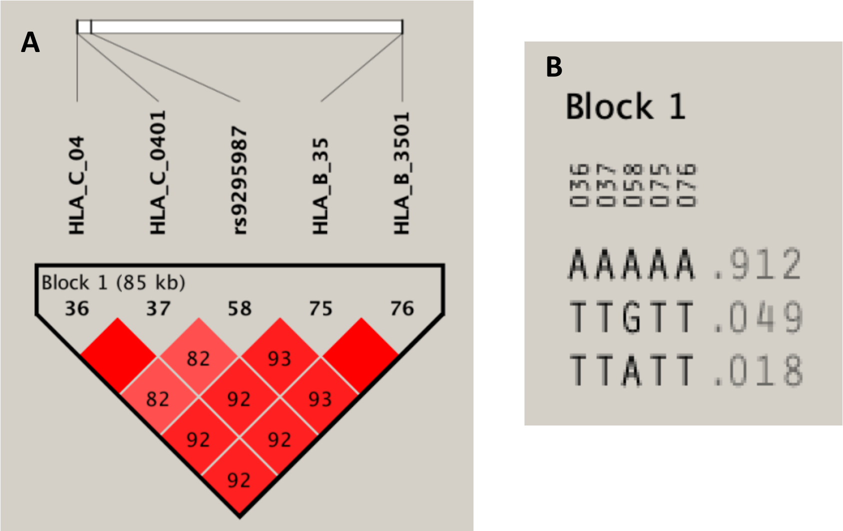


Figure 2-6: Haploview plot indicating LD between rs9295987 and SNP2HLA imputed HLA alleles.
rs9295987 is in LD with HLA_C_04:01 and HLA_B_35:01. The minor G allele is correlated with the presence of these alleles, while the major A allele is correlated with absence of these alleles.

The three SNP2HLA SNPs that were in LD with the top GWAS variant and 2019 meta-analysis meQTL were searched for published QTL properties. As these three SNPs are in strong LD, the results were similar. Both the GTEx and QTLbase databases indicated they could be eQTLs for HLA-DQA2 and/or HLA-DQA2. The top 10 GTEx results for each were all associations with HLA-DQA2 expression, and the top QTLbase result for each was association HLA-DQA1 expression (Tables 16-21).

Gene	P Value	Effect Size	Tissue
HLA-DQA2	1.2e-49	0.84	Whole Blood
HLA-DQA2	2.9e-47	0.86	Muscle - Skeletal
HLA-DQA2	3.2e-42	0.94	Adipose - Subcutaneous
HLA-DQA2	4.5e-42	0.77	Skin - Sun Exposed (Lower leg)
HLA-DQA2	1.8e-39	0.83	Thyroid
HLA-DQA2	2.7e-39	0.91	Nerve - Tibial
HLA-DQA2	4.1e-39	0.80	Artery - Tibial
HLA-DQA2	1.4e-36	0.85	Lung
HLA-DQA2	1.9e-35	0.91	Esophagus - Muscularis
HLA-DQA2	4.6e-31	0.75	Esophagus - Mucosa

Table 16: Top 10 results from the GTEx database for rs3793127

Trait	P Value	Effect Size	Tissue	Source	Sample Size
HLA-DQA1	3.30E-165	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	4.40E-165	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	4.50E-135	0.85	Blood	Lloyd-Jones (2017)	2765
HLA-DQA2	2.57E-120	-0.08	Blood	Yao (2017)	5257
HLA-DQA2	2.57E-120	-0.08	Blood	Yao (2017)	5257

Table 17: Top results from the QTLbase database for rs3793127

Gene	P Value	Effect Size	Tissue
HLA-DQA2	1.2e-49	0.84	Whole Blood
HLA-DQA2	2.9e-47	0.86	Muscle - Skeletal
HLA-DQA2	3.2e-42	0.94	Adipose - Subcutaneous
HLA-DQA2	4.5e-42	0.77	Skin - Sun Exposed (Lower leg)
HLA-DQA2	1.8e-39	0.83	Thyroid
HLA-DQA2	2.7e-39	0.91	Nerve - Tibial
HLA-DQA2	4.1e-39	0.80	Artery - Tibial
HLA-DQA2	1.4e-36	0.85	Lung
HLA-DQA2	1.9e-35	0.91	Esophagus - Muscularis
HLA-DQA2	4.6e-31	0.75	Esophagus - Mucosa

Table 18: Top 10 results from the GTEx database for rs3763309

Gene	P-Value	Effect Size	Tissue	Source	Sample Size
HLA-DQA1	5.00E-165	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	6.70E-165	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	4.70E-135	0.85	Blood	Lloyd-Jones (2017)	2765
HLA-DQA2	7.93E-121	-0.08	Blood	Yao (2017)	5257
HLA-DQA2	7.93E-121	-0.08	Blood	Yao (2017)	5257

Table 19: Top results from the QTLbase database for rs3763309

Gene	P Value	Effect Size	Tissue
HLA-DQA2	1.2e-49	0.84	Whole Blood
HLA-DQA2	2.9e-47	0.86	Muscle - Skeletal
HLA-DQA2	3.2e-42	0.94	Adipose - Subcutaneous
HLA-DQA2	4.5e-42	0.77	Skin - Sun Exposed (Lower leg)
HLA-DQA2	1.8e-39	0.83	Thyroid
HLA-DQA2	2.7e-39	0.91	Nerve - Tibial
HLA-DQA2	4.1e-39	0.80	Artery - Tibial
HLA-DQA2	1.4e-36	0.85	Lung
HLA-DQA2	1.9e-35	0.91	Esophagus - Muscularis
HLA-DQA2	4.6e-31	0.75	Esophagus - Mucosa

Table 20: Top 10 results from the GTEx database for rs3763312

Trait	P-Value	Effect Size	Tissue	Source	Sample Size
HLA-DQA1	1.10E-164	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	1.50E-164	0.94	Blood	Lloyd-Jones (2017)	2765
HLA-DQA1	1.70E-134	0.84	Blood	Lloyd-Jones (2017)	2765
HLA-DQA2	7.00E-121	-0.08	Blood	Yao (2017)	5257
HLA-DQA2	7.00E-121	-0.08	Blood	Yao (2017)	5257

Table 21: Top results from the QTLbase database for rs3763312

2.3.5 HLA Loci Selection

Based upon these results, the selected loci to perform long-read sequencing on were HLA-B, HLA-C, HLA-DQA1, and HLA-DRB1. (Table 22). Despite the evidence for HLA-DQA2 and HLA-DRB5 association, this was not included due to a lack of primers in the case of HLA-DQA2, and limited sample numbers carrying HLA-DRB5.

Loci	Reasons for Sequencing
HLA-B HLA-C	• Top SNP from conditional analysis is in partial LD with HLA-B/HLA-C alleles, and is a potential QTL for HLA-B/HLA-C
HLA-DQA1	• DISH results indicate most associated allele is HLA-DQA1*03 • Top SNP from SNP2HLA has mixed eQTL result for HLA-DQA1
HLA-DQA2	• Top association from GWAS could be an eQTL for HLA-DQA2 • Top variant from SNP2HLA had mixed eQTL result for HLA-DQA2
HLA-DRB1	• SNP2HLA results indicate most associated allele is HLA-DRB1*13:01 • HIBAG results indicate most associated allele is HLA-DRB1*04:01
HLA-DRB5	• Nalls (2019) meta-analysis results indicate top HLA SNP could be an eQTL for HLA-DRB5, and the top HLA meQTL is within HLA-DRB5

Table 22: HLA Loci selected for sequencing

2.4 Discussion

This chapter describes the investigation of PD associations within the HLA region using a range of bioinformatics techniques applied to a large case-control PD dataset. It is important to examine the demographic properties of the samples prior to bioinformatic investigations. Age, sex, and genetic ancestry are all sample properties that can have an impact on the detection of genetic factors associated with PD. Age and sex are both risk factors for PD; the disease twice as common in men than women (3,94), and primarily affects the elderly population.

As far as possible, age and sex of samples should be matched across cases and controls, with samples differing only by the phenotype, as these covariates can explain some of the phenotypic variation. However, this is not often possible; as seen in the samples used here, whilst there is a relatively similar age and sex divide, there is a higher proportion of men in cases (59%) compared to controls (47%). There is also some missing demographic data, which reduces the ability to precisely control these factors. Correcting for age and sex here will increase the precision of the study, accounting for sample stratification and reducing the residual variance of the outcome, which could increase statistical power to detect effect size of true associations.

However, the differences here still introduce some potential for issues. Conversely, correcting for known covariates can also reduce the power to detect associations in rare diseases. One study found when conducting GWAS of rare diseases (prevalence <20%), controlling for known covariates reduced the power to detect genome wide significant associations whilst the opposite was the case for common diseases (95). Whilst this is the case, type I error rate would also be higher when not controlling for known covariates. Having a more closely age and sex matched sample would improve this study and remove the uncertainty of whether controlling for these characteristics will negatively affect power.

Whilst age and sex can be described as non-confounding covariates as they are not confounders (associated with both the genotype and the disease outcome), genetic ancestry is a confounding factor that can explain some phenotypic variation. Not adjusting for this will increase the type I error rate by leading to false associations between the disease status and SNPs whose frequency is dependent upon ancestry. Controlling for the principal components associated with disease status as has been done here will avoid the confounding effect of population structure and is a necessary step in the quality control process. Therefore whilst quality control procedures could have been improved, the most appropriate steps were taken to control for confounding factors within the dataset used here.

Firstly, a GWAS and conditional analysis were performed, identifying a PD associated HLA SNP that is in partial LD with the top HLA SNP from the latest PD meta-analysis, indicating the same association could be identified. Further exploration of the QTL properties of both the top HLA SNP from the current data and from the meta-analysis indicated that these had ambiguous QTL properties, but both showed associations with HLA-DQA1/DQA2 loci as well as the HLA-DRB1 loci. Furthermore, Haploview analysis indicated HLA-DRB1*04:01 is correlated with the protective allele, whilst HLA-DRB1*15, HLA-DQA1*01, and HLA-DQA1*05 are correlated with the PD risk allele. These results indicated that this association has broader associations than the HLA-DRB5 locus, as suggested in the meta-analysis. The conditional analysis revealed a further independent association at the HLA locus, which upon exploration was found to be associated with class I loci HLA-C and potentially HLA-B, rather than the established class II association.

These results are in agreement with previous studies that have identified a class I HLA association with PD as well as class II, such as in the original 2013 imputation study. In comparison, the more recent imputation results published during this investigation suggest class II loci alone were associated with PD risk, and that the HLA-DRB1*04 allele was driving this association, with no class I association (87). However, the present results indicate it can be worthwhile to continue the investigation of a broader class of HLA loci associations, particularly when basing evidence upon imputation studies which may be more likely to be flawed in their conclusions compared to sequencing studies.

Subsequently, three separate HLA imputation methods were applied to this dataset for comparison. SNP2HLA and HIBAG have been applied to different PD datasets previously, yet the more novel method DISH has not previously been used to assess PD HLA associations. Within the GWAS and conditional analysis, both minor and major SNP alleles were tested for association with PD, providing one allele that is associated risk, and one allele associated with protection. At the HLA loci, up to thousands of potential alleles were tested for PD association, with each tested for the presence (P) against the absence (A). This allowed identification of further HLA alleles correlated with PD risk and protection.

Results from the SNP2HLA method indicated that the HLA-DRB1*13:01 allele was associated with PD risk. Whilst this allele has not been indicated previously to be associated with PD, HLA-DRB1*13 alleles have been shown to have a protective effect in neurodegenerative disease. In a recent study of age-related brain atrophy, absence of the DRB1*13:02 allele was found to be significantly association with reduction in total grey matter, cerebrocortical grey matter, and subcortical grey matter in a small number of healthy individuals (96). Consequently, this allele was then also found to have a protective effect against dementia in a larger study of western European samples (97). As the observed PD association of DQA1*13 alleles in this data is protective, this imputation result indicates there could also be a protective effect of this allele against PD neurodegeneration. However, this result would need to be repeated in order to assess the significance in PD, as this could be a result of imputation error.

Alternatively, HIBAG imputation results replicated an association from previous studies indicating HLA-DRB1*04:01 is the significant DRB1 allele of interest, which also had a protective effect. The results from the DISH imputation on the other hand indicated that the HLA-DQA1*03 allele was the most associated, which has also been observed in the previous PD imputation studies (87). This result could be due to the high LD between the DQA1*03 and DRB1*04 alleles, as suggested in the previous imputation study.

Different computational approaches have been taken with each of these imputation methods, and the results here demonstrate that alterations in these methods or reference panels applied to the same dataset can yield varying results. Due to the imperfect nature of the imputation approach and inability to identify novel alleles, inaccuracies can be introduced. This is especially the case with the class II alleles, and it has previously been observed in a PD imputation study that the DRB1 locus was most poorly imputed compared to others (81). Given that this is the case for the most commonly associated PD HLA allele, it is not appropriate to rely on one method to provide a reliable association at this locus. The comparison of imputation results here has indicated that it is worthwhile to follow up both DRB1 and DQA1 loci, and that alternative sequencing methods are necessary to identify the specific PD associated alleles. Any rare or non-classical HLA alleles that are correlated with PD risk and protection, but not present in the reference panels used here, would also not have been identified in this investigation. Therefore, sequencing is still required to best identify the most significant alleles of interest.

Applying a combination of GWAS, imputation and QTL property analysis to this PD dataset has outlined the most significant HLA loci that are candidates for driving the PD HLA risk, replicating certain established associations, and revealing potential new associations. Consequently, it was decided that the class I alleles HLA-B and HLA-C, and the class II alleles HLA-DRB1 and HLA-DQA1, are the loci that will be taken forward to further investigation with a long-read sequencing approach using the PacBio method.

3 Long-Read Sequencing of the HLA Locus

3.1 Introduction

3.1.1 HLA Imputation

HLA imputation is the process by which a samples HLA alleles are predicted using common SNP data. This has been the primary method applied in large PD datasets to identify HLA alleles associated with risk of disease (87,98,99). This approach has the potential to achieve high levels of accuracy at resolving risk alleles, with the popular SNP2HLA imputation method applied in the previous chapter reporting an average accuracy of 96.7% at 4-field resolution (100). As imputation methods can be applied to large cohorts at low cost, they are instrumental to conducting well powered HLA association analysis and fine mapping HLA disease risk.

Certain conditions must be met to ensure high quality imputation. Due to the significant differences in HLA allele frequencies between different ancestral populations, HLA reference panels must be well-matched to the sample population. Issues can arise when there is broader allele diversity in the sample than in the reference panel. To demonstrate, imputation of HLA alleles in a cohort from the Human Genome Diversity Project was conducted using a variety of different methods including HIBAG and HLA*IMP:02, each using a European reference panel from the 1000 genomes project (101). Correct imputation across all HLA loci was achieved in less than 27.8% of the sample, with most imputation errors occurring at HLA-B and HLA-DRB1. This indicates the difficulties faced with imputing HLA alleles in more diverse samples using currently available panels. In the case of imputing alleles in PD datasets, this issue could lead to errors if allele frequency in disease differs from standard European populations.

The difficulty with imputation of specific HLA loci, such as HLA-DRB1, is reflected in other HLA imputation software. For example, HLA*IMP and SNP2HLA reported accuracy of up to 98% at individual loci with appropriate reference panel of 5,225 individuals, yet HLA-DRB1 alleles were consistently the least well imputed across both methods, with 92% and 93% accuracy respectively (100). HLA-DRB1 was also found to be the most poorly imputed of the HLA alleles in a previous PD HLA association analysis, which implemented a 1000 genomes sample reference panel of 1,092 individuals (81). Given the suggested significance of the HLA-DRB1 and potentially HLA-B loci in PD HLA associations discussed in the previous chapter, this limitation of HLA imputation approaches can be particularly restrictive in fine mapping the PD HLA risk.

Where reference panels are well matched and of appropriate size, imputation software is still limited in that it can only impute alleles that are present in the reference panel. No association information will be available for rare or novel alleles in the samples, which can introduce inaccuracies when disease risk is associated with alleles not commonly observed in the population.

The current largest reference panel is part of the lately released HLA-TAPAS software, released in October 2021, which provides a multi-ancestry panel consisting of 21,546 diverse individuals (102). This method demonstrated high accuracy imputation at G-group (2-field) resolution, with the multi-ancestry nature addressing issues with poor imputation of alleles not commonly observed in European ancestry populations. However, this method is still limited to identifying G-group sequence of known alleles, meaning it has limited use in identifying variants outside of the binding domain and is unable to identify novel variants. To achieve unambiguous HLA typing at full four-field resolution, methods outside of imputation must be considered.

3.1.2 Sequencing of the HLA Region

In comparison to imputation, sequencing of the HLA region can provide information on full length alleles and infer novel or rare variants. Different sequencing approaches are used to achieve this.

Next generation sequencing (NGS) methods such as Illumina sequencing are popular for HLA sequencing. A short-read approach is taken that either uses short-range PCR to produce exon only amplicons, or long-range PCR to cover multiple exons, which are then fragmented to into stretches of approximately 500bp. Reads of these short fragments are then either mapped individually onto reference alleles from an HLA database to identify the best match, or de novo consensus sequences are formed which are then fully mapped as one to the best match allele; the former is the most common approach.

NGS provides an improvement on the previously used Sanger sequencing method (Figure 3-1). This method resulted in mixed data from the two alleles, limiting accurate phasing of alleles. In contrast, the clonal nature of NGS allows for resolution of phasing ambiguities, as multiple overlapping short reads can provide high coverage and be sorted into phased alleles (103). However, there are still major limitations for using NGS approaches to sequence HLA regions.

HLA genes are long and highly repetitive, with 49.5% of the HLA genomic sequence composed of interspersed repeat elements (49). When polymorphisms are separated by long stretches of SNP poor regions and cannot be covered by the same short read or read pair, allele phasing becomes more challenging and results in inaccuracies with this approach. This issue is particularly relevant when it comes to phasing alleles of the HLA locus, as these contain long repetitive regions separating variants, resulting in poor phase resolution between the two chromosomes (103).

The HLA region also contains multiple paralogues of several genes, and there are highly conserved repetitive regions shared between alleles and different loci. This presents another challenge when mapping of short reads to correct regions, as they can correctly map to many regions or alleles. This can be an added issue when attempting to characterise the whole allele and not just the core exons, as non-coding regions include long homopolymer stretches and short tandem repeats (STRs); these not only provide multiple stretches for short reads to map to, but if a read does not cover the whole region of an STR then it can be difficult to ascertain the exact length.

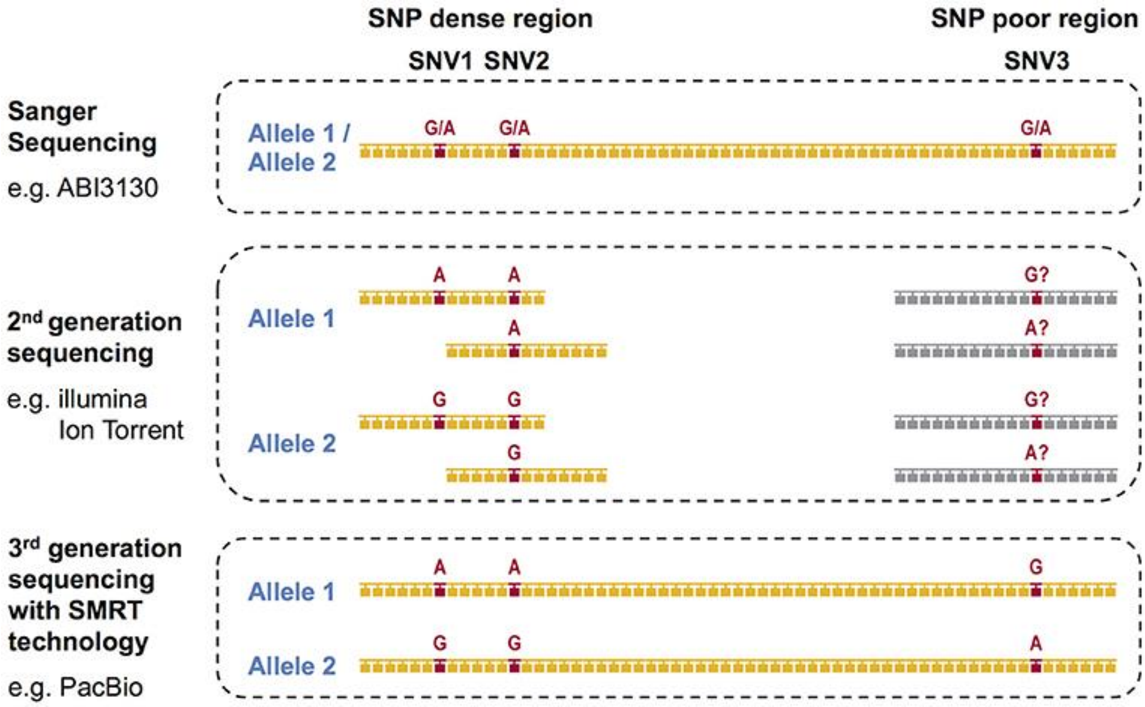


Figure 3-1: HLA sequencing methods

Demonstration of ease of unambiguous allele phasing using PacBio sequencing approach in comparison to sanger sequencing or NGS. The SNP poor regions demonstrated here for which NGS methods struggle to phase are long in HLA-DRB1 and other class II alleles. From Suzuki et al (2018)

Some HLA read mapping software tackle these computational issues by discarding reads that do not uniquely map, which results in errors by biasing against alleles containing conserved sequences. Similarly, duplicated reads can also be removed, which can bias results by reducing coverage of repetitive sequences. Some approaches consider only exact read matches, but this can result in incorrect results if the allele is novel. By ignoring rare alleles (no frequency on AlleleFrequency.net), or partially known alleles in the database, mapping software can address the issue of the high number of alleles to sort through, however this will bias against rare alleles which could have biological importance. Many algorithms only report at G group resolution to reduce computational issues with long stretches in non-coding regions. This can be most useful in clinical contexts, but less so when sequencing to identify disease associated variants. Current computational approaches for short-read mapping have limited ability to address the range of potential issues with this method (103).

The main issues associated with NGS sequencing can be addressed with long-read sequencing, using technologies developed by companies such as Pacific Biosciences (PacBio) or Oxford Nanopore. This approach involves generating long-range PCR products covering the entirety (or majority) of the gene, then sequencing these amplicons in full without breaking into short segments. This is the only approach capable of providing fully phased, unambiguous reads at 4-field resolution. Reads cannot be assigned to the incorrect allele, and length and position of repetitive sequences require no extra computational approaches to determine (104). The majority of entries in the IGMT HLA database are now being sequenced with the PacBio system, which has allowed resolution of discrepancies and identification of novel intronic and exonic polymorphisms of known alleles (105). The application of this HLA typing method has significant consequences for the understanding of HLA risk alleles for certain inflammatory disease, such as identification of an intronic risk SNP in HLA-DRB1 for rheumatoid arthritis (104), amongst many other conditions. For example, an investigation into the HLA-G association with preeclampsia (PE) used PacBio sequencing of the whole locus to determine genetic associations. Two novel alleles at four-field resolution were identified, as well as a poly-T stretch downstream of HLA-G, the length of which appeared to be associated with onset of maternal PE (106). Short-read sequencing cannot achieve the same full-length high-resolution HLA reads to enable such discoveries, especially when it comes to correctly identifying the length of repetitive sequences. PacBio HLA typing has also been demonstrated to provide significant improvement for donor matching, with the accuracy from PacBio reads resulting in improved survival rates for patients receiving hematopoietic cell transplantation (107).

Long-read sequencing is a more expensive method than other sequencing approaches, often resulting in fewer subjects being tested. As well as short-read sequencing, mapping long-reads also relies on incomplete database reads to assign alleles to samples. However, discrepancies between sequencing data and database allele records can be more easily identified through long-read sequencing, and the ability to overcome phasing ambiguities is important for novel allele discovery in disease association.

3.1.3 PD HLA Sequencing

At the onset of this investigation, limited studies had been conducted which involved sequencing of the HLA region in PD samples. The largest sequencing study that aimed to characterise the HLA alleles associated with PD involved sequencing of 11 HLA genes in 1,597 PD cases and 1,606 controls (108). Results indicated that HLA-DRB1*04:01 and HLA-DQB1*03:02 alleles had protective effects, with these alleles in high LD forming part of the same haplotype; HLA-DQA1*03:01~HLA-DQB1*03:02~HLA-DRB1*04:01. DRB1*04:01 was identified as the source of this protective effect, with other haplotypes containing this allele also strongly protective. HLA-DRB1*01:01 was also found to be a borderline risk allele, in strong LD with HLA- DQA1*01:01 and HLA-DQB1*05:01 as a risk haplotype. However, no interaction between these associated haplotypes and known PD risk SNPs was identified.

The amino acid residues within DRB1 were then focused on for functional investigation. Four positions were identified, three of which are in the binding motif, that were significantly associated with PD: 11-V, 13-H, and 26-F. The protective effects of these were due to their association with HLA-DRB1*04, whilst a risk amino acid 11-L was similarly specific to HLA-DRB1*01. A set of amino acids (shared epitope) at positions 70-74 was identified, which is shared across HLA-DRB1*01:01, HLA-DRB1*04:01 and HLA-DRB1*10:01. It was observed that only HLA-DRB1*04 alleles with the SE were associated with reduced risk for PD. This SE in combination with 11-V explained the protective effects of DRB1 alleles, while the SE without 11-V conferred risk (108).

Though the risk alleles and their functional properties identified here could explain some of the PD HLA risk, the fact that the alleles identified in this investigation were not associated with any established PD HLA risk SNPs at the time suggests this does not explain the entire HLA risk. There are likely to be other alleles that individually or collectively confer PD risk. A larger sample and improved methods of HLA typing could contribute to expanding the understanding of these risk alleles. Furthermore, this study implemented a short-read sequencing approach with long-range PCR products broken down into 300-500bp fragments, with reads then compared to multiple HLA references and the matching allele manually selected. As discussed, any novel variants would be ambiguously phased by this approach, and identification of structural variants that differ from alleles within the reference database would be limited.

Identifying the HLA alleles that confer risk for PD is important to understand the functional consequence of this established risk locus. The best approaches to typing HLA alleles must be implemented to understand those variants that could have a role in PD pathology. In particular, the HLA alleles tagged by the most recent HLA risk variant identified in the large-scale meta-analysis should be identified. To best identify any potential novel alleles, long-read sequencing of this region is preferred.

3.1.4 Aims

Performing PacBio long-read sequencing can aid unambiguous identification of HLA alleles that are associated with the established PD risk variants, and what functional qualities of these alleles could potentially explain this risk. This investigation aims to conduct an exploratory analysis of variants within the HLA region of PD samples, in particular examining whether the top PD associated HLA SNP from the latest meta-analysis GWAS (Nalls 2019) is associated with a particular HLA variant.

Due to the recent release of a large and diverse HLA imputation panel which improves upon the accuracy of methods applied in the initial study, this investigation is also able to implement up to date HLA imputation with a large multi-ancestry reference panel that has not yet been applied to a PD dataset. Alongside the long-read sequencing approach, this will allow for a comparison of methods within the sequenced sample. Furthermore, improved imputation will be to the larger case-control dataset to investigate HLA alleles independent of the Nalls 2019 top PD associated HLA SNP.

The aims for this investigation are as follows:

1. To generate PacBio long-read sequencing data for the HLA loci that were previously identified as potential risk loci; HLA-B, HLA-C, HLA-DQA1, HLA-DRB1.
2. Compare long-read data for these loci in PD patients that carry the risk or protective allele for the top PD associated HLA SNP identified in the 2019 meta-analysis.
3. If associated alleles are identified, establish which of these can also be identified as PD associated in a larger case-control dataset, and so further explore associated alleles independent of this SNP.

3.2 Methods

3.2.1 PacBio Samples

PD patient DNA samples were selected from the Proband cohort (109). Within the cohort, samples within the recent onset (<3 years) group were selected. Samples were either homozygous for the risk allele (C) of the top HLA SNP from the Nalls 2019 GWAS (rs112485576), or the protective allele (A). A total of 70 samples were sequenced, 31 homozygous for C and 39 homozygous for A. Age of onset and sex were evenly distributed across the groups (Table 23)

rs112485576 genotype	No. of samples	Age at onset average (years)	Percent male (%)
C/C	31	66	45
A/A	39	66.4	51
All	70	66	50

Table 23: Characteristics of PD samples selected for long-read sequencing

3.2.2 Amplicon preparation

HLA-B, HLA-C, HLA-DRB1, HLA-DQA1 loci were amplified using GenDx NGSgo-AmpX or NGSgo-AmpX v2 primers. The method for selection of these HLA loci based on PD HLA association data and QTL data is described in Chapter 2. Separate PCR reactions amplified each of the HLA loci. Primers amplified the entirety of the HLA-B, HLA-C and HLA-DQA1 loci in one reaction, producing amplicons of 3.4 kb, 3.4 kb and 5.8 kb respectively. To amplify the entirety of the DRB1 locus, two primer pairs were used which amplified exon 1 and exons 2-6 respectively, resulting in 2 amplicons of approximately 2.5 and 5 kb in length covering the whole gene. Figure 3-2 demonstrates the amplicon ranges.

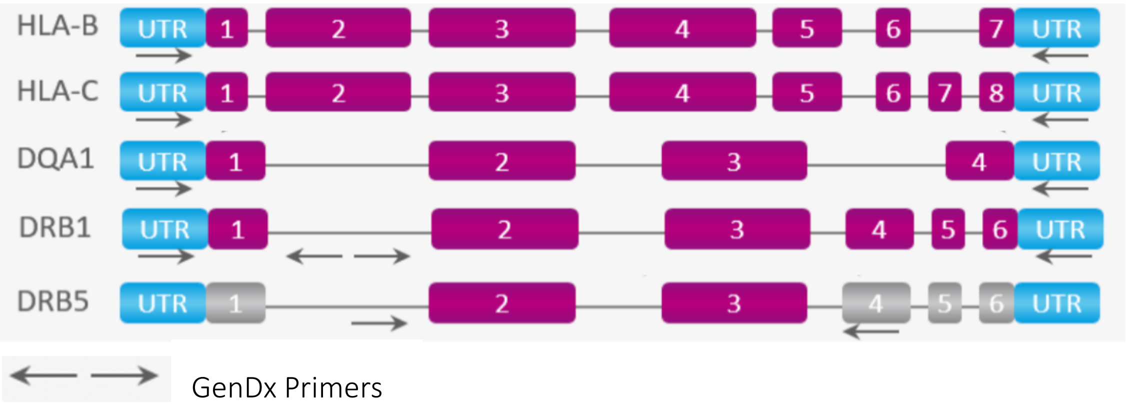


Figure 3-2: Length of amplicon products from PCR reactions to be processed for PacBio library.
Adapted from https://www.gendx.com/product_line/ngsgo-ampx-v2/.

Each PCR reaction contained approximately 40ng genomic DNA, or 60ng for HLA-DRB1 amplification. HLA-B and HLA-C were amplified using the following PCR settings: initial denaturation of 95°C/3 min, followed by 25 cycles of 95°C/15 sec, 65°C/30 sec, 67°C/4 min, followed by final extension of 67°C/10 min. For HLA-DRB1 and HLA-DQA1 the following touchdown PCR protocol was used: initial denaturation of 95°C/3 min, followed by 10 cycles of 95°C/15 sec, 68 → 63°C/30 sec touchdown, 67°C/5 min, followed by 15 cycles of 95°C/15 sec, 63°C/30 sec, 67°C/5 min, followed by final extension of 67°C/10 min. PCRs were carried out in Bio Rad T100 Thermal Cyclers.

Success of amplification was determined using gel electrophoresis. 0.5µl of PCR product was loaded into a 0.5% agarose gel and run at 80V for 4 hours, alongside a 10kb ladder. Figure 3-3 demonstrates an example of PCR product separation

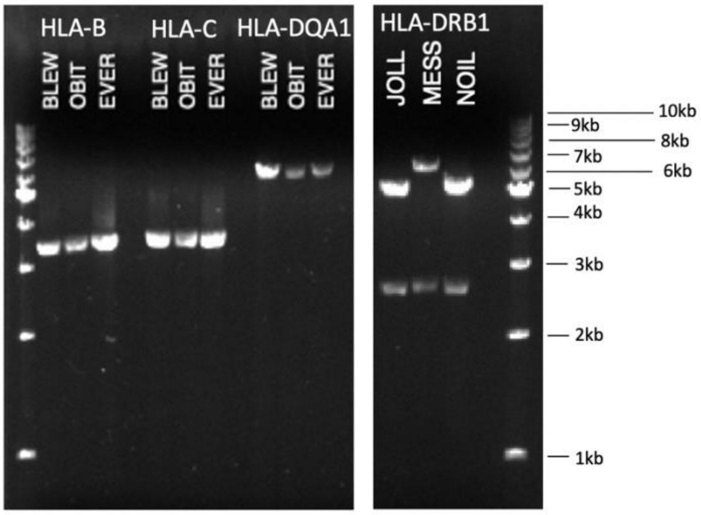


Figure 3-3: Gel electrophoresis of PCR products from amplification of HLA loci.

3.2.3 PacBio library preparation

Amplicons were first purified using AMPure PB beads (Pacific Biosciences) to remove excess PCR reaction materials, at a ratio of 0.45X or 0.6X for DRB1 PCR products. Purified samples were then quantified using a PicoGreen assay (Thermo Fisher) using a Tecan Fluorometer plate reader. The purified amplicons for each sample were then pooled following an equimolar pooling strategy, and further purified with AMPure PB beads 0.6X to achieve a total volume of 5µl for each sample containing at least 250ng.

Each sample underwent library preparation using the PacBio SMRTbell Express Template Prep Kit 2.0. Firstly, DNA damage repair was conducted, followed by end repair and

A-tailing, and barcoded adapter ligation. Barcoded adapters from the PacBio Barcoded Overhang Adapter kits 8A and 8B were used, which result in a circular DNA SMRTBell template appropriate for PacBio sequencing (Figure 3-4). Barcoded samples were pooled into libraries of 16 samples, and then each library was further purified with 0.6X AMPure PB beads. Final libraries were quantified using the Qubit dsDNA BR Assay Kit (ThermoFisher) with the Qubit Fluorometer and 260/230 ratios verified using the NanoDrop.

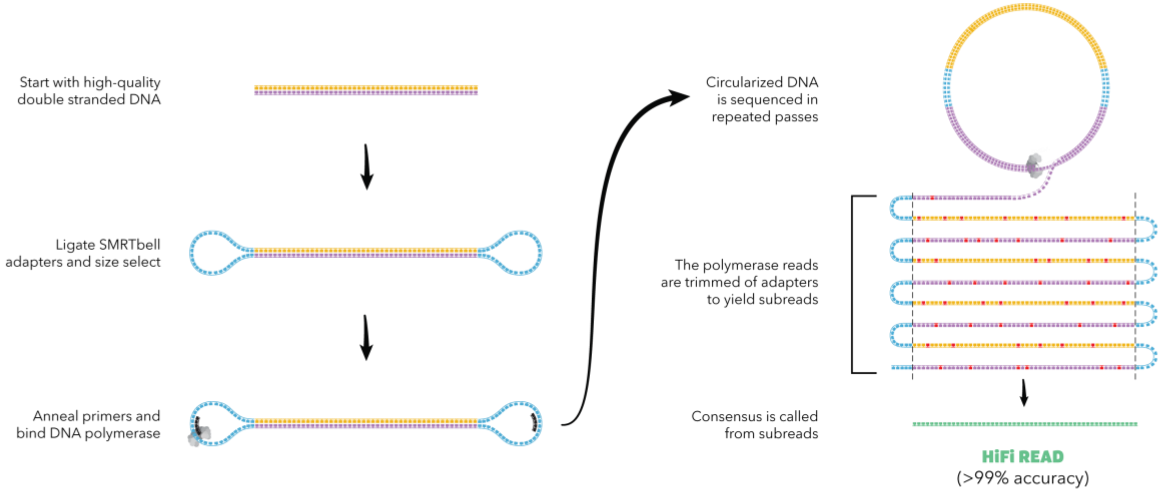

Start with high-quality double stranded DNA
Ligate SMRTbell adapters and size select
Anneal primers and bind DNA polymerase
Circularized DNA is sequenced in repeated passes
The polymerase reads are trimmed of adapters to yield subreads
Consensus is called from subreads
HiFi READ
(>99% accuracy)

Figure 3-4: Generation of PacBio SMRTbell template appropriate for long-read sequencing.
From https://www.pacb.com/technology/hifi-sequencing/.

3.2.4 PacBio Sequencing

Libraries were sequenced on the PacBio Sequel II system. This system uses Single Molecule, Real-Time (SMRT) long-read sequencing technology in which amplicons are ligated to adapters to create a circular SMRTbell template. DNA polymerase is bound to the template before addition to the SMRT Cell, which contain millions of ‘zero mode waveguides’ (ZMWs). Each ZMW holds a single molecule SMRTbell template. The polymerase is fixed to the bottom of the ZMW, where it incorporates nucleotides to the circular template. As each nucleotide is added a light is emitted corresponding to the base, which allows for real time sequencing of the library (Figure 3-5).

First, libraries were bound to the Sequel polymerase using Sequel Binding Kit 3.0. An internal control of 2kb bound to the sequel polymerase was included. Bound libraries were then sequenced using the Sequel Sequencing Kit reagent plate on a Sequel SMRT Cell 1M v3. Sequencing was run for 10 hours of collection time, producing circular consensus reads.

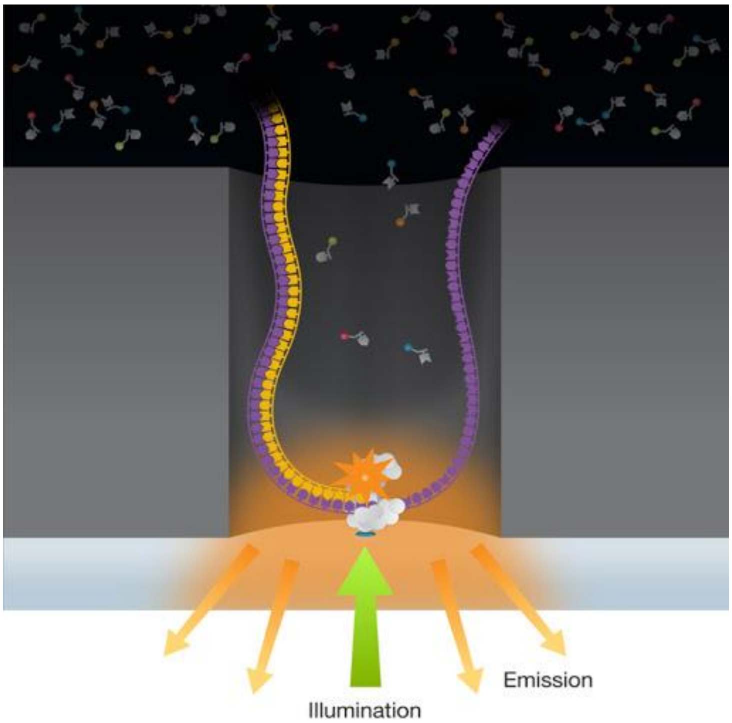


Figure 3-5: SMRT Template sequencing within the Zero Mode Waveguides.

3.2.5 Analysis of PacBio data

SMRTLink platform version 10.2.0.133434 was used to assess quality (Q20 score) and read length and depth for all samples. PacBio data was exported as FastQ files and analysed using the GenDx NGSEngine software version 2.23. The settings used consisted of ‘PacBio subread’ as sequencing instrument, with no quality trimming measures. The cluster phasing algorithm was used for all samples except HLA-DRB1, where classic phasing algorithm was used. A maximum of 10,000 reads per sample were analysed. Reads were mapped to the IGMT HLA database version 3.43.

NGS engine reports the best match allele from the database for each sample, also showing all incomplete allele entries that exactly match the data and further close matches. Where novel alleles occur, the best match alleles are presented with the number of mismatches between the data and database allele record. These are listed in order of best match, with a scoring strategy that penalises mismatches in ARD (antigen recognition domain) regions higher than other exons and in other exons higher than non-coding regions. The best match allele provided from this analysis was used for association analysis.

3.2.6 Association analysis

Using R studio version 1.4.1106 Fishers Exact test was applied to test if there was a significant difference between the total number of each HLA allele carried between the rs112485576 genotype groups.

3.2.7 HLA Imputation

Imputation of the HLA alleles for each sequenced sample and for the PD case-control dataset consisting of 5,322 cases and 10,018 controls (described in Chapter 2) was conducted using the full multi-ancestry reference panel of 21,546 individuals available on the Michigan imputation server (102). This imputation approach is a recently available improvement upon the imputation approaches applied in the previous chapter, allowing for comparison between sequencing and imputation to a more accurate level. Pre-imputation quality control for the case-control sample was applied by filtering samples for missingness > 0.02, minor allele frequency (MAF) < 0.01, and HWE P value < 1e−6, and excluding individuals according to missing genotype data (mind) > 0.01.

3.2.8 Association analysis and conditional analysis

Association of imputed HLA alleles with PD in the case-control dataset was measured by logistic regression in Plink version 1.9. Regression was performed with principal components associated with PD phenotype and sex as covariates, previously described in Chapter 2. Conditional analysis was then using the same approach with specific HLA alleles as a condition.

3.3 Results

3.3.1 HLA Allele Determination

PacBio long-read sequencing of full-length HLA-B, HLA-C, HLA-DQA1 and HLA-DRB1 loci was successful for 70 samples. The mean barcode quality (Q score) for all samples was between 90-96, indicating high single-molecule accuracy and a base call error rate of less than 0.0001%. Average read depth was 1,460, 812, 7,578 and 2,600 for HLA-B, HLA-C, HLA-DQA1 and HLA-DRB1 respectively. The average read lengths were 3,365, 3,397, 4,812, and 5,648 kb for HLA-B, HLA-C, HLA-DQA1 and HLA-DRB1 respectively.

Reads were mapped to IGMT HLA database using GenDx NGSengine. Most sequences matched known HLA alleles within the database. Extrapolation was applied in several cases where for example there is only allele data for the core exons that encode the binding pocket, and sequence data from a similar allele is used to match the whole read. The best match alleles for all samples at four-field resolution are displayed in Supplementary Table 1.

3.3.1.1 Novel Allele Identification

Most alleles sequenced completely matched with the database entry HLA alleles, with no novel indels, repeat expansions, inversions or other variants detected. The one novel allele identified was at the HLA-DQA1 locus of a sample in the PD risk allele group. The best match allele was DQA1*01:01:01:01, yet the sample differed from the entry for this allele by two SNPs within the first intron; a G to A change at gDNA position 3828 (rs9272687) and an A to T change at position 3832 (rs28654242). These were upstream of the second exon, which begins at gDNA 3858. rs28654242 has been associated with an increased chance of immune-related adverse events in patients with melanoma (110). With this being the only example of novel variation at this locus, it is not possible to suggest whether this is associated with any risk or protective group in PD.

3.3.2 Identification of HLA Alleles correlated with PD associated SNP (rs112485576)

To determine which 2-field HLA alleles were correlated with the risk and protective allele of rs112485576, a Fishers Exact test was performed for each (Supplementary Table 2). Those alleles that showed a significant difference between C/C (PD protective) and A/A (PD risk) groups in which sequencing was conducted (Fishers Test P value < 0.05) are listed in Table 24.

These results indicate that the alleles that show significantly higher representation within the risk group are: B*44:03, C*07:01, DQA1*01:01, DQA1*01:02, DQA1*02:01, DQA1*04:01, DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*11:01, DRB1*15:01. The alleles that are significantly higher within the protective group are: B*40:01, B*44:02, C*03:04, DQA1*03:01, DQA1*03:03, DRB1*04:01, DRB1*04:04.

Locus	Allele	No. of allele copies		Fishers Exact P Val
		Risk group	Protective group	
B	B*40:01	3	13	3.36E-02
	B*44:02	4	19	5.30E-03
	B*44:03	6	0	6.60E-03
C	C*03:04	4	15	4.48E-02
	C*07:01	9	3	3.34E-02
DQA1	DQA1*01:01	9	0	5.00E-04
	DQA1*01:02	12	1	2.00E-04
	DQA1*02:01	11	0	1.00E-04
	DQA1*03:01	0	38	< 0.00001
	DQA1*03:03	1	38	< 0.00001
	DQA1*04:01	4	0	3.64E-02
	DQA1*05:01	12	0	< 0.00001
DRB1	DRB1*01:01	9	0	5.00E-04
	DRB1*03:01	13	0	< 0.00001
	DRB1*04:01	2	39	< 0.00001
	DRB1*04:04	0	24	< 0.00001
	DRB1*07:01	9	0	5.00E-04
	DRB1*11:01	5	0	1.55E-02
	DRB1*15:01	12	0	< 0.00001

Table 24: Distribution of HLA alleles in PacBio dataset

Distribution of alleles between the risk group (Nalls SNP A/A) and protective group (Nalls SNP C/C) of all alleles with significantly different allele distribution. Fishers Exact test P val < 0.05 indicates significant difference between the risk and protective groups. Full allele distribution between groups is provided in Supplementary Table 2.

3.3.3 Comparison of HLA Imputation Using the Most Recent Reference Panel with Best-Match Alleles Determined by PacBio Sequencing

For many HLA association studies, to maximise sample size and reduce costs HLA imputation is used rather than sequencing methods. A recently published multi-ancestry reference panel aims to address some of the issues involved with HLA imputation by improving upon allele diversity and panel size. To assess the quality of imputation of HLA alleles using this panel compared to PacBio sequencing, HLA imputation was also conducted from genotype data for these sequenced samples using the latest imputation approach. Whilst this method does implement the largest and most diverse available dataset to support high quality imputation, it can only provide G-group resolution, and so the ability of imputation to correctly capture the first two fields of the PacBio sequenced samples was assessed.

The HLA alleles imputed by this method for all samples are provided in Supplementary Table 1, under 'HLA-TAPAS'. These showed discrepancies when compared to the PacBio best match alleles, especially evident in the class II loci. Table 25 indicates the percentage of samples where the first two fields of the PacBio best match allele did not match the same fields of the imputation results.

Locus	% Mismatch between sequencing and imputation (risk group)	% Mismatch between sequencing and imputation (protective group)	% Mismatch between sequencing and imputation (total)
HLA-B	6	10	9
HLA-C	3	4	4
HLA-DQA1	19	48	37
HLA-DRB1	18	38	31

Table 25: Percent mismatch between best match allele to PacBio long-reads and imputed result from HLA-TAPAS.

Percentage is a calculation of each allele incidence (2n) in which the first 2 fields do not match between the methods. The group homozygous for the protective alleles showed greater mismatch across all alleles excepting HLA-C.

At the DQA1 locus, this mismatch primarily came from any allele of the DQA1*03 group being imputed as DQA1*03:01, which disregarded any of the 03:03 and 03:02 alleles (Supplementary Table 1). At the DRB1 locus, this primarily arose from imputation of a high number of alleles within the HLA-DRB1*04 group to various incorrect alleles of the DRB1*04 or DRB1*07 group, although with a less consistent pattern.

These errors are weighted more in the protective allele group than the risk group (Table 25), with up to 48% of all alleles being incorrectly imputed at 2-field resolution in the case of DQA1. This demonstrates the issue with relying on imputation alone to understand which alleles are associated with this genome wide significant SNP, as association of certain alleles may be inflated such as in the case of DQA1*03:01. Confirmation with sequencing studies can aid in identification of true associations.

Despite these errors, the overall accuracy of this method is an improvement on other approaches, and most alleles are imputed with high accuracy. When common repeated errors in the DQA1 locus or errors where the second field was not assigned were removed, the remaining two field alleles are imputed to a high level of accuracy as demonstrated in Table 26. Furthermore, of the alleles that were significantly associated with either the risk or protective group in the PacBio sequenced data, the majority were imputed to either 100% accuracy or with one instance of a sample with this allele being incorrectly imputed. DRB1 is still the most poorly imputed allele, so caution should be taken when interpreting results of DRB1 allele associations within the imputed data.

Locus	% Mismatch between sequencing and imputation (risk group)	% Mismatch between sequencing and imputation (protective group)	% Mismatch between sequencing and imputation (total)
HLA-B	3	9	6
HLA-C	2	4	3
HLA-DQA1	8	3	5
HLA-DRB1	10	20	16

Table 26: Percent mismatch between best match allele to PacBio long-reads and imputed result from HLA-TAPAS, ignoring selective errors.

Percentage is a calculation of each allele incidence (2n) in which the first 2 fields do not match between the methods. The imputation of the majority of alleles at each loci show accurate imputation, with alleles at the DRB1 locus still showing the most errors.

3.3.4 Case/Control Association Analysis of HLA Alleles Correlated With rs28654242

Despite these discrepancies with imputation identified at the individual sample level, imputation quality was still demonstrated to be high for most alleles, and in larger case-control datasets will be valuable in providing information on allele groups that are associated with disease. Identification of which alleles associated with PD risk/protection in the PacBio analysis are also associated with PD in a larger case-control sample will help identify which are true associations with these risk/protective alleles.

To conduct a case-control association analysis of the HLA alleles implicated by the PacBio sequencing, HLA imputation of the PD dataset with 10,017 cases and 5,322 controls was performed with the latest large multi-ancestry panel. This data was used to perform an association analysis of HLA alleles with PD. Table 27 shows the P value and odds ratio from this analysis for each HLA allele implicated by PacBio sequencing dataset.

Allele	**Case-Control Imputation PD association**	
	Odds Ratio	**P value**
B*40:01	0.96	0.45
B*44:02	0.93	0.13
B*44:03	1.04	0.49
C*03:04	***0.90***	***0.029***
C*07:01	0.95	0.33
DQA1*01:01	***1.08***	***0.037***
DQA1*01:02	1.04	0.24
DQA1*02:01	1.00	0.96
DQA1*03:01	***0.83***	***4.38E-08***
DQA1*03:03	N/A	N/A
DQA1*04:01	1.13	0.13
DQA1*05:01	1.02	0.70
DRB1*01:01	1.05	0.34
DRB1*03:01	1.03	0.59
DRB1*04:01	***0.88***	***0.0059***
DRB1*04:04	0.85	0.026
DRB1*07:01	0.99	0.90
DRB1*11:01	1.06	0.29
DRB1*15:01	1.06	0.14

Table 27: Comparison of associated alleles from PacBio sequencing approach with allele association results from imputation in larger case-control dataset.

*Odds ratio and P value of all alleles that showed significantly different distribution between Nalls risk and protective group are provided. DQA1*03:03 is not provided as HLA-TAPAS failed to impute this allele.*

The results from this data indicate which of the alleles implicated by the PacBio dataset are also putatively associated ($P < 0.05$) in the case-control dataset. These include DQA1*03:01(P = 4.38 x 10^{-8}, OR = 0.83), DRB1*04:01 (P = 0.005786, OR = 0.88), C*03:04 (P = 0.029, OR = 0.90) associated with PD protection, and DQA1*01:01 (P = 0.037, OR = 1.08) associated with risk.

The remainder of HLA alleles that were implicated did not translate into significance in a case-control association, which indicates that could have been a product of the small sample used for the PacBio analysis. These alleles were demonstrated to be imputed to a high level of accuracy, so it is unlikely to be a result of imputation error. However, within the significant results which are replicated, imputation errors of DQA1*03:01 and DRB1*04:01 could have had an impact on biasing results to inflate or deflate the significance of the association.

3.3.5 Case-Control Association Analysis of HLA Alleles Independent of rs28654242

These imputation results indicate the potential true associations from the PacBio analysis, however this approach focused only on those alleles and variants that were correlated with the most significantly associated PD HLA SNP. This approach overlooks the other associations that are either independent of rs112485576 or those that were not captured by this sample. To fully examine the PD associations in this latest imputation of case-control data, the complete PD case-control association results were examined independently (Table 28)

The results indicate that the top associated allele is the protective effect of the DQA1*03:01 allele ($P = 4.38 \times 10^{-8}$, OR = 0.83). This association analysis also identified multiple putatively associated alleles ($P < 0.05$), including the protective alleles DRB1*04 ($P = 3.8 \times 10^{-6}$, OR=0.85), DQB1*03:02 ($P = 1.1 \times 10^{-5}$, OR = 0.82), and DRB1*11:04 ($P = 3.0 \times 10^{-3}$, OR = 0.77) and the risk alleles DQA1*01 ($P = 2.1 \times 10^{-4}$, OR = 1.11), C*02:02 risk ($P = 2.0 \times 10^{-3}$, OR = 1.19), DPB1*01:01 (P =0.013, OR=1.15), B*37:01 (P = 0.017, OR = 1.28).

To correct for multiple comparisons, the Bonferroni corrected P value was calculated to account for testing 1,781 HLA alleles; in this case $P < 3 \times 10^{-5}$. Thus DQA1*03:01, DRB1*04, DQB1*03:02 associations remained significant after correcting for multiple comparisons. However, uncorrected putative associations of $P < 0.05$ are also considered in this analysis, due to the conservative nature of the Bonferroni correction and considering the high level of correlation of alleles within the HLA locus.

Indicated in bold are those alleles that were found to be correlated with rs112485576 in the PacBio sequenced dataset; these alleles being the protective DQA1*03:01, DRB1*04, and C*03:04 alleles, and the risk allele DQA1*01. The replication of these alleles gives further confidence in these being true associations with PD risk and protection. The presence of other associated alleles not correlated with risk/protection in the PacBio dataset suggests that these associations are independent of rs112485576.

HLA Allele	Frequency	PD Association		Conditional on DQA1*03:01		Conditional on DQA1*01		Conditional on C*02:02		Conditional on DPB1*01:01	
		OR	P	OR	P	OR	P	OR	P	OR	P
DQA1*03:01	***0.18***	***0.83***	***4.4E-08***	***NA***	***NA***	***0.85***	***1.49E-05***	***0.83***	***4.10E-08***	***0.83***	***6.00E-08***
DRB1*04	***0.16***	***0.85***	***3.8E-06***	***1.05***	***0.55***	***0.87***	***5.78E-04***	***0.85***	***3.26E-06***	***0.85***	***4.87E-06***
DQB1*03:02	0.10	0.82	1.1E-05	0.95	0.45	0.85	6.42E-04	0.82	8.74E-06	0.83	1.47E-05
DQB1*03	0.27	0.89	8.2E-05	1.01	0.87	0.92	0.02	0.89	8.21E-05	0.89	1.30E-04
DQA1*01	***0.42***	***1.11***	***2.1E-04***	***1.05***	***0.14***	***NA***	***NA***	***1.11***	***2.20E-04***	***1.12***	***9.89E-05***
C*02:02:02:01	0.04	1.22	1.0E-03	1.22	1.00E-03	1.22	1.00E-03	3.26	0.05	1.22	1.16E-03
C*02:02	0.04	1.20	2.0E-03	1.20	2.00E-03	1.20	3.00E-03	NA	NA	1.20	2.52E-03
DRB1*11:04	0.03	0.77	3.0E-03	0.75	1.00E-03	0.80	0.01	0.78	4.00E-03	0.77	3.31E-03
DRB1*04:01	***0.09***	***0.88***	***6.0E-03***	***1.07***	***0.28***	***0.92***	***0.07***	***0.88***	***6.00E-03***	***0.88***	***6.38E-03***
DPB1*01:01	0.06	1.15	0.013	1.14	0.02	1.17	6.00E-03	1.15	0.01	NA	NA
DPB1*01:01:01:01	0.06	1.15	0.016	1.14	0.02	1.17	7.00E-03	1.15	0.02	0.82	0.67
B*37:01	0.02	1.28	0.017	1.26	0.02	1.25	0.03	1.28	0.02	1.28	0.01
B*35	0.08	0.88	0.018	0.87	0.01	0.88	0.02	0.89	0.03	0.87	0.01
C*03:04:01:01	***0.07***	***0.89***	***0.019***	***0.92***	***0.11***	***0.90***	***0.03***	***0.90***	***0.04***	***0.89***	***0.02***
DQA1*01:03	0.06	1.13	0.025	1.09	0.12	1.07	0.26	1.12	0.03	1.13	0.02
DRB1*04:04	0.03	0.85	0.026	0.99	0.89	0.89	0.12	0.85	0.02	0.85	0.03
DQB1*05:01:01:01	0.12	1.09	0.028	1.05	0.25	1.04	0.42	1.09	0.03	1.09	0.03
C*01:02	0.04	0.86	0.028	0.84	0.01	0.85	0.02	0.87	0.04	0.86	0.03
DQB1*05:01	0.12	1.09	0.029	1.05	0.25	1.04	0.43	1.09	0.03	2.14	0.03
C*03:04	0.07	0.90	0.029	0.93	0.16	0.90	0.04	0.91	0.05	0.89	0.02
DQA1*01:01	0.15	1.08	0.037	1.04	0.35	1.02	0.65	1.08	0.04	1.08	0.04
C*01	0.04	0.87	0.038	0.85	0.02	0.85	0.02	0.87	0.05	0.86	0.04
DQB1*06	0.25	1.07	0.042	1.02	0.57	0.98	0.62	1.07	0.03	1.07	0.02
DQB1*05	0.16	1.08	0.042	1.03	0.40	1.01	0.78	1.08	0.05	1.08	0.04
DQB1*06:03	0.06	1.11	0.048	1.07	0.20	1.05	0.41	1.11	0.06	1.11	0.04

Table 28: PD case-control association analysis of imputed HLA alleles.

Association analysis of HLA alleles imputed by the HLA-TAPAS method with PD phenotype. Association analysis was conducted with 10,018 cases and 5,322 controls. Following this, association was repeated whilst conditioning on the top risk and protective alleles. Odds ratio (OR) and P value are displayed. Those passing putative significant threshold (P < 0.05) in the initial association analysis and with a frequency greater than 0.01 are included. In bold are those alleles that were implicated by the PacBio long-read sequencing analysis.

To test the independence of these alleles, various conditional analyses were performed on the most significantly associated risk and protective HLA alleles (DQA1*03:01, DQA1*01, DRB1*11:04, DPB1*01:01, C*02:02). Conditioning on DQA1*03:01 showed DRB1*11:04 remained significant as a protective allele (P =1.00 x 10^{-3}), while DRB1*04 (P = 0.55) and DQB1*03:02 (P = 0.45) were insignificant (Table 28). This indicates the LD between the latter alleles and DQA1*03:01 was driving their protective association, whilst DRB1*11:04 is independently associated. Out of the risk alleles C*02:02 (P = 2.00 x 10^{-3}), DPB1*01:01 (P = 0.02), and B*37:01 (P = 0.02) remained significant while DQA1*01 (P = 0.14) became insignificant. Conditioning on C*02:02 showed DPB1*01:01 (P = 0.01) and B*37:01 (P = 0.02) remained significant as risk alleles, and after conditioning on DPB1*01:01, B*37:01 (P = 0.01) remained significant still.

These results indicate that the top independent PD associations include DQA1 (DQA1*03:01 protective and DQA1*01 risk), DRB1*11:04 (protective), and C*02:02 (risk), with further putative risk or protective alleles.

3.3.6 Interaction of Imputed Alleles Associated with PD Risk and Protection

To assess the interaction between the risk and protective alleles, frequencies of haplotypes including the top identified risk and protective alleles (DQA1*03:01, DQA1*01, DRB1*11:04, DPB1*01:01, C*02:02) were assessed (Table 29). For the purpose of this investigation, those that had excess frequency in the cases were considered risk haplotypes, while those that had excess frequency in controls were protective haplotypes. All haplotypes that had a frequency of > 0.01 are included.

The results indicate that in most instances, any haplotypes containing the DQA1*03 allele will always be protective, while carrying the DQA1*01 allele will only confer risk in the absence of DQA1*03 (Table 29). The protective effect of DQA1*03 is only overcome when the haplotype also contains C*02:02 as well as DQA1*01. C*02:02 and DRB1*11:04 alleles are only present in risk haplotypes at this frequency, while it can be observed that the protective effect of DPB1*01:01 is carried in risk and protective haplotypes, suggesting this allele is less influential on outcome than other HLA loci. The frequencies of these interactions are small, so these suggestions are observational rather than conclusive. This does however indicate the nature of the DQA1*03:01 risk overpowering alternative protective effect.

	C		DRB1		DQA1		DPB1		Control count	Control Frq	Case count	Case Frq	Frequency difference
Risk haplotypes	-	-	-	-	01	-	-	-	2549	0.25	1436	0.27	-0.015
	-	-	-	-	01	01	-	-	1479	0.15	844	0.16	-0.011
	-	-	-	-	-	-	-	-	1039	0.10	595	0.11	-0.008
	02:02	-	-	-	01	-	-	-	194	0.02	136	0.03	-0.006
	-	-	-	-	01	-	01:01	-	393	0.04	237	0.04	-0.005
	-	-	11:04	-	-	-	-	-	132	0.01	88	0.02	-0.003
	02:02	-	-	-	01	03:01	-	-	122	0.01	80	0.02	-0.003
	-	-	11:04	-	01	-	-	-	151	0.02	90	0.02	-0.002
	02:02	-	-	-	01	01	-	-	147	0.01	87	0.02	-0.002
Protective haplotypes	-	-	-	-	-	-	01:01	-	306	0.03	156	0.03	0.001
	-	-	-	-	03:01	-	01:01	-	163	0.02	67	0.01	0.004
	-	-	-	-	03:01	03:01	-	-	298	0.03	114	0.02	0.008
	-	-	-	-	01	03:01	-	-	1332	0.13	589	0.11	0.022
	-	-	-	-	03:01	-	-	-	1161	0.12	461	0.09	0.029

Table 29: Most frequent haplotypes containing any combination of risk and protective alleles.

*All haplotypes with frequency > 0.01 within the imputed case-control dataset are listed. – corresponds to carrying of alleles other than C*02:02, DRB1*11:04, DQA1*01/03 or DPB1*01:01. Frequency in case and control groups are displayed, with the difference between the two indicating risk or protection.*

3.3.6.1 Interaction between DQA1 and DQB1 alleles

As DQA1 alleles form functional heterodimers with DQB1 alleles, frequency of DQA1/DQB1 haplotypes was also examined (Table 30). DQA1*03:01 was most frequently in haplotypes with DQB1*03 alleles, forming the DQ8 HLA molecule, whilst DQA1*01 was most frequently observed in haplotypes with DQB1*05 and 06, forming the DQ5 and DQ6 molecules respectively. Other haplotypes at lower frequencies with DQA1*01 were also observed. This observation is in line with other data suggesting that these are the most frequent DQ molecules combinations for these DQA1 alleles, forming stable surface heterodimers (111).

DQA1		DQB1		Control count	Control Frq	Case count	Case Frq
03:01	03:01	03	03	342	0.03	140	0.03
01	01	05	06	748	0.07	439	0.08
01	01	06	06	600	0.06	315	0.06
01	01	05	05	214	0.02	138	0.03
01	03:01	03	06	907	0.09	405	0.08
01	03:01	03	05	529	0.05	254	0.05

Table 30: Combinations of DQA1 and DQB1 alleles

*DQB1 combinations are listed for all samples either homozygous for DQA1*03:01 or DQA1*01 or heterozygous for both alleles, at a frequency greater than 0.01. Frequencies demonstrate that DQA1*03:01 is most often in a haplotype with a DQB1*03 allele, whilst DQA1*01 alleles are most often in a haplotype with a DQB1*05 or 06 allele.*

3.3.7 Amino Acids Associated with PD Risk or Protection

To examine the potential functional properties of the top protective and risk alleles at the DQA1 locus, the amino acid residues that are most associated with PD were examined. The top amino acid positions that were associated with disease are listed in Table 31. This shows that the association of the DQA1*03:01 allele was driven primarily by positions 187, 47, 56 and 76 carrying a threonine, glutamine, arginine, and valine respectively.

The protective association of the top 4 most significant positions (187 T, 47 Q, 56 R, 76 V) are consistent with the association of DQA1*03:01 allele, as these amino acids are unique to DQA1*03 alleles. The three positions which reach a lower significance threshold (215 L, 50 L, 53 R) are also shared by the DQA1*03 group alleles but are not unique to this group, for instance DQA1*02 alleles also carry a leucine at positions 215 and 50, and an arginine at position 53.

Amino acid residue	Odds ratio	P value	Allele group
DQA1 position 187 (exon4) T (Threonine)	0.83	4.38E-08	Unique to DQA1*03:01 alleles
DQA1 position 47 (exon2) Q (Glutamine)	0.83	4.38E-08	
DQA1 position 56 (exon2) R (Arginine)	0.83	4.38E-08	
DQA1 position 76 (exon2) V (Valine)	0.83	4.38E-08	
DQA1 position 215 (exon4) L (Leucine)	0.87	1.66E-06	Not unique to DQA1*03:01
DQA1 position 50 (exon2) L (Leucine)	0.87	1.66E-06	
DQA1 position 53 (exon2) R (Arginine)	0.87	1.66E-06	

Table 31: The top associated amino acid positions with PD protection.

The differences at these positions therefore account for most of the protective effect of the DQA*03:01 alleles. Figure 3-6 demonstrates the location of these unique positions in green, and the respective amino acids within the DQA1*01:01 allele for reference. Within the DQA1*01 group, these consist of arginine, glycine and methionine at positions 47, 56 and 76 respectively. These amino acid residues are associated with disease risk.

Amino acid residue	Odds ratio	P value	Allele group
DQA1 position -16 (exon1) L (Leucine)	1.13	1.07E-05	Shared with DQA1*04 and 06 alleles
DQA1 position 69 (exon2) AT	1.13	1.12E-05	
DQA1 position 175 (exon3) Q	1.11	2.05E-04	Unique to DQA1*01 alleles
DQA1 position 218 (exon4) Q	1.11	2.13E-04	
DQA1 position 47 (exon2) R	1.11	2.13E-04	
DQA1 position 50 (exon2) E	1.11	2.13E-04	
DQA1 position 52 (exon2) S	1.11	2.13E-04	
DQA1 position 53 (exon2) K	1.11	2.13E-04	
DQA1 position 55 (exon2) G	1.11	2.13E-04	
DQA1 position 56 (exon2) G	1.11	2.13E-04	
DQA1 position 61 (exon2) G	1.11	2.13E-04	
DQA1 position 64 (exon2) R	1.11	2.13E-04	
DQA1 position 66 (exon2) M	1.11	2.13E-04	
DQA1 position 69 (exon2) A	1.11	2.13E-04	
DQA1 position 76 (exon2) M	1.11	2.13E-04	
DQA1 position 80 (exon2) Y	1.11	2.13E-04	

Table 32: The top associated amino acid positions with PD risk.

At a lesser significance, amino acids are associated with PD risk. These are listed in Table 32. The two most significant position are not unique to DQA1*01 alleles, with a Leucine at position -16 also being shared by DQA1*04/06 alleles, and a Threonine at 69 being carried by DQA1*04/06 alleles, whilst DQA1*01 alleles carry an Alanine. The remainder of the risk DQA1 amino acids associated are all unique to the DQA1*01 allele group, and so explain most of the PD risk association of this locus. Figure 6 highlights the position of these unique risk amino acids in red. The three positions for which there is one unique risk amino acid and one unique protective amino acid are demonstrated in Figure 6 C-H.

The differences in PD risk and protection conferred by these amino acid positions can be due to the binding and presentation of specific residues that they enable, or in the way that they impact stability and expression of the DQ heterodimer molecule. The amino acid at position 76 is part of the helix forms the peptide binding pocket P9 with the DQB molecule, and is involved in stability of the heterodimer at this pocket (111). Position 56 is also part of peptide binding structure within the P1 binding pocket (112). The amino acid at position 47 lies outside of the peptide binding pocket domains but is an important stability regulatory site for the heterodimer formation. An arginine in this position such as in the DQA1*01 alleles forms hydrogen bonds with the alpha 2 domain, whereas these are not present when a glutamine is substituted as in DQA1*03 alleles, negatively impacting heterodimer stability (113). It is less clear how amino acid changes within the peptide binding pockets or position 187 which lies outside of the DQ molecule domain could have different functional properties in the context of PD pathology.

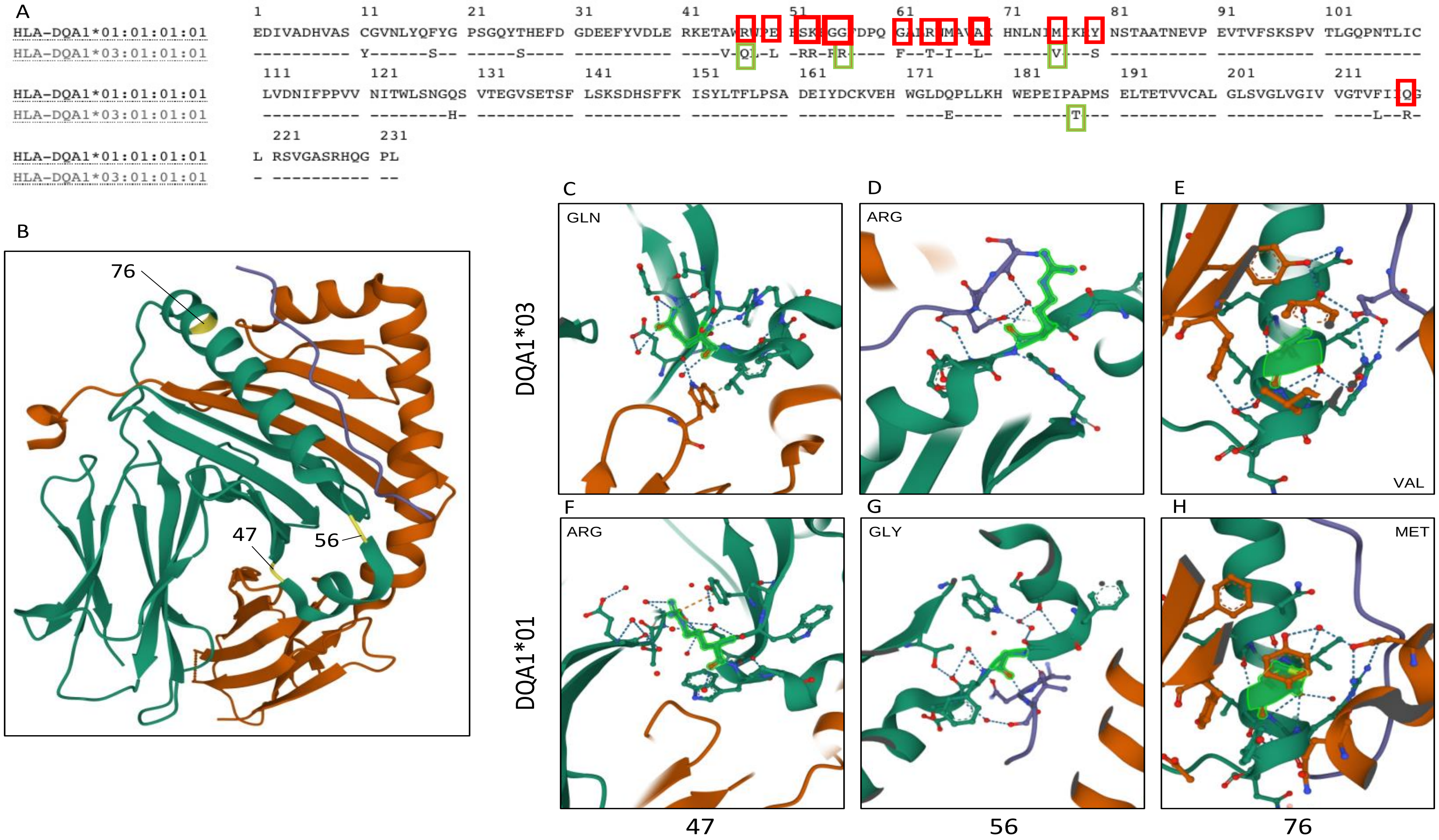

A
1 11 21 31 41 51 61 71 81 91 101
HLA-DQA1*01:01:01:01 EDIVADHVAS CGVNLYQFYG PSGQYTHEFD GDEEFYVDLE RKETAWRWPE FSKFGGFDPQ GALRNMAVAK HNLNIMIKRY NSTAATNEVP EVTVFSKSPV TLGQPNTLIC
HLA-DQA1*03:01:01:01 ---------- Y------S-- -----S---- ---------- ----V-QL-L -RR-FR---- F--T-I--L- -----V---S ---------- ---------- ----------
111 121 131 141 151 161 171 181 191 201 211
HLA-DQA1*01:01:01:01 LVDNIFPPVV NITWLSNGQS VTEGVSETSF LSKSDHSFFK ISYLTFLPSA DEIYDCKVEH WGLDQPLLKH WEPEIPAPMS ELTETVVCAL GLSVGLVGIV VGTVFIIQG
HLA-DQA1*03:01:01:01 ---------- --------H- ---------- ---------- ---------- ---------- ----E----- ------T--- ---------- ---------- ----L--R-
221 231
HLA-DQA1*01:01:01:01 L RSVGASRHQG PL
HLA-DQA1*03:01:01:01 - ---------- --
B
76
47
56
C
GLN
D
ARG
E
VAL
DQA1*03
F
ARG
G
GLY
H
MET
DQA1*01
47
56
76

Figure 3-6: Amino acid residue differences between HLA-DQA1*01:01 and HLA-DQA1*03:01 alleles.

*(A) Amino acid sequence of both alleles, with amino acids unique to DQA1*03 highlighted in red. (B) Location of top associated amino acids within the DQ heterodimer. (C-H) Differences in amino acid residues at associated positions; (C,F) position 47, (D,G) position 56, (E,H) position 76 for DQA1*03 (C-E) and DQA1*01 (F-H). DQA1 molecule is represented in green, DQB1 molecule in orange, and bound peptide in purple. Images generated using the RCSB protein data bank entries 2NNA (DQ8) and 1UVQ (DQA1*01:02/DQ0602).*

3.4 Discussion

3.4.1 Approach

In this chapter, PacBio long-read sequencing was applied to PD patient samples to determine if there were any unknown structural variants within the HLA locus that were correlated with the main SNP associated with PD risk/protection. Long-read sequencing is an especially important method to apply to the HLA region as its complex patterns of LD, conserved repetitive sequencies, and highly polymorphic nature can make it challenging to identify associated alleles and novel variants using short-read sequencing methods.

PD samples from the proband cohort were selected that were homozygous for either the protective or the risk allele of rs112485576, which is the SNP at the HLA locus showing strongest evidence for association with PD in the latest meta-analysis (80). HLA loci determined to be of most relevance (HLA-B, -C, -DQA1, -DRB1) were amplified and full-length long-read sequencing was conducted. Whilst this method did reveal one novel allele at the DQA1 locus that differed from a known HLA allele in the IGMT database by two intronic SNPs, there were no other structural variants that were observed in these samples and no novel variants associated with this PD risk SNP. However, it was possible to observe which alleles were associated with the risk or protective group samples.

To determine which alleles associated in this group were also associated at a case control level, comparison of PacBio results to imputation results from the latest HLA imputation method was conducted. This allowed observation of both protective and risk alleles of interest.

3.4.2 What protective alleles were identified?

Within the PacBio dataset, the top HLA alleles associated with PD protective group were the class II alleles DQA1*03:01, DQA1*03:03, DRB1*04:01, DRB1*04:04. This result was in agreement with results reported from the previous sequencing project, which identified the HLA-DRB1*04_HLA-DQA1*03 haplotype as the source of PD protection (108). When compared with the imputed HLA case-control association, the DQA1*03 and DRB1*04 allele groups were also identified amongst the most associated protective alleles, with DQA1*03:01 the most associated PD allele. Conditioning on DQA1*03:01 showed DRB1*04 and

DQB1*03:02 alleles were no longer significantly associated, demonstrating their LD with the DQA1*03 allele potentially caused their association. One independent risk allele DRB1*11:04 was also identified, however as this is a novel identification it would need to be repeated to ensure this is not an imputation error.

The identification of DQA1*03:01 as the main protective allele in this dataset is of interest, as previous association studies have identified the DRB1*04 allele as the most significant protective factor, with any association from DQA1*03 being explained the LD between these two alleles. This result suggests there could potentially be an alternative explanation to this protective factor, with the role of the DQ alleles as well as the DR of functional consequence. The imputation method applied here could potentially have inflated this association due to imputation of other DQA1*03 alleles as DQA1*03:01, but this still demonstrates the significance of this allele group.

3.4.3 What risk alleles were identified?

Previously, protective alleles have been the focus of PD HLA association, with the top associated alleles acting in protective haplotypes and the top SNP rs112485576 being associated with PD protection (80,87). However, the PacBio analysis identified several alleles associated with the risk group, including DQA1*01:01, DQA1*01:02, DQA1*02:01, DQA1*04:01, DRB1*01:01, DRB1*03:01, DRB1*07:01, DRB1*11:01, and DRB1*15:01. In previous sequencing results, DRB1*01:01 and DQA1*01:01 had been identified as risk associated alleles only. Comparison to the imputation case control results demonstrated that the DQA1*01 group was the most associated allele with PD risk, and so together with the sequencing results this gives confidence in this risk allele.

Beyond this established risk, the C*02:02 allele was also associated with risk despite not being associated with the PacBio sample risk group. Conditioning on DQA1*01 confirmed C*02:02 remained significant, indicating that this is an independent risk allele outside of association with the rs112485576 SNP. DPB1*01:01 was also identified as a potential independent risk allele. This locus was not included in the sequencing data as DPB1 had not previously been identified as a locus of interest. Confirmation that this association is not a product of poor imputation would need to be conducted before further investigating this potential PD risk.

3.4.4 How do these interact?

Exploring the properties of these risk and protective alleles, and their interaction with one another, is important to understanding their role in PD pathology. Identification of the most common haplotype pairs containing the main risk and protective alleles observed here demonstrated this interaction. When carrying at least one copy of the DQA1*03:01 allele, these diplotypes had a protective effect even when also carrying risk allele DQA1*01, unless a second risk allele C*02:02 is also carried. This, whilst the effect of this main protective allele is dominant compared to DQA1*01 risk effect, the data suggests multiple risk alleles can overcome this effect. Due to small frequencies of these haplotypes caution should be taken when observing interactions, however this dominance can explain why protective alleles have commonly been the most readily observed associations in PD studies.

3.4.5 What are the biological consequences of these alleles?

The functional role of these HLA alleles in the pathology of PD, and how they confer risk and protection, has yet to be determined. Alleles from both class I and class II have been identified as conferring risk or protection for PD, despite their different mechanisms of action. Class II alleles are the most associated group, which act by forming heterodimers of one alpha and one beta molecule to present peptides on the cell surface. DQA1 will form a heterodimer with the polymorphic DQB1 molecule, while DRB1 forms a heterodimer with the non-polymorphic DRA molecule, both of which present engulfed extracellular molecules to CD4+ T cells. Class I alleles on the other hand do not form heterodimers, and act presentation of intracellular molecules to activate CD8+ T cells.

As the DQA1*03 and DQA1*01 groups have been identified as protective and risk alleles respectively, the DQB1 alleles within their haplotypes were also examined. DQA1*03 almost exclusively formed haplotypes with DQB1*03, forming the DQ8 heterodimer molecule. On the other hand, DQA1*01 was commonly in haplotypes with DQB1*05 or 06, forming the DQ5 and DQ6 heterodimers, however other haplotypes at lower frequencies were also observed. This aligns with previously observed haplotypes in other populations (113).

Both trans- and cis- heterodimers can be expressed on the cell surface, meaning heterodimers formed from alpha and beta subunits on the same chromosome or from different chromosomes. The formation of trans-heterodimers has been indicated to be of importance in disease susceptibility, for example in the case of type I diabetes (114). However, current evidence suggests that expression of trans-heterodimers between DQA1*03 and DQB1*05/06 or DQA1*01 and DQB1*03 is not readily detected at the cell surface (113,115). This suggests that it is competition between the DQ8 molecule and alternative class II molecules such DQ5 and DQ6 that results in either a risk or protective effect, rather than interaction with trans-heterodimers.

Relative expression levels of these DQ molecules could contribute to their differing influence on PD risk. The level of cell surface density of DQ8 molecules formed of DQA1*03 and DQB1*03 molecules has been observed to be lower than DQ 5/6 molecules formed with DQA1*01 and DQB1*05/06 for example (113). This difference was suggested to be due to the stabilising effect of amino acid position 47, with a glutamine at this position in DQA1*03:01 conferring less stability and so reduced heterodimer expression. This could indicate that reduced DQ-TCR interaction and CD4+ T cell activity could be protective in PD.

Whilst the mechanism by which these HLA alleles confer risk or protection is still unknown, one possibility is the involvement of cell surface expression of α-syn. It has been demonstrated that T cells from PD patients can recognise α-syn (66), however it is currently unclear which HLA alleles present α-syn and whether the expression of these alleles is significant to PD development. One study investigating the recognition of α-syn molecules identified two antigen regions which elicited a response from class II stimulated CD4+ cells (116). One antigenic peptide near the N-terminus which was bound by DRB1*15:01 and DRB5*01:01 with high affinity, but also by DQB1*03:01 with slightly lower affinity. A different peptide near the C-terminus was found to be weakly bound to most class II molecules except from DQB1*05:01, to which it bound strongly. These results suggest that DQ molecules formed by both the risk and protective alleles could be involved in antigen presentation of α-syn molecules, and so doesn’t present a clear picture of whether this process is involved in reduced or increased risk of PD. It is also not yet clear whether the specific amino acids of significant identified in the DQA1*03/01 alleles could impact the presentation of these α-syn molecules. By further exploring potential differences in the immune response elicited by these molecules and the antigens that they preferentially bind, this could present a pathway of modulated disease risk.

It was also observed that the response resulting from α-syn molecule recognition was mostly mediated by CD4+ T cells, indicating that class I alleles are not involved in this pathway. The PD risk conferred by HLA C*02:02 would be a result of its interaction with CD8+ T cells. Although consistently less significantly associated than the role of the class II HLA loci, this interaction with CD8+ molecules could be of importance in PD pathology. CD8+ rather than CD4+ T tell infiltration was observed to be increased in within the brain of PD patients and correlated with neuronal cell death, and studies in post-mortem tissue indicate this infiltration can occur in early stages of disease prior to α-syn aggregation (117). CD8+ infiltration was also observed to be increased in the brain in PD mouse models after increase in oxidative stress and expression of class I HLA molecules (118). This indicates there could be alternative biological pathways that separately influence PD risk, independent class I PD associations acting through influence of CD8+ activity. The specific role of HLA-C or other class I alleles in this pathway however is yet to be determined.

3.4.6 Conclusions

The results of this investigation indicate that both risk and protective HLA alleles are associated with PD, primarily at the DQA1 locus. The protective effects of DQA1*03:01 are the most influential on PD outcome, so further investigation into the role of this allele in PD pathology will be important to understand this association. Previous short-read sequencing studies and imputation studies have identified the DRB1 alleles in the risk/protective haplotype as the driving factor of impact on PD risk, with a specific shared epitope identified at DRB1*04 alleles conferring PD protection. Any association of DQA1 alleles was suggested to be due to the high LD between the class II alleles. However, due to the complex patterns of LD at the HLA region it difficult to pinpoint the exact functional mechanism behind this association at linked alleles, and the results from this study suggest that investigation of the role of the DQ molecules should also be further explored.

4 Investigating the Causal Relationship Between Pain and Depression in Parkinson’s Disease

4.1 Introduction

4.1.1 Experience of Non-Motor Symptoms in PD

While Parkinson’s disease (PD) is characterised by motor symptoms, patients also suffer from a host of other non-motor PD symptoms, such as sleep disturbances and cognitive changes. The experience of these symptoms differs across patients but can have a significant impact on quality of life. An investigation into the experience of those living with PD aimed to establish the most impactful symptoms. Responses showed that for early PD patients, pain was the 4th most troubling PD related symptom behind slowness, tremor and stiffness (4), with 9.8% ranking it as their worst symptom. On the other hand, patients with advanced PD reported that mood disorders were one of the most troublesome symptoms, with 7.5% listing this as the most significant symptom, and second only to fluctuating response to medication.

This demonstrates how for many, it is these non-motor symptoms that form the worst aspects of living with PD. Pain is a common issue for PD patients, with a recent investigation finding that 85% of PD patients report experiencing some pain, with 42% having moderate to severe pain. This study also found that pain had a greater impact on quality of life compared to motor symptoms in this cohort (8). Mood disorders are similarly a very common issue, with depression being the most common psychiatric symptom experienced by PD patients (119). An estimated 35% of patients experiencing clinically significant depressive symptoms (120), with experiences of sadness, pessimism, and increased anxiety the most reported experiences from PD patients who complained of mood changes (4).

Whilst traditional PD therapies which focus on dopamine replacement have benefit in treating the core disease characteristics, it is questionable how much they can impact these non-motor symptoms. To have a greater impact on quality of life of PD patients an improved understanding of the underlying causes of these non-motor symptoms is required so that they may be targeted appropriately.

4.1.1.1 PD Pain

The most common types of pain experienced in PD are musculoskeletal, radicular, and dystonic. It has been suggested that motor impairments and related stiffness could contribute to these symptoms, however, severity of motor impairment has not been shown to be correlated with levels of PD pain (8). Alterations in central or peripheral pain processing pathways could also be underlying causes. Severity of PD symptoms of anxiety and depression were found to predict the pain levels experience, suggesting potential shared central mechanisms such as monoamine depletion caused by PD degeneration could also be driving PD pain.

One previous GWAS of pain in PD has been conducted (13), aiming to identify the genetic factors influencing PD patients to experience no/low pain or high pain levels. This study implicated the TRPM8 locus as a risk factor, which is a cold sensing ion channel also involved in inflammation and analgesia. TRPM8 is highly expressed in the caudate, potentially implicating it in central pain pathway regulation (121), and also in DRG neurons, which are involved in the development of neuropathic pain (20). However, pathway through which this risk factor could impact pain is still uncertain.

Current treatment options for pain in PD, including traditional analgesic drugs such as NSAIDs, seem to have limited effect. 28% of PD patients experiencing pain reported paracetamol to be effective, 12% found NSAIDs effective, 10% found opioids effective and 3% found drugs targeting central pain (gabapentin, pregabalin etc) effective (4). Trials of other therapies have yet to indicate a promising alternative, however with greater knowledge of factors which result in the experience of pain in PD, more effective therapies could be achieved.

4.1.1.2 PD Depression

Likewise with pain in PD, there is still an incomplete understanding for why the incidence of mood disorders in PD is so high. A PD diagnosis as an adverse life event itself can be a risk factor for developing depression, and biological changes in PD can also directly affect mood. Depression in PD is associated with degeneration of dopaminergic neurons, noradrenergic limbic and brainstem structures, supporting the impact of these structures on mood (28). This is also supported by patients 'off period' depression, demonstrating the impact of dopamine dysfunction on mood (30). It has also been observed that inflammatory factors are increased in PD patients with depression (35). Whether inflammatory factors act on mood via increasing degeneration of dopaminergic neurons or through a different avenue is uncertain.

Whilst no GWAS of depression in PD has been conducted to date, potential genetic associations have been investigated. One study previously identified a CB1 (cannabinoid receptor 1) gene polymorphism that could impact the expression of this gene that was associated with depression in PD (37). The endocannabinoid system has been shown to be a target in both pain and depression therapies, so this could be an avenue of further investigation. Furthermore, another study investigating the variable effects of PD associated SNPs on different clinical features of disease found one SNP near the BRIP1 gene which was associated with depression (38). Further exploration of genetic associations with PD depression is necessary to determine if these are the primary factors.

SSRIs are the most common approach for treatment of PD depression. These may not be the most appropriate approach however, with trials showing they are less effective at treating PD depression than non-PD (28). There is evidence antidepressants targeting both serotonergic and noradrenergic systems could be more effective for these patients (7). Furthermore, given the potential impact of dopamine on depression in PD, dopamine replacement therapies could also be effective anti-depressants as well as treating motor symptoms. Levodopa, the gold standard treatment, has not been demonstrated to have antidepressant effects, but a trial of dopamine agonist Pramipexole showed a reduction in depressive symptoms compared with placebo in PD patients (46). Improved understanding of various causes of depression in PD could lead to greater improvements for treatments of this symptom.

4.1.2 Pain and Depression Comorbidity

Whilst non-motor symptoms can impact PD patients to varying degrees, an association between pain and depression in PD has been observed in multiple investigations (33). This is reflected in the general population, where chronic pain and depression are often comorbid conditions. Approximately 85% of chronic pain patients also experience depression (122). Similarly, patients with depression can be twice as likely to develop a pain condition compared to the general population (123). Common neurological changes have been suggested as the reason for this (122), with similar shared factors as studied in PD observed in the general population. In particular, monoamine activity including dysregulation of the dopaminergic system has been found to be significantly impacted in both chronic pain and depression.

Treatment of the symptoms for both conditions can be limited, with pain killers and physical therapy often having inadequate effect on chronic pain conditions, and about 10-30% of patients with major depression showing treatment resistance (124). It has been indicated that the comorbidity of pain and depression could cause worse response to treatment, highlighting the need to investigate these conditions jointly (125). A greater understanding of the causes and risk factors of these disorders and how they interconnect can make targets for treatments improved and help develop preventative measures.

4.1.3 Mendelian Randomisation

4.1.3.1 MR Method

Given their comorbidity, an understanding of whether there exists a causal relationship between pain and depression or whether these are independent factors is important. It can be misleading to rely upon observed correlation to determine causation, as there can be multiple other unmeasured phenotypes also correlating with pain and depression (known as confounding factors). Correlation observed will also not indicate a direction of effect between the two. Ideally, to determine the effect of one exposure on a health outcome (e.g. the effect of the exposure of depression on later development of pain), a randomised control trial (RCT) would be conducted so that randomly selected participants differ only by their exposure. For evident ethical reasons, this cannot be done for risk factors such as chronic pain and depression in the same way it is done for therapies. Therefore, comparison of the outcomes of patients with pain or depression and those without will be potentially impacted by confounding factors.

To address this issue, the Mendelian randomisation (MR) method was developed as a way of conducting a version of an RCT using genetic data of samples. MR allows the detection of a causal relationship between a risk factor (exposure) and an outcome of interest, using genetic variants associated with exposure as instrumental variables for the exposure. As these genetic factors are randomly distributed and less susceptible to confounding bias, they can be used to differentiate samples in a way that reflects the RCT approach, with individuals differing only by allele. This approach also has the benefit of reflecting lifelong patterns of the exposure. The association of these variables can then be tested on the outcome, with any association presumed to be acting through the exposure, and therefore reflecting a direct causal pathway.

The strength of an MR study will be improved by having instrumental variables that closely represent the exposure, either by accounting for a high proportion of the variability or by representing a known biological pathway. The three key assumptions for MR results to be valid are:

- The relevance assumption, that the genetic variants have a true association with the exposure
- The independence assumption, that the genetic variants are not associated with confounders of the exposure-outcome association
- The exclusion restriction assumption, that variants affect the outcome only through the effect on the risk factor.

When these assumptions hold true, the test will indicate if a significant causative relationship exists between the exposure and outcome. Although it is impossible to prove that all potential confounders have been measured, use of reliable genetic variants and application of different sensitivity analyses can contribute to assessing these assumptions in MR tests.

4.1.3.2 MR of Chronic Pain and Depression

The MR method has been applied to test the relationship between chronic pain and depression in the general population. A recent GWAS of multisite chronic pain (MCP) found multiple genetic loci that were risk factors for development of MCP (126). It was also observed that MCP was most closely genetically correlated to major depressive disorder (MDD) out of several psychiatric disorders tested. To determine if a causative relationship existed, an MR test was performed in which the effect of these genetic factors on MDD was tested, with both MCP and MDD tested as exposures. A positive and significant causal effect of MDD on MCP (*ß* = 0.019 and P = 0.0006) was observed, however there were inconsistencies between methods with regards to whether this had a positive or negative causal effect on the outcome, indicating this was not a true causal effect. Alternatively, with MCP as the exposure and MDD as the outcome, a consistent positive and significant causal effect (*ß* = 0.16 and P = 0.047) was observed, indicating that MCP has a causal effect on MDD.

This result indicates that not only is there a correlation between these phenotypes, but that a causative relationship could exist. This result is consistent with observations that sufferers of chronic pain go on to develop depression at a higher rate than general population. However, this study also observed that the genetic instruments used showed evidence of pleiotropic effects, indicating they could be influencing MDD through a pathway other than MCP. Therefore, this outcome would need to be replicated in another dataset to have more certainty of a true causal relationship. A greater understanding of this pathway can be useful to identify those most at risk of developing depression.

Given a causative pathway potentially exists for MCP on MDD in the general population, the question remains whether there is a causative relationship between these non-motor symptoms in PD. While PD pain and PD depression could have common causes with non-PD equivalents, it cannot be assumed that PD pain will also be causative for PD depression. These could have different biological basis, which is indicated by the differing responses to treatments. Therefore, it needs to be investigated whether these factors share a causative relationship in PD which could be targeted, or whether these are independent symptoms.

4.1.4 Aims and Hypotheses

The aims of this investigation are:

1. To investigate the correlation between pain and depression in PD cohorts
2. To identify the possible genetic associations influencing pain and depression in PD in these cohorts via GWAS
3. To establish if a causative relationship exists between these PD phenotypes via Mendelian randomisation.

It is hypothesised that

1. There will be a correlation between pain and depression phenotypes in PD.
2. A causative relationship of chronic pain in PD on depression in PD will exist, reflecting that in the general population.

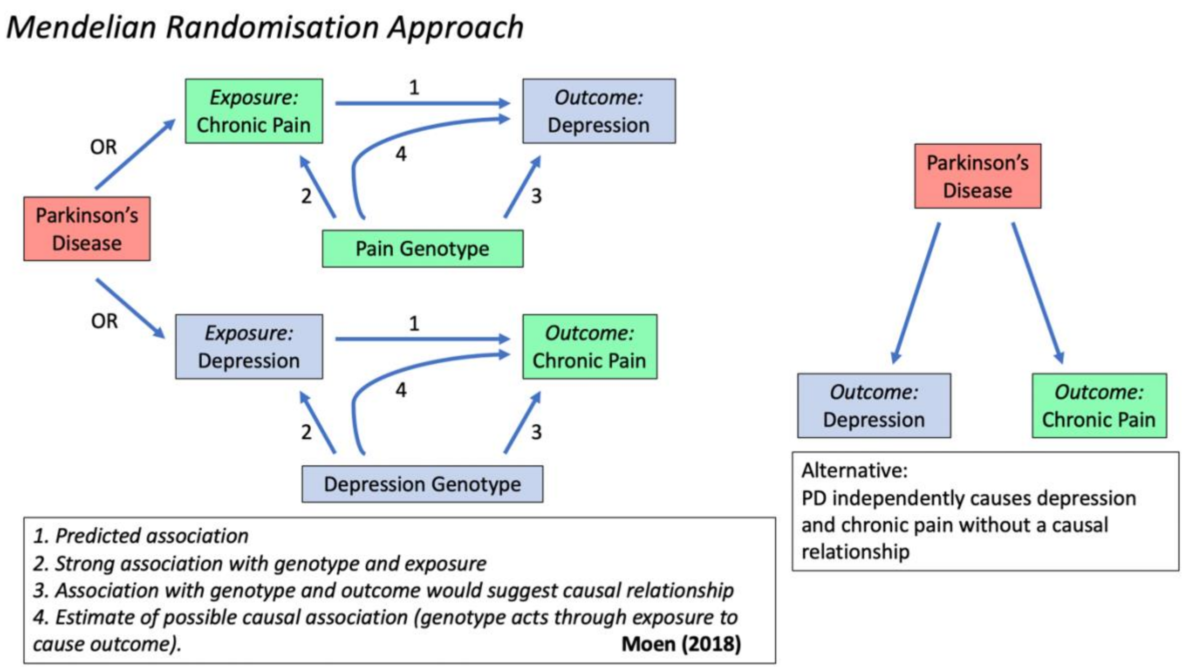

Mendelian Randomisation Approach
Exposure: Chronic Pain
Outcome: Depression
1
4
2
3
OR
Parkinson's Disease
Pain Genotype
OR
Exposure: Depression
Outcome: Chronic Pain
1
4
2
3
Depression Genotype
1. Predicted association
2. Strong association with genotype and exposure
3. Association with genotype and outcome would suggest causal relationship
4. Estimate of possible causal association (genotype acts through exposure to cause outcome).
Moen (2018)
Parkinson's Disease
Outcome: Depression
Outcome: Chronic Pain
Alternative:
PD independently causes depression and chronic pain without a causal relationship

Figure 4-1: Schematic of the hypotheses to be tested in the MR analysis.

Adapted from Moen (2018)

4.2. Methods

4.2.1 PD Datasets

4.2.1.1 Proband

PD samples were obtained from the Proband (Tracking Parkinson's) patient cohort (109). This cohort consists of 2,247 patient samples, 1,987 with recent onset (<3.5 years) and 260 young onset (diagnosed <50 years of age). DNA samples were genotyped using the Illumina Human ExomeCore-12 v1.1 array.

Data collected for pain and depression phenotypes of these samples was obtained for this study. Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) was used to assess PD pain, with a scale from 0 for no pain to 4 for severe pain. Leeds Anxiety and Depression Scale (LADS) depression score (LADS-D) was used to assess depression characteristics in these participants. This ranges from 0-18, with 0 being a self-assessment of no depression symptoms and 18 being severe depression.

4.2.1.2 UK BioBank

PD samples were extracted from the UK Biobank (UKBB) database. All samples with self-reported PD (field: 20002) or with an ICD10 main or secondary diagnosis of PD (field: 41202 and 41204) were used, with 1,566 PD patients extracted in total.

Depression phenotype data was extracted according to the 'broad depression' phenotype previously used in the UKBB GWAS (Howard 2018), consisting of patients who had answered 'yes' to "Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?" (field: 2090) or "Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?" (field: 2010), or who has a primary or secondary diagnosis of a depressive mood disorder from linked hospital admission records (F32—Single Episode Depression, F33—Recurrent Depression, F34—Persistent mood disorders, F38—Other mood disorders and F39—Unspecified mood disorders).

Chronic pain phenotype data was extracted as the number of pain sites at which pain has persisted for over 3 months (0 to 7 sites). Patients responded to the question "pain types experienced in the last month" (field: 6159), with either 'None of the above'; 'Prefer not to answer'; pain at seven different body sites (head, face, neck/shoulder, back, stomach/abdomen, hip, knee); or 'all over the body'. Patients could select any number of pain sites, and then were additionally asked (category: 100048) whether this pain had lasted for 3 months or longer. Those who answered 'all over the body' could not also select from the seven sites, so were excluded.

4.2.2 Association Analysis

Association between pain and depression phenotypes in both PD cohorts was tested to observe potential correlation between PD symptoms. In the Proband cohort, correlation between UPDRS PD pain score and Leeds depression score was measured. In the UKBB cohort, correlation between the number of chronic pain sites and presence/absence of a broad depression phenotype was measured. Linear regression analysis was performed using RStudio 1.4.1106, with age and sex included as confounding factors.

4.2.3 GWAS

All GWAS and principal component analysis (PCA) were performed using Plink version 1.9. PCA was performed to generate principal components (PCs) to use as covariates in the association analysis to control for genetic variation within the population. For each PCA, the Proband or UKBB dataset was merged with the FIN, CHB and YRI samples from 1000 genomes reference genotype data, and the resulting dataset LD pruned. This ensures the PCAs represent population structure and not LD. This merged dataset was then used for a PCA for each cohort. These principal components represent population stratification, and are important to use as covariates so that association with the phenotype is not explained by population structure.

4.2.3.1 GWAS of Depression in Proband cohort

The Proband cohort was used to conduct a GWAS of depression in PD. Depression phenotype data was available for 1,820 samples, with their LADS depression scores used for a linear regression GWAS. SNPs were excluded according to the following QC thresholds: imputation quality (INFO) > 0.8, missingness (GENO) > 0.01, minor allele frequency (MAF) < 0.001, and HWE P value < 1 x 10^{-5}. A linear regression of the principal components was conducted to test association with the depression phenotype. The three principal components that were associated ($P < 0.05$) with depression phenotype (PC7, PC13 and PC16) were used as covariates to perform the GWAS.

4.2.3.2 GWAS of MCP in UKBB cohort

UKBB PD samples were used to conduct a GWAS of multisite chronic pain in PD. 1,394 samples in total were included in this analysis. SNPs were excluded according to the following QC thresholds: imputation quality (INFO) > 0.4, missingness (GENO) > 0.05, minor allele frequency (MAF) < 0.01, and HWE P value < 1 x 10^{-6}. The number of chronic pain sites was used as a phenotype to perform a linear regression analysis. A linear regression of the principal components was conducted to test association with the MCP phenotype. As no principal components were associated with the MCP phenotype, PCs 1-5 were included as covariates in the GWAS along with age and sex.

4.2.3.3 GWAS of Depression in UKBB cohort

UKBB PD samples were also used for a GWAS of broad depression (BD) in PD. There were 563 PD cases with BD and 831 PD cases without BD. SNPs were excluded according to the following QC thresholds: imputation quality (INFO) > 0.4, missingness (geno) > 0.05, minor allele frequency (MAF) < 0.01, and HWE P value < 1 x 10^{-6}. The BD phenotype was used to perform a logistic regression analysis. A logistic regression of the principal components was conducted to test association with the BD phenotype As PC18 was the only principal component associated with the BD phenotype, this was included as a covariate in the GWAS along with age and sex.

4.2.4 Polygenic Risk Score Analysis

To test correlation between genetic risk factors for MDD and depression in PD, and MCP and pain in PD, a polygenic risk score (PRS) analysis was conducted. This involves generating a PRS for MDD/MCP for each PD sample using the summary statistics from a ‘discovery’ MDD/MCP GWAS dataset containing effect size and association P value data for each SNP allele. The combined effect size of all SNPs carried in individuals from the PD ‘target’ sample constitutes their PRS for the given phenotype in a discovery sample. This PRS can then be tested for association with a different phenotype in the target sample. For example, if PRS for MDD is associated with the depression phenotype in the PD samples from the Proband or UKBB samples, this would indicate a shared genetic association between MDD and PD depression, and likewise with chronic pain.

PRS were calculated using Plink version 1.9. GWAS summary statistics from the Wray 2018 MDD meta-analysis and the Johnston 2019 MCP GWAS were used as discovery samples, with those excluding 23 and me used for the Proband PRS and those excluding both 23 and me and the UKBB sample used or the UKBB PRS. To identify independent causal SNPs to use to calculate the PRS, SNP LD clumping was performed so that for SNPs within 250kb of an index SNP, those with $R^2 > 0.1$ were removed. PRS scores were calculated using Plink –score function, using SNPs from a range of SNP P thresholds (P </= 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5). Regression analysis between PRS and target sample phenotype was also calculated using the *glm* regression function in R studio, with principal components 1-5, age and sex included as control variables. This was repeated for PRS at each of the P value thresholds. The best fit P value threshold (pT) was generated in R Studio by calculating the Nagelkerke R^2 value for each, and the pT with the higher R^2 taken as explaining the greatest proportion of variance.

4.2.5 Mendelian Randomisation

4.2.5.1 Two-sample Mendelian Randomisation

Two sample Mendelian randomisation tests were conducted using the TwoSampleMR R package (version 3.5). Instrumental variables (IVs) were selected based on association with the exposure, at genome-wide significance ($P < 5 \times 10^{-8}$) for the meta-analysis or at a suggestive significance threshold ($P < 5 \times 10^{-6}$) for the remaining GWAS. This allowed inclusion of suggestive instruments from GWAS where genome-wide significant results were not produced, and reduced winners curse bias.

Exposure instruments were LD clumped according to the recommended parameters in the TwoSampleMR package to ensure all were independent instruments before harmonising with outcome data. These parameters included clumping within a window of 10,000kb at an R^2 of 0.001.

The Inverse Variance Weighted (IVW) method was used as the primary MR test, with sensitivity analysis applied including the MR-Egger, Simple Mode, Weighted Mode, Weighted Median and MR-RAPS methods. These tests each have different assumptions, and so test the validity of the IV assumptions under different conditions.

- IVW: The IVW method tests association based on the assumption that all instruments are valid, or that if pleiotropic effects are present they are balanced.

Whilst this test has the greatest statistical power, it can be biased if the pleiotropic effect is unbalanced (does not equal zero).

- MR-Egger: This method uses the slope from the weighted regression of the variant-outcome associations on the variant-exposure associations to estimate the causal effect. This takes the intercept as the average pleiotropic effect, based on the assumption that pleiotropy is independent of the variant-exposure associations (InSIDE assumption). This results in bias when this assumption is violated and is also sensitive to outliers.
- Mode: Mode based methods assume more variants estimate true causal effect than any other quantity (plurality valid). This is robust to outliers, yet a more conservative approach
- Median: Median based methods are similar to mode except they assume the majority of instruments are valid. This is also robust to outliers.

Other measures used to assess the reliability of the tests conducted included the MR Egger intercept, a test of potential pleiotropic effects; the Q statistic, a measure of heterogeneity between IVs of the exposure; and the F statistic to assess the strength of the instruments used, with a value >10 indicating absence of weak instrument bias.

The following exposure and outcome combinations were tested in this way:

Exposure	Outcome
Pain in PD (Proband)	MDD
	Depression in PD (UKBB)
Broad Depression in PD (UKBB)	Pain in PD (Proband)
	Multisite chronic pain
Multisite chronic pain (MCP)	Depression in PD (Proband)
	Depression in PD (UKBB)
	MDD
MDD	Multisite chronic pain
	Pain in PD (Proband)
	MCP in PD (UKBB)
MCP in PD (UKBB)	Depression in PD (Proband)
	MDD
Depression in PD (Proband)	MCP in PD (UKBB)
	Multisite Chronic Pain

Table 33: Combinations of exposures and outcomes tested in the two sample MR approach

4.2.5.2 GWAS datasets

Summary statistics for different existing GWAS were obtained to complete the samples needed for the two sample MR approach.

A GWAS of pain in PD was previously performed using the Proband and Oxford PD cohort (13). PD patients had been divided into 2 groups that represented patients with no/low pain (McGill score < 3 and Visual Analog Scale severity <2) and high pain (McGill Score ≥ 3 and Visual Analog Scale severity ≥2). The GWAS was performed with 898 PD cases with pain and 420 PD cases with no/low pain.

For the multisite chronic pain (MCP) phenotype, data was extracted from the recent MCP GWAS (126). 387,649 UKBB samples were included, with number of chronic pain sites used as the phenotype ranging from 0-7.

For the major depressive disorder (MDD) phenotype, data from a recent meta-analysis was used excluding 23 and Me samples (127). 135,458 cases with an MDD diagnosis or equivalent were used with 344,901 controls from various cohorts for this GWAS analysis. For two-sample MR tests including UKBB PD samples, results from the MDD meta-analysis excluding both the 23 and Me and UKBB samples were used.

4.3 Results

4.3.1 Association analysis

Regression analysis was first conducted to attempt to replicate the association found between pain and depression in PD observed previously (4,10). Within the Proband cohort, association was tested between the Leeds depression score and UPDRS pain score of samples, adjusted for age and sex. These were significantly positively associated ($ß$ = 1.20, P = 2 x 10^{-16}, adjusted R^2 = 0.17). Within the UKBB PD cohort, association was tested between the number of chronic pain sites and the presence of a broad depression phenotype in samples, adjusted for age and sex. These phenotypes were also significantly positively associated ($ß$ = 0.05, P = 5.56 x 10^{-6}, adjusted R^2 = 0.03). This indicates that there is a positive correlation between experiencing pain and depression in PD. However, this does not indicate presence of causative relationship.

4.3.2 GWAS

4.3.2.1 GWAS of Depression in PD (Proband)

To test the genetic associations with depression in PD, a GWAS was conducted on the depression phenotype of PD Proband samples. Table 34 shows the number of samples with each self-reported LADS depression score, from 0 for healthy to 18 for severely depressed.

There were no genome-wide significant associations identified in the results from this GWAS, but multiple that crossed the suggestive significance threshold (Figure 4-2). These are distributed across various chromosomes, with the top 20 of these associated SNPs listed in Table 35. There was no sign of genomic inflation (Genomic Inflation Factor = 0.99). The most significantly associated SNP is located within the ASXL2 gene. Among the most significantly associated SNPs is rs12765904, an intronic variant in ZNF33B that also acts as an sQTL (splicing QTL) for the same gene. Expression levels of ZNF33B have been associated with suicide attempts in major depressive disorder (128), indicating this could be of potential biological significance for PD depression.

Leeds Depression Score	Sample no.
0	252
1	180
2	221
3	222
4	221
5	172
6	143
7	115
8	87
9	76
10	58
11	30
12	15
13	9
14	10
15	2
16	3
17	0
18	0

Table 34: Distribution of Leeds Depression Score (LADS) across the Proband cohort*.*

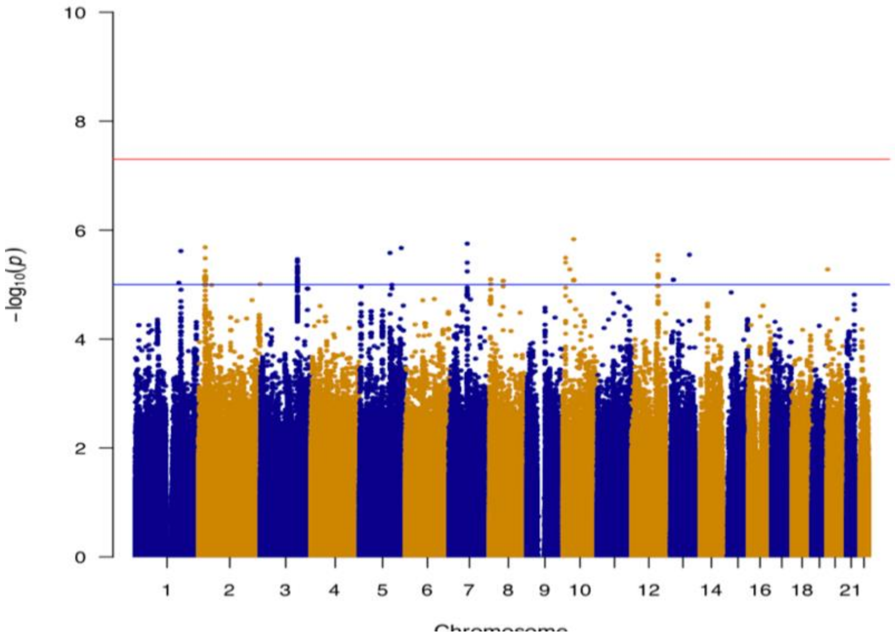


Figure 4-2: Manhattan plot for the Proband GWAS of depression in PD:

Association results from the linear GWAS of depression scores in the Proband cohort. The red line indicates genome wide significant threshold, and the blue line indicates the suggestive significance threshold. No genome wide significant associations *were identified*

.

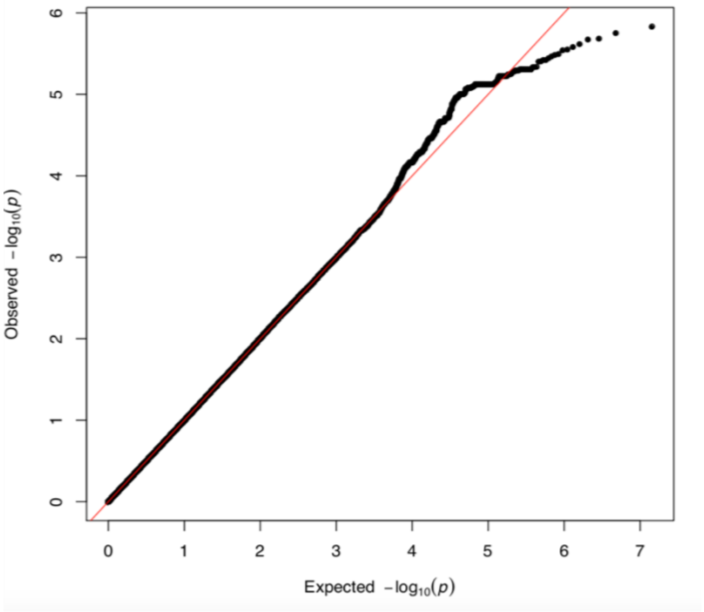


Figure 4-3: QQ plot for the Proband GWAS of depression in PD

Quantile-quantile plot of the GWAS, showing observed P values vs expected P values

CHR	SNP	BP	A1	BETA	P
2	2:25981549	25981549	GT	-0.5595	1.80E-06
13	rs368229332	93050447	T	1.96	1.91E-06
10	rs12765904	43073561	A	5.031	2.25E-06
5	rs187185731	163976625	T	5.315	2.42E-06
2	2:26022849	26022849	CA	-0.5431	2.54E-06
1	rs2504032	179701403	T	-0.5179	2.63E-06
10	rs7909331	11205224	G	-0.6694	3.20E-06
3	rs9846248	144992508	C	0.7544	3.48E-06
10	10:11207166	11207166	TC	-0.6651	3.55E-06
3	rs9818054	145006500	A	0.752	3.81E-06
3	rs9855632	145006499	T	0.752	3.81E-06
7	7:71051664	71051664	G	-0.5488	3.88E-06
5	rs1191746	119412658	A	-0.4908	4.27E-06
2	2:26032165	26032165	GAA	-0.5324	4.34E-06
3	rs16856636	145025038	C	0.7489	4.55E-06
3	rs1822905	145024118	T	0.7489	4.55E-06
3	rs1596719	144988130	C	0.7475	4.66E-06
12	rs12422530	103482733	A	0.6072	4.72E-06
2	2:25979334	25979334	G	-0.5256	4.84E-06
3	3:145013732	145013732	G	0.7438	4.93E-06

Table 35 : Results of Proband depression GWAS

Top 20 suggestive significant SNPs from the Proband GWAS of depression in PD

4.3.2.2 GWAS of Depression in PD (UKBB)

To test the genetic association of depression in PD in an alternative dataset, a GWAS of depression in PD was also conducted in the UKBB PD samples. The phenotype tested was 'broad depression' in PD, with 563 PD cases with BD and 831 PD cases without BD. This GWAS also found no genome wide significant associations, but again several SNPs passing the suggested significance threshold (Figure 4-4). The top 20 associated SNPs are listed in Table 36. There was no evidence of genome inflation (Genomic Inflation Factor = 1.01).

The top associated SNPs include rs8021933 at chromosome 14, which is located within a long noncoding RNA, rs113628522 at chromosome 3 which is located within GTF2E1, and rs116725867 at chromosome 12 which is located within KDM5A, and also an eQTL for KDM5A. This genetic association is of biological interest, as hypothalamic levels of KDM5A have been linked to post-partum depression in a rat model (129).

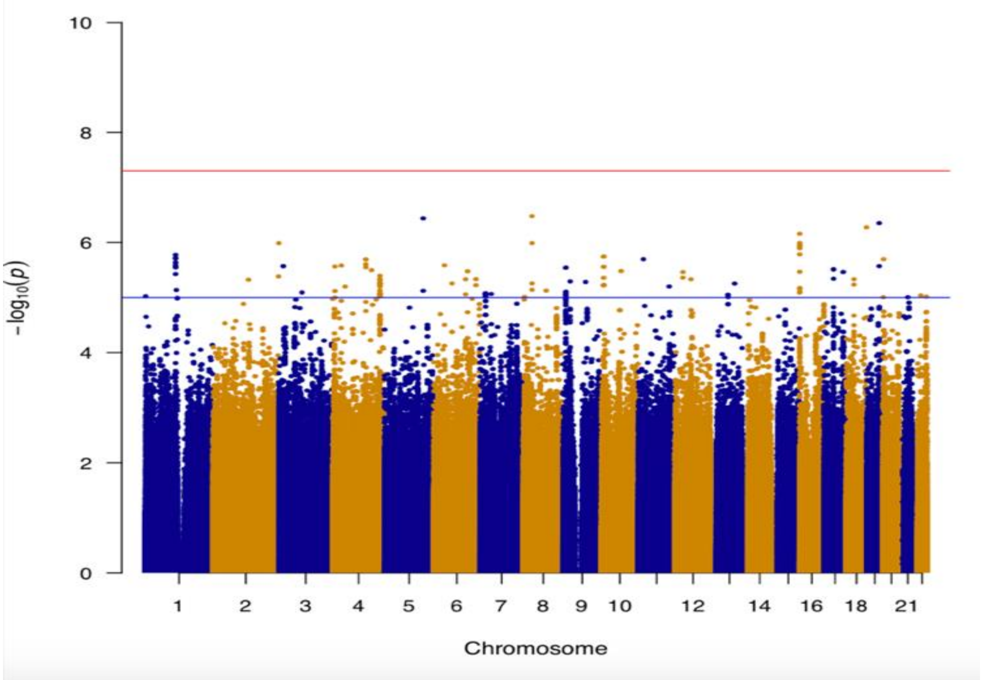


Figure 4-4: Manhattan plot for the UKBB GWAS of depression in PD

Association results from the logistic GWAS of broad depression in the UKBB cohort. The red line indicates genome wide significant threshold, and the blue line indicates the suggestive significance threshold. No genome wide significant associations were identified

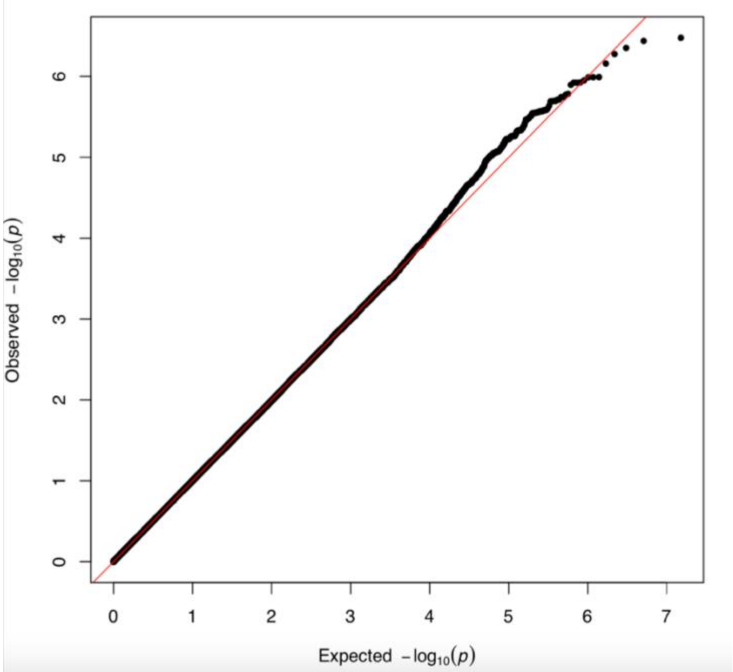


Figure 4-5: Figure 4: QQ plot for the UKBB GWAS of depression in PD

Quantile-quantile plot of the GWAS, showing observed p values vs expected p values.

CHR	SNP	BP	A1	OR	P
14	rs8021933	50475405	C	1.546	2.09E-07
14	rs8020990	50475426	A	1.537	2.98E-07
3	rs73031851	14542762	A	1.484	1.84E-06
3	rs113628522	120482991	T	2.764	2.92E-06
14	rs2355655	50474531	A	1.523	3.14E-06
3	rs4684228	14556175	A	1.452	3.26E-06
14	rs11157717	50474754	A	1.506	4.13E-06
17	rs79830835	39164353	C	1.602	4.44E-06
17	rs6503759	54826065	G	0.6619	4.57E-06
14	rs2526935	73076595	C	1.434	4.57E-06
17	rs6503760	54826089	A	0.6615	4.66E-06
12	rs116725867	433630	C	2.868	5.42E-06
12	rs145520591	431276	C	2.868	5.42E-06
12	rs148698001	457146	C	2.868	5.42E-06
12	rs16929140	389801	A	2.868	5.42E-06
12	rs16929352	417321	A	2.868	5.42E-06
12	rs16929362	418046	G	2.868	5.42E-06
12	rs7297011	451918	A	2.868	5.42E-06
12	rs75755418	418424	C	2.868	5.42E-06
12	rs76971761	436293	A	2.868	5.42E-06

Table 36: Results of UKBB depression GWAS

Top 20 suggestive significant SNPs from the UKBB GWAS of depression in PD.

4.3.2.3 GWAS of MCP in PD (UKBB)

The UKBB PD sample was also used for a GWAS of multi-site chronic pain in PD. Table 37 lists the number of PD samples with each number of self-reported chronic pain sites.

Chronic pain sites	Samples
0	623
1	363
2	220
3	133
4	40
5	10
6	5
7	0
Total:	1394

Table 37: Number of chronic pain sites in UKBB PD samples

Distribution of multisite chronic pain (MCP) phenotype within the UKBB PD samples

There were again no genome wide significant associations resulting from this GWAS, but several variants that crossed the suggestive significant threshold (Figure 4-6). The top associated SNPs are listed in Table 38. There was no evidence of genomic inflation (Genomic inflation factor = 1.00). The top associated SNP rs11787328 within chromosome 8 is an intronic variant in NRG1, levels of which have been related to neuropathic pain in a rat model (130).

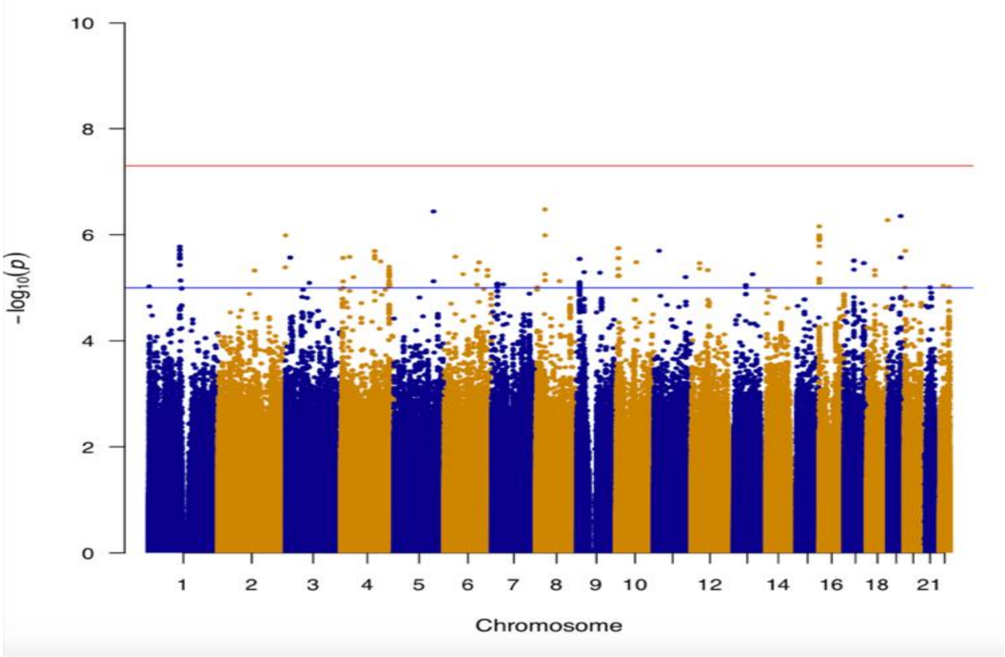


Figure 4-6: Manhattan plot for the UKBB GWAS of MCP in PD

Association results from the logistic GWAS of multisite chronic pain (MCP) in the UKBB cohort. The red line indicates genome wide significant threshold, and the blue line indicates the suggestive significance threshold. No genome wide significant associations were identified.

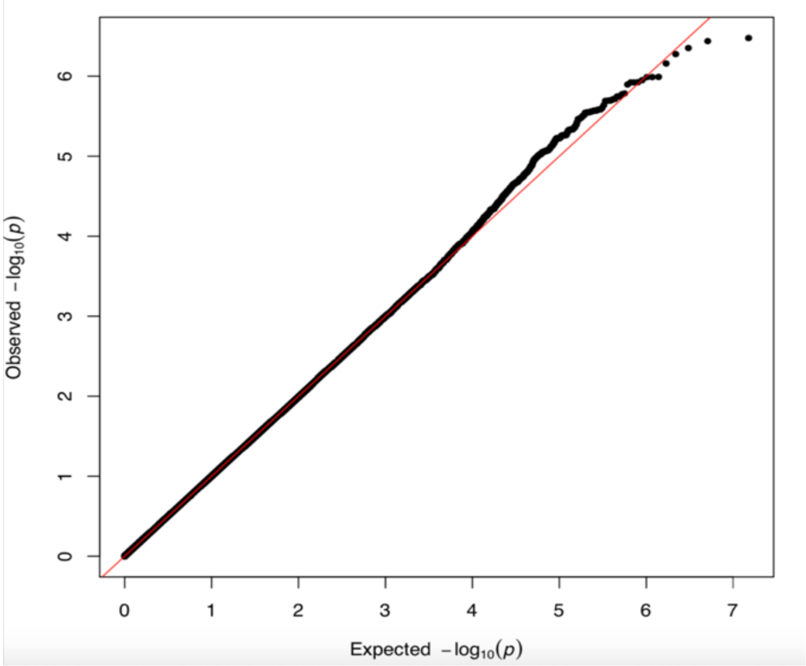


Figure 4-7: QQ plot for the UKBB GWAS of MCP in PD

Quantile-quantile plot of the GWAS, showing observed p values vs expected p values.

CHR	SNP	BP	A1	BETA	P
8	rs11787328	32024297	T	1.021	2.88E-07
5	rs111848133	142472408	T	1.163	3.99E-07
8	rs149157041	32055503	A	0.968	8.38E-07
2	rs181629690	243013288	T	1.944	1.46E-06
3	rs145972844	16533323	A	0.7314	1.60E-06
3	rs113031786	16511074	T	0.7284	1.66E-06
9	rs73642017	10573215	A	0.4304	1.85E-06
9	rs73642018	10573216	G	0.4304	1.85E-06
4	rs35368229	122163771	A	0.3768	1.99E-06
4	rs73843561	122165084	T	0.3768	1.99E-06
6	rs2475509	39890217	G	-0.235	2.35E-06
1	rs7532316	112176753	G	0.2143	2.36E-06
4	rs34150800	122169908	G	0.3749	2.43E-06
4	rs2877748	174102397	G	-0.2276	2.64E-06
1	rs11102320	112169145	G	0.2134	2.70E-06
1	rs3904831	112169812	A	0.213	2.75E-06
4	rs13125905	122175720	G	0.3734	2.76E-06
8	rs184752533	83997828	A	1.007	2.98E-06
4	rs6826412	174109622	C	-0.2265	2.98E-06
4	rs6851632	174109611	A	-0.2265	2.98E-06

Table 38: Results of UKBB MCP GWAS

Top 20 suggestive significant SNPs from the UKBB GWAS of MCP in PD

4.3.3 Polygenic Risk Score

4.3.3.1 PRS of MDD in PD Depression cohorts

To test the genetic correlation between MDD and depression in PD, PRS for MDD was calculated within the Proband and UKBB PD cohorts. The PRS was then tested for association with depression phenotype. PRS was calculated by combining the effect sizes of all variants carried by a sample with an MDD association P value below a certain threshold (pT). The number of independent SNPs used to calculate the PRS at each pT is provided in tables 39 and 40. Effect size and P value provided are the result of association analysis to calculate the association between the PRS and depression in PD phenotype.

pT	N SNPS	EFFECT SIZE	P VALUE	GOODNESS OF FIT*
0.001	4814	4.14	0.92	6.68E-06
0.05	109334	389.04	0.09	2.04E-03
0.1	187750	445.62	0.16	1.42E-03
0.2	***315082***	***906.79***	***0.02***	***3.61E-03***
0.3	419730	841.69	0.08	2.16E-03
0.4	507228	735.38	0.17	1.36E-03
0.5	582366	708.19	0.22	1.07E-03

Table 39: PRS analysis for BD in PD (UKBB sample)

Results of MDD PRS association with broad depression phenotype in the UKBB PD sample. PRS for pT <0.2 has the highest R^2, indicating it represents the greatest proportion of variance. The p value = 0.024, indicating a potential significant association between MDD PRS and depression phenotype. However this is not repeated across other tests.

Table 39 shows the results of this analysis for broad depression (BD) in PD as measured in the UKBB PD sample. The pT threshold which provided the best proportion of variance explained by the PRS is 0.2. At this pT, PRS explained 0.36% difference between the depressed and non-depressed PD samples. Depressed PD samples had a significantly higher PRS for major depressive disorder (P = 0.024, effect size = 906.79*)*. However, this was the only test in which P < 0.05 was reached, with P values at other pT thresholds all consistently larger. This indicates a lack of strong evidence for a shared genetic background between MDD and BD in PD.

pT	N SNPS	EFFECT SIZE	P VALUE	GOODNESS OF FIT*
0.001	3326	230	0.33	2.15E-04
0.05	57594	369.9	0.74	2.58E-05
0.1	94392	1134	0.43	1.41E-04
0.2	152260	847	0.64	4.82E-05
0.3	198594	2200	0.29	2.48E-04
0.4	***237402***	***2880***	***0.20***	***3.66E-04***
0.5	269568	1716	0.49	1.07E-04

Table 40: PRS analysis for depression in PD (Proband sample)

Results of MDD PRS association with depression phenotype in the Proband PD sample. PRS for pT <0.4 has the highest R^2, indicating it represents the greatest proportion of variance. The p value = 0.20, indicating no significant association between MDD PRS and depression phenotype.

Table 40 shows the results of this analysis for depression in PD as measured in the Proband sample. The pT level which provided the best proportion of variance explained by the PRS is 0.4. At this pT, PRS explained 0.0037% of the difference between the depression score of the PD samples. PD samples with higher depression scores did not show a significantly higher PRS for MDD in this pT PRS test (P = 0.20, effect size = 2880*)*, indicating no shared genetic risk between depression in PD and MDD in this cohort.

4.3.3.2 PRS of MCP in PD Pain cohort

To test the genetic correlation between MCP and pain in PD, PRS for MCP was calculated within the Proband PD cohorts. The PRS was then tested for association with UPDRS pain phenotype. The UKBB cohort was not tested to avoid sample overlap between the base and target data, which can inflate the association Table 41 shows the results from the PRS analysis.

pT	N SNPS	EFFECT SIZE	P VALUE	GOODNESS OF FIT*
0.001	5788	816.965	0.33	1.95E-04
0.05	54500	6.611e+03	*0.046*	8.00E-04
0.1	85034	9.391e+03	*0.035*	8.97E-04
0.2	126332	1.026e+04	0.085	5.99E-04
0.3	158594	1.289e+04	0.070	6.64E-04
0.4	185662	1.831e+04	*0.024*	1.03E-03
0.5	***208280***	***2.094e+04***	***0.020***	***1.09E-03***

Table 41: PRS analysis for pain in PD (Proband sample)

Results of MCP PRS association with pain phenotype in the Proband PD sample. PRS for pT <0.5 has the highest R^2, indicating it represents the greatest proportion of variance. The p value = 0.024, indicating no significant association between MDD PRS and depression phenotype.

The pT level which provided the best proportion of variance explained by the PRS is pT < 0.5. At this pT, PRS explained 0.0011% of the difference between the depression score of the PD samples. PD samples with higher pain scores did show a significantly higher PRS for MCP in this pT PRS test (P = 0.024, effect size = 1.831 x 10^4). Whilst all tests did not indicate a significant association between MCP PRS and pain phenotype, the majority did align with the pT < 0.5 result which indicates a shared genetic risk between pain in PD and MCP in this cohort.

4.3.4 Two-sample Mendelian Randomisation

To establish if a causal relationship exists between pain and depression in PD, multiple two sample Mendelian randomisation tests were carried out. Each of the following sections details the results from each phenotype tested as an exposure.

4.3.4.1 Multisite Chronic Pain

General MCP was tested as an exposure with MDD and depression in PD (Proband) as outcomes. There were 37 independent IVs selected (F stat = 36.14). Two SNPs were missing from the MDD dataset with no appropriate proxies, and 3 were missing from the Proband dataset, with a further 3 excluded for incompatible alleles. MCP was found to have a positive significant causal effect on MDD (*ß* = 0.69, P = 3.79 x 10^{-7}, SE = 0.14). This was replicated across the MR RAPS and weighted median tests performed. However, the MR Egger test showed insignificant association, and although the *ß* was in the same direction, the effect was largely reduced compared to the sensitivity tests (Table 42, Figure 4-8A). This indicates that these instruments could be pleiotropic and affect MDD through pathways other than MCP, as originally suggested by Johnston et al (126). Significant heterogeneity was also detected (Table 43), which also indicates potential violation of the IV assumptions and that this is not a true causative association.

MCP was not observed to have a significant causal effect on depression in PD as measured in the Proband cohort (IVW *ß* = 1.60, P = 0.22, SE = 1.30) (Table 42, Figure 4-8B).

Samples		**MR**				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
Multisite chronic pain	Depression in PD (Proband)	MR Egger	31	2.45	6.27	0.70
		Weighted median	31	1.30	1.81	0.47
		IVW	31	1.60	1.30	0.22
		Simple mode	31	2.45	3.66	0.51
		Weighted mode	31	2.26	3.84	0.56
	MDD	MR Egger	35	0.01	0.68	0.99
		Weighted median	35	0.64	0.16	***9.42E-05***
		IVW	35	0.69	0.14	***3.79E-07***
		Simple mode	35	1.09	0.38	***7.00E-03***
		Weighted mode	35	1.00	0.41	***0.021***

Table 42: MR results for MCP exposure:

Two sample MR results for the effect of multisite chronic pain (MCP exposure on both depression in PD (UKBB and Proband) and major depressive disorder (MDD) outcomes. Significant causal relationships were observed for MDD as an outcome; however this is likely not a true causative association.

Samples		**Heterogeneity**			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
Multisite chronic pain	Depression in PD (Proband)	MR Egger	28.41	29	0.5
		IVW	28.43	30	0.55
	MDD	MR Egger	53.14	33	***0.015***
		IVW	54.85	34	***0.013***

Table 43: Heterogeneity results for multisite chronic pain (MCP) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW and MR Egger tests

Samples		Pleiotropy test			
Exposure	**Outcome**	**Method**	**Egger intercept**	**SE**	**P val**
Multisite chronic pain	Depression in PD (Proband)	MR Egger	-0.015	0.11	0.89
	MDD	MR Egger	0.012	0.011	0.31

Table 44: MR Egger intercept for MCP exposure

MR Egger intercept tests potential pleiotropic effects of instrumental variables in each two sample MR test.

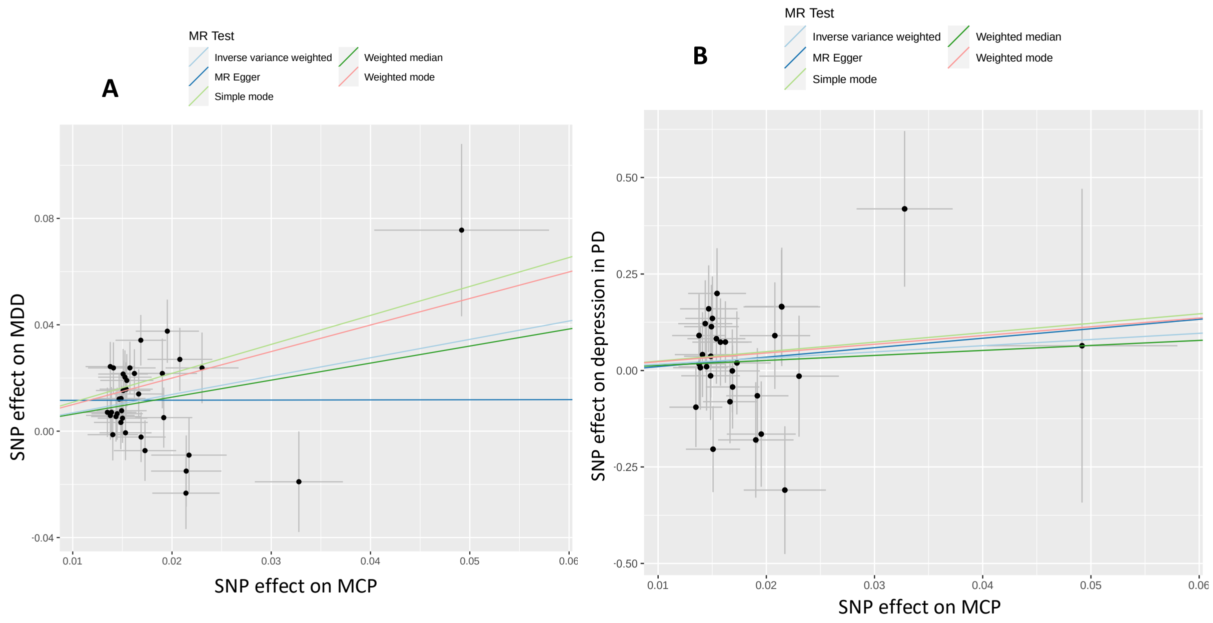


Figure 4-8 Two Sample MR Scatter plot for MCP exposure:

The SNP effect on the exposure (MCP) is plotted against the effect on each outcome, with lines fitted for each of the MR tests conducted. The outcomes of major depressive disorder (MDD) (A), depression in PD (Proband) (B) are shown. The slope of each line corresponds to the causal effect estimated by the test.

4.3.4.2 Major Depressive Disorder

MDD was also tested as an exposure on chronic pain outcomes, with 35 independent IVs selected for MDD including UKBB samples, and 44 independent IVs selected for the MDD exposure excluding UKBB samples (F stat = 37.01). All IVs were present in the MCP dataset, but the MCP in PD UKBB dataset had 9 missing SNPs and the Proband dataset had 7, with no suitable proxies. Similarly to the Johnston study, MDD was found to show a significant positive causative effect on MCP as tested in the IVW (*ß* =0.05, P = 1.86 x 10^{-4}, SE = 0.01). However, the MR Egger approach showed an insignificant effect and a reversal of the *ß* effect sign, indicating pleiotropic instruments and inconsistent causal estimates (Table 45, Figure 4-9A).

The test of heterogeneity also indicated significant heterogeneity amongst the instruments, also indicating that this is not a true causal effect (Table 46).

MDD also did not show any significant causal effect on multisite chronic pain in PD as measured in the UKBB cohort (IVW *ß* = 0.28, P = 0.12, SE = 0.18) (Figure 4-9B) or on pain in PD as measured in the Proband cohort (IVW *ß* = 0.24, P = 0.66, SE = 0.55) (Figure 4-9C).

Samples		**MR**				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
MDD	Multisite chronic pain	MR Egger	44	-0.03	0.05	0.53
		Weighted median	44	0.03	0.01	***0.021***
		IVW	44	0.05	0.01	***1.86E-04***
		Simple mode	44	0.03	0.03	0.34
		Weighted mode	44	0.03	0.03	0.30
	Pain in PD (Proband)	MR Egger	28	-2.65	3.52	0.46
		Weighted median	28	-0.13	0.77	0.87
		IVW	28	0.24	0.55	0.66
		Simple mode	28	2.760	1.82	0.14
		Weighted mode	28	-0.18	1.55	0.91
	MCP in PD (UKBB)	MR Egger	35	0.18	0.61	0.77
		Weighted median	35	-0.03	0.26	0.91
		IVW	35	0.28	0.18	0.12
		Simple mode	35	-0.43	0.60	0.48
		Weighted mode	35	-0.41	0.56	0.48

Table 45: MR results for MDD exposure:

Two sample MR results for the effect of major depressive disorder (MDD) exposure on both pain in PD (UKBB and Proband) and multisite chronic pain (MCP) outcomes. Significant causal relationships were observed for MCP as an outcome, however inconsistent ß estimate was observed in the MR Egger test of sensitivity.

Samples		**Heterogeneity**			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
MDD	Multisite chronic pain	MR Egger	112.99	42	***2.00E-08***
		IVW	121.81	43	***1.83E-09***
	Pain in PD (Proband)	MR Egger	25.04	26	0.52
		IVW	25.73	27	0.53
	MCP in PD (UKBB)	MR Egger	35.60	33	0.35
		IVW	35.63	34	0.39

Table 46: Heterogeneity results for major depressive disorder (MDD) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW and MR Egger tests. Significant heterogeneity amongst instruments used for the MR including MCP as the outcome was detected.

Samples		Pleiotropy test			
Exposure	Outcome	Method	Egger intercept	SE	P val
MDD	Multisite chronic pain	MR Egger	0.005	0.003	0.08
	Pain in PD (Proband)	MR Egger	0.09	0.11	0.41
	MCP in PD (UKBB)	MR Egger	0.006	0.035	0.87

Table 47: MR Egger intercept for MDD exposure

MR Egger intercept tests potential pleiotropic effects of instrumental variables in each two sample MR test.

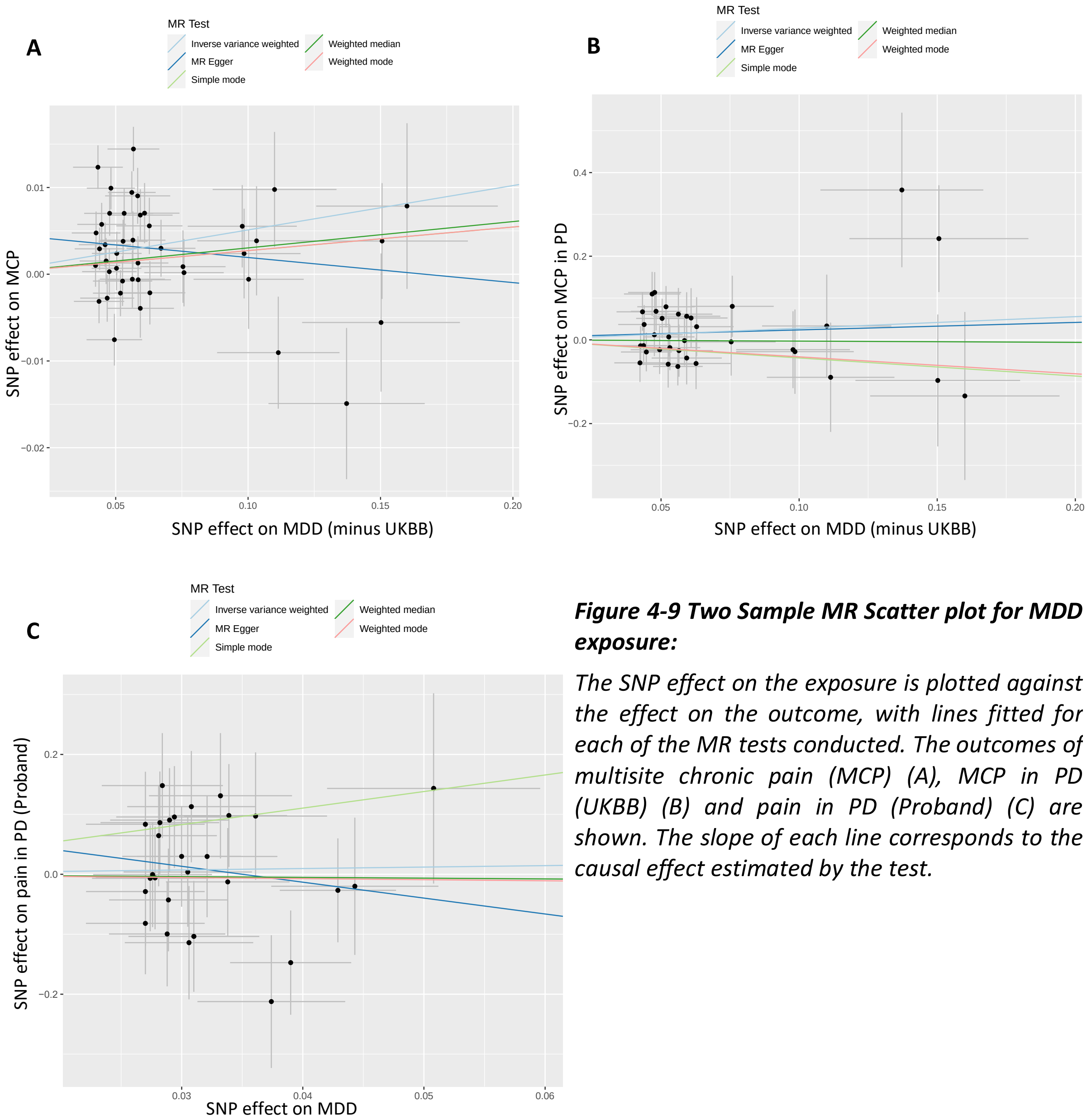


Figure 4-9 Two Sample MR Scatter plot for MDD exposure:

The SNP effect on the exposure is plotted against the effect on the outcome, with lines fitted for each of the MR tests conducted. The outcomes of multisite chronic pain (MCP) (A), MCP in PD (UKBB) (B) and pain in PD (Proband) (C) are shown. The slope of each line corresponds to the causal effect estimated by the test.

4.3.4.3 Pain in PD (Proband)

Pain in PD as measured in the Proband cohort was used as an exposure, with 7 independent SNPs out of the 106 passing the P value threshold ($P < 5 \times 10^{-6}$) used as instrumental variables (F stat = 24.51). One SNP was missing from the MDD dataset, with no proxies available. The results showed no evidence for a causative effect on either general MDD or depression in PD as measured in the UKBB cohort. The IVW test was used to assess the effect on MDD as the outcome (*ß* = 2.12×10^{-5}, P = 1, SE = 0.01), and also with depression in PD as the outcome (*ß* = 7.17×10^{-3}, P = 0.94, SE = 0.09), both of which were insignificantly associated (Table 48) (Figure 4-10).

Samples		MR				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
Pain in PD (Proband)	BD in PD (UKBB)	MR Egger	7	-1.02E-02	0.6	0.99
		Weighted median	7	-8.86E-03	0.1	0.93
		IVW	7	7.17E-03	0.09	0.94
		Simple mode	7	-9.76E-02	0.16	0.56
		Weighted mode	7	-2.31E-02	0.15	0.88
	MDD	MR Egger	6	1.95E-02	0.06	0.77
		Weighted median	6	-4.66E-03	0.01	0.67
		IVW	6	2.12E-05	0.01	1
		Simple mode	6	-1.03E-02	0.02	0.53
		Weighted mode	6	-9.52E-03	0.01	0.49

Table 48: MR results for Pain in PD (Proband) exposure:

Two sample MR results for the effect of the pain in PD (Proband) exposure on both major depressive disorder (MDD) and depression in PD (UKBB) outcomes indicate no causative associations are present.

Samples		Heterogeneity			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
Pain in PD (Proband)	BD in PD (UKBB)	MR Egger	8.21	5	0.14
		IVW	8.21	6	0.22
	MDD	MR Egger	5.99	4	0.19
		IVW	6.14	5	0.29

Table 49: Heterogeneity results for Pain in PD (Proband) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW test indicate no significant heterogeneity within the instrumental variables.

Samples		Pleiotropy test			
Exposure	**Outcome**	**Method**	**Egger intercept**	**SE**	**P val**
Pain in PD (Proband)	BD in PD (UKBB)	MR Egger	0.01	0.34	0.98
	MDD	MR Egger	-0.01	0.03	0.77

Table 50: MR Egger intercept for Pain in PD (Proband) exposure

MR Egger intercept to test potential pleiotropic effects of instrumental variables in each two sample MR test indicated no significant pleiotropy within the two MR tests.

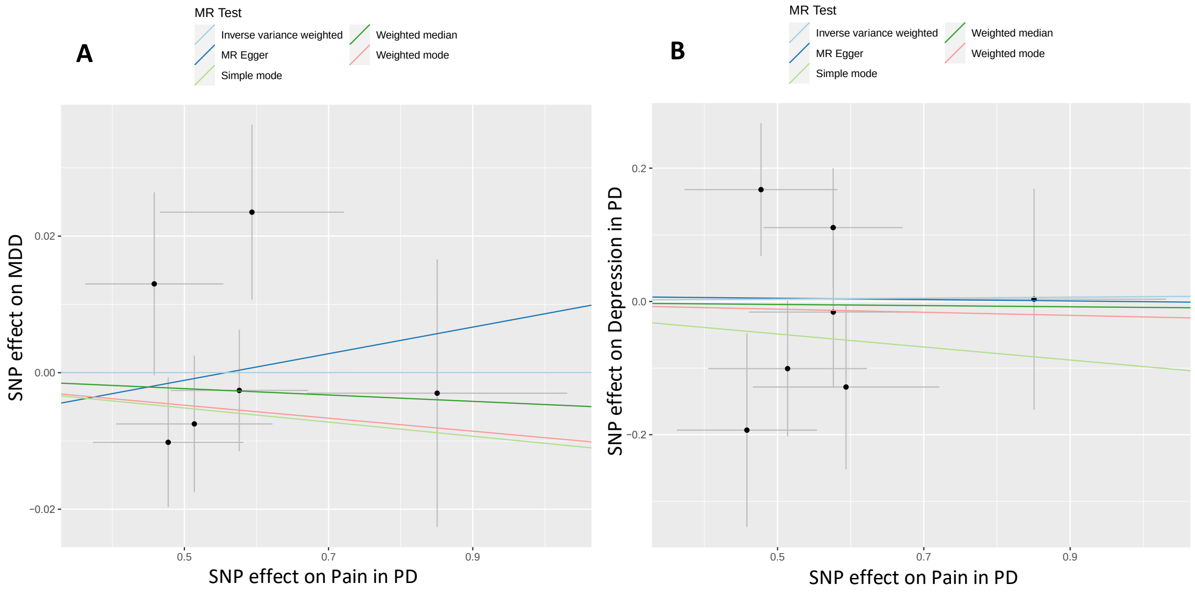


Figure 4-10: Two Sample MR Scatter plot for Pain in PD (Proband) exposure:

The SNP effect on the exposure is plotted against the effect on the outcome, with lines fitted for each of the MR tests used. The outcomes of major depressive disorder (MDD) (A) and depression in PD (B) are shown. The slope of each line corresponds to the causal effect estimated by the test. No significant causal association was identified.

4.3.4.4 Multisite Chronic Pain in PD (UKBB)

MCP in PD as measured in the UKBB cohort was then used as an exposure to test the causal effect on MDD (excluding UKBB samples) and Depression in PD (Proband) as outcomes. 16 independent IVs out of 49 passing the P value threshold (P < 5 x 10^{-6}) were selected for this exposure (F stat = 22.15). 8 SNPs were missing from the MDD data with 4 appropriate proxy variants available, so 12 IVs were used in total. 9 SNPs were present in the depression in PD Proband samples with one suitable proxy variants present, so 10 IVs were used in total. MCP in PD was not found to have a causal effect on MDD as measured by the IVW (*ß* = -2.47 x 10^{-3}, P = 0.87, SE = 0.01), (Table 51) (Figure 4-11A), and likewise on depression in PD (*ß* = 0.16, P = 0.43, SE = 0.21). (Table 51) (Figure 4-11B).

Samples		MR				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
MCP in PD (UKBB)	Depression in PD (Proband)	MR Egger	10	0.21	0.63	0.74
		Weighted median	10	-0.07	0.23	0.76
		IVW	10	0.16	0.21	0.43
		Simple mode	10	-0.17	0.38	0.66
		Weighted mode	10	-0.17	0.35	0.64
	MDD	MR Egger	12	-1.54E-02	0.03	0.66
		Weighted median	12	6.63E-03	0.02	0.73
		IVW	12	-2.47E-03	0.01	0.87
		Simple mode	12	4.70E-03	0.03	0.86
		Weighted mode	12	1.14E-02	0.03	0.66

***Table 51: MR results for MCP in PD (UKBB) exposure**:*

Two sample MR results for the effect of the multisite chronic pain (MCP) in PD (UKBB) exposure on both major depressive disorder (MDD) and depression in PD (Proband) outcomes indicate no causal associations are present.

Samples		Heterogeneity			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
MCP in PD (UKBB)	Depression in PD (Proband)	MR Egger	15.71	8	0.047
		IVW	15.73	9	0.073
	MDD	MR Egger	8.49	10	0.58
		IVW	8.67	11	0.65

Table 52: Heterogeneity results for MCP in PD (UKBB) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW test indicate no significant heterogeneity within the instrumental variables.

Samples		Pleiotropy test			
Exposure	**Outcome**	**Method**	**Egger intercept**	**SE**	**P val**
MCP in PD (UKBB)	Depression in PD (Proband)	MR Egger	-0.016	0.19	0.94
	MDD	MR Egger	0.005	0.012	0.68

Table 53: MR Egger intercept for MCP in PD (UKBB) exposure

MR Egger intercept to test potential pleiotropic effects of instrumental variables in each two sample MR test indicated no significant pleiotropy within the two MR tests.

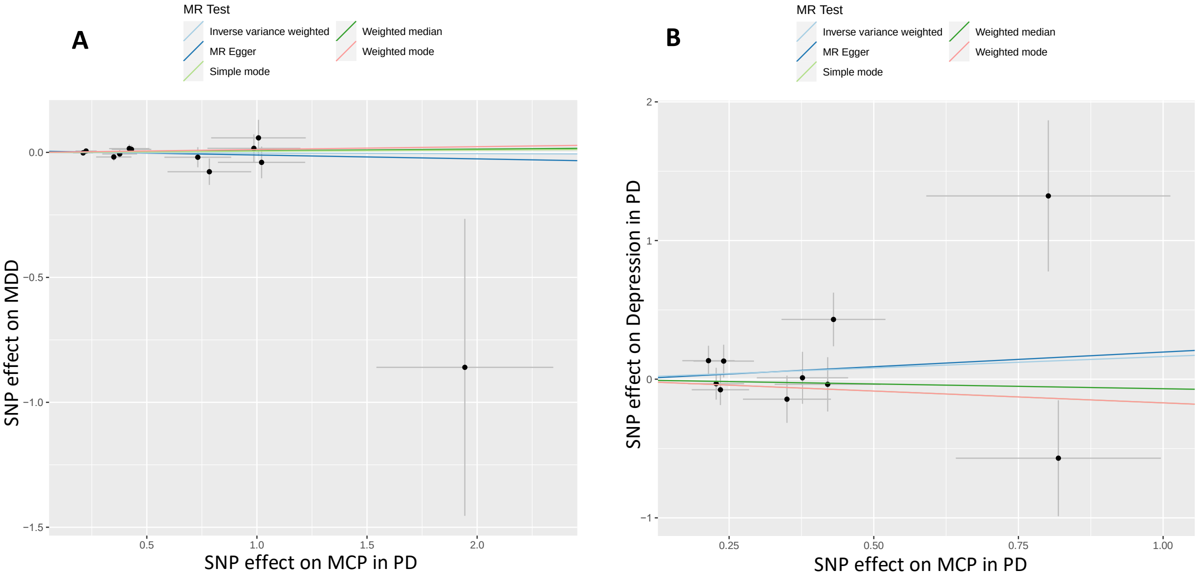


Figure 4-11: Two Sample MR Scatter plot for MCP in PD (UKBB) exposure:

The SNP effect on the exposure is plotted against the effect on the outcome, with lines fitted for each of the MR tests conducted. The outcomes of major depressive disorder (MDD) (A) and depression in PD (B) are shown. The slope of each line corresponds to the causal effect estimated by the test. No significant causal association was identified.

4.3.4.5 Depression in PD (UKBB)

Six independent IVs were selected that passed the P value threshold (P < 5 x 10^{-6}) for the exposure of broad depression (BD) in PD as measured by the UKBB cohort (F stat = 23.56). Three of these were present in the Proband dataset, with no appropriate proxies available. This exposure was found to not have a causative effect on pain in PD as measured in Proband cohort (*ß* = 0.10, P = 0.61, SE = 0.19) (Table 54) (Figure 4-12). This was reflected across all other sensitivity analyses.

Samples		**MR**				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
BD in PD (UKBB)	Pain in PD (Proband)	MR Egger	3	-3.33	1.69	0.30
		Weighted median	3	0.03	0.17	0.87
		IVW	3	0.10	0.19	0.61
		Simple mode	3	7.41E-04	0.23	1.00
		Weighted mode	3	-0.02	0.22	0.92

Table 54: MR results for Depression in PD (UKBB) exposure:

Two sample MR results for the effect of the Broad depression (BD) in PD (UKBB) exposure on pain in PD (Proband). No significant causal relationships were observed.

Samples		**Heterogeneity**			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
BD in PD (UKBB)	Pain in PD (Proband)	MR Egger	0.12	1	0.76
		IVW	6.58	2	0.12

Table 55: Heterogeneity results for Depression in PD (UKBB) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW and MR Egger tests indicate no significant heterogeneity within the instrumental variables.

Samples		**Pleiotropy test**			
Exposure	**Outcome**	**Method**	**Egger intercept**	**SE**	**P val**
BD in PD (UKBB)	Pain in PD (Proband)	MR Egger	1.36	0.67	0.29

Table 56: MR Egger intercept for Depression in PD (UKBB) exposure

MR Egger intercept tests potential pleiotropic effects of instrumental variables in each two sample MR test indicated no significant pleiotropy within the two MR tests.

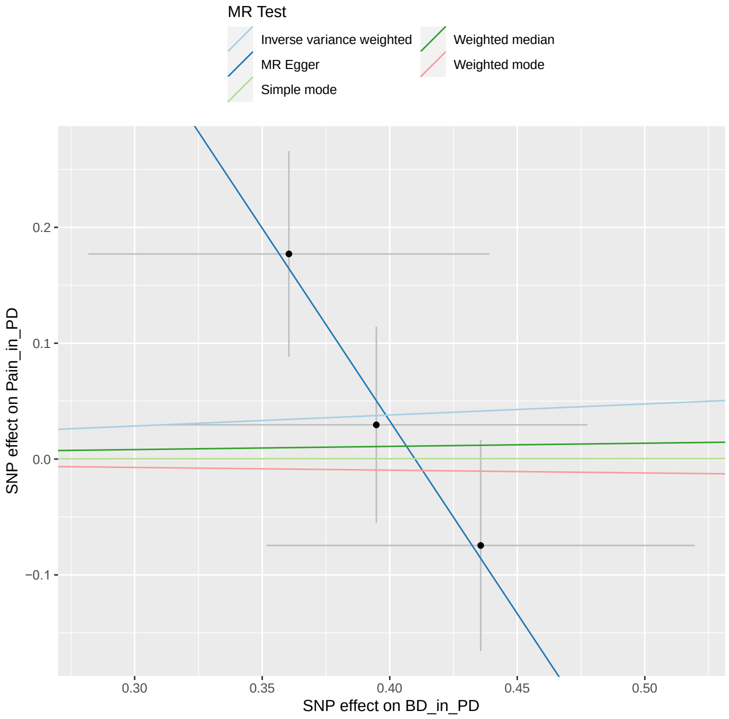


Figure 4-12: Two Sample MR Scatter plot for Depression in PD (UKBB) exposure:

The SNP effect on the exposure (Broad depression in PD) is plotted against the effect on the outcome (Pain in PD), with lines fitted for each of the MR tests used. The slope of each line corresponds to the causal effect estimated by the test.

4.3.4.6 Depression in PD (Proband)

8 independent IVs were selected out of the 29 SNPs that passed the P value threshold ($P < 5 \times 10^{-6}$) for the exposure of depression in PD as measured in the Proband cohort (F stat = 21.97). In both MCP outcome datasets, two IVs were missing with no appropriate proxies available. This exposure was found not to have a significant causative effect on MCP (IVW *ß* = -3.88×10^{-3}, P = 0.059, SE = 2.05×10^{-3}). This was reflected in all other tests performed, with *ß* sign swapping in half of the sensitivity tests reflecting an inability to identify a causative pathway (Table 57, Figure 4-13B).

When MCP in PD was tested as the outcome, there was similarly no evidence for a causative effect of Depression in PD on MCP in PD (IVW ß = 0.05, P = 0.20, SE = 0.04). (Table 57, Figure 4-13A).

Samples		**MR**				
Exposure	**Outcome**	**Method**	**No. SNP**	**Beta**	**SE**	**P val**
Depression in PD (Proband)	MCP in PD (UKBB)	MR Egger	6	0.14	0.12	0.31
		Weighted median	6	0.06	0.05	0.22
		IVW	6	0.05	0.04	0.20
		Simple mode	6	0.08	0.08	0.35
		Weighted mode	6	0.08	0.07	0.34
	MCP	MR Egger	6	-6.74E-03	6.84E-03	0.38
		Weighted median	6	-2.65E-03	2.69E-03	0.32
		IVW	6	-3.88E-03	2.05E-03	0.059
		Simple mode	6	-1.77E-03	4.20E-03	0.69
		Weighted mode	6	-1.77E-03	4.27E-03	0.70

Table 57: MR results for Depression in PD (Proband) exposure:

Two sample MR results for the effect of the depression in PD (Proband) exposure on both multisite chronic pain (MCP) in PD (UKBB) and MCP outcomes. Significant causal relationships were observed for MCP in PD as an outcome.

Samples		**Heterogeneity**			
Exposure	**Outcome**	**Method**	**Q**	**Q df**	**Q pval**
Depression in PD (Proband)	MCP in PD (UKBB)	MR Egger	2.17	4	0.70
		IVW	2.79	5	0.73
	Multisite chronic pain	MR Egger	4.46	4	0.35
		IVW	4.68	5	0.46

Table 58: Heterogeneity results for Depression in PD (Proband) exposure

Q statistic results to test heterogeneity of instrumental variables used in the IVW and MR Egger tests

Samples		Pleiotropy test			
Exposure	**Outcome**	**Method**	**Egger intercept**	**SE**	**P val**
Depression in PD (Proband)	MCP in PD (UKBB)	MR Egger	-0.06	0.08	0.48
	Multisite chronic pain	MR Egger	1.91E-03	4.33E-03	0.68

Table 59: MR Egger intercept for Depression in PD (Proband) exposure

MR Egger intercept tests potential pleiotropic effects of instrumental variables in each two sample MR test.

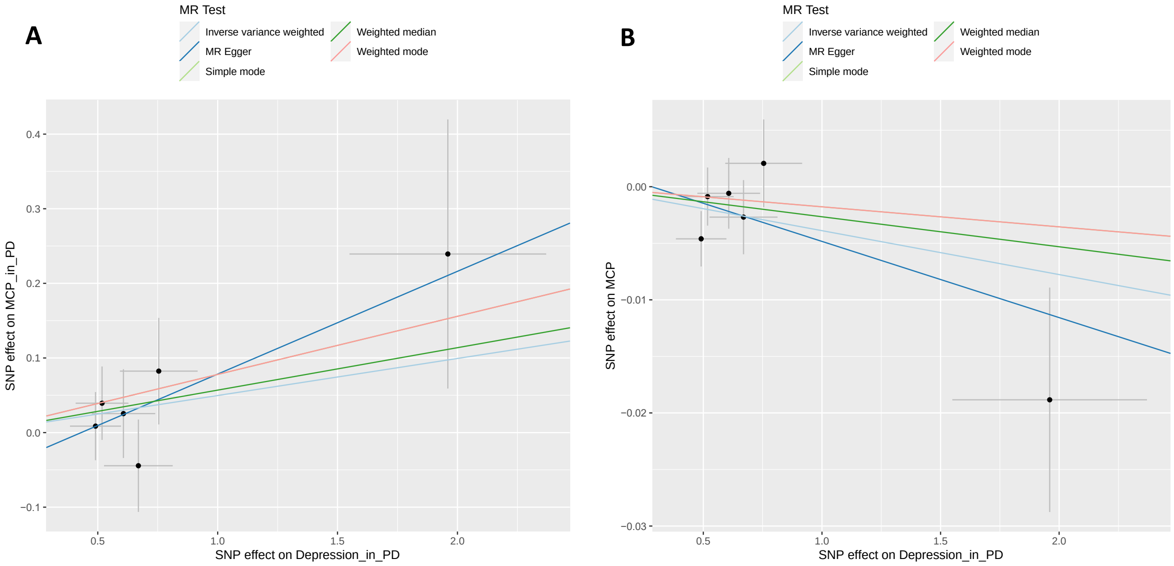


Figure 4-13: Two Sample MR Scatter plot for depression in PD (Proband) exposure:

The SNP effect on the exposure is plotted against the effect on the outcome, with lines fitted for each of the MR tests conducted. The outcomes of multisite chronic pain (MCP) in PD (A) and MCP (B) are shown. The slope of each line corresponds to the causal effect estimated by the test.

4.4 Discussion

The aim of this investigation was to examine the genetic factors influencing pain and depression as non-motor symptoms in Parkinson's disease and explore the possibility of a causal relationship between these symptoms. Association analysis found a replication in both UKBB and Proband PD cohorts of a correlation existing between these symptoms, an observation that has been previously observed in other PD cohorts (4). It was then investigated whether these were independent yet correlated symptoms or if a causative pathway existed, with the aim of furthering our understanding of causes of non-motor symptoms in PD.

Initially, three separate GWAS studies were conducted to identify genetic factors influencing pain and depression in PD. These tested PD depression in the Proband cohort, and PD multisite chronic pain (MCP) and broad depression (BD) in the UKBB cohort. In each of these tests there were no genome wide significant associations identified, which is potentially due to the small sample sizes that were available for these cohorts resulting in underpowered GWAS.

However, putative associated SNPs of potential biological interest were identified across these GWAS. In the pain in PD GWAS in the UKBB cohort, the top associated SNP is an intronic variant within NRG1, expression levels of which have been related to neuropathic pain in a rat model (130). The pathway proposed by this model describes how a reduction in NRG1 secretion from large dorsal root ganglion (DRG) neurons following nerve injury can result in disinhibition of C-fibre-mediated nociceptive signalling, causing chronic pain conditions. This model of neuropathic pain focusing on peripheral disinhibition could be inconsistent with the proposal that central pain processes related to the monoamine dysregulation in PD are more significantly impacting pain processing in PD. However, this is worth further investigation, and considering TRPM8 which has been previously implicated in PD pain from the pain in PD GWAS (13) is also expressed in DRG neurons, it is worth establishing the potential role of the DRG neuron circuit in PD pain.

Similarly with depression in PD, the two GWAS performed revealed potential associations of interest. In the PROBAND cohort GWAS which tested genetic associations with self-reported depression scores, the most significantly associated SNP is an intronic variant in

ZNF33B, expression levels of which have been associated with suicide attempts in major depressive disorder. A study into MDD patients with and without suicide attempts (SA) found that a copy number variant (CNV) was more associated with SA subjects, with this CNV impacting expression levels of ZNF33B in the pituitary and cerebellum (128). This observation is of interest, as zinc finger proteins have been previously associated with various neurological disorders. ZNF33B variants have been associated with bipolar disorder risk, and further members of the zinc finger family have been identified as risk genes for various disorders. ZNF804A has been established as a risk gene for schizophrenia, with recent evidence suggesting ZNF804A polymorphisms are involved in maintenance of neuronal function and integrity of white matter fibre bundles (131). Whether the processes that zinc finger proteins regulate in these neurological disorders is also relevant in PD depression is yet to be determined, but could be an important avenue of investigation.

In the UKBB broad depression GWAS, the association at chromosome 12 located within KDM5A, which is also an eQTL for KDM5A, is of particular interest. Hypothalamic expression levels of this gene have been linked to post-partum depression in a rat model (129). Hypothalamic transcript levels of KDM5A, which is involved in circadian rhythm regulation, were greater in the depression model rats, which also demonstrated a loss of diurnal rhythms. This change in diurnal patterns and circadian rhythm gene expression could be important in PD depression. Sleep disturbances in PD are recognised as one of the key non-motor symptoms, and a recent investigation observed that poorer sleep in PD patients was correlated with depression severity (132), suggesting this could be having an impact on mood. It is therefore worth investigating this genetic risk factor and establishing whether PD sleep disturbances impact PD depression.

The lack of genome wide significant associations across these GWAS could be attributed to the small sample sizes used. The power of a GWAS to detect genome wide significant associations is dependent upon the sample size, the MAF of the associated SNP, prevalence of the disease, and the Genotype Relative Risk (GRR). GRR refers to how the number of copies of the risk allele carried (0, 1, or 2) affects the probability of an individual being in the case group i.e. whether the risk allele is additive, dominant, recessive, or multiplicative. Smaller GWAS samples can detect some genome wide significant associations, which was demonstrated in the recent Pain in PD GWAS (13). This study, using 898 PD patients with high pain and 420 PD patients with no/low pain, was able to detect the SNP association at the TRPM8 locus (P = 1.45 x 10^{-9}, OR = 1.78). While it is therefore not inappropriate to conduct GWAS with relatively small sample sizes, this had the potential to be a common limiting factor across the GWAS conducted in the present study. For example, the GWAS for depression in PD in the UKBB sample was conducted with 563 PD patients with a broad depression phenotype, and 890 PD patients with no history of depression. For a hypothetical associated allele with minor allele frequency of 0.1, disease prevalence of 0.4, and genotype relative risk of 1.5, the power (probability of rejecting the null hypothesis) to detect an associated allele at the significance level of P < 5 x 10^{-8} using this sample is 0.653 (133). If the sample size were to be increased to 800 PD cases with depression and 1000 without, this power would be increased to 0.914. This demonstrates the greater potential to identify genetic factors influencing these non-motor symptoms when even marginally larger samples are obtained.

Polygenic Risk Scores (PRS) were calculated within the Proband and UKBB PD cohorts for MDD and MCP, which could then be tested for association with their non-motor pain and depression phenotypes. No strong evidence was found for a correlation of genetic factors influencing depression in PD and MDD, yet there was more evidence for correlation between pain in PD and MCP. The lack of evidence within the depression phenotype test could also be due to the underpowered sample size. Given the top genetic risk factors for pain and depression in PD identified in the GWAS indicated some potential shared biological mechanisms, it could be expected that a greater amount of polygenic risk factors would be shared between depression in PD and MDD than could be detectable in this study. If future studies with improved power can more decisively indicate whether polygenic risk is shared between PD non-motor symptoms and non-PD related pain and depression, this would help determine whether investigating shared biological pathways and treatment options is appropriate.

Two sample Mendelian randomisation was performed using the results from the GWAS conducted here along with existing GWAS data for MCP, MDD and pain in PD to establish whether causal relationships exist between these PD symptoms. The putative associations identified in the GWAS conducted were sufficient to select instrumental variables for use in MR analysis. Although genome-wide significant instruments can be preferable as they are more strongly associated with the exposure, selection of all associated SNPs above a suggestive significant threshold can be implemented in MR studies to avoid winners curse bias in selection and represent a greater share of genetic influences. Furthermore, identification of potential genes of biological significance at these associated SNPs suggests some biological relevance to the exposure is present in these instruments, which adds confidence to their selection as IVs.

Similarly to the results from Johnston et al, significant and positive causative associations were observed for both MCP on MDD and MDD on MCP. It was also observed that opposite *ß* signs of the causative estimate occurred for MDD as the exposure, suggesting this is also not a causative estimate. However, whilst the effect of MCP on MDD was more consistent, the MR Egger analysis also suggested the presence of pleiotropic effects, in line with the earlier observations of Johnston. Rather than confirming this reported causative association, this suggests that there may not be a true causative association of MCP on MDD, but rather these instrumental variables act through other confounding pathways to impact MDD. Significant heterogeneity amongst the IVs for this MR test further suggests that this could be the case, as this is an indicator of a breach of the IV assumptions. However, this would need to be investigated further to establish the nature of these pleiotropic effects.

The results from the remainder of the MR tests also indicated no causative pathways exist between pain in PD and depression in PD, or between these non-motor PD symptoms and non-PD related pain and depression. These negative results could also be due to the small sample sizes of the GWAS leading to insufficient power to detect all the appropriate IVs, and consequently performing MR tests with a small number of IVs.

Several limitations of this investigation could possibly render the results from this MR approach misleading, so should be explored. As discussed, the instrumental variables used potentially do not represent a significant proportion of variance in the exposure phenotype, which could be the case given there were not genome wide significant associations in the GWAS performed. The F statistics indicate the tests do all avoid weak instrument bias (all F statistics > 10), but these could still not explain significant variance. However, weak instrument bias from underpowered GWAS would bias in favour of the null in this instance, so positively associated results can be better trusted and indicate this bias, if present, has been insignificant.

Selection or collider bias may also be an issue here, which will instead bias in favour of type I errors. This can occur if there is significant bias due to sample selection i.e. if selection into the sample is dependent on the risk factor or the outcome, or depends on a 'collider' of the instrumental variable and potential confounding factors. This is mainly an issue where the bias due to selection is large, meaning the samples used differ from the populations they represent. It has been observed that for moderate influence of risk factor on sample selection, selection bias is small, and the type I error rate is not increased (134). As the PD populations recruited can be representative of the general PD population rather than those with specific risk of non-motor symptoms, this bias should be minimal.

Whilst the consistent negative results from these tests could reflect the disadvantages of the current method as described above, consideration should be given to the potential that these are fully valid and indicate independent symptoms that are correlated but not causative. This conclusion can also apply to chronic pain and depression in the general population, given the conflicting results of the previous MR study.

The correlation between these symptoms can be due to some of the overlapping potential causes arising from PD pathology as discussed previously; monoamine depletion, neuroinflammation, and degeneration of limbic structures. Taking into account the pleiotropic effects suggested in the previous MR study, no evidence to date adequately supports a causative pathway of pain on depression or vice versa in PD or in the general population. The most compelling evidence of a causative pathway remains animal studies, where models of neuropathic and inflammatory pain consistently induce depression-like behavior (135). However, these observations are under very different conditions to human conditions and there is doubt as to the extent to which depressive like behaviors in animals can fully reflect depression in humans. In comparison, depression models have shown mixed results in the resulting influence they have on pain sensitivity (135). Combined with the results from human genetic studies here, evidence currently indicates these are correlative but not causative.

As the genetic factors influencing pain and depression in PD identified here also indicate these symptoms are independent rather than causative, focusing on the separate biological pathways that these indicate are the most associated with these symptoms could be the best approach to improving therapeutic targets. Evidence indicates pain in PD is associated with the TRPM8 locus encoding an ion channel expressed in DRG neurons that acts as a cannabinoid receptor amongst other roles. Exploring cannabinoid therapies as well as other TRPM8 ligands could be the appropriate next step in improving PD pain targets, with the benefit of being able to repurpose existing drugs. Depression in PD alternatively is potentially influenced by KDM5A and ZNF33B expression levels, which would require further investigation for an appropriate therapeutic intervention. Improving the therapy options for these symptoms by focusing on the most appropriate target rather than on reduction of pain or depression as a causative factor could be the best approach to tackling these symptoms. Whilst it is still the case that responses to treatment for pain or depression could be affected by experience of the other symptoms as indicated in the general population, until there is more compelling evidence for a causative pathway it cannot be assumed that targeting one will alleviate the other.

It is worthwhile working to increase the power of these GWAS and MR tests with greater PD samples so that the true genetic risk factors for these symptoms can be substantiated, and any true causal effects, if present, could be estimated. This is especially of importance as it relates to the experiences and therapies received by PD patients. Monitoring of non-motor symptoms would be of clear benefit if for example those reporting high PD were then able to receive preventive interventions for mood disorders. However, greater understanding of the biological mechanisms behind both non-motor symptoms is still required so that the most suitable intervention targets can be identified, and preventative PD pain and depression therapies applied in an appropriate way. This can then go towards addressing some of the most troublesome aspects of the disease for PD patients.

5 General Discussion

5.1 Overview

Parkinson's disease is a neurodegenerative disorder which primarily affects the elderly population. It is characterised by loss of motor control caused by degeneration of the basal ganglia; however PD also impacts many aspects of health including pain, neuropsychiatric symptoms, and inflammation. The aim of the work in this thesis was to investigate the genetic factors influencing depression, pain, and inflammation in PD, and to study the associations between these PD non-motor symptoms. Two main experimental approaches were pursued: the first was to generate novel long-read sequencing data of the HLA locus in PD to allow observation of any polymorphisms detectable by this superior HLA sequencing application, and the second aim was to conduct GWAS to establish genetic associations with depression and pain in PD, and consequently establish if a causative relationship exists between these symptoms using an MR approach. These aims were pursued in order to better understand the underlying causes of these PD symptoms which are currently imperfectly understood, and aid future work in targeting therapies more appropriately for these symptoms.

The following sections will summarise the findings from each experimental chapter, and how these have been interpreted.

5.2 Summary and Interpretation of Findings

5.2.1 Chapter 2

The main aim of Chapter 2 was to establish which HLA loci are most associated with PD risk, to subsequently select which loci to focus on for HLA sequencing. Previous publications had suggested specific HLA loci of interest, but without a thorough comparison between bioinformatics approaches or incorporating the most recent imputation application. Here, a range of different bioinformatics approaches were applied to a PD case-control dataset, which allowed comparison with other published association results from different PD datasets. Firstly a GWAS and conditional analysis were conducted, indicating that there were potentially two independent associations at the HLA locus; rs9268926 was the most associated GWAS result ($P = 3.67 \times 10^{-7}$, OR = 0.84), which is in partial LD with the main SNP from the Nalls 2019 meta-analysis (rs112485576 (R^2 = 0.76, D' = 0.89)) (80), and rs9295987 was the most associated SNP when conditioning on this result *(*P = 9.82 x 10^{-5}, OR = 0.80). The top association is located near the class II DR loci, while the top independent association from the conditional analysis is located near the class I B and C loci. QTL properties of these SNPs also indicated association with these loci, with rs9268926 a potential QTL for HLA-DQA1/DRB1 loci, and rs9295987 a QTL for HLA-B and HLA-C.

The HLA loci and specific HLA alleles that these SNPs were associated with were analysed using imputation and QTL approaches. Varying results were achieved across imputation methods, which applied different computational approaches. Whilst the SNP2HLA approach indicated a DRB1 allele that had not been previously associated with PD was the top associated allele (HLA-DRB1*13:01, P = 1.49 x 10^{-5}, OR = 0.83), the HIBAG approach replicated a previously associated allele as the top associated (HLA-DRB1*04:01 P = 1.56 x 10^{-6}, OR = -0.03) (87). The novel DISH method which has not previously been applied to a PD dataset also indicated HLA-DQA1*03:01 as the top associated allele, which is in strong LD with DRB1*04:01 and has been previously associated with PD (87). The conditional analysis result was also observed to be in LD with HLA-B*35:01 and HLA-C*04:01 alleles.

The overall interpretation of these findings was that they provide further evidence from a new dataset for the associations at DRB1 and DQA1 loci associated with PD protection, that had been identified in previous imputation studies. However, comparison between multiple methods showed that discrepancies can occur when different panels are applied, highlighting the limitations with relying on any single imputation approach. Whilst the DRB1*13 was a novel association identified, and evidence exists for the importance of this association in other neurodegeneration conditions (96,136), this could also be a product of imputation error as the smallest panel was applied in this situation. Furthermore, whilst the top associated result from the conditional analysis here is of interest, it does not reach a sufficiently high significance level to indicate a definite independent association, demonstrating that a greater powered study needs to be applied to identify the true association here. Despite the limitations of the methods used here, DRB1, DQA1, B and C loci were repeatedly found to be associated with PD risk across methods, giving confidence that these were the most likely candidates to further explore via sequencing.

5.2.2 Chapter 3

The aim of Chapter 3 was to conduct long-read sequencing of the HLA region in PD patient samples, which would allow testing for the presence of PD-associated polymorphisms that were undetectable by imputation or by short-read sequencing methods, and to fully characterise which HLA alleles are associated with the Nalls 2019 SNP. Sequencing of the HLA locus in PD samples has only been conducted in one previous study, which applied a short-read sequencing approach with limited ability to capture HLA structural variations(108). In this chapter, 70 PD samples were used for PacBio long-read sequencing, with data generated for full length HLA-B, C, DQA1, and DRB1 loci. Samples were either homozygous for the risk allele (C) or protective allele (A) of the top associated HLA SNP.

Following on from this, the secondary aim was to conduct a further PD association analysis using HLA allele data imputed using the latest panel, released while these experiments were being carried out (102). This imputation approach is a large multi-ancestry panel that can improve upon imputation quality and address some of the previous issues with panel sample size and allele variety. This data allowed comparison of the best method sequencing approach to a case-control situation in a larger cohort.

Analysis of sequencing data demonstrated allele associations with the rs112485576 protective and risk alleles. It was observed that DQA1*03:01, DQA1*03:03, DRB1*04:01, DRB1*04:04 were correlated with the minor protective allele, as previously reported. HLA alleles were also identified as correlated with the risk allele, with DQA1*02:01 the most associated. One novel allele was identified via sequencing that differed from the HLA-DQA1*01:01:01:01 IPD-IMGT/HLA database entry by two intronic SNPs, however as this was an individual identification it could not be tested for association with PD risk.

Comparison with the case-control imputation results revealed which correlated alleles were also associated with PD in this analysis. It was observed that DQA1*03:01 was the top protective allele (P = 4.4 x 10^{-8}, OR = 0.83), and DQA1*01 (P = 2.1 x 10^{-4}, OR= 1.11) the top risk allele in this case-control dataset. These associations are consistent across the sequencing and imputation data, giving greater confidence that these alleles are the most significant. HLA-DQA1*03:01 allele diplotypes were observed to consistently confer protection for the disease, whilst risk alleles only conferred risk in the absence of DQA1*03:01.

The sequencing results describe alleles that were observed to be correlated with the rs112485576 alleles, the most associated SNP in the latest meta-analysis. This did not analyse potentially independent associated SNPs. However, in the case-control dataset, associations independent of rs112485576 were also able to be observed. This conditional investigation found that C*02:02 (P = 2.52 x 10^{-3}, OR = 1.15) and DPB1*01:01 (P = 0.013, OR = 1.15) were top candidates for independent associations. This suggests that the C locus allele could be driving the class I association. DPB1 has not previously been associated with PD, and as this was a suggestive significant finding, it would need to be replicated before deciding if this is of importance.

Overall, these findings provided evidence from a sequencing approach that corroborates the association of the DQA1*03/DRB1*04 haplotype, whilst providing novel data to suggest which alleles can confer the greatest PD risk. No unknown structural variants were observed in these PD samples; it is possible the sample size was too small, however there is no evidence here for unknown structural variants affecting PD risk or protection. The imputation results here do suggest that the main locus that is driving this association, previously suggested to be DRB1*04, could require further investigation. Multiple previous methods have identified the DRB1 shared epitope as driving PD protection (87,108), yet this study indicated that the DQA1*03 amino acid positions are most associated. Overall, this implies the properties of the DQA1*03 association merit further investigation.

5.2.3 Chapter 4

The aim of this chapter was to investigate the genetic factors influencing pain and depression in PD, and their potential causal interactions. Firstly, regression analyses were conducted which indicated that pain and depression in PD are associated symptoms in two PD cohorts, a replication of previous findings (4,10). Following this, various GWAS of pain and depression in PD were conducted to identify the genetic associations with these symptoms in PD samples from Proband and UKBB cohorts. Whilst no genome-wide significant associations were observed, certain putative genetic associations indicated genetic loci that are potentially associated with these symptoms: NRG1 in multisite chronic pain in PD, and KDM5A, ASXL2, and ZNF33B in depression in PD. The role of NRG1 in DRG mediated neuropathic pain suggests a potential shared pathway between neuropathic pain and MCP in PD (130). Furthermore, NRG1-ErbB4 signalling has been suggested to mediate inflammatory pain symptoms in an animal model (137). Similarly, the role of KDM5A in regulating sleep patterns and its association with post-partum depression (129), as well as the association of ZNF33B with suicide attempts (128), also suggests pathways of interest for depression in PD.

These findings indicate a genetic underpinning of these two non-motor symptoms, which could potentially determine which PD patients are more likely to develop pain and depression symptoms. By identifying which of these genetic factors have also been identified as risk factors with MDD and pain in the general population, the potential shared biological processes could be identified. However, it can also be the case that factors influencing these symptoms are independent of non-PD related disorders.

To further investigate this question of shared genetic risk between PD non-motor symptoms and pain and depression in the general population, a polygenic risk score (PRS) analysis was conducted. This identified that the broad depression phenotype tested in the UKBB sample had a shared genetic risk with MDD (P = 0.024, *ß* = 906.79*)*, however this was not replicated across other PRS tests. Furthermore, LADS score depression phenotype as measured in the Proband phenotype did not indicate a shared genetic risk (P = 0.20*, ß* = 2880*)*. This was therefore inconclusive as to the extent to which genetic risk is shared. In the case of pain, there was greater evidence for the shared genetic risk between pain in PD as measured by the UPDRS pain score and MCP in the general population (P = 0.024, *ß* = 18,310)

The results obtained in the first part of this investigation were then used for a two sample Mendelian randomisation study. Previously a causative association had been suggested for multisite chronic pain on MDD in the general population using a similar method (126). Here, a TSMR study was conducted which aimed to replicate this finding and then determine if this was also the case for PD pain and depression.

Interestingly, the results did not indicate a causative association between depression and pain in the general or the PD population. Whilst this study did replicate the causative association observed in the general population of MCP on MDD using a different MR approach (P = 3.79 x 10^{-7}, *ß* = 0.69), tests of heterogeneity indicated that the IV principals had been violated and that the results were influenced by pleiotropic effects. No significant causative associations were observed when testing the relationship between PD symptoms.

Taken together, these results build upon existing knowledge of genetic associations of pain in PD, and introduce preliminary results to understand the genetic influences on depression in PD. There is no evidence for a causative association from this data, indicating these symptoms can be treated independently.

5.3 Limitations of Work

It is important to consider the limitations of this work and how they may affect the interpretation of these findings. One of the main weaknesses is the size of the samples used for these investigations, which has limited the power of the association studies to detect genetic loci contributing to phenotypic variation. This is the case in the first results chapter when considering the number of PD samples used; this sample size meant that genome-wide significance was not achieved at the HLA loci as was previously observed (79). Furthermore, within this sample the current top HLA association was poorly imputed and removed in quality control. Sample size is an important factor determining the ability of a GWAS to detect genome wide significant results; for example, it was observed that a sample of over 50,000 cases enabled substantially greater ability to observe genome-wide significant associations in an MDD study (138). Of the factors that limit the power of an association study, especially when detecting smaller effect sizes as is the case with HLA associations, the sample size is most straightforwardly addressed. Obtaining a larger and more current PD sample for this purpose would help to overcome this issue. One of the main consequences of this limitation was a lack of confidence in the independent class I association identified, so a larger sample would help determine if this would reach genome-wide significance. This is important for basing expansion of data collection into class I variants in PD samples.

Sample size was also a limitation with the imputation reference panels applied to this dataset, which impacted the quality of HLA imputation. Since this work was completed, it was observed that applying a much larger multi-ancestry reference panel can improve imputation quality significantly, and so this was able to be addressed in the subsequent work completed. Whilst the methods applied worked well for the purpose of identifying loci of interest to further explore via sequencing, the poorer quality reference panels introduced ambiguity in the alleles of interest when looking at imputation results alone. In future work, this multi-ancestry panel should be used on the larger PD samples available to improve the power of this work.

The PacBio long-read sequencing method was applied to samples that were homozygous for the top Nalls 2019 HLA SNP risk allele or protective allele. Whilst this allowed a focused exploration of potential polymorphisms associated with this top associations, it was a limiting factor in the scope of the sequencing work as it prevented observation of other polymorphisms not associated with this SNP. This could be improved upon by sequencing a larger cohort of PD samples that also carried alleles associated with independent SNPs. However, this would firstly require a greater powered investigation into independent associations to ascertain which SNPs are also worth investigating in a similar manner. Furthermore, sequencing focused solely on the HLA-B, C, DQA1, DRB1 loci, which disregarded the multiple further genes within the HLA loci such as DPB1. As imputation is an imperfect association approach, it is possible other loci have been overlooked at this stage, and so data was missed when taking these selected loci forwards for sequencing analysis. A more thorough investigation would cover all the primary class I and class II loci, which would allow for comparison between DPB1 associations observed in imputation and sequencing results.

The main limitations of the final results chapter on non-motor symptoms also include sample size issues, which potentially contributed towards no observation of genome-wide significant associations in the various GWAS conducted. Genome-wide significant results had previously been observed with small PD sample sizes, such as the pain in PD GWAS which used 1,318 samples and detected a genetic association at the TRPM8 locus. However, the smaller sample sizes clearly impacted the current study's ability to detect genetic associations. It was calculated that a sample size increase from 1,453 to 1,800 in the UKBB PD depression GWAS could have increased the power to detect a genome wide significant association from 0.65 to 0.91. This power limitation could also be addressed by conducting a meta-analysis of pain and depression in PD combining the two samples used. This was not attempted here due to the different phenotype approaches used, but future work could prioritise using more similar phenotypes instead to create more power for these GWAS.

This issue of power also had a knock-on effect on the Mendelian randomisation approach, as no instrumental variables from these samples had genome-wide significance, and in some cases there were only a small number of independent IVs to be taken to the MR test. This can have a significant impact on the ability of Mendelian randomisation to detect causative associations, as the IVs used were potentially insufficient to represent phenotypic variance. Increasing the power of this study could be key to further understanding this causative association by providing superior IVs to use in a larger MR study. In addition, a two sample Mendelian randomisation approach was used instead of a one sample approach, which increased the bias in favour of type II error. Combined with the low power, this could explain the lack of significant causative associations observed.

5.4 Strengths of Work

Despite these limitations described, there were multiple strengths to the approaches taken in this work. In the first results chapter, whilst sample sizes were small for some of the imputation panels applied, a broader range of imputation techniques than usual were applied to allow for comparison between results that used different reference panels and computational approaches. In particular, the novel DISH method which uses direct imputation from summary statistics had not previously been applied to PD association analysis, so was an important comparison to more established methods. This novel data allowed corroboration of existing results, as the DISH results were in partial agreement with the HIBAG results in that the correlated alleles HLA-DQA1*03 and HLA-DRB1*04 were most associated with PD. Comparison with SNP2HLA also showed additional potential associations, indicating the HLA-DQA1*13:01 allele could be associated with PD risk. Thorough investigation of associated loci within existing QTL databases was an improvement upon existing published results, which omitted this or relied upon a single source with potentially misleading results. Overall, this chapter also highlighted the limitations of relying on a single bioinformatics approach when it comes to the HLA locus, as differing techniques gave varying results.

The main strength of the second results chapter was application of a long-read sequencing method to the HLA locus in PD. No long-read sequencing of the HLA locus has previously been applied to this genetic risk factor in PD, yet it is highly preferable over short-read methods due to its ability to obtain unambiguous data of the long, complex, and highly repetitive HLA region. Whilst no structural variants or repeat regions were identified, this method allowed identification of one new allele identified in PD samples. Also, the use of the novel large imputation panel allowed results to be compared between the most up-to-date imputation approach and the long-read sequencing data. Application of these gold-standard methods for HLA sequencing and imputation allowed discrepancies to be highlighted between current results and previously published results, and provide a novel insight into the HLA genetic influence on PD.

The various approaches applied in the final results chapter permitted a thorough investigation into genetic factors influencing PD non-motor symptoms across different cohorts. A GWAS of depression in PD has not previously been carried out, so this provided novel insight into this PD symptom which can be built upon further in future work. The MR methods used are well-established computational approaches to assess causation whilst avoiding bias from confounding. Despite the small numbers of IVs, weak instrument bias was avoided, and the replication of previously results indicates suitable methods were applied.

5.5 Implications

The results from this work have several implications with regards to assessment of the genetic factors influencing pain, depression, and the immune system in PD. Firstly, the results from the HLA association studies have shown some deviation from previous published results as to which variants are driving the primary HLA genetic association. The one previous sequencing study and latest imputation results published are in agreement that the DRB1*04 allele is the top association, driven by specific amino acids 11V, 13H, and 33H. Any association at DQA1*03 was not significant after adjustment for these amino acids. However, the present results from the more up-to-date and powerful imputation approach suggest the DQA1*03 allele is the leading association, with amino acids unique to DQA1*03 driving this association. When conditioning on this result, DRB1*04 did not remain significantly associated. The DQA1 and DRB1 loci are in strong linkage disequilibrium, which can make identifying the causative allele difficult. This has also had an impact on identifying the causative DQA1/DRB1 variants in other disorders such as type I diabetes (139,140). Whilst both molecules act to activate CD4+ T cells through antigen presentation, they differ in the heterodimers that they form and the molecules that they can present, and therefore their effect on immune responses. Current data is unclear what these differences could mean for molecular pathways in PD. Analysis of HLA alleles that present α-syn peptides has so far indicated that DRB1*15:01 and DRB5*01:01 alleles can present one epitope with high affinity, whilst DQB1*03:01, DQB1*05:01, and

DQB1*04:02 bound to three different epitopes with varying affinity (66). This indicates that there is currently no evidence for DRB1*04:01 having affinity to bind α-syn, but there is for DQB1 alleles that can form heterodimers with DQA1*03:01. This may not be the most significant HLA-related pathway influencing PD immune dysfunction; however, it is worth investigating further whether the DQA1*03 allele influences this process.

This has implications on the development of new therapeutics targeting inflammatory processes in PD. Due to their potential as early intervention therapies, and ability to repurpose currently available drugs, these are promising targets for new therapeutic trials. However, a greater understanding of the immune processes involved will help in more appropriately selecting targets. Recent examples include a trial of MCC950, an NLRP3 inflammasome inhibitor, following evidence that this protects against neuroinflammation in a PD mouse model (141). NLRP3 activation has been found to activate upon a-syn binding to TLR2 and TLR5 (142). Whilst these do not interact with HLA molecules in the same way TCRs do, HLA-DR molecules have been shown to interact with and enhance TLR activity (143). Understanding if the HLA-DR risk variants identified here modulate this process can help identify which patients would best benefit from this therapy, if it is proven to be useful. Other therapeutic interventions targeting the immune system in PD include Sargramostim, a granulocyte macrophage colony stimulating factor (GM-CSF). This works to increase the number of Tregs, demonstrated to be reduced in PD, which can then counteract the proinflammatory T cell responses. Initial results published recently have been positive for the potential of this therapy to regulate the PD immune response (144) As HLA allele type can determine T cell repertoire, a greater understanding of whether class II risk alleles influence Treg levels can also aid in appropriate targeting of this therapy.

This work also has implications for the treatment of other non-motor symptoms. As there is no evidence for a similar causative effect in PD symptoms as there is in the general population currently, these should be treated as associated symptoms but not causative. This observation could change with a better powered study, but this implies that treatment could be approached separately. The results from the MCP in PD GWAS indicate that as well as TRPM8, NRG1 signalling could be targeted. Both have been demonstrated to be involved in inflammatory pain; therefore, the link with inflammation and pain could be more relevant for this symptom and anti-inflammatory therapeutics could have a greater effect. More data is needed before exploring this possibility. KDM5A has a more varied impact; however, as it is suggested to be involved in sleep-dysregulation, targeting this pathway could also benefit this additional non-motor symptom. ZNF33B also implicates a range of zinc-finger related processes, that could be more appropriate to target in the case of PD neuropsychiatric symptoms. An important conclusion is the need for greater statistical power to identify which, if any, of these targets is worthy of future investigation.

5.6 Future Directions and Further Work

Whilst this work made progress in answering some of the questions outlined in the aims, more work can be conducted to improve upon and further the conclusions made here. It was not within the scope of this research project to determine how these HLA loci or risk variants for other symptoms function in the process of PD, but further work could help elucidate this.

With more time and funding, this would include the following:

1. Increase the sample size of the PD cohort to use for the HLA imputation method using the large multi-ancestry panel, so this latest method can be analysed with greater power.
2. Conduct long-read sequencing on a larger cohort of PD patients and across more HLA loci (e.g. including HLA-DPB1). Conduct sequencing in PD cohorts that are not selected based on rs112485576 allele carried, but a wider selection that could include independent associations.
3. Conduct HLA imputation and sequencing across populations of different ancestry. For example, South Indian PD patients have demonstrated certain different PD HLA allele associations within the class II HLA alleles (145). This can be built upon to improve the understanding of risk across more diverse populations.
4. To explore the potential interaction of the DRB1 PD protective allele with TLR and NLRP3 activity, test co-expression of differing HLA-DR molecules with TLR3/TLR4 in vitro to observe the effect on NLRP3 activation.
5. To explore the potential function of the DQA1 risk and protective alleles, use flow cytometry to test T cell repertoire differentiation in response to various antigens in samples with differing HLA-DQ alleles. This will help identify if, for example, Tregs are

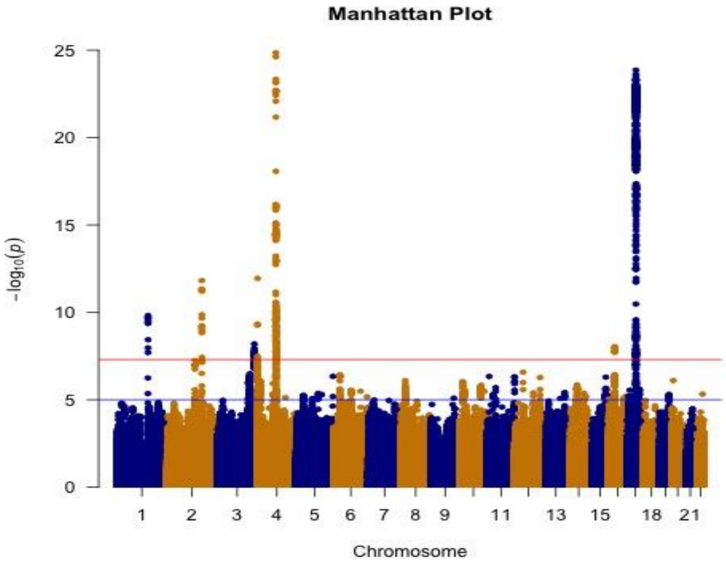
more affected by activity of specific DQ molecules in PD patients. Being able to correlate specific HLA DQ alleles with T cell subset activity will help focus the targets for PD immune therapies.

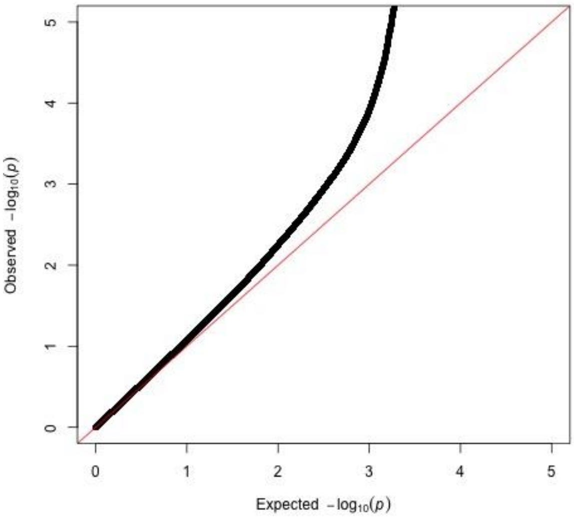
6. Repeat the GWAS for pain and depression in PD with more samples and conduct a meta-analysis using a combination of Proband and UKBB samples. This increase in power can lead to improved ability to identify genetic risk factors for these symtpoms.

5.7 Concluding Remarks

Overall, the work in this thesis made contributions to furthering our understanding of the genetic underpinnings of these non-motor and biological characteristics of PD. They highlight the need for further work to be conducted on these issues, and where this could be applied. Given the estimated growth in the burden of PD on the global population, and the current lack of preventative measures and adequate therapies, focusing on these elements will be beneficial to address some of the worst aspects of living with this disease.

Appendix





Appendix 1: *Manhattan plot and QQ plot for PD GWAS. Top associated SNP at the HLA locus is rs9268926 (P = 3.67 x 10^{-7}, OR = 0.84).*

References

1. GBD 2016 Parkinson's Disease Collaborators ER, Elbaz A, Nichols E, Abd-Allah F, Abdelalim A, Adsuar JC, et al. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2018 Nov 1;17(11):939–53.

2. Dorsey ER, Bloem BR. The Parkinson Pandemic—A Call to Action. JAMA Neurol. 2018 Jan 1;75(1):9–10.

3. Reeve A, Simcox E, Turnbull D. Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? Ageing Res Rev [Internet]. 2014 Mar 1 [cited 2019 Jul 23];14:19–30. Available from: https://www.sciencedirect.com/science/article/pii/S1568163714000051

4. Politis M, Wu K, Molloy S, G. Bain P, Chaudhuri KR, Piccini P. Parkinson's disease symptoms: The patient's perspective. Movement Disorders [Internet]. 2010 Aug 15 [cited 2019 Apr 4];25(11):1646–51. Available from: http://doi.wiley.com/10.1002/mds.23135

5. Braak H, Del Tredici K, Rüb U, De Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging. 2003 Mar 1;24(2):197–211.

6. Emamzadeh FN, Surguchov A. Parkinson's disease: Biomarkers, treatment, and risk factors. Vol. 12, Frontiers in Neuroscience. Frontiers Media S.A.; 2018. p. 612.

7. Jacobs BM, Belete D, Bestwick J, Blauwendraat C, Bandres-Ciga S, Heilbron K, et al. Parkinson's disease determinants, prediction and gene-environment interactions in the UK Biobank. J Neurol Neurosurg Psychiatry. 2020 Oct 1;91(10):1046–54.

8. Silverdale MA, Kobylecki C, Kass-Iliyya L, Martinez-Martin P, Lawton M, Cotterill S, et al. A detailed clinical study of pain in 1957 participants with early/moderate Parkinson's disease. Parkinsonism Relat Disord [Internet]. 2018 Nov [cited 2019 Apr 4];56:27–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29903584

9. Beiske AG, Loge JH, Rønningen A, Svensson E. Pain in Parkinson's disease: Prevalence and characteristics. Pain. 2009 Jan 1;141(1–2):173–7.

10. Nègre-Pagès L, Regragui W, Bouhassira D, Grandjean H, Rascol O. Chronic pain in Parkinson's disease: The cross-sectional French DoPaMiP survey. Movement Disorders [Internet]. 2008 Jul 30 [cited 2019 Apr 12];23(10):1361–9. Available from: http://doi.wiley.com/10.1002/mds.22142

11. Taylor AMW, Becker S, Schweinhardt P, Cahill C. Mesolimbic dopamine signaling in acute and chronic pain: implications for motivation, analgesia, and addiction. Pain. 2016 Jun 1;157(6):1194.

12. Florin E, Koschmieder KC, Schnitzler A, Becker S. Recovery of Impaired Endogenous Pain Modulation by Dopaminergic Medication in Parkinson's Disease. Movement Disorders. 2020 Dec 1;35(12):2338–43.

13. Williams NM, Hubbard L, Sandor C, Webber C, Hendry H, Lawton M, et al. Genome-Wide Association Study of Pain in Parkinson's Disease Implicates TRPM8 as a Risk Factor [Internet]. Movement Disorders. John Wiley and Sons Inc.; 2020 [cited 2020 Mar 5]. p. mds.28001. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.28001

14. Muller C, Morales P, Reggio PH. Cannabinoid Ligands Targeting TRP Channels. Front Mol Neurosci. 2019 Jan 15;11:487.

15. Gómez-Gálvez Y, Palomo-Garo C, Fernández-Ruiz J, García C. Potential of the cannabinoid CB2 receptor as a pharmacological target against inflammation in Parkinson's disease. Prog Neuropsychopharmacol Biol Psychiatry. 2016 Jan 4;64:200–8.

16. Weyer AD, Lehto SG. Development of TRPM8 Antagonists to Treat Chronic Pain and Migraine. Pharmaceuticals (Basel). 2017 Mar 30;10(2).

17. Sałat K, Filipek B. Antinociceptive activity of transient receptor potential channel TRPV1, TRPA1, and TRPM8 antagonists in neurogenic and neuropathic pain models in mice. Journal of Zhejiang University-SCIENCE B. 2015 Mar 11;16(3):167–78.

18. Lashinger ESR, Steiginga MS, Hieble JP, Leon LA, Gardner SD, Nagilla R, et al. AMTB, a TRPM8 channel blocker: evidence in rats for activity in overactive bladder and painful bladder syndrome. American Journal of Physiology-Renal Physiology [Internet]. 2008 Sep [cited 2019 Jul 25];295(3):F803–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18562636

19. Le Foll B, French L. Transcriptomic Characterization of the Human Habenula Highlights Drug Metabolism and the Neuroimmune System. Front Neurosci. 2018 Oct 31;12.

20. Chen Y, Wang Z, Jiang C, Y-h Z, Y-h J, Z-y T, et al. Oxaliplatin Depolarizes the IB4 − Dorsal Root Ganglion Neurons to Drive the Development of Neuropathic Pain Through TRPM8 in Mice. 2021; Available from: www.frontiersin.org

21. Gerdelat-Mas A, Simonetta-Moreau M, Thalamas C, Ory-Magne F, Slaoui T, Rascol O, et al. Levodopa raises objective pain threshold in Parkinson's disease: a RIII reflex study. J Neurol Neurosurg Psychiatry. 2007 Oct;78(10):1140.

22. De Michele G, Rizos A, Chan P, B-f Z, Z-q G, S-s M, et al. Musculoskeletal Pain in Parkinson's Disease. Frontiers in Neurology | www.frontiersin.org. 2022;1.

23. Kassubek J, Chaudhuri KR, Zesiewicz T, Surmann E, Boroojerdi B, Moran K, et al. Rotigotine transdermal system and evaluation of pain in patients with Parkinson's disease: a post hoc analysis of the RECOVER study. BMC Neurol. 2014 Mar 6;14(1).

24. Brunner R GA. Managing Pain in Parkinson's Disease. Pract Pain Manag. 2021;21(1).

25. Freo U, Furnari M, Ori C. Effects of tapentadol on pain, motor symptoms and cognitive functions in Parkinson's disease. J Pain Res. 2018;11–1849.

26. Trenkwalder C, Chaudhuri KR, Martinez-Martin P, Rascol O, Ehret R, Vališ M, et al. Prolonged-release oxycodone–naloxone for treatment of severe pain in patients with Parkinson's disease (PANDA): a double-blind, randomised, placebo-controlled trial. Lancet Neurol. 2015 Dec 1;14(12):1161–70.

27. Edinoff A, Sathivadivel N, Mcbride T, Parker A, Okeagu C, Kaye AD, et al. Chronic Pain Treatment Strategies in Parkinson's Disease.

28. Timmer MHM, Beek MHCT van, Bloem BR, Esselink RAJ. What a neurologist should know about depression in Parkinson's disease. Pract Neurol. 2017 Oct 1;17(5):359–68.

29. Ehgoetz Martens KA, Lewis SJG. Pathology of behavior in PD: What is known and what is not? J Neurol Sci. 2017 Mar 15;374:9–16.

30. Richard IH, Justus AW, Kurlan R. Relationship between mood and motor fluctuations in parkinson’s disease. Journal of Neuropsychiatry and Clinical Neurosciences. 2001;13(1):35–41.

31. Vriend C, Pattij T, Van Der Werf YD, Voorn P, Booij J, Rutten S, et al. Depression and impulse control disorders in Parkinson’s disease: Two sides of the same coin? Vol. 38, Neuroscience and Biobehavioral Reviews. Elsevier Ltd; 2014. p. 60–71.

32. Kano O, Ikeda K, Cridebring D, Takazawa T, Yoshii Y, Iwasaki Y. Neurobiology of depression and anxiety in Parkinson’s disease. Parkinsons Dis. 2011;2011:143547.

33. Ehrt U, Larsen JP, Aarsland D. Pain and Its Relationship to Depression in Parkinson Disease. The American Journal of Geriatric Psychiatry. 2009 Apr 1;17(4):269–75.

34. Hannestad J, DellaGioia N, Bloch M. The effect of antidepressant medication treatment on serum levels of inflammatory cytokines: a meta-analysis. Neuropsychopharmacology. 2011 Nov;36(12):2452–9.

35. Lian T hong, Guo P, Zhang Y nan, Li J hui, Li L xia, Ding D yu, et al. Parkinson’s Disease With Depression: The Correlations Between Neuroinflammatory Factors and Neurotransmitters in Cerebrospinal Fluid. Front Aging Neurosci. 2020 Oct 23;0:298.

36. Miyajima M, Zhang B, Sugiura Y, Sonomura K, Guerrini MM, Tsutsui Y, et al. Metabolic shift induced by systemic activation of T cells in PD-1-deficient mice perturbs brain monoamines and emotional behavior. Nat Immunol. 2017 Dec 23;18(12):1342–52.

37. Barrero FJ, Ampuero I, Morales B, Vives F, de Dios Luna del Castillo J, Hoenicka J, et al. Depression in Parkinson’s disease is related to a genetic polymorphism of the cannabinoid receptor gene (CNR1). The Pharmacogenomics Journal 2005 5:2. 2005 Jan 25;5(2):135–41.

38. Markopoulou K, Chase BA, Premkumar AP, Schoneburg B, Kartha N, Wei J, et al. Variable Effects of PD-Risk Associated SNPs and Variants in Parkinsonism-Associated Genes on Disease Phenotype in a Community-Based Cohort. Front Neurol. 2021 Apr 14;12:529.

39. Pankratz N, Marder KS, Halter CA, Rudolph A, Shults CW, Nichols WC, et al. Clinical correlates of depressive symptoms in familial Parkinson's disease. Movement Disorders [Internet]. 2008 Nov 15 [cited 2022 Jul 27];23(15):2216–23. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mds.22285

40. Olgiati S, Thomas A, Quadri M, Breedveld GJ, Graafland J, Eussen H, et al. Early-onset parkinsonism caused by alpha-synuclein gene triplication: Clinical and genetic findings in a novel family. Parkinsonism Relat Disord. 2015;21(8).

41. Piredda R, Desmarais P, Masellis M, Gasca-Salas C. Cognitive and psychiatric symptoms in genetically determined Parkinson's disease: a systematic review. Vol. 27, European Journal of Neurology. 2020.

42. Richard IH, McDermott MP, Kurlan R, Lyness JM, Como PG, Pearson N, et al. A randomized, double-blind, placebo-controlled trial of antidepressants in Parkinson disease. Neurology. 2012 Apr 17;78(16):1229–36.

43. Menza M, Dobkin RD, Marin H, Mark MH, Gara M, Buyske S, et al. A controlled trial of antidepressants in patients with Parkinson disease and depression. 2009.

44. Weintraub D, Aarsland D, Chaudhuri KR, Dobkin RD, Leentjens AF, Rodriguez-Violante M, et al. The neuropsychiatry of Parkinson's disease: advances and challenges. Lancet Neurol. 2022 Jan 1;21(1):89–102.

45. Mills KA, Greene MC, Dezube R, Goodson C, Karmarkar T, Pontone GM. Efficacy and tolerability of antidepressants in Parkinson's disease: A systematic review and network meta-analysis. Int J Geriatr Psychiatry. 2018 Apr 1;33(4):642–51.

46. Barone P, Poewe W, Albrecht S, Debieuvre C, Massey D, Rascol O, et al. Pramipexole for the treatment of depressive symptoms in patients with Parkinson's disease: a randomised, double-blind, placebo-controlled trial. Lancet Neurol [Internet]. 2010 Jun 1 [cited 2021 Jul 15];9(6):573–80. Available from: http://www.thelancet.com/article/S147444221070106X/fulltext

47. Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. Principles of innate and adaptive immunity. 2001;

48. Williams TM. Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. Vol. 3, Journal of Molecular Diagnostics. Association of Molecular Pathology; 2001. p. 98–104.

49. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: Expression, interaction, diversity and disease. Vol. 54, Journal of Human Genetics. Springer Japan; 2009. p. 15–39.

50. La Gruta NL, Gras S, Daley SR, Thomas PG, Rossjohn J. Understanding the drivers of MHC restriction of T cell receptors. Vol. 18, Nature Reviews Immunology. Nature Publishing Group; 2018. p. 467–78.

51. Traherne JA. Human MHC architecture and evolution: Implications for disease association studies [Internet]. Vol. 35, International Journal of Immunogenetics. John Wiley & Sons, Ltd; 2008 [cited 2020 Jun 23]. p. 179–92. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1744-313X.2008.00765.x

52. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020;48.

53. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection | Molecular Biology and Evolution | Oxford Academic [Internet]. [cited 2020 Jul 7]. Available from: https://academic.oup.com/mbe/article/35/9/2145/5034935

54. Jin P, Wang E. Polymorphism in clinical immunology - From HLA typing to immunogenetic profiling. Vol. 1, Journal of Translational Medicine. BioMed Central; 2003. p. 8.

55. Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. Current Biology. 2005 Jun 7;15(11):1022–7.

56. Reale M, Iarlori C, Thomas A, Gambi D, Perfetti B, Di Nicola M, et al. Peripheral cytokines profile in Parkinson's disease. Brain Behav Immun. 2009 Jan;23(1):55–63.

57. Bas J, Calopa M, Mestre M, Molleví DG, Cutillas B, Ambrosio S, et al. Lymphocyte populations in Parkinson's disease and in rat models of parkinsonism. J Neuroimmunol. 2001 Feb 1;113(1):146–52.

58. Baba Y, Kuroiwa A, Uitti RJ, Wszolek ZK, Yamada T. Alterations of T-lymphocyte populations in Parkinson disease. Parkinsonism Relat Disord. 2005 Dec;11(8):493–8.

59. Saunders JAH, Estes KA, Kosloski LM, Allen HE, Dempsey KM, Torres-Russotto DR, et al. CD4+ regulatory and effector/memory T cell subsets profile motor dysfunction in Parkinson's disease. Journal of Neuroimmune Pharmacology. 2012 Dec;7(4):927–38.

60. McGeer PL, McGeer EG. Glial reactions in Parkinson's disease. Movement Disorders. 2008 Mar 15;23(4):474–83.

61. Cunningham C, Campion S, Lunnon K, Murray CL, Woods JFC, Deacon RMJ, et al. Systemic Inflammation Induces Acute Behavioral and Cognitive Changes and Accelerates Neurodegenerative Disease. Biol Psychiatry. 2009 Feb 15;65(4):304–12.

62. Kortekaas R, Leenders KL, Van Oostrom JCH, Vaalburg W, Bart J, Willemsen ATM, et al. Blood-brain barrier dysfunction in Parkinsonian midbrain in vivo. Ann Neurol. 2005 Feb;57(2):176–9.

63. Seo J, Park J, Kim K, Won J, Yeo HG, Jin YB, et al. Chronic Infiltration of T Lymphocytes into the Brain in a Non-human Primate Model of Parkinson's Disease. Neuroscience. 2020 Apr 1;431:73–85.

64. Brochard V, Combadière B, Prigent A, Laouar Y, Perrin A, Beray-Berthat V, et al. Infiltration of CD4+ lymphocytes into the brain contributes to neurodegeneration in a mouse model of Parkinson disease. Journal of Clinical Investigation. 2009 Jan 5;119(1):182–92.

65. Theodore S, Cao S, McLean PJ, Standaert DG. Targeted Overexpression of Human α-Synuclein Triggers Microglial Activation and an Adaptive Immune Response in a Mouse Model of Parkinson Disease. J Neuropathol Exp Neurol. 2008 Dec 1;67(12):1149–58.

66. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, et al. T cells from patients with Parkinson's disease recognize α-synuclein peptides. Nature. 2017 Jun 29;546(7660):656–61.

67. Campos-Acuña J, Elgueta D, Pacheco R. T-cell-driven inflammation as a mediator of the gut-brain axis involved in Parkinson's disease. Vol. 10, Frontiers in Immunology. Frontiers Media S.A.; 2019. p. 239.

68. Harms AS, Thome AD, Yan Z, Schonhoff AM, Williams GP, Li X, et al. Peripheral monocyte entry is required for alpha-Synuclein induced inflammation and Neurodegeneration in a model of Parkinson disease. Exp Neurol. 2018 Feb 1;300:179–87.

69. Martin HL, Santoro M, Mustafa S, Riedel G, Forrester J V., Teismann P. Evidence for a role of adaptive immune response in the disease pathogenesis of the MPTP mouse model of Parkinson’s disease. Glia [Internet]. 2016 Mar 1 [cited 2020 May 1];64(3):386–95. Available from: http://doi.wiley.com/10.1002/glia.22935

70. Harm AS, Cao S, Rowse AL, Thome AD, Li X, Mangieri LR, et al. MHCII is required for α-Synuclein-induced activation of microglia, CD4 T cell proliferation, and dopaminergic neurodegeneration. Journal of Neuroscience [Internet]. 2013 Jun 5 [cited 2020 May 1];33(23):9592–600. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23739956

71. Williams GP, Schonhoff AM, Jurkuvenaite A, Thome AD, Standaert DG, Harms AS. Targeting of the class II transactivator attenuates inflammation and neurodegeneration in an alpha-synuclein model of Parkinson’s disease. J Neuroinflammation [Internet]. 2018 Aug 30 [cited 2020 Jul 6];15(1):244. Available from: https://jneuroinflammation.biomedcentral.com/articles/10.1186/s12974-018-1286-2

72. Cebrián C, Zucca FA, Mauri P, Steinbeck JA, Studer L, Scherzer CR, et al. MHC-I expression renders catecholaminergic neurons susceptible to T-cell-mediated degeneration. Nat Commun. 2014 Apr 16;5:3633.

73. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. Nat Genet. 2009 Nov 15;41(12):1308–12.

74. Dissection of the genetics of Parkinson’s disease identifies an additional association 5’ of SNCA and multiple associated haplotypes at 17q21 | Human Molecular Genetics | Oxford Academic [Internet]. [cited 2020 Apr 19]. Available from: https://academic.oup.com/hmg/article/20/2/345/654559

75. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet. 2010 Aug 15;42(9):781.

76. Hill-Burns EM, Wissemann WT, Hamza TH, Factor SA, Zabetian CP, Payami H. Identification of a novel Parkinson's disease locus via stratified genome-wide association study. BMC Genomics. 2014 Feb 10;15(1):118.

77. Kannarkat GT, Cook DA, Lee JK, Chang J, Chung J, Sandy E, et al. Common genetic variant association with altered HLA expression, synergy with pyrethroid exposure, and risk for Parkinson's disease: An observational and case-control study. Parkinsons Dis. 2015;1.

78. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet. 2009 Nov 6;124(6):593–605.

79. Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, Saad M, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: A meta-analysis of genome-wide association studies. The Lancet. 2011;377(9766):641–9.

80. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019 Dec 1;18(12):1091–102.

81. Wissemann WT, Hill-Burns EM, Zabetian CP, Factor SA, Patsopoulos N, Hoglund B, et al. Association of parkinson disease with structural and regulatory variants in the hla region. Am J Hum Genet. 2013 Nov 7;93(5):984–93.

82. Ahmed I, Tamouza R, Delord M, Krishnamoorthy R, Tzourio C, Mulot C, et al. Association between Parkinson's disease and the HLA-DRB1 locus. Movement Disorders. 2012 Aug;27(9):1104–10.

83. Shigenari A, Ota M, Kulski JK, Ozaki Y, Suzuki S, Shigenari A, et al. HLA-DRB1,-DRB3,-DRB4 and-DRB5 genotyping at a super-high resolution level by long range PCR and

high-throughput sequencing. 2014 [cited 2020 Oct 15]; Available from: https://www.researchgate.net/publication/261180886

84. Pierce S, Coetzee GA. Parkinson's disease-associated genetic variation is linked to quantitative expression of inflammatory genes. 2017;

85. Coetzee SG, Pierce S, Brundin P, Brundin L, Hazelett DJ, Coetzee GA. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. Sci Rep. 2016 Jul 27;6(1):1–11.

86. Latourelle JC, Dumitriu A, Hadzi TC, Beach TG, Myers RH. Evaluation of Parkinson Disease Risk Variants as Expression-QTLs. Lewis P, editor. PLoS One. 2012 Oct 5;7(10):e46199.

87. Yu E, Ambati A, Andersen MS, Krohn L, Estiar MA, Saini P, et al. Fine mapping of the HLA locus in Parkinson's disease in Europeans. npj Parkinson's Disease 2021 7:1 [Internet]. 2021 Sep 21 [cited 2022 Mar 9];7(1):1–7. Available from: https://www.nature.com/articles/s41531-021-00231-5

88. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. Tang J, editor. PLoS One. 2013 Jun 6;8(6):e64683.

89. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG - HLA genotype imputation with attribute bagging. Pharmacogenomics Journal. 2014 May 28;14(2):192–200.

90. Lim J, Bae SC, Kim K. Understanding HLA associations from SNP summary association statistics. Sci Rep. 2019 Dec 1;9(1):1–5.

91. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv [Internet]. 2018 Oct 19 [cited 2020 Oct 15];18:447367. Available from: https://doi.org/10.1101/447367

92. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979). 2020 Sep 1;369(6509):1318–30.

93. Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Res [Internet]. 2020 Jan 8 [cited 2022 Jul 22];48(D1):D983–91. Available from: https://academic.oup.com/nar/article/48/D1/D983/5584691

94. Cerri S, Mus L, Blandini F. Parkinson's Disease in Women and Men: What's the Difference? Vol. 9, Journal of Parkinson's Disease. 2019.

95. Pirinen M, Donnelly P, Spencer CCA. Including known covariates can reduce power to detect genetic effects in case-control studies. Nat Genet. 2012;44(8).

96. James LM, Christova P, Lewis SM, Engdahl BE, Georgopoulos A, Georgopoulos AP. Protective Effect of Human Leukocyte Antigen (HLA) Allele DRB1*13:02 on Age-Related Brain Gray Matter Volume Reduction in Healthy Women. EBioMedicine. 2018 Mar 1;29:31–7.

97. James L, Georgopoulos A. The Human Leukocyte Antigen (HLA) DRB1*13:02 Allele Protects against Dementia in Continental Western Europe. J Neurol Neuromedicine. 2019;4(5).

98. Steele NZR, Carr JS, Bonham LW, Geier EG, Damotte V, Miller ZA, et al. Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: A case-control study. PLoS Med. 2017 Mar;14(3):e1002272.

99. Naito T, Satake W, Ogawa K, Suzuki K, Hirata J, Foo JN, et al. Trans-Ethnic Fine-Mapping of the Major Histocompatibility Complex Region Linked to Parkinson's Disease. Movement Disorders. 2021 Aug 1;36(8):1805–14.

100. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. Tang J, editor. PLoS One. 2013 Jun 6;8(6):e64683.

101. Pappas DJ, Lizee A, Paunic V, Beutner KR, Motyer A, Vukcevic D, et al. Significant variation between SNP-based HLA imputations in diverse populations: The last mile is the hardest. Pharmacogenomics J. 2018 May 22;18(3):367.

102. Luo Y, Kanai M, Choi W, Li X, Sakaue S, Yamamoto K, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-

mapping in HIV host response. Nature Genetics 2021 53:10 [Internet]. 2021 Oct 5 [cited 2021 Dec 2];53(10):1504–16. Available from: https://www.nature.com/articles/s41588-021-00935-7

103. Klasberg S, Surendranath V, Lange V, Schöfl G. Bioinformatics Strategies, Challenges, and Opportunities for Next Generation Sequencing-Based HLA Genotyping. Transfusion Medicine and Hemotherapy [Internet]. 2019 Oct 1 [cited 2020 Aug 20];46(5):312–25. Available from: https://www.karger.com/Article/FullText/502487

104. Suzuki S, Ranade S, Osaki K, Ito S, Shigenari A, Ohnuki Y, et al. Reference Grade Characterization of Polymorphisms in Full-Length HLA Class I and II Genes With Short-Read Sequencing on the ION PGM System and Long-Reads Generated by Single Molecule, Real-Time Sequencing on the PacBio Platform. Front Immunol. 2018 Oct 4;9(OCT):2294.

105. Turner TR, Hayhurst JD, Hayward DR, Bultitude WP, Barker DJ, Robinson J, et al. Single molecule real-time DNA sequencing of HLA genes at ultra-high resolution from 126 International HLA and Immunogenetics Workshop cell lines. HLA. 2018 Feb 1;91(2):88–101.

106. Nishizawa A, Kumada K, Tateno K, Wagata M, Saito S, Katsuoka F, et al. Analysis of HLA-G long-read genomic sequences in mother–offspring pairs with preeclampsia. Scientific Reports 2020 10:1. 2020 Nov 18;10(1):1–10.

107. Mayor NP, Hayhurst JD, Turner TR, Szydlo RM, Shaw BE, Bultitude WP, et al. Recipients Receiving Better HLA-Matched Hematopoietic Cell Transplantation Grafts, Uncovered by a Novel HLA Typing Method, Have Superior Survival: A Retrospective Study. Biology of Blood and Marrow Transplantation. 2019 Mar 1;25(3):443–50.

108. Hollenbach JA, Norman PJ, Creary LE, Damotte V, Montero-Martin G, Caillier S, et al. A specific amino acid motif of HLA-DRB1 mediates risk and interacts with smoking history in Parkinson's disease. Proc Natl Acad Sci U S A. 2019 Apr 9;116(15):7419–24.

109. Malek N, Swallow DMA, Grosset KA, Lawton MA, Marrinan SL, Lehn AC, et al. Tracking Parkinson's: Study Design and Baseline Patient Data. J Parkinsons Dis. 2015 Nov 21;5(4):947.

110. Abdel-Wahab N, Diab A, Yu RK, Futreal A, Criswell LA, Tayar JH, et al. Genetic determinants of immune-related adverse events in patients with melanoma receiving immune checkpoint inhibitors. Cancer Immunology, Immunotherapy. 2021 Jul 1;70(7):1939–49.

111. Bondinas GP, Moustakas AK, Papadopoulos GK. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. Immunogenetics [Internet]. 2007;59(7):539–53. Available from: https://doi.org/10.1007/s00251-007-0224-8

112. Sarri CA, Papadopoulos GE, Papa A, Tsakris A, Pervanidou D, Baka A, et al. Amino acid signatures in the HLA class II peptide-binding region associated with protection/susceptibility to the severe West Nile Virus disease. PLoS One. 2018 Oct 1;13(10).

113. Miyadera H, Ohashi J, Lernmark Å, Kitamura T, Tokunaga K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. Journal of Clinical Investigation. 2015 Jan 2;125(1):275–91.

114. Koeleman B, Lie BA, Undlien DE, Dudbridge F, Thorsby E, De Vries R, et al. Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. Genes Immun. 2004;5:381–8.

115. Kwok WW, Kovats S, Thurtle P, Nepom GT. HLA-DQ allelic polymorphisms constrain. 1993;

116. Arlehamn CSL, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, et al. Immune response in Parkinson's disease driven by HLA display of α-synuclein peptides. The Journal of Immunology. 2017;198(1 Supplement).

117. Galiano-Landeira J, Torra A, Vila M, Bové J. CD8 T cell nigral infiltration precedes synucleinopathy in early stages of Parkinson's disease. Brain. 2020 Dec 1;143(12):3717–33.

118. Wang BY, Ye YY, Qian C, Zhang HB, Mao HX, Yao LP, et al. Stress increases MHC-I expression in dopaminergic neurons and induces autoimmune activation in Parkinson's disease. Neural Regen Res. 2021 Dec 1;16(12):2521–7.

119. Aarsland D, Påhlhagen S, Ballard CG, Ehrt U, Svenningsson P. Depression in Parkinson disease—epidemiology, mechanisms and management. Nature Reviews Neurology 2011 8:1. 2011 Dec 26;8(1):35–47.

120. Reijnders JSAM, Ehrt U, Weber WEJ, Aarsland D, Leentjens AFG. A systematic review of prevalence studies of depression in Parkinson's disease. Movement Disorders. 2008 Jan 30;23(2):183–9.

121. Wunderlich AP, Klug R, Stuber G, Landwehrmeyer B, Weber F, Freund W. Caudate nucleus and insular activation during a pain suppression paradigm comparing thermal and electrical stimulation. Open Neuroimag J. 2011;5:1–8.

122. Sheng J, Liu S, Wang Y, Cui R, Zhang X. The Link between Depression and Chronic Pain: Neural Mechanisms in the Brain. 2017;

123. Inequalities in health care for people with depression and/or anxiety - The Health Foundation [Internet]. [cited 2022 Mar 29]. Available from: https://www.health.org.uk/publications/long-reads/inequalities-in-health-care-for-people-with-depression-and-anxiety

124. Al-Harbi KS. Treatment-resistant depression: therapeutic trends, challenges, and future directions. Patient Prefer Adherence. 2012;6:369.

125. Karp JF, Scott J, Houck P, III CFR, Kupfer DJ, Frank E. Pain Predicts Longer Time to Remission During Treatment of Recurrent Depression. J Clin Psychiatry. 2005 May 15;66(5):0–0.

126. Johnston KJA, Adams MJ, Nicholl BI, Ward J, Strawbridge RJ, Ferguson A, et al. Genome-wide association study of multisite chronic pain in UK biobank. PLoS Genet [Internet]. 2019 Jun 1 [cited 2020 Oct 1];15(6):e1008164. Available from: https://doi.org/10.1371/journal.pgen.1008164

127. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature Genetics 2018 50:5. 2018 Apr 26;50(5):668–81.

128. Rao S, Shi M, Han X, Lam MHB, Chien WT, Zhou K, et al. Genome-wide copy number variation-, validation- and screening study implicates a new copy number polymorphism associated with suicide attempts in major depressive disorder. Gene. 2020 Sep 10;755:144901.

129. Luo W, Lim PH, Wert SL, Gacek SA, Chen H, Redei EE. Hypothalamic Gene Expression and Postpartum Behavior in a Genetic Rat Model of Depression. Front Behav Neurosci [Internet]. 2020 Oct 22 [cited 2021 Jul 14];14:190. Available from: www.frontiersin.org

130. Wang G, Dai D, Chen X, Yuan L, Zhang A, Lu Y, et al. Upregulation of neuregulin-1 reverses signs of neuropathic pain in rats. Int J Clin Exp Pathol [Internet]. 2014 [cited 2021 Jul 10];7(9):5916. Available from: /pmc/articles/PMC4203206/

131. Zhou J, Bao Q, Liang S, Guo H, Meng X, Zhang G, et al. rs1344706 polymorphism of zinc finger protein 804a (ZNF804a) gene related to the integrity of white matter fiber bundle in schizophrenics. Exp Ther Med. 2021 Jul 1;22(1):1–8.

132. Kay DB, Tanner JJ, Bowers D. Sleep disturbances and depression severity in patients with Parkinson’s disease. Brain Behav. 2018 Jun 1;8(6).

133. Li Johnson J, Abecasis GR. Genetics and Population Analysis GAS Power Calculator: web-based power calculator for genetic association studies. [cited 2022 Aug 11]; Available from: https://doi.org/10.1101/164343

134. A G, S B. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? Int J Epidemiol. 2019 Jun 1;48(3):691–701.

135. Li JX. Pain and depression comorbidity: A preclinical perspective. Vol. 276, Behavioural Brain Research. 2015.

136. James LM, Dolan S, Leuthold AC, Engdahl BE, Georgopoulos A, Georgopoulos AP. The effects of human leukocyte antigen DRB1*13 and apolipoprotein E on age-related variability of synchronous neural interactions in healthy women. EBioMedicine. 2018 Sep 1;35:288–94.

137. Wan C, Xu Y, Cen B, Xia Y, Yao L, Zheng Y, et al. Neuregulin1-ErbB4 Signaling in Spinal Cord Participates in Electroacupuncture Analgesia in Inflammatory Pain. Front

Neurosci [Internet]. 2021 Jan 28 [cited 2022 Jul 14];15. Available from: /pmc/articles/PMC7875897/

138. Nishino J, Ochi H, Kochi Y, Tsunoda T, Matsui S. Sample size for successful genome-wide association study of major depressive disorder. Front Genet. 2018 Jun 1;9.

139. Ilonen J, Kiviniemi M, Lempainen J, Simell O, Toppari J, Veijola R, et al. Genetic susceptibility to type 1 diabetes in childhood – estimation of HLA class II associated disease risk and class II effect in various phases of islet autoimmunity. Pediatr Diabetes [Internet]. 2016 Jul 1 [cited 2022 Jul 15];17:8–16. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/pedi.12327

140. Zhao LP, Papadopoulos GK, Moustakas AK, Bondinas GP, Carlsson A, Larsson HE, et al. Nine residues in HLA-DQ molecules determine with susceptibility and resistance to type 1 diabetes among young children in Sweden. Scientific Reports | [Internet]. 123AD [cited 2022 Jul 15];11:8821. Available from: https://doi.org/10.1038/s41598-021-86229-8

141. Huang S, Chen Z, Fan B, Chen Y, Zhou L, Jiang B, et al. A selective NLRP3 inflammasome inhibitor attenuates behavioral deficits and neuroinflammation in a mouse model of Parkinson’s disease. J Neuroimmunol. 2021 May 15;354.

142. Scheiblich H, Bousset L, Schwartz S, Griep A, Latz E, Melki R, et al. Microglial NLRP3 Inflammasome Activation upon TLR2 and TLR5 Ligation by Distinct α-Synuclein Assemblies. The Journal of Immunology. 2021 Oct 15;207(8):2143–54.

143. Frei R, Steinle J, Birchler T, Loeliger S, Roduit C, Steinhoff D, et al. MHC Class II Molecules Enhance Toll-Like Receptor Mediated Innate Immune Responses. PLoS One [Internet]. 2010 Jan 20 [cited 2022 Jul 13];5(1):e8808. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008808

144. Olson KE, Namminga KL, Lu Y, Schwab AD, Thurston MJ, Abdelmoaty MM, et al. Safety, tolerability, and immune-biomarker profiling for year-long sargramostim treatment of Parkinson’s disease. EBioMedicine. 2021 May 1;67:103380.

145. Pandi S, Chinniah R, Sevak V, Ravi PM, Raju M, Vellaiappan NA, et al. Association of HLA–DRB1, DQA1 and DQB1 alleles and haplotype in Parkinson’s disease from South India. Neurosci Lett. 2021;765.