# HLA Class I and II Associations with Renal Function and Associated Conditions in an Aging Population

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2022

Marcus P. Lowe

Division of Population Health, Health Services Research & Primary Care

School of Health Sciences

# Contents

## List of figures



## List of tables



Word count: 45,488

# 1 Abstract


## 1.1 Introduction

Kidney dysfunction is a highly significant disease, both in the United Kingdom and globally. Human leukocyte antigens (HLA) are thought to be associated with kidney function; this study aims to collate all previously reported associations between HLA and renal function, and to identify novel associations (and replicate previous findings) by analysing a cohort of subjects recruited in the UK. Secondary analysis investigated whether there is a link between Covid-19 and either HLA or renal function.

## 1.2 Methods

A systematic review was performed to identify all studies that had looked for associations between HLA and renal function. Additionally, regression analyses were performed to test for HLA associations with renal function in a cohort of around 500,000 UK Biobank subjects aged 39-73. Seven different ethnic groups and 362 HLA types were included. The primary outcome was estimated glomerular filtration rate (eGFR), with clinical outcomes used in secondary analyses. Finally, regression analysis was performed to test for associations between Covid-19, HLA, and renal function in cohorts of up to 6,000 Covid-19 positive subjects aged 50-86.

## 1.3 Results

35 papers investigating the link between HLA and renal function were identified, with over 100 HLA types and haplotypes reported to be associated with kidney function. These findings were collated and published.

This study revealed 33 HLA types that were linked to kidney function in white British subjects (11 hazardous and 22 beneficial). Studies of other ethnicities revealed nine significant associations, e.g., four alleles were linked to decreased function in Black African subjects.

The findings relating to Covid-19 were limited. Previously reported associations were replicated (e.g., increased age and male sex were linked to increased mortality), but there were no associations found between Covid-19 and either HLA or renal function.

## 1.4 Discussion

This thesis provides a list of all previously reported associations between HLA and renal function, and also presents a number of novel associations. Many of the associated alleles are commonly inherited together as haplotypes: *HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DQB1*02:01* has 9.5% frequency in England and each allele was associated

with decreased renal function in white British subjects. This may have clinical applications and could affect the UK's organ allocation algorithm.

# 2 Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

There are a number of papers included in this thesis. For the five that are included in the main text, I was the first author. I was responsible for the design of the studies, acquisition and analysis of the data, writing the copy, submitting to journals (in four out of five cases; the other was not submitted for peer-review), and making revisions based on reviewer feedback. I had help and advice from co-authors at every stage. Three of these papers were published in peer-reviewed journals, and two were not.

There are a further three papers included in the Appendix; in one case I was the first author (with the same responsibilities as above), and in two I was a co-author. As a co-author, I was responsible for advising on the statistical methods and interpretation of the results, as well as proof-reading the copy and providing feedback. These three papers were all published in peer-reviewed journals.

# 3 Copyright statement

# 4 Acknowledgements

I would like to thank my supervisors Kay Poulton, Tony Payton, Arpana Verma, Isla Gemmell, and Bill Ollier (plus de facto supervisor Judith Worthington) for all the guidance they have provided over the last six years. I wouldn't have known where to begin with this project but their experience and advice made the journey as smooth as it could possibly have been.

I am hugely grateful to Irene Chambers, Victoria Thomson, and all at Kidneys for Life for providing funding and tuition fees: the project would not have been possible without the generous backing of KFL and its benefactors.

I have had massive amounts of help along the way. Thanks to (to name a few): Ste Jervis, James Cashin, Debs Cook, Amy Bedford, Dan Eggleston, Marie Hampson, Patrick Hamilton, Muir Morton, Titus Augustine, David van Dellen, Videha Sharma, Lesley Patterson, Nisha Nair, and John Blaikley. Thanks also to everyone in the TLab (especially those in the techno bay) for their patience when explaining the most basic H&I concepts.

Thanks also to the personal tutors who provided advice and feedback throughout my studies (Christine Greenhalgh, Kenneth Muir, and Annemarie Money), to my first-year viva assessors John Bowes and Simon Rothwell, and to my final viva assessors Janine Lamb and Brendan Clarke.

On a personal note, I would like to thank my family (shout out HGB) and friends (shout out RVC), my cats Daisy, Pablo, and Bonnie, my mentor Pep Guardiola, and my best mate Phil Foden. Finally, thanks to my partner Lizzy for all her support, positivity, and roast dinners.

# 5 Preface

I graduated from the University of Sheffield in 2013 with a BA in Philosophy. Since 2015, I have worked as Data & Audit Manager for the Transplantation Laboratory at Manchester University NHS Foundation Trust. The trust performs around 300 kidney, pancreas, and islet transplants each year, and has data on over 7,500 transplants performed in Manchester dating back to the 1960s. Part of my role involves analysis of this data for audit purposes and to identify trends in transplantation activity and outcomes. When I started, I was able to perform some basic analysis, but further study and skills were needed to get the best out of this rich data source.

In 2017, I began studying part-time for a PhD at the University of Manchester. The plan was to acquire a large dataset and look for associations between HLA (immune-related genes which are known to be important in many aspects of nephrology) and renal function. The results of this work are presented in this thesis. Using data provided by UK Biobank, I published two peer-reviewed original articles which found HLA associations with renal function in a variety of ethnic groups. I also published a systematic review of previous research in this area. During the Covid-19 pandemic, I investigated the relationship between Covid-19, HLA, and renal function. Although this work was never published, I believe it is a useful contribution which replicated many observations that were already known or suspected about Covid-19.

To some extent, this research project is a “means to an end”. It has provided me with the ability to perform more sophisticated analysis of the clinical data available at MFT. This will be useful for future audits and research projects, for example to identify factors that affect transplant outcomes and to better understand the causes and pathways of renal failure. The applications of this, for patients and for healthcare providers, are clear.

However, I believe the work also stands alone as a valuable source of information about the relationship between HLA and renal function. The systematic review compiles all discoveries made in this field, allowing people to see the results of dozens of studies in a single place. The two original articles present novel findings which help to explain why certain people are susceptible to kidney dysfunction and others are not.

I intend to continue researching genetic associations with kidney function and transplantation in the future, building on the skills, findings, and connections that I acquired throughout my time working on this project.

# 6 Introduction

## 6.1 Kidney function

### 6.1.1 Measures of kidney function

The function of the human kidney is to filter toxins out of the blood and to process waste into urine[1]. A kidney, which is composed of nephrons, is functioning well when it has a high glomerular filtration rate (GFR). GFR is the rate at which each nephron is filtering toxins, multiplied by the number of nephrons present in the kidney[2]. Low GFR (and therefore poor kidney function) could be due to an overall lack of nephrons, or the nephrons not performing effectively, or a combination of these factors.

GFR can be measured or estimated in a number of ways. Most of these involve the measurement of biomarkers in the blood or urine of a subject; the presence (or prevalence, or concentration) of these biomarkers gives an indication of how well the nephrons are functioning[3]. Inulin is considered the "gold standard" biomarker when measuring GFR because its prevalence in the urine of a subject is affected only by the subject's GFR. That is, no factors other than GFR contribute to how much inulin is present in urine. For this reason, it is a very reliable way of measuring the GFR of a subject. However, inulin is exogenous to humans (that is, it does not occur naturally) so must be injected prior to measurement. This means the subject must be present for several hours while the inulin is injected, absorbed, filtered, and passed in urine. Also, inulin is not always readily available and can be expensive. These limitations mean that inulin is not appropriate to be the standard biomarker used to measure GFR in the clinical setting[2].

Creatinine is a practical alternative to inulin as it is easy to measure in a subject's serum, it is inexpensive, and it is endogenous to humans (that is, it occurs naturally). It is not a perfect measure of GFR as it is affected by muscle mass (which in turn is related to age, sex, race, diet, and lifestyle, among other factors). Also, creatinine is not exclusively filtered by the kidneys so its prevalence in serum is not a perfect representation of how well the nephrons are filtering[2].

Cystatin is a potential alternative to creatinine, especially for subjects whose GFR is in the normal range[4]. It is endogenous to humans, and it is thought not to be related to muscle mass[2]. However, it may be associated with age, gender, weight, and height. It is also not excreted in urine, meaning measurements must be carried out using serum[3]. Albumin is a protein that can be measured in urine to test kidney function. In clinical settings, the ratio of albumin to creatinine in urine is used to test for kidney disease[5].

Urea was used as a marker for kidney function even before the concept of GFR was established[2]. However, it is no longer used (either in serum or in urine) because both its generation and its excretion by the kidney are subject to non-renal factors which vary between individuals, such as age and dietary protein[6]. This means that measuring a subject's urea does not give a reliable indication of the function of the subject's kidney.

Healthy GFR levels vary between people. Generally, it is normal for young adults to have GFR between 120-130ml/minute/1.73m² of height, and this rate declines with age. Over a quarter of people over 70 years old have GFR under 60[2].

Formulae have been developed to produce an "estimated GFR" (eGFR), usually based on serum creatinine levels and other factors. These give a more accurate indication of kidney function than the biomarkers alone because they control for other factors which affect the biomarker's prevalence. For example, the Cockcroft-Gault equation controls for age, weight, and sex[7], while the Modification of Diet in Renal Disease (MDRD) study equation controls for age, sex, and ethnicity[8].

A limitation of these eGFR formulae is that they are often developed with reference to groups which do not adequately represent the general population. For example, Cockcroft-Gault was developed using a group of 249 males, so applying the formula to the general population may not give accurate results. MDRD was developed using a group of predominantly white subjects with chronic kidney disease (CKD), though it has since been validated in other groups such as African Americans and kidney transplant recipients[9]. The use of these formulae also requires that the creatinine measurement technique used must be the same as the technique used by the laboratory which developed the formula. According to Stevens[2], Cockcroft-Gault is used most often but MDRD gives a more accurate estimation of GFR. Cockcroft-Gault systematically overestimates GFR.

### 6.1.2 Effects of poor kidney function

There are many health risks associated with poor kidney function. Decreased kidney function is associated with increased mortality[10], increased hospitalisation[11], and increased risk of cardiac events such as myocardial infarction[12].

In extreme cases, poor kidney function requires renal replacement therapy (RRT), usually in the form of dialysis or transplantation. These often have a significant effect on the patient's quality of life due to pain, fatigue, inconvenience, and risk of surgical complications. At the end of 2019, over 68,000 adults in the UK were either transplant recipients or were on dialysis, and the rate of RRT (1,293 per million population) was increasing by 2.0% per year[13]. It remains to be seen how the Covid-19 pandemic will affect these figures, though a 2020 study

(to which I contributed) estimated that at least 700 kidney transplant opportunities would be lost in the UK, and that the dialysis population would increase accordingly[14].

CKD is a diagnosis which refers to structural or functional damage to the kidney. At its most severe, CKD is known as end-stage renal disease (ESRD), which requires RRT. There are five stages of CKD, which are based on the patient's GFR as well as the physical state of their kidneys (i.e., whether they are damaged or not). Table 1 summarises the definitions of the five stages.

Table 1: Stages of chronic kidney disease

| Stage | Description | GFR | Notes |
|---|---|---|---|
| 1 | Normal or increased GFR | >90 | Kidney damage also required |
| 2 | Mild decrease in GFR | 60-89 | |
| 3 | Moderate decrease in GFR | 45-59 | Stage 3a |
| | | 30-44 | Stage 3b |
| 4 | Severe decrease in GFR | 15-29 | |
| 5 | Kidney failure, usually requiring RRT | <15 | Also known as ESRD |

This table summarises the definitions of the five stages of chronic kidney disease. It was adapted from Assessing Kidney Function – Measured and Estimated Glomerular Filtration[3]. GFR – glomerular filtration rate; RRT – renal replacement therapy; ESRD – end-stage renal disease

Stage 1 refers to evidence of kidney damage but a GFR ≥90 (that is, normal or increased GFR). Stage 2 is defined as GFR between 60-89 (a mild decrease) with evidence of kidney damage, which can progress to stage 3 (GFR 30-59, a moderate decrease in GFR), stage 4 (GFR 15-29, a severe decrease), and ultimately stage 5 (GFR <15, which usually requires renal replacement). Stages 3, 4 and 5 can be diagnosed without evidence of kidney damage. Stage 5 is synonymous with ESRD. Stage 3 is often split into stage 3a (GFR 45-59) and stage 3b (GFR 30-44).

CKD affects over 1.8m people in England, and it is estimated that over 40,000 of these people die prematurely each year[15]. The cost of treating CKD in England in 2009/10 was around £1.45bn, or around 1.3% of all NHS expenditure. CKD is associated with decreased quality of life, increased risk of stroke and myocardial infarction, increased hospital stays, and increased risk of infection. In addition to increased healthcare spending, kidney failure may also be linked to increased social care and social welfare costs, as well as decreased productivity and tax revenue due to sufferers and (in some cases) their carers being unable to work[15].

A number of studies have linked renal dysfunction to adverse health outcomes. Go et al.[11] studied over 1.1m adults who had not undergone dialysis or kidney transplant. They found that as GFR decreased, risk of death increased. Adjusted hazard ratio for death was 1.2 times higher for subjects with eGFR between 45-59 (stage 3a CKD), 1.8 times higher for subjects

with eGFR 30-44 (stage 3b CKD), 3.2 times higher for subjects with eGFR 15-29 (stage 4 CKD), and 5.9 times higher for subjects with eGFR <15 (stage 5 CKD or ESRD). The hazard ratio for hospitalisation and cardiovascular events also increased as eGFR decreased. Tonelli[10] performed a systematic review of 16 studies (including Go) which had similar findings to Go's study.

Weiner[12] combined data from four different studies (totalling over 22,000 subjects) and found that CKD was an independent risk factor for cardiovascular disease and all-cause mortality. Outcome measures were myocardial infarction, fatal coronary heart disease, stroke, and death. CKD (defined as subjects having a GFR between 15-60) had a hazard ratio of 1.2 for a composite of these outcomes. For black subjects (who comprised 18% of the study group), the hazard ratio was 1.8, while it was only 1.1 in white subjects.

Fried et al.[16] studied over 4,000 subjects aged over 65 and found that increased cystatin was associated with increased mortality from non-cardiovascular causes, even after adjustment for several factors. They concluded that kidney function is a predictor of non-cardiovascular mortality in the elderly. Finally, Hemmelgarn[17] studied over 900,000 subjects in Canada. The majority had healthy levels of GFR. They found that worse kidney function (as measured by proteinuria or albumin-creatinine ratio) was linked with increased mortality and myocardial infarction.

## 6.2 History of genome-wide association studies

Genome-wide association studies (GWAS) analyse the genetic information of a group of people and identify single nucleotide polymorphisms (SNPs – small differences in DNA between individuals) which can be associated with traits like disease susceptibility or responsiveness to drugs[18]. Associating SNPs with particular diseases or outcomes is useful as it can help to diagnose symptoms, predict the development of diseases before they occur, identify novel drug development targets or help with effective treatment pathways.

Before GWAS became prevalent in medical research, inheritance studies of genetic linkage were used to identify genes associated with diseases. This was effective for single gene disorders but less so for complex diseases[19].

Since 2003, when the Human Genome Project successfully mapped human DNA at a cost of around $2.7bn[20], the price of determining a subject's DNA sequence has reduced dramatically. Genome sequencing can now be performed for under $1000 depending on resolution[21]. Affordable genome sequencing has helped make genetic association studies prevalent.

The first ever GWAS discovered a susceptibility gene for myocardial infarction[22]. This 2002 study of 3,011 Japanese subjects (1,133 with myocardial infarction, and two control groups

totalling 1,872 people) looked at 92,788 SNP markers and found five SNPs on chromosome 6p21 to be associated with higher risk of myocardial infarction. Since then, GWAS have advanced considerably in terms of sample size, genetic scope, and precision. They typically cover the entire genomes of hundreds of thousands of subjects (often encompassing 1m SNPs or more), and over 3,700 GWAS have been published, identifying over 89,000 SNP-trait associations[23].

Many genetic-focused companies, charities, academic institutions, and other organisations map genomes, which means data is more plentiful and accessible than before. The rise of biobanks (discussed in section 6.6) which share genetic and phenotype data means that individual research projects may not need to collect data themselves, which may have been prohibitively expensive. Instead, researchers can apply for access to existing samples and data from a third party which may reduce overall costs. Biobanks may also be publicly or charitably funded, which can reduce costs further. HapMap, for example, was a collaborative biobank project that looked at one million individuals from four populations[24]. It looked for differences between individuals and blocks of low haplotype diversity (that is, areas of the genome where people are generally similar). These can highlight other genomic areas which may be influential in determining traits and disease.

Findings from GWAS tend not to be useful on an individual diagnostic level as they are often relatively weak associations that have been discovered due to the high power of GWAS studies. The SNPs identified in these studies tend to have odds ratios of around 1.1-1.2[25]. This is not large enough that the presence or absence of a particular SNP can be used as a sole diagnostic tool. The SNP may have no detectable effect (or only a very small effect) on one individual but may be associated with a particular outcome at a population level.

The large sample sizes of GWAS also mean that it is often impossible for all subjects in the study to have all their genotype and phenotype data checked and verified, so some of it may be inaccurate. In case-control studies, this can lead to subjects being included in the control group when they should be in the study group, leading to less accurate findings. Gathering data from multiple sources can lead to study heterogeneity[26], with consequences such as inconsistent or inaccurate definitions and classifications (for example, of ethnic groups or of diseases).

GWAS do not generally isolate causation: the variant genetic marker may be associated with a disease but may not be the cause of it. They are also at risk of finding false positives[26] (though of course this problem is not unique to GWAS). This is because they often investigate a large number of SNPs and outcomes, so apparent associations can be found which are merely due to chance. This problem can be largely mitigated by using a low P value for

significance (though if the adjusted P value is too stringent, there is an increased risk of false negatives), increasing sample size or using an independent replication cohort.

Despite these limitations, GWAS have uncovered many significant, replicated findings[27]. For example, they have revealed thousands of genetic variants associated with diabetes[28], auto-immune diseases[29], educational attainment[30], depression[31], and many other conditions and outcomes.

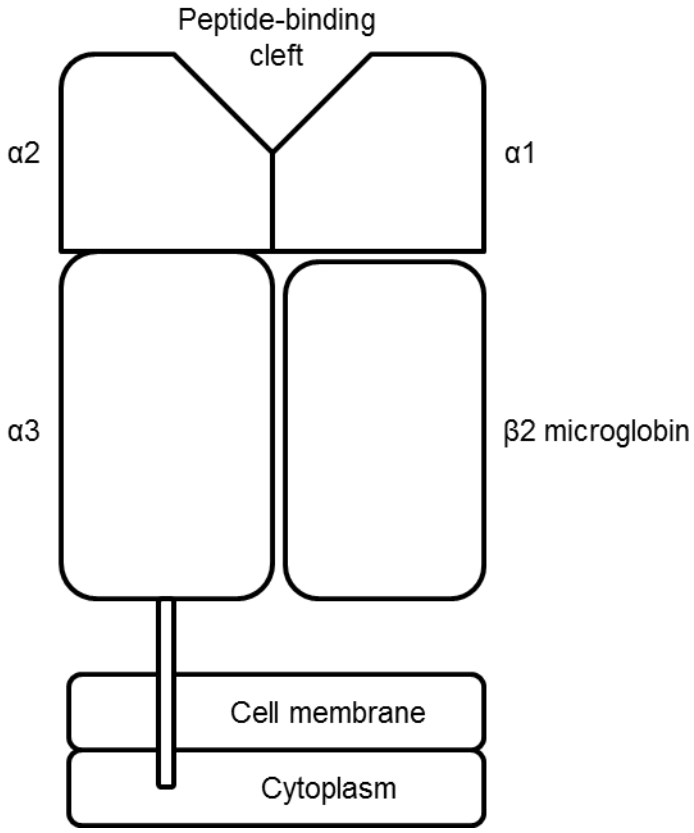
## 6.3 Major histocompatibility complex: organisation of HLA genes

The major histocompatibility complex (MHC) is a genetic region of chromosome 6p21 containing a series of immune-related genes which encode molecules or structures found on cell surfaces that are instrumental in the identification and destruction of foreign molecules or pathogens. In humans, these structures are known as human leukocyte antigens and form part of the adaptive immune system[32]. There are two primary types of MHC antigens: class I, and class II. There is also a class III region, but no class III antigens.

The HLA genes are encoded on chromosome 6 in the MHC genetic region. There are three classical class I loci (HLA-A, -B, and -C) and three class II loci (HLA-DR, -DQ, and -DP) which are involved in graft acceptance, and therefore have an impact on transplantation. These are highly polymorphic, meaning there are many different forms or “alleles” they can take[33]. The genes are inherited in a co-dominant fashion from parents, with every individual having two genes for each locus. As of March 2022, 24,308 class I alleles and 9,182 class II alleles have been identified[34].

Class I antigens have a four-part structure and are displayed on the surface of most nucleated cells[35] associated with β2 microglobulin, which is encoded outside the MHC. They facilitate the recognition of antigens or foreign bodies by leukocytes. To do this, during its synthesis, the class I protein binds a peptide from within the cell into a cleft within its structure. If the cell is healthy, the protein will be a “self-antigen”, but if the cell is infected with a pathogen, a pathogen-derived peptide could be displayed by the cleft. During immune surveillance, the leukocyte approaches the cell; if it recognises the protein as a self-antigen it takes no action. If, however, it recognises a pathogen, it will bind to the cell and destroy it. Figure 1 shows the structure of class I antigens.

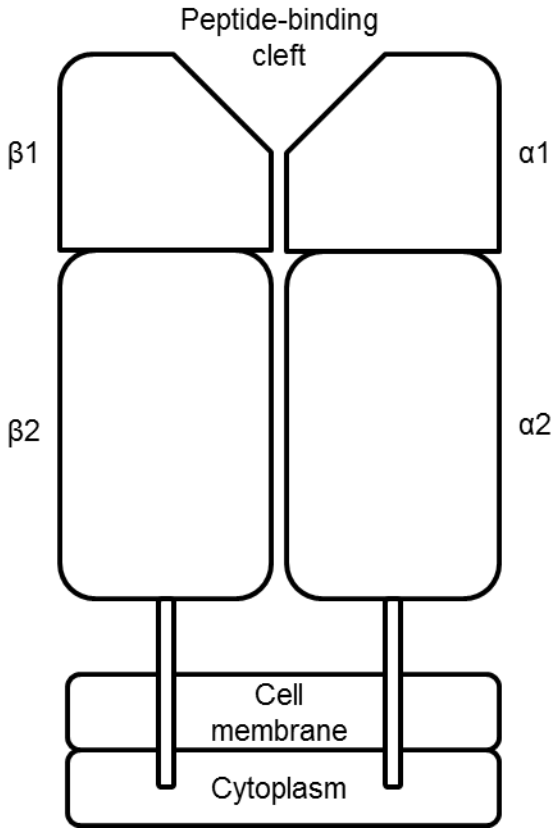
Figure 1: Structure of HLA class I antigens




This figure shows the structure of HLA class I antigens. It was adapted from Immunobiology: the Immune System in Health and Disease[36].

Class II antigens also have a four-part structure, and support leukocyte communication to bring about the elimination of pathogens. These antigens do not associate with the β2 microglobulin and all components are encoded within the MHC. Class II antigens are only present on specific immune cells known as “antigen-presenting cells”. Examples of these include monocytes, dendritic cells, and B lymphocytes. When a pathogen is encountered by one of these cells, it is engulfed and internalised by the cell, then broken down, and part of the foreign body (a peptide) is inserted into a cleft within the MHC class II protein. Other immune cells (T lymphocytes, helper T cells, and B cells) can then learn to recognise that particular pathogen so it can be destroyed more quickly in the future. They do this by producing antibodies which specifically target the pathogen[35]. Figure 2 shows the structure of class II antigens.

Figure 2: Structure of HLA class II antigens




This figure shows the structure of HLA class II antigens. It was adapted from Immunobiology: the Immune System in Health and Disease[36].

Macrophages are directed towards pathogens by cytokines and chemokines, which are released as part of the inflammatory response. When pathogenic bacteria invade the body, more peripheral blood (which contains macrophages) is recruited to the infection site. Macrophages engulf the pathogen, forming a phagosome with the pathogen inside it, and digesting it to break it down. This is known as opsonisation.

Dendritic cells are found close to the outside environment such as in the skin, the nose, the lungs, the stomach, and other places. They have spiky projections known as dendrites. They also engulf pathogens, break them down, and display these for immune recognition through the HLA class II proteins. Dendritic cells can move into the spleen and the lymph nodes, which are rich in B and T cells, amplifying the efficiency of the immune response. They are a bridge between innate and adaptive immunity.

B lymphocytes are instrumental in promoting and driving humoral immunity, and producing antibodies rather than using cell-mediated immunity observed in T cells. They identify pathogens, bind and engulf them, then display their constituent peptides in HLA class II

molecules. A CD4+ lymphocyte then causes the B lymphocyte to multiply. This creates B lymphocytes (memory cells) which are already capable of destroying the particular pathogen.

T lymphocytes are incapable of binding directly to pathogens, but can bind to peptides derived from the pathogen which are displayed by HLA class II molecules. When they bind, they activate helper T and B lymphocytes, and release co-stimulatory molecules which effect the destruction of the pathogen, without the need to engulf it. These lymphocytes are adaptive, meaning they must be produced in response to a certain stimulus, rather than innate, meaning that they already exist in the body. Innate immune cells are quick to generate and can destroy any pathogen, but they do so slowly. The adaptive immune response takes longer to produce, and is more efficient, but is highly specialised and only destroys one type of pathogen[35].

Each individual's combination of HLA alleles directs which peptides they are able to present and mount an immune response to[37]. The wide range of polymorphism is essential for humans as a species as it is unlikely that there could be a single pathogen which evades presentation. Such an evasion of the immune response by a pathogen could mean that the whole species could be wiped out by a single disease. However, this polymorphism also means that individuals vary in their susceptibility or response to certain diseases as a result of their combination of HLA genes.

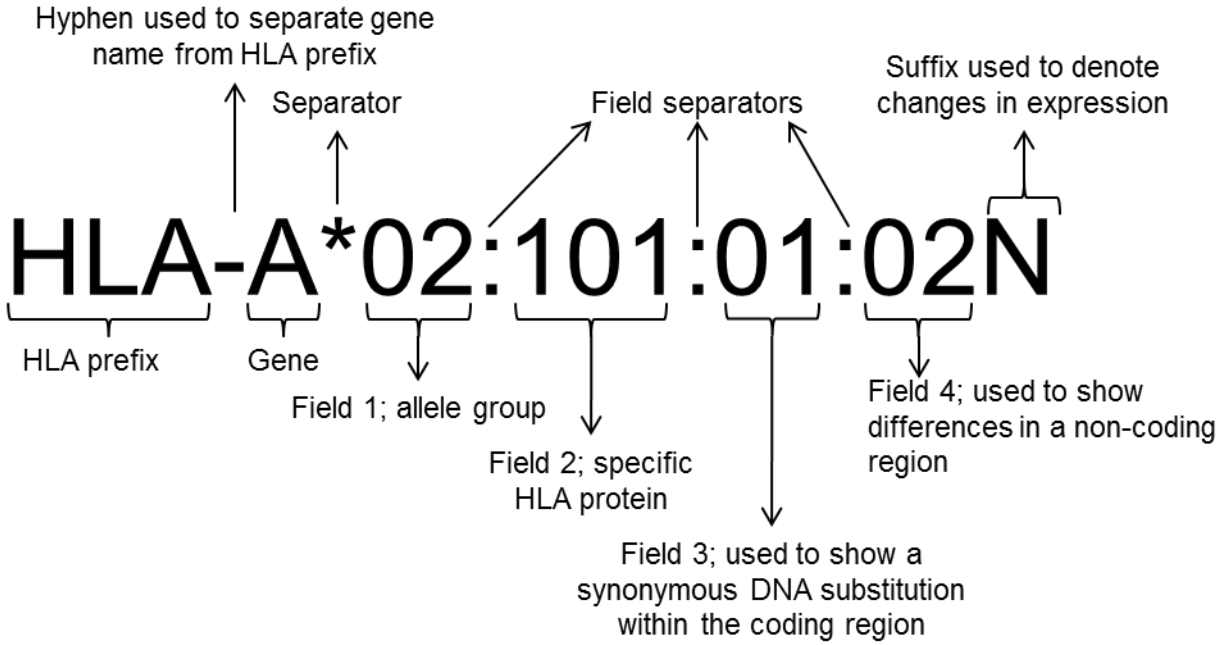
### 6.4 Traditional HLA typing techniques

The number of different HLA loci and the extensive polymorphism of each locus mean that there is a highly diverse range of HLA phenotypes in humans. Determining an individual's HLA type has clinical applications in terms of donor-recipient matching for transplantation, and discovering susceptibility to certain diseases[38,39].

Over the years, a number of different techniques have been developed for HLA typing. These have various benefits and drawbacks in terms of accuracy (the proportion of alleles which are identified correctly), precision (the resolution to which subjects can be typed), throughput efficiency (the number of samples which can be tested in a single run), financial cost (the cost of equipment and other consumables required, as well as the cost of employing staff to conduct the testing), speed (the time taken to complete the testing), and ease (the skill required to perform the test or analyse the results).

In terms of precision, there are a number of different levels of resolution possible when typing HLA[40]. "Allelic resolution" is the highest possible: this refers to a result which identifies an individual HLA allele. "High-resolution" typing can identify a set of specific HLA proteins. "Low-resolution" typing can identify only an allele group. Figure 3 shows the nomenclature of an HLA allele.

Figure 3: Nomenclature of HLA




This figure shows the nomenclature of an HLA allele. It was adapted from Prof. Steven Marsh[41].

### 6.4.1 Serological

The first HLA typing method was serological. Developed in the 1960s[42], this method involves combining a sample (usually T and B lymphocytes isolated from peripheral blood) from the subject with a number of sera each containing an antibody of known specificity. Some antibodies will bind to the HLA antigens present on the lymphocytes, and some will not, depending on the subject's HLA type. By observing which antibodies cause a reaction, the subject's HLA type can be deduced[43].

This method relies on having a high-quality viable sample, which is not always possible (for example, for trauma victims and those close to death). There is also a possibility of false negative results, as serological typing is unable to detect some antigens such as null or low-expressed antigens. Serological typing is somewhat subjective, as the fact of whether a reaction has occurred is sometimes ambiguous. It is therefore not used for typing deceased organ donors as the accuracy of typing is of paramount importance.

### 6.4.2 Polymerase chain reaction

As a result of the drawbacks mentioned in section 6.4.1, serology has been surpassed in accuracy and precision by molecular typing methods which were developed subsequently. Some of these DNA-based methods (but not all) also allow many patients to be tested

simultaneously, which increases efficiency, can type to a higher resolution, can be performed more quickly, and analysis of results can be more easily automated[44].

The development of polymerase chain reaction (PCR) in 1983[45] led to improvements in HLA typing. PCR is the process of amplifying a segment of DNA many times, giving thousands or millions of copies of a DNA sequence. In the case of HLA typing, the segments amplified are the genetic regions containing HLA class I and II loci. These can then be analysed to determine a subject's type. There are three main applications of PCR used in HLA typing: sequence specific primers (SSP), reverse sequence specific oligonucleotide probing (SSO), and sequence based typing (SBT).

### 6.4.3 Sequence specific primers

In PCR-SSP typing, a number of primers are used to identify which segments of DNA to amplify[46]. A primer is a short strand of DNA (that is, a sequence of genetic code) which is complementary to a particular sequence of DNA. Primers which allow amplification indicate that the subject has the allele as their phenotype, while primers do not amplify indicate that the allele is not present in the subject. When that particular strand or sequence is present in the subject's DNA, the primer will bind, which gives PCR its starting place, and the corresponding DNA sequence is amplified. When the sequence is not present, there is no primer binding and it is not amplified. Each primer matches the genetic code of a particular HLA allele or group of alleles. By assessing patterns of reactivity, the researcher or scientist can conclude which HLA alleles are present in the subject.

This method provides high accuracy when compared to serological typing. It is also comparatively fast, taking under three hours[44]. However, a single PCR-SSP test is most suited to single-sample testing, as the capacity of thermal cyclers limits the number of samples which can be included in a single run[47], even with the introduction of 384-well platforms. PCR-SSP also does not provide the highest possible resolution, because a different primer pair is required to test for each possible type and it is not practical to include a separate primer pair for each of the thousands of HLA specificities. PCR-SSP also cannot detect new or novel alleles because tests must be designed for a known DNA sequence.

### 6.4.4 Reverse sequence specific oligonucleotide probing

Reverse PCR-SSO typing also involves amplification using primers, which are specific for multiple HLA alleles at a particular locus. Specific DNA probes are covalently linked to a fluorescent microsphere (known as a "bead") and hybridised with the PCR-amplified DNA. The test DNA binds to the beads and probes, and a laser detects hybridised product which

reflects the subject's HLA type for that locus. By performing this process for all loci simultaneously, the subject's complete HLA type can be defined.

A major advantage of PCR-SSO is its efficiency. A single run can simultaneously test 100 or more subjects, compared to PCR-SSP which can only be carried out for a single subject at a time. It is also relatively inexpensive. However, PCR-SSO takes approximately one working day to run[48], so is less useful in emergency situations such as typing a deceased organ donor. PCR-SSO generally produces an HLA type to an intermediate resolution, though high resolution kits are available[49].

### 6.4.5 Sequence based typing

Another method of HLA typing using PCR is SBT, in which the nucleotide sequence of the encoding regions of the HLA molecule are defined, whether using Sanger sequencing, or by next generation sequencing methods (NGS). In Sanger sequencing, encoding regions (known as exons) for each HLA locus are amplified, and added to a reaction mixture with DNA polymerase (an enzyme), a supply of nucleotides, and a chain-terminating variant of a nucleotide. Using capillary gel electrophoresis in a semi-quantitative fluorescent analyser, the DNA sequence can be deduced. By comparing the DNA sequence obtained with a reference database of HLA allele sequences, the subject's type can be found.

The main advantage of SBT is its precision[47]. Because it involves analysing the exact nucleotide sequence, rather than analysing which primers enabled amplification, it can type subjects to a higher resolution. It is also relatively easy to automate the process, meaning it can be performed by less skilled technicians. However, PCR-SBT is unable to resolve some ambiguities as some sequences can be shared by multiple HLA alleles. When this occurs, more testing is required using specific primers. SBT is also able to identify novel or previously unidentified alleles, which other typing methods are incapable of achieving.

### 6.4.6 Next generation sequencing

In the 2000s, a new approach to HLA typing has been commercialised: NGS[50]. In NGS, a long PCR-amplified strand of DNA is fragmented into smaller sequences and denatured. The next step is to sequence the DNA using fluorescent nucleotides. When these nucleotides are added to a denatured strand of DNA, different fluorochromes are excited, emitting light depending on which nucleotide is present. By analysing these emission spectra, the DNA sequence can be deduced.

A key advantage of NGS is its throughput. It is possible to type 96 patients in a single run. It also types to a higher resolution than previous PCR-based methods[51], and with higher accuracy than SBT due to less ambiguity. However, typing by NGS takes a long time to

complete (generally at least two days), and requires skilled staff at a number of points in the process. Designated instruments are also required to process the tests, which, together with high-cost consumables, makes NGS expensive compared to other typing techniques.

## 6.5 HLA typing by imputation: history, uses, benefits, how to

Imputation is the process of estimating missing data. It is more reliable and efficient when more context is available to inform the prediction of the missing values. Imputation is particularly useful for meta-analysis of GWAS[29], where SNP arrays may not be the same for each sample. It allows like for like comparison.

In terms of genotyping and HLA, a person's HLA type can be imputed by comparing SNPs in their DNA to a reference panel of similar people, whose SNPs and HLA types are known. People in the reference panel who display a certain nucleotide or combination of nucleotides may all display a particular HLA allele. In this case, study subjects displaying these nucleotides can be inferred to also display this allele. In some cases, combinations of nucleotides may be, for example, 95% likely to express as a particular HLA type and 5% likely to express as a different HLA type. In this case, subjects with these nucleotides' HLA types can be imputed with 95% certainty.

A key benefit of imputation is its low cost. Typing subjects by traditional methods may cost between $50 per genome (for chip-typing) and $1000 per genome (for high coverage sequencing), whereas imputation can be done for under $0.50 per genome[52]. Imputation is also less time-consuming than traditional typing, particularly when performed on large datasets (e.g., 1m subjects).

A key drawback of imputation is its inaccuracy. It may be accurate at a population level (giving allele frequencies consistent with the reference panel) but be wrong for individuals. Karnes[53] suggested that imputation is good enough for research studies but not sufficiently accurate for clinical practice.

Another drawback of imputation is that the accuracy of imputed data is dependent on the quality of the reference panel. If a particular allele is missing in the reference panel, or if the reference panel is not representative of the study group, alleles may be missing or inaccurately imputed. Recent papers have mentioned that European Caucasian reference panels are easily available but it is more difficult to obtain reference panels for other ethnicities[53,54].

There are a number of different programs and methods available to impute HLA information, including HIBAG[55], SNP2HLA[56], and HLA*IMP:02[57]. Karnes[53] assessed these three competing methods by using them to impute data on 3,265 subjects. Outcome measures of these methods were: concordance (that is, the proportion of imputed values that were the

same as the results of sequencing); call-rate (the proportion of the time that the imputation program was able to predict the allele to a certain level of confidence); and number of alleles imputed.

The best program in terms of call-rate and the number of alleles imputed was SNP2HLA. Concordance was lower for African Americans (62-93% concordance depending on program used) than for European Americans (94-98% concordance) for all imputation programs[53]. This may be due to low availability of reference panels for African Americans or greater HLA diversity in this population. If the reference panel does not closely represent the demographic, imputed HLA values may appear accurate when in fact they are not (and vice versa). All methods performed similarly for low frequency alleles. Every method gave more accurate imputations when using GWAS platforms (either Illumina HumanOmni1-QUAD BeadChip[58] or Illumina HumanOmni5-QUAD BeadChip[59], which contain 11,675 and 26,952 SNPs in the HLA region respectively) compared to Illumina HumanExome BeadChip[60] (which contains only 2,061 “HLA tagging” SNPs). Most notably, HLA*IMP:02 had a concordance of 89% for Illumina HumanExome BeadChip compared to 94% for GWAS platforms. This suggests that high genomic coverage is better for imputation.

Neville[61] also recommends the use of SNP2HLA. They used the program to impute HLA types for 5,553 British people and compared the results to a subset of 70 who were typed using PCR, as well as to a large US resource of Caucasian people. Imputed allele frequencies were very similar to frequencies in the reference panel: there was a strong correlation between an allele’s frequency in the imputed dataset and the allele’s frequency in the reference panel ($r^2$=0.96-0.99 depending on locus). The level of confidence in the imputed values was good, especially in HLA class I (where, depending on locus, 88-94% of alleles were imputed with >95% certainty). DRB1 and DPB1 were imputed with lower confidence (respectively, 67% and 72% were imputed with >95% certainty). Other class II loci were imputed with >95% confidence for 95-99% of individuals. Neville concluded that SNP2HLA gives high confidence at low and intermediate resolution, especially when used for small haplotype blocks (rather than the entire haplotype). However, they did not use an alternative imputation method as a control. So, while they conclude that SNP2HLA is sufficient for research purposes, they cannot comment on whether it is better than any alternatives.

The HLA data used in this study was imputed using HLA-IMP:02. Its accuracy has been estimated to be >96%[62]. See section 6.6.4 for more details on the methods and accuracy of this imputation process.

## 6.6 UK Biobank

A biobank is a collection of medical data and biological samples which exists for research purposes[63]. They are often large collections which are shared among different groups of researchers and collaborators, though these are not unanimously considered essential characteristics of biobanks[64].

### 6.6.1 History and purpose

UK Biobank (UKB) is a health resource and charity based in Greater Manchester, UK which holds data on over 500,000 participants. Its purpose is to help prevent, diagnose, and treat a range of illnesses[65]. Researchers apply to use the data and, if their project is deemed to be in the public interest, can analyse it to identify genetic and environmental factors that are associated with particular diseases.

The participants were recruited in the UK between 2006-2010, and were aged between 39-73 years. They completed a questionnaire about their lifestyle and medical history, had physical characteristics measured, and gave blood, urine, and saliva samples. Subsequent diseases, drug prescriptions, and mortality information is provided by the NHS and incorporated into the database. Since the initial recruitment, some extra measures have been added, including 20,000 participants repeating the initial assessments, 200,000 participants providing more information on their diets and nutrition, and 100,000 wearing an activity monitor for a week. Hospital inpatient records as well as information from death and cancer registries have also been included. The initial funding was £62m, and the project has funding to remain active (including storing samples and making the data accessible online) until at least 2022. There are also plans to provide some further measurements[65].

Researchers have been able to apply for access to the data since 2012. It is open to researchers worldwide, from the public or private sector, as well as charities and universities. The research projects must, however, be in the public interest, and findings must be reported back to UKB so other researchers can benefit. As of June 2022, over 2,800 publications, abstracts and presentations have been produced using UKB data[66].

UKB is one of the world's biggest biobanks. It has generally been praised, for example an independent review panel said when it opened that it had the potential to "support a wide range of research" in ways that were "not currently available elsewhere"[67]. The chief executive of the Medical Research Council said it would provide "extraordinary information" and would offer "unprecedented opportunities to improve people's lives"[68].

### 6.6.2 Data accuracy and biases

The reliability of UKB data can be assessed by comparing UKB's measurement of a particular variable to a well-validated measurement of the same variable. If there is a strong correlation between UKB's measurements and the validated measurements, this increases confidence in UKB's data. Fawns-Ritchie compared UKB's "measure of general cognitive ability" with a measure created using well-validated cognitive tests, and found good correlation ($r=0.83$, $P<0.001$)[69]. They also measured the "test-retest reliability" in cases where subjects had the same cognitive measurements taken by UKB on separate occasions, four weeks apart. The test-retest reliability was "moderate-to-high". They recommended that researchers use the UKB data fields that were found to have both moderate-to-high test-retest reliability and moderate-to-high agreement with well-validated measurements. They also noted that subjects' self-rated assessments of their own memory were not significantly associated with their performance on cognitive tests. This suggests that self-reported measures in the UKB dataset may not be as reliable as more objective measures. This study included only 160 subjects and focused only on measurements of cognitive ability; it may not be possible to generalise these findings to all subjects and data fields within the biobank. Similarly, Lyall investigated over 20,000 subjects who had repeated cognitive assessments an average of four years after their initial assessment[70]. They found "reasonable stability" for measures of reasoning and reaction time, but a measurement of visual memory "did not show good stability across time". Nevertheless, they concluded that UKB is a "valuable resource" for future studies.

The subjects in the UKB cohort may not be representative of the general population. 9.2m individuals aged 40-69 were invited to enter the cohort, but only 5.5% of these participated in the baseline assessments. People who died before they reached 40 years old were not considered, meaning the cohort may suffer from survival bias. Fry found that those included in the cohort were older, more likely to be female, and less likely to live in a deprived area than those who did not respond[71]. When compared with the general population, UKB subjects were less likely to be obese, less likely to smoke, and less likely to drink alcohol every day. They also had fewer self-reported health conditions and lower mortality rates when they reached 70-74 years. This suggests that there is a "healthy volunteer" selection bias, which may reduce the applicability of findings. UKB participants are more likely to own their home outright than the general population, suggesting increased wealth. 94.6% of the cohort reported their ethnicity as "white"; this is in line with 40-69-year-olds in the 2001 UK census (94.5%) but higher than the 2011 census (91.3%); recruitment was performed between these censuses (2006-2010). Due to these biases, Fry advised that the biobank could not be used to measure

disease prevalence or incidence. However, they noted that "valid assessment of exposure-disease relationships may be widely generalisable and does not require participants to be representative of the population at large"[71], suggesting that UKB may be an appropriate source of data for the present study. Section 6.6.4 summarises the sample storage procedures[72] and HLA imputation methods[62] that impact on this study.

### 6.6.3 Ethics

UKB is subject to an Ethics and Governance Framework and accountable to the independent UKB Ethics Governance Council. The Framework defines the standards which UKB must meet in the creation, maintenance, and use of the resource, and outlines its commitments to the participants, researchers, and the general public. There were a number of groups consulted in the creation of this Framework, including experts in research ethics, the law, and science, as well as the public. The Council advises the project and monitors conformity to the Framework. The Council also provides information on the project to participants and the public[73].

UKB has been approved to conduct research by a number of bodies such as the North West Multi-Centre Research Ethics Committee, the Community Health Index Advisory Group, the Research Tissue Bank (RTB), and the Human Tissue Authority (HTA). RTB approval means that applicants do not need ethics approval for each individual project, but rather UKB has generic approval for most research. HTA approval represents a licence for researchers to use samples without needing a separate HTA licence[73].

The project was criticised by the director of GeneWatch UK, who claimed it would "not accurately measure environmental factors" and that there was a "real danger of spurious results". GeneWatch also expressed concern about commercial companies accessing participants' data, saying that without "clear controls" the database would be "open to abuse" such as gene patenting[74].

### 6.6.4 Use for this study

The main reasons for using UKB for this project were its sample size, the scope of its data, and its affordability. As independent researchers, it would not have been possible to get such rich data on so many aspects of people's lives and genetics without incurring prohibitive costs, particularly for such a large sample of people. Genome data from all 500,000 participants were available. Individually typing this number of people would have been far too time-consuming and expensive to be practical, but imputed HLA data were available from UKB for a nominal cost. Information on the general health, physical characteristics, lifestyle, and environment of

these participants was also available, which would have been very difficult to obtain independently.

The accuracy of UKB's biomarker measurements was essential for this study as the primary outcome measures were based on the concentration of cystatin and creatinine in each subject's serum. Peakman tested UKB's processes for sample handling and storage; they found that the processes were "robust" and the tests provided "justification" for the procedures used by UKB[72]. This increases confidence in the accuracy of these measurements.

The accuracy of UKB's imputed HLA types was also vitally important for this study. Bycroft reviewed UKB's imputation using a "cross-validation experiment" and found the accuracy to be "better than 96%" for a "typical use"[62]. More detail is provided in the Supplementary Information of Bycroft's paper (section S5). Assuming a posterior probability threshold of 0.7 (that is, excluding types that were imputed with <70% confidence), they estimate the accuracy in European subjects to be between 96.1% (HLA-DPB1) and 99.6% (HLA-DPA1). 4.9% of HLA-DPB1 values were excluded due to the posterior probability threshold (i.e., excluded due to lack of confidence); all other loci had fewer values excluded. For African/African American subjects, accuracy estimates range from 89.3% (HLA-B) to 99.6% (HLA-DRB5). The locus with the most values excluded was HLA-B (12.5%). For Asian subjects, accuracy ranged from 89.5% (HLA-DPB1) to 100% (HLA-DRB4). The locus with the most values excluded was HLA-B (11.3%). The research presented in this thesis used a more stringent posterior probability threshold (0.8), so the accuracy is probably higher than the estimates above (but more imputed values were excluded due to lack of confidence).

Bycroft also provides information on the reference panels used by UKB[62]. There was a different reference panel for each locus, each combining individuals from various datasets. For each locus, only individuals who had a laboratory-derived HLA type were included in the reference panel. Imputation models were based on markers that had been genotyped in both the reference panels and the UKB dataset. Each locus had 623-927 SNPs contributing to its imputation, each of which was polymorphic and was typed for >98% of the individuals included. There were 5,615-8,112 Europeans in the reference panels depending on locus (excluding HLA-DRB3, DRB4, and DRB5, which are much less polymorphic so don't need as many subjects). There were 0-346 Africans/African Americans depending on locus; eight of the 11 loci had over 200 African/African American individuals. There were 307-503 Asians included (again excluding HLA-DRB3, DRB4, and DRB5).

Bycroft also used the imputed HLA types to attempt to replicate known associations between HLA and 11 immune-mediated diseases, and their findings were "consistent with previous reports"[62]. Overall, this paper suggests that the HLA types imputed by UKB are sufficiently

accurate for this study. In particular, the HLA typing accuracy of white British subjects is expected to be over 96% for every locus. As far as I am aware, no other validations of the accuracy of UKB's imputed data have been performed, though Naito recently developed a different imputation method using UKB data[75]. They claim this method can impute rare alleles with greater accuracy than HLA-IMP:02. Nevertheless, UKB's imputed data has been used in many disease association studies[76-78], suggesting that researchers and peer-reviewers consider the data to be suitable for this sort of study.

## 6.7 Covid-19

Covid-19 is an infectious respiratory disease caused by the SARS-CoV-2 virus. It was first identified in December 2019, and was declared a global pandemic by the World Health Organization in March 2020[79]. By March 2022, over 6m people had died of Covid-19 and the societal, economic, and healthcare implications of the pandemic are expected to last for many years[80].

Symptoms include cough, fever, shortness of breath, weakness, muscle pain, sore throat, loss of taste and/or smell, vomiting, diarrhoea, and others[81,82]. Some infected individuals develop no symptoms, some recover within a week, while others take longer. One third of survivors experience "long covid"[83], meaning their symptoms persist for more than four weeks[84].

The disease is highly infectious, and is usually spread via respiratory droplets during human-to-human contact within two metres[85]. Contagion rates vary based on viral load[86] and infection control measures; the UK's highest estimated "reproduction rate" (the average number of people that an infected person will subsequently infect[87]) was between 1.3-1.6 in October 2020[88].

Mortality rates also vary based on multiple factors. A UK study published in July 2020 found the following factors to be associated with increased risk of death from Covid-19: male sex; increased age; increased deprivation; diabetes; asthma; Black ethnicity (compared with white); South Asian ethnicity (compared with white); and various other medical conditions[89]. Vaccination reduces both the chance of being infected with Covid-19 and the chance of severe illness and death[90,91].

Covid-19 appears to be linked to kidney function, with kidney transplant recipients more likely to test positive and more likely to die of the disease[92]. A study of people hospitalised due to Covid-19 found kidney disease to be associated with increased mortality[93]. The association appears to go both ways: as well as kidney disease contributing to Covid-19 susceptibility and severity, Covid-19 may also be a cause of kidney disease. Survivors display increased risk of

eGFR decline and “major adverse kidney events” compared with people who had not had Covid-19[94].

There may be HLA associations with Covid-19. A systematic review published in August 2021 found that a number of studies have reported certain HLA alleles to be linked to increased risk of Covid-19[95], including (among others): *HLA-C*01* and *HLA-B*44* (in a study of Italian subjects)[96]; and *HLA-A*11:01, B*51:01*, and *C*14:02* (in a study of Chinese subjects)[97]. However, there was no clear agreement as to which alleles were associated, and the authors of the systematic review recommended further investigations using larger cohorts.

# 7 Aims of this study

The aims of this study are:

- to compile a comprehensive list of all previously published associations between HLA and renal function.

This was achieved by performing a systematic review of a number of databases (see section 10), searching for all peer-reviewed papers on the subject, and creating a resource so that researchers can see all associations that have been published “at a glance”.

- to discover HLA alleles which impact upon renal function in a population of middle-aged people in the UK.

This was achieved with the publication of two peer-reviewed papers, one investigating white British people (see section 11) and one investigating people of six other ethnicities (see section 12). Some of these findings were novel (i.e., associations that had not been previously reported), while some were replications of findings that had already been published.

- to investigate the link between HLA, Covid-19, and renal function.

This was achieved, to some extent, by producing two papers. The first investigated associations between HLA and Covid-19 in a cohort of 700 Covid-19 positive white British subjects and 1,400 controls (see section 13). The second looked for associations between Covid-19 and a variety of factors including kidney function in a cohort of 6,000 Covid-19 positive subjects and 30,000 controls (see section 14). However, these analyses had limited findings and they were not published in peer-reviewed journals.

It is possible that, rather than HLA types being the cause of renal failure, they are in fact markers of closely associated genes which may contribute towards the onset of renal disease. However, whether causative or not, associations between HLA type and renal failure can still have important medical applications, and it was necessary to both collate the findings that had already been published and to contribute new research which added to these findings.

There are a number of ways in which the findings of this research can be applied within clinical and academic settings. The study highlighted immunogenetic associations which previously had not been discovered, and also replicated findings from previous studies. This is of particular value in renal diseases with immune-related causes such as lupus or membranous nephropathy. The age of the participants is also significant as renal function is known to deteriorate with age; studying patients in middle- to old-age is therefore important as older people are more likely to suffer from renal failure and to benefit from the findings. The research

has the potential to inform the clinical management of patients with renal disease, as well as to help diagnose patients or identify susceptibility to renal disease.

# 8 Methods

The analysis used data from UKB (which comprises over 500,000 subjects who were aged between 39-73 years when they were recruited in 2006-2010). These subjects' HLA types were imputed by UKB. The data were analysed with respect to clinical measures of renal function such as eGFR, and outcomes such as need for renal replacement therapy (e.g., dialysis or transplantation). Associations between HLA type and incidence of renal dysfunction were then discovered.

Prior to analysis, the integrity of the data was ensured by applying quality control measures to remove subjects for reasons such as discrepancies and inaccurate data, related pairs, subjects with large amounts of missing data, and subjects who withdrew their consent. Association analyses between HLA type and renal function were performed using Plink software. The significance threshold was based on the number of tests performed, with an initial standard of $1.4 \times 10^{-4}$. This section gives methodological details that were not included in the published papers.

## 8.1 Acquisition of data

In order to receive data from UKB, researchers must submit an application and show that they meet certain criteria. Specifically, the research must be health-related and in the public interest, and results should be returned to UKB. The applicant must provide details of the aims of the project, how the project meets UKB's purpose, and a description of how the research will be carried out (including a proposed timetable). The data requested (i.e., which participants, fields, and samples are required) must also be specified. A copy of the form submitted for this project is shown in section 19.2.1.

Once the application was approved (16th March 2018), the data was downloaded from UKB's website, decrypted using a unique password, and converted into a usable file format (in this case, Stata format). This process was repeated several times as supplementary data files were provided after the initial data release. The separate files were merged to ensure all data was available in a single file. The steps used to download, decrypt, and format the data are shown in section 19.2. Quality control was performed as described in the papers. This involved excluding subjects who had ambiguous genetic sex, who were close relations of other subjects, and who had missing genetic data. Subjects were also stratified based on ethnicity.

## 8.2 Plink analysis

Genetic association analysis was performed in Plink. Stata and Plink commands can be seen in section 19.4. Prior to analysis, a version of the dataset was created in Stata containing only

a small number of essential phenotype fields. This was to reduce the size of the file, which resulted in faster processing. The fields chosen included age, sex, and measures of kidney function such as eGFR and kidney transplant status. A unique patient identifier was also included. Missing values were converted to -9, as per the Plink standard. The phenotype file was then exported in tab-delimited .csv format to a folder which also contained the Plink programme, and UKB HLA data stored in a binary (.bed) format.

In all analyses, significance was calculated by dividing 0.05 by the number of HLA types included in the analysis. For example, when all 362 types were included, a P value of 0.00014 or lower was considered significant and when only 100 types were included, $P<0.0005$ was considered significant. A complete list of all 362 types included in the analysis, along with their frequencies, can be seen in section 19.5.

## 8.3 HLA and renal function analysis

Regression analyses were performed to test for associations between HLA types and renal function. The primary outcome measure was eGFR (continuous data; the standard clinical test of kidney function). Secondary outcome measures were also used: ESRD; CKD status; dependence on renal replacement therapy; and kidney transplant status (all of which are categorical data relating to clinical diagnoses and interventions). When the outcome measure was continuous, linear regression was used. When the outcome measure was categorical, logistic regression was used.

# 9 Results

## 9.1 First author publications

Following the methods outlined above, I published three peer-reviewed papers as first author. One is a systematic review which collated all previous attempts to find associations between HLA and renal function[98]. Over 100 associations were found in 35 different studies. This paper was published in International Journal of Immunogenetics in December 2021. Prior to this, I wrote two original articles: one investigating HLA associations with renal function in white British subjects[99], and another performing the same analysis in six different ethnic groups[100]. These were published in Scientific Reports (February 2021) and HLA (September 2020) respectively. A number of associations were found in each study, including a common haplotype (*HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DQB1*02:01*, frequency 9.5% in England) which appears to be associated with reduced kidney function in white British subjects. The three papers are included in full in this thesis.

## 9.2 Covid-19 research

At the beginning on the Covid-19 pandemic, I contributed to a study detailing the likely impact of the pandemic on the UK's renal transplantation program[14]. We concluded that the closure of transplantation centres would result in at least 700 kidney transplant opportunities being lost in the UK, and a commensurate increase in the dialysis population. The paper advised that clearing the backlog of transplants should be a high priority once the immediate threat of the pandemic receded. This paper was published in Clinical Medicine in July 2020 and can be seen in section 19.6.2.

I also used the UKB dataset to assess whether there were any associations between Covid-19 and either HLA or kidney function. Firstly, I tested for associations between 50 HLA types and Covid-19 status in white British UKB subjects (756 Covid-19 positive subjects, 1,449 Covid-19 negative subjects, and 387,356 subjects who were not tested). However, none of the alleles analysed was significantly associated with Covid-19 status. One allele trended towards significance: *HLA-DQA1*03:01* may be less frequent in Covid-19 positive subjects than in subjects who had not been tested for Covid-19 (P=0.0038, odds ratio 0.82, 95% confidence interval 0.72-0.94). However, after Bonferroni correction for multiple tests, this association was not statistically significant. This research was not submitted for peer-review as a similar study using the same dataset was released in pre-print form before we were ready to submit[101]. The draft of my paper can be seen in section 13.

Later, I investigated whether Covid-19 status was associated with eGFR. eGFR values for 6,223 Covid-19 positive subjects were compared with eGFR for 30,531 Covid-19 negative

controls. Death within 28 days of Covid-19 diagnosis was used as a secondary outcome (n=247). Kidney function was not associated with either of these outcomes. This research was submitted to Kidney International Reports in February 2021 but was not accepted for publication. The full version, as submitted, can be seen in section 14. It may be worth updating these two papers now that more Covid-19 cases have been identified; the increased statistical power may result in significant findings.

## 9.3 Other work

Throughout the course of my degree, I undertook other research that was either not published or did not directly relate to the research project. This section summarises these projects.

### 9.3.1 Seasonal variance in renal transplantation

I was the first author on a paper investigating the seasonal variance of renal transplantation in the UK[102]. This found that deceased donor transplant activity is significantly increased in November and December compared to other times of year, possibly due to higher mortality rates resulting in greater availability of donors. There was also variance in living donor rates throughout the year. The paper recommended that transplantation centres be aware of and prepare for this predictable increase in activity. It was published in BMJ Open in September 2019 and is included in full in section 19.6.1.

### 9.3.2 HLA mismatch in cardiothoracic transplantation

I contributed to a paper assessing how HLA epitope mismatch between donor and recipient was associated with the development of donor-specific antibodies (DSA) in cardiothoracic transplantation[103]. The paper compared three software programs which calculate mismatch scores for donor-recipient pairs. One of these scores (HLA-EMMA) was significantly associated with DSA production, while the other two (HLAMatchmaker, and PIRCHE-II) were not. This suggests that HLA-EMMA is the most useful tool for predicting DSA production. This tool may therefore be used to help inform donor selection and to identify recipients who are at increased risk of producing DSAs. This paper was published in International Journal of Immunogenetics in December 2021 and is included in full in section 19.6.3.

I intend to perform a similar study investigating whether these tools can predict kidney transplant outcomes (see section 16.5.2). The scores from the three tools will be analysed with respect to the outcomes of kidney transplants performed at Manchester Royal Infirmary. Outcome measures will be eGFR, graft survival, and patient survival. If any of the tools is a good predictor of transplant success, it could be used to help donor selection in future.

# 10 Paper 1: Systematic review of associations between HLA and renal function

This paper was published in the International Journal of Immunogenetics on 17th December 2021.

Received: 27 July 2021 | Revised: 1 November 2021 | Accepted: 17 November 2021

DOI: 10.1111/iji.12566

INTERNATIONAL JOURNAL OF IMMUNOGENETICS WILEY

REVIEW

# Systematic review of associations between HLA and renal function

Marcus Lowe[1,3] | Steven Jervis[1,3] | Antony Payton[2] | Kay Poulton[1,3] | Judith Worthington[1] | Isla Gemmell[3] | Arpana Verma[3]

[1] Transplantation Laboratory, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK

[2] Faculty of Biology, Medicine and Health, Division of Informatics, Imaging & Data Sciences, School of Health Sciences, University of Manchester, Manchester, England

[3] Faculty of Biology, Medicine and Health, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, University of Manchester, Manchester, England

**Correspondence**
Marcus Lowe, Transplantation Laboratory, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK.
Email: Marcus.Lowe@mft.nhs.uk

## Abstract

**Introduction:** Kidney dysfunction is a highly significant disease, both in the United Kingdom and globally. Many previous studies have reported associations between human leukocyte antigens (HLA) and renal function; this systematic review attempts to identify, summarize and appraise all published studies of these associations.

**Methods:** A literature search was performed using Medline, Embase and Cochrane Central Register of Controlled Trials to identify papers whose keywords included each of the following concepts: HLA, renal failure and genetic association. A total of 245 papers were identified and assessed for eligibility; 35 of these were included in the final study.

**Results:** A total of 95 HLA types and 14 three-locus haplotypes were reported to be associated with either increased or decreased renal function. A number of these findings were replicated by independent studies that reported 16 types were protective against renal dysfunction and 15 types were associated with reduced renal function. A total of 20 HLA types were associated with both increased risk of renal disease and decreased risk by independent studies.

**Discussion:** There is very little consensus on which HLA types have a protective or deleterious effect on renal function. Ethnicity may play a role, with HLA types possibly having different effects among different populations, and it is possible that the different primary diseases that lead to ESRD may have different HLA associations. Some of the studies may contain type I and type II errors caused by insufficient sample sizes, cohort selection and statistical methods. Although we have compiled a comprehensive list of published associations between renal function and HLA, in many cases, it is unclear which associations are reliable. Further studies are required to confirm or refute these findings.

KEYWORDS
chronic, genetic predisposition to disease, genome-wide association study, histocompatibility antigens, HLA antigens, kidney failure, renal insufficiency, systematic review

## 1 | INTRODUCTION AND AIMS

Chronic kidney disease (CKD) is diagnosed either when a person has glomerular filtration rate (GFR) < 60 ml/min/1.73 $m^2$ or when they have structural or functional abnormalities of the kidney that could lead to renal failure (The Cochrane Collaboration, 2019). In England, over 1.8 m people are affected and it is estimated that over 40,000 of these people die prematurely each year (NHS, 2012). The cost of treating CKD in England in 2019 was approximately £1.5bn (Health Service Journal, 2020), representing 1% of all NHS expenditure. Lower kidney function is associated with decreased quality of life and increased risks of mortality, stroke, myocardial infarction, infection and hospital admission (Go et al., 2004; NHS, 2012; Tonelli et al., 2006; Weiner et al., 2004).

According to the Global Burden of Disease 2015 Study (Mortality & Causes of Death, 2016), 1.2 million people worldwide died due to CKD in 2015 (an increase of 32% on 2005). CKD is now 17th in the list of diseases which cause the most 'years of lost life' (having been 21st in 2005 and 25th in 1990). Understanding the causes of kidney dysfunction could have important applications in terms of diagnosis and treatment of one of the most globally significant diseases.

CKD has a high heritability index (30–75%) (Canadas-Garre et al., 2019), with over 100 genomic regions reported to contribute towards it (Xu et al., 2018). Many previous studies have identified human leukocyte antigen (HLA) types that are associated with increased risk of end-stage renal disease (ESRD), which is when renal replacement therapy (such as dialysis or transplantation) becomes necessary. Other HLA associations indicate a protective effect. Most of these studies focus on subjects from specific ethnic cohorts or particular geographical regions, and HLA associations reported may be contradictory due to variation in disease prevalence and population stratification.

This study performed a systematic search of existing literature. The aim was to identify, summarize and appraise all previous studies which have attempted to find associations between HLA type and renal function. The HLA region was selected for analysis because it is closely linked to many disorders of the kidney (Robson et al., 2018). This suggests that it may have an impact on kidney function.

## 2 | METHODS

This literature review investigates HLA genotypes associated with renal function in global populations. It was defined in terms of the PICO framework, a process which advises clear definitions of the participants, intervention, comparison and outcome that are to be studied (Huang et al., 2006). The participants (i.e. the subjects to be included) were global populations aged 18 years and over. The interventions (the independent variable) were HLA class I and II types. The comparisons (controls) were subjects without renal disease. The outcome (the dependent variable) was renal function (either increased or decreased). As of 19 September 2019 there is no review protocol for studies of HLA and renal function (as per a search of PROSPERO), though there is one for studies of HLA mismatches and kidney transplant outcomes (Shi et al., 2017).

A literature search was carried out on 30 November 2018 and updated on 17 February 2021 using Medline, Embase and Cochrane Central Register of Controlled Trials on the Ovid platform. These databases are a comprehensive and relevant source of papers and reviews which date back to 1946, 1974 and 2005 respectively. Only primary studies which had been published as full, peer-reviewed papers (rather than abstracts only) were considered. This helped to ensure that only high-quality, reliable evidence was used. Papers were also required to be written in the English language, to guarantee comprehension by the researchers. Their database thesaurus (or 'index') terms were required to include words or phrases associated with the following concepts: HLA, renal failure and genetic association. Medline scope notes were checked and all terms associated with key terms within the MeSH thesaurus tree were examined.

We originally intended to consider only papers investigating subjects of white ethnicity, as this most closely reflected the subjects of a research project that we were planning at the time and have since conducted (Lowe et al., 2021). However, ultimately any ethnicity was included as ethnicity was not reliably indexed in the databases searched, and filtering using this criterion would have lost relevant studies. Medline, for example, had 62,789 results for the search term 'European Continental Ancestry Group' (the index term for the word 'Caucasian'). If the term had been properly indexed, the number of papers identified would have been much higher. The search criterion for white ethnicity was therefore not used. For similar reasons, the review initially intended to consider only subjects of middle-age or older, but age was ultimately not selected as a search term because requiring the index terms 'aged' or 'middle-aged' reduced the yield of publications by approximately 75%. As a result of excluding the concepts of ethnicity and age from the search, papers were identified which related to participants with a wide range of ethnic origins and ages. The observations contained within these papers, therefore, cannot necessarily be applied directly to the middle-aged white population as had originally been intended. A number of more sensitive searches were attempted, using broader search terms. For example, the search term 'haplotypes' was added to the genetic association concept and to the HLA concept. However, the additional papers highlighted through this strategy were not relevant to the research question, so the term was released.

A total of 242 papers were identified through the database search, but an earlier scoping exercise had revealed three additional publications which were not included in the database search (Davood et al., 2008; Mosaad et al., 2014; Nassar et al., 2015), possibly due to their selected keywords. The keywords of two papers did not include the concept of genetic association (Mosaad et al., 2014; Nassar et al., 2015). The keywords of the third paper suggested that this publication should have been captured by the search (Davood et al., 2008); possible explanations for this omission are that either the keywords were not indexed correctly or the journal (Research Journal of Biological Sciences) was not included in the databases searched. All three papers were included, giving a total of 245 publications selected. Of

**TABLE 1** Primary reason for exclusion after screening of titles and abstracts

| Reason for exclusion | Number of results excluded |
|---|---|
| Not related to kidneys | 42 |
| Investigating transplantation but not kidney failure | 38 |
| Not related to HLA class I and II | 33 |
| Only tangentially related to kidney failure | 16 |
| Case study | 14 |
| Study of children | 6 |
| New HLA allele discovery | 4 |
| Not an article | 2 |

these 245 papers, 32 were duplicates of the same publication found in different databases, leaving 213 unique papers for screening. The titles and abstracts of these 213 papers were read to assess their relevance to the research question. A total of 155 were deemed irrelevant to the research question and were removed. The primary reason for each exclusion is documented in Table 1. After this initial screening exercise, 58 papers remained for consideration.

In 17 cases, no full text version of the paper was available, either because the reference related to a conference poster or abstract which had not been published as a paper or because the full text was inaccessible. The University of Manchester library was consulted for help in obtaining these papers, but was unsuccessful. This left 41 full text papers for assessment. A further six publications were removed after the full text articles were assessed, either because the paper was not related to HLA class I or II ($n = 3$), the paper did not measure renal failure subjects against healthy controls ($n = 2$), or because the paper was not a primary study ($n = 1$). Thirty-five articles, therefore, were deemed relevant to the research question. Figure 1 summarizes the entire process of exclusion following a strategy adapted from Moher et al. (2009), which outlines the process for completing a systematic review.

The 35 papers were analysed and the results extracted from tables in the Section 3 and prose in the Section 4. No contact with original researchers was made to verify their data or request further information such as funding sources. Each paper was also subject to an assessment using the Critical Appraisal Skills Programme to ensure it was of sufficient quality to be included (Critical Appraisal Skills Programme, 2018). The search was carried out by a single researcher with consultation and advice from two librarians. All screening and assessments of eligibility were performed independently by two researchers who discussed any discrepancies until consensus was reached. Extraction of data from the papers was performed by a single researcher. Different studies used different methods to determine whether HLA types were associated with renal function. The principal summary measures were odds ratio, relative risk and hazard ratio.

## 3 | RESULTS

Associations between HLA and renal function were identified in 30 of the 35 papers. In all, there were a total of 181 associations reported, relating to 58 different HLA class I types, 37 class II types and 14 three-locus haplotypes. None of the 14 findings which related to a haplotype was independently replicated, but 31 of the findings relating to a single HLA type were replicated. There were 20 types which were found to be associated with increased risk of kidney disease by at least one study, but protective against kidney disease by at least one other.

Sixteen HLA types (HLA-A*24, A*26, A*29, A*30, A*32, B*07, B*40, B*44, C*02, DRB1*03, DRB1*04, DRB1*08, DRB1*11, DRB1*13, DQA1*03 and DQB1*06) were found to be protective against ESRD by multiple studies (though nine of these were also found to be associated with increased risk of ESRD by at least one other study: A*24, B*07, B*40, C*02, DRB1*03, DRB1*04, DRB1*11, DRB1*13 and DQA1*03). In total, 38 class I types, 24 class II types and 8 haplotypes were found to be protective against ESRD, though 11 of the class I types and 9 of the class II types were in conflict and also associated with ESRD. Similarly, 15 HLA types (HLA-A*11, B*08, B*15, B*18, B*49, B*50, B*51, B*53, B*55, C*01, DRB1*03, DRB1*04, DRB1*11, DRB1*12 and DQB1*02) were associated with ESRD in at least two studies. However, seven of these were protective against ESRD according to a different study (HLA-A*11, B*08, B*50, DRB1*03, DRB1*04, DRB1*11 and DRB1*12). In total, 31 HLA class I types, 22 class II types and 6 haplotypes were found to be associated with renal dysfunction (though, as mentioned above, 11 of the class I types and 9 of the class II types were also found to be protective against ESRD by different studies).

Many of the findings were directly refuted by other studies. There does not appear to be a consensus around which HLA types have a protective effect and which incur additional risk of ESRD; of the 95 HLA types with a reported association, 20 had the finding refuted by at least one other independent study (21%). Only 10 HLA associations were reported in three or more studies (HLA-A*11, B*07, B*08, B*53, DRB1*03, DRB1*04, DRB1*08, DRB1*11, DQB1*02 and DQB1*06), and 6 of these were refuted by another study (the exceptions being HLA-B*53, DRB1*08, DQB1*02 and DQB1*06). Table 2 shows all reported associations between HLA and renal function (some associations with increased incidence of renal dysfunction, and some with decreased incidence). The table illustrates that there are a large number of HLA types which may have an effect on renal failure. The subjects' primary diseases are noted in the table (where possible) to allow for comparison of associations based on the underlying cause of renal dysfunction. There are a number of abbreviations in Table 2; these are expanded below the table. Five of the papers did not report any significant associations. These are listed in Table 3.

## 4 | DISCUSSION

The 35 papers identified 95 HLA types and 14 HLA haplotypes which were associated with renal function. Some of these types appear to confer protection from renal failure, while others appear to confer susceptibility to renal failure. Indeed, 20 HLA types were found to have a protective effect in at least one study but a hazardous effect in at least one other. Only 10 of the associations were replicated by three or more independent studies. This suggests that there may be type I and type II

**TABLE 2** All HLA types and three-locus haplotypes found to be associated with renal function

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| HLA-A | A*01 | Karahan et al. (2009) | Protective against CGN, HTN nephrosclerosis | Turkish | 3230 (587 ESRD patients, 2643 controls). Primary diseases: unknown 27%; CGN 23%; HTN nephrosclerosis 17%; PKD 9%; pyelonephritis 9%; VUR nephropathy 4%; diabetic nephropathy 4%; amyloidosis 3%; urologic abnormalities 2%; other 3% |
| | | Lowe et al. (2021) | Reduced eGFR, increased ESRD | British | 401,307 (with a range of kidney function, eGFR calculated for each) |
| | A*02 | Karahan et al. (2009) | Protective against CGN, HTN nephrosclerosis | Turkish | See above |
| | A*03 | Karahan et al. (2009) | Protective against CGN, HTN nephrosclerosis | Turkish | See above |
| | | Lowe et al. (2021) | Reduced eGFR | British | See above |
| | A*09 | Rivera et al. (2012) | Protective against ESRD | Venezuelan | 390 (188 ESRD patients, 202 controls). Majority post-streptococcal or other glomerulonephritis origin |
| | A*11 | Davood et al. (2008) | Associated with ESRD | Azerbaijani | 77 (26 ESRD patients awaiting transplantation, primary disease not specified, 51 controls) |
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | 3326 (1965 ESRD patients awaiting transplantation, primary disease not specified, 1361 potential kidney donors) |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | 2083 (499 ESRD patients awaiting transplantation, 1584 controls). Primary diseases: unknown 63%; glomerulonephritis 13%; HTN nephropathy 8%; diabetic nephropathy 6%; interstitial nephritis 5%; PKD 2%; HSPN 2%; obstructive nephropathy 1%; autoimmune diseases 1% |
| | | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | A*23 | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | A*24 | Cao et al. (2014) | Associated with ESRD | Cantonese | 8285 (4541 ESRD patients awaiting transplantation, 3744 controls). Primary disease: 57% not specified; 43% glomerulonephritis |
| | | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | | Uygun et al. (2015) | Protective against ESRD | Turkish | 677 (144 highly PRA positive ESRD patients, primary disease not specified, 533 controls) |
| | A*25:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | A*26 | Hamdi et al. (2014) | Protective against ESRD | Saudi Arabian | 455 (350 ESRD patients, primary disease not specified, 105 controls) |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | A*28 | Mosaad et al. (2014) | Protective against ESRD | Kuwaiti | 525 (334 ESRD patients, primary disease not specified, 191 controls) |
| | A*29 | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | | Lowe et al. (2021) | Increased eGFR | British | See above |
| | A*30 | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | | Nassar et al. (2017) | Protective against ESRD | Yemeni | 381 (187 CRF patients on haemodialysis, primary disease not specified, 194 controls) |
| | A*31:01 | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | A*32 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | A*33 | Davood et al. (2008) | Associated with ESRD | Azerbaijani | See above |
| | A*66 | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | A*68 | Karahan et al. (2009) | Protective against ESRD, CGN, HTN nephrosclerosis | Turkish | See above |
| | A*69 | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | A*78 | Crispim et al. (2008) | Associated with ESRD | Sao Paolo (Brazil) | 265 (105 ESRD patients awaiting transplantation, 160 controls). Primary diseases: undetermined 33%; HTN 25%; diabetes 10%; renal cystic disease 8%; Berger disease 5%; glomerulosclerosis 5%; lupus 4%; other 11% |
| HLA-B | Bw*4 | Prakash et al. (2013) | Associated with ESRD | North Indian | 1024 (512 ESRD patients, primary disease not specified, 512 controls) |
| | B*07 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | 2831 (1620 IgA nephropathy patients awaiting kidney transplant, 1211 controls) |
| | | Lowe et al. (2021) | Reduced eGFR | British | See above |
| | | Hieu et al. (2019) | Protective against ESRD | Vietnamese | 383 (196 ESRD patients, 187 controls). Primary diseases: CGN 76%; HTN 10%; diabetes 6%; PKD 4%; other 4% |
| | | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | B*08 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | | Lowe et al. (2021) | Reduced eGFR, increased ESRD, increased CKD | British | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | | Mosaad et al. (2014) | Associated with ESRD | Kuwaiti | See above |
| | B*12 | Rivera et al. (2012) | Protective against ESRD | Venezuelan | See above |
| | B*14 | Crispim et al. (2008) | Protective against ESRD | Sao Paolo (Brazil) | See above |
| | B*14:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | B*14:02 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | B*15 | Hamdi et al. (2014) | Associated with ESRD | Saudi Arabian | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | B*17 | Rivera et al. (2012) | Protective against ESRD | Venezuelan | See above |
| | B*18 | Hamdi et al. (2014) | Associated with ESRD | Saudi Arabian | See above |
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | B*35 | Doxiadis et al. (2001) | Associated with IgA nephropathy | European | See above |
| | B*38 | Rivera et al. (2012) | Associated with ESRD | Venezuelan | See above |
| | B*39 | Hamdi et al. (2014) | Protective against ESRD | Saudi Arabian | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | B*40 | Cao et al. (2014) | Associated with ESRD | Cantonese | See above |
| | | Noureen et al. (2020) | Protective against ESRD | Pakistani | 1169 (497 ESRD patients, primary disease not specified, 672 controls) |
| | | Rivera et al. (2012) | Protective against ESRD | Venezuelan | See above |
| | B*42 | Yamakawa et al. (2014) | Associated with ESRD | Brazilian | 183 ESRD patients on haemodialysis, primary disease not specified, multiple control groups used |
| | B*44 | Yamakawa et al. (2014) | Protective against ESRD | Brazilian | See above |
| | | Lowe et al. (2021) | Increased eGFR | British | See above |
| | B*45 | Yamakawa et al. (2014) | Associated with ESRD | Brazilian | See above |
| | B*48 | Rivera et al. (2012) | Protective against ESRD | Venezuelan | See above |
| | B*49 | Hamdi et al. (2014) | Associated with ESRD | Saudi Arabian | See above |
| | | Davood et al. (2008) | Associated with ESRD | Azerbaijani | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | B*50 | Hamdi et al. (2014) | Protective against ESRD | Saudi Arabian | See above |
| | | Noureen et al. (2020) | Associated with ESRD | Pakistani | See above |
| | | Uygun et al. (2015) | Associated with ESRD | Turkish | See above |
| | B*51 | Rivera et al. (2012) | Associated with ESRD | Venezuelan | See above |
| | | Yamakawa et al. (2014) | Associated with ESRD | Brazilian | See above |
| | B*52 | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | B*53 | Rivera et al. (2012) | Associated with ESRD | Venezuelan | See above |
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | | Lowe et al. (2020) | Reduced eGFR | Black African | 3038 subjects (with a range of kidney function, eGFR calculated for each) |
| | B*54 | Cao et al. (2014) | Associated with ESRD | Cantonese | See above |
| | B*55 | Cao et al. (2014) | Associated with ESRD | Cantonese | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | B*57 | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | B*58 | Karahan et al. (2009) | Associated with ESRD. Protective against amyloidosis | Turkish | See above |
| | B*62 | Rivera et al. (2012) | Associated with ESRD | Venezuelan | See above |
| HLA-C | C*01 | Fejzic et al. (2017) | Associated with ESRD | Bosnian | 245 (186 ESRD patients, primary disease not specified, 59 controls) |
| | | Prakash et al. (2013) | Associated with ESRD | North Indian | See above |
| | C*02 | Prakash et al. (2013) | Associated with ESRD | North Indian | See above |
| | | Lowe et al. (2021) | Increased eGFR | British | See above |
| | | Almogren et al. (2012) | Protective against ESRD | Saudi Arabian | 295 (235 ESRD patients awaiting transplantation, 60 controls). Primary diseases: diabetic nephropathy 74%; CGN 16%; 11% HTN |
| | C*03 | Nassar et al. (2015) | Protective against hypertensive ESRD | Yemeni | 100 (50 HTN ESRD patients, 50 controls) |
| | C*04:01 | Lowe et al. (2020) | Reduced eGFR | Black African | See above |
| | C*05:01 | Lowe et al. (2021) | Increased eGFR | British | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | C*06:02 | Pan et al. (2019) | Protective against ESRD | Chinese | See above |
| | C*07:01 | Lowe et al. (2021) | Reduced eGFR, increased CKD | British | See above |
| | | Pan et al. (2019) | Protective against ESRD | Chinese | See above |
| | C*07:02 | Lowe et al. (2021) | Reduced eGFR | British | See above |
| | C*08:02 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | C*12:03 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | C*16:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| HLA-DR | DRB1*01 | Uygun et al. (2015) | Protective against ESRD | Turkish | See above |
| | DRB1*02 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | DRB1*03 | Dai et al. (2015) | Associated with ESRD | Taiwanese | 331 (141 ESRD patients, 331 controls). Primary diseases: unknown 63%; diabetes 21%; IgA nephropathy 6%; PKD 3%; other 6% |
| | | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | | Hamdi et al. (2014) | Associated with ESRD | Saudi Arabian | See above |
| | | Karahan et al. (2009) | Associated with diabetic nephropathy | Turkish | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | | Lowe et al. (2020) | Increased eGFR | Black African | See above |
| | | Lowe et al. (2020) | Reduced eGFR | Indian | 5475 subjects (with a range of kidney function, eGFR calculated for each) |
| | | Lowe et al. (2021) | Reduced eGFR, increased ESRD, increased CKD | British | See above |
| | | Yamakawa et al. (2014) | Associated with ESRD | Brazilian | See above |
| | DRB1*04 | Chang et al. (2012) | Associated with treatment failure and decreased renal survival | Chinese | 152 patients with AAV, no controls |
| | | Cao et al. (2014) | Associated with ESRD | Cantonese | See above |
| | | Perez-Luque et al. (2000) | Protective against ESRD | Mexican | 240 (139 patients with combinations of diabetes and ESRD, 101 controls) |
| | | Lowe et al. (2021) | Increased eGFR | British | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | | Lowe et al. (2020) | Increased eGFR and protective against CKD | Irish | 12,191 subjects (with a range of kidney function, eGFR calculated for each) |
| | | Pan et al. (2019) | Associated with ESRD (DRB1*04:03, 04:04 and 04:05) | Chinese | See above |
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | | Gencik et al. (1999) | Associated with WG | Bavaria (Germany) | 102 patients with AASV, size of control group not specified |
| | DRB1*05 | Doxiadis et al. (2001) | Associated with IgA nephropathy | European | See above |
| | DRB1*07:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | DRB1*08 | Dai et al. (2015) | Protective against ESRD | Taiwanese | See above |
| | | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | | Nassar et al. (2015) | Protective against hypertensive ESRD | Yemeni | See above |
| | DRB1*09 | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | DRB1*10 | Liu et al. (2007) | Associated with renal failure | Dalian Han, China | 101 (20 patients with renal failure, primary disease not specified, 81 controls) |
| | | Hieu et al. (2019) | Protective against ESRD | Vietnamese | See above |
| | DRB1*11 | Crispim et al. (2008) | Associated with ESRD | Sao Paolo (Brazil) | See above |
| | | Dai et al. (2015) | Associated with ESRD | Taiwanese | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | | Karahan et al. (2009) | Protective against ESRD | Turkish | See above |
| | | Mosaad et al. (2014) | Protective against ESRD | Kuwaiti | See above |
| | DRB1*12 | Shao et al. (2018) | Associated with ESRD | Dalian Han, China | 14,692 (163 ESRD patients, 14,529 controls). Primary disease: CGN 76%; HTN nephrosclerosis 15%; diabetic nephropathy 4%; other 4% |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | | Noureen et al. (2020) | Protective against ESRD | Pakistani | See above |
| | DRB1*13 | Gencik et al. (1999) | Protective against AASV | Bavaria (Germany) | See above |
| | | Noureen et al. (2020) | Protective against ESRD | Pakistani | See above |
| | | Gerhardsson et al. (2015) | Associated with ESRD | Swedish | 110 (16 patients with HAPLN, 94 patients with LN) |
| | DRB1*15:01 | Pan et al. (2019) | Protective against ESRD | Chinese | See above |
| | DRB1*15:02 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | | Perez-Luque et al. (2000) | Associated with ESRD | Mexican | See above |
| | DRB1*16 | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | DRB1*17 | Hernandez-Rivera et al. (2019) | Associated with ESRD | Mexican | See above |
| | DRB3*01:01 | Lowe et al. (2021) | Reduced eGFR | British | See above |
| | DRB3 – no gene | Lowe et al. (2021) | Reduced eGFR, increased CKD | British | See above |
| | DRB4*01:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | DRB4*01:03 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | DRB4 – no gene | Lowe et al. (2021) | Increased eGFR | British | See above |
| HLA-DP | DPA1*01:03 | Lowe et al. (2020) | Increased eGFR | Black African | See above |
| | DPA1*02:01 | Lowe et al. (2020) | Reduced eGFR | Black African | See above |
| | DPA1:02:02 | Lowe et al. (2020) | Reduced eGFR | Black African | See above |
| | DPB1*04:02 | Chang et al. (2012) | Associated with increased mortality | Chinese | See above |
| HLA-DQ | DQA1*02:01 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | DQA1*03 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | | Noureen et al. (2020) | Protective against ESRD | Pakistani | See above |
| | | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | DQA1*04 | Hernandez-Rivera et al. (2019) | Protective against ESRD | Mexican | See above |
| | DQA1*05:01 | Lowe et al. (2021) | Reduced eGFR, increased CKD | British | See above |
| | DQA1*06 | Noureen et al. (2020) | Protective against ESRD | Pakistani | See above |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | DQB1*02 | Liu et al. (2007) | Associated with renal failure | Dalian Han, China | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | | Lowe et al. (2021) | Reduced eGFR, increased ESRD, increased CKD | British | See above |
| | | Lowe et al. (2020) | Reduced eGFR | Indian | See above |
| | DQB1*03:01 | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | DQB1*03:02 | Lowe et al. (2021) | Increased eGFR | British | See above |
| | | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | DQB1*03:08 | Almogren et al. (2012) | Associated with ESRD | Saudi Arabian | See above |
| | DQB1*04:01 | Pan et al. (2019) | Associated with ESRD | Chinese | See above |
| | DQB1*05:01 | Perez-Luque et al. (2000) | Associated with ESRD | Mexican | See above |
| | DQB1*06 | Lowe et al. (2021) | Increased eGFR (DQB1*06:01) | British | See above |
| | | Noureen et al. (2020) | Protective against ESRD | Pakistani | See above |
| | | Pan et al. (2019) | Protective against ESRD (DQB1*06:02 and 06:09) | Chinese | See above |
| | | Gencik et al. (1999) | Protective against AASV (DQB1*06:03) | Bavaria (Germany) | See above |
| | | Lowe et al. (2021) | Increased eGFR (DQB1*06:09) | British | See above |
| Haplotypes | A*01-B*8-DR*03 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*01-B*15-DR*04 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*02-B*05-DR*05 | Doxiadis et al. (2001) | Associated with IgA nephropathy | European | See above |
| | A*02-B*07-DR*02 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*02-B*40-DRB1*09 | Shao et al. (2018) | Associated with ESRD | Dalian Han, China | See above |
| | A*02-B*40-DRB1*12 | Shao et al. (2018) | Associated with ESRD | Dalian Han, China | See above |
| | A*02-B*44-DRB1*07 | Chowdhry et al. (2016) | Protective against ESRD | Indian | 339 (148 ESRD patients, primary disease not specified, 191 controls) |

(Continues)

**TABLE 2** (Continued)

| HLA Locus | HLA type or haplotype | Study | Effect | Population | Number of subjects |
|---|---|---|---|---|---|
| | A*03-B*07-DR*02 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*09-B*12-DR*07 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*10-B*18-DR*2 | Doxiadis et al. (2001) | Protective against IgA nephropathy | European | See above |
| | A*11-B*27-DRB1*04 | Cao et al. (2014) | Associated with ESRD | Cantonese | See above |
| | A*24-B*15-DRB1*12 | Shao et al. (2018) | Associated with ESRD | Dalian Han, China | See above |
| | A*33-B*44-DRB1*07 | Chowdhry et al. (2016) | Protective against ESRD | Indian | See above |
| | DRB1*03:01-DQA1*05:01-DQB1*02:01 | Levine et al. (2020) | Associated with PMG | European | 6588 (146 PMG patients, 6442 controls) |

Antineutrophil cytoplasmic antibodies-associated systemic vasculitis (AASV); antineutrophil cytoplasmic antibodies-associated vasculitis (AAV); chronic glomerulonephritis (CGN); chronic renal failure (CRF); histopathological antiphospholipid-associated nephropathy (HAPLN); hypertensive (HTN); Henoch-Schönlein purpura nephritis (HSPN); immunoglobulin A (IgA); polycystic kidney disease (PKD); primary membranoproliferative glomerulonephritis (PMG); panel reactive antibody (PRA); vesicoureteral reflux (VUR); Wegener's granulomatosis (WG).

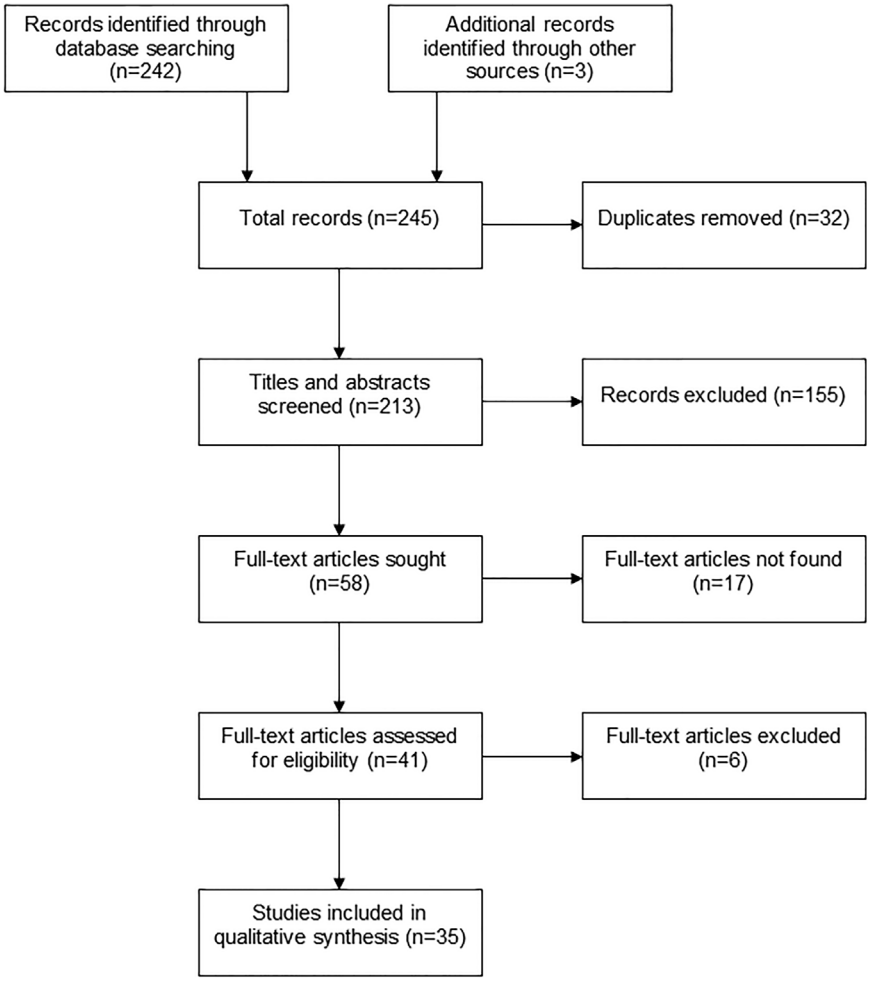



**FIGURE 1** Process of including and excluding papers

**TABLE 3** Studies which reported no significant associations between HLA type and renal function

| Study | No associations reported between | Population | Number of subjects |
|---|---|---|---|
| Demaine et al. (1982) | HLA and ESRD | Caucasian | 275 (163 ESRD patients, 112 controls). Primary diseases: non-immunological eg PKD, HTN 53%; immunological, e.g. glomerolunephritis 47% |
| Hanna et al. (2015) | HLA and ESRD | Egyptian | 94 (50 ESRD patients, 44 controls). Primary diseases: diabetic nephropathy 26%; HTN nephrosclerosis 22%; PKD 10%; lupus 8%; unknown 34% |
| Regueiro and Arnaiz-Villena (1984) | HLA and CKD | Spanish | 216 (20 CKD patients, 196 controls). Primary diseases: 15% glomerulonephritis 15%; lupus 5%; uropathy 5%; unknown 75% |
| Zachary et al. (1996) | HLA and ESRD | USA | 38,582 (20,069 ESRD patients, primary diseases not specified, 18,513 controls) |
| Spriewald et al. (2005) | HLA and WG | German | 123 (32 WG patients, 91 controls) |

errors in the findings: true HLA associations with renal function may be more likely to have been replicated in independent cohorts.

Generally, the studies in this review included a group of subjects with ESRD or some other criteria of kidney dysfunction who were compared with a control group of healthy subjects. The subjects were HLA typed by a particular method (such as PCR-SSO or PCR-SSP typing) and the frequencies of the HLA specificities in the different groups compared. HLA specificities which were found significantly more commonly in the disease group were said to confer risk of susceptibility to the disease (or be 'associated' with the disease), while those which were more commonly found in the healthy group were said to confer protection from the disease.

The majority of these studies focused on a specific ethnicity or nationality. HLA allele and haplotype frequencies vary considerably between groups of people of different ethnicities. An allele that confers a risk of (or protection from) ESRD in one population may not have the same effect for people of another population; this may explain some of the discrepant findings from different studies. Each of these studies' findings can only be applied to the ethnic population used in the study. However, such an association may be a good indicator of the same association in other groups. It is important, then, to investigate as many different cohorts of patients as possible, as results do not appear to be universal. There is also often an environmental component to disease onset, which may vary between ethnic groups. Due to the difficulty of comparing HLA associations among different ethnic groups, it was not possible to perform a meta-analysis based on the papers identified by this literature search. Similarly, there are many different primary diseases or underlying causes of CKD (NHS, 2019), and these may each have different HLA associations (or no HLA associations at all), as well as having different frequencies in different ethnic populations (Stratton et al., 2000). This could explain the contradictory associations reported: an allele could possibly lead to increased chance of one underlying disease but decreased chance of another. Depending on the composition of each study's cohort, the cumulative effect of these differences (ethnicity, primary disease, allele frequency and environmental factors) may be responsible for some of the contradictory and discrepant findings. Some of the studies in this review clarified which primary disease they were investigating; where available, the primary diseases can be seen in Table 2. Some studies focused on a particular underlying cause of renal dysfunction, while others provided the proportions of each primary disease but analysed all subjects together in a heterogeneous group, and still others did not mention the cause of renal failure. For this reason, comparing associations based on primary disease is difficult.

One problem with these studies is that the sample sizes are often small, which may lead to insufficient power to detect the true genetic effect size. For example, Liu et al. (2007) included only 20 patients and 81 controls, and other studies such as Regueiro and Arnaiz-Villena (1984), Davood et al. (2008) and Spriewald et al. (2005) all used study groups of ≤32 patients. The largest case-control study was performed by Zachary et al. (1996), who used a cohort of 20,069 ESRD patients and 18,513 controls. However, this study did not report any significant associations between HLA and ESRD. Although the paper reported the frequencies of HLA types in both the patient and control groups, it did not investigate whether there were significant differences between the two groups. It is possible that there were significant associations but they were not explicitly mentioned in the paper.

A criticism of some studies is the selection of the case and control groups. For example, in the studies by Nassar et al. (2015) and Hamdi et al. (2014), the control groups of healthy patients were made up of kidney transplantation donors; many of these may have been relatives of people with ESRD (some of whom may even have been included in the case group). The hereditary nature of HLA types means that the HLA haplotypes of living kidney donors may not be representative of a typical group of healthy people from their respective populations; they may have been genetically more similar to the subjects with ESRD. Some studies mentioned that their control group was made up of unrelated people, so this criticism does not apply.

A further drawback of a proportion of these studies is their statistical methodologies. It is not clear in a number of papers whether the researchers applied a correction for multiple tests. As each study tests a large number of HLA alleles for association with ESRD, using a value of $p < .05$ to indicate significance is likely to lead to a number of apparently significant associations which are simply due to chance. To resolve this, an adjusted $p$ value should be calculated to correct for multiple tests. Cao et al. (2014) and Mosaad et al. (2014), for example, mentioned that they used the Bonferroni formula to achieve this correction, and only reported corrected $p$ values as significant. Many studies, though, did not clarify whether they adjusted the significance levels to accommodate multiple testing. A final criticism of these studies is that they often treat renal function as a categorical variable, with subjects classed as either healthy or unhealthy depending on their GFR. In reality, there is a wide range of levels of renal function; future studies should be advised to use continuous measures such as GFR or estimated GFR (eGFR) if possible in order to obtain a more precise measurement of the subjects' kidney functions. Although categorizations based on continuous data have useful clinical applications (e.g. for diagnosis), they lead to reduced statistical power and increased risk of type I errors when used for research (Altman & Royston, 2006). For this reason, the practice is 'rarely defensible and often … misleading' (MacCallum et al., 2002) and 'unnecessary for statistical analysis' (Naggara et al., 2011). A strength of these studies is that the HLA genotyping methods used are generally very accurate. Subjects were commonly typed by PCR-SSO (Fejzic et al., 2017; Perez-Luque et al., 2000) or PCR-SSP (Cao et al., 2014; Nassar et al., 2015) methods, though some used less accurate serological methods or imputation.

A limitation of the systematic review process is that some of the process was carried out by a single researcher, so the methods were not independently validated. Additionally, only full articles were considered so it is possible that some relevant findings (such as abstracts and posters from conferences) were omitted due to not having been peer-reviewed. Initially there were an additional two concepts which were intended to be included in the review but were ultimately dropped; originally, the review intended to consider only studies of white subjects of middle age or older. Age is related to kidney function (Weinstein & Anderson, 2010) and it is known that genetic associations with

renal function vary according to age. Removing the concept of age increased the sensitivity of the search but reduced the specificity. It is possible also that there is reporting bias, which would mean that studies which did not find any associations between HLA and renal function were less likely to be published. However, five of the studies included in this review reported no associations, so it appears that a lack of findings is not a complete contraindication to publication.

## 5 | CONCLUSION

This paper has performed a systematic review of previous literature investigating associations between HLA type and renal function. A total of 245 papers were considered from a wide range of sources, and 35 were considered relevant. These papers revealed a large number of associations between renal function and HLA types. The findings strongly suggest that there is a link between HLA type and renal function, though the exact nature of this link is unclear. There is very little consensus on which HLA types are protective, or which confer risk of ESRD. Ethnicity appears to have a role in whether particular genes are associated with renal function, and the underlying cause of renal failure may complicate this further. Further research is required to confirm or refute the associations found so far, and possibly to reveal novel associations.

### ACKNOWLEDGEMENTS

Funding was provided by Kidneys for Life. We would also like to acknowledge Lesley Patterson, who provided valuable knowledge and advice on how to search the various databases.

### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

### ORCID

*Marcus Lowe* https://orcid.org/0000-0002-0799-1674
*Kay Poulton* https://orcid.org/0000-0002-5608-8107

### REFERENCES

Almogren, A., Shakoor, Z., & Hamam, K. D. (2012). Human leucocyte antigens: Their association with end-stage renal disease in Saudi patients awaiting transplantation. *British Journal of Biomedical Science*, *69*(4), 159–163.

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *Bmj (Clinical Research Ed.)*, *332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

Canadas-Garre, M., Anderson, K., Cappa, R., Skelly, R., Smyth, L. J., McKnight, A. J., & Maxwell, A. P. (2019). Genetic susceptibility to chronic kidney disease – Some more pieces for the heritability puzzle. *Frontiers in Genetics*, *10*, 453. https://doi.org/10.3389/fgene.2019.00453

Cao, Q., Xie, D., Liu, J., Zou, H., Zhang, Y., Zhang, H., Zhang, Z., Xue, H., Zhou, J., & Chen, P. (2014). HLA polymorphism and susceptibility to end-stage renal disease in Cantonese patients awaiting kidney transplantation. *Plos One*, *9*(6), e90869. https://doi.org/10.1371/journal.pone.0090869

Chang, D. Y., Luo, H., Zhou, X., Chen, M., & Zhao, M. H. (2012). Association of HLA genes with clinical outcomes of ANCA-associated vasculitis. *Clinical Journal of the American Society of Nephrology*, *7*(8), 1293–1299. https://doi.org/10.2215/CJN.13071211

Chowdhry, M., Makroo, R. N., & Kumar, M. (2016). Genetic diversity through human leukocyte antigen typing in end-stage renal disease patients and prospective donors of North India. *Indian Journal of Pathology & Microbiology*, *59*(1), 59–62. https://doi.org/10.4103/0377-4929.178221

Crispim, J. C., Mendes-Junior, C. T., Wastowski, I. J., Palomino, G. M., Saber, L. T., Rassi, D. M., & Donadi, E. A. (2008). HLA polymorphisms as incidence factor in the progression to end-stage renal disease in Brazilian patients awaiting kidney transplant. *Transplantation Proceedings*, *40*(5), 1333–1336. https://doi.org/10.1016/j.transproceed.2008.02.086

Critical Appraisal Skills Programme. (2018). CASP case control study checklist. Retrieved from https://casp-uk.net/wp-content/uploads/2018/03/CASP-Case-Control-Study-Checklist-2018_fillable_form.pdf

Dai, C. S., Chu, C. C., Chen, S. F., Sun, C. Y., Lin, M., & Lee, C. C. (2015). Association between human leucocyte antigen subtypes and risk of end stage renal disease in Taiwanese: A retrospective study. *BMC Nephrology*, *16*(1), 177. https://doi.org/10.1186/s12882-015-0165-7

Davood, P. P., Farhadi, N., & Najafizadeh, M. (2008). Protective and susceptible HLA class I genes in patients with end-stage renal disease. *Research Journal of Biological Sciences*, *3*(11), 1344–1346.

Demaine, A. G., Vaughan, R. W., & Welsh, K. I. (1982). Immunoglobulin (Gm) allotype frequencies in patients with renal failure. *Transplantation*, *34*(3), 153–154.

Doxiadis, I. I., De Lange, P., De Vries, E., Persijn, G. G., & Claas, F. H. (2001). Protective and susceptible HLA polymorphisms in IgA nephropathy patients with end-stage renal failure. *Tissue Antigens*, *57*(4), 344–347.

Fejzic, E., Karamehic, J., Eminovi, I., Suljevic, D., Alijagic, A., Becirevic, S., Šahović, A., & Sisic, S. (2017). HLA genotyping in patients with end-stage renal disease waiting for cadaveric renal transplantation in federation of Bosnia and Herzegovina. *Macedonian Journal of Medical Sciences*, *5*(1), 1–5. https://doi.org/10.3889/oamjms.2017.015

Gencik, M., Borgmann, S., Zahn, R., Albert, E., Sitter, T., Epplen, J. T., & Fricke, H. (1999). Immunogenetic risk factors for anti-neutrophil cytoplasmic antibody (ANCA)-associated systemic vasculitis. *Clinical and Experimental Immunology*, *117*(2), 412–417. https://doi.org/10.1046/j.1365-2249.1999.00969.x

Gerhardsson, J., Sundelin, B., Zickert, A., Padyukov, L., Svenungsson, E., & Gunnarsson, I. (2015). Histological antiphospholipid-associated nephropathy versus lupus nephritis in patients with systemic lupus erythematosus: An observational cross-sectional study with longitudinal follow-up. *Arthritis Research and Therapy*, *17*(1), 109. https://doi.org/10.1186/s13075-015-0614-5

Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E., & Hsu, C. Y. (2004). Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *New England Journal of Medicine*, *351*(13), 1296–1305. https://doi.org/10.1056/NEJMoa041031

Hamdi, N. M., Al-Hababi, F. H., & Eid, A. E. (2014). HLA class I and class II associations with ESRD in Saudi Arabian population. *Plos One*, *9*(11), e111403. https://doi.org/10.1371/journal.pone.0111403

Hanna, M. O. F., Shahin, R. M. H., Meshaal, S. S., & Kostandi, I. F. (2015). Susceptibility and progression of end stage renal disease are not associated with angiotensin II type 1 receptor gene polymorphism. *Journal of Receptors and Signal Transduction*, *35*(5), 381–385. https://doi.org/10.3109/10799893.2014.956757

Health Service Journal. (2020). Tackling the £1.5bn a year cost of chronic kidney disease. Retrieved from https://www.hsj.co.uk/the-hsj-awards/tackling-the-15bn-a-year-cost-of-chronic-kidney-disease/7027154.article

Hernandez-Rivera, J. C. H., Salazar-Mendoza, M., Perez-Lopez, M. J., Gonzalez-Ramos, J., Espinoza-Perez, R., Martinez-Alvarez, J. C., Trejo-Villeda, M., & Paniagua-Sierra, R. (2019). Alelos HLA mas comunes y asociados con riesgo o proteccion en enfermedad renal cronica de etiologia no determinada. *Gaceta Medica de Mexico*, *155*(3), 243–248.

Hieu, H. T., Ha, N. T., Song, L. H., & Nghi, T. H. (2019). Association of human leukocyte antigen haplotypes with end-stage renal disease in Vietnamese patients prior to first transplantation. *Transplantation Proceedings*, *51*(8), 2549–2554. https://doi.org/10.1016/j.transproceed.2019.03.067

Huang, X., Lin, J., & Demner-Fushman, D. (2006). Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annual Symposium Proceedings [Electronic Resource]*, 2006, 359–363.

Karahan, G. E., Seyhun, Y., Oguz, F. S., Kekik, C., Onal, A. E., Yazici, H., Turkmen, A., Aydin, A. E., Sever, M. S., Eldegez, U., & Carin, M. N. (2009). Impact of HLA on the underlying primary diseases in Turkish patients with end-stage renal disease. *Renal Failure*, *31*(1), 44–49. https://doi.org/10.1080/08860220802546297

Levine, A. P., Chan, M. M. Y., Sadeghi-Alavijeh, O., Wong, E. K. S., Cook, H. T., Ashford, S., Carss, K., Christian, M. T., Hall, M., Harris, C. L., Mcalinden, P., Marchbank, K. J., Marks, S. D., Maxwell, H., Megy, K., Penkett, C. J., Mozere, M., Stirrups, K. E., Tuna, S., … Gale, D. P. (2020). Large-scale whole-genome sequencing reveals the genetic architecture of primary membranoproliferative GN and C3 glomerulopathy. *Journal of the American Society of Nephrology*, *31*(2), 365–373. https://doi.org/10.1681/ASN.2019040433

Liu, H., Wang, B., Liu, D., Cheng, Y., Yu, W., Liang, X., & An, W. (2007). Molecular genetic studies on relationships among longevity, diseases, and HLA-DRB1/DQB1 allelic polymorphism. *Experimental Aging Research*, *33*(2), 123–125.

Lowe, M., Payton, A., Verma, A., Gemmell, I., Worthington, J., Hamilton, P., Ollier, W., Augustine, T., & Poulton, K. (2020). Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom. *HLA*, *96*(6), 697–708. https://doi.org/10.1111/tan.14078

Lowe, M., Payton, A., Verma, A., Worthington, J., Gemmell, I., Hamilton, P., Ollier, W., Augustine, T., & Poulton, K. (2021). Associations between human leukocyte antigens and renal function. *Scientific Reports*, *11*(1), 3158. https://doi.org/10.1038/s41598-021-82361-7

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40. https://doi.org/10.1037/1082-989x.7.1.19

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Bmj (Clinical Research Ed.)*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Mortality, G. B. D., & Causes of Death, C. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, *388*(10053), 1459–1544. https://doi.org/10.1016/S0140-6736(16)31012-1

Mosaad, Y. M., Mansour, M., Al-Muzairai, I., Al-Otabi, T., Abdul-Moneam, M., Al-Attiyah, R., & Shahin, M. (2014). Association between Human Leukocyte Antigens (HLA-A, -B, and -DR) and end-stage renal disease in Kuwaiti patients awaiting transplantation. *Renal Failure*, *36*(8), 1317–1321. https://doi.org/10.3109/0886022x.2014.937672

Naggara, O., Raymond, J., Guilbert, F., Roy, D., Weill, A., & Altman, D. G. (2011). Analysis by categorizing or dichotomizing continuous variables is inadvisable: An example from the natural history of unruptured aneurysms. *American Journal of Neuroradiology*, *32*(3), 437–440. https://doi.org/10.3174/ajnr.A2425

Nassar, M. Y., Al-Shamahy, H. A., & Masood, H. A. (2015). The association between human leukocyte antigens and hypertensive end-stage renal failure among Yemeni patients. *Sultan Qaboos University Medical Journal*, *15*(2), e241–e249.

Nassar, M. Y., Al-Shamahy, H. A., Al-Samawi, A. S., Abu Asba, N. W., El-Nono, I. H., & Masood, H. A. (2017). Human leukocyte antigen class I and II variants in Yemeni patients with chronic renal failure. *Iranian Journal of Immunology: IJI*, *14*(3), 240–249.

NHS. (2012). *Chronic kidney disease in England: The human and financial cost*. Retrieved from https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Chronic-Kidney-Disease-in-England-The-Human-and-Financial-Cost.pdf

NHS. (2019). Chronic kidney disease. Retrieved from https://www.nhs.uk/conditions/kidney-disease/

Noureen, N., Shah, F. A., Lisec, J., Usman, H., Khalid, M., Munir, R., & Zaidi, N. (2020). Revisiting the association between human leukocyte antigen and end-stage renal disease. *PLoS ONE [Electronic Resource]*, *15*(9), e0238878.

Pan, Q., Ma, X., Chen, H., Fan, S., Wang, X., You, Y., Wang, Z., Sun, L., Zhao, X., Shen, J., Tan, R., Gong, G., & Gu, M. (2019). A single center study of protective and susceptible HLA alleles and haplotypes with end-stage renal disease in China. *Human Immunology*, *80*(11), 943–947. https://doi.org/10.1016/j.humimm.2019.09.001

Pérez-Luque, E., Malacara, J. M., Olivo-Dıáz, A., Aláez, C., Debaz, H., Vázquez-Garcia, M., Garay, M. E., Nava, L. E., Burguete, A., & Gorodezky, C. (2000). Contribution of HLA class II genes to end stage renal disease in Mexican patients with type 2 diabetes mellitus. *Human Immunology*, *61*(10), 1031–1038. https://doi.org/10.1016/S0198-8859(00)00174-9

Prakash, S., Alam, S., Sharma, R. K., Sonawane, A., Imran, M., & Agrawal, S. (2013). Distribution of Killer cell immunoglobulin like receptor genes in end stage renal disease among North Indian population. *Human Immunology*, *74*(10), 1399-1345. https://doi.org/10.1016/j.humimm.2013.06.015

Regueiro, J. R., & Arnaiz-Villena, A. (1984). C3 polymorphism, HLA and chronic renal failure in Spaniards. *Human Genetics*, *67*(4), 437–440.

Rivera P, S., Márquez, G., Cipriani, A. M., Hassanhi, M., Villalobos, C. C., Fuenmayor, A., Chavez, M., & Montiel, M. (2012). HLA class i association with progression to end-stage renal disease in patients from Zulia, Venezuela. *Inmunología*, *31*(2), 37–42. https://doi.org/10.1016/j.inmuno.2011.12.001

Robson, K. J., Ooi, J. D., Holdsworth, S. R., Rossjohn, J., & Kitching, A. R. (2018). HLA and kidney disease: From associations to mechanisms. *Nature Reviews Nephrology*, *14*(10), 636–655. https://doi.org/10.1038/s41581-018-0057-8

Shao, L.-N., Yang, Y., Zhang, S.-T., Zhou, S.-H., Duan, Y., Meng, Q.-L.i, Liang, X.-H., & Yu, W.-J. (2018). Association between the polymorphism of HLA and ESRD in Dalian Han population located in north of China. *Immunological Investigations*, *47*(2), 212–219. https://doi.org/10.1080/08820139.2017.1416397

Shi, X., Han, W., & Ding, J. (2017). The impact of human leukocyte antigen mismatching on graft survival and mortality in adult renal transplantation: A protocol for a systematic review and meta-analysis. *Medicine*, *96*(49), e8899. https://doi.org/10.1097/MD.0000000000008899

Spriewald, B. M., Witzke, O., Wassmuth, R., Wenzel, R. R., Arnold, M. L., Philipp, T., & Kalden, J. R. (2005). Distinct tumour necrosis factor alpha, interferon gamma, interleukin 10, and cytotoxic T cell antigen 4 gene polymorphisms in disease occurrence and end stage renal disease in Wegener's granulomatosis. *Annals of the Rheumatic Diseases*, *64*(3), 457–461. https://doi.org/10.1136/ard.2004.025809

Stratton, I. M., Adler, A. I., Neil, H. A., Matthews, D. R., Manley, S. E., Cull, C. A., Hadden, D., Turner, R. C., & Holman, R. R. (2000). Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): Prospective observational study. *Bmj (Clinical Research Ed.)*, *321*(7258), 405–412. https://doi.org/10.1136/bmj.321.7258.405

The Cochrane Collaboration. (2019). Glossary of kidney disease terms. Retrieved from https://kidneyandtransplant.cochrane.org/cochrane-renal-group-glossary

Tonelli, M., Wiebe, N., Culleton, B., House, A., Rabbat, C., Fok, M., Mcalister, F., & Garg, A. X. (2006). Chronic kidney disease and mortality risk: A systematic review. *Journal of the American Society of Nephrology*, *17*(7), 2034–2047. https://doi.org/10.1681/ASN.2005101085

Uygun, D. F. K., Ozbay, M., Sallakci, N., Kilinc, Y., Filiz, S., & Yegin, O. (2015). Human leukocyte antigen frequencies in highly sensitized patients. *Turkish Journal of Immunology*, *3*(1), 11–14. https://doi.org/10.5606/tji.2015.333

Weiner, D. E., Tighiouart, H., Amin, M. G., Stark, P. C., MacLeod, B., Griffith, J. L., Salem, D. N., Levey, A. S., & Sarnak, M. J. (2004). Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: A pooled analysis of community-based studies. *Journal of the American Society of Nephrology*, *15*(5), 1307–1315. https://doi.org/10.1097/01.ASN.0000123691.46138.E2

Weinstein, J. R., & Anderson, S. (2010). The aging kidney: Physiological changes. *Advances in Chronic Kidney Disease*, *17*(4), 302–307. https://doi.org/10.1053/j.ackd.2010.05.002

Xu, X., Eales, J. M., Akbarov, A., Guo, H., Becker, L., Talavera, D., Ashraf, F., Nawaz, J., Pramanik, S., Bowes, J., Jiang, X., Dormer, J., Denniff, M., Antczak, A., Szulinska, M., Wise, I., Prestes, P. R., Glyda, M., Bogdanski, P., … Tomaszewski, M. (2018). Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nature Communications*, *9*(1), 4800. https://doi.org/10.1038/s41467-018-07260-4

Yamakawa, R. H., Saito, P. K., da Silva Junior, W. V., de Mattos, L. C., & Borelli, S. D. (2014). Polymorphism of leukocyte and erythrocyte antigens in chronic kidney disease patients in southern Brazil. *Plos One*, *9*(1), e84456. https://doi.org/10.1371/journal.pone.0084456

Zachary, A. A., Steinberg, A. G., Bias, W. B., & Leffell, M. S. (1996). The frequencies of HLA alleles and haplotypes and their distribution among donors and renal patients in the UNOS registry. *Transplantation*, *62*(2), 272–283.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Lowe, M., Jervis, S., Payton, A., Poulton, K., Worthington, J., Gemmell, I., & Verma, A. (2021). Systematic review of associations between HLA and renal function. *International Journal of Immunogenetics*, 1–17. https://doi.org/10.1111/iji.12566

# 11 Paper 2: Associations between human leukocyte antigens and renal function

This paper was published in Scientific Reports on 4th February 2021.

scientific reports

Check for updates

OPEN

# Associations between human leukocyte antigens and renal function

Marcus Lowe[1,3]✉, Antony Payton[2], Arpana Verma[3], Judith Worthington[1], Isla Gemmell[3], Patrick Hamilton[4], William Ollier[3,5], Titus Augustine[4,6] & Kay Poulton[1,3]

**Human leukocyte antigens (HLA) have been associated with renal function, but previous studies report contradictory findings with little consensus on the exact nature or impact of this observation. This study included 401,307 white British subjects aged 39–73 when they were recruited by UK Biobank. Subjects' HLA types were imputed using HLA*IMP:02 software. Regression analysis was used to compare 362 imputed HLA types with estimated glomerular filtration rate (eGFR) as a primary outcome and clinical indications as secondary outcome measures. 22 imputed HLA types were associated with increased eGFR (and therefore increased renal function). Decreased eGFR (decreased renal function) was associated with 11 imputed HLA types, seven of which were also associated with increased risk of end-stage renal disease and/or chronic kidney disease. Many of these HLA types are commonly inherited together in established haplotypes, for example: HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DQB1*02:01. This haplotype has a population frequency of 9.5% in England and each allele was associated with decreased renal function. 33 imputed HLA types were associated with kidney function in white British subjects. Linkage disequilibrium in HLA heritance suggests that this is not random and particularly affects carriers of established haplotypes. This could have important applications for the diagnosis and treatment of renal disease and global population health.**

Approximately 1.2 million people died due to chronic kidney disease (CKD) worldwide in 2015[1], representing an increase of 32% since 2005. CKD is now ranked 17th in the list of diseases which cause the most "years of lost life", rising from 21st in 2005 and 25th in 1990[1]. 2.6 million people received dialysis in 2010, and treatment of end-stage renal disease (ESRD) accounts for 2–3% of the healthcare budgets of high-income countries[2]. Understanding the genetic influences which predispose people to kidney dysfunction will have important applications for the diagnosis and treatment of a globally significant disease.

A number of human leukocyte antigens (HLA) encoded within the major histocompatibility complex are associated with increased or decreased risk of renal failure[3]. For example, HLA-B*51 was associated with ESRD in Venezuelan[4] and Brazilian[5] subjects, while A*26 was protective against ESRD in Saudi Arabia[6] and Turkey[7]. Approximately 100 different HLA types have been linked to renal function by various studies worldwide[8], including studies of European or white populations[9–13]. These were mostly reported in case–control studies comprising fewer than 500 subjects of a single nationality, and many associations remain unreplicated or have been contradicted by other studies. Despite mounting evidence that an association between HLA and renal function exists, there is no supportive confirmation in sufficiently-powered studies. We interrogated a large cohort of white British subjects to test the hypothesis that the HLA region is associated with renal function.

[1]Transplantation Laboratory, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK. [2]Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [3]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK. [4]Department of Renal and Pancreatic Transplantation, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK. [5]Centre for Bioscience, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, UK. [6]Division of Diabetes, Endocrinology and Gastroenterology, University of Manchester, Manchester, UK. ✉email: marcus.lowe@mft.nhs.uk

## Methods

**Study population and quality control.** This is a UK Biobank (UKB) retrospective cohort study using data from 502,616 subjects aged 39–73 years at the time of recruitment between 2006 and 2010[14]. 88% of the cohort self-identifies as "white British", and principal component analysis conducted by UKB concluded that 82% of the UKB cohort is white British[15]. Analysis was restricted to this group to reduce population stratification; 92,858 subjects who were not white British were analysed separately[16].

Individuals within the cohort whom UKB deemed to be related[17] (kinship coefficient ≥ 0.044) were also excluded (n = 7318) to avoid HLA frequency bias[18]. Where subjects were related, the individual with the most complete set of genetic data, based on a set of "high-quality markers"[17], was included. Genetic sex influences kidney function[19], so only individuals whose sex could be clearly assigned were included. Subjects identified by UKB to have sex chromosome karyotypes other than XX or XY[20] and those whose genetic sex, as calculated by UKB, did not match their self-reported sex[21] were removed (n = 786 in total). Finally, 347 subjects were excluded at UKB's recommendation due to missing genetic data[22]. A total of 101,309 subjects were excluded during quality control, leaving 401,307 subjects for analysis. All quality control was performed using Stata/SE 13.0 (StataCorp).

**HLA typing.** Imputation estimates a person's most likely HLA type based on the presence of particular single nucleotide polymorphisms[23]. HLA types were imputed for each subject by UKB using HLA*IMP:02 software[24] at the following loci: HLA-A, B, C (Class I) and DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4, DRB5 (Class II)[25] at a level equivalent to high resolution typing using eight reference datasets[26]. 362 HLA types were imputed. Two of these (HLA-DQB1*02:02 and DPB1*03:01) were not in Hardy–Weinberg equilibrium (HWE, $P < 0.00014$) so were excluded from this study; the remaining 360 alleles were included. Table 1 shows the 100 HLA types with frequency > 1% in the cohort.

**Measures of renal function.** Renal function was determined using estimated glomerular filtration rate (eGFR), a measure of toxin filtration calculated using serum biomarkers such as creatinine and cystatin[27]. High levels of these biomarkers are indicative of poor renal function and manifest in a lower eGFR. This study used the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) eGFR calculation which adjusted for age and sex[28]. Three eGFR values were calculated for each subject, using measures of: creatinine; cystatin; and both creatinine and cystatin[29]. Pairwise correlation confirmed that the three eGFR values were similar (Pearson's correlation coefficients > 0.6; $P < 0.0001$). The cystatin-based eGFR value provided the most complete dataset; only this value was used for analysis to avoid repetition of testing using closely correlated variables.

Clinical histories for each subject were used as secondary outcomes. Subjects with kidney dysfunction were identified by examining self-reported questionnaires in addition to data relating to clinical diagnosis and procedures undertaken. These were deduced using a combination of the International Classification of Diseases[30] (ICD)-9 and -10, Office of Population Censuses and Surveys (OPCS)[31]-3 and -4, and UKB's own coding systems[32,33]. Subjects were categorised as: ESRD patients (yes or no); kidney transplant recipients (yes or no); dependent on renal replacement therapy (RRT) including transplantation at any point (yes or no); and CKD patients of any stage (yes or no) (see Table 6).

**Statistical analysis.** Linear regression analysis was performed to test for associations between HLA alleles and eGFR as a continuous variable. All 360 HLA alleles which were in HWE were included, with a Bonferroni threshold of $P < 0.00014$ considered significant[34] (0.05/360). Subjects who had ever received RRT were excluded as their eGFR values may have suggested healthy renal function even though their native function was poor.

Logistic regression was used to test for associations between HLA types and adverse clinical outcomes (ESRD, RRT, CKD, and kidney transplantation; binary variables). Age at recruitment and sex were included as covariates, and only alleles in HWE with minor allele frequency > 5% were considered (n = 50) in order to increase statistical power. $P < 0.001$ was considered significant after Bonferroni correction. All regression analysis was performed using Plink software[35].

**Ethical approval.** All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by UKB's Research Ethics Committee. Informed consent was obtained for all subjects. UKB has obtained Research Tissue Bank approval from its ethics committee that covers the majority of proposed uses of the resource, so researchers do not typically need to obtain separate ethics approval.

## Results

**Variation in renal function.** Variation in renal function within the UKB cohort is outlined in Table 2. eGFR values could not be calculated for around 18,000 subjects (< 5%) due to missing creatinine and/or cystatin measurements. Subjects dependent on RRT were excluded from the analysis, although their eGFR values are listed in Table 2, which shows calculated eGFR values and the corresponding CKD stages[36] as well as the number of subjects in the final analysis.

The calculated eGFR values were compared to average values to check that they were plausible. Average eGFR for different age categories were taken from the National Kidney Foundation[37]. The values calculated by this study were in line with NKF's estimates, as shown in Table 3. This increases confidence in the calculated values.

11,379 subjects were identified with ESRD. Of these, 437 were renal transplant recipients and 1412 subjects had RRT. 4794 subjects had a clinical diagnosis of CKD (36 stage 1, 300 stage 2, 3557 stage 3, 397 stage 4, and 504 stage 5).

| Class I | | Class II | |
|---|---|---|---|
| HLA– | Frequency (%) | HLA– | Frequency (%) |
| A*02:01 | 27.44 | DPB1*04:01 | 43.55 |
| A*01:01 | 19.42 | DRB4—no gene | 35.02 |
| C*07:01 | 17.64 | DRB3—no gene | 34.33 |
| C*07:02 | 15.82 | DRB4*01:03 | 25.35 |
| B*07:02 | 14.82 | DQA1*05:01 | 23.00 |
| A*03:01 | 14.48 | DQA1*03:01 | 20.35 |
| B*08:01 | 14.44 | DQA1*01:02 | 19.09 |
| C*05:01 | 11.38 | DPA1*01:03 | 18.44 |
| B*44:02 | 11.20 | DQB1*03:01 | 17.53 |
| C*06:02 | 8.97 | DRB3*01:01 | 16.74 |
| C*04:01 | 8.26 | DRB5—no gene | 15.27 |
| C*03:04 | 7.97 | DQB1*02:01 | 14.94 |
| A*24:02 | 7.25 | DRB1*03:01 | 14.87 |
| B*15:01 | 6.11 | DPA1*02:01 | 14.60 |
| A*11:01 | 6.04 | DQA1*02:01 | 14.58 |
| B*44:03 | 5.85 | DRB1*07:01 | 14.56 |
| B*40:01 | 5.56 | DRB1*15:01 | 14.47 |
| C*03:03 | 5.54 | DRB5*01:01 | 14.44 |
| B*35:01 | 4.52 | DQB1*06:02 | 14.19 |
| C*16:01 | 4.43 | DQA1*01:01 | 14.14 |
| A*29:02 | 4.24 | DRB3*02:02 | 13.24 |
| B*27:05 | 3.95 | DQB1*05:01 | 12.00 |
| B*57:01 | 3.92 | DRB1*04:01 | 11.14 |
| C*08:02 | 3.62 | DPB1*02:01 | 10.76 |
| B*51:01 | 3.60 | DPB1*04:02 | 10.69 |
| B*18:01 | 3.59 | DQB1*03:02 | 10.37 |
| C*02:02 | 3.59 | DRB1*01:01 | 9.33 |
| A*32:01 | 3.47 | DQB1*02:02a | 9.27 |
| C*01:02 | 3.41 | DPB1*03:01a | 9.11 |
| A*68:01 | 3.01 | DRB4*01:01 | 8.06 |
| C*12:03 | 2.82 | DPB1*01:01 | 5.92 |
| A*31:01 | 2.68 | DQA1*01:03 | 5.44 |
| B*14:02 | 2.52 | DQB1*03:03 | 5.23 |
| B*13:02 | 1.93 | DQB1*06:03 | 5.06 |
| A*26:01 | 1.85 | DRB1*13:01 | 5.00 |
| B*55:01 | 1.83 | DRB1*04:04 | 3.98 |
| C*07:04 | 1.76 | DRB1*13:02 | 3.79 |
| C*15:02 | 1.73 | DRB3*03:01 | 3.77 |
| A*23:01 | 1.70 | DRB1*11:01 | 3.19 |
| A*25:01 | 1.62 | DPA1*02:02 | 3.10 |
| B*37:01 | 1.33 | DQB1*06:04 | 2.74 |
| B*14:01 | 1.16 | DPB1*11:01 | 2.60 |
| B*49:01 | 1.15 | DQB1*05:03 | 2.17 |
| A*30:01 | 1.01 | DPB1*05:01 | 2.07 |
| | | DQB1*04:02 | 2.00 |
| | | DQA1*04:01 | 1.95 |
| | | DRB1*14:01 | 1.91 |
| | | DRB1*08:01 | 1.79 |
| | | DRB1*01:03 | 1.69 |
| | | DPB1*10:01 | 1.61 |
| | | DPB1*13:01 | 1.49 |
| | | DRB1*12:01 | 1.42 |
| | | DRB1*09:01 | 1.31 |
| | | DPB1*17:01 | 1.11 |
| | | DRB1*11:04 | 1.04 |
| | | DQB1*06:09 | 1.02 |

**Table 1.** HLA types with frequency > 1%. This shows the 44 Class I and 56 Class II types which have frequency > 1% in the cohort. They are split into Class I and Class II and sorted in descending order of frequency. [a]Not in Hardy–Weinberg equilibrium so excluded from analysis.

| Subjects with eGFR | CKD stage | Kidney function determined by | | |
|---|---|---|---|---|
| | | Creatinine | Cystatin | Both |
| 0–14.9 | 5 (ESRD requiring RRT) | 95 | 91 | 92 |
| 15–29.9 | 4 | 277 | 538 | 374 |
| 30–44.9 | 3b | 1147 | 2611 | 1436 |
| 45–59.9 | 3a | 7164 | 14,777 | 7575 |
| 60–89.9 | 2 (when combined with other evidence of kidney damage) | 148,885 | 183,525 | 172,037 |
| 90–119.9 | 1 (when combined with other evidence of kidney damage) | 225,142 | 180,139 | 199,325 |
| ≥ 120 | | 694 | 1877 | 2287 |
| Data missing | | 17,903 | 17,749 | 18,181 |
| Excluded due to RRT | | 1357 | 1354 | 1351 |
| Included in eGFR analysis | | 382,047 | 382,204 | 381,775 |

**Table 2.** Distribution of subjects' eGFR. This shows the distribution of subjects' eGFRs for each of the three eGFR formulae, as well as the number of subjects included and excluded. eGFR thresholds are based on thresholds used for CKD diagnosis. Some subjects were dependent on RRT so were not included in analysis of eGFR, though their eGFRs are included in this table.

| Age | Average eGFR according to | | | |
|---|---|---|---|---|
| | National Kidney Foundation | This study (creatinine) | This study (cystatin) | This study (both) |
| 40 s | 99 | 99.6 | 100.0 | 100.7 |
| 50 s | 93 | 92.4 | 90.2 | 92.1 |
| 60 s | 85 | 85.0 | 80.5 | 83.5 |

**Table 3.** eGFR values by age. This shows average eGFR values by decade of life for the general population (provided by National Kidney Foundation) and for the three eGFR formulae used in this study.

**Regression analysis.** 33 HLA types were significantly associated with renal function after correction for multiple testing. Table 4 lists the 11 HLA alleles linked with decreased renal function (defined by either decreased eGFR or the presence of CKD or ESRD). Table 5 shows the 22 HLA alleles associated with increased renal function. No HLA associations were identified with kidney transplant status or RRT status. Tables 4 and 5 also show the population frequency of the alleles, the beta value or odds ratio (OR) of each effect, and the level of significance of the associations.

**Associations with decreased renal function.** HLA types are inherited in maternal and paternal haplotypes and are not randomly distributed. Of the 11 HLA associations with decreased eGFR, seven were also linked to development of CKD, ESRD or both. 10 of these 11 HLA alleles are inherited in two well-documented haplotypes: (HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DRB3*01:01, DQA1*05:01, DQB1*02:01; and A*03:01, B*07:02, C*07:02[38]. All genes in the former are associated with decreased eGFR, and all but HLA-DRB3*01:01 are also linked to increased risk of CKD, ESRD, or both. This haplotype is seen in 9.5% of the English population[39]. The "absence of DRB3 genes" was also associated with decreased eGFR, which may either indicate increased susceptibility in subjects homozygous for this common haplotype, or may reflect individuals with the latter haplotype, which is in linkage disequilibrium (LD) with HLA-DRB1*15:01 and therefore has no associated DRB3 genes present. Alternatively, a closely linked haplotype, HLA-A*03:01, B*07:02, C*07:02, DRB1*03:01, DQB1*02:01 (present in 0.5% of the English population[39]) may be implicated here.

**Associations with increased renal function.** The HLA associations with increased eGFR values do not appear to belong to full length haplotypes, but can be separated into groups of two or three HLA alleles which are often co-inherited. For example: HLA-DRB1*04:01, DQA1*03:01, DQB1*03:02 (seen in 8.2% of the English population[40]); DRB1*07:01, DQA1*02:01 (10.5% of the English population[40]); A*29:02, B*14:02, C*08:02 (2.1% of the Northern Irish population[40]); and B*44:03, C*16:01 (4.7% of the Northern Irish population[40], also commonly associated with A*25:01) were all linked to increased eGFR. None of the 22 alleles associated with increased eGFR was shown to reduce the risk of adverse renal-related clinical outcomes.

| HLA- | Frequency (%) | eGFR based on cystatin | | CKD | | ESRD | |
|---|---|---|---|---|---|---|---|
| | | Beta | P | OR | P | OR | P |
| A*01:01 | 19.4 | −0.26 | 9.98E−09 | | | 1.062 | 0.00033 |
| A*03:01 | 14.5 | −0.22 | 2.65E−05 | | | | |
| B*07:02 | 14.8 | −0.20 | 8.84E−05 | | | | |
| B*08:01 | 14.4 | −0.48 | 3.26E−20 | 1.109 | 0.00027 | 1.066 | 0.00072 |
| C*07:01 | 17.6 | −0.35 | 1.68E−13 | 1.098 | 0.00038 | | |
| C*07:02 | 15.8 | −0.21 | 2.35E−05 | | | | |
| DQA1*05:01 | 23.0 | −0.38 | 2.03E−18 | 1.105 | 3.30E−05 | | |
| DQB1*02:01 | 14.9 | −0.49 | 7.06E−22 | 1.121 | 4.43E−05 | 1.084 | 1.44E−05 |
| DRB1*03:01 | 14.9 | −0.49 | 1.14E−21 | 1.122 | 3.61E−05 | 1.077 | 6.33E−05 |
| DRB3—no gene | 34.3 | −0.31 | 6.09E−16 | 1.090 | 6.35E−05 | | |
| DRB3*01:01 | 16.7 | −0.42 | 7.57E−18 | | | | |

**Table 4.** Alleles which are associated with decreased kidney function. This shows the alleles significantly associated with decreased kidney function as well as the frequency of the alleles, the beta value or odds ratio (OR) of each effect, and the P values.

| HLA- | Frequency (%) | eGFR based on cystatin | |
|---|---|---|---|
| | | Beta | P |
| A*25:01 | 1.6 | 0.77 | 1.07E−07 |
| A*29:02 | 4.2 | 0.39 | 1.37E−05 |
| A*32:01 | 3.5 | 0.38 | 0.00014 |
| B*14:01 | 1.2 | 0.67 | 8.62E−05 |
| B*14:02 | 2.5 | 0.64 | 3.43E−08 |
| B*44:03 | 5.9 | 0.34 | 1.31E−05 |
| C*12:03 | 2.8 | 0.63 | 7.99E−09 |
| C*16:01 | 4.4 | 0.40 | 5.79E−06 |
| C*02:02 | 3.6 | 0.39 | 8.64E−05 |
| C*05:01 | 11.4 | 0.23 | 4.57E−05 |
| C*08:02 | 3.6 | 0.63 | 1.04E−10 |
| DQA1*02:01 | 14.6 | 0.24 | 4.23E−06 |
| DQA1*03:01 | 20.4 | 0.31 | 8.12E−12 |
| DQB1*03:02 | 10.4 | 0.29 | 1.60E−06 |
| DQB1*06:01 | 0.4 | 1.15 | 4.63E−05 |
| DQB1*06:09 | 1.0 | 0.72 | 7.39E−05 |
| DRB1*15:02 | 0.4 | 1.10 | 9.85E−05 |
| DRB1*04:01 | 11.1 | 0.30 | 3.23E−07 |
| DRB1*07:01 | 14.6 | 0.22 | 1.46E−05 |
| DRB4—no gene | 35.0 | 0.33 | 5.64E−18 |
| DRB4*01:01 | 8.1 | 0.38 | 1.43E−08 |
| DRB4*01:03 | 25.4 | 0.23 | 5.85E−08 |

**Table 5.** Alleles which are associated with increased kidney function. This shows the alleles significantly associated with increased kidney function as well as the frequency of the alleles, the beta value of each effect, and the P values.

## Discussion

We identified significant HLA associations with renal function in the largest reported study to date. 22 HLA alleles were associated with increased renal function and 11 with decreased function. The HLA associations with increased renal function did not suggest a protective effect against CKD or ESRD, but the 11 associations with decreased renal function (seven of which were also linked to ESRD and/or CKD) were of particular interest. HLA genes are inherited through maternal and paternal haplotypes, which suggests a high probability that these alleles are not independently associated with renal function, but rather that this observation is non-random within the population. Specifically, individuals who carry the haplotypes listed are at increased risk of developing renal dysfunction, and may carry sub-clinical levels of impairment even in the absence of identifiable disease. This clustering of the HLA genes within well-documented haplotypes adds validity, which is reinforced as the

primary and secondary outcome measures were calculated using the independent phenotypes of biomarkers and clinical outcomes. It should be noted that some significant alleles appear to be alone in significance (that is, the alleles that they are in LD with were not significant). Examples include HLA-A*32:01 and B*14:01, among others. In these cases, it is possible that the allele itself is linked to kidney function, independent of its haplotype, or it is possible that the other alleles in LD with this allele are also significant, and this study failed to detect this. The CKD-EPI calculation of eGFR was selected rather than MDRD[41] or Cockcroft-Gault[42] due to its increased accuracy when assessing subjects with normal renal function (eGFR > 60)[43]. Using only one eGFR value avoided multiple testing of closely related variables; the formula based on cystatin was selected as it had the fewest missing values. For comparison, the two other CKD-EPI eGFR formulae (one based on creatinine, and another based on both creatinine and cystatin) were used and the data re-analysed. In addition to the associations already described, three additional associations were identified as significant (assuming the same Bonferroni threshold of $P < 0.00014$): HLA-A*23:01 and DRB3*02:02 were linked to decreased renal function, and B*27:05 was linked to increased function.

**Comparison with previous research.** Previous literature has reported conflicting HLA associations with renal function in populations of different ethnic origin. Potentially, these contradictory findings may include false positives arising from inadequate statistical power, multiple testing, publication bias or methodological differences. Alternatively, it is possible that HLA associations with kidney function differ between populations due to varied heritage. Limiting this study to only white British subjects reduced any likelihood of bias due to population stratification.

Almost 100 HLA associations with ESRD have been described, only 11 of which have been confirmed by two or more independent studies. Our study replicated one of these 11 observations but refuted two. HLA-DRB1*03 was previously associated with renal dysfunction by four groups with a combined total of 1261 ESRD subjects and over 3000 controls[5–7,44]. We found not only HLA-DRB1*03:01 but an entire haplotype to be associated with decreased eGFR and increased risk of poor clinical outcome. However, HLA-B*07 was reported to be protective against ESRD in 1620 ESRD patients and 1211 controls by Doxiadis et al.[10], and Karahan's study of 587 patients and 2643 controls[7]. In this population, HLA-B*07:02 was associated with decreased renal function. Furthermore, HLA-DRB1*04 was associated with adverse renal outcomes in three previous studies with over 4000 ESRD subjects[12,45,46], but here, DRB1*04:01 was linked to increased renal function. The remaining eight previously replicated HLA associations were not significant in this study. Overall, 14 of our associations confirmed previous observations[5–7,12,44,45,47–49], while 12 of our findings refuted previous results[7,10,12,45,46,49,50].

It is worth noting, however, that this study is much larger than any previous study. Most previous studies used case–control methodology (see "Strengths and weaknesses" below) and many failed adjust for multiple testing. Therefore, the findings reported here, which have undergone more stringent statistical testing, may be less prone to type I or II error.

**Implications.** This study is unique in that some of the HLA alleles associated with decreased renal function form a well-characterised haplotype. Both this and individual component HLA alleles have been associated with multiple diseases which result in CKD or ESRD, including systemic lupus erythematosus and IgA deficiency[51]. Our study indicates that even within a healthy population, renal function may be sub-clinically impaired in subjects with these alleles. These findings have the potential to impact upon clinical practice. HLA typing is already used as a diagnostic tool for disorders with strong HLA associations such as coeliac disease[52], ankylosing spondylitis[53], and actinic prurigo[54]. It may be advisable for clinicians to use HLA disease association typing to aid the diagnosis of renal failure, which could ensure timely therapeutic intervention. However, HLA associations with these diseases are much stronger than those reported here: the association between B*27 and ankylosing spondylitis has an odds ratio of 171[55], while HLA associations with coeliac disease have OR > 10[56], compared to ORs < 1.13 in this study. Clinicians and national kidney transplantation programmes may also use the HLA types associated with increased renal function to help identify suitable kidney donors.

**Strengths and weaknesses.** A key advantage of this study is the cohort size, which is larger than any previously published research. 382,204 subjects were included in the analysis of the primary outcome measure (eGFR), and the secondary analysis consisted of 11,379 cases of ESRD (and 389,928 controls). This study uses a variety of measures of renal function, most of which are calculated independently and are therefore unlikely to be subject to systemic bias. eGFR is a useful outcome measure because it provides a continuous scale, giving an accurate and precise estimate of renal function. Many previous studies used case–control methodology, reducing kidney function from a spectrum to binary categorisations such as "ESRD or healthy". Measuring renal function on a spectrum may strengthen the statistical and clinical significance of this study.

A limitation of this investigation is that the HLA typing was performed by imputation rather than direct genotyping, which is more accurate[57]. This is because the cost of HLA typing a cohort of this size using traditional methods is prohibitively expensive. The imputation program used for the UKB population was HLA*IMP:02, though Karnes' review[57] of competing programs suggests that SNP2HLA is more accurate. Nevertheless, the review stated that HLA*IMP:02 is 94% accurate when imputing white subjects which, given the size of our cohort, is acceptable within the scope of this study. Furthermore, 360 of the 362 imputed alleles were in HWE ($P > 0.00014$), suggesting that the majority of imputed allele frequencies were consistent with frequencies that might be expected in a stable population. The two alleles which were not in HWE (HLA-DQB1*02:02 and DPB1*03:01) were excluded from the analysis.

Some HLA associations found in this study do not appear to be part of a haplotype. These alleles may be independently associated with renal function, or they may be false positives caused by inaccurate imputation. For

| Coding system | Code | Decoded | ESRD patient | RRT recipient | Kidney transplant recipient | CKD patient |
|---|---|---|---|---|---|---|
| Coding 5 | 1195 | Renal/kidney transplant | Yes | Yes | Yes | |
| | 1476 | Fistula for dialysis | Yes | Yes | | |
| | 1580 | Dialysis access surgery | Yes | Yes | | |
| | 1581 | Haemodialysis access/fistula surgery | Yes | Yes | | |
| | 1582 | Peritoneal dialysis (CAPD) access surgery | Yes | Yes | | |
| Coding 6 | 1192 | Renal/kidney failure | Yes | | | |
| | 1193 | Renal failure requiring dialysis | Yes | Yes | | |
| | 1194 | Renal failure not requiring dialysis | Yes | | | |
| ICD-10 | I120 | Hypertensive renal disease with renal failure | Yes | | | |
| | I131 | Hypertensive heart and renal disease with renal failure | Yes | | | |
| | I132 | Hypertensive heart and renal disease with both (congestive) heart failure and renal failure | Yes | | | |
| | N17 | Acute renal failure | Yes | | | |
| | N170 | Acute renal failure with tubular necrosis | Yes | | | |
| | N172 | Acute renal failure with medullary necrosis | Yes | | | |
| | N178 | Other acute renal failure | Yes | | | |
| | N179 | Acute renal failure, unspecified | Yes | | | |
| | N18 | Chronic renal failure | Yes | | | |
| | N180 | End-stage renal disease | Yes | | | |
| | N181 | Chronic kidney disease, stage 1 | | | | Yes |
| | N182 | Chronic kidney disease, stage 2 | | | | Yes |
| | N183 | Chronic kidney disease, stage 3 | | | | Yes |
| | N184 | Chronic kidney disease, stage 4 | | | | Yes |
| | N185 | Chronic kidney disease, stage 5 | Yes | | | Yes |
| | N188 | Other chronic renal failure | Yes | | | |
| | N19 | Unspecified renal failure | Yes | | | |
| | P960 | Congenital renal failure | Yes | | | |
| | T861 | Kidney transplant failure and rejection | Yes | Yes | Yes | |
| | Y841 | Kidney dialysis | Yes | Yes | | |
| | Z49 | Care involving dialysis | Yes | Yes | | |
| | Z491 | Extracorporeal dialysis | Yes | Yes | | |
| | Z492 | Other dialysis | Yes | Yes | | |
| | Z940 | Kidney transplant status | Yes | Yes | Yes | |
| | Z992 | Dependence on renal dialysis | Yes | Yes | | |
| ICD-9 | 584 | Acute renal failure | Yes | | | |
| | 585 | Chronic renal failure | Yes | | | |
| | 586 | Renal failure, unspecified | Yes | | | |
| | 5845 | Acute renal failure with lesion of tubular necrosis | Yes | | | |
| | 5846 | Acute renal failure with lesion of renal cortical necrosis | Yes | | | |
| | 5847 | Acute renal failure with lesion of renal medullary (papillary) necrosis | Yes | | | |
| | 5848 | Acute renal failure with other specified pathological lesion in kidney | Yes | | | |
| | 5849 | Acute renal failure, unspecified | Yes | | | |
| | 5859 | Chronic renal failure | Yes | | | |
| | 5869 | Renal failure, unspecified | Yes | | | |
| | 77980 | Congenital renal failure | Yes | | | |
| | E8791 | Abn. reaction to kidney dialysis without misadventure at time | Yes | Yes | | |
| | V451 | Renal dialysis status | Yes | Yes | | |
| | V56 | Aftercare involving intermittent dialysis | Yes | Yes | | |
| | V560 | Aftercare involving extracorporeal dialysis | Yes | Yes | | |
| | V568 | Aftercare involving other dialysis | Yes | Yes | | |
| N/A | <15 | eGFR based on creatinine | Yes | | | |
| | <15 | eGFR based on cystatin | Yes | | | |
| | <15 | eGFR based on both creatinine and cystatin | Yes | | | |

Continued

| Coding system | Code | Decoded | ESRD patient | RRT recipient | Kidney transplant recipient | CKD patient |
|---|---|---|---|---|---|---|
| OPCS3 | 566 | Transplantation of kidney | Yes | Yes | Yes | |
| | 4013 | Paracentesis abdomini: peritoneal dialysis | Yes | Yes | | |
| | 4695 | Other operations on intestine, not elsewhere classified: isolation loop for dialysis | Yes | Yes | | |
| | 5661 | Transplantation of kidney: donor | Yes | Yes | Yes | |
| | 5662 | Transplantation of kidney: cadaver | Yes | Yes | Yes | |
| | 9503 | Other vascular procedures: haemodialysis | Yes | Yes | | |
| OPCS4 | L746 | Creation of graft fistula for dialysis | Yes | Yes | | |
| | M01 | Transplantation of kidney | Yes | Yes | Yes | |
| | M011 | Autotransplantation of kidney | Yes | Yes | Yes | |
| | M012 | Allotransplantation of kidney from live donor | Yes | Yes | Yes | |
| | M013 | Allotransplantation of kidney from cadaver NEC | Yes | Yes | Yes | |
| | M014 | Allotransplantation of kidney from cadaver heart beating | Yes | Yes | Yes | |
| | M015 | Allotransplantation of kidney from cadaver heart non-beating | Yes | Yes | Yes | |
| | M018 | Other specified transplantation of kidney | Yes | Yes | Yes | |
| | M019 | Unspecified transplantation of kidney | Yes | Yes | Yes | |
| | M172 | Pre-transplantation of kidney work-up—recipient | Yes | Yes | Yes | |
| | M174 | Post-transplantation of kidney examination—recipient | Yes | Yes | Yes | |
| | X40 | Compensation for renal failure | Yes | Yes | | |
| | X401 | Renal dialysis | Yes | Yes | | |
| | X402 | Peritoneal dialysis NEC | Yes | Yes | | |
| | X403 | Haemodialysis NEC | Yes | Yes | | |
| | X404 | Haemofiltration | Yes | Yes | | |
| | X405 | Automated peritoneal dialysis | Yes | Yes | | |
| | X406 | Continuous ambulatory peritoneal dialysis | Yes | Yes | | |
| | X407 | Haemoperfusion | Yes | Yes | | |
| | X408 | Other specified compensation for renal failure | Yes | Yes | | |
| | X409 | Unspecified compensation for renal failure | Yes | Yes | | |
| | X41 | Placement of ambulatory apparatus for compensation for renal failure | Yes | Yes | | |
| | X411 | Insertion of ambulatory peritoneal dialysis catheter | Yes | Yes | | |
| | X412 | Removal of ambulatory peritoneal dialysis catheter | Yes | Yes | | |
| | X418 | Other specified placement of ambulatory apparatus for compensation for renal failure | Yes | Yes | | |
| | X419 | Unspecified placement of ambulatory apparatus for compensation for renal failure | Yes | Yes | | |
| | X42 | Placement of other apparatus for compensation for renal failure | Yes | Yes | | |
| | X421 | Insertion of temporary peritoneal dialysis catheter | Yes | Yes | | |
| | X428 | Other specified placement of other apparatus for compensation for renal failure | Yes | Yes | | |
| | X429 | Unspecified placement of other apparatus for compensation for renal failure | Yes | Yes | | |
| | X431 | Extracorporeal albumin haemodialysis | Yes | Yes | | |
| UKB data | 42026 | End stage renal disease report | Yes | | | |
| | 42027 | Date of end stage renal disease report | Yes | | | |

**Table 6.** Codes included in definitions of ESRD, kidney transplant status, RRT status, and CKD. This shows the various coding systems and codes used to identify subjects with adverse clinical outcomes.

example, HLA-B*44:02 was not significant in this study, but is commonly associated with C*05:01, DRB1*04:01, DQA1*03:01 and DQB1*03:02 in combination with either A*25:01, A*29:02, or A*32:01, which were all linked to increased renal function. Our study identified HLA-B*44:03, rather than B*44:02, to be associated with increased renal function. This discrepancy may suggest that HLA-B*44:02 alleles were incorrectly imputed as B*44:03, though given that C*16:01 (which is often seen in LD with B*44:03) was also associated with increased renal function may validate the observation regarding B*44:03. Any dubious observations may be resolved by repeating the imputation using an alternative imputation programme or additional reference panels.

It is possible that the strategy employed to identify subjects with adverse kidney-related clinical outcomes was insufficiently comprehensive to capture all cases. If data held by UKB were incomplete, or if relevant codes were not included (see Table 6), subjects with poor renal outcomes would be mischaracterised as healthy. This could be averted by obtaining a peer-reviewed validation of the coding systems that documents exactly which codes are representative of adverse renal outcomes, but to the best of our knowledge no such validation exists. Clinical outcomes were secondary outcome measures in this study; the primary outcome of eGFR is not affected by this limitation.

A final limitation of this study is that the sizes of the associations with eGFR were smaller than previously published HLA disease associations[55,56] and possibly too small to be considered clinically relevant. 25 out of 33 (76%) significant associations with eGFR had a beta value between −0.5 and 0.5, suggesting that the presence of the allele has only a minor effect on kidney function. However, in seven cases, these apparently small effects were corroborated by associations with adverse clinical outcomes, implying that small beta values are not a contraindication of clinical relevance.

## Conclusions

This study has identified 22 HLA types which are associated with increased kidney function, and 11 which are linked to decreased kidney function in a large UK population. Many of these are commonly inherited together in haplotypes. Importantly, seven alleles, which are each seen in between 14–34% of the cohort, were linked to both decreased eGFR and increased incidence of adverse clinical outcomes. Due to the constitution of the cohort, the results of this study can only be applied to white British people aged 39–73. Repeating the analyses with alternative cohorts may add considerably to our current knowledge and allow a better assessment on the implications for population health.

Received: 11 September 2020; Accepted: 9 December 2020
Published online: 04 February 2021

## References

1. Mortality GBD, Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459–1544 (2016).
2. Luyckx, V. A., Tonelli, M. & Stanifer, J. W. The global burden of kidney disease and the sustainable development goals. *Bull. World Health Organ.* **96**, 414–422 (2018).
3. Robson, K. J., Ooi, J. D., Holdsworth, S. R., Rossjohn, J. & Kitching, A. R. HLA and kidney disease: From associations to mechanisms. *Nat. Rev. Nephrol.* **14**, 636–655 (2018).
4. Rivera, P. S. *et al.* HLA class I association with progression to end-stage renal disease in patients from Zulia, Venezuela. *Inmunologia* **31**, 37–42 (2012).
5. Yamakawa, R. H., Saito, P. K., da Silva Junior, W. V., de Mattos, L. C. & Borelli, S. D. Polymorphism of leukocyte and erythrocyte antigens in chronic kidney disease patients in southern Brazil. *PLoS ONE* **9**, e84456 (2014).
6. Hamdi, N. M., Al-Hababi, F. H. & Eid, A. E. HLA class I and class II associations with ESRD in Saudi Arabian population. *PLoS ONE* **9**, e111403 (2014).
7. Karahan, G. E. *et al.* Impact of HLA on the underlying primary diseases in Turkish patients with end-stage renal disease. *Ren. Fail.* **31**, 44–49 (2009).
8. Xu, X. *et al.* Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nat. Commun.* **9**, 4800 (2018).
9. Davood, P. P., Farhadi, N. & Najafizadeh, M. Protective and susceptible HLA class I genes in patients with end-stage renal disease. *Res. J. Biol. Sci.* **3**, 1344–1346 (2008).
10. Doxiadis, I. I., De Lange, P., De Vries, E., Persijn, G. G. & Claas, F. H. Protective and susceptible HLA polymorphisms in IgA nephropathy patients with end-stage renal failure. *Tissue Antigens* **57**, 344–347 (2001).
11. Fejzic, E. *et al.* HLA genotyping in patients with end-stage renal disease waiting for cadaveric renal transplantation in federation of Bosnia and Herzegovina. *Maced. J. Med. Sci.* **5**, 1–5 (2017).
12. Gencik, M. *et al.* Immunogenetic risk factors for anti-neutrophil cytoplasmic antibody (ANCA)-associated systemic vasculitis. *Clin. Exp. Immunol.* **117**, 412–417 (1999).
13. Gerhardsson, J. *et al.* Histological antiphospholipid-associated nephropathy versus lupus nephritis in patients with systemic lupus erythematosus: An observational cross-sectional study with longitudinal follow-up. *Arthr. Res. Ther.* **17**, 109 (2015).
14. Sudlow, C. *et al.* UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
15. Data-Field 22011: Genetic relatedness pairing (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22011 (Accessed 22 Nov 2019).
16. Lowe, M. *et al.* Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom. *HLA* **96**, 697–708 (2020).
17. Data-Field 22005: Missingness (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22005 (Accessed 22 Nov 2019).
18. Choo, S. Y. The HLA system: Genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* **48**, 11–23 (2007).
19. Sabolic, I. *et al.* Gender differences in kidney function. *Pflugers Arch.* **455**, 397–429 (2007).
20. Data-Field 22001: Genetic sex (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22001 (Accessed 22 Nov 2019).
21. Data-Field 22006: Genetic ethnic grouping (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22006 (Accessed 22 Nov 2019).
22. Data-Field 22010: Recommended genomic analysis exclusions (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22010 (Accessed 3 Dec 2019).
23. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
24. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
25. Dilthey, A. *et al.* Multi-population classical HLA type imputation. *PLoS Comput. Biol.* **9**, e1002877 (2013).
26. Motyer, A. *et al.* Practical use of methods for imputation of HLA alleles from SNP genotype data. *bioRxiv* https://doi.org/10.1101/091009 (2016).
27. Stevens, L. A. & Levey, A. S. Measurement of kidney function. *Med. Clin. N. Am.* **89**, 457–473 (2005).
28. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
29. Inker, L. A. *et al.* Estimating glomerular filtration rate from serum creatinine and cystatin C. *N. Engl. J. Med.* **367**, 20–29 (2012).
30. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders* (Switzerland, Geneva, 1992).
31. NHS Classifications OPCS-4 (2019). https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/10 (Accessed 28 Nov 2019).
32. Data-Coding 5: Operation (2019). http://biobank.ctsu.ox.ac.uk/crystal/coding.cgi?id=5 (Accessed 28 Nov 2019).
33. Data-Field 22019: Sex chromosome aneuploidy (2019). http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22019 (Accessed 22 Nov 2019).
34. Armstrong, R. A. When to use the Bonferroni correction. *Ophthalmic Physiol. Opt.* **34**, 502–508 (2014).
35. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

36. Stevens, L. A., Coresh, J., Greene, T. & Levey, A. S. Assessing kidney function–measured and estimated glomerular filtration rate. *N. Engl. J. Med.* **354**, 2473–2483 (2006).
37. Estimated Glomerular Filtration Rate (eGFR) (2020). https://www.kidney.org/atoz/content/gfr (Accessed 26 Nov 2020).
38. Degli-Esposti, M. A. *et al.* Ancestral haplotypes: Conserved population MHC haplotypes. *Hum. Immunol.* **34**, 242–252 (1992).
39. Haplotype Frequency Search (USA European American) (2019). http://www.allelefrequencies.net/hla6003a.asp?hla_population=1699 (Accessed 3 Dec 2019).
40. Haplotype Frequency Search (Ireland Northern) (2019). http://www.allelefrequencies.net/hla6003a.asp?hla_population=1243 (Accessed 3 Dec 2019).
41. Levey, A. S. *et al.* A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann. Intern. Med.* **130**, 461–470 (1999).
42. Cockcroft, D. W. & Gault, M. H. Prediction of creatinine clearance from serum creatinine. *Nephron* **16**, 31–41 (1976).
43. Florkowski, C. M. & Chew-Harris, J. S. Methods of estimating GFR—Different equations including CKD-EPI. *Clin. Biochem. Rev.* **32**, 75–79 (2011).
44. Dai, C. S. *et al.* Association between human leucocyte antigen subtypes and risk of end stage renal disease in Taiwanese: A retrospective study. *BMC Nephrol.* **16**, 177 (2015).
45. Chang, D. Y., Luo, H., Zhou, X., Chen, M. & Zhao, M. H. Association of HLA genes with clinical outcomes of ANCA-associated vasculitis. *Clin. J. Am. Soc. Nephrol.* **7**, 1293–1299 (2012).
46. Cao, Q. *et al.* HLA polymorphism and susceptibility to end-stage renal disease in Cantonese patients awaiting kidney transplantation. *PLoS ONE* **9**, e90869 (2014).
47. Crispim, J. C. *et al.* HLA polymorphisms as incidence factor in the progression to end-stage renal disease in Brazilian patients awaiting kidney transplant. *Transpl. Proc.* **40**, 1333–1336 (2008).
48. Almogren, A., Shakoor, Z. & Hamam, K. D. Human leucocyte antigens: Their association with end-stage renal disease in Saudi patients awaiting transplantation. *Br. J. Biomed. Sci.* **69**, 159–163 (2012).
49. Perez-Luque, E. *et al.* Contribution of HLA class II genes to end stage renal disease in Mexican patients with type 2 diabetes mellitus. *Hum. Immunol.* **61**, 1031–1038 (2000).
50. Prakash, S. *et al.* Distribution of Killer cell immunoglobulin like receptor genes in end stage renal disease among North Indian population. *Hum. Immunol.* **74**, 1399–1445 (2013).
51. Price, P. *et al.* The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. *Immunol. Rev.* **167**, 257–274 (1999).
52. Kaukinen, K., Partanen, J., Maki, M. & Collin, P. HLA-DQ typing in the diagnosis of celiac disease. *Am. J. Gastroenterol.* **97**, 695–699 (2002).
53. Chen, B. *et al.* Role of HLA-B27 in the pathogenesis of ankylosing spondylitis (Review). *Mol. Med. Rep.* **15**, 1943–1951 (2017).
54. Sheridan, D. P., Lane, P. R., Irvine, J., Martel, M. J. & Hogan, D. J. HLA typing in actinic prurigo. *J. Am. Acad. Dermatol.* **22**, 1019–1023 (1990).
55. Brown, M. A. *et al.* HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. *Ann. Rheum. Dis.* **55**, 268–270 (1996).
56. De Silvestri, A. *et al.* HLA-DQ genetics in children with celiac disease: A meta-analysis suggesting a two-step genetic screening procedure starting with HLA-DQ beta chains. *Pediatr. Res.* **83**, 564–572 (2018).
57. Karnes, J. H. *et al.* Comparison of HLA allelic imputation programs. *PLoS ONE* **12**, e0172444 (2017).


## Author contributions

The study was devised by K.P., W.O., A.V., A.P., T.A.U., and M.L. Data was provided by U.K.B. Quality control and data analysis was performed by M.L. with advice from A.P., I.G., and P.H. Results were interpreted by M.L., A.P., K.P., and J.W. The paper was written by M.L. with advice from all the authors.

## Funding

Funding was provided by Kidneys for Life.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# 12 Paper 3: Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom

This paper was published in HLA on 28th September 2020. There is a typographical error in section 4.1 of this paper. The sentence reading, “Absolute effect sizes range from 0.62 to 3.20 mL/min/1.73 $m^2$” should read, “Absolute effect sizes range from 1.08 to 3.20 mL/min/1.73 $m^2$”.

Received: 29 April 2020 | Revised: 10 September 2020 | Accepted: 21 September 2020

DOI: 10.1111/tan.14078

ORIGINAL ARTICLE

HLA Immune Response Genetics WILEY

# Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom

**Marcus Lowe**[1,2] | **Antony Payton**[3] | **Arpana Verma**[2] | **Isla Gemmell**[2] | **Judith Worthington**[1] | **Patrick Hamilton**[4] | **William Ollier**[2,5] | **Titus Augustine**[4,6] | **Kay Poulton**[1,2]

[1]Transplantation Laboratory, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK

[2]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK

[3]Faculty of Biology, Medicine and Health, Division of Informatics, Imaging & Data Sciences, School of Health Sciences, University of Manchester, Manchester, UK

[4]Department of Renal and Pancreas Transplantation, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK

[5]Faculty of Science and Engineering, Centre for Bioscience, Manchester Metropolitan University, Manchester, UK

[6]Division of Diabetes, Endocrinology and Gastroenterology, University of Manchester, Manchester, UK

**Correspondence**
Marcus Lowe, Transplantation Laboratory, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK
Email: marcus.lowe@mft.nhs.uk

**Funding information**
Kidneys for Life

Human leukocyte antigens (HLA) have been associated with renal function, but previous studies report contradictory findings. There has been a lack of research into how HLA affects renal function in Black, Asian and Minority Ethnic (BAME) people in the UK, despite BAME people being disproportionately affected by renal dysfunction. This study included >27 000 UK Biobank subjects of six ethnicities (>12 100 Irish, >5400 Indian, >4000 Black Caribbean, >3000 Black African, >1600 Pakistani, and >1400 Chinese) aged 39 to 73. Subjects' high-resolution HLA genotypes were imputed using *HLA*IMP:02* software. Regression analysis was used to compare 108 imputed HLA alleles with two measures of estimated glomerular filtration rate (eGFR): one based on serum creatinine; one based on serum cystatin. Secondary analysis compared CKD stage 2 subjects to healthy controls. Nine imputed HLA alleles were associated with eGFR (adjusted $P < .05$). Six associations were based on creatinine in Black African subjects: *HLA-B*53:01* (beta = −2.628, adjusted $P = 4.69 \times 10^{-4}$); *C*04:01* (beta = −1.667, adjusted $P = .0269$); *DPA1*02:01* (beta = −1.569, adjusted $P = .0182$); and *DPA1*02:02* (beta = −1.716, adjusted $P = .0251$) were linked to decreased renal function, *while DRB1*03:01* (beta = 3.200, adjusted $P = 3.99 \times 10^{-3}$) *and DPA1*01:03* (beta = 2.276, adjusted $P = 2.31 \times 10^{-5}$) were linked to increased renal function. Two of these (*HLA-B*53:01 and C*04:01*) are commonly inherited together. In Irish subjects, *HLA-DRB1*04:01* (beta = 1.075, adjusted $P = .0138$) was linked to increased eGFR (based on cystatin); in Indian subjects, *HLA-DRB1*03:01* (beta = −1.72, adjusted $P = 4.78 \times 10^{-3}$) *and DQB1*02:01* (beta = −1.755, adjusted $P = 2.26 \times 10^{-3}$)were associated with decreased eGFR (based on cystatin). No associations were found in the other three ethnic groups. Nine HLA alleles appear to be associated with kidney function in BAME people in the UK. This could have applications for the diagnosis and treatment of renal disease and could help reduce health inequalities in the UK.

**KEYWORDS**

genome-wide association study, glomerular filtration rate, HLA antigens, imputation, kidney failure, chronic, polymorphism, single nucleotide, renal insufficiency

## 1 | INTRODUCTION

Chronic kidney disease (CKD) is 17th in the worldwide list of diseases which cause the most "years of lost life", according to the Global Burden of Disease 2015 study; it was 21st in 2005 and 25th in 1990.[1] The same study reported that CKD caused 1.2 million deaths in 2015, an increase of 32% compared with 2005. In developed countries, treatment for end-stage renal disease (ESRD) typically accounts for around 2-3% of the overall healthcare budget.[2] In England, CKD affects over 1.8 million people and over 40 000 people are on some form of renal replacement therapy (RRT, usually either dialysis or kidney transplantation).[3]

Renal dysfunction is a problem that disproportionately affects Black, Asian and Minority Ethnic (BAME) people in the UK. Although BAME people represent only 11% of the UK population,[4] they accounted for 24% of patients dependent on renal replacement therapy in 2017.[5] BAME people are more likely to develop CKD than white people,[6] and those who have CKD experience faster decline in renal function than white people.[7] Understanding the genetic causes of renal dysfunction in BAME people in the UK could play an important role in reducing these health inequalities.

Heritability of CKD is thought to be between 30% and 75%.[8] The Human Leukocyte Antigen (HLA) system, encoded within the major histocompatibility complex, has been reported to be associated with renal function.[9] This may be because unique HLA alleles confer different levels of protection from different diseases associated with kidney dysfunction.[10] Almost 100 HLA alleles and haplotypes have been linked to increased or decreased risk of renal failure.[11] However, conflicting associations have been reported, for example *HLA-A*11* was linked to increased ESRD in a study of Azerbaijani subjects[12] but protective against ESRD in a study of Turkish subjects.[13] This and other conflicting findings may be in part due to population stratification, as well as differing patterns of linkage disequilibrium (which lead to different combinations of genes among different ethnicities), and environmental effects.

This study uses a cohort of over 27 000 subjects aged 39 to 73 who were recruited in the UK but do not identify as ethnically "British". The study aims to discover high-resolution HLA alleles which are associated with either increased or decreased renal function. Most previous studies used case-control methodology with a smaller number of subjects, usually all of a single ethnicity or nationality. Here we present the findings of a retrospective cohort study using subjects of six different ethnicities who are mostly healthy.

## 2 | METHODS

### 2.1 | Quality control and identification of study population

This study used data from UK Biobank (UKB) to test for associations between HLA and renal function in over 27 000 subjects from six different self-reported ethnic groups (Irish, Indian, Black Caribbean, Black African, Pakistani, and Chinese). UKB holds data on 502 616 subjects who were aged 39-73 when they were recruited in the UK between 2006 and 2010. 409 692 subjects (81.5%) reported their ethnic background as "British" and genetic principal component analysis performed by UKB concurred; these subjects were analysed separately and the results reported elsewhere. The remaining 92 924 subjects (those who did not report their ethnic background to be "British", and those who self-identified as "British" but UKB's principal component analysis suggested were not British) were considered for this study.

Seven hundred and fourteen individuals were removed during quality control. Of these: 129 had ambiguous genetic sex (a confounder in the calculation of subjects' kidney function); 420 were related to other subjects in the cohort (kinship coefficient $\geq 0.044$; the individual with the most complete set of genetic data was included); 132 were removed at UKB's recommendation due to "missing genetic data"; and 33 withdrew their consent to be included in the study. There were 92 210 subjects remaining after quality control.

As previous studies have suggested that HLA associations with renal function vary depending on ethnicity, subjects were analysed only within their ethnic groups. There were 23 different groups depending on subjects' answers to the question, "What is your ethnic background?" Only groups of 1000 or more subjects were analysed in this study in order to reduce type I and type II errors. A minimum cohort size of 1000 unrelated individuals provides 94% power to detect a genetic effect size of 3%, assuming a 5% minor allele frequency (minimum used in our analysis) and additive model. We used a significance threshold of unadjusted $P < .001$ (correction for multiple testing). Furthermore, ethnic backgrounds which were unknown, mixed, or categorised as something "other" than the options provided (eg, "any other white background") were excluded from the analysis due to possible heterogeneity. Finally, the subjects who self-reported as "British" but whom UKB deemed not to be British were also excluded. This left a total of six ethnic backgrounds to be analysed: Irish ($N = 13\,071$); Indian ($N = 5919$); Black Caribbean ($N = 4488$); Black African ($N = 3386$); Pakistani ($N = 1822$); and Chinese ($N = 1563$). Table 1 shows the 23 ethnic groups, the

number of subjects in each group, and whether the group was included in the analysis (along with reasons for exclusion if applicable).

## 2.2 | HLA typing

Subjects were HLA typed by UKB using single nucleotide polymorphism imputation. *HLA*IMP:02* software[14] was used to impute the following HLA loci: HLA-A, B, C (Class I) and DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4, and DRB5 (Class II) at a level equivalent to high-resolution typing using eight reference datasets. 360 high-resolution HLA genotypes were imputed, including a group of subjects who do not have genes at HLA-DRB3/4/5.[15] Some alleles were excluded due to their frequencies not being in Hardy-Weinberg equilibrium (HWE), $P < .00014$. Overall, nine alleles were excluded from particular ethnic groups due to HWE analysis (one Irish, four Indian, one Black Caribbean, one Black African, two Pakistani, and zero Chinese). Only alleles with >5% frequency were considered for analysis. There were a total of 108 alleles with >5% frequency in at least one cohort (45 for Irish subjects, 56 Indian, 54 Black Caribbean, 55 Black African, 54 Pakistani, and 48 Chinese). Table 2 shows the alleles which had >5% frequency in each ethnic group, as well as which alleles were excluded due to HWE analysis.

## 2.3 | Measuring renal function

Renal function was primarily measured using estimated glomerular filtration rate (eGFR), based on two

**TABLE 1** Subjects by self-reported ethnic background

| Self-reported ethnic group | Self-reported ethnic background | *N* | Included in analysis |
|---|---|---|---|
| White | Prefer not to answer | 567 | No[a] |
| | British | 32 662 | No[b] |
| | Irish | 13 071 | Yes |
| | Any other white background | 16 252 | No |
| Mixed | Prefer not to answer | 48 | No[a,c] |
| | White and Black Caribbean | 619 | No[a,c] |
| | White and Black African | 423 | No[a,c] |
| | White and Asian | 831 | No[a,c] |
| | Any other mixed background | 1029 | No[c] |
| Asian or Asian British | Prefer not to answer | 42 | No[a] |
| | Indian | 5919 | Yes |
| | Pakistani | 1822 | Yes |
| | Bangladeshi | 236 | No[a] |
| | Any other Asian background | 1809 | No[c] |
| Black or Black British | Prefer not to answer | 27 | No[a] |
| | Caribbean | 4488 | Yes |
| | African | 3386 | Yes |
| | Any other Black background | 123 | No[a,c] |
| Chinese | | 1563 | Yes |
| Prefer not to answer | | 1651 | No[c] |
| Do not know | | 216 | No[a,c] |
| Other ethnic group | | 4537 | No[c] |
| Data missing | | 889 | No[a,c] |
| Total | | 92 210 | |

[a]Excluded as group size was less than 1000.
[b]Excluded as these subjects self-reported their ethnic background as "British", but UKB's principal component analysis suggested they were not British.
[c]Excluded as ethnic group was mixed or "other" or missing.

**TABLE 2** HLA alleles by ethnicity with >5% frequency and HLA alleles excluded due to HWE analysis

| HLA- | Irish (%) | Indian (%) | Black Caribbean (%) | Black African (%) | Pakistani (%) | Chinese (%) |
|---|---|---|---|---|---|---|
| *A*01:01* | 23.1 | 15.2 | – | – | 15.9 | – |
| *A*11:01* | 6.9 | 13.4 | – | – | 13.3 | 24.0 |
| *A*11:02* | – | – | – | – | –[a] | – |
| *A*02:01* | 25.9 | 5.2 | 10.6 | 9.4 | 5.8 | 8.7 |
| *A*02:02* | – | – | – | 5.1 | – | – |
| *A*02:07* | – | – | – | – | – | 8.6 |
| *A*23:01* | – | – | 11.3 | 10.7 | – | – |
| *A*24:02* | 6.3 | 11.6 | – | – | 11.4 | 14.9 |
| *A*26:01* | – | 5.5[a] | – | – | 7.1 | – |
| *A*30:01* | – | – | 7.6 | 8.1 | – | – |
| *A*30:02* | – | – | 6.7 | 8.1 | – | – |
| *A*03:01* | 14.2 | 6.4 | 7.9 | 7.1 | 5.8 | – |
| *A*33:03* | – | 9.3 | 5.9 | 6.0 | 7.2 | 11.1 |
| *A*68:01* | – | 6.7 | – | – | 7.4 | – |
| *A*68:02* | – | – | 6.6 | 7.1 | – | – |
| *A*74:01* | – | – | 6.5 | 6.1 | – | – |
| *B*13:01* | – | – | – | – | – | 5.9 |
| *B*15:02* | – | – | – | – | – | 6.3 |
| *B*15:03* | – | – | – | 5.4 | – | – |
| *B*35:01* | – | – | 6.6 | 6.8 | – | – |
| *B*35:03* | – | 5.6 | – | – | 5.3 | – |
| *B*40:01* | – | – | – | – | – | 13.7 |
| *B*42:01* | – | – | 5.9 | 6.4 | – | – |
| *B*44:02* | 12.1 | – | – | – | – | – |
| *B*44:03* | 5.6 | 6.1 | – | – | – | – |
| *B*46:01* | – | – | – | – | – | 13.4 |
| *B*51:01* | – | 7.7 | – | – | 9.1 | 5.1 |
| *B*52:01* | – | 7.2 | – | – | 7.4 | – |
| B*53:01 | – | – | 13.0 | 14.4 | – | – |
| *B*57:01* | – | 5.2 | – | – | – | – |
| *B*58:01* | – | – | – | – | – | 9.3 |
| *B*07:02* | 17.8 | – | 6.8 | 6.2 | – | – |
| *B*08:01* | 17.5 | 6.4 | – | – | 9.6 | – |
| *C*01:02* | – | – | – | – | – | 20.1 |
| *C*12:02* | – | 8.2 | – | – | 7.9 | – |
| *C*12:03* | – | 5.9 | – | – | 6.9 | – |
| *C*15:02* | – | 10.4 | – | – | 10.3 | – |
| *C*16:01* | – | – | 9.2 | 9.5 | – | – |
| *C*17:01* | – | – | 7.1 | 8.1 | – | – |
| *C*03:02* | – | – | – | – | – | 9.4 |
| *C*03:04* | 5.4 | – | 5.1 | 5.1 | – | 10.5 |
| C*04:01 | 6.6 | 13.3 | 22.4 | 23.4 | 12.5 | – |
| *C*05:01* | 12.0 | – | – | – | – | – |

**TABLE 2** (Continued)

| HLA- | Irish (%) | Indian (%) | Black Caribbean (%) | Black African (%) | Pakistani (%) | Chinese (%) |
|---|---|---|---|---|---|---|
| *C*06:02* | 9.3 | 13.3 | 6.9 | 6.5 | 13.6 | – |
| *C*07:01* | 20.3 | 9.4 | 12.1 | 11.5 | 8.1 | – |
| *C*07:02* | 18.9 | 13.6 | 6.7 | 6.0 | 14.2 | 17.7 |
| *C*08:01* | – | – | – | – | – | 10.6 |
| *C*08:02* | 5.7 | – | – | – | – | – |
| *DPA1*01:03* | 18.0 | 33.9 | 34.7 | 28.0 | 32.7 | 32.2 |
| *DPA1*02:01* | 14.3 | 24.6 | 32.6 | 32.7 | 24.7 | 9.0 |
| *DPA1*02:02* | – | 8.2 | 18.7 | 22.3 | 6.8 | 45.1 |
| *DPA1*03:01* | – | – | 9.7 | 11.7 | – | – |
| *DPB1*01:01* | – | – | 33.5 | 37.9[a] | – | – |
| *DPB1*13:01* | – | 6.2 | – | – | 5.6 | 5.6 |
| *DPB1*17:01* | – | – | 6.3 | 6.9 | – | – |
| *DPB1*18:01* | – | – | 5.4 | – | – | – |
| *DPB1*02:01* | 10.1 | 18.4 | 8.9 | 7.9 | 16.5 | 15.3 |
| *DPB1*02:02* | – | – | – | – | – | 6.3 |
| *DPB1*26:01* | – | 5.1 | – | – | 5.4 | – |
| *DPB1*03:01* | 9.5 | – | – | – | – | – |
| *DPB1*04:01* | 47.2 | 36.0 | 8.2 | – | 38.1 | 8.1 |
| *DPB1*04:02* | 8.8 | 7.4 | 11.6 | 12.4 | 7.5 | – |
| *DPB1*05:01* | – | – | – | – | – | 40.4 |
| *DQA1*01:01* | 12.4 | 18.0 | 13.1 | 13.4 | 16.2 | 11.4 |
| *DQA1*01:02* | 23.4 | 9.9 | 30.4 | 30.1 | 9.2 | 17.5 |
| *DQA1*01:03* | – | 20.6 | – | – | 18.5 | 7.4 |
| *DQA1*02:01* | 15.7 | 15.0 | 10.4 | 8.9 | 14.6 | – |
| *DQA1*03:01* | 18.1 | 9.5 | 7.9 | 7.4 | 7.3 | 28.1 |
| *DQA1*04:01* | – | – | 11.6 | 12.8 | – | – |
| *DQA1*05:01* | 23.3 | 21.6[a] | 17.8 | 18.0 | 30.0 | 16.8 |
| *DQA1*06:01* | – | – | – | – | – | 10.2 |
| DQB1*02:01 | 17.5 | 12.2 | 12.4 | 12.9 | 18.1 | 7.1 |
| *DQB1*02:02* | 9.6[a] | 8.4 | – | – | 8.9 | – |
| *DQB1*03:01* | 15.4 | 13.0 | 16.9 | 16.9 | 13.9[a] | 20.8 |
| *DQB1*03:02* | 8.5 | 7.9 | – | – | 5.9 | 6.3 |
| *DQB1*03:03* | 5.1 | 5.6 | – | – | 5.5 | 16.2 |
| *DQB1*04:01* | – | – | – | – | – | 5.4 |
| *DQB1*04:02* | – | – | 7.3 | 8.0 | – | – |
| *DQB1*05:01* | 10.9 | 9.8 | 14.3 | 15.0 | 8.9 | – |
| *DQB1*05:02* | – | – | – | – | – | 10.2 |
| *DQB1*05:03* | – | 8.6 | – | – | 7.3 | – |
| *DQB1*06:01* | – | 14.6 | –[a] | – | 12.8 | 10.9 |
| *DQB1*06:02* | 18.8 | – | 17.3 | 16.1 | – | – |
| *DQB1*06:03* | – | 6.8 | – | – | 6.3 | – |
| *DRB1*10:01* | – | 6.4 | – | – | 5.3 | – |
| *DRB1*01:01* | 7.1 | – | – | – | – | – |

(Continues)

**TABLE 2** (Continued)

| HLA- | Irish (%) | Indian (%) | Black Caribbean (%) | Black African (%) | Pakistani (%) | Chinese (%) |
|---|---|---|---|---|---|---|
| *DRB1*01:02* | – | – | 5.1 | 5.7 | – | – |
| *DRB1*11:01* | – | 5.3 | 7.7 | 8.0 | 5.6 | – |
| *DRB1*12:02* | – | – | – | – | – | 10.6 |
| *DRB1*13:01* | – | 6.5 | – | 5.7 | 6.0 | – |
| *DRB1*13:02* | – | – | 6.5 | 7.0 | – | – |
| *DRB1*15:01* | 19.2 | 9.8 | – | – | 8.3 | 8.8 |
| *DRB1*15:02* | – | 10.1 | – | – | 9.5 | – |
| *DRB1*15:03* | – | – | 12.7 | 13.1 | – | – |
| DRB1*03:01 | 17.5 | 11.6 | 6.4 | 6.6 | 17.7 | 7.2 |
| *DRB1*03:02* | – | – | 6.2 | 7.3 | – | – |
| *DRB1*04:01* | 9.8 | – | – | – | – | – |
| *DRB1*04:05* | – | – | – | – | – | 6.0 |
| *DRB1*07:01* | 15.7 | 14.9 | 8.9 | 7.7 | 14.6 | – |
| *DRB1*08:03* | – | – | – | – | – | 6.2 |
| *DRB1*08:04* | – | –[a] | 6.5 | 7.0 | – | – |
| *DRB1*09:01* | – | – | – | – | – | 15.6 |
| *DRB3*01:01* | 19.5 | 5.8 | 13.4 | 14.5 | 5.1 | – |
| *DRB3*02:02* | 9.3 | 27.8[a] | 24.0 | 24.6 | 35.3 | 20.3 |
| *DRB3*03:01* | – | – | 9.4 | 10.1 | – | 11.6 |
| *DRB4*01:01* | 9.1 | – | 8.4 | 8.0 | – | – |
| *DRB4*01:03* | 23.2 | 20.5 | – | – | 19.1 | 27.8 |
| *DRB5*01:01* | 19.1 | 8.8 | 15.0 | 13.8 | 7.8 | 12.0 |
| DRB3/4/5—no gene | 12.5 | 10.7 | 15.0 | 15.6 | 9.4 | 8.7 |
| No. alleles included in analysis | 44 | 53 | 54 | 54 | 53 | 48 |

*Note:* "–"Allele had <5% frequency among this ethnic group so was excluded from analysis.
[a]Allele was not in HWE in this ethnic group so was excluded from analysis.

biomarkers: creatinine and cystatin. High levels of these biomarkers in a subject's serum are a sign of decreased renal function, while low levels are a sign of increased renal function. Subjects with decreased renal function have lower eGFRs than those with normal or increased renal function. This study used the CKD-EPI method of calculating eGFR,[16] which adjusts for age, sex, and ethnicity. Two separate formulae were used: one based on creatinine, and one based on cystatin. 2333 subjects were missing creatinine values (7.8%) and 2322 were missing cystatin values (7.7%), and were excluded from the corresponding analyses. For both formulae, eGFR is measured in millilitres per minute and the unit of any eGFR value reported in this text is "mL/min/1.73 $m^2$". eGFR ≥120 mL/min/1.73 $m^2$ is considered healthy, with lower eGFRs being indicative of decreasing renal function. eGFR <60 is indicative of CKD stage 3, while eGFR <15 mL/min/1.73 $m^2$ is indicative of CKD stage 5 (ESRD, requiring RRT). A table summarising eGFRs by ethnicity can be seen in Section 3.

Subjects were also sorted into categories based on their eGFR and secondary "case-control" analysis was performed. Subjects with eGFR (based on cystatin) <90 mL/min/1.73 $m^2$ (a "mild" decrease in renal function, indicative of at least CKD stage 2 when combined with evidence of kidney damage[17]) were compared with subjects with eGFR ≥90 mL/min/1.73 $m^2$ ("normal or increased" renal function).

### 2.3.1 | Regression analysis

Linear regression analysis was used to test for associations between HLA alleles (the predictor variable) and eGFR (a continuous outcome variable). Results of linear regression are reported in terms of beta values; an

HLA allele with beta >0 indicates that the allele is associated with increased eGFR (compared with subjects who do not have the allele), while beta <0 indicates that the allele is linked to decreased eGFR. The magnitude of the beta value shows the size of the association: the further from 0, the larger the association. 188 subjects who had received RRT were excluded from this analysis, as their eGFR values may have suggested healthy renal function even though their native function was poor. The total number of subjects included in analysis of eGFR based on creatinine was: 12140 Irish; 5427 Indian; 4068 Black Caribbean; 3012 Black African; 1649 Pakistani; and 1432 Chinese (27 728 total). For analysis of eGFR based on cystatin, the number of subjects included was: 12143 Irish; 5428 Indian; 4068 Black Caribbean; 3017 Black African; 1649 Pakistani; and 1434 Chinese (27 739 total). Separate regression analyses were performed for each of the six ethnic groups and both eGFR formulae.

Logistic regression analysis was performed to test for HLA associations in subjects with eGFR (based on cystatin) <90 mL/min/1.73 $m^2$ vs controls with eGFR ≥90 mL/min/1.73 $m^2$. After excluding subjects with missing eGFR and those who had received RRT, there were 12 752 cases and 14 987 controls (Irish: 5964 cases and 6179 controls; Indian: 3199 cases and 2229 controls; Black Caribbean: 1406 cases and 2662 controls; Black African: 951 cases and 2066 controls; Pakistani: 905 cases and 744 controls; Chinese: 327 cases and 1107 controls). Results of logistic regression are reported in terms of odds ratio (OR): HLA alleles with OR > 1 are associated with increased risk of CKD, while OR < 1 indicates decreased risk of CKD.

In all analyses, age and sex were included as covariates. Bonferroni correction thresholds were applied for each ethnic group based on the number of alleles analysed within that group. Unadjusted *P* values for significance ranged from $P < .00114$ (0.05/44, Irish subjects) to $P < .00093$ (0.05/54, Black African subjects and Black Caribbean subjects). Analysis of all genes was performed using Plink software,[18] except the analysis of subjects who have an absence of genes at HLA-DRB3/4/5, which was performed using Stata/SE 13.0 (StataCorp).

## 3 | RESULTS

### 3.1 | eGFR values by ethnicity

Table 3 summarises the eGFR values calculated, stratified by ethnicity. All calculated eGFR values are included, although some subjects were excluded from analysis due to receiving RRT.

### 3.2 | Regression analysis

After correction for multiple testing, nine associations were found between HLA and renal function in UK ethnic minority groups. Six of these relate to Black African subjects: *HLA-B*53:01* (beta = −2.628, adjusted $P = 4.69 \times 10^{-4}$); *C*04:01* (beta = −1.667, adjusted $P = .0269$); *DPA1*02:01* (beta = −1.569, adjusted $P = .0182$); and *DPA1*02:02* (beta = −1.716, adjusted $P = .0251$) were linked to decreased eGFR (based on creatinine) in Black Africans, while *DRB1*03:01* (beta = 3.200, adjusted $P = 3.99 \times 10^{-3}$) and *DPA1*01:03* (beta = 2.276, adjusted $P = 2.31 \times 10^{-5}$) were linked to increased eGFR (also based on creatinine). In Irish subjects, *HLA-DRB1*04:01* (beta = 1.075, adjusted $P = .0138$) was associated with increased eGFR (based on cystatin)and decreased risk of CKD (stage 2 or worse; odds ratio = 0.836, adjusted $P = 8.16 \times 10^{-3}$). In Indian subjects, *HLA-DRB1*03:01* (beta = −1.72, adjusted $P = 4.78 \times 10^{-3}$) and *DQB1*02:01* (beta = −1.755, adjusted $P = 2.26 \times 10^{-3}$) were associated with decreased eGFR (based on cystatin). There were no significant associations found between HLA and eGFR for any of the other three ethnic groups (Black Caribbean, Pakistani, and Chinese). Table 4 shows all alleles significantly associated with eGFR. The table also shows the frequency of the allele within the ethnic group, the direction of the allele's association with kidney function (increase or decrease), which eGFR formula the allele was associated with (creatinine or cystatin), the beta value (the size and direction of the association with eGFR) and (if applicable) the odds ratio (size and direction of the association with CKD stage 2 or worse). Adjusted and unadjusted *P* values of the associations are provided.

## 4 | DISCUSSION

The findings suggest that there may be HLA alleles which are associated with kidney function in BAME groups in the UK, especially Black Africans. Some findings replicated associations which had previously been reported. *HLA-DRB1*03* has been found to be protective against IgA nephropathy in European people[19]; this study found that *DRB1*03:01* was linked to increased renal function in Black African subjects. On the other hand, the same gene has been reported to be linked to increased risk of kidney dysfunction in Taiwanese,[20] Saudi Arabian,[21] Turkish,[13] and Brazilian[22] subjects. *This study*

**TABLE 3** Summary of calculated eGFR values

| Ethnicity | No. subjects with eGFR based on creatinine calculated | Average eGFR based on creatinine in mL/min/1.73 $m^2$ (SD) | No. subjects with eGFR based on cystatin calculated | Average eGFR based on cystatin in mL/min/1.73 $m^2$ (SD) |
|---|---|---|---|---|
| Irish | 12 191 | 92.4 (13.3) | 12 194 | 89.1 (16.2) |
| Indian | 5475 | 94.6 (14.4) | 5475 | 84.5 (18.4) |
| Black Caribbean | 4101 | 98.3 (17.7) | 4100 | 94.5 (16.7) |
| Black African | 3038 | 102.0 (18.1) | 3041 | 96.1 (16.8) |
| Pakistani | 1662 | 97.0 (17.6) | 1662 | 86.5 (17.7) |
| Chinese | 1435 | 98.2 (12.6) | 1437 | 98.9 (13.5) |

*also found that HLA-DRB1*03:01* is linked to decreased kidney function in Indian subjects. Similarly, *HLA-DRB1*04:01* has been linked to increased renal function in a study of Mexican subjects,[23] but decreased renal function in two Chinese studies[24,25] and one German study.[26] This study found *HLA-DRB1*04:01* to be associated with increased kidney function in Irish subjects. *HLA-DRB1*03:01* and *DRB1*04:01*, therefore, appear to have complex interactions with kidney function which may differ depending on ethnicity. Furthermore, our observation that *HLA-B*53:01* is linked to decreased kidney function in Black Africans replicated the findings of a study of Venezuelan subjects, for whom the allele was associated with increased ESRD.[27] Finally, our finding that *HLA-DQB1*02:01* was linked to decreased kidney function in Indians replicated a study of Chinese subjects, for whom the gene was associated with increased renal failure.[28]

Two of the alleles linked to decreased kidney function in Black African subjects (*HLA-B*53:01 and C*04:01*) form a common (9.2% frequency) haplotype in African Americans.[29] $D' = 0.824$ for the linkage disequilibrium between these genes. This suggests that the genes may not be independently associated with renal function, but represent an associated haplotype. *HLA-C*04:01* was analysed in every ethnic group except Chinese in this study but was only significant in the analysis of Black Africans. This may suggest that the gene itself does not contribute to decreased kidney function, but its apparent association is a result of its strong linkage disequilibrium with *HLA-B*53:01*. Alternatively, it may be associated with renal function in other ethnicities but this association may not have been detected due to lower frequency (and therefore reduced power) in non-Black cohorts. The other genes associated with renal function in Black African subjects do not appear to be part of well-documented haplotypes.

Six of the nine significant associations relate to creatinine in Black African subjects. A possible reason for this may be that creatinine levels are partially dependent on muscle mass and are therefore affected by ethnicity.[16] This may particularly affect estimates of GFR for Black subjects, who have higher muscle mass than other ethnic groups on average.[30] The CKD-EPI formula does adjust for ethnicity, but it is possible that this adjustment is not adequate, which could lead to false positive findings when analysing Black subjects.

The different populations showed different allele frequencies. 108 different alleles had >5% frequency in at least one ethnic group, but only 15 had >5% frequency in all six groups. As allele frequencies were so different among the different ethnic groups, it was essential to analyse ethnic groups separately in this study. The differing HLA frequencies across different populations may explain why renal dysfunction is more prevalent in some ethnic groups than others. However, Black African and Black Caribbean cohorts showed similar frequencies for many alleles, but none of the findings were replicated across the two groups. This is typical of association studies between HLA and renal function; findings are often not replicated even among apparently similar cohorts, possibly due to environmental differences or other genetic factors.

## 4.1 | Implications

These findings suggest that, even within a broadly healthy population of people, HLA alleles may have an impact on renal function. Absolute effect sizes range from 0.62 to 3.20 mL/min/1.73 $m^2$. One finding (*HLA-DRB1*04:01* being linked to increased eGFR in Irish subjects) was replicated in the secondary analysis of clinical outcomes (the gene was linked to decreased risk of CKD stage 2 or worse); this suggests that small beta values are not a contraindication of clinical relevance. HLA typing may, therefore, help predict which people are at

**TABLE 4** Alleles associated with renal function

| | | | eGFR | | | | CKD stage 2 or worse | | |
|---|---|---|---|---|---|---|---|---|---|
| **HLA-** | **Freq** | **Effect on renal function** | **Based on** | **Beta (mL/min/1.73 m$^2$)** | **Unadjusted *P* value** | **Adjusted *P* value** | **Odds ratio** | **Unadjusted *P* value** | **Adjusted *P* value** |
| Black African subjects (unadjusted $P < 9.3 \times 10^{-4}$) | | | | | | | | | |
| *B*53:01* | 14.4% | Decrease | Creatinine | −2.628 | $8.38 \times 10^{-6}$ | $4.7 \times 10^{-4}$ | Not significant | | |
| *C*04:01* | 23.4% | Decrease | Creatinine | −1.667 | $4.80 \times 10^{-4}$ | .027 | | | |
| *DRB1*03:01* | 6.6% | Increase | Creatinine | 3.2 | $7.13 \times 10^{-5}$ | .004 | | | |
| *DPA1*01:03* | 28.0% | Increase | Creatinine | 2.276 | $4.13 \times 10^{-7}$ | $2.3 \times 10^{-5}$ | | | |
| *DPA1*02:01* | 32.7% | Decrease | Creatinine | −1.569 | $3.25 \times 10^{-4}$ | .018 | | | |
| *DPA1*02:02* | 22.3% | Decrease | Creatinine | −1.716 | $4.49 \times 10^{-4}$ | .025 | | | |
| Irish subjects (unadjusted $P < 1.1 \times 10^{-3}$) | | | | | | | | | |
| *DRB1*04:01* | 9.8% | Increase | Cystatin | 1.075 | $3.01 \times 10^{-4}$ | .014 | .836 | $1.77 \times 10^{-4}$ | .008 |
| Indian subjects (unadjusted $P < 9.4 \times 10^{-4}$) | | | | | | | | | |
| *DRB1*03:01* | 11.6% | Decrease | Cystatin | −1.72 | $8.69 \times 10^{-5}$ | .005 | Not significant | | |
| *DQB1*02:01* | 12.2% | Decrease | Cystatin | −1.755 | $4.11 \times 10^{-5}$ | .002 | | | |

increased or decreased risk of developing adverse clinical outcomes. HLA association typing is already used to help diagnose a number of diseases[31-33]; it may be possible to use the findings of this study to help identify patients at risk of developing renal dysfunction and provide them with accelerated clinical pathways. It may particularly be of importance in combination with other concomitant clinical conditions like diabetes and hypertension which have a direct impact on kidney function. Likewise, it may be possible to use this data to reassure potential kidney donors with very good and stable renal function about long-term renal function. However, the findings are limited and additional research is required before these changes could be enacted.

### 4.2 | Strengths and weaknesses

A strength of this study is that the primary outcome measure, eGFR, is an accurate and precise way to measure kidney function. Renal function is a continuous spectrum, and eGFR reflects this. Previous studies, which often used binary categorisations such as "ESRD or healthy", sacrificed precision by using these broad categories. The CKD-EPI method is particularly accurate when assessing renal function in healthy individuals,[34] so was the most appropriate formula for this study. Another strength of this study may be the size of the study groups (ranging from >1400 Chinese subjects to >12 000 Irish), which are larger than many previous studies which have attempted to find associations between HLA and renal function. However, most previous studies used a case-control methodology, rather than a retrospective cohort methodology, so it is difficult to directly compare sample sizes. The adjustment for multiple testing applied in this study was stringent, with low P values required for results to be declared significant. For this reason, significant results are less likely to include type I errors. Age and sex were adjusted as covariates, which further reduces the chance of type I errors.

A limitation of this study is the HLA typing method (imputation), which is not as accurate as direct genotyping.[35] The reason that imputation was used is that it is more cost-effective than laboratory-based typing, which would have been prohibitively expensive. The imputation program used by UKB, *HLA*IMP:02*, was reported to be only 62% accurate when imputing African American subjects. The same study found another HLA imputation program, SNP2HLA, to be 92% accurate.[35] However, only nine alleles were not in HWE for particular ethnic groups (while 2151 were in HWE, 99.6%), demonstrating that the frequencies of alleles were consistent with expected frequencies. This contributed to an increased confidence in HLA assignment.

Although this study includes six different ethnic groups (whereas most of the existing literature focuses on only a single ethnicity), there were many ethnic groups excluded from the analysis. Repeating the testing with more ethnicities may lead to further discoveries, since it appears that HLA associations with renal function vary according to ethnic background. The study also used self-reported ethnicity, which may be less accurate than evaluations of ethnicity based on more thorough genetic analysis. The UKB dataset may also not be representative of the respective populations. This would limit the extent to which findings can be applied to the populations as wholes. Finally, although study groups were relatively large, the lack of significant findings suggests that larger samples may be required in order to reliably detect associations, particularly if the associations are small.

## 5 | CONCLUSION

This study identified nine HLA associations with eGFR in people from BAME backgrounds in the UK. Six of these relate to Black African, two to Indian and one to Irish populations. One finding was replicated in the secondary analysis of clinical outcomes. No associations were found for any of the other three ethnicities studied. All of the alleles studied were in HWE and were present in >5% of the relevant population. It is possible that there are further HLA associations with renal function but this study's cohort may not have been large enough to detect these; future analysis with larger study groups and a broader range of ethnicities may reveal additional associations.

### CONFLICT OF INTEREST

The authors have declared no conflicting interests.

### AUTHOR CONTRIBUTIONS

The study was devised by Kay Poulton, William Ollier, Arpana Verma, Antony Payton, Titus Augustine, and Marcus Lowe. Data was provided by UK Biobank. Quality control and data analysis was performed by Marcus Lowe with advice from Arpana Verma, Isla Gemmell, and Patrick Hamilton. Results were interpreted by Marcus Lowe, Arpana Verma, Kay Poulton, and Judith Worthington. The paper was written by Marcus Lowe with advice from all the authors.

### DATA AVAILABILITY STATEMENT

Data was provided by UK Biobank. Funding was provided by Kidneys for Life. Data cannot be shared as

access is restricted to individuals named on the application to UK Biobank.

## ORCID

*Marcus Lowe* https://orcid.org/0000-0002-0799-1674

## REFERENCES

1. Mortality GBD. Causes of death C. global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the global burden of disease study 2015. *Lancet*. 2016; 388(10053):1459-1544.
2. Luyckx VA, Tonelli M, Stanifer JW. The global burden of kidney disease and the sustainable development goals. *Bull World Health Organ*. 2018;96(6):414-422D.
3. NHS. Chronic kidney disease in england: the human and financial cost. 2012.
4. NHS. Organ donation and transplantation data for Black, Asian and Minority Ethnic (BAME) communities 2019; https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/17496/organ-donation-and-transplantation-bame-activity-report-2018-2019.pdf. Accessed 12 March, 2020.
5. UK Renal Registry. UK Renal Registry 21st Annual Report 2019; https://www.renalreg.org/wp-content/uploads/2019/05/21st_UKRR_Annual_Report.pdf. Accessed 2 June, 2020.
6. Dreyer G, Hull S, Aitken Z, Chesser A, Yaqoob MM. The effect of ethnicity on the prevalence of diabetes and associated chronic kidney disease. *QJM*. 2009;102(4):261-269.
7. Mathur R, Dreyer G, Yaqoob MM, Hull SA. Ethnic differences in the progression of chronic kidney disease and risk of death in a UKdiabetic population: an observational cohort study. *BMJ Open*. 2018;8(3):e020145.
8. Canadas-Garre M, Anderson K, Cappa R, et al. Genetic susceptibility to chronic kidney disease: some more pieces for the heritability puzzle. *Front Genet*. 2019;10:453.
9. Robson KJ, Ooi JD, Holdsworth SR, Rossjohn J, Kitching AR. HLA and kidney disease: from associations to mechanisms. *Nat Rev Nephrol*. 2018;14(10):636-655.
10. Crux NB, Elahi S. Human leukocyte antigen (HLA) and immune regulation: how do classical and non-classical HLA alleles modulate immune response to human immunodeficiency virus and hepatitis C virus infections? *Front Immunol*. 2017;8:832.
11. Xu X, Eales JM, Akbarov A, et al. Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nat Commun*. 2018;9(1):4800.
12. Davood PP, Farhadi N, Najafizadeh M. Protective and susceptible HLA class I genes in patients with end-stage renal disease. *Res J Biol Sci*. 2008;3(11):1344-1346.
13. Karahan GE, Seyhun Y, Oguz FS, et al. Impact of HLA on the underlying primary diseases in Turkish patients with end-stage renal disease. *Ren Fail*. 2009;31(1):44-49.
14. Bycroft C, Freeman C, Petkova D, et al. The UKBiobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
15. Kotsch K, Blasczyk R. The noncoding regions of HLA-DRB uncover interlineage recombinations as a mechanism of HLA diversification. *J Immunol*. 2000;165(10):5664-5670.
16. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-29.
17. Levey AS, Eckardt KU, Tsukamoto Y, et al. Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). *Kidney Int*. 2005;67(6):2089-2100.
18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.
19. Doxiadis II, de Lange P, de Vries E, Persijn GG, Claas FH. Protective and susceptible HLA polymorphisms in IgA nephropathy patients with end-stage renal failure. *Tissue Antigens*. 2001; 57(4):344-347.
20. Dai CS, Chu CC, Chen SF, Sun CY, Lin M, Lee CC. Association between human leucocyte antigen subtypes and risk of end stage renal disease in Taiwanese: a retrospective study. *BMC Nephrol*. 2015;16(1):177.
21. Hamdi NM, Al-Hababi FH, Eid AE. HLA class I and class II associations with ESRD in Saudi Arabian population. *PLoS One*. 2014;9(11):e111403.
22. Yamakawa RH, Saito PK, da Silva Junior WV, de Mattos LC, Borelli SD. Polymorphism of leukocyte and erythrocyte antigens in chronic kidney disease patients in southern Brazil. *PLoS One*. 2014;9(1):e84456.
23. Perez-Luque E, Malacara JM, Olivo-Diaz A, et al. Contribution of HLA class II genes to end stage renal disease in Mexican patients with type 2 diabetes mellitus. *Hum Immunol*. 2000;61 (10):1031-1038.
24. Cao Q, Xie D, Liu J, et al. HLA polymorphism and susceptibility to end-stage renal disease in Cantonese patients awaiting kidney transplantation. *PLoS One*. 2014;9(6):e90869.
25. Chang DY, Luo H, Zhou X, Chen M, Zhao MH. Association of HLA genes with clinical outcomes of ANCA-associated vasculitis. *Clin J Am Soc Nephrol*. 2012;7(8):1293-1299.
26. Gencik M, Borgmann S, Zahn R, et al. Immunogenetic risk factors for anti-neutrophil cytoplasmic antibody (ANCA)-associated systemic vasculitis. *Clin Exp Immunol*. 1999;117(2): 412-417.
27. Rivera PS, Marquez G, Cipriani AM, et al. HLA class I association with progression to end-stage renal disease in patients from Zulia, Venezuela. *Inmunologia*. 2012;31(2):37-42.
28. Liu H, Wang B, Liu D, et al. Molecular genetic studies on relationships among longevity, diseases, and HLA-DRB1/DQB1 allelic polymorphism. *Exp Aging Res*. 2007;33(2):123-125.
29. Allele Frequencies. USA African American pop 4. 2006; http://www.allelefrequencies.net/pop6001c.asp?pop_id=2419. Accessed 12 March, 2020.
30. Udler MS, Nadkarni GN, Belbin G, et al. Effect of genetic African ancestry on eGFR and kidney disease. *J Am Soc Nephrol*. 2015;26(7):1682-1692.
31. Kaukinen K, Partanen J, Maki M, Collin P. HLA-DQ typing in the diagnosis of celiac disease. *Am J Gastroenterol*. 2002;97(3): 695-699.
32. Sheridan DP, Lane PR, Irvine J, Martel MJ, Hogan DJ. HLA typing in actinic prurigo. *J Am Acad Dermatol*. 1990;22(6 Pt 1): 1019-1023.
33. Chen B, Li J, He C, et al. Role of HLA-B27 in the pathogenesis of ankylosing spondylitis. *Mol Med Rep*. 2017;15(4):1943-1951.

34. Florkowski CM, Chew-Harris JS. Methods of estimating GFR: different equations including CKD-EPI. *Clin Biochem Rev*. 2011;32(2):75-79.
35. Karnes JH, Shaffer CM, Bastarache L, et al. Comparison of HLA allelic imputation programs. *PLoS One*. 2017;12(2): e0172444.

**How to cite this article:** Lowe M, Payton A, Verma A, et al. Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom. *HLA*. 2020; 1–12. https://doi.org/10.1111/tan.14078

# 13 Human leukocyte antigen associations with Covid-19 in the United Kingdom

Marcus Lowe, Dr Antony Payton, Prof Arpana Verma, Dr Judith Worthington, Prof Kay Poulton

This paper was not submitted for peer-review as a similar study using the same data was released in pre-print form in May 2020[101], before we were ready to submit. There have been minor updates to the manuscript since then (such as the number of Covid-19 cases and deaths) but most of the content has not changed.

## 13.1 Introduction

The UK has been severely affected by the Covid-19 pandemic. As of September 2022, over 22m cases had been reported, with over 200,000 deaths associated with the disease[104]. The disease is particularly dangerous to elderly people, those with underlying health conditions[105], and certain ethnic groups[106]. Understanding the factors that make people more or less susceptible to Covid-19 could help to shield particularly vulnerable people without placing unnecessary restrictions on people who are at reduced risk of the disease.

Given the importance of an individual's immune response in determining disease susceptibility and severity, it is possible that genetic variations in immune response genes may be implicated. Human leukocyte antigens (HLA), encoded within the major histocompatibility complex, play a role in immunity and susceptibility towards a number of diseases including respiratory[107-109]. HLA types may be associated with Covid-19 outcomes, although research in this area is extremely limited. Nguyen found that *HLA-B*46:01* may be linked to increased risk of Covid-19 as this allele had the fewest predicted binding peptides for SARS-CoV-2, while *HLA-B*15:03* may confer a protective effect as it was more able to present highly conserved SARS-CoV-2 peptides[110]. Giamarellos-Bourboulis found that Covid-19 patients who developed severe respiratory failure had lower HLA-DR expression than healthy controls[111]. In a recent paper, Shi concluded that it is "imperative" to study whether HLA loci are associated with the disease[112]. Here we present the results of a case-control study of over 700 Covid-19 cases and 387,000 controls. Subjects are of white British ethnicity and aged 50-86. The study aims to discover HLA types which are associated with either increased or decreased risk of Covid-19 infection.

## 13.2 Methods

### 13.2.1 Quality control and identification of study population

This study used data from UKB to test for associations between HLA type and Covid-19 in 756 Covid-19 positive subjects, 1,449 Covid-19 negative subjects, and 387,356 controls who

had not been tested. UKB holds data on 502,616 subjects who were aged 39-73 when they were recruited in the UK between 2006-2010. In the entire cohort, 3,002 subjects had been tested for Covid-19 infection (though some were removed for quality control reasons); 1,073 had at least one positive test result, while 1,929 tested only negative. Many of the subjects with negative test results were probably tested because they were suspected of having Covid-19, perhaps because they were displaying symptoms of the disease or because they had had close contact with a person who had the disease. Some may have tested negative because they had already recovered from Covid-19, and some may have been false negatives caused by equipment error, human error, or the subject being tested at the wrong stage of their infection. These subjects were therefore analysed as a third category, alongside subjects who tested positive and subjects who were never tested.

Of the entire cohort (n=502,616), 23,568 subjects were removed for the following quality control reasons. 14,423 subjects were known to have died prior to the Covid-19 pandemic, so could not be included in the healthy control group. 109 subjects withdrew their consent to be included in the study. 806 had ambiguous genetic sex; these subjects were removed as sex was used as a covariate in the analysis. 7,761 were related to other subjects in the cohort (kinship coefficient ≥0.044; the individual with the most complete set of genetic data was included), while 469 were removed at UKB's recommendation due to "missing genetic data". This left 479,048 subjects in the cohort.

After quality control, 72.0% of the Covid-19 positive group was of white British ethnicity, compared to 76.8% of the Covid-19 negative group and 81.4% of the control group, highlighting ethnicity variation. To reduce population stratification influence and maintain statistical power, all subjects who were not white British according to principal component analysis were excluded from the analysis (n=89,487). In total, 389,561 subjects were included in the final study group for analysis (756 Covid-19 positive, 1,449 Covid-19 negative, and 387,356 controls). Demographic data for the three groups can be seen in Table 2.

Table 2: Demographic data for subjects included in analysis

| | Covid-19 positive | Covid-19 negative | Healthy controls |
|---|---|---|---|
| Number of subjects | 756 | 1,449 | 387,356 |
| Mean age (SD) | 69.5 (9.1) | 69.3 (8.7) | 68.6 (8.0) |
| Male | 54.6% | 46.4% | 45.5% |

This table shows the demographics of subjects in each group. SD – standard deviation

### 13.2.2 Covid-19 testing

Data on subjects tested for Covid-19 infection was provided by UKB. 5,356 tests were conducted on a total of 3,002 individuals between 16th March 2020-3rd May 2020. According to UKB, the "vast majority" of tests were carried out by polymerase chain reaction[113]. 93

independent laboratories performed the tests. Samples were taken from 26 locations on subjects' bodies; the most common location was upper respiratory tract (31.0%).

### 13.2.3 HLA typing

Subjects were HLA typed by UKB using single nucleotide polymorphism imputation. HLA*IMP:02 software was used to impute the following HLA loci: HLA-A, B, C (class I) and DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4, DRB5 (class II) at a level equivalent to high resolution typing using eight reference datasets. 362 HLA types were imputed. To avoid testing rare alleles, which could lead to misleading results, only alleles with >5% frequency were considered for analysis (n=52). However, two of these were excluded due to their frequencies not being in Hardy-Weinberg equilibrium (HWE), P<0.00014.

### 13.2.4 Regression analysis

Logistic regression analysis was used to test for associations between HLA alleles and Covid-19 infection, adjusting for age and sex as covariates. Three tests were performed using the cohorts described above: i. comparison of HLA frequencies of Covid-19 positive subjects (n=756) with the frequencies of controls who had not had a Covid-19 test (n=387,356); ii. comparison of Covid-19 positive subjects (n=756) with subjects who had tested negative for Covid-19 (n=1,449); iii. comparison of Covid-19 negative subjects (n=1,449) with controls who had not had a test (n=387,356). A Bonferroni threshold for significance of P<0.001 was applied to reflect the fact that 50 alleles were analysed (0.05/50). Analysis was performed using Plink software.

## 13.3 Results

### 13.3.1 HLA allele frequencies

52 alleles imputed by UKB had >5% frequency in the cohort as a whole. Two of these were excluded from the analysis due to not being in HWE (P<0.00014): *HLA-DQB1*02:02* (9.3% frequency in the cohort) and *DPB1*03:01* (9.1% frequency). The remaining 18 class I alleles and 32 class II alleles were included in all analyses. The alleles with >5% frequency are shown in Table 3, along with their frequencies in each of the three test groups (Covid-19 positive subjects, Covid-19 negative subjects, and subjects not tested for Covid-19).

Table 3: HLA types with frequency >5%

| Class I | | | |
|---|---|---|---|
| | Frequency among | | |
| *HLA-* | Covid-19 pos | Covid-19 neg | Not tested |
| *A*02:01* | 27.0% | 27.9% | 27.5% |

| Class II | | | |
|---|---|---|---|
| | Frequency among | | |
| *HLA-* | Covid-19 pos | Covid-19 neg | Not tested |
| *DPB1*04:01* | 43.0% | 43.2% | 43.5% |

| | | | |
|---|---|---|---|
| *A*01:01* | 20.1% | 18.8% | 19.4% |
| *C*07:01* | 17.6% | 17.3% | 17.6% |
| *C*07:02* | 16.4% | 16.1% | 15.8% |
| *B*07:02* | 14.9% | 15.5% | 14.8% |
| *A*03:01* | 15.7% | 14.0% | 14.5% |
| *B*08:01* | 14.4% | 14.4% | 14.4% |
| *C*05:01* | 10.6% | 11.4% | 11.4% |
| *B*44:02* | 9.9% | 11.4% | 11.2% |
| *C*06:02* | 9.9% | 9.6% | 9.0% |
| *C*04:01* | 8.1% | 9.2% | 8.3% |
| *C*03:04* | 7.8% | 7.5% | 8.0% |
| *A*24:02* | 6.6% | 7.4% | 7.3% |
| *B*15:01* | 6.2% | 6.2% | 6.1% |
| *A*11:01* | 6.1% | 6.1% | 6.0% |
| *B*44:03* | 5.4% | 6.2% | 5.9% |
| *B*40:01* | 5.5% | 5.3% | 5.6% |
| *C*03:03* | 4.8% | 5.3% | 5.5% |

| | | | |
|---|---|---|---|
| *DRB4 – no gene* | 32.2% | 35.0% | 35.0% |
| *DRB3 – no gene* | 34.7% | 33.8% | 34.3% |
| *DRB4*01:03* | 23.5% | 24.7% | 25.4% |
| *DQA1*05:01* | 22.9% | 23.3% | 23.0% |
| *DQA1*03:01* | 17.3% | 20.4% | 20.4% |
| *DQA1*01:02* | 20.2% | 18.3% | 19.1% |
| *DPA1*01:03* | 18.3% | 18.2% | 18.5% |
| *DQB1*03:01* | 16.6% | 17.8% | 17.5% |
| *DRB3*01:01* | 16.7% | 16.3% | 16.7% |
| *DRB5 – no gene* | 16.4% | 15.4% | 15.3% |
| *DQB1*02:01* | 15.2% | 15.4% | 14.9% |
| *DRB1*03:01* | 15.1% | 15.2% | 14.9% |
| *DPA1*02:01* | 14.2% | 14.7% | 14.6% |
| *DQA1*02:01* | 14.4% | 14.9% | 14.6% |
| *DRB1*07:01* | 14.1% | 14.9% | 14.6% |
| *DRB1*15:01* | 15.3% | 14.7% | 14.5% |
| *DRB5*01:01* | 15.2% | 14.6% | 14.4% |
| *DQB1*06:02* | 15.3% | 14.5% | 14.2% |
| *DQA1*01:01* | 15.0% | 14.4% | 14.2% |
| *DRB3*02:02* | 13.7% | 14.2% | 13.3% |
| *DQB1*05:01* | 12.5% | 12.2% | 12.0% |
| *DRB1*04:01* | 10.1% | 11.3% | 11.1% |
| *DPB1*02:01* | 12.2% | 10.9% | 10.8% |
| *DPB1*04:02* | 9.9% | 11.6% | 10.7% |
| *DQB1*03:02* | 8.5% | 10.2% | 10.4% |
| *DRB1*01:01* | 9.5% | 9.0% | 9.4% |
| *DQB1*02:02* † | 9.4% | 9.9% | 9.3% |
| *DPB1*03:01* † | 9.2% | 8.4% | 9.1% |
| *DRB4*01:01* | 7.3% | 8.4% | 8.1% |
| *DPB1*01:01* | 5.4% | 5.7% | 5.9% |
| *DQA1*01:03* | 6.3% | 5.4% | 5.4% |
| *DQB1*03:03* | 4.9% | 5.1% | 5.2% |
| *DQB1*06:03* | 6.0% | 5.0% | 5.1% |

†not in Hardy-Weinberg equilibrium so excluded from analysis

This table shows the 18 class I and 34 class II types which have frequency >5% in the cohort. They are split by class and sorted in descending order of frequency, and alleles which were not in Hardy-Weinberg equilibrium are noted.

### 13.3.2 Regression analysis

After correction for multiple testing, no HLA types were associated with Covid-19 infection in any of the analyses performed. This suggests that either HLA is not associated with likelihood of Covid-19 infection, or that this study was unable to detect associations that may be present. One allele trended towards significance (P<0.01): *HLA-DQA1*03:01*, which may be lower in Covid-19 positive subjects than subjects who had not been tested for Covid-19 (P=0.0038, odds ratio 0.82, 95% confidence interval 0.72-0.94). However, due to Bonferroni correction for multiple tests, this association cannot be considered statistically significant. The allele is seen

in 20.4% of subjects not tested for Covid-19 and 20.4% of subjects who tested negative for Covid-19, but only 17.3% of Covid-19 positive subjects.

## 13.4 Discussion

The findings suggest that no HLA types appear to be associated with increased or decreased risk of Covid-19 infection in white British subjects aged 50-86. This may be due to insufficient power; as more cases of Covid-19 are discovered and reported to UKB, the case group may become large enough to detect HLA associations.

This study includes only white British subjects as there was not enough data on any other ethnicity to conduct a meaningful analysis. Other than white British, the ethnicities with the largest number of subjects positive for Covid-19 were Caribbean (35 positive, 30 negative), Irish (34 positive, 60 negative), and “any other white background” (32 positive, 67 negative). As the study only included white British people aged 50-86, the findings can only be applied to this group. However, people from ethnic minorities are worse affected by the disease[106]; future studies should include a wider range of ethnicities to ensure that findings can be applied to more of the population, and possibly reveal associations which are specific to particular ethnicities. Similarly, studies of a wider range of ages may help to increase representativeness of any findings.

### 13.4.1 *HLA-DQA1*03:01*: a nonsignificant trend

*HLA-DQA1*03:01* trended towards significance and may prove to be linked to reduced risk of Covid-19. This allele is more common among white people than ethnic minority groups in the UK, except for Chinese people. Within the UKB cohort, the allele was seen in 20.4% of white British subjects and 18.1% of Irish subjects, compared to subjects who self-reported their ethnicity as Indian (9.5% frequency of *HLA-DQA1*03:01*), Caribbean (7.9%), African (7.4%), Pakistani (7.3%), and Chinese (28.1%). If this allele does confer a protective effect against Covid-19 then its frequency in these populations may contribute towards the observations that Black, Bangladeshi/Pakistani, and Indian people are all more likely to die of the disease in the UK than white people, while Chinese people are not at increased risk[106].

*HLA-DQA1*03:01* is associated with increased risk of autoimmune diseases such as type 1 diabetes[114] and coeliac disease[115,116]. It is possible that the allele is not protective against Covid-19 itself, but rather against the symptoms of the disease. The allele may be linked to increased immune response: this may explain both the lower rates of Covid-19 and the increased risk of autoimmune disease[117,118].

HLA association typing is already used to help diagnose several diseases; it may be possible to use association studies to help identify patients at increased risk of developing Covid-19

and provide them with accelerated clinical pathways. However, the current findings are limited as no allele was associated with increased risk of Covid-19; additional research is required before such applications could be enacted. Such work may be able to confirm or refute the possibility that *HLA-DQA1*03:01* is protective against Covid-19.

### 13.4.2 Strengths and weaknesses

A strength of this study is that the statistical methods are robust enough to limit the risk of type I errors. Including only HLA alleles with >5% frequency reduces the risk of misleading results caused by a small number of subjects, which could happen if rare alleles were analysed. Excluding alleles which were not in HWE also reduces the risk of inaccurate results by ensuring that alleles are only analysed if their frequencies are consistent with frequencies that might be expected in the general population. The adjustment for multiple tests and adjustment for covariates (age and sex) also reduces the risk of type I errors. The statistical methods might be improved in future analysis by matching cases to controls to ensure that the control group accurately reflects the case group.

Another strength is that the ethnicity of the subjects was confirmed by genetic principal component analysis; this is more accurate than self-reported measures of ethnicity. Furthermore, the study focuses on subjects aged over 50; older people have higher mortality rates from Covid-19[105] so it is particularly important to focus research on how Covid-19 affects middle-aged and elderly people. Similarly, males are more likely than females to die from Covid-19[106]; this is reflected in the fact that the Covid-19 positive group was 54.6% male, compared to 45.5% in the control group.

A limitation of this study is that the HLA typing method (imputation) is not as accurate as direct genotyping[53]. The reason that imputation was used is that it is more cost-effective than laboratory-based typing, which would have been prohibitively expensive. The imputation program used was HLA*IMP:02, though Karnes' review of competing programs suggests that SNP2HLA is more accurate[53]. Nevertheless, the review stated that HLA*IMP:02 is 94% accurate when imputing white subjects which is acceptable within the scope of this study, and Bycroft estimated the accuracy of UKB's imputed data to be >96%[62]. Furthermore, 360 of the 362 imputed alleles were in HWE (P>0.00014), suggesting that the majority of imputed allele frequencies were consistent with frequencies that might be expected in a stable population. The two alleles which were not in HWE (*HLA-DQB1*02:02* and *DPB1*03:01*) were excluded from the analysis. Another limitation is that the cohort may not have been large enough to detect small genetic effects. As more data emerges, both from UKB and other sources, studies may have greater power to find associations; in particular, they may be able to confirm or refute an association with *HLA-DQA1*03:01*.

It is probable that some subjects included in the healthy control group were miscategorised. Some of them may have been susceptible to the disease but had not been exposed at the time of analysis, some may have had the disease but not had a test (for example, because they were asymptomatic, or because tests were not widely available so early in the pandemic), and some positive test results may not have been received by UKB. These confounding factors may mean that this study assesses merely which subjects were exposed to or tested for the disease, rather than which subjects were susceptible to it. Furthermore, some subjects may have died prior to the pandemic but UKB may have been unaware of this; these subjects should have been removed during the quality control stage. As time progresses and UKB receive more data, it may be possible to identify more subjects who belong in the case group and reduce the number of miscategorisations in the control group. Larger cohorts may be required to find associations (particularly if effect sizes are small). More accurate identification of cases and controls may be valuable; ideally, everyone in the cohort should be tested simultaneously, whether they are symptomatic or not.

Rather than searching for HLA associations with Covid-19 infection (based on PCR test results, as presented here), it may be preferable to investigate whether any HLA types are associated with increased or decreased presence of antibodies against Covid-19 (which would suggest an increased or decreased ability to respond to the disease). HLA enables immune response: it may not have any impact on the presence of the disease itself (in which case no association would be found between HLA and Covid-19 test results), but certain HLA types may nevertheless be associated with increased or decreased immune response to the disease. Analysing the presence of antibodies would reveal whether subjects had had a Covid-19 infection at any point, whereas the method presented here was only capable of identifying subjects who were infected at the time of testing. Analysing antibodies, therefore, would be expected to expand the “case” group and provide a more accurate reflection of which subjects were susceptible to the disease. Unfortunately, data on whether subjects had antibodies to Covid-19 was not available from UKB so it was not possible to perform this analysis.

Finally, the outcome data was quite basic. Subjects were categorised based on their Covid-19 test results, but there was no survival data, nor any distinction between subjects based on disease severity. More detailed data, such as death rates, hospital records, and categorisation based on symptoms, may be useful in gaining a more detailed understanding of any link between HLA and Covid-19.

## 13.5 Conclusion

This study did not identify any HLA alleles which are associated with risk of Covid-19 infection in white British people aged 50-86 in the UK. All alleles studied were in HWE and were present

in >5% of the cohort. It is possible that there are HLA associations with Covid-19 but this study's cohort may not have the statistical power to detect these; in particular, *HLA-DQA1*03:01* may be associated with reduced risk of testing positive for the disease. As more cases emerge, studies using larger cohorts, more accurate and sophisticated categorisation of subjects, and a broader range of ages and ethnicities may reveal associations.

# 14 Kidney function is not associated with Covid-19 susceptibility or mortality: a UK Biobank study of middle and old age, white British individuals

Marcus Lowe, Dr Antony Payton, Prof Arpana Verma, Dr Judith Worthington, Prof Kay Poulton

This paper was submitted to Kidney International Reports on 12th February 2021 and to Nephron on 17th February 2021 but was not selected for publication. There have been minor updates since then (such as the number of Covid-19 cases and deaths) but most of the content has not changed.

## 14.1 Abstract

There may be a link between Covid-19 and kidney function, but this has not yet been clearly established.

This study uses UKB data on over 6,000 white British Covid-19 positive cases and over 30,000 white British Covid-19 negative controls. Logistic regression was used to test for associations between Covid-19 infection and the following factors: kidney function (measured primarily by eGFR) and associated susceptibility factors including age; sex; body mass index (BMI); smoking status; and household deprivation (measured by Townsend deprivation index). Two outcome measures were used: death within 28 days of Covid-19 diagnosis (247 cases, 5,976 controls); and Covid-19 test result (6,223 Covid-19 positive cases, 30,531 Covid-19 negative controls).

Kidney function was not linked to susceptibility to Covid-19, nor mortality from Covid-19. Factors associated with increased Covid-19 mortality were: older age (OR 1.15, $P<1\times10^{-10}$); male sex (OR 1.51, P=0.006); and increased BMI (OR 1.037, P=0.013). Factors associated with increased susceptibility to Covid-19 were: younger age (OR 0.946, $P<1\times10^{-10}$); male sex (OR 1.166, $P=2.4\times10^{-7}$); increased deprivation (OR 1.028, $P=1.0\times10^{-8}$); and increased BMI (OR 1.018, $P=1.5\times10^{-9}$).

After adjusting for confounding variables, kidney function was not shown to be linked to Covid-19 susceptibility or severity. A number of factors were linked to the disease, replicating previously published findings.

## 14.2 Introduction

The UK has been severely affected by the Covid-19 pandemic. By September 2022, over 22m cases had been reported[119], resulting in over 200,000 deaths[120]. Elderly people, those with underlying health conditions[105], and certain ethnic groups[106] are most likely to experience more

severe symptoms and a higher rate of mortality. Understanding the factors that determine susceptibility to Covid-19 could highlight vulnerable people who would benefit from shielding without placing unnecessary restrictions on others who are at reduced risk.

Poor kidney function is known to be linked to a reduced immune response[121,122]. Research suggests that individuals with lower levels of kidney function or kidney disease are at increased risk of mortality from Covid-19[93], and kidney transplant recipients with Covid-19 showed faster disease progression than the general population[123]. Although kidney disease appears to be a predictor of Covid-19-related mortality, research into whether decreased kidney function is associated with increased risk of Covid-19 infection is limited. Similarly, Covid-19 may lead to, or exacerbate, kidney disease[124,125], although long-term studies of this effect have not yet been conducted.

Here we present the results of a case-control study of over 6,000 Covid-19 cases and 30,000 Covid-19-negative controls from UKB[126]. Subjects are of white British ethnicity and aged 50-86 years. The study aims to investigate whether lower kidney function (determined by eGFR) is linked to increased risk of testing positive for the disease and worse disease outcome.

## 14.3 Methods

### 14.3.1 Quality control and identification of study population

This study used data from UKB to test for associations between Covid-19 and kidney function in 6,223 Covid-19 positive subjects (247 of whom died within 28 days of testing positive)[113]. UKB holds data on 502,616 subjects who were aged 39-73 when they were recruited in the UK between 2006-2010. 43,013 of these had been tested for Covid-19; 7,394 had at least one positive test result, while 35,619 tested negative.

94,367 subjects were removed for the following quality control reasons. 27,953 subjects were known to have died prior to the Covid-19 pandemic. 105 withdrew their consent to be studied. 849 had ambiguous genetic sex. 8,334 were related to other subjects in the cohort (kinship coefficient ≥0.044), while 479 were removed at UKB's recommendation due to "missing data". Finally, 56,647 subjects whose self-reported ethnicity was not "white British" were excluded, as ethnicity is known to affect both kidney function and Covid-19 susceptibility. Ethnic groups other than "white British" had insufficient Covid-19 cases to conduct a sufficiently powered independent analysis.

Of the remaining 408,249 subjects in the cohort, 371,495 were never tested (91%). These subjects were excluded from the analysis since their Covid-19 status was unknown. This left a cohort of 36,754 subjects: 30,531 tested negative (83.1%), 5,976 tested positive and were not known to have died within 28 days (16.3%), and 247 tested positive and died within 28

days (0.67%). The demographics of these three groups can be seen in Table 4. All subjects were aged 50-86 in October 2020.

Table 4: Characteristics of subjects studied

| | Tested negative (n=30,531) | Tested positive, survived (n=5,976) | Tested positive, died (n=247) |
|---|---|---|---|
| Average age (SD) | 70.6 (7.9) | 66.4 (8.7) | 75.2 (6.0) |
| % male | 47.0% | 48.7% | 62.8% |
| % never smoked | 37.0% | 37.7% | 29.1% |
| Average BMI (SD) | 27.9 (5.0) | 28.3 (4.9) | 29.5 (5.8) |
| Average Townsend deprivation index (SD) | -1.42 (3.0) | -1.06 (3.1) | -0.60 (3.3) |
| Average eGFR (SD) | 87.8 (15.0) | 91.7 (14.8) | 81.2 (15.5) |
| % eGFR >90 (good kidney function) | 50.1% | 61.0% | 35.2% |
| % eGFR <60 (poor kidney function) | 3.8% | 2.4% | 7.3% |

This table summarises the age, sex, smoking status, BMI, deprivation, and kidney function of the three cohorts (subjects who tested negative, subjects who tested positive and survived, and subjects who tested negative and died). SD – standard deviation; BMI – body mass index; eGFR – estimated glomerular filtration rate

### 14.3.2 Covid-19 testing and death data

Data on subjects tested for Covid-19 was provided by Public Health England (via UKB[113]). 75,395 tests were performed on a total of 43,013 individuals between 16th March 2020 and 30th November 2020. 9,903 positive results were reported, relating to 7,394 individuals. According to UKB, the "vast majority" of tests were carried out by polymerase chain reaction[113]. 143 laboratories performed the tests. Samples were taken from 38 locations on subjects' bodies; the most common specimen type was nasal swab (31.9%). UKB received notification and cause of death for deceased subjects directly from national death registries[127].

### 14.3.3 Regression analysis

Logistic regression analyses were used to test for associations between Covid-19 and kidney function. The first analysis used Covid-19 mortality as an outcome measure: 247 subjects who died within 28 days of a positive Covid-19 test were compared with 5,976 subjects who tested positive but did not die. This analysis aimed to discover factors associated with worse or better disease outcome. The second analysis used Covid-19 test result as an outcome measure: 6,223 Covid-19 positive subjects (regardless of mortality) were compared with 30,531 subjects who tested negative for Covid-19. This aimed to discover factors which are linked to risk of being infected with the disease.

Kidney function was measured using eGFR (at the time of recruitment). eGFR is a measure of kidney function based on levels of creatinine and cystatin in serum[128,129]. It is a continuous scale measured in ml/min/1.73m$^2$. The volunteers' eGFRs ranged from 4 (low kidney function)

to 159 (high kidney function) and the average was 91 (SD 13.8). eGFR >90 is generally considered to be healthy, while eGFR <60 is indicative of chronic kidney disease (at least stage 3)[130]. eGFR <15 is indicative of end-stage renal disease, requiring renal replacement therapy such as dialysis or transplantation[3]. The two regression analyses described above used eGFR as a continuous measurement of kidney function. In a secondary analysis, subjects were split into categories based on their kidney functions: subjects with CKD stage 3 or worse (eGFR <60) were compared to healthy controls (subjects with eGFR >90) using logistic regression. The outcome measure was Covid-19 test result (6,223 positive subjects, 30,531 negative subjects).

Five confounders were included in the model. These comprised: age (as of October 2020); sex; BMI (measured at the time of recruitment, 2006-2010, the average was 27.4, SD 4.7); whether the subject had ever smoked (at the time of recruitment, 60% had smoked and 40% had not); and the Townsend deprivation index (at the time of recruitment). The Townsend deprivation index is a measure of deprivation in the subject's home postcode, based on factors such as unemployment, non-car ownership, non-home ownership, and household overcrowding[131,132]. It is a continuous scale; this study's subjects ranged from -6.3 (least deprived) to 10.8 (most deprived), with an average of -1.57 (SD 2.9). A Bonferroni threshold for significance[133] of $P<0.025$ was applied to reflect that two main analyses were performed (investigating disease progression and disease susceptibility). Analysis was performed using Stata/SE 13.0 (StataCorp).

## 14.4 Results

### 14.4.1 Disease progression

In subjects who tested positive for Covid-19, three factors were independent predictors of death within 28 days ($P<0.025$). Age is linked to increased risk of death (OR 1.15, suggesting that each extra year of age is associated with 15% higher chance of dying after a Covid-19 diagnosis, $P<1\times10^{-10}$). Males are 51% more likely than females to die from Covid-19 (OR 1.51, $P=0.006$). Increased BMI was linked to increased mortality (OR 1.037, suggesting each extra BMI point increased risk by 3.7%, $P=0.013$).

Higher deprivation showed a non-significant trend linked to increased risk of death (OR 1.036, $P=0.105$). There was no association found between risk of death and either kidney function ($P=0.20$) or smoking status ($P=0.99$).

### 14.4.2 Disease susceptibility

When Covid-19 positive subjects were compared to the 30,531 subjects who tested negative, all confounders but one were independently associated with susceptibility to Covid-19

(P<0.025). Subjects who tested positive were younger than those who tested negative (OR 0.946, suggesting each extra year of age reduces risk by 5.4%, $P<1\times10^{-10}$). Males were at 17% greater risk of testing positive than females (OR 1.166, $P=2.4\times10^{-7}$). Increased deprivation increased risk (OR 1.028, suggesting each extra point on the Townsend deprivation index increased risk by 2.8%, $P=1.0\times10^{-8}$), as did increased BMI (OR 1.018, suggesting each extra BMI point increased risk by 1.8%, $P=1.5\times10^{-9}$). However, having smoked in the past did not increase risk of contracting the disease compared to people who have never smoked (P=0.57).

After adjusting for these factors, subjects who tested positive were not shown to have significantly different eGFRs than subjects who tested negative for Covid-19 (P=0.16). This suggests that kidney function is not linked to risk of contracting Covid-19. Similarly, the secondary analysis did not find that subjects with CKD stage 3-5 were more likely to test positive for Covid-19 than healthy controls (P=0.22).

## 14.5 Discussion

### 14.5.1 Implications

The findings suggest that, among white British people aged 50-86, kidney function was not linked to risk of testing positive for Covid-19, nor to mortality rates after a positive test. We suspected that there may have been associations, since decreased kidney function is known to have an adverse effect on the immune system[121]. It is possible that there is a small effect that was not detected by this study; larger studies may be able to confirm or refute this. However, based on the evidence presented here, people with reduced kidney function are not at increased risk of becoming infected with Covid-19, nor increased risk of mortality if they are infected.

A number of factors were linked to Covid-19 infection, including age, sex, BMI, and deprivation. Older people (within the cohort of 50-86-year-olds) were less likely to have a positive test. This may be because they took greater precautions to avoid the disease, knowing older people are more vulnerable to the disease[105]. This study replicated the finding that older people who test positive are more likely than younger people to die within 28 days. Males were both more likely to test positive for Covid-19 and more likely to die if they tested positive (by 16% and 51% respectively). A meta-analysis also found that males are more likely to die from Covid-19 than females[134]. Living in a deprived area was a significant risk factor for testing positive for Covid-19: people who lived in areas with high levels of unemployment and household overcrowding, and low levels of car- and home-ownership were more likely to test positive. They may be more likely to die soon after diagnosis, though this trend was not

statistically significant. Higher BMI was linked to an increased proportion of positive tests, and also increased mortality rates.

Smoking status was not associated with Covid-19. The data presented in Table 4 appears to suggest that smokers were more likely to die than non-smokers: 71% of people who died had smoked in the past, compared to only 62% of people who survived Covid-19. Before adjusting for covariates, this was a significant association (OR 1.463, P=0.008). However, when the regression analysis adjusted for covariates (age, sex, BMI, eGFR, and deprivation), this association was not significant (P=0.99). The apparent discrepancy appears to be caused by the fact that men are more likely to smoke than women (66% v 59% among people who tested positive for Covid-19), and smokers are older than non-smokers on average (67.4 years v 65.6 years among Covid-19 positive subjects). Male sex and increased age are both linked to increased Covid-19 mortality but smoking alone does not appear to be an independent predictor of Covid-19 death.

### 14.5.2 Strengths and limitations

A strength of this study is the large cohort size. Over 6,000 cases and 30,000 controls gave the study considerable statistical power to detect associations, with high confidence. The use of multiple outcome measures (testing positive for Covid-19, and mortality after a positive test) allowed us to draw conclusions about both susceptibility to, and progression of the disease. In some cases, the associations were similar (being male is linked to higher risk of getting the disease and higher risk of dying), while others appear to be contrasting (older subjects are less likely to get the disease but more likely to die from it). These two outcome measures together contribute to a clearer understanding of the disease pathway than either measure alone.

A limitation of the study is that data on eGFR, smoking, BMI and deprivation (collected 2006-2010) may have changed in the intervening years. In some cases, this is not a problem: genetic sex does not change, and current age was calculated. In other cases, this may reduce the accuracy and reliability of the findings. Furthermore, subjects' smoking status was self-reported, based upon responses to a questionnaire, which may also impact on accuracy. However, the fact that our results support previous findings suggests that the data is valid.

Another potential weakness is the possibility of missing or inaccurate outcome data. Covid-19 test results are provided to UKB by PHE. If PHE data is incomplete, or subject to delay, then some subjects who should have been categorised as "cases" of Covid-19 may not have been included in the study. Similarly, others may have had Covid-19 but were not recorded as "cases" due to not having been tested at all, or false negative test results, or tested at the

wrong stage of their infection. Others may be susceptible to Covid-19 but not exposed in the time period studied (up to the end of November 2020). As more cases are discovered and reported, the accuracy of the “case” and “control” groups is expected to increase.

## 14.6 Conclusion

This study has identified a number of factors which are linked to risk of testing positive for Covid-19 and risk of dying within 28 days of a positive test. Notably, kidney function was not linked to risk of testing positive nor to mortality. Factors that increased the risk of having a positive test were: male sex; younger age; increased BMI; and living in a deprived area. Risk factors for dying within 28 days of testing positive were: older age; male sex; and increased deprivation. Future studies may benefit from more comprehensive data and may confirm these findings and identify other factors which are linked to either susceptibility or mortality from Covid-19.

# 15 Discussion

## 15.1 What has been found?

### 15.1.1 Comparison with previously published HLA associations with renal function

Over 100 HLA types and haplotypes were found to be linked to kidney function in 35 peer-reviewed papers, including two that were published as part of this project. This study collated these findings and published them in a single place for the first time[98].

A number of the associations published previously were replicated by this study. For example, *HLA-DRB1*03* was associated with decreased kidney function in white British subjects by this study, and was also linked to kidney dysfunction by four other studies. Multiple other findings were replicated in other studies; section 10 gives a comprehensive list of these. Findings which are independently replicated are generally more valid than findings which are not replicated, particularly if the findings are discovered in well-designed studies with high statistical power.

However, some of this study's findings contradicted previous findings. For example, *HLA-DRB1*04* was linked to poor renal function by multiple previous studies (though one study found it to be protective against ESRD); in this study it was associated with increased eGFR in white British subjects and Irish subjects. *HLA-DRB1*04* has also been linked to other diseases such as Vogt-Koyanagi-Harada's disease[135] (which affects many bodily systems including vision and the nervous system) and rheumatoid arthritis[136], which is also linked to kidney disease[137]. Again, section 10 lists all contradictory findings.

Finally, some of my findings neither replicated nor refuted previous literature. A lack of concordance with previous studies does not necessarily mean that the findings are inaccurate. HLA types seem to have different effects in different populations, so it may be that the associations found can be applied only to the population studied. It may also be the case that this study is uncovering novel associations with renal function which previous studies were unable to detect.

The fact that most previously reported findings have never been replicated by any other study suggests that many may be type I errors rather than true associations, possibly due to poor methodology. Issues such as low statistical power, sub-optimal measures of renal function, and failure to adjust for multiple tests could all contribute to such errors. The discussion section of the systematic review paper (section 10) outlines these problems, while section 15.5 details the steps I took to avoid them in my study.

### 15.1.2 White British subjects

In my study of over 400,000 white British subjects, 22 HLA types were linked to increased eGFR and 11 HLA types were linked to decreased eGFR[99]. Seven of these were also associated with increased risk of end-stage renal disease and/or chronic kidney disease. Many of these types are in linkage disequilibrium (LD) with each other and are commonly inherited together. As mentioned above, some of the findings replicated previous results while some refuted them.

### 15.1.3 Subjects of other ethnicities

I investigated six further ethnicities (>12,100 Irish subjects, >5,400 Indian, >4,000 Black Caribbean, >3,000 Black African, >1,600 Pakistani, and >1,400 Chinese)[100]. There were six associations between HLA and renal function found in Black African subjects (four hazardous and two protective), while two types were found to be hazardous in Indian subjects, and one type was protective in Irish subjects. No HLA associations were detected in the remaining three ethnic groups. As above, some findings replicated previous publications (e.g., *HLA-B*53*, which was linked to reduced renal function in Black Africans as well as two previous studies), while others had not been seen before or refuted previous results (including *HLA-DRB1*03:01*: this study linked the allele to increased kidney function in Black Africans, but reduced renal function in white British and Indians).

### 15.1.4 Covid-19

The analysis of the relationship between HLA, kidney function, and Covid-19 revealed limited findings. Although I was able to replicate well-known observations regarding risk factors for Covid-19 (such as age, male sex, and increased deprivation), there were no significant associations detected between Covid-19 and HLA, nor Covid-19 and eGFR, in my studies of white British subjects aged 50-86. Kidney disease is thought to be linked to Covid-19 risk (see section 6.7), but my study did not find a link between eGFR and Covid-19. It may be that severe kidney dysfunction is hazardous but a small decrease in eGFR is not enough to contribute to Covid-19 in people whose kidneys are broadly healthy. Or it may be that my study had insufficient power to detect an effect.

Similarly, HLA type is thought to be associated with Covid-19 (though the exact alleles and their effects are not widely agreed upon, see section 6.7), but my study of white British subjects aged 50-86 did not find any alleles to be linked to Covid-19. One allele appeared to be associated with reduced risk of testing positive for Covid-19 (*HLA-DQA1*03:01*, OR 0.82, P=0.0038), but this was not significant after Bonferroni adjustment (the threshold for significance was P<0.001). A different study of 96 Covid-19 positive Indian subjects found that

this allele was linked to more severe symptoms[138]. However, the sample size was very small and it is not clear whether there was any adjustment for multiple testing. In any case, this finding does not necessarily contradict my (tentative) finding that the allele may reduce the risk of testing positive: it merely suggests that the allele is linked to worse outcomes for people who do test positive.

## 15.2 Notes

### 15.2.1 Type I and type II errors

Much of section 15 refers to “type I” and “type II” errors. “Type I error” describes a “false positive” finding. In this study, a false positive would mean that an HLA type which is not associated with renal function is incorrectly found to have an association. This could be a result of subjects being wrongly classified as a “case” of renal dysfunction, when they should be considered a “control”. This could lead to their HLA types being found to have an association with renal failure. Inversely, it is also possible that a type is wrongly found to have a protective effect against renal failure. This would be a result of subjects being wrongly considered “controls” when they should be considered “cases”.

“Type II error” describes a “false negative” finding. In this study, a false negative would mean that a type which is associated with renal function is incorrectly found to have no association. This could be a result of subjects being wrongly classified as a “control”, when they should be considered “cases” of renal dysfunction. This could lead to their HLA types being found to have no association with renal function. Inversely, a type could be found to have no association with renal function when in fact it has a protective effect against renal failure. This would be a result of subjects being wrongly considered “cases” when they should be considered “controls”. Insufficient statistical power is a major cause of both type I and type II errors; the large sample size of this study reduces the chance of these errors, though they can also result from poor study design or methodology, and biased or inaccurate data.

### 15.2.2 P values

Different P value thresholds were used to denote significance (for example, $P<0.00014$, $P<0.001$, and others). These were calculated by dividing 0.05 (the standard P value for significance) by the number of HLA types included in the analysis. This means that, when “raw” or “unadjusted” P values for HLA types are reported, it may not be clear whether they are significant without having the threshold P value in mind. Paper 3 avoided this problem by presenting the raw P value alongside an “adjusted” P value. The adjusted P value was calculated by multiplying the raw P value by the number of HLA types included in the analysis. An adjusted P value of $<0.05$ was always significant, regardless of the analysis being

performed. This made the interpretation of results simpler as it was clear “at a glance” whether the adjusted P value is significant.

## 15.3 Attempt to validate findings

### 15.3.1 Introduction

In order to validate the associations reported in this study, I investigated the HLA type frequencies of patients who received a kidney alone transplant at Manchester Royal Infirmary and Royal Manchester Children’s Hospital. For each of the alleles that this study reported to be associated with kidney function, its frequency among transplant recipients was compared to its frequency observed in the UKB cohort, which served as a “healthy control” group. Alleles that were protective according to the published papers were expected to be less common in the transplant recipient group than in the UKB cohort, while alleles that were hazardous were expected to be more frequent among transplant recipients than in the UKB cohort.

### 15.3.2 Methods

602 kidney transplant recipients had been HLA typed by NGS. Of these, seven did not have ethnicity data recorded (1.2%), while a further 76 ethnicities were recorded as “prefer not to say” (12.6%). The self-reported ethnicities of the remaining 519 patients can be seen in Table 5. Ethnicities with fewer than five patients are reported as “<5” to preserve patients’ anonymity.

Table 5: ethnicities of kidney transplant recipients

| Ethnicity | Count | Percent |
|---|---|---|
| White British | 340 | 65.5% |
| Asian Pakistani | 56 | 10.8% |
| Black African | 38 | 7.3% |
| Asian Indian | 18 | 3.5% |
| Gypsy or Irish Traveller | 11 | 2.1% |
| Other Ethnic Group | 11 | 2.1% |
| White Irish | 10 | 1.9% |
| Other Asian Background | 8 | 1.5% |
| Black Caribbean | 7 | 1.3% |
| Any White Background Not Listed | 5 | 1.0% |
| Asian Bangladeshi | <5 | <1.0% |
| Chinese | <5 | <1.0% |
| Mixed White and Asian | <5 | <1.0% |
| Mixed White and Black Caribbean | <5 | <1.0% |
| Other Black Background | <5 | <1.0% |
| Other Mixed Background | <5 | <1.0% |
| Total | 519 | 100% |

This table shows the ethnicities of kidney transplant recipients. Patients without ethnicity data are excluded (data missing n=7, “prefer not to say” n=76).

There were 33 HLA alleles linked to kidney function in white British subjects, six linked to kidney function in Black African subjects, two linked to kidney function in Indian subjects, and one linked to kidney function in white Irish subjects. Only the findings in white British subjects were validated as there was insufficient data for the other ethnic groups. There were 340 white British transplant recipients included in the validation. I performed a logistic regression for each allele to test for associations between HLA and the need for renal transplantation.

Associations relating to HLA-DRB3 and DRB4 were not included in the validation; differences in typing methodology between the UKB method (imputation) and the NGS method used for the transplant recipients meant that frequencies for alleles at these loci could not be directly compared. In the UKB cohort, the sums of the frequencies of all HLA-DRB3 and DRB4 alleles (plus the frequencies of "no gene" at these loci) were 68.1% and 68.4% respectively (rather than 100%). This is because some imputed allele calls were excluded due to the posterior probability threshold of 0.8 (see section 6.6.4). The sum of the allele frequencies in the transplant recipient population (who were typed by NGS) was >98.8% for each locus (indicating a very small number of missing values). As a result, the frequencies in the transplant population appear to be higher than the frequencies in the UKB cohort for HLA-DRB3 and DRB4, but this is misleading because the UKB frequencies are artificially reduced by allele calls being excluded. Therefore, for these loci, the frequencies in the UKB cohort and the transplant population cannot be directly compared. This problem does not apply to other loci (HLA-A, B, C, DRB1, DQA1, and DQB1) because these loci had fewer allele calls removed due to posterior probability, so the sums of the allele frequencies for these loci were all >95.0%.

### 15.3.3 Results

The allele frequencies in the UKB cohort and among transplant recipients can be seen in Table 6. The table also shows the results of the logistic regression (odds ratio and P value). An odds ratio >1 suggests that the allele is more common among transplant recipients than in the UKB cohort, while an odds ratio <1 suggests the opposite. P<0.0018 was considered significant after Bonferroni correction (0.05/28).

Table 6: allele frequencies in UK Biobank cohort and among renal transplant recipients

| *HLA-* | Allele frequency in UK Biobank cohort | Allele frequency among transplant recipients | Odds ratio | P value |
|---|---|---|---|---|
| Alleles previously associated with decreased renal function | | | | |
| *A*01:01* | 19.4% | 20.1% | 1.05 | NS |
| *A*03:01* | 14.5% | 13.4% | 0.91 | NS |
| *B*07:02* | 14.8% | 15.1% | 1.03 | NS |

| | | | | |
|---|---|---|---|---|
| *B*08:01* | 14.4% | 14.4% | 1.00 | NS |
| *C*07:01* | 17.6% | 17.4% | 0.98 | NS |
| *C*07:02* | 15.8% | 16.2% | 1.03 | NS |
| *DQA1*05:01* | 23.0% | 14.4% | 0.56 | <0.0001 |
| *DQB1*02:01* | 14.9% | 14.6% | 0.97 | NS |
| *DRB1*03:01* | 14.9% | 14.6% | 0.98 | NS |
| Alleles previously associated with increased renal function | | | | |
| *A*25:01* | 1.6% | 2.8% | 1.75 | NS |
| *A*29:02* | 4.2% | 2.9% | 0.68 | NS |
| *A*32:01* | 3.5% | 3.5% | 1.02 | NS |
| *B*14:01* | 1.2% | 1.0% | 0.89 | NS |
| *B*14:02* | 2.5% | 2.1% | 0.81 | NS |
| *B*44:03* | 5.9% | 4.4% | 0.74 | NS |
| *C*02:02* | 3.6% | 3.5% | 0.98 | NS |
| *C*05:01* | 11.4% | 10.0% | 0.87 | NS |
| *C*08:02* | 3.6% | 2.6% | 0.73 | NS |
| *C*12:03* | 2.8% | 3.5% | 1.26 | NS |
| *C*16:01* | 4.4% | 2.8% | 0.62 | NS |
| *DQA1*02:01* | 14.6% | 12.4% | 0.83 | NS |
| *DQA1*03:01* | 20.4% | 11.0% | 0.49 | <0.0001 |
| *DQB1*03:02* | 10.4% | 12.4% | 1.22 | NS |
| *DQB1*06:01* | 0.4% | 0.3% | 0.70 | NS |
| *DQB1*06:09* | 1.0% | 0.3% | 0.29 | NS |
| *DRB1*04:01* | 11.1% | 13.4% | 1.23 | NS |
| *DRB1*07:01* | 14.6% | 12.1% | 0.80 | NS |
| *DRB1*15:02* | 0.4% | 0.3% | 0.71 | NS |

This table shows allele frequencies in the UK Biobank cohort and among renal transplant recipients. It also shows the results of logistic regression between these two groups. NS – not significant

Two significant associations were found. *HLA-DQA1*05:01* and *DQA1*03:01* were both less common in the transplant cohort than in the UKB cohort (P<0.0001), suggesting that these alleles are protective against renal failure. In the case of *HLA-DQA1*03:01*, this replicates the finding reported in this thesis. However, the finding relating to *HLA-DQA1*05:01* contradicts the finding reported in this thesis. The other associations reported in this thesis were neither replicated nor refuted by this validation.

### 15.3.4 Discussion

It is concerning that a finding from this thesis (linking *HLA-DQA1*05:01* to decreased renal function) was refuted by this validation. Confidence in this finding was initially high because two separate outcome measures found the same association: the allele was linked to lower eGFR and increased risk of CKD (see section 11). The allele was therefore expected to have higher frequency among transplant recipients than the UKB cohort. However, contrary to this expectation, the allele had 23.0% frequency among white British subjects in the UKB cohort and only 14.4% frequency among the 340 white British transplant recipients in this validation (OR 0.56, P<0.0001). I attempted to confirm the frequency of this allele among white British

people but found conflicting reports. One study of 177 “Caucasoid” subjects from England reported a frequency of 22.0%[139], but another study of 114 “Caucasoid” subjects from England reported a frequency of 15.4%[140]. It is therefore unclear which of the frequencies observed in this validation (23.0% in the UKB group, or 14.4% in the renal transplant recipient group) is closer to the true frequency in the entire white British population. It is possible that this validation underestimated the allele’s frequency among transplant recipients (14.4% observed), though there is no particular reason to suspect this. Such an error could be caused by the relatively small sample size (n=340). Future studies may be able to clarify whether this allele is truly associated with renal function (see section 16.4). On the other hand, it is reassuring that the finding relating to *HLA-DQA1*03:01* (i.e., that it is linked to increased renal function) was replicated by this validation. Other findings were neither refuted nor replicated.

There are significant limitations which mean that drawing conclusions from this validation may not be possible. Firstly, the two groups were HLA typed by different methods, which may introduce errors and/or bias; future studies should use the same methods for both groups to ensure that the analysis compares HLA types on a “like for like” basis. This would allow validation of the findings relating to HLA-DRB3 and DRB4, which were excluded from this validation (see section 15.3.2). Secondly, not all subjects in the UKB cohort have good kidney function, so they cannot all be considered “healthy controls”. Future studies should ensure that the control group is comprised entirely of subjects with good kidney function. Thirdly, not all transplants performed at Manchester Royal Infirmary were included as some recipients did not have high-resolution HLA typing data; future studies should ensure that all transplants are included (subject to other inclusion criteria such as a specific time period) to reduce the risk of bias. A small number of subjects were missing ethnicity data, though there is no reason to think that there is a bias relating to which subjects did not have ethnicity recorded.

### 15.3.5 Conclusion

To conclude, the majority of the findings reported in this thesis were not replicated by this validation. One finding was replicated but one was refuted. The analysis should be repeated with bigger, better-defined case and control groups, with fewer subjects excluded due to missing data. This future study should include more HLA types than the 28 that were analysed here, and the case and control groups should be treated identically (including using the same HLA typing method for both). This recommendation is also mentioned in section 16.4.2.

## 15.4 Implications and applications

### 15.4.1 Haplotypes and linkage disequilibrium

During recombination, the cutting of DNA is not random, resulting in groups of genetic variations being inherited together more often than would be expected if a process of random pairing was occurring. When two loci are inherited together more often than they would be by random chance, they are said to be in LD. For example, Vogel and Motulsky[141] showed that Danish subjects who expressed HLA-A1 were significantly more likely than the general Danish population to express HLA-B8.

The reliability and validity of some of the findings of this study are increased by the fact that some of the HLA types that were associated with kidney function are in LD with each other. For example, every allele in the haplotype *HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DRB3*01:01, DQA1*05:01, DQB1*02:01* was linked to decreased kidney function in white British subjects. The haplotype has 9.5% frequency in the English population[142]; if the alleles were distributed randomly throughout the population then the frequency of this haplotype would be much lower.

The reason that the tight LD between these types may increase the validity of the findings is that it might suggest that the analysis is identifying associations with an area of the genome, rather than with a single HLA type. This is because these (apparently independent) observations are identifying alleles which are commonly inherited together more often than would be expected by chance. However, it could also be the case that a type I error for one HLA type would increase the likelihood of a type I error for another type if the types are in moderate/strong LD.

### 15.4.2 Clinical applications

The findings made in this study may have an impact on clinical practice. Genotyping patients (to assess whether they express genes associated with renal failure) could be used as a screening tool to help predict renal failure even before symptoms have begun to manifest. Subjects with genetic variations which are associated with renal failure could potentially have their renal function tested more regularly than subjects without such variations, in order to effect faster diagnoses and inform treatment pathways while the kidney failure is at its least severe. Although potential renal transplant recipients are HLA typed prior to entry on the national renal transplant waiting list, their HLA type does not influence the course of their treatment at present.

Another possible application of this study's findings could be a change in the way organs are allocated for transplantation. Patients with HLA types linked to reduced renal function are more

likely to need a kidney transplant, and they may also be less likely to receive an offer for a deceased donor kidney. This is because the HLA match between donor and recipient is an important factor in organ allocation[143] as it is known to affect graft survival[144]. For patients with hazardous HLA types, kidneys that match their HLA type have lower function on average (by definition) and thus may be less likely to proceed to donation. This means that there may be fewer well-matched kidneys available for patients who have hazardous HLA types. If this is the case, these patients could be given priority or increased weighting when kidneys become available for transplantation to redress this imbalance. However, it is not clear that people with hazardous HLA types are in fact less likely to proceed to donation. More research is required.

In order for this study to have an effect in a clinical setting, the findings will need to be replicated using an independent cohort. Any changes to clinical practice would also need to be assessed in terms of efficacy and cost-effectiveness, as well as any possible ethical, legal, or political issues regarding the equality and equity of access to healthcare.

## 15.5 Strengths of this study

### 15.5.1 Sample size

UKB's dataset of over 500,000 subjects meant that large sample sizes were possible in these analyses. The study of white British subjects included over 380,000 subjects for the primary outcome measure of eGFR, while the secondary outcome measures saw case groups of 11,379 (ESRD), 4,794 (CKD), 1,412 (RRT), and 437 (kidney transplant recipients). The study of subjects from ethnic minorities had smaller study groups: 12,191 Irish, 5,475 Indian, 4,100 Black Caribbean, 3,038 Black African, 1,662 Pakistani, and 1,435 Chinese subjects were included in the eGFR analysis, while the CKD (stage 2 or higher) groups ranged from 327-5,964 depending on ethnicity.

This is the largest study of its kind. The analysis of eGFR in white British subjects had 80% power to detect a difference of ~0.35 eGFR points in an allele with 10% frequency compared to subjects without the allele, given a P value threshold of 0.00014, an average eGFR of 90 (±14), and a sample size of 380,000. The analysis of ESRD in white British people had 80% power to detect an odds ratio of 1.137 in an allele with 10% frequency, given a P value threshold of 0.001 and a group of 11,000 cases. The analyses of other ethnicities and other outcomes, which all used smaller study groups, had less power than this.

The power of this study compares favourably to those that have been published before. In the studies included in the systematic review (other than my own), the median number of subjects in the study group (i.e., subjects with renal dysfunction) was 163, while the median control group size was 191. A case-control study of 163 cases and 191 controls would have 80%

power to detect an odds ratio of ~2.4 in an allele with 10% frequency (assuming a P value threshold of 0.05, i.e., without any correction for multiple testing). That is much less statistically powerful than this study, and adjusting for multiple testing would reduce the power further. Of course, half of the studies had even smaller sample sizes, including seven which included only 16-50 subjects with kidney dysfunction. Given that the largest odds ratio reported in my study was 1.122 (for the association between *HLA-DRB1*03:01* and CKD, $P=3.6x10^{-5}$, frequency=14.9%, CKD case group=4,794), it appears that many of the previous studies used sample sizes that were too small to reliably detect associations, particularly for low-frequency alleles. The problem was averted in this study by the large sample sizes used. The high statistical power of this study reduced the risk of type I and type II errors.

### 15.5.2 Measures of renal function

The primary outcome measure of this study was eGFR. This is the standard clinical measure of kidney function. Previous studies tended to use inferior outcome measures, which could increase the risk of errors. The majority treated renal function as a categorical variable, with subjects classed as either healthy or unhealthy based on eGFR or clinical diagnoses. In fact, renal function is a spectrum with a wide range of possible values; a continuous measure such as eGFR is preferable as it provides a more precise measurement of kidney function. In this study, categorical measures of kidney function were used as secondary outcomes only. In a number of cases, analysis of these secondary outcomes successfully replicated findings from the primary outcome.

### 15.5.3 Statistical methods

The statistical methods and quality control were of a high standard, reducing the risk of errors. Bonferroni correction was used to account for the fact that many different statistical tests were performed; this reduced the chance of type I error. Many previous studies did not make such an adjustment. However, this correction is very stringent and may increase the chance of type II errors[145].

As an alternative to Bonferroni correction, I applied Benjamini-Hochberg correction to my analysis of eGFR in white British subjects. This method would have resulted in an additional 13 alleles being considered statistically significant. This would have raised the number of alleles linked to increased kidney function from 22 to 30, and the number of alleles linked to decreased kidney function from 11 to 16. The additional protective alleles identified were: *HLA-A*11:01, A*26:01, B*27:05, B*35:01, B*38:01, B:40:02, B*44:02,* and *DRB1*04:04*. The additional hazardous alleles identified were: *HLA-B*40:01, C*03:04, DRB1*14:01, DRB3*02:02,* and *DQB1*05:03*. The analyses of other outcomes and other ethnicities may

also have seen an increase in significant findings if Benjamini-Hochberg correction had been used instead of Bonferroni.

Quality control also ensured that the data analysed was of a high standard. This included removal of certain subjects based on poor quality or missing data, and the exclusion of HLA alleles which were not in Hardy-Weinberg equilibrium.

## 15.6 Criticisms and limitations

While the analysis has revealed a number of HLA types which may be associated with renal function, there are some limitations which mean that further research is required in order to improve certainty of the findings.

### 15.6.1 Study population

The cohort of subjects used in the analysis is not representative of the general population, so the findings cannot be applied to the entire population. All subjects were recruited in the UK, were aged between 39-73 at recruitment, and were stratified based on their ethnicity. For this reason, findings can only be applied to people in the UK who are middle-aged or approaching middle-age and who are of the ethnic group specified.

It may be possible to improve the representativeness of the study and to validate the study's findings by using other cohorts. This could include subjects from a different biobank, or NHS patients from either Manchester University Foundation Trust (MFT) or NHS Blood and Transplant. This would be subject to ethical approval. There are over 1.8m patients with renal failure in the UK. Acquiring genetic data on all of these is not a realistic target, but it may be possible to analyse a subset of these patients (see section 16.4.2). A smaller dataset than the one used so far might even have greater statistical power, depending on the number of cases of renal dysfunction it contains and the genetic diversity of its subjects. This is discussed in more detail in section 15.6.5.

### 15.6.2 HLA data

The HLA type for each subject was gleaned by imputation. Imputation is not sufficiently accurate to be used clinically (see section 6.5), and its accuracy depends on the software and reference panel that were used. If HLA data is inaccurate, the analysis is susceptible to both type I and type II errors.

The HLA data for this project was imputed by UKB using the HLA*IMP:02 program[146]. A reference panel comprised of multiple datasets was used; there were between 800-9,000 subjects included in this reference panel, typed on "various SNP arrays". The exact number included varied for different HLA loci. Though it is not feasible to use a traditional HLA typing

method due to financial constraints, it may be possible to improve on the imputed data used so far. SNP2HLA software may provide more accurate HLA data[53,61] if a reference panel which represents the study cohort can be found. However, the imputation used is estimated to be over 96% accurate for European subjects[62], which is sufficient for this study. This study used a posterior probability threshold of 0.8, ensuring that only HLA types that were imputed with high confidence were included in the analysis. See section 6.6.4 for more details.

Another issue with the HLA information used in the initial analysis is that there was a danger of type I and II errors due to low type frequencies. Of the 362 types included in the initial analysis, 72% had frequency ≤1% and 85% had frequency ≤5% (see section 19.5). Including types with low frequency could lead to errors due to the small sample conferring low statistical power. This problem was averted (to some extent) by excluding low-frequency types from some analyses. This had the added benefit of reducing the adjustment of the P value required for significance in these analyses as the adjustment for multiple testing was less stringent than if all HLA types were included. However, these analyses could not possibly detect associations in low-frequency alleles (since they were excluded), meaning that true associations may have gone undiscovered (type II error).

### 15.6.3 Limited applications

The findings presented in this study have limited clinical applications. Although HLA testing is used as a diagnostic tool for a number of diseases, these diseases have much stronger HLA associations than the associations presented here. For example, the association between *HLA-B*27* and ankylosing spondylitis has an odds ratio of over 100[147], compared with odds ratios of around 1.1 presented in this study. The associations presented here are not strong enough to be useful in diagnosing renal disorders. Other applications of these findings, such as changes to NHSBT's organ allocation process, are possible but would require further analysis to ensure that they are beneficial and equitable (see section 15.4.2).

The analysis of Covid-19 did not produce any novel findings. There was some value in replicating previously known associations (e.g., increased age and male sex are linked to increased Covid-19 mortality), but there were no significant associations discovered between Covid-19 and either HLA or renal function. The lack of new findings was probably a factor in this research not being selected for publication.

### 15.6.4 Limited genetic scope

There are many factors other than HLA which are associated with renal function, including other genetic factors. The analysis performed in this study so far is unable to detect any of these genetic associations as it considers only HLA type as the independent variable. The

UKB dataset includes the entire genotype for each subject so it would be possible to extend the genetic scope of the research to find non-HLA genetic associations with renal function (see section 16.2).

### 15.6.5 Statistical power

Despite being the largest ever investigation into HLA associations with renal function (see section 15.5.1), this study may lack statistical power to detect associations, particularly in analyses involving outcomes with small numbers of cases or in analyses of lower frequency HLA types. The power of a study depends not only on the number of subjects included but also on the size of the effect that the study aims to detect and the distribution of subjects among the “case”/”control” groups and “exposed”/”not exposed” groups. Outcomes which are particularly rare may not have a difference between experimental group and control group which is sufficiently large to be detected, particularly when the HLA type being tested is rare.

For example, only 258 out of 401,307 white British subjects are kidney transplant recipients. This means that, for a type to be found to have a significant association with kidney transplant status, the genetic effect size would have to be large: this analysis had 80% power to detect an odds ratio of ~2.0 in an allele with 10% frequency (given a P value threshold of 0.001). For less common HLA types, this problem is compounded. The power of the tests is better for common types, but for lower-frequency types and/or rare outcomes, the study is only able to detect relatively large effects. To address this problem, multiple outcome measures have been used, and the statistical power for the primary analysis is high. In addition, rare HLA alleles are less likely to be important in diagnostic tests.

### 15.6.6 Control for covariates

Since there are many factors other than HLA which are associated with renal function, it was necessary to account for these using covariate analysis. The analysis presented here included adjustments for the age and sex of the subjects, but future work should also adjust for other factors such as lifestyle and other health conditions associated with kidney disease (see section 16.1).

# 16 Summary of future work

## 16.1 More control for covariates

The analysis performed so far had minimal adjustment for confounding variables and stratification of the cohort. Age and sex were factored in as covariates in the analysis, and the analyses were performed separately on different self-reported ethnic groups. However, there may be many other factors which interact with kidney function, for example variables relating to lifestyle, demographics, and clinical history.

Adjusting for these variables may help to ensure that any genetic associations found are in fact associations between the gene and renal function, rather than a result of the interaction of a confounding variable. The use of these variables is dependent upon sample sizes; the number of individuals who display certain confounding variables may be too small to analyse in a meaningful way.

## 16.2 Non-HLA genes

The UKB dataset includes genotypes covering the majority of the genome. The analysis so far has tested only for associations between renal function and the HLA region of the genome, which includes only a tiny proportion of all genes. The reason for such specific testing so far is that previous research has suggested that HLA type is linked to renal function.

Future work should expand the genetic scope of the study to search other areas of the genome for associations with renal function, since it is possible that there are many genes which are associated with renal function or likelihood of developing renal failure.

A possible drawback of a wider search of the genome is that, by including a larger number of genes in the analysis, the correction for multiple testing will be more severe. This means that, for a gene to be found to be significantly associated with renal function, the effect size may have to be very large. It is possible that the study will lack sufficient power to detect small effects, particularly in genetic variants which are uncommon.

It may be necessary, therefore, to analyse only specific genetic regions (that have been identified in previous literature as significant or possibly significant) rather than an indiscriminate search of the entire genome. Testing pre-determined genes which are already suspected of having an association is known as “candidate-gene searching”, while a broader search of all genes is known as a “hypothesis-generating” search.

## 16.3 Check accuracy of imputation

The HLA data used so far in the project was imputed by UKB using HLA*IMP:02 software and a reference panel of 800-9,000 subjects (see section 15.6.2). Previous literature has suggested that SNP2HLA software may be superior to HLA*IMP:02 (see section 6.5).

The quality of this imputation should be verified. This can be done by imputing the data using the SNP2HLA program and comparing the newly imputed data to the existing dataset. If this verification shows inaccuracies in the dataset used so far, it may be necessary to re-run the analysis using the newly imputed data.

However, imputing HLA types requires significant time and resources. A cohort of 500,000 is too large to impute simultaneously, so must be split into approximately 100 runs of around 5,000 subjects at a time. Each run takes up to ten hours to complete, so it may not be feasible to re-run the entire imputation on this dataset, particularly if initial results suggest that the data imputed by UKB is accurate. As described in section 6.6.4, a 2018 study estimated the accuracy of UKB's imputation to be >96%[62], which is sufficient for this study.

## 16.4 Replication using different populations

It may be possible to repeat the analysis using a different sample of subjects. This could replicate any findings, which would increase the validity of the study. It could also mitigate any bias present in the cohort used so far. Additionally, it could reveal whether findings can be applied to populations other than people in the UK aged 39-73. Finally, it could reveal new associations relevant to the population being studied. Possible groups of subjects could include subjects from a different biobank, or a set of NHS patients.

### 16.4.1 Different biobanks

Analysing data from another biobank would be a good method of validating and possibly replicating the findings made in the project. The analytical techniques would be similar, and large cohorts already exist. On the other hand, acquiring the data may be costly, and limitations of UKB such as a lack of "cases" of renal failure (see section 15.6.1) may also apply to other biobanks.

Potential biobanks which could be used include the 100,000 Genomes Project (a UK government initiative which achieved its target of sequencing 100,000 genomes in December 2018[148]) and All of Us (a National Institutes of Health project in USA which aims to capture genomic data on 1m subjects by 2024[149]).

### 16.4.2 NHS patients

It may be possible to study a group of NHS patients with renal failure. This could be provided by NHS Blood and Transplant (NHSBT), or by the Transplantation Laboratory at MFT. This would be subject to ethical approval and data-release protocols specific to the organisations.

An advantage of this approach would be that, although the dataset would be smaller overall than the UKB cohort, there would be a large number of "cases" of subjects with varying degrees of renal failure. MFT, for example, has recorded the HLA types of over 7,500 transplant recipients, as well as other patients with renal dysfunction such as those on the waiting list and those on dialysis. A control group of healthy individuals would also be required. This could include transplant donors or another cohort of subjects with healthy renal function. See section 15.3 for recommendations on the methodology of such a study. Since 2018, kidney donors and recipients at MFT have been NGS typed, meaning high-resolution HLA typing data is available. Some donor-recipient pairs have also been GWAS typed, although this was performed recently and therefore was not included in my research. Transplant outcome data is provided regularly by NHSBT. Pending ethical approval, this dataset could be a valuable resource for many different studies.

## 16.5 Analysis of transplant outcomes

### 16.5.1 HLA type

The work so far has investigated associations between HLA type and renal function. There may also be a link between HLA type and renal transplant outcomes. The degree of matching between donor and recipient HLA type is known to affect graft success[144] (see section 16.5.2), but there has been very little research into whether the HLA types themselves (independent of mismatch) impact on graft success. Transplant recipients with certain HLA types may have better (or worse) graft survival than others, and kidneys from donors with certain HLA types might perform differently than others.

If a particular HLA type is linked with both increased incidence of renal failure and decreased graft survival, this could lead to a "vicious cycle": patients with this type would be more likely to need a transplant, but they would have to receive either a well-matched kidney (which would have an HLA type associated with poor transplant outcomes) or a poorly-matched kidney (which is known to lead to poor outcomes). It may also be that people with alleles associated with poor renal function are less likely to become kidney donors, meaning the recipients with these types are less likely to receive a well-matched kidney.

Investigating the impact of HLA type on transplant outcomes (irrespective of mismatch) may reveal findings with important applications for organ allocation. An analysis of HLA's impact

on graft outcomes could be performed using the NHS cohort mentioned in section 16.4.2. Outcome measures could include graft survival, patient survival, and eGFR at set time periods post-transplant.

### 16.5.2 HLA mismatch

It has been widely documented that HLA mismatch affects transplant outcomes. Deceased donor kidney allocation in the UK is based on a number of factors including "ABDR" mismatch (a three-digit code showing the number of mismatches between donor and recipient at HLA-A, -B, and -DR)[143,150]. Recently, however, tools have been developed to provide a more precise measurement of donor-recipient HLA mismatch[151,152]. I was involved in a study comparing three such tools' ability to predict donor-specific antibodies in cardiothoracic transplantation[103], which found HLA-EMMA to be superior to HLAMatchmaker and PIRCHE-II. I intend to perform a similar study assessing the ability of these tools to predict kidney transplant success, with outcome measures such as graft survival, patient survival, and eGFR. This could ultimately lead to the UK's organ allocation algorithm using a more accurate and precise measurement of HLA mismatch.

# 17 Conclusion

The research presented in this thesis includes substantial new findings regarding the relationship between HLA and kidney function. 33 HLA types were linked to kidney function in white British subjects. The 11 alleles associated with decreased kidney function are particularly interesting, as many are in LD with each other and are often inherited together as a haplotype (*HLA-A*01:01, B*08:01, C*07:01, DRB1*03:01, DRB3*01:01, DQA1*05:01, DQB1*02:01*). This suggests that people with this haplotype (9.5% of the English population) may be at increased risk of kidney disease. There were a further nine associations found relating to other ethnic groups, most notably four alleles which appear to be hazardous for Black African people in the UK (*HLA-B*53:01, C*04:01, DPA1*02:01,* and *DPA1*02:02*). Black people are already at increased risk of kidney disease compared to white people, and these alleles may contribute to the risk.

The systematic review compiles all previously published findings (including the results of my own publications) into a single place. This resource is both unique and useful, allowing researchers and clinicians to see details of all known HLA associations with kidney function and providing a list of references for further reading.

Although the research into Covid-19 did not reveal novel findings, I believe the methodology was of a high standard and the results are a useful replication of the previously known risk factors. As more data becomes available regarding the long-term effects of Covid-19, I hope to see more insightful and influential studies describing the relationship between Covid-19 and both HLA and kidney function.

This work has laid the foundation for future studies, which may include attempts to replicate the findings with different participants, and studies of associations between HLA and kidney transplant outcomes. I aim to contribute to such studies, and await other research groups' publications with interest.

# 18 References


1. Kidney Research UK. The Kidneys – a Basic Guide. 2018; https://www.nhs.uk/Livewell/Kidneyhealth/Documents/kidney%20guide.pdf. Accessed 19 Nov, 2018.
2. Stevens LA, Levey AS. Measurement of kidney function. *Med Clin North Am.* 2005;89(3):457-473.
3. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function--measured and estimated glomerular filtration rate. *N Engl J Med.* 2006;354(23):2473-2483.
4. Levin A. Cystatin C, serum creatinine, and estimates of kidney function: searching for better measures of kidney function and cardiovascular risk. *Ann Intern Med.* 2005;142(7):586-588.
5. NHS. Urine albumin to creatinine ratio (ACR). 2018; https://www.nhs.uk/conditions/acr-test/. Accessed 12 Dec, 2018.
6. Higgins C. Urea and the clinical value of measuring blood urea concentration. 2016; https://acutecaretesting.org/en/articles/urea-and-the-clinical-value-of-measuring-blood-urea-concentration. Accessed 20 May, 2022.
7. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron.* 1976;16(1):31-41.
8. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med.* 1999;130(6):461-470.
9. Lewis J, Agodoa L, Cheek D, et al. Comparison of cross-sectional renal function measurements in African Americans with hypertensive nephrosclerosis and of primary formulas to estimate glomerular filtration rate. *Am J Kidney Dis.* 2001;38(4):744-753.
10. Tonelli M, Wiebe N, Culleton B, et al. Chronic kidney disease and mortality risk: a systematic review. *J Am Soc Nephrol.* 2006;17(7):2034-2047.
11. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med.* 2004;351(13):1296-1305.
12. Weiner DE, Tighiouart H, Amin MG, et al. Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality: a pooled analysis of community-based studies. *J Am Soc Nephrol.* 2004;15(5):1307-1315.
13. UK Renal Registry. 23rd Annual Report. 2021; https://ukkidney.org/sites/renal.org/files/publication/file-attachments/23rd_UKRR_ANNUAL_REPORT_0.pdf. Accessed 12 May, 2022.
14. Sharma V, Shaw A, Lowe M, Summers A, van Dellen D, Augustine T. The impact of the COVID-19 pandemic on renal transplantation in the UK. *Clin Med (Lond).* 2020;20(4):e82-e86.
15. NHS. Chronic Kidney Disease in England: The Human and Financial Cost. 2012; https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Chronic-Kidney-Disease-in-England-The-Human-and-Financial-Cost.pdf. Accessed 7 Mar, 2023.
16. Fried LF, Katz R, Sarnak MJ, et al. Kidney function as a predictor of noncardiovascular mortality. *J Am Soc Nephrol.* 2005;16(12):3728-3735.

17. Hemmelgarn BR, Manns BJ, Lloyd A, et al. Relation between kidney function, proteinuria, and adverse outcomes. *Jama.* 2010;303(5):423-429.

18. Genetics Home Reference. What are genome-wide association studies? 2018; https://ghr.nlm.nih.gov/primer/genomicresearch/gwastudies. Accessed 21 Sep, 2018.

19. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet.* 2001;69(5):936-950.

20. National Human Genome Research Institute. The Human Genome Project Completion: Frequently Asked Questions. 2010; https://www.genome.gov/11006943/human-genome-project-completion-frequently-asked-questions/ Accessed 17 Nov, 2017.

21. National Human Genome Research Institute. The Cost of Sequencing a Human Genome. 2016; https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/. Accessed 21 Sep, 2018.

22. Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature genetics.* 2002;32(4):650-654.

23. EMBL-EBI. GWAS Catalog. 2018; https://www.ebi.ac.uk/gwas/. Accessed 3 Jan, 2019.

24. Ikegawa S. A short history of the genome-wide association study: where we were and where we are going. *Genomics & informatics.* 2012;10(4):220-225.

25. Flint J. Gwas. *Curr Biol.* 2013;23(7):R265-266.

26. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *Jama.* 2008;299(11):1335-1344.

27. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101(1):5-22.

28. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature.* 2016;536(7614):41-47.

29. Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature genetics.* 2016;48(5):510-518.

30. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics.* 2018;50(8):1112-1121.

31. Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci.* 2019;22(3):343-352.

32. Kirk A, Knechtle SJ, Larsen CP, Madsen JC, Pearson TCMD, Webber SA, eds. *Textbook of organ transplantation.* 2014.

33. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J.* 2007;48(1):11-23.

34. Anthony Nolan Research Institute. HLA Alleles Numbers. 2019; http://hla.alleles.org/nomenclature/stats.html. Accessed 25 Jan, 2019.

35. Marsh SGE, Parham P, Barber LD, eds. *The HLA factsbook.* 2000. Factsbook series.

36. Janeway C, ed *Immunobiology 5 : the immune system in health and disease.* 2001.

37. Williams TM. Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *J Mol Diagn.* 2001;3(3):98-104.

38. Moalic V, Ferec C. Typage HLA, méthodes d'analyses et applications cliniques. *La Presse Médicale.* 2005;34(15):1101-1108.

39. Perkins HA. Clinical applications of HLA typing. *Clin Lab Med.* 1982;2(1):123-135.

40. Nunes E, Heslop H, Fernandez-Vina M, et al. Definitions of histocompatibility typing terms. *Blood.* 2011;118(23):e180-183.

41. Anthony Nolan Research Institute. Nomenclature for Factors of the HLA System. 2010; http://hla.alleles.org/nomenclature/naming.html. Accessed 21 Jan, 2019.

42. Dyer PA, Martin S. Techniques used to define human MHC antigens: serology. *Immunology letters.* 1991;29(1-2):15-21.

43. Dyer P, Middleton D, eds. *Histocompatibility testing : a practical approach.* 1993. The practical approach series.

44. Woszczek G, Borowiec M, Mis M, Gorska M, Kowalski ML. Comparison of serological and molecular (PCR-SSP) techniques of HLA-DR typing in clinical laboratory routine. *Ann Transplant.* 1997;2(1):39-42.

45. Bartlett J, Stirling D, eds. *PCR protocols.* 2002. Methods in molecular biology 226.

46. Jones DC, Bunce M, Fuggle SV, Young NT, Marshall SE. Human platelet alloantigens (HPAs): PCR-SSP genotyping of a UK population for 15 HPA alleles. *Eur J Immunogenet.* 2003;30(6):415-419.

47. Bontadini A. HLA techniques: typing and antibody detection in the laboratory of immunogenetics. *Methods.* 2012;56(4):471-476.

48. Dunckley H. HLA typing by SSO and SSP methods. *Methods in molecular biology (Clifton, NJ).* 2012;882:9-25.

49. Kallon D, ed *Histocompatibility & Immunogenetics.* 2011.

50. Kulski J, ed *Next-Generation Sequencing — An Overview of the History, Tools, and " Omic " Applications.* 2016.

51. Gabriel C, Furst D, Fae I, et al. HLA typing by next-generation sequencing - getting closer to reality. *Tissue Antigens.* 2014;83(2):65-75.

52. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet.* 2016;98(1):116-126.

53. Karnes JH, Shaffer CM, Bastarache L, et al. Comparison of HLA allelic imputation programs. *PLoS One.* 2017;12(2):e0172444.

54. Khor SS, Yang W, Kawashima M, et al. High-accuracy imputation for HLA class I and II genes based on high-resolution SNP data of population-specific references. *Pharmacogenomics J.* 2015;15(6):530-537.

55. Zheng X, Shen J, Cox C, et al. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* 2014;14(2):192-200.

56. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One.* 2013;8(6):e64683.

57. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA*IMP--an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics.* 2011;27(7):968-972.

58. Illumina. HumanOmni1-Quad BeadChip. 2009; https://www.illumina.com/documents/products/datasheets/datasheet_humanomni1_quad.pdf. Accessed 9 Mar, 2023.

59. Illumina. HumanOmni5-Quad BeadChip. 2011; https://www.illumina.com/documents/products/product_information_sheets/product_info_humanomni5.pdf. Accessed 9 Mar, 2023.

60. Illumina. HumanExome BeadChips. 2012; https://bioresource.nihr.ac.uk/media/43dpibvc/datasheet_humanexome_beadchips.pdf. Accessed 9 Mar, 2023.

61. Neville MJ, Lee W, Humburg P, et al. High resolution HLA haplotyping by imputation for a British population bioresource. *Hum Immunol.* 2017;78(3):242-251.

62. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.

63. Oxford Dictionaries. biobank. 2018; https://en.oxforddictionaries.com/definition/biobank. Accessed 21 Sep, 2018.

64. Shaw DM, Elger BS, Colledge F. What is a biobank? Differing definitions among biobank stakeholders. *Clinical genetics.* 2014;85(3):223-227.

65. UK Biobank. About UK Biobank. 2018; http://www.ukbiobank.ac.uk/about-biobank-uk/. Accessed 21 Sep, 2018.

66. UK Biobank. Published Papers. 2018; http://www.ukbiobank.ac.uk/published-papers/ Accessed 21 Sep, 2018.

67. BBC News. Biobank set for national roll out. 2006; http://news.bbc.co.uk/1/hi/health/5270478.stm. Accessed 18 Nov, 2017.

68. The Guardian. £61m medical experiment begins. 2006; https://www.theguardian.com/science/2006/aug/22/medicineandhealth.lifeandhealth Accessed 18 Nov, 2017.

69. Fawns-Ritchie C, Deary IJ. Reliability and validity of the UK Biobank cognitive tests. *PLoS One.* 2020;15(4):e0231627.

70. Lyall DM, Cullen B, Allerhand M, et al. Correction: Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One.* 2016;11(5):e0156366.

71. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol.* 2017;186(9):1026-1034.

72. Peakman TC, Elliott P. The UK Biobank sample handling and storage validation studies. *Int J Epidemiol.* 2008;37 Suppl 1:i2-6.

73. UK Biobank. Ethics. 2018; http://www.ukbiobank.ac.uk/ethics/ Accessed 21 Sep, 2018.

74. GeneWatch UK. GeneWatch PR: Response to UK Biobank recruitment drive. 2018; http://www.genewatch.org/article.shtml?als[cid]=507674&als[itemid]=556435. Accessed 18 Nov, 2017.

75. Naito T, Suzuki K, Hirata J, et al. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nat Commun.* 2021;12(1):1639.

76. Venkataraman GR, Olivieri JE, DeBoever C, et al. Pervasive additive and non-additive effects within the HLA region contribute to disease risk in the UK Biobank. *bioRxiv.* 2020:2020.2005.2028.119669.

77. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nature genetics.* 2018;50(11):1593-1599.

78. Green HD, Beaumont RN, Thomas A, et al. Genome-Wide Association Study of Microscopic Colitis in the UK Biobank Confirms Immune-Related Pathogenesis. *J Crohns Colitis.* 2019;13(12):1578-1582.

79. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R, eds. *Features, Evaluation, and Treatment of Coronavirus (COVID-19).* Treasure Island (FL)2022. StatPearls.

80. Nicola M, Alsafi Z, Sohrabi C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int J Surg.* 2020;78:185-193.

81. Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili SM, Bahreini E. A comprehensive review of COVID-19 characteristics. *Biol Proced Online.* 2020;22:19.

82. Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: An overview. *Diabetes Metab Syndr.* 2021;15(3):869-875.

83. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLoS Med.* 2021;18(9):e1003773.

84. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. *BMJ.* 2021;374:n1648.

85. Lotfi M, Hamblin MR, Rezaei N. COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clin Chim Acta.* 2020;508:254-266.

86. Marc A, Kerioui M, Blanquart F, et al. Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *Elife.* 2021;10.

87. GOV.UK. The R value and growth rate. 2022; https://www.gov.uk/guidance/the-r-value-and-growth-rate. Accessed 18 Aug, 2022.

88. GOV.UK. Time series of published reproduction number (R) estimates. 2022; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1093056/220720_R_and_growth_rate_time_series_for_publication_v1.0-1__1_.ods. Accessed 18 Aug, 2022.

89. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature.* 2020;584(7821):430-436.

90. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med.* 2020;383(27):2603-2615.

91. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med.* 2021;384(5):403-416.

92. Elias M, Pievani D, Randoux C, et al. COVID-19 Infection in Kidney Transplant Recipients: Disease Incidence and Clinical Outcomes. *J Am Soc Nephrol.* 2020;31(10):2413-2423.

93. Cheng Y, Luo R, Wang K, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int.* 2020;97(5):829-838.

94. Bowe B, Xie Y, Xu E, Al-Aly Z. Kidney Outcomes in Long COVID. *J Am Soc Nephrol.* 2021;32(11):2851-2862.

95. Migliorini F, Torsiello E, Spiezia F, Oliva F, Tingart M, Maffulli N. Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature. *Eur J Med Res.* 2021;26(1):84.

96. Correale P, Mutti L, Pentimalli F, et al. HLA-B*44 and C*01 Prevalence Correlates with Covid19 Spreading across Italy. *Int J Mol Sci.* 2020;21(15).

97. Wang F, Huang S, Gao R, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov.* 2020;6(1):83.

98. Lowe M, Jervis S, Payton A, et al. Systematic review of associations between HLA and renal function. *Int J Immunogenet.* 2022;49(1):46-62.

99. Lowe M, Payton A, Verma A, et al. Associations between human leukocyte antigens and renal function. *Sci Rep.* 2021;11(1):3158.

100. Lowe M, Payton A, Verma A, et al. Human leukocyte antigen associations with renal function among ethnic minorities in the United Kingdom. *HLA.* 2020.

101. Kachuri L, Francis SS, Morrison ML, et al. The landscape of host genetic factors involved in immune response to common viral infections. *Genome Med.* 2020;12(1):93.

102. Lowe M, Maidstone R, Poulton K, et al. Monthly variance in UK renal transplantation activity: a national retrospective cohort study. *BMJ Open.* 2019;9(9):e028786.

103. Bedford A, Jervis S, Worthington J, Lowe M, Poulton K. Human leukocyte antigen epitope mismatch loads and the development of de novo donor-specific antibodies in cardiothoracic organ transplantation. *Int J Immunogenet.* 2022;49(1):30-38.

104. GOV.UK. Coronavirus (COVID-19) in the UK. 2020; https://coronavirus.data.gov.uk/. Accessed 3 Jun, 2020.

105. World Health Organization. Older people are at highest risk from COVID-19, but all must act to prevent community spread. 2020; http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/statements/statement-older-people-are-at-highest-risk-from-covid-19,-but-all-must-act-to-prevent-community-spread. Accessed 3 Jun, 2020.

106. Office for National Statistics. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 10 April 2020. 2020; https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronavirusrelateddeathsbyethnicgroupenglandandwales/2march2020to10april2020. Accessed 3 Jun, 2020.

107. Lin M, Tseng HK, Trejaut JA, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet.* 2003;4:9.

108. Boyton RJ, Smith J, Jones M, et al. Human leucocyte antigen class II association in idiopathic bronchiectasis, a disease of chronic lung infection, implicates a role for adaptive immunity. *Clin Exp Immunol.* 2008;152(1):95-101.

109. Kubo K, Yamazaki Y, Hanaoka M, et al. Analysis of HLA antigens in Mycobacterium avium-intracellulare pulmonary infection. *Am J Respir Crit Care Med.* 2000;161(4 Pt 1):1368-1371.

110. Nguyen A, David JK, Maden SK, et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J Virol.* 2020.

111. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host Microbe.* 2020.

112. Shi Y, Wang Y, Shao C, et al. COVID-19 infection: the perspectives on immune responses. *Cell Death Differ.* 2020;27(5):1451-1454.

113. UK Biobank. COVID-19 test results data. 2020; http://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=COVID19_tests. Accessed 3 Jun, 2020.

114. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr Diab Rep.* 2011;11(6):533-542.

115. Megiorni F, Pizzuti A. HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *J Biomed Sci.* 2012;19:88.

116. Farina F, Picascia S, Pisapia L, et al. HLA-DQA1 and HLA-DQB1 Alleles, Conferring Susceptibility to Celiac Disease and Type 1 Diabetes, are More Expressed Than Non-Predisposing Alleles and are Coordinately Regulated. *Cells.* 2019;8(7).

117. Blach-Olszewska Z, Leszek J. Mechanisms of over-activated innate immune system regulation in autoimmune and neurodegenerative disorders. *Neuropsychiatr Dis Treat.* 2007;3(3):365-372.

118. Bachmann MF, Kopf M. On the role of the innate immunity in autoimmune disease. *J Exp Med.* 2001;193(12):F47-50.

119. GOV.UK. Cases in United Kingdom. 2020; https://coronavirus.data.gov.uk/details/cases. Accessed 18 Dec, 2020.

120. GOV.UK. Deaths in United Kingdom. 2020; https://coronavirus.data.gov.uk/details/deaths. Accessed 18 Dec, 2020.

121. Kurts C, Panzer U, Anders HJ, Rees AJ. The immune system and kidney disease: basic concepts and clinical implications. *Nat Rev Immunol.* 2013;13(10):738-753.

122. National Kidney Foundation. Chronic Kidney Disease and Pneumococcal Disease: Do You Know the Facts? 2020; https://www.kidney.org/atoz/chronic-kidney-disease-and-pneumococcal-disease-do-you-know-facts. Accessed 18 Dec, 2020.

123. Akalin E, Azzi Y, Bartash R, et al. Covid-19 and Kidney Transplantation. *N Engl J Med.* 2020;382(25):2475-2477.

124. National Kidney Foundation. Kidney disease & COVID-19. 2020; https://www.kidney.org/coronavirus/kidney-disease-covid-19. Accessed 18 Dec, 2020.

125. Johns Hopkins Medicine. Coronavirus: Kidney Damage Caused by COVID-19. 2020; https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-kidney-damage-caused-by-covid19. Accessed 18 Dec, 2020.

126. MacGillivray TJ, Cameron JR, Zhang Q, et al. Suitability of UK Biobank Retinal Images for Automatic Analysis of Morphometric Properties of the Vasculature. *PLoS One.* 2015;10(5):e0127914.

127. UK Biobank. Mortality data: linkage to death registries. 2020; https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/DeathLinkage.pdf. Accessed 18 Dec, 2020.

128. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612.

129. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med.* 2012;367(1):20-29.

130. Levey AS, Coresh J. Chronic kidney disease. *Lancet.* 2012;379(9811):165-180.

131. Townsend P, Phillimore P, Beattie A, eds. *Health and Deprivation: Inequality and the North.* 1987.

132. Adams J, Ryan V, White M. How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data. *J Public Health (Oxf).* 2005;27(1):101-106.

133. Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt.* 2014;34(5):502-508.

134. Peckham H, de Gruijter NM, Raine C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat Commun.* 2020;11(1):6317.

135. Shindo Y, Ohno S, Yamamoto T, Nakamura S, Inoko H. Complete association of the HLA-DRB1*04 and -DQB1*04 alleles with Vogt-Koyanagi-Harada's disease. *Hum Immunol.* 1994;39(3):169-176.

136. Varga E, Palkonyai E, Temesvari P, Toth F, Petri IB. The role of HLA-DRB1*04 alleles and their association with HLA-DQB genes in genetic susceptibility to rheumatoid arthritis in Hungarian patients. *Acta Microbiol Immunol Hung.* 2003;50(1):33-41.

137. Chiu HY, Huang HL, Li CH, et al. Increased Risk of Chronic Kidney Disease in Rheumatoid Arthritis Associated with Cardiovascular Complications - A National Population-Based Cohort Study. *PLoS One.* 2015;10(9):e0136508.

138. Vishnubhotla R, Sasikala M, Ketavarapu V, Reddy DN. High-resolution HLA genotyping identifies alleles associated with severe COVID-19: A preliminary study from India. *Immun Inflamm Dis.* 2021;9(4):1781-1785.

139. Allele Frequencies. England pop 6. 2019; http://www.allelefrequencies.net/hla6003a.asp?hla_population=1939. Accessed 3 Dec, 2019.

140. Allele Frequencies. England pop 2. 2005; http://www.allelefrequencies.net/pop6001c.asp?pop_id=2245. Accessed 17 Apr, 2023.

141. Vogel F, Motulsky AG, eds. *Human genetics : problems and approaches.* 1997.

142. Allele Frequencies. England North West. 2019; http://www.allelefrequencies.net/hla6003a.asp?hla_population=2837. Accessed 3 Dec, 2019.

143. Johnson RJ, Fuggle SV, Mumford L, et al. A New UK 2006 National Kidney Allocation Scheme for deceased heart-beating donor kidneys. *Transplantation.* 2010;89(4):387-394.

144. Opelz G, Wujciak T, Dohler B, Scherer S, Mytilineos J. HLA compatibility and organ transplant survival. Collaborative Transplant Study. *Rev Immunogenet.* 1999;1(3):334-342.

145. Perneger TV. What's wrong with Bonferroni adjustments. *BMJ.* 1998;316(7139):1236-1238.

146. UK Biobank. Imputation of classical HLA types from UK Biobank genotype data. 2016; http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/HLA_imputation.pdf. Accessed 20 Dec, 2018.

147. Brown MA, Pile KD, Kennedy LG, et al. HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. *Ann Rheum Dis.* 1996;55(4):268-270.

148. BBC News. Faster diagnosis from 'transformational' gene project. 2018; https://www.bbc.co.uk/news/health-46456984. Accessed 8 Jan, 2019.

149. All of Us Research Program I, Denny JC, Rutter JL, et al. The "All of Us" Research Program. *N Engl J Med.* 2019;381(7):668-676.

150. NHSBT. Kidney Transplantation: Deceased Donor Organ Allocation. 2019; https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/16915/kidney-allocation-policy-pol186.pdf. Accessed 16 Jun, 2022.

151. Geneugelijk K, Spierings E. PIRCHE-II: an algorithm to predict indirectly recognizable HLA epitopes in solid organ transplantation. *Immunogenetics.* 2020;72(1-2):119-129.

152. Kramer CSM, Koster J, Haasnoot GW, Roelen DL, Claas FHJ, Heidt S. HLA-EMMA: A user-friendly tool to analyse HLA class I and class II compatibility on the amino acid level. *HLA.* 2020;96(1):43-51.

# 19 Appendices

## 19.1 Search strategies

This section shows searches conducted in Medline, Embase, and Cochrane Central Register of Controlled Trials for Paper 1. Search terms were chosen with reference to the indexes of the respective databases. Emtree thesaurus terms (for Embase) that were equivalent to Medline's Medical Subject Headings (MeSH) terms were also selected. A composite of Emtree and MeSH terms (in one search strategy) was used to search Medline and Embase on the Ovid platform.

Line 5 shows searches in which papers were required to be about HLA, renal function, and genetic association. Line 6 shows searches with the same requirements as line 5, plus the requirement that the paper be about "aged" or "middle-aged" participants. There are considerably more results for line 5, so the concept of age was dropped from the search (see section 10).

### 19.1.1 Medline

Database: Ovid MEDLINE(R) <1946 to November Week 4 2018>

Search Strategy:

---

1 exp aged/ or middle aged/ (4682332)

2 "genetic association"/ or "genetic polymorphism"/ or "genetic susceptibility"/ or "Gene Frequency"/ or "Genetic Association Studies"/ or exp Genetic Predisposition to Disease/ or Genome-Wide Association Study/ or Alleles/ or allele/ or Polymorphism, Single Nucleotide/ or single nucleotide polymorphism/ (371561)

3 "end stage renal disease"/ or "kidney failure"/ or "Kidney Failure, Chronic"/ or Renal Insufficiency, Chronic/ (119455)

4 HLA antigen/ or hla system/ or exp HLA Antigens/ or exp histocompatibility antigens class ii/ (94274)

5 2 and 3 and 4 (64)

6 1 and 5 (29)

7 5 not 6 (35)

### 19.1.2 Embase

Database: Embase <1974 to 2018 December 03>

Search Strategy:

---

1 exp aged/ or middle aged/ (3414721)

2 "genetic association"/ or "genetic polymorphism"/ or "genetic susceptibility"/ or "Gene Frequency"/ or "Genetic Association Studies"/ or exp Genetic Predisposition to Disease/ or Genome-Wide Association Study/ or Alleles/ or allele/ or Polymorphism, Single Nucleotide/ or single nucleotide polymorphism/ (604529)

3 "end stage renal disease"/ or "kidney failure"/ or "Kidney Failure, Chronic"/ or Renal Insufficiency, Chronic/ (190668)

4 HLA antigen/ or hla system/ or exp HLA Antigens/ or exp histocompatibility antigens class ii/ (114885)

5 2 and 3 and 4 (134)

6 1 and 5 (20)

7 5 not 6 (114)

### 19.1.3 Cochrane Central Register of Controlled Trials

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <October 2018>, EBM Reviews - Cochrane Database of Systematic Reviews <2005 to November 30, 2018>

Search Strategy:

---

1 exp aged/ or middle aged/ (316344)

2 "genetic association"/ or "genetic polymorphism"/ or "genetic susceptibility"/ or "Gene Frequency"/ or "Genetic Association Studies"/ or exp Genetic Predisposition to Disease/ or Genome-Wide Association Study/ or Alleles/ or allele/ or Polymorphism, Single Nucleotide/ or single nucleotide polymorphism/ (3766)

3 "end stage renal disease"/ or "kidney failure"/ or "Kidney Failure, Chronic"/ or Renal Insufficiency, Chronic/ (6488)

4 HLA antigen/ or hla system/ or exp HLA Antigens/ or exp histocompatibility antigens class ii/ (662)

5 2 and 3 and 4 (0)

6 1 and 5 (0)

7 5 not 6 (0)

## 19.2 UK Biobank application

### 19.2.1 Application form

This is the application form that was submitted to UKB in 2017.

# Project Details

19/07/2017

**Name:** Kay Poulton

**Institute Name:** Central Manchester University Hospitals NHS Trust

**Application:** 28539

**Current Stage:** Main - Adjudication in Progress

**Application Name:** HLA Class I and II Associations with Renal Dysfunction in an Aging Population

## Preliminary Application Details

### 1a. The aims of the proposed research including the research question(s) that you are aiming to answer and the health condition(s) under investigation

This project aims to establish a new way of identifying immune-related genes which impact upon renal disease.

This project will “impute” HLA types from Biobank data. This will indicate extended genomic coverage, allowing the assessment of any impact beyond the HLA genes. These data will be compared with markers of renal decline and dysfunction with the aim of identifying additional and novel genetic associations in renal impairment.

This would be of particular value in renal diseases with immune-mediated components in their aetiology such as lupus or membranous nephropathy.

### 1b. How does the proposed research meet UK Biobank's stated purpose?

This research will identify previously unknown associations with renal function and will inform the clinical management of patients with renal disease.

The research team is composed of clinical and academic professionals and as such this is targeted research with a strong likelihood of patient benefit.

### 1c. Please give a non-technical description of how the research will be undertaken

The Human Leucocyte Antigen (HLA) genes are encoded on chromosome 6 which is included in the Biobank's data resource. There are known HLA associations with some renal diseases and renal function is known to deteriorate with age.

The Biobank data will be translated ("imputed") into HLA types and compared with clinical measures of renal function such as urinary biomarkers (microalbumin, creatinine, potassium and sodium). Analysis will be performed to identify immune-related influences on renal function to extend our understanding of the causes of renal failure and any impact of aging which may hasten renal decline.

### 1d. Please state the approximate number of participants to be included (i.e. whether the full cohort or a subset)

Full cohort

### 2. Please describe the methodology and timetable for the proposed Research Project. Please include a description of the data and/or the quantity and type of samples required

Download, storage and analyses of data

Data will be stored on the secure University of Manchester Research Data Storage (RDS) system http://ri.itservices.manchester.ac.uk/rds/. Only the named applicants for this project will have access to the data. Quality control and analyses will be performed using The University of Manchester Computational Shared Facility (CSF), which comprises a high performance cluster of almost 10,000 cores with nodes of up to 1TB RAM available (see http://ri.itservices.manchester.ac.uk/csf/ for details). The applicants have contributed funds into the CSF and therefore have priority access. 15TB storage space has currently been allocated to this project.

Imputation of MHC alleles and amino acids

We will use imputation software (SNP2HLA) to impute the Major Histocompatibility Complex (MHC) region (Human Leukocyte Antigens and amino acids; chromosome 6) using UK Biobank imputed genetic data from all 500k participants. SNP2HLA has been developed for imputing high resolution MHC alleles from GWAS data. See http://software.broadinstitute.org/mpg/snp2hla/ for details.

Quality control

Quality control of the imputed GWAS data and HLA imputation will be performed under the supervision of Dr Tony Payton. Dr Payton is currently analysing UKB genetic data for age-related hearing loss (Dawes, Payton, Williams, Dawson) and cancer (Muir, Payton, et al). In brief, individuals will be excluded from analyses for reasons that may include (and assuming that QC of next genetic release did not perform these):
Self-reported and genetic sex do not match
Related pairs (EKC >0.044)
High missingness and unexplained heterozygosity
Non-Caucasian
Any participant who requests removal before data is published
QC of imputed data will depend on the UKB QC pipeline already performed.

Analyses

An estimation of G, E and GxE variance may be performed using restricted maximum likelihood analyses implemented in GCTA (http://cnsgenomics.com/software/gcta/). Phenotypes will be excluded if the estimated variance is 0 or lower than the standard error. Linear regression analysis will be performed on Caucasian participants and the imputed MHC alleles using Plink software (Version 1.9). All analyses will be adjusted for participant age and sex. The genome-wide significance threshold will be based on the number of tests performed with an initial standard of 5x10-8 assumed in the first instance.

Timeline

Preliminary work – apply for data, pay fees, register with University of Manchester
Day 0 (~May 2017) – download data
Year 1 – data quality control, HLA imputation, validation of HLA types
Years 2 and 3 – analysis of factors affecting renal decline
Year 4 – publication of findings in journals and at conferences. Feedback data to Biobank
Years 5 and 6 – complete PhD thesis and design future projects based on outcome data

**2.1. What results will you return to UK Biobank, i.e. what will you produce that will be useful to other researchers?**

It is anticipated that at the end of this project, all imputed HLA types for the full cohort will be submitted to the Biobank for free access by other investigators.

Any significant observations will also be submitted and full recognition of the Biobank will be acknowledged in any publications.

**Research is in the Public's Interest?** Y **Research is Health-Related?** Y

**Did the Principal Applicant foresee any Ethical issues with the Application?** N

**Did the Principal Applicant believe that Re-contact was potentially required?** N

**Did the Principal Applicant indicate that the project requires sample analysis?** N

**"Tags" provided by Principal Applicant**

HLA
Renal
Creatinine
Microalbumin
Imputation
Age

---

## Main Application Details

**1a. The aims of the proposed research including the research question(s) that you are aiming to answer and the health condition(s) under investigation**

**1b. How does the proposed research meet UK Biobank's stated purpose?**

**1c. Please give a non-technical description of how the research will be undertaken**

**1d. Please state the approximate number of participants to be included (i.e. whether the full cohort or a subset)**

**Provided scientific rationale**

The aim of this project is to establish a novel way to identify immune-related genes which may impact upon renal disease and transplant outcome.

It is well established that matching for Human Leukocyte Antigens (HLA) between donor and recipient confers optimal kidney survival. HLA antigens are also associated with some disorders which lead to renal failure, although it is possible that the HLA types are markers of closely associated genes which are actually responsible for the onset of renal disease. HLA typing is comparatively expensive, and can become dated when newer tests allow higher level definition of the type. This project aims to validate the ability to interpret, or “impute”, high resolution HLA types from a series of genetic markers across the genomic region encoding the HLA genes which can be translated into HLA types.

Using a system by which HLA types are imputed using genomic marker profiles allows extremely high definition HLA typing at a fraction of the cost. It also carries an advantage of extended genomic coverage, which allows the researcher to analyse any impact of matching beyond the HLA genes. This would be of particular value in a disease where there is an immune-mediated component in the aetiology of renal impairment such as lupus or membranous nephropathy.

By comparing imputed HLA types with markers of renal function, it may be possible to identify putative genes which influence transplant outcome or rapid progression of decline of renal function.

**Detail of pilot studies undertaken**

Using UK Biobank data, Dr Anthony Payton has demonstrated the ability to impute HLA types from the data in a population of 1559 non-pathological elderly volunteers and demonstrated an HLA association with cognitive decline. See reference below.

Payton A, Dawes P, Platt H, Morton CC, Moore DR, Massey J, Horan M, Ollier W, Munro KJ, Pendleton N. 2016. A Role for HLA-DRB1*1101 and DRB1*0801 in Cognitive Ability and Its Decline With Age. Am J Med Genet Part B 171B:209–214.

**Please provide details of the methodology to be used**

Storage of Data

Data will be stored on the secure University of Manchester Research Data Storage (RDS) system http://ri.itservices.manchester.ac.uk/rds/. Only the named applicants for this project will have access to the data. Quality control and analyses will be performed using The University of Manchester Computational Shared Facility (CSF), which comprises a high performance cluster of almost 10,000 cores with nodes of up to 1TB RAM available (see http://ri.itservices.manchester.ac.uk/csf/ for details). The applicants have contributed funds into the CSF and therefore have priority access. 15TB storage space has currently been allocated to this project.

Imputation of MHC alleles and amino acids

We will use imputation software (SNP2HLA) to impute the Major Histocompatibility Complex (MHC) region (Human Leukocyte Antigens and amino acids; chromosome 6) using UK Biobank imputed genetic data from all 500k participants. SNP2HLA has been developed for imputing high resolution MHC alleles from GWAS data. See http://software.broadinstitute.org/mpg/snp2hla/ for details.

Quality control

Quality control of the imputed GWAS data and HLA imputation will be performed under the supervision of Dr Tony Payton. Dr Payton is currently analysing UKB genetic data for age-related hearing loss (Dawes, Payton, Williams, Dawson) and cancer (Muir, Payton, et al). In brief, individuals will be excluded from analyses for reasons that may include (and assuming that QC of next genetic release did not perform these):

Self-reported and genetic sex do not match
Related pairs (EKC >0.044)
High missingness and unexplained heterozygosity
Any participant who requests removal before data is published

QC of imputed data will depend on the UKB QC pipeline already performed.

Analyses

An estimation of G, E and GxE variance may be performed using restricted maximum likelihood analyses implemented in GCTA (http://cnsgenomics.com/software/gcta/). Phenotypes will be excluded if the estimated variance is 0 or lower than the standard error. Linear regression analysis will be performed on Caucasian participants and the imputed MHC alleles using Plink software (Version 1.9). All analyses will be adjusted for participant age and sex. The genome-wide significance threshold will be based on the number of tests performed with an initial standard of 5x10-8 assumed in the first instance.

Timeline

Preliminary work – apply for data, pay fees, register with University of Manchester
Day 0 (~May 2017) – download data
Year 1 – data quality control, HLA imputation, validation of HLA types
Years 2 and 3 – analysis of factors affecting renal decline
Year 4 – publication of findings in journals and at conferences. Feedback data to Biobank
Years 5 and 6 – complete PhD thesis and design future projects based on outcome data

**Expected value of results?**

The Transplantation Laboratory holds DNA, isolated lymphocytes and serum samples from all transplant recipients and their donors processed at this centre, and gathers outcome data within the Transplant Audit Sub Committee (TASC) database. Manchester is the largest individual centre for transplantation and it is therefore essential that we maximise our outcome from this resource, while keeping expenditure to a minimum. If the ability to impute reliable high resolution HLA types from data obtained using a series of genetic markers is proven, it would be a valuable asset to carry out marker testing on all our renal transplant recipients using the Infinium ImmunoArray-24 v2 BeadChip2 which has the capacity to test simultaneously for 253,702 markers involved in immune regulation.
It would then be possible not only to impute high resolution HLA types at minimal cost, but also to search for additional associations in other genes which have not previously been associated with renal disorders, or transplant survival. It would significantly progress our understanding of post-transplant immunological complications, and would enable individualised post-transplant care in response to the patient's susceptibility profile. This would enhance our Transplant Resource in a future-proof way, as it will always be possible to relate these genome-wide markers to associated genes.

Once this facility is established, we aim to
1. Use the Infinium ImmunoArray-24 v2.0 BeadChip to obtain data on 253,702 markers for 1000 consenting renal transplant recipients and their corresponding donors transplanted within the last 5 years.
2. Extend the TASC database to interface with the marker data, allowing analysis to identify putative associations with graft outcome or disease recurrence.
3. Identify potential areas of interest for future research into immune based complications of renal transplantation.
4. Extend the genome wide marker testing to all patients with end stage renal disease who are entered onto the transplant waiting list.

| | |
|---|---|
| **Were references supporting justification for the project provided?** | Y |
| **Will any Sample analysis be performed as part of the research project?** | N |
| **If yes, which sample(s), how many, what volume and what assay?** | |
| **Details of any power calculations made to support this application** | |
| **Confirmed procedures in place to manage access to samples?** | N |
| **Proposal for managing access to samples** | |
| **Confirmed that all samples will be stored in a secure manner at all times?** | N |
| **Proposal for securing samples storage** | |
| **Confirmed that suitable storage facilities for samples are available** | N |
| **Proposal for providing suitable sample storage** | |
| **Please confirm that the data will be stored in a secure manner at all times (e.g., behind a firewall, use of anti-virus software)** | Y |
| **If no, please describe how you will rectify this** | |
| **Expected Project Start Date:** | 01/07/2017 |
| **Expected Project End Date:** | 31/03/2023 |
| **Estimated publication date?** | 31/03/2023 |

| | |
|---|---|
| **Has funding been granted?** | Y |

# biobank uk Project Details

19/07/2017

**Showcase Notes:**

Measures of renal function eg urea and creatinine levels in blood would be useful
Alcohol use info if preview fields become available

Regarding our request for pancreatic scans and pancreatic fat, as a high proportion of people with failure have diabetes, we feel it would be advantageous to have this additional information to indicate the extent of their pancreatic dysfunction in relation to their renal failure.

## Collaborators

| Full Name | Institute | Email Address | Status |
|---|---|---|---|
| Dr. Antony Payton | University of Manchester | tony.payton@manchester.ac.uk | Approved for App |
| Mr. Marcus Lowe | Central Manchester University Hospitals NHS Trust | marcus.lowe@cmft.nhs.uk | Approved for App |
| Dr. Kay Poulton | University of Manchester | kay.v.poulton@manchester.ac.uk | Approved for App |
| Mr. Marcus Lowe | University of Manchester | marcus.lowe@postgrad.manchester.ac.uk | Approved for App |
| Prof William Ollier | Not Known | Bill.Ollier@manchester.ac.uk | Pending |

## Materials

| Field Name | Type | Customised | Exclude Missing Values ? |
|---|---|---|---|
| Private healthcare | Categorical (single) | N | N |
| Ever smoked | Categorical (single) | N | N |
| Light smokers, at least 100 smokes in lifetime (pilot) | Categorical (single) | N | N |
| Smoking status | Categorical (single) | N | N |
| Current tobacco smoking | Categorical (single) | N | N |
| Past tobacco smoking | Categorical (single) | N | N |
| Light smokers, at least 100 smokes in lifetime | Categorical (single) | N | N |
| Age started smoking in current smokers | Integer | N | N |
| Type of tobacco currently smoked | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Previously smoked cigarettes on most/all days | Categorical (single) | N | N |
| Number of cigarettes currently smoked daily (current cigarette smokers) | Integer | N | N |
| Age stopped smoking cigarettes (current cigar/pipe or previous cigarette smoker) | Integer | N | N |
| Number of cigarettes previously smoked daily (current cigar/pipe smokers) | Integer | N | N |
| Time from waking to first cigarette | Categorical (single) | N | N |
| Difficulty not smoking for 1 day | Categorical (single) | N | N |
| Ever tried to stop smoking | Categorical (single) | N | N |
| Wants to stop smoking | Categorical (single) | N | N |
| Smoking compared to 10 years previous | Categorical (single) | N | N |
| Why reduced smoking | Categorical (multiple) | N | N |
| Age started smoking in former smokers | Integer | N | N |
| Type of tobacco previously smoked | Categorical (single) | N | N |
| Number of cigarettes previously smoked daily | Integer | N | N |
| Age stopped smoking | Integer | N | N |
| Ever stopped smoking for 6+ months | Categorical (single) | N | N |
| Ever stopped smoking for 6+ months (pilot) | Categorical (single) | N | N |
| Why stopped smoking | Categorical (multiple) | N | N |

| | | | |
|---|---|---|---|
| Why stopped smoking (pilot) | Categorical (single) | N | N |
| Number of unsuccessful stop-smoking attempts | Integer | N | N |
| Likelihood of resuming smoking | Categorical (single) | N | N |
| Smoking/smokers in household | Categorical (single) | N | N |
| Exposure to tobacco smoke at home | Integer | N | N |
| Exposure to tobacco smoke outside home | Integer | N | N |
| Cooked vegetable intake | Integer | N | N |
| Salad / raw vegetable intake | Integer | N | N |
| Fresh fruit intake | Integer | N | N |
| Dried fruit intake | Integer | N | N |
| Oily fish intake | Categorical (single) | N | N |
| Non-oily fish intake | Categorical (single) | N | N |
| Processed meat intake | Categorical (single) | N | N |
| Poultry intake | Categorical (single) | N | N |
| Beef intake | Categorical (single) | N | N |
| Lamb/mutton intake | Categorical (single) | N | N |
| Pork intake | Categorical (single) | N | N |
| Age when last ate meat | Integer | N | N |

| | | | |
|---|---|---|---|
| Never eat eggs, dairy, wheat, sugar | Categorical (multiple) | N | N |
| Never eat eggs, dairy, wheat, sugar (pilot) | Categorical (multiple) | N | N |
| Cheese intake | Categorical (single) | N | N |
| Milk type used | Categorical (single) | N | N |
| Spread type | Categorical (single) | N | N |
| Non-butter spread type details | Categorical (single) | N | N |
| Spread type (pilot) | Categorical (single) | N | N |
| Bread intake | Integer | N | N |
| Bread type | Categorical (single) | N | N |
| Bread type/intake (pilot) | Categorical (single) | N | N |
| Cereal intake | Integer | N | N |
| Cereal type | Categorical (single) | N | N |
| Salt added to food | Categorical (single) | N | N |
| Tea intake | Integer | N | N |
| Coffee intake | Integer | N | N |
| Coffee type | Categorical (single) | N | N |
| Illnesses of father | Categorical (single) | N | N |
| Age at recruitment | Integer | N | N |
| Month of birth | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Year of birth | Integer | N | N |
| Sex | Categorical (single) | N | N |
| Townsend deprivation index at recruitment | Continuous | N | N |
| Year immigrated to UK (United Kingdom) | Integer | N | N |
| Ethnic background | Categorical (single) | N | N |
| Attendance/disability/mobility allowance | Categorical (multiple) | N | N |
| Average monthly champagne plus white wine intake | Integer | N | N |
| Job coding | Categorical (single) | N | N |
| Work hours - lumped category | Categorical (single) | N | N |
| Work hours per week - exact value | Continuous | N | N |
| Workplace full of chemical or other fumes | Categorical (single) | N | N |
| Workplace had a lot of cigarette smoke from other people smoking | Categorical (single) | N | N |
| Worked with materials containing asbestos | Categorical (single) | N | N |
| Worked with paints, thinners or glues | Categorical (single) | N | N |
| Worked with pesticides | Categorical (single) | N | N |
| Workplace had a lot of diesel exhaust | Categorical (single) | N | N |
| Average monthly beer plus cider intake | Integer | N | N |
| Average monthly spirits intake | Integer | N | N |

| | | | |
|---|---|---|---|
| Average monthly fortified wine intake | Integer | N | N |
| Average monthly intake of other alcoholic drinks | Integer | N | N |
| Average weekly red wine intake | Integer | N | N |
| Average weekly champagne plus white wine intake | Integer | N | N |
| Average weekly beer plus cider intake | Integer | N | N |
| Average weekly spirits intake | Integer | N | N |
| Average weekly fortified wine intake | Integer | N | N |
| Average weekly intake of other alcoholic drinks | Integer | N | N |
| Alcohol usually taken with meals | Categorical (single) | N | N |
| Alcohol intake versus 10 years previously | Categorical (single) | N | N |
| Reason for reducing amount of alcohol drunk | Categorical (single) | N | N |
| Reason for reducing amount of alcohol drunk (pilot) | Categorical (single) | N | N |
| Reason former drinker stopped drinking alcohol | Categorical (single) | N | N |
| Reason former drinker stopped drinking alcohol (pilot) | Categorical (single) | N | N |
| Time spend outdoors in summer | Integer | N | N |
| Time spent outdoors in winter | Integer | N | N |
| Skin colour | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Ease of skin tanning | Categorical (single) | N | N |
| Childhood sunburn occasions | Integer | N | N |
| Hair colour (natural, before greying) | Categorical (single) | N | N |
| Facial ageing | Categorical (single) | N | N |
| Use of sun/uv protection | Categorical (single) | N | N |
| Frequency of solarium/sunlamp use | Integer | N | N |
| Answered sexual history questions | Categorical (single) | N | N |
| Age first had sexual intercourse | Integer | N | N |
| Lifetime number of sexual partners | Integer | N | N |
| Ever had same-sex intercourse | Categorical (single) | N | N |
| Lifetime number of same-sex sexual partners | Integer | N | N |
| Country of birth (UK/elsewhere) | Categorical (single) | N | N |
| Breastfed as a baby | Categorical (single) | N | N |
| Comparative body size at age 10 | Categorical (single) | N | N |
| Comparative height size at age 10 | Categorical (single) | N | N |
| Handedness (chirality/laterality) | Categorical (single) | N | N |
| Adopted as a child | Categorical (single) | N | N |
| Part of a multiple birth | Categorical (single) | N | N |
| Maternal smoking around birth | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Illnesses of mother | Categorical (single) | N | N |
| Illnesses of siblings | Categorical (single) | N | N |
| Eye problems/disorders | Categorical (multiple) | N | N |
| Which eye(s) affected by diabetes-related eye disease | Categorical (single) | N | N |
| Which eye(s) affected by glaucoma | Categorical (single) | N | N |
| Which eye(s) affected by injury or trauma resulting in loss of vision | Categorical (single) | N | N |
| Which eye(s) are affected by cataract | Categorical (single) | N | N |
| Which eye(s) affected by macular degeneration | Categorical (single) | N | N |
| Age when diabetes-related eye disease diagnosed | Integer | N | N |
| Age glaucoma diagnosed | Integer | N | N |
| Age when loss of vision due to injury or trauma diagnosed | Integer | N | N |
| Age cataract diagnosed | Integer | N | N |
| Age macular degeneration diagnosed | Integer | N | N |
| Age other serious eye condition diagnosed | Integer | N | N |
| Overall health rating | Categorical (single) | N | N |
| Long-standing illness, disability or infirmity | Categorical (single) | N | N |
| Chest pain or discomfort | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Had major operations | Categorical (single) | N | N |
| Had other major operations | Categorical (single) | N | N |
| Vascular/heart problems diagnosed by doctor | Categorical (multiple) | N | N |
| Age heart attack diagnosed | Integer | N | N |
| Age angina diagnosed | Integer | N | N |
| Age stroke diagnosed | Integer | N | N |
| Age high blood pressure diagnosed | Integer | N | N |
| Blood clot, DVT, bronchitis, emphysema, asthma, rhinitis, eczema, allergy diagnosed by doctor | Categorical (multiple) | N | N |
| Age deep-vein thrombosis (DVT, blood clot in leg) diagnosed | Integer | N | N |
| Age pulmonary embolism (blood clot in lung) diagnosed | Integer | N | N |
| Age emphysema/chronic bronchitis diagnosed | Integer | N | N |
| Age asthma diagnosed | Integer | N | N |
| Age hay fever, rhinitis or eczema diagnosed | Integer | N | N |
| Diabetes diagnosed by doctor | Categorical (single) | N | N |
| Gestational diabetes only | Categorical (single) | N | N |
| Gestational diabetes only (pilot) | Categorical (single) | N | N |
| Age diabetes diagnosed | Integer | N | N |

| | | | |
|---|---|---|---|
| Started insulin within one year diagnosis of diabetes | Categorical (single) | N | N |
| Cancer diagnosed by doctor | Categorical (single) | N | N |
| Fractured/broken bones in last 5 years | Categorical (single) | N | N |
| Fractured bone site(s) | Categorical (multiple) | N | N |
| Fracture resulting from simple fall | Categorical (single) | N | N |
| Other serious medical condition/disability diagnosed by doctor | Categorical (single) | N | N |
| Chest pain or discomfort walking normally | Categorical (single) | N | N |
| Chest pain due to walking ceases when standing still | Categorical (single) | N | N |
| Chest pain or discomfort when walking uphill or hurrying | Categorical (single) | N | N |
| Medication for cholesterol, blood pressure or diabetes | Categorical (multiple) | N | N |
| Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones | Categorical (multiple) | N | N |
| Taking other prescription medications | Categorical (single) | N | N |
| Medication for pain relief, constipation, heartburn | Categorical (multiple) | N | N |
| Medication for pain relief, constipation, heartburn (pilot) | Categorical (multiple) | N | N |
| Medication for smoking cessation, constipation, heartburn, allergies (pilot) | Categorical (multiple) | N | N |
| Vitamin and mineral supplements | Categorical (multiple) | N | N |

| | | | |
|---|---|---|---|
| Vitamin and mineral supplements (pilot) | Categorical (multiple) | N | N |
| Mineral and other dietary supplements | Categorical (multiple) | N | N |
| Vitamin supplements (pilot) | Categorical (multiple) | N | N |
| Other dietary supplements (pilot) | Categorical (multiple) | N | N |
| Number of live births | Integer | N | N |
| Ever had stillbirth, spontaneous miscarriage or termination | Categorical (single) | N | N |
| Ever taken oral contraceptive pill | Categorical (single) | N | N |
| Ever used hormone-replacement therapy (HRT) | Categorical (single) | N | N |
| Place of birth in UK - north co-ordinate | Integer | N | N |
| Place of birth in UK - east co-ordinate | Integer | N | N |
| Country of Birth (non-UK origin) | Categorical (single) | N | N |
| Number of self-reported cancers | Integer | N | N |
| Number of self-reported non-cancer illnesses | Integer | N | N |
| Cancer code, self-reported | Categorical (multiple) | N | N |
| Non-cancer illness code, self-reported | Categorical (multiple) | N | N |
| Operation year/age first occurred | Integer | N | N |
| Number of operations, self-reported | Integer | N | N |
| Pace-maker | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Operation code | Categorical (multiple) | N | N |
| Interpolated Year when operation took place | Continuous | N | N |
| Interpolated Age of participant when operation took place | Continuous | N | N |
| Method of recording time when operation occurred | Categorical (single) | N | N |
| Height | Integer | N | N |
| Waist circumference | Continuous | N | N |
| Weight | Continuous | N | N |
| Hip circumference | Continuous | N | N |
| Body mass index (BMI) | Continuous | N | N |
| Pancreatic fat - DICOM | Text | N | N |
| Measurements of pancreas volume - DICOM | Text | N | N |
| Microalbumin in urine acquisition time | Date | N | N |
| Microalbumin in urine device ID | Categorical (single) | N | N |
| Creatinine (enzymatic) in urine acquisition time | Date | N | N |
| Creatinine (enzymatic) in urine device ID | Categorical (single) | N | N |
| Potassium in urine acquisition time | Date | N | N |
| Potassium in urine device ID | Categorical (single) | N | N |
| Sodium in urine acquisition time | Date | N | N |

| | | | |
|---|---|---|---|
| Sodium in urine device ID | Categorical (single) | N | N |
| Chromosome 1 genotype calls and imputed values | Text | N | N |
| Chromosome 2 genotype calls and imputed values | Text | N | N |
| Chromosome 3 genotype calls and imputed values | Text | N | N |
| Chromosome 4 genotype calls and imputed values | Text | N | N |
| Chromosome 5 genotype calls and imputed values | Text | N | N |
| Chromosome 6 genotype calls and imputed values | Text | N | N |
| Chromosome 7 genotype calls and imputed values | Text | N | N |
| Chromosome 8 genotype calls and imputed values | Text | N | N |
| Chromosome 9 genotype calls and imputed values | Text | N | N |
| Chromosome 10 genotype calls and imputed values | Text | N | N |
| Chromosome 11 genotype calls and imputed values | Text | N | N |
| Chromosome 12 genotype calls and imputed values | Text | N | N |
| Chromosome 13 genotype calls and imputed values | Text | N | N |
| Chromosome 14 genotype calls and imputed values | Text | N | N |

| | | | |
|---|---|---|---|
| Chromosome 15 genotype calls and imputed values | Text | N | N |
| Chromosome 16 genotype calls and imputed values | Text | N | N |
| Chromosome 17 genotype calls and imputed values | Text | N | N |
| Chromosome 18 genotype calls and imputed values | Text | N | N |
| Chromosome 19 genotype calls and imputed values | Text | N | N |
| Chromosome 20 genotype calls and imputed values | Text | N | N |
| Chromosome 21 genotype calls and imputed values | Text | N | N |
| Chromosome 22 genotype calls and imputed values | Text | N | N |
| Chromosome X genotype calls | Text | N | N |
| Chromosome Y genotype calls | Text | N | N |
| Mitochondrial genotype calls | Text | N | N |
| HLA imputation values and quality | Curve | N | N |
| Chromosome 1 genotype probabilities | Text | N | N |
| Chromosome 2 genotype probabilities | Text | N | N |
| Chromosome 3 genotype probabilities | Text | N | N |
| Chromosome 4 genotype probabilities | Text | N | N |

| | | | |
|---|---|---|---|
| Chromosome 5 genotype probabilities | Text | N | N |
| Chromosome 6 genotype probabilities | Text | N | N |
| Chromosome 7 genotype probabilities | Text | N | N |
| Chromosome 8 genotype probabilities | Text | N | N |
| Chromosome 9 genotype probabilities | Text | N | N |
| Chromosome 10 genotype probabilities | Text | N | N |
| Chromosome 11 genotype probabilities | Text | N | N |
| Chromosome 12 genotype probabilities | Text | N | N |
| Chromosome 13 genotype probabilities | Text | N | N |
| Chromosome 14 genotype probabilities | Text | N | N |
| Chromosome 15 genotype probabilities | Text | N | N |
| Chromosome 16 genotype probabilities | Text | N | N |
| Chromosome 17 genotype probabilities | Text | N | N |
| Chromosome 18 genotype probabilities | Text | N | N |
| Chromosome 19 genotype probabilities | Text | N | N |

| | | | |
|---|---|---|---|
| Chromosome 20 genotype probabilities | Text | N | N |
| Chromosome 21 genotype probabilities | Text | N | N |
| Chromosome 22 genotype probabilities | Text | N | N |
| Chromosome X genotype probabilities | Text | N | N |
| Chromosome Y genotype probabilities | Text | N | N |
| Mitochondrial genotype probabilities | Text | N | N |
| Chromosome 1 genotype intensities | Text | N | N |
| Chromosome 2 genotype intensities | Text | N | N |
| Chromosome 3 genotype intensities | Text | N | N |
| Chromosome 4 genotype intensities | Text | N | N |
| Chromosome 5 genotype intensities | Text | N | N |
| Chromosome 6 genotype intensities | Text | N | N |
| Chromosome 7 genotype intensities | Text | N | N |
| Chromosome 8 genotype intensities | Text | N | N |
| Chromosome 9 genotype intensities | Text | N | N |
| Chromosome 10 genotype intensities | Text | N | N |
| Chromosome 11 genotype intensities | Text | N | N |

| | | | |
|---|---|---|---|
| Chromosome 12 genotype intensities | Text | N | N |
| Chromosome 13 genotype intensities | Text | N | N |
| Chromosome 14 genotype intensities | Text | N | N |
| Chromosome 15 genotype intensities | Text | N | N |
| Chromosome 16 genotype intensities | Text | N | N |
| Chromosome 17 genotype intensities | Text | N | N |
| Chromosome 18 genotype intensities | Text | N | N |
| Chromosome 19 genotype intensities | Text | N | N |
| Chromosome 20 genotype intensities | Text | N | N |
| Chromosome 21 genotype intensities | Text | N | N |
| Chromosome 22 genotype intensities | Text | N | N |
| Chromosome X genotype intensities | Text | N | N |
| Chromosome Y genotype intensities | Text | N | N |
| Mitochondrial genotype intensities | Text | N | N |
| Average X chromosome intensities for determining sex | Continuous | N | N |

| | | | |
|---|---|---|---|
| Average Y chromosome intensities for determining sex | Continuous | N | N |
| Genotype measurement batch | Categorical (single) | N | N |
| CEL files | Text | N | N |
| Genotype measurement plate | Text | N | N |
| Genotype measurement well | Text | N | N |
| UKBiLEVE Affymetrix quality control for samples | Categorical (single) | N | N |
| UKBiLEVE genotype quality control for samples | Categorical (single) | N | N |
| Tobacco smoking | Categorical (single) | N | N |
| Age of stopping smoking | Integer | N | N |
| Amount of tobacco currently smoked | Integer | N | N |
| Date first recorded at location | Date | N | N |
| Home location - east co-ordinate (rounded) | Integer | N | N |
| Home location - north co-ordinate (rounded) | Integer | N | N |
| Nitrogen dioxide air pollution; 2010 | Continuous | N | N |
| Nitrogen oxides air pollution; 2010 | Continuous | N | N |
| Particulate matter air pollution (pm10); 2010 | Continuous | N | N |
| Particulate matter air pollution (pm2.5); 2010 | Continuous | N | N |

| | | | |
|---|---|---|---|
| Particulate matter air pollution (pm2.5) absorbance; 2010 | Continuous | N | N |
| Particulate matter air pollution 2.5-10um; 2010 | Continuous | N | N |
| Traffic intensity on the nearest road | Integer | N | N |
| Inverse distance to the nearest road | Continuous | N | N |
| Traffic intensity on the nearest major road | Integer | N | N |
| Inverse distance to the nearest major road | Continuous | N | N |
| Total traffic load on major roads | Integer | N | N |
| Close to major road | Categorical (single) | N | N |
| Sum of road length of major roads within 100m | Continuous | N | N |
| Nitrogen dioxide air pollution; 2005 | Continuous | N | N |
| Nitrogen dioxide air pollution; 2006 | Continuous | N | N |
| Nitrogen dioxide air pollution; 2007 | Continuous | N | N |
| Particulate matter air pollution (pm10); 2007 | Continuous | N | N |
| Source of death report | Categorical (single) | N | N |
| Date of death | Date | N | N |
| Age at death | Continuous | N | N |
| Underlying (primary) cause of death: ICD10 | Categorical (single) | N | N |

| | | | |
|---|---|---|---|
| Contributory (secondary) causes of death: ICD10 | Categorical (single) | N | N |
| Description of cause of death | Curve | N | N |
| Hot drink temperature | Categorical (single) | N | N |
| Water intake | Integer | N | N |
| Major dietary changes in the last 5 years | Categorical (single) | N | N |
| Variation in diet | Categorical (single) | N | N |
| Variation in diet (pilot) | Categorical (single) | N | N |
| Alcohol drinker status | Categorical (single) | N | N |
| Alcohol intake frequency. | Categorical (single) | N | N |
| Former alcohol drinker | Categorical (single) | N | N |
| Average monthly red wine intake | Integer | N | N |
| Date of attending assessment centre | Date | N | N |
| UK Biobank assessment centre | Categorical (single) | N | N |
| Age when attended assessment centre | Integer | N | N |
| Episodes containing "Date of admission to hospital" data | Integer | N | N |
| Episodes containing "Date of discharge from hospital" data | Integer | N | N |
| Episodes containing "Episode end date" data | Integer | N | N |

| | | | |
|---|---|---|---|
| Episodes containing "Episode start date" data | Integer | N | N |
| Episodes containing "Source of inpatient record" data | Integer | N | N |

### 19.2.2 Download

Log in to https://bbams.ndph.ox.ac.uk/ams/resApplications and click Projects>View/Update>Data>Go to Showcase download page>Datasets and click the relevant ID number. Enter the MD5 checksum (32 digit password provided by UKB) and click Generate, then click Fetch on the next page. A file will be downloaded in .enc format.

Download files from http://biobank.ctsu.ox.ac.uk/showcase/download.cgi. There are five files in the “File Handlers” tab (ukbmd5, ukbconv, ukbunpack, ukbfetch, and ukbgene) and one in the “Miscellaneous Utility” tab (encoding.ukb).

### 19.2.3 Decryption

With all the files above in one directory, use the following Linux command. Numbers are specific to the application.

./ukbunpack ukb22142.enc dbd6620a8c5c66999a25fdca0378c052c3fb32b9a24f8ca38d4ad748e8e0889c

UKBiobank ukb_unpack_lx (c) CTSU. Compiled Mar 14 2018 14:21:31.

Attempting Unpack of "ukb22142.enc"

Password: "dbd...89c"

Unpacking, 3 chunks

...

Unpack output as "ukb22142.enc_ukb"

Bytes written: 25827814

MD5 computed: 2993b103a6344dce5478ca21a93d7fa4

### 19.2.4 Formatting

This data must be converted to a format that can be opened by a statistical software package. Supported formats are CSV, SAS, Stata, r, Docs, and Bulk. The command below converts to Stata format.

./ukbconv ukb22142.enc_ukb stata

UKBiobank ukbconv_lx (c) CTSU. Compiled Mar 14 2018 14:21:31.

Opened logfile "ukb22142.log"

4 distinct data fields (9 columns) present in dataset

Field list output to "fields.ukb"

Headers loaded for 9 columns

Done  10,000

Done  20,000

…

Done 490,000

Done 500,000

Starting data processing

Done  10,000

Done  20,000

…

Done 490,000

Done 500,000

Read 502616 rows

502616/502616 data rows processed

Output finished

Open and run the .do file that was generated. This will open a Stata dataset with all the fields downloaded and column headings included. As multiple datasets were downloaded from UKB, it was then necessary to use the “Combine datasets” feature in Stata to merge all data into a single file.

## 19.3 Quality control

### 19.3.1 Sex

Open dataset in Stata and generate new column.

. gen sex_discrepancy=.

(502616 missing values generated)

Label unmatched as 1.

. replace sex_discrepancy = 1 if n_31_0_0 ==1 & n_22001_0_0 == 0

(235 real changes made)

. replace sex_discrepancy = 1 if n_31_0_0 ==0 & n_22001_0_0 == 1

(143 real changes made)

Label matched as 2.

. replace sex_discrepancy = 2 if n_31_0_0 ==0 & n_22001_0_0 == 0

(264622 real changes made)

. replace sex_discrepancy = 2 if n_31_0_0 ==1 & n_22001_0_0 == 1

(223363 real changes made)

Drop subjects whose genetic sex and self-reported sex do not match.

. drop if sex_discrepancy==1

(378 observations deleted)

Drop subjects with sex chromosome aneuploidy.

. drop if n_22019_0_0==1

(408 observations deleted)

### 19.3.2 Ethnicity

Drop subjects whose genetic ethnicity is not Caucasian.

. drop if n_22006_0_0 !=1

(92858 observations deleted)

### 19.3.3 Related individuals

Identify related subjects and sort by missingness.

. sort n_22012_0_0

. gen dups=n_22011_0_0

(392818 missing values generated)

. sort dups n_22005_0_0

. by dups: gen j=_n

For related pairs, drop those with highest missingness.

. drop if j==2

(7122 observations deleted)

Remove columns created for this task.

. drop dups

. drop j

Repeat process for each of the other four “genetic relatedness factor” fields.

. sort n_22012_0_1

. gen dups=n_22011_0_1

(400394 missing values generated)

. sort dups n_22005_0_0

. by dups: gen j=_n

. drop if j==2

(179 observations deleted)

. drop dups

. drop j

. sort n_22012_0_2

. gen dups=n_22011_0_2

(401219 missing values generated)

. sort dups n_22005_0_0

. by dups: gen j=_n

. drop if j==2

(11 observations deleted)

. drop dups

. drop j

. sort n_22012_0_3

. gen dups=n_22011_0_3

(401296 missing values generated)

. sort dups n_22005_0_0

. by dups: gen j=_n

. drop if j==2

(5 observations deleted)

. drop dups

. drop j

. sort n_22012_0_4

. gen dups=n_22011_0_4

(401306 missing values generated)

. sort dups n_22005_0_0

. by dups: gen j=_n

. drop if j==2

(1 observation deleted)

. drop dups

. drop j

### 19.3.4 Other exclusions

Drop subjects with “poor heterozygosity/missingness” as advised by “genomic analysis exclusions” field.

. drop if n_22010_0_0 ==1

(347 observations deleted)

## 19.4 Initial Plink analysis: all types

Remove superfluous variables to reduce file size.

Open dataset in Stata and remove unneeded columns.

. drop n_3005_0_0 n_3005_1_0 n_3005_2_0 [etc]

Combine columns if necessary before removing data.

. replace n_2976_0_0= n_2976_1_0 if missing(n_2976_0_0)

(288 real changes made)

Change missing values to -9 per plink standard.

. mvencode  FID IID n_eid n_31_0_0 n_34_0_0 … n_30520_0_0 n_30530_0_0 waisthipratio kidneytx, mv(-9)

s_20202_2_0: string variable ignored

s_20206_2_0: string variable ignored

…

n_30530_0_0: 11715 missing values recoded

waisthipra~o: 750 missing values recoded

Export file as .csv (tab delimited) to directory with plink and .bed/.bim/.fam files

. export delimited using "reducedpheno.csv", delimiter(tab) nolabel

### 19.4.1 Analysing white British subjects

In Command Prompt, use Plink commands:

plink --bfile ukb_cauc --linear --pheno reducedpheno.csv --pheno-name n_30510_0_0 --out creatinine

PLINK v1.90b3.39 32-bit (5 Aug 2016) https://www.cog-genomics.org/plink2

(C) 2005-2016 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to creatinine.log.

Options in effect:

--bfile ukb_cauc

--linear

--out creatinine

--pheno reducedpheno.csv

--pheno-name n_30510_0_0

32648 MB RAM detected; reserving 2047 MB for main workspace.

Allocated 1535 MB successfully, after larger attempt(s) failed.

362 variants loaded from .bim file.

408972 people (187902 males, 221070 females) loaded from .fam.

390397 phenotype values present after --pheno.

Using 1 thread (no multithreaded calculations invoked).

Before main variant filters, 408972 founders and 0 nonfounders present.

Calculating allele frequencies... done.

Total genotyping rate is 0.998533.

362 variants and 408972 people pass filters and QC.

Phenotype data is quantitative.

Writing linear model association results to creatinine.assoc.linear ... done.

plink --bfile ukb_cauc --logistic --1 --pheno reducedpheno.csv --pheno-name kidneytx --ci 0.95 --out kidneytx

PLINK v1.90b3.39 32-bit (5 Aug 2016) https://www.cog-genomics.org/plink2

(C) 2005-2016 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to kidneytx.log.

Options in effect:

--1

--bfile ukb_cauc

--logistic

--out kidneytx

--pheno reducedpheno.csv

--pheno-name kidneytx

32648 MB RAM detected; reserving 2047 MB for main workspace.

Allocated 1535 MB successfully, after larger attempt(s) failed.

362 variants loaded from .bim file.

408972 people (187902 males, 221070 females) loaded from .fam.

401307 phenotype values present after --pheno.

Using 1 thread (no multithreaded calculations invoked).

Before main variant filters, 408972 founders and 0 nonfounders present.

Calculating allele frequencies... done.

Total genotyping rate is 0.998533.

362 variants and 408972 people pass filters and QC.

Among remaining phenotypes, 258 are cases and 401049 are controls. (7665 phenotypes are missing.)

Writing logistic model association results to kidneytx.assoc.logistic ...done.

### 19.4.2 Covariate analysis

plink --bfile ukb_cauc --linear --pheno reducedpheno.csv --pheno-name n_30510_0_0 --covar reducedpheno.csv --covar-name n_34_0_0, n_31_0_0 --out creatinine_covar

PLINK v1.90b3.39 32-bit (5 Aug 2016) https://www.cog-genomics.org/plink2

(C) 2005-2016 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to creatinine_covar.log.

Options in effect:

--bfile ukb_cauc

--covar reducedpheno.csv

--covar-name n_34_0_0, n_31_0_0

--linear

--out creatinine_covar

--pheno reducedpheno.csv

--pheno-name n_30510_0_0

32648 MB RAM detected; reserving 2047 MB for main workspace.

Allocated 1535 MB successfully, after larger attempt(s) failed.

362 variants loaded from .bim file.

408972 people (187902 males, 221070 females) loaded from .fam.

390397 phenotype values present after --pheno.

Using 1 thread (no multithreaded calculations invoked).

--covar: 2 out of 7 covariates loaded.

7665 people were not seen in the covariate file.

Before main variant filters, 408972 founders and 0 nonfounders present.

Calculating allele frequencies... done.

Total genotyping rate is 0.998533.

362 variants and 408972 people pass filters and QC.

Phenotype data is quantitative.

Writing linear model association results to creatinine_covar.assoc.linear ... done.

plink --bfile ukb_cauc --logistic --pheno reducedpheno.csv --pheno-name kidneytx --1 --covar reducedpheno.csv --covar-name n_34_0_0, n_31_0_0 --ci 0.95 --out kidneytx_covar

PLINK v1.90b3.39 32-bit (5 Aug 2016) https://www.cog-genomics.org/plink2

(C) 2005-2016 Shaun Purcell, Christopher Chang GNU General Public License v3

Logging to kidneytx_covar.log.

Options in effect:

--1

--bfile ukb_cauc

--covar reducedpheno.csv

--covar-name n_34_0_0, n_31_0_0

--logistic

--out kidneytx_covar

--pheno reducedpheno.csv

--pheno-name kidneytx

32648 MB RAM detected; reserving 2047 MB for main workspace.

Allocated 1535 MB successfully, after larger attempt(s) failed.

362 variants loaded from .bim file.

408972 people (187902 males, 221070 females) loaded from .fam.

401307 phenotype values present after --pheno.

Using 1 thread (no multithreaded calculations invoked).

--covar: 2 out of 7 covariates loaded.

7665 people were not seen in the covariate file.

Before main variant filters, 408972 founders and 0 nonfounders present.

Calculating allele frequencies... done.

Total genotyping rate is 0.998533.

362 variants and 408972 people pass filters and QC.

Among remaining phenotypes, 258 are cases and 401049 are controls. (7665 phenotypes are missing.)

Writing logistic model association results to kidneytx_covar.assoc.logistic ...done.

## 19.5 HLA types and frequencies

Table 7: HLA types and frequencies

| Class I | |
|---|---|
| *HLA-* | Frequency |
| *A*02:01* | 26.24% |
| *A*01:01* | 18.71% |
| *C*07:01* | 17.15% |
| *C*07:02* | 15.40% |
| *B*07:02* | 14.14% |
| *A*03:01* | 13.94% |
| *B*08:01* | 13.69% |
| *C*05:01* | 10.65% |
| *B*44:02* | 10.47% |
| *C*06:02* | 9.06% |
| *C*04:01* | 8.87% |
| *C*03:04* | 7.53% |
| *A*24:02* | 7.40% |
| *A*11:01* | 6.20% |
| *B*44:03* | 5.73% |
| *B*15:01* | 5.69% |
| *B*40:01* | 5.24% |
| *C*03:03* | 5.21% |
| *B*35:01* | 4.51% |
| *C*16:01* | 4.33% |
| *A*29:02* | 4.01% |
| *B*57:01* | 3.83% |
| *B*27:05* | 3.70% |
| *B*51:01* | 3.66% |
| *C*08:02* | 3.66% |
| *B*18:01* | 3.62% |
| *C*02:02* | 3.52% |
| *A*32:01* | 3.44% |
| *C*01:02* | 3.40% |
| *C*12:03* | 3.19% |
| *A*68:01* | 3.07% |
| *A*31:01* | 2.61% |
| *B*14:02* | 2.58% |
| *A*26:01* | 2.08% |

| Class II | |
|---|---|
| *HLA-* | Frequency |
| *DPB1*04:01* | 42.36% |
| *DRB3 – no gene* | 35.17% |
| *DRB4 – no gene* | 34.03% |
| *DRB4*01:03* | 24.50% |
| *DQA1*05:01* | 23.03% |
| *DPA1*01:03* | 20.02% |
| *DQA1*03:01* | 19.56% |
| *DQA1*01:02* | 19.14% |
| *DQB1*03:01* | 17.61% |
| *DRB3*01:01* | 16.19% |
| *DRB5 – no gene* | 15.25% |
| *DPA1*02:01* | 15.12% |
| *DQB1*02:01* | 14.62% |
| *DRB1*03:01* | 14.41% |
| *DQA1*02:01* | 14.39% |
| *DRB1*07:01* | 14.34% |
| *DRB3*02:02* | 14.23% |
| *DQA1*01:01* | 14.18% |
| *DRB5*01:01* | 14.18% |
| *DRB1*15:01* | 13.95% |
| *DQB1*06:02* | 13.75% |
| *DQB1*05:01* | 11.94% |
| *DPB1*02:01* | 11.12% |
| *DPB1*04:02* | 10.64% |
| *DRB1*04:01* | 10.28% |
| *DQB1*03:02* | 10.08% |
| *DQB1*02:02* | 9.11% |
| *DPB1*03:01* | 8.83% |
| *DRB1*01:01* | 8.78% |
| *DRB4*01:01* | 7.82% |
| *DPB1*01:01* | 6.24% |
| *DQA1*01:03* | 5.85% |
| *DQB1*03:03* | 5.13% |
| *DQB1*06:03* | 5.07% |

| | |
|---|---|
| *B*13:02* | 1.93% |
| *A*23:01* | 1.89% |
| *C*15:02* | 1.89% |
| *B*55:01* | 1.76% |
| *C*07:04* | 1.71% |
| *A*25:01* | 1.58% |
| *B*37:01* | 1.37% |
| *B*49:01* | 1.20% |
| *A*30:01* | 1.18% |
| *B*14:01* | 1.12% |
| *A*30:02* | 1.09% |
| *C*14:02* | 1.00% |
| *B*35:03* | 0.97% |
| *B*50:01* | 0.92% |
| *B*38:01* | 0.90% |
| *A*02:05* | 0.83% |
| *B*40:02* | 0.78% |
| *C*12:02* | 0.74% |
| *A*68:02* | 0.73% |
| *B*52:01* | 0.73% |
| *C*17:01* | 0.72% |
| *B*58:01* | 0.72% |
| *B*45:01* | 0.69% |
| *B*39:01* | 0.69% |
| *B*39:06* | 0.64% |
| *A*33:01* | 0.61% |
| *A*33:03* | 0.57% |
| *B*53:01* | 0.49% |
| *B*41:02* | 0.36% |
| *A*66:01* | 0.32% |
| *C*03:02* | 0.32% |
| *B*35:02* | 0.29% |
| *B*47:01* | 0.29% |
| *B*56:01* | 0.29% |
| *C*16:02* | 0.24% |
| *B*41:01* | 0.24% |
| *A*02:06* | 0.24% |
| *B*15:17* | 0.23% |
| *B*15:18* | 0.22% |
| *B*07:05* | 0.21% |
| *A*29:01* | 0.18% |
| *B*15:03* | 0.17% |
| *A*74:01* | 0.16% |
| *A*34:02* | 0.16% |
| *A*03:02* | 0.16% |
| *A*02:02* | 0.15% |
| *A*02:03* | 0.14% |
| *B*42:01* | 0.14% |
| *C*08:01* | 0.13% |
| *B*35:08* | 0.13% |
| *B*44:05* | 0.13% |

| | |
|---|---|
| *DRB1*13:01* | 5.01% |
| *DRB3*03:01* | 4.02% |
| *DRB1*13:02* | 3.85% |
| *DPA1*02:02* | 3.79% |
| *DRB1*04:04* | 3.72% |
| *DRB1*11:01* | 3.42% |
| *DQB1*06:04* | 2.70% |
| *DPB1*11:01* | 2.48% |
| *DQB1*05:03* | 2.32% |
| *DPB1*05:01* | 2.18% |
| *DQA1*04:01* | 2.13% |
| *DQB1*04:02* | 2.12% |
| *DRB1*14:01* | 1.90% |
| *DRB1*08:01* | 1.71% |
| *DPB1*13:01* | 1.68% |
| *DRB1*01:03* | 1.59% |
| *DPB1*10:01* | 1.55% |
| *DRB1*12:01* | 1.44% |
| *DRB1*09:01* | 1.33% |
| *DPB1*17:01* | 1.30% |
| *DRB1*11:04* | 1.25% |
| *DQB1*06:09* | 1.10% |
| *DQB1*05:02* | 0.97% |
| *DRB1*01:02* | 0.93% |
| *DRB1*13:03* | 0.90% |
| *DQB1*06:01* | 0.85% |
| *DPB1*14:01* | 0.84% |
| *DRB5*02:02* | 0.83% |
| *DRB1*10:01* | 0.77% |
| *DPB1*15:01* | 0.74% |
| *DPB1*02:02* | 0.74% |
| *DRB1*15:02* | 0.72% |
| *DRB1*16:01* | 0.60% |
| *DRB1*04:05* | 0.54% |
| *DPB1*19:01* | 0.50% |
| *DPA1*01:04* | 0.50% |
| *DRB1*04:07* | 0.44% |
| *DPB1*09:01* | 0.43% |
| *DRB1*04:02* | 0.42% |
| *DPB1*16:01* | 0.39% |
| *DQA1*06:01* | 0.37% |
| *DPB1*06:01* | 0.30% |
| *DRB1*08:03* | 0.28% |
| *DRB1*11:03* | 0.27% |
| *DRB1*15:03* | 0.26% |
| *DRB1*11:02* | 0.24% |
| *DPA1*03:01* | 0.23% |
| *DRB1*08:04* | 0.22% |
| *DRB1*04:03* | 0.17% |
| *DRB1*14:04* | 0.14% |
| *DRB1*03:02* | 0.14% |

| | |
|---|---|
| *B*40:06* | 0.12% |
| *C*15:05* | 0.11% |
| *A*30:04* | 0.11% |
| *A*02:11* | 0.10% |
| *B*15:10* | 0.09% |
| *B*27:02* | 0.09% |
| *C*18:01* | 0.08% |
| *B*51:08* | 0.08% |
| *A*36:01* | 0.08% |
| *B*15:16* | 0.07% |
| *B*15:02* | 0.07% |
| *A*26:08* | 0.07% |
| *C*16:04* | 0.07% |
| *B*58:02* | 0.06% |
| *B*57:03* | 0.06% |
| *B*46:01* | 0.06% |
| *A*24:03* | 0.05% |
| *B*13:01* | 0.05% |
| *C*04:03* | 0.05% |
| *A*69:01* | 0.05% |
| *B*38:02* | 0.05% |
| *B*44:04* | 0.04% |
| *B*48:01* | 0.04% |
| *A*02:07* | 0.04% |
| *B*57:02* | 0.04% |
| *B*81:01* | 0.04% |
| *A*24:07* | 0.04% |
| *A*01:02* | 0.03% |
| *A*80:01* | 0.02% |
| *B*15:24* | 0.02% |
| *B*39:10* | 0.02% |
| *A*66:02* | 0.02% |
| *C*08:04* | 0.02% |
| *B*73:01* | 0.02% |
| *B*54:01* | 0.02% |
| *A*11:02* | 0.02% |
| *B*35:05* | 0.02% |
| *B*27:07* | 0.02% |
| *A*34:01* | 0.02% |
| *B*18:03* | 0.02% |
| *C*15:04* | 0.01% |
| *B*27:04* | 0.01% |
| *B*15:25* | 0.01% |
| *B*78:01* | 0.01% |
| *B*27:03* | 0.01% |
| *B*51:07* | <0.01% |
| *B*51:06* | <0.01% |
| *C*07:26* | <0.01% |
| *C*14:03* | <0.01% |
| *A*11:03* | <0.01% |
| *B*15:13* | <0.01% |

| | |
|---|---|
| *DQB1*05:04* | 0.14% |
| *DQB1*03:04* | 0.12% |
| *DRB1*12:02* | 0.12% |
| *DPB1*23:01* | 0.12% |
| *DPB1*26:01* | 0.11% |
| *DPB1*18:01* | 0.10% |
| *DQA1*03:03* | 0.09% |
| *DRB1*16:02* | 0.09% |
| *DPB1*20:01* | 0.08% |
| *DQA1*05:05* | 0.07% |
| *DRB1*04:08* | 0.06% |
| *DPA1*01:05* | 0.06% |
| *DRB1*08:06* | 0.06% |
| *DRB1*04:06* | 0.05% |
| *DPA1*04:01* | 0.04% |
| *DRB1*08:02* | 0.04% |
| *DQB1*04:01* | 0.04% |
| *DRB1*13:05* | 0.03% |
| *DPB1*85:01* | 0.03% |
| *DPB1*40:01* | 0.03% |
| *DPB1*28:01* | 0.02% |
| *DRB1*13:04* | 0.02% |
| *DPB1*21:01* | 0.01% |
| *DRB1*14:02* | 0.01% |
| *DRB3*02:24* | 0.01% |
| *DRB1*11:11* | <0.01% |
| *DQA1*05:09* | <0.01% |
| *DRB1*14:06* | <0.01% |
| *DRB1*04:10* | <0.01% |
| *DRB1*14:05* | <0.01% |
| *DRB3*02:10* | <0.01% |
| *DQA1*03:02* | <0.01% |
| *DRB1*13:12* | <0.01% |
| *DQB1*06:05* | <0.01% |
| *DRB1*14:03* | <0.01% |
| *DQA1*01:04* | <0.01% |
| *DPB1*34:01* | <0.01% |
| *DQA1*05:03* | <0.01% |
| *DRB1*14:15* | <0.01% |
| *DRB1*08:10* | <0.01% |
| *DRB1*14:10* | <0.01% |
| *DRB1*04:11* | <0.01% |
| *DRB1*08:07* | <0.01% |
| *DRB1*08:11* | <0.01% |
| *DPB1*30:01* | <0.01% |
| *DPB1*33:01* | <0.01% |
| *DRB1*11:06* | <0.01% |
| *DRB1*13:15* | <0.01% |
| *DRB1*13:21* | <0.01% |
| *DPB1*31:01* | <0.01% |
| *DRB1*14:09* | <0.01% |

| | |
|---|---|
| *B*15:21* | <0.01% |
| *A*02:20* | <0.01% |
| *A*02:16* | <0.01% |
| *B*35:43* | <0.01% |
| *A*74:03* | <0.01% |
| *B*51:05* | <0.01% |
| *A*26:03* | <0.01% |
| *B*15:07* | <0.01% |
| *B*35:12* | <0.01% |
| *A*43:86* | <0.01% |
| *A*24:04* | <0.01% |
| *A*31:04* | <0.01% |
| *B*15:26* | <0.01% |
| *B*40:12* | <0.01% |
| *A*24:10* | <0.01% |
| *B*15:04* | <0.01% |
| *B*15:05* | <0.01% |
| *B*39:02* | <0.01% |
| *B*55:02* | <0.01% |
| *A*68:17* | <0.01% |
| *B*15:37* | <0.01% |
| *B*15:27* | <0.01% |
| *B*35:22* | <0.01% |
| *B*15:08* | <0.01% |
| *B*15:20* | <0.01% |
| *B*35:10* | <0.01% |
| *B*27:06* | <0.01% |
| *B*35:09* | <0.01% |
| *B*56:04* | <0.01% |
| *B*39:24* | <0.01% |
| *B*40:09* | <0.01% |
| *B*50:02* | <0.01% |
| *C*03:05* | <0.01% |
| *A*01:03* | <0.01% |
| *B*51:23* | <0.01% |
| *B*59:01* | <0.01% |
| *C*03:06* | <0.01% |
| *B*15:06* | <0.01% |
| *B*41:04* | <0.01% |
| *B*51:02* | <0.01% |
| *B*35:41* | <0.01% |
| *B*15:09* | <0.01% |
| *B*15:15* | <0.01% |
| *B*18:04* | <0.01% |
| *B*39:11* | <0.01% |
| *C*02:10* | <0.01% |
| *B*15:29* | <0.01% |
| *B*44:07* | <0.01% |
| *B*39:03* | <0.01% |
| *B*82:01* | <0.01% |
| *A*40:27* | <0.01% |

| | |
|---|---|
| *DPB1*09:02* | <0.01% |
| *DPB1*45:01* | <0.01% |
| *DRB1*14:44* | <0.01% |
| *DPA1*01:07* | <0.01% |
| *DPB1*22:01* | <0.01% |
| *DPB1*35:01* | <0.01% |
| *DPB1*48:01* | <0.01% |
| *DPB1*05:02* | <0.01% |
| *DPB1*81:01* | <0.01% |
| *DRB1*11:43* | <0.01% |
| *DRB1*14:07* | <0.01% |
| *DRB5*01:02* | <0.01% |

| | |
|---|---|
| *B*18:02* | <0.01% |
| *B*35:20* | <0.01% |
| *B*56:03* | <0.01% |
| *A*02:10* | <0.01% |
| *B*35:17* | <0.01% |
| *B*39:05* | <0.01% |
| *B*48:02* | <0.01% |
| *B*67:01* | <0.01% |
| *B*70:20* | <0.01% |
| *A*02:14* | <0.01% |
| *A*02:64* | <0.01% |
| *A*31:03* | <0.01% |
| *B*15:11* | <0.01% |
| *B*15:12* | <0.01% |
| *B*15:30* | <0.01% |
| *B*15:31* | <0.01% |
| *B*35:04* | <0.01% |
| *B*35:19* | <0.01% |
| *B*35:21* | <0.01% |
| *B*39:08* | <0.01% |
| *B*39:09* | <0.01% |
| *B*39:15* | <0.01% |
| *B*40:04* | <0.01% |
| *B*40:10* | <0.01% |
| *B*48:03* | <0.01% |
| *B*82:02* | <0.01% |
| *C*04:07* | <0.01% |

This table shows all 362 HLA types included in the analysis, split into class I and II and sorted by frequency in the study population. 101 types have frequency >1%, and 53 types have frequency >5%.

## 19.6 Other published work

This section shows three peer-reviewed papers (one that I authored and two that I co-authored) that were published while this project was ongoing. They are discussed in section 9.3 but they are not included in the main text of the thesis as they are not directly related to the research question.

### 19.6.1 Monthly variance in UK renal transplantation activity: a national retrospective cohort study

This paper was published in BMJ Open on 17th September 2019.

BMJ Open

# Monthly variance in UK renal transplantation activity: a national retrospective cohort study

Marcus Lowe,[1] Robert Maidstone,[2,3] Kay Poulton,[1] Judith Worthington,[1] Hannah J Durrington,[2,4] David W Ray,[3,5] David van Dellen,[2,6] Argiris Asderakis,[7] John Blaikley,[2,4] Titus Augustine[2,6]

**To cite:** Lowe M, Maidstone R, Poulton K, *et al*. Monthly variance in UK renal transplantation activity: a national retrospective cohort study. *BMJ Open* 2019;**9**:e028786. doi:10.1136/bmjopen-2018-028786

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-028786).

Received 23 December 2018
Revised 14 July 2019
Accepted 19 August 2019



For numbered affiliations see end of article.

**Correspondence to**
Dr John Blaikley;
john.blaikley@manchester.ac.uk

Prof. Titus Augustine;
titus.augustine@mft.nhs.uk

## ABSTRACT

**Objective** To identify whether renal transplant activity varies in a reproducible manner across the year.

**Design** Retrospective cohort study using NHS Blood and Transplant data.

**Setting** All renal transplant centres in the UK.

**Participants** A total of 24 270 patients who underwent renal transplantation between 2005 and 2014.

**Primary outcome** Monthly transplant activity was analysed to see if transplant activity showed variation during the year.

**Secondary outcome** The number of organs rejected due to healthcare capacity was analysed to see if this affected transplantation rates.

**Results** Analysis of national transplant data revealed a reproducible yearly variance in transplant activity. This activity increased in late autumn and early winter (p=0.05) and could be attributed to increased rates of living (October and November) and deceased organ donation (November and December). An increase in deceased donation was attributed to a rise in donors following cerebrovascular accidents and hypoxic brain injury. Other causes of death (infections and road traffic accidents) were more seasonal in nature peaking in the winter or summer, respectively. Only 1.4% of transplants to intended recipients were redirected due to a lack of healthcare capacity, suggesting that capacity pressures in the National Health Service did not significantly affect transplant activity.

**Conclusion** UK renal transplant activity peaks in late autumn/winter in contrast to other countries. Currently, healthcare capacity, though under strain, does not affect transplant activity; however, this may change if transplantation activity increases in line with national strategies as the spike in transplant activity coincides with peak activity in the national healthcare system.

## Strengths and limitations of this study

- The study was a 10-year retrospective study involving all the registered renal transplant recipients in the UK over this time period.
- The national transplant database is filled using data submitted by each transplant centre; therefore, the data has not been independently corroborated.
- A minority of patients will have refused to be enrolled in the national database potentially affecting our findings.
- The database is used to audit transplant provision and outcomes in the UK; therefore, it was not set up specifically for this research project.

## INTRODUCTION

When organ donors and families of potential donors consent to donation, they make a very valuable gift which is life-changing for the recipient. Despite recent improvements in the UK[1] and other countries regarding access to transplantation, demand for renal transplantation still exceeds the number of available donors.[2] This results in increased patient mortality and morbidity.[3] Since many potential donors cannot be used for various reasons, it is vital that these reasons are minimised so that every potential organ for transplant is used for the primary intended recipient. One of the reasons transplantation does not proceed is due to a lack of clinical capacity. Therefore, healthcare planning plays a key role in ensuring that sufficient capacity exists so that all transplants are used for the primary intended recipients. This could be potentially difficult as the National Health Service (NHS) often works near or at maximum capacity,[4] especially in winter months.

Renal transplantation uses both living and deceased donors. Human mortality rates are known to oscillate in a seasonal manner for some diseases.[5] This can be attributed to the effects of endogenous seasonal rhythms and climatic factors on human performance and activity patterns. In the UK, seasonal variations are commonly observed for infectious diseases, such as influenza,[6] cerebrovascular disease[7–12] and myocardial infarctions.[13] Several studies have recently shown how understanding these oscillations are crucial for planning and delivering optimum

healthcare delivery.[14] For example, the USA covers many different climate zones, but depite this cerebrovascular accidents in the USA peak in the winter[15] in a similar pattern to the UK.[7–12] Surprisingly, transplant activity in the USA is lowest in the winter months for both renal[16] and heart transplantation, despite these deaths being a leading cause of organ donation.[17] A similar pattern is also seen in Italy,[18] suggesting that this seasonal pattern is conserved in other countries.

To the best of our knowledge, seasonal fluctuations in transplant activity have not been investigated in the UK. We, therefore, examined the national cohort over a 10-year period to establish both seasonal fluctuations and whether healthcare capacity influences activity.

## METHODS

All UK renal adult transplants, performed between 2005 and 2014 to recipients over 18 years old, were included in the study. Data were provided by NHSBT (NHS Blood and Transplant) who maintain a comprehensive national database on the 24 adult kidney transplant centres in the UK. This data was combined with data collected on donor activity, provided by the National Organ Retrieval Service. Deceased donation is categorised into donation after brain death and donation after circulatory death. All donor deaths are further classified according to the cause of mortality. 'Cerebrovascular event' was defined as death due to intracranial haemorrhage, intracranial thrombus or unclassified intracranial event. 'Infection' was defined as death due to meningitis, septicaemia, pneumonia or unclassified infection. Deceased donor transplants form about two-thirds of all UK kidney transplant activity. The other one-third of kidney transplants are transplants from living donors. Living donor transplants are logistically different as they are planned and scheduled elective cases depending on several factors which can usually be controlled. Deceased donor transplants, on the other hand, are unplanned occurring when organs become available after the death of a suitable donor. Recipients are then allocated organs according to agreed national allocation criteria.

Data were analysed according to the month and year the transplant occurred. Donation details were analysed in a similar manner. The observed data were compared with the expected transplant activity calculated by measuring the total number of transplants divided by the number of days in the same year.

Data were also analysed according to organs offered to different centres for named patients and declined, for various reasons by the centre, including capacity issues. These issues for the purposes of this paper were subclassified into different categorical values (centre already transplanting, no beds, no staff, no theatre and no time).

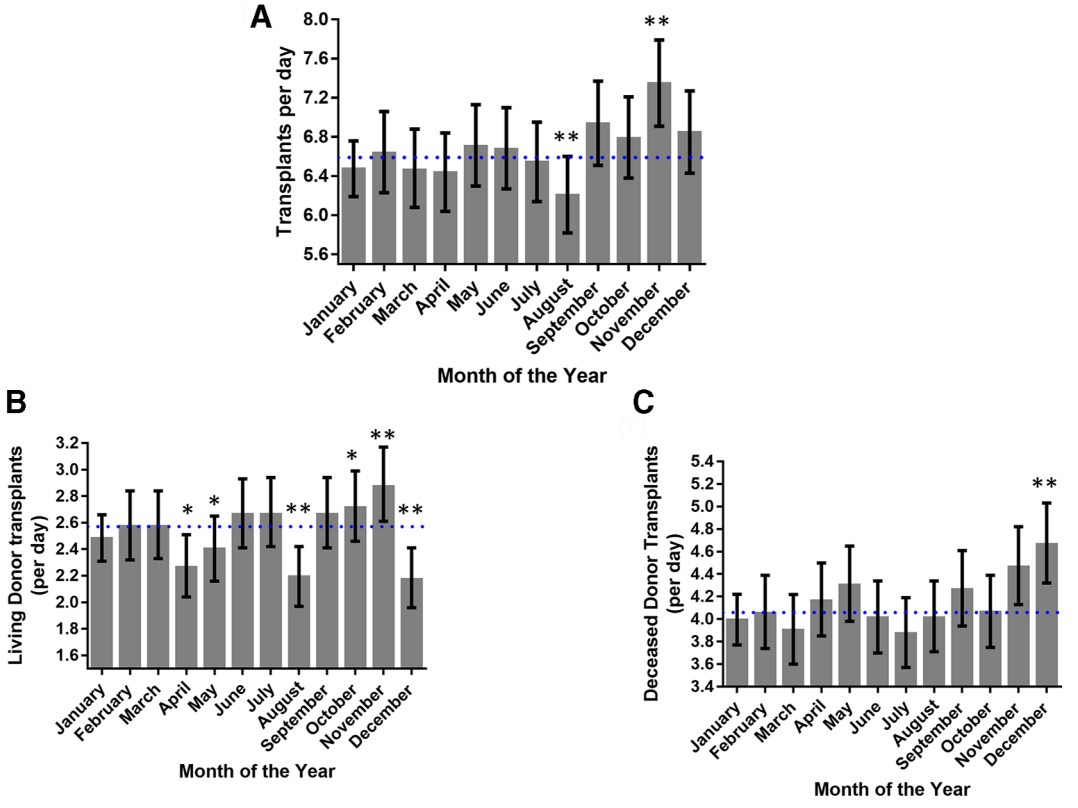



**Figure 1** Changes in transplant activity during the year. (A) The number of UK renal transplants per month varied significantly during the year (p<0.01) peaking in November. (B) Renal transplants arising from living donors varied significantly throughout the year peaking in October and November. (C) Transplant activity using deceased donors varied significantly over the year peaking in December. (*P<0.05, **p<0.01 Pearson residual, dotted line shows average activity).

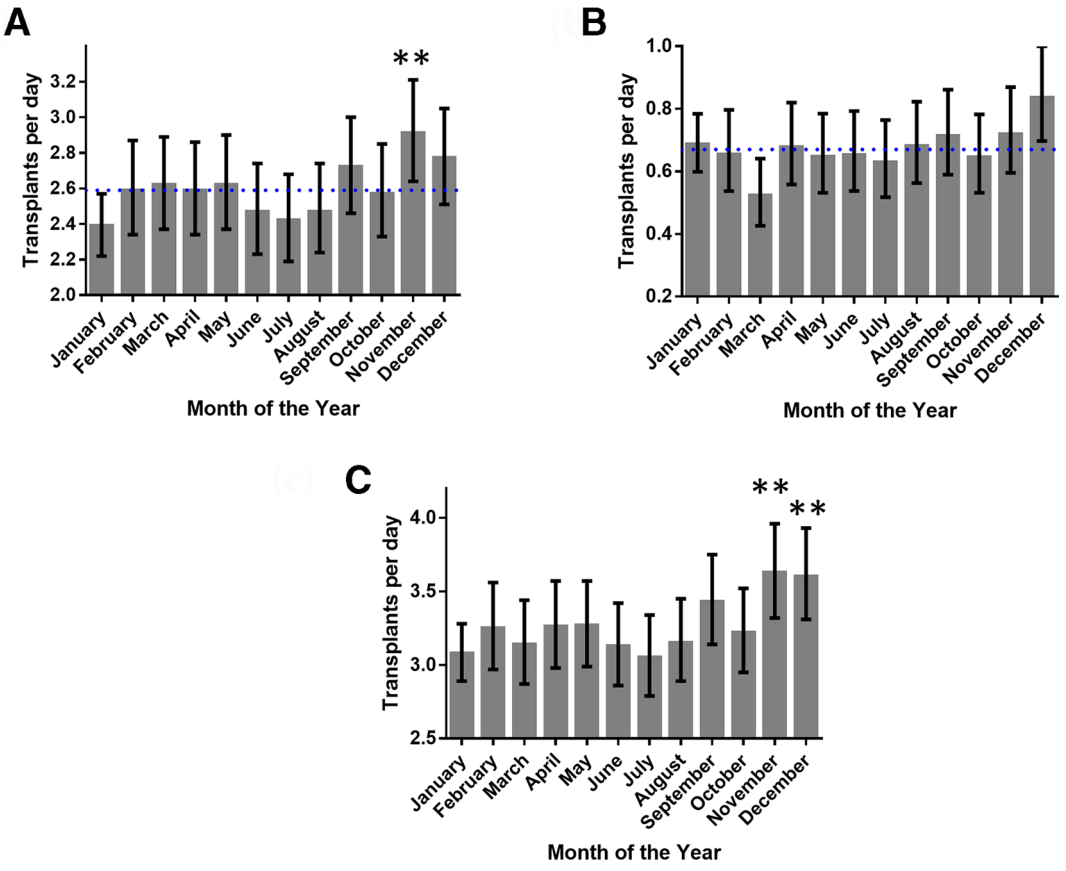



**Figure 2** Changes in transplant activity using donors after cerebrovascular events or hypoxic brain injury. (A) Transplant activity using donors following cerebrovascular events peaks in November. (B) Transplant activity using donors with hypoxic brain damage also increased in activity at the end of the year (November and December). (C) By combining both these causes, the graph mirrors the fluctuation seen in transplant activity for all deceased donors (shown in figure 1C). (*P<0.05, **p<0.01 Pearson residual, dotted line shows average activity).

### Statistical methods

Generalised linear models (GLMs), with a Poisson link function, were used to model the effect that month of the year had on our data. Offset variables were used to account for the slight differences in the number of days per month. To account for the unknown correlation structure in our data, generalised estimating equations (GEEs) per year were used to estimate the GLM parameters.[19]

The $\chi^2$ goodness-of-fit tests assessed whether the observed number of transplants per month differed from the expected value if there was no seasonal variability. When this test was significant (ie, the expected values were significantly different from the observed values), Pearson residuals were used to identify the particular months which caused this mismatch. Residuals that are greater than two in magnitude suggest some degree of lack of fit.[20]

Sine and cosine curves were fitted to the transplant data to investigate whether repeating patterns occurred across the months. All analyses were performed using SPSS V.22 (IBM corp.).

### Patient and public involvement

Patients and the public were not involved in the design or analysis of data for this study.

### Ethics

Data were collected by NHSBT and analysed after obtaining approval from them in accordance with their policies. Since the data were fully anonymised, no ethical permission was sought.

## RESULTS

During the 10-year study period, 24270 adult kidney transplants were performed in the UK. A total of 15094 (62%) were from deceased donors and 9166 (38%) were from living donors.

### Kidney transplant activity changes in a consistent manner throughout the year

Kidney transplant activity varied by 17.1% within a year (p<0.01, $\chi^2$, figure 1A). Transplant activity was highest in November (11% increase from mean, p<0.01) and lowest in August (6% decrease from mean, p<0.01). Investigating seasonal variation, transplant activity increased in the autumn compared with spring and summer (p<0.05, one-way analysis of variance); these patterns were consistent every year during the 10-year study period (p<0.01, GEE).

Kidney transplantation uses both living and deceased donors. Our data set was, therefore, examined to see

**Table 1** Seasonal variation in kidney transplants using organs from deceased donors

| Cause of death in the donor | n | P value for whether donation varied across the year |
|---|---|---|
| Intracranial/cerebrovascular causes | 9452 | <0.01 |
| Hypoxic brain damage | 2448 | <0.01 |
| Trauma | 1532 | <0.01 |
| Infective causes | 367 | <0.01 |
| Cardiovascular | 237 | 0.13 |
| Respiratory/pulmonary causes | 192 | 0.17 |
| Brain tumours | 180 | 0.14 |
| Poisoning/drug overdose | 55 | 0.03 |
| Organ failure (various causes) | 53 | <0.01 |
| Other/unknown | 578 | 0.01 |

All deceased donations over a 10-year period were split into groups according to aetiology defined by NHS Blood and Transplant. The incidence for the majority (7/10) of causes varied significantly during the year.

**Table 3** Seasonal variation in kidney transplants using organs from donors after circulatory death

| Cause of death in the donor | n | P value for whether donation varied across the year |
|---|---|---|
| Intracranial/cerebrovascular causes | 2467 | 0.60 |
| Hypoxic brain damage | 1216 | 0.16 |
| Trauma | 562 | <0.01 |
| Infective causes | 107 | 0.05 |
| Cardiovascular | 166 | 0.88 |
| Respiratory/pulmonary causes | 185 | 0.25 |
| Brain tumours | 31 | 0.07 |
| Poisoning/drug overdose | 29 | <0.01 |
| Other/unknown | 228 | <0.01 |

All donations after circulatory death over a 10-year period were split into groups according to aetiology defined by NHS Blood and Transplant. The incidence for a minority (3/9) of causes varied significantly during the year.

if one of these types was responsible for the variance in activity. Transplant activity using organs from living donors (figure 1B) significantly increased in October and November and decreased in December, April, May and August. Transplant activity using organs from deceased donors (figure 1C) increased in December and a similar trend was observed in November. Therefore, the increase in activity for November, which has the highest activity, is due to an increase in transplants using kidneys from both living and deceased organ donors. This contrasts with August, which had the lowest activity, and is solely explained by a fall in living donor activity.

**Table 2** Seasonal variation in kidney transplants using organs from donors after brain death

| Cause of death in the donor | n | P value for whether donation varied across the year |
|---|---|---|
| Intracranial/cerebrovascular causes | 6985 | <0.01 |
| Hypoxic brain damage | 1232 | <0.01 |
| Trauma | 970 | <0.01 |
| Infective causes | 260 | <0.01 |
| Cardiovascular | 71 | <0.01 |
| Respiratory/pulmonary causes | 7 | 0.18 |
| Brain tumours | 149 | 0.57 |
| Poisoning/drug overdose | 26 | <0.01 |
| Other/unknown | 350 | 0.21 |

All donations after brain death over a 10-year period were split into groups according to aetiology defined by NHS Blood and Transplant. The incidence for the majority (6/9) of causes varied significantly during the year.

### Incidence for brain injury donors fluctuates over the year

Donation from cerebrovascular and hypoxic brain injury accounted for 78.8% of all deceased donor transplants; therefore, we hypothesised that the spike in deceased donor transplants is due to an increase in donors from these categories during November and December. Cerebrovascular deaths, as a cause of donor death, were significantly increased in November (12% increase from the mean, p<0.01) (figure 2A). Donations from donors who had experienced a hypoxic brain injury tended to increase during this period with the highest spike being in December (25% increase from the mean, p=0.15) (figure 2B). When these results are combined, they mirror the fluctuations seen for all deceased organ transplant activity (figure 2C) with significant elevations in transplant activity during November and December (p<0.01).

### Seasonal variance in the incidence of donors after infection

We also noted that other causes of donor death significantly varied over the year (table 1). Interestingly, this variance was mainly observed in donations after brain death (table 2) rather than circulatory death (table 3). Donations after road traffic accidents (figure 3A) were higher during late spring, summer and early autumn (April–September) compared with other times of the year. Donations from donors dying from infective causes peaked in the winter and declined in the summer (figure 3B); this relationship could be explained by a cosine curve (figure 3B, p<0.05) suggesting a potential underlying seasonal oscillation.

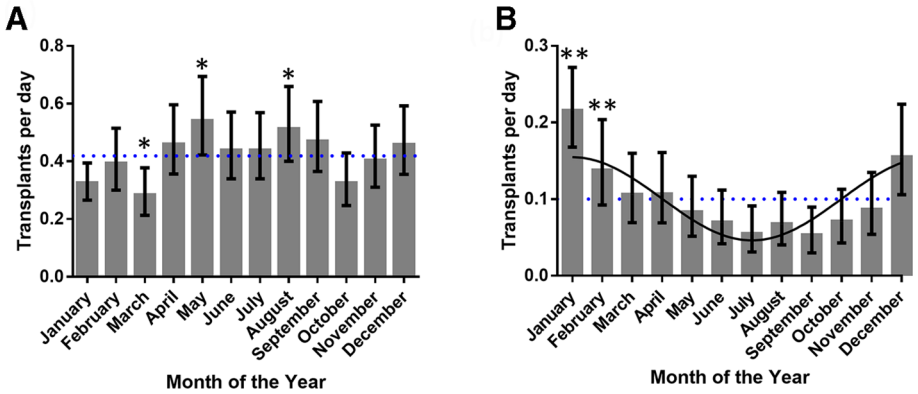



**Figure 3** Seasonal variation in transplant activity arising after infection or trauma. (A) Transplant activity using donors following traumatic incidents increased in the summer (B) as opposed to transplant activity using donors following infection, which oscillated in a sinusoidal manner peaking in the winter. (*P<0.05, **p<0.01 Pearson residual, best fit line for infection is a cosine (p<0.05), dotted line shows average activity).

### The effect of healthcare capacity on transplant activity

The UK healthcare system routinely works at high capacity,[21] potentially resulting in reduced transplantation rates due to bed and staffing shortages. We, therefore, examined whether the reproducible monthly changes in transplant activity could be attributed to a lack of capacity in the healthcare system. Although a large number 28 789 of deceased donor offers were declined for transplantation over the study period only a small fraction of these were due to lack of capacity (n=480, mean 1.22% (±0.22% SD)).

## DISCUSSION

The NHS in the UK is coping with increased pressures, especially in the winter.[21] This study reveals that UK kidney transplant activity peaks in November and December, at the beginning of the winter, in contrast with a number of European[18 22] and North American centres.[16 17] This unexpected finding is due to variations in living and deceased donation. Although detailed causal analysis was beyond the scope of the study, the winter increases in deceased donation could be attributed to increases in both cerebrovascular and hypoxic brain events. This is consistent with findings from previous studies studying the incidence of these events both in the UK[9] and other northern hemisphere countries.[11 12]

The winter surge in transplant activity has important implications for the UK health system (NHS). During the UK winter, there is also a surge in emergency admissions to hospitals, placing the system under significant strain, sometimes resulting in cancellation of elective and semi-elective operations for up to several months.[23] It was, therefore, reassuring that we found at the national level no solid evidence of transplant surgery using deceased donors being cancelled during the winter months. Despite this, however, it is important that individual transplant departments plan for this predictable and reproducible surge in transplant activity making sure that their own activity is not affected. We are confident that this surge is likely to continue into the future as two different statistical tests ($\chi^2$ and GEE) produced similar results. This is especially important as UK transplant activity is likely to increase due to recent changes in legislation.

Certain limitations should be noted when interpreting the results of this study. The donor rates reported in the paper are from used donations, and therefore can be influenced by changes in donor conversion rates; the proportion of donor offers used in a transplant operation. This is unlikely, however, as these changes would have to be consistently occurring in the same way each year for 10 years; furthermore, the peak of deceased donation activity coincided with the epidemiological peaks for the underlying diseases.[7–12] Since this was a retrospective study, there is always the potential for inherent bias, but since this is relatively a large data set, it should minimise this effect. The data from each centre could not be independently verified, due to anonymisation in the data set. Finally, the effect of healthcare capacity on living donations was not investigated, but this would be an important area to investigate in future research.

Our study clearly shows that UK renal transplant activity increases in the winter in contrast to previous studies investigating seasonal transplant activity in other countries. This could have implications for the UK health system since winter is when the British health system is placed under maximal strain. This seasonal variation should, therefore, be considered for any future planning especially with the potential impact of the opt-out legislation and strategies to increase organ donation and transplantation.

**Author affiliations**
[1]Transplantation Laboratory, Manchester University NHS Foundation Trust, Manchester, UK
[2]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[3]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK
[4]Department of Respiratory Medicine, Manchester University NHS Foundation Trust, Manchester, UK
[5]NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[6]Transplant and Endocrine Surgery, Manchester University NHS Foundation Trust, Manchester, UK
[7]Cardiff Transplant Unit, University Hospital of Wales, Cardiff, UK

**Acknowledgements** The authors would like to acknowledge Phil Foden for his contribution to the statistical analysis of the results, and NHSBT and UK renal transplant centres for collecting and providing the data.

**Contributors** TA and JB conceived the study. TA, JB and ML obtained and analysed the data. RM did the statistical analysis. KP, JW, HJD, DvD, DWR and AA substantially contributed to the interpretation of the data. TA, JB and AA drafted the manuscript, with the other authors who revised it critically.

**Funding** JB holds an MRC Clinician Scientist Award (MR/L006499/1). DWR holds a Wellcome Trust Investigator Award (107851/Z/15/Z) and a Medical Research Council grant (MR/P023576/1). HJD is supported by an Asthma UK Senior Clinical Academic Development Award (AUK-SCAD-2013-229). JB and HJD are partially supported by the Manchester NIHR Biomedical Research Centre.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request, but may be subject to approval from NHSBT. The statistical code has been uploaded onto a data repository (doi: https://dx.doi.org/10.17632/48nxwvcfnh.1).



## REFERENCES

1. Transplant NBa. *Organ donation and transplantation activity report 2017/18*, 2018.
2. Pyart R, Wong E, Sharples E, *et al*. Chapter 3 demographic and biochemistry profile of kidney transplant recipients in the UK in 2016. *Nephron* 2018;139:75–104.
3. Steenkamp R, Pyart R, Fraser S. Chapter 5 survival and cause of death in UK adult patients on renal replacement therapy in 2016. *Nephron* 2018;139:117–50.
4. Scriven N. Capacity issues need to be tackled by looking at all of acute provision. *Br J Hosp Med* 2018;79:304–5.
5. Rolden HJA, Rohling JHT, van Bodegom D, *et al*. Seasonal variation in mortality, medical care expenditure and institutionalization in older people: evidence from a Dutch cohort of older health insurance clients. *PLoS One* 2015;10:e0143154.
6. Paules C, Subbarao K. Influenza. *Lancet* 2017;390:697–708.
7. Anlar O, Tombul T, Unal O, *et al*. Seasonal and environmental temperature variation in the occurrence of ischemic strokes and intracerebral hemorrhages in a Turkish adult population. *Int J Neurosci* 2002;112:959–63.
8. Steenhuijsen Piters WAA, Algra A, Broek MFM, *et al*. Seasonal and meteorological determinants of aneurysmal subarachnoid hemorrhage: a systematic review and meta-analysis. *J Neurol* 2013;260:614–9.
9. Haberman S, Capildeo R, Clifford Rose F. The seasonal variation in mortality from cerebrovascular disease. *J Neurol Sci* 1981;52:25–36.
10. Mao Y, Schnytzer Y, Busija L, *et al*. "MOONSTROKE": Lunar patterns of stroke occurrence combined with circadian and seasonal rhythmicity—A hospital based study. *Chronobiol Int* 2015;32:881–8.
11. Sipilä JOT, Ruuskanen JO, Kauko T, *et al*. Seasonality of stroke in Finland. *Ann Med* 2017;49:310–8.
12. Telman G, Sviri GE, Sprecher E, *et al*. Seasonal variation in spontaneous intracerebral hemorrhage in northern Israel. *Chronobiol Int* 2017;34:563–70.
13. Yang L, Li L, Lewington S, *et al*. Outdoor temperature, blood pressure, and cardiovascular disease mortality among 23 000 individuals with diagnosed cardiovascular diseases from China. *Eur Heart J* 2015;36:1178–85.
14. Stevenson TJ, Visser ME, Arnold W, *et al*. Disrupted seasonal biology impacts health, food security and ecosystems. *Proceedings Biological sciences* 1817;2015.
15. Kalkstein AJ. Regional similarities in seasonal mortality across the United States: an examination of 28 metropolitan statistical areas. *PLoS One* 2013;8:e63971.
16. Rosendale JD. Organ donation in the United States: 1988-2003. *Clin Transpl* 2004:41–50.
17. Grodin JL, Ayers CR, Thibodeau JT, *et al*. Variation of heart transplant rates in the United States during holidays. *Clin Transplant* 2014;28:877–82.
18. Santori G, Fontana I, Bertocchi M, *et al*. A time series analysis performed on a 25-year period of kidney transplantation activity in a single center. *Transplant Proc* 2010;42:1080–3.
19. Liang K-YEE, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
20. Menard S. *Applied logistic regression analysis*. 2nd ed. USA: Sage, 2002.
21. Iacobucci G. NHS "winter" pressure could become norm in summer, experts warn. *BMJ* 2016;352.
22. Leal J, Caballero F, García-Sousa S, *et al*. Distribution of organ donors on a Weekly and annual basis according to cause of death. *Transplant Proc* 1999;31:2602–3.
23. Iacobucci G. NHS cancels planned surgery and outpatient appointments in response to winter crisis. *BMJ* 2018;360.

### 19.6.2 The impact of the COVID-19 pandemic on renal transplantation in the UK

This paper was published in Clinical Medicine on 16th July 2020.

# The impact of the COVID-19 pandemic on renal transplantation in the UK

**Authors:** Videha Sharma,[A] Alex Shaw,[A] Marcus Lowe,[B] Angela Summers,[C] David van Dellen[D] and Titus Augustine[E]

ABSTRACT

**COVID-19 is impacting provision of renal transplantation in the UK with a reduction in clinical activity. Publicly available Renal Registry and NHS Blood and Transplant reports were analysed to model the number of missed transplant opportunities, waiting list size and change in dialysis population over a six-month period starting 5 March 2020. An estimated 1,670 kidney transplant opportunities may be lost, which will lead to 6,317 active patients on the kidney-alone waiting list, compared to 4,649 based on usual activity estimates. This will result in 1,324 additional patients on dialysis who would otherwise have been transplanted. COVID-19 will lead to a marked loss of transplant opportunities and a significantly larger national waiting list. The existing strain on dialysis capacity will be exacerbated as patients remain on dialysis as the only available form of renal replacement therapy. These findings will help inform policy and service specific strategies.**

**KEYWORDS:** COVID-19, kidney transplantation, NHS, renal insufficiency, renal replacement therapy

**DOI:** 10.7861/clinmed.2020-0183

## Introduction

Kidney transplantation is the most common form of renal replacement therapy in the UK and provides a sustained cure for end-stage renal disease (ESRD).[1] Transplant activity is coordinated centrally by NHS Blood and Transplant (NHSBT) and over 3,000 renal transplants across 23 centres are performed each year.[2] Transplantation provides significant benefits compared to dialysis and is the renal replacement therapy of choice in all eligible patients.[3] In spite of increasing access to transplantation in the UK, deceased donor waiting times remain high with a median wait of 3 years, predominantly due to a paucity of suitable organ donors.[2] Most patients are therefore established on dialysis for some time prior to transplantation.[4] Several national strategies are in place to increase transplant rates including the use of marginal donors and the opt-out consent system.[5]

The global COVID-19 pandemic reached the UK on 31 January 2020, with the first recorded mortality on 5 March 2020.[6] Caused by the novel coronavirus (SARS-CoV-2), it mostly presents as a respiratory illness, although 25% can have gastrointestinal symptoms.[7] Disease severity can range from mild flu-like symptoms to an acute respiratory distress syndrome with the requirement for ventilatory support.[8] It is highly transmissible and carries a significantly higher mortality than other coronaviruses.[9] As the pandemic establishes itself in the UK, the NHS has been forced to rapidly evolve existing care models in preparation for significant front-line service pressures. In particular, NHS England mandated the suspension of all non-urgent elective surgery from 15 March 2020 to create hospital capacity for COVID-19 related activity. The majority of UK transplant centres appropriately suspended kidney transplant activity in response to this national challenge. In particular, living donor transplantation was halted at all centres and a limited number of units continue with selected deceased donor activity.[10] Decisions to continue deceased donor transplantation may be based on local severity of the pandemic as there is a geographic variation across the UK. Intensive care unit (ICU) capacity, emergency theatre access and staff resources further play a role in centres' capacity to continue to deliver services. The uncertainty in risks combined with resource and workforce challenges make decisions around transplantation highly complex at present.

The consequences of infection with novel coronavirus for renal transplant recipients remain poorly characterised. Post-transplant immunosuppressed patients are known to be susceptible to opportunistic viral infection, with coronaviruses previously reported as a causative agent.[11] The role of immunosuppression in the context of COVID-19 is still being explored and potential donor transmission rates are unestablished. Early reports suggest elective surgery on asymptomatic patients during the incubation period of COVID-19 carries a significant risk of early post-operative mortality.[12–14] The overall impact for the UK is an acute reduction in

**Authors:** [A]clinical research fellow in transplantation, Manchester University Hospitals NHS Foundation Trust, Manchester, UK and University of Manchester, Manchester, UK; [B]data and audit manager, Manchester University Hospitals NHS Foundation Trust, Manchester, UK; [C]senior research fellow in transplantation, Manchester University Hospitals NHS Foundation Trust, Manchester, UK and University of Manchester, Manchester, UK; [D]consultant transplant surgeon, Manchester University Hospitals NHS Foundation Trust, Manchester, UK and University of Manchester, Manchester, UK; [E]professor of clinical transplantation and consultant transplant surgeon, Manchester University Hospitals NHS Foundation Trust, Manchester, UK and University of Manchester, Manchester, UK

renal transplant activity, but the effects on waiting list times and dialysis provision are unquantified. The immediate national focus is appropriately on pandemic response, but consideration should be given to planning and resource allocation for medium- and long-term effects, which include challenges for service planning and resource allocation.

This study aims to help quantify the impact of the COVID-19 pandemic on renal transplantation in the UK.[15] The primary aim is to predict the monthly number of transplants that will not be performed ('missed opportunities') due to suspension of transplant activity secondary to COVID-19 and model the size of the prospective waiting list. The secondary aim is to predict how many additional patients may require dialysis due to lack of transplantation.

## Methods

The latest NHSBT and UK Renal Registry (UKRR) annual reports were included in the study.[2,16,17] These national reports are collated by NHSBT and the UK Renal Association respectively. Both reports include patient-level data submitted by all 23 transplant units and 84 renal centres in the UK. NHSBT also maintains an overview of national waiting list figures and living and deceased donor transplant activity.

The day of the first reported death in the UK attributed to COVID-19 (5 March 2020) was used as 'day 0' to model the loss of transplant opportunities during the pandemic.[6] The UK government has modelled a minimum of a 3-month period of impact, but this remains an estimation. For this reason, we decided to model a period of 3–6 months. The modelling was performed on 2 April 2020.

### Lost transplant opportunities

All adult and paediatric kidney transplants from 1 April 2018 to 31 March 2019 were included. This included kidney-alone transplants, and kidney transplants forming part of a multi-organ transplant such as simultaneous pancreas–kidney transplant. We utilised the reported activity of the 12-month period as stated in the latest report to model the expected activity that would have taken place normally without any pandemic effect. For purposes of this study, we worked on the assumption that transplant activity would be uniformly even across the 12 months. This provided the number of 'missed opportunities' per month, and allowed modelling of the size of the prospective transplant waiting list and dialysis population.

### Kidney transplant waiting list

The kidney-alone transplant waiting list at 'day 0' was used as a starting point for modelling the size of the waiting list at 3–6 months. Over the course of the last decade the kidney transplant waiting list has been gradually declining. This is a result of increasing transplant activity, though this appears to have plateaued over the last two years. A number of patients die on the waiting list and a number are removed, typically because they become too unwell for a transplant.

### UK dialysis population

The UKRR records the haemodialysis and peritoneal dialysis capacity of all UK renal centres. These were collated to provide an estimate on prevalent dialysis numbers on 'day 0', as well as new incident patients starting dialysis over the study period. These figures were used to model the size of the additional UK dialysis population as a result of curtailed transplant activity during the pandemic.

### Modelling analysis

The modelling assumed that activity during the study period would be identical to the figures reported in the latest NHSBT and UKRR reports. We identified the key coefficients such as number of transplants performed, waiting list size and prevalent dialysis numbers and modelled these with the variable of number of months since 'day 0'. Modelling was performed on Microsoft Excel.® Detailed formulae are available in the supplementary data.

### Ethical approval

This study analysed data from publicly available NHSBT and UKRR reports. No explicit consent to use this data was deemed necessary as they are publicly available and the reports were cited accurately. Since the data in the reports are fully anonymised, no ethical permission was required.

## Results

We included data on 3,597 kidney transplants in the 2018/2019 NHSBT annual report over a 12-month period.[2,16] There were 1,020 living donor kidney transplants and 2,577 deceased donor transplants. The overall active kidney transplant waiting list at the end of the financial year (31 March 2019) was 4,954 patients, which was a 2% decrease from the previous year.

### Lost transplant opportunities

There was a mean of 300 kidney transplants each month between 1 April 2018 and 31 March 2019 (85 living donor and 215 deceased donor transplants). This indicates that routine activity over a 3-month period would predict 900 kidney transplants. Between 'day 0' and the time of the analysis, 128 kidney transplants were performed. The COVID-19 pandemic is thus expected to lead to a potential loss of up to 1,672 transplant opportunities between 5 March 2020 and the 5 September 2020 (6 months). Table 1 shows the number of transplants projected based on the anticipated activity in normal and pandemic circumstances, as well as the number of lost opportunities.

### Kidney transplant waiting list

There were 4,748 patients active on the list on 5 March 2020 ('day 0'). The modelling has shown that between 772 and 1,672 lost opportunities will result in a 16.3%–35.2% increase in waiting list size (Fig 1). Projecting the trend of a 4.2% reduction between 31 March 2019 and 5 March 2020 ('day 0'), the size of the list would have normally decreased by 1–2% over the same period. If activity continued as normal, the list would have expected to reduce by 17 patients a month; however, the modelling predicts a relative increase of up to 283 patients per month.

### UK dialysis population

Many transplant opportunities missed during the COVID-19 pandemic would have benefitted patients currently on dialysis.

**Table 1. Expected cumulative number of transplants performed in normal times and during the COVID-19 pandemic**

| | 3 months | 4 months | 5 months | 6 months | Lost opportunities |
|---|---|---|---|---|---|
| **Normal activity** | 900 | 1200 | 1500 | 1800 | n/a |
| **COVID-19 pandemic*** | 128 | 128 | 128 | 128 | 772–1672 |

*Only includes transplants performed from 5 March 2020 until time of analysis (2 April 2020).

62% of living donor transplants go to dialysis patients and a 3–6-month suspension of the living donor programme would lead to 159–317 additional dialysis patients. The reduction of the deceased donor activity has even greater impact, as 87% of deceased donor transplants are for established dialysis patients. With nil deceased donor transplantation, a further 1,121 patients may be on dialysis who would have been transplanted under normal circumstances. Table 2 shows the number of additional dialysis patients there will be, based on the projected number of missed transplant opportunities.

The reduction in transplant activity will result in a total of 720–1,438 patients continuing to require dialysis who would otherwise have been transplanted. The UKRR states there were 29,064 dialysis patients in the UK on 31 December 2017, and the number had been increasing by 2.0% per year over the previous two years. Assuming this trend continued until 'day 0', there would have been approximately 30,337 dialysis patients in the UK. The figures from our modelling would thus increase the dialysis population in addition to normal growth by a further 4.7% over 6 months.

**Table 2. Additional dialysis patients in the UK over the next 3–6 months**

| Months after 'day 0' | Additional dialysis patients due to missed transplants from: | | |
|---|---|---|---|
| | Deceased donors | Living donors | Total |
| **3** | 561 | 159 | 720 |
| **4** | 748 | 211 | 959 |
| **5** | 935 | 264 | 1199 |
| **6** | 1,121 | 317 | 1,438 |

*This has not been adjusted for the 128 transplants recorded from 'day 0' till time of analysis (2 April 2020).

## Discussion

This study has quantified the impact of the COVID-19 pandemic on renal transplant activity in the UK. The modelled loss of transplant opportunities will result in substantially increased waiting list numbers and times. This will impact dialysis capacity and provision, with additional inevitable morbidity and mortality. This study demonstrates the value of publicly available UK-based national data, allowing rapid analysis, predict service impact and inform urgent strategic policy-making.

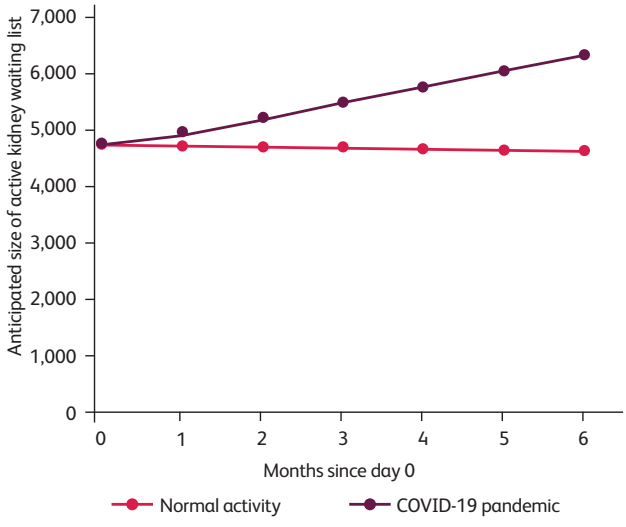



**Fig 1. Predicted change in the size of the kidney-alone waiting list during the COVID-19 pandemic.**

### Transplantation during the COVID-19 pandemic

There has been no national policy to halt organ donation and transplantation across the UK. At the time of analysis, however, national activity has come to a near stand-still.[10] This study shows that from 5 March 2020 onwards, between 772 and 1672 transplant opportunities will be missed over 3–6 months. This will greatly impact patients and families. Concerted efforts must be made to provide communication and support during the pandemic. Using digital technology and telemedicine is proving crucial in maintaining patient contact and health service delivery.[18] Transplant patients have been shown to have favourable attitudes towards using health information technology and COVID-19 may be a stimulus to drive change in the way services are delivered.[19,20]

Transplantation provides superior outcomes and quality of life parameters for ESRD; however, the adjuvant immunosuppression required during and after the surgical procedure may expose recipients to additional risk.[21,22] Recent case series from the UK and Italy have reported a mortality of 7–25% of COVID-19 in post-transplant patients, adding to the concern that these patients suffer significantly adverse outcomes.[23,24] A post-mortem analysis of native kidneys from six patients with confirmed COVID-19 demonstrated a high tropism of SARS-CoV-2 for the human kidney.[25] Across Europe, many national bodies continue to support transplantation but stress the importance of donor COVID-19

testing.[26,27] This is advocated to reduce the risk of transmission to recipients, but also to protect organ retrieval teams. The primary challenge appears to be that a clear risk/benefit assessment of transplantation during the active COVID-19 pandemic is still unclear. Guidance published thus far provides limited objective data for detailed risk/benefit discussions with patients.[10,28] While the current onus is on individual centres to formulate local practice, national strategies are mandated to provide clear guidance to both maintain activity but more importantly provide quantifiable risk stratification. This requires a focus on understanding the consequences of COVID-19 in the context of immunosuppression, and establishing safe approaches to donation. It is imperative that the transplant community shares knowledge and experience rapidly to drive research and form a readily accessible evidence base.

### Post-pandemic transplant activity

Following COVID-19, service review with cogent planning and updated resource allocation will be necessary to minimise the effects of lost activity on individual patients. While patients are not currently being added to the transplant waiting list, the rate of patients developing ESRD will be the same or even higher as a result of COVID-19. Therefore, it is envisaged that a backlog of patients will be added post-pandemic, resulting in an acute rise in the waiting list size. This study has shown the transplant waiting list may increase by 720–1,438 active dialysis patients over 3–6 months. Dialysis significantly impacts health, with an increased risk for cardiovascular events and sepsis.[29] There is emerging evidence that kidney disease is associated with higher in-hospital mortality secondary to COVID-19.[30] It is postulated that the COVID-19 pandemic may disproportionally affect patients on dialysis, resulting in a reduced transplant waiting list. Under normal circumstances 22 patients die each month on the kidney-alone waiting list, and prospective studies determining the additional mortality among wait-listed patients will help characterise the impact more fully.

Halting kidney transplantation will affect living and deceased donation equally. However, deceased donor transplants will be missed opportunities due to the temporal nature of organ offers. Living donor kidney transplantation is a semi-elective procedure and may be recommenced when deemed safe. This study has shown that 255–510 living donor pairs will have been postponed by the end of 3–6 months. It will require significant resources and efforts to clear this backlog, including operating theatre capacity and anaesthetic and intensive care support. The wellbeing of potential living donors is paramount and evidence-based recommendations will be required for counselling and decision-making. Early post-transplant patients are regularly reviewed in the outpatient clinic. In the interest of minimising exposure and reducing the risk of contracting COVID-19, changes in ways of working will have to be introduced, with increased remote monitoring and local blood investigations. This envisaged increased and altered working will require rationalisation along with the several other services where activity is currently halted. Healthcare workers normally involved in transplantation are being redeployed to work in emergency departments or intensive care units, which has further contributed to cessation in transplant activity. To cope with the significant increase in workload, this trend may need to reversed with additional infrastructural support post-pandemic.

### Impact on workforce and resource allocation

UK kidney transplant services function under considerable pressure during ‘normal’ circumstances, with periodic surges in activity. The service is delivered by highly specialised multi-disciplinary staff including surgeons, nephrologists, specialist nurses and coordinators. They provide a round-the-clock national organ retrieval and implantation service of organs from deceased donors along with elective living donation. While delivering transplant care is highly rewarding, the frequency of on-calls and high volume of out-of-hours work makes recruitment and retention challenging. This study has quantified the considerable rise in workload for transplant staff post-pandemic. Reverting back to ‘normal’ work practices is unlikely to be a safe and effective strategy for delivering the service. This study further shows a potential unaccounted rise in the UK dialysis population, which will require additional service provision and resources. The projected increase in demand on dialysis will further stretch a recovering NHS. Per patient costs for dialysis are over £24,000 per year, and a potential increase in the dialysis population modelled in this study would incur an additional cost of £13.4 to £31.8 million.[31] This is caveated by the aforementioned, as yet unknown, additional mortality risk from COVID-19 for patients on the transplant waiting list.

### Limitations of this study

Due to the novelty and unprecedented impact of the of the COVID-19 pandemic, there are no other studies to act as comparators in terms of health care provision modelling. Though the methodology for modelling this is reproducible, it was dependent on the availability of national data, which may not be the case globally. We have based our modelling on the assumption that activity during the pandemic would have been similar to the last financial year. Though activity has been near-identical between the last two years, the change in the organ donation law coming in to place in May 2020 may have positively impacted on deceased donor transplantation.[2,5] The missed transplant opportunities predicted in this study may therefore underestimate the projected shortfall in activity. On an international scale, however, these databases are among the largest and most accurate sources for transplant-related population level data, which would add to the validity of our findings. In 2018, Lowe et al showed a seasonal variation in kidney transplant activity by analysing 10 years of NHSBT data.[32] We have not included this variation in our modelling at present. The impact of loss of activity, however, is nonetheless expected to be significant, especially if the UK kidney transplant programme is halted entirely. Additional mortality on the waiting list and the number of patients becoming too unwell to remain on the list due to COVID-19 is unknown. This would influence our overall modelling and impact the size of the waiting list as well as dialysis capacity further. We chose the 5 March 2020 as a defining point of a change in national, regional and local strategy to shift priorities in NHS services towards preventing and treating COVID-19. Retrospectively, it may be argued that a different ‘day 0’ represents this change more accurately; however, given the limitations and unpredictability of the situation this was considered a fair representation.

## Conclusion

Kidney transplantation provides a life-enhancing procedure for patients with ESRD and is a flagship service of the NHS. While

the unequivocal focus is battling COVID-19 and maintaining vital frontline services, kidney transplantation will be adversely affected. This will result in ancillary pressures on national dialysis capacity and the transplant waiting list. We have quantified this impact and provided evidence for health providers to recognise this upcoming change in landscape. It is critical that kidney transplant services develop innovative solutions and expedite research to deliver transplantation as soon as possible. ■

## Supplementary material

Additional supplementary material may be found in the online version of this article at www.rcpjournals.org/clinmedicine:

S1 – Formulae, coefficients and variables used in our modelling.

## Acknowledgements

The data reported here have been supplied by the UK Renal Registry of The Renal Association. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the UKRR or The Renal Association.

## References

1 Schnuelle P, Lorenz D, Trede M, Van Der Woude FJ. Impact of renal cadaveric transplantation on survival in end-stage renal failure: evidence for reduced mortality risk compared with hemodialysis during long-term follow-up. *J Am Soc Nephrol* 1998;9:2135–41.
2 NHS Blood and Transplant. *Organ donation and transplantation: activity report 2018/19*. NHSBT, 2020. www.organdonation.nhs.uk/helping-you-to-decide/about-organ-donation/statistics-about-organ-donation/transplant-activity-report/.
3 Oniscu GC, Brown H, Forsythe JLR. Impact of cadaveric renal transplantation on survival in patients listed for transplantation. *J Am Soc Nephrol* 2005;16:1859–65.
4 Meier-Kriesche HU, Kaplan B. Waiting time on dialysis as the strongest modifiable risk factor for renal transplant outcomes: a paired donor kidney analysis. *Transplantation* 2002;74:1377–81.
5 Department of Health Social Care. *Opt-out organ donation: Max and Keira's Bill passed into law*. DHSC, 2019. www.gov.uk/government/news/opt-out-organ-donation-max-and-keira-s-bill-passed-into-law.
6 Marsh S. Coronavirus: first UK death confirmed as cases surge to 116. *The Guardian*, 6 March 2020.
7 Sohrabi C, Alsafi Z, O'Neill N *et al*. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 2020;76:71–6.
8 Zhou F, Yu T, Du R *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
9 Yang X, Yu Y, Xu J *et al*. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020;8:475–81.
10 NHS Blood and Transplant. *COVID-19 Bulletin Number 5*. NHSBT, 2020. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/18190/covid-19-bulletin-5-02042020.pdf.
11 Kotton CN, Fishman JA. Viral infection in the renal transplant recipient. *J Am Soc Nephrol* 2005;16:1758–74.
12 Lei S, Jiang F, Su W *et al*. Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection. *EClinicalMedicine* 2020;21:100331.
13 Aminian A, Safari S, Razeghian-Jahromi A, Ghorbani M, Delaney CP. COVID-19 outbreak and surgical practice: unexpected fatality in perioperative period. *Ann Surg* 2020, in press (doi: 10.1097/SLA.0000000000003925).
14 Li Y-K, Peng S, Li L-Q *et al*. Clinical and transmission characteristics of Covid-19 – a retrospective study of 25 cases from a single thoracic surgery department. *Curr Med Sci* 2020;40:295–300.
15 Kumar D, Manuel O, Natori Y *et al*. COVID-19: A global transplant perspective on successfully navigating a pandemic. *Am J Transplant* 2020, in press (doi: 10.1111/ajt.15876).
16 NHS Blood and Transplant. *Annual report on kidney transplantation: report for 2018/2019 (1 April 2009 – 31 March 2019)*. NHSBT, 2019.
17 UK Renal Registry. *UK Renal Registry 21st Annual Report – data to 31/12/2017*. Renal Association, 2019. www.renalreg.org/reports/data_to_end_2017/.
18 Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. *N Engl J Med* 2020;382:1679–81.
19 Browning RB, McGillicuddy JW, Treiber FA, Taber DJ. Kidney transplant recipients' attitudes about using mobile health technology for managing and monitoring medication therapy. *J Am Pharm Assoc* 2016;56:450–4.
20 Zanetti-Yabur A, Rizzo A, Hayde N *et al*. Exploring the usage of a mobile phone application in transplanted patients to encourage medication compliance and education. *Am J Surg* 2017;214:743–7.
21 Laupacis A, Keown P, Pus N *et al*. A study of the quality of life and cost-utility of renal transplantation. *Kidney Int* 1996;50:235–42.
22 Mehta P, McAuley DF, Brown M *et al*. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* 2020;395:1033–4.
23 Alberici F, Delbarba E, Manenti C *et al*. A single center observational study of the clinical characteristics and short-term outcome of 20 kidney transplant patients admitted for SARS-CoV2 pneumonia. *Kidney Int* 2020;pii:S0085-2538(20)30365-3.
24 Banerjee D, Popoola J, Shah S et al. COVID-19 infection in kidney transplant recipients. *Kidney Int* 2020, in press (doi: 10.1016/j.kint.2020.03.018).
25 Diao B, Feng Z, Wang C *et al*. Human kidney is a target for novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. *medRxiv* 2020.03.04.20031120 (doi: 10.1101/2020.03.04.20031120).
26 Ahn C, Amer H, Anglicheau D *et al*. Global Transplantation COVID Report, March 2020. *Transplantation* 2020, in press (doi: 10.1097/TP.0000000000003258).
27 NHS Blood and Transplant. *Covid-19 donation – a summary of international correspondence*. NHSBT, 2020. https://nhsbtdbe.blob.core.windows.net/umbraco-assets-corp/18016/donation-international-reflections-17032020.pdf.
28 American Society of Transplantation. *COVID-19 (Coronavirus): FAQs for organ transplantation*. AST, 2020.
29 Bray BD, Boyd J, Daly C *et al*. How safe is renal replacement therapy? A national study of mortality and adverse events contributing to the death of renal replacement therapy recipients. *Nephrol Dial Transplant* 2014;29:681–7.
30 Cheng Y, Luo R, Wang K *et al*. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020;97:829–38.
31 Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. *Nephrol Dial Transplant* 2012;27(suppl_3):iii73–iii80.
32 Lowe M, Maidstone R, Poulton K *et al*. Monthly variance in UK renal transplantation activity: a national retrospective cohort study. *BMJ Open* 2019;9:e028786.

**Address for correspondence: Mr Videha Sharma, Department of Renal and Pancreatic Transplantation, Manchester University NHS Foundation Trust, Oxford Road, Manchester M13 9WL, UK.**
**Email: videha.sharma@mft.nhs.uk**

### 19.6.3 Human leukocyte antigen epitope mismatch loads and the development of de novo donor-specific antibodies in cardiothoracic organ transplantation

This paper was published in the International Journal of Immunogenetics on 14th December 2021.

Received: 22 August 2021 | Revised: 18 October 2021 | Accepted: 4 November 2021

DOI: 10.1111/iji.12563

INTERNATIONAL JOURNAL OF IMMUNOGENETICS WILEY

INVITED REVIEW

# Human leukocyte antigen epitope mismatch loads and the development of de novo donor-specific antibodies in cardiothoracic organ transplantation

Amy Bedford[1,2] | Steven Jervis[1] | Judith Worthington[1] | Marcus Lowe[1] | Kay Poulton[1]

[1] Manchester Royal Infirmary, Manchester University NHS Foundation Trust, Manchester, UK

[2] Faculty of Biology, Medicine and Health, Division of Medical Education, School of Medical Sciences, University of Manchester, Manchester, UK

**Correspondence**
Amy Bedford, Manchester Royal Infirmary, Manchester University NHS Foundation Trust, 2nd Floor, Purple Zone, Oxford Road, Manchester, M13 9WL, UK.
Email: Amy.Bedford@ouh.nhs.uk

## Abstract

De novo donor-specific human leucocyte antigen (HLA) antibodies (dnDSA) are associated with increased risk of rejection and mortality in solid organ transplantation. Such dnDSA is produced in some recipients upon allorecognition of mismatched HLA post-transplant. HLA matching is not currently considered in the allocation of deceased donor hearts and lungs and pre-transplant immunological risk stratification is based entirely on the mean fluorescence intensity (MFI) of circulating donor-directed HLA antibodies. HLA epitope-based matching tools predict B-cell or T-cell HLA epitopes that are present in the donor's HLA but absent in the recipient's HLA. We hypothesized that patients with higher epitope mismatch loads would be at increased risk of dnDSA development. We retrospectively analysed 73 heart and/or lung transplant recipients who were tested for DSA between 2015 and 2020. HLAMatchmaker, PIRCHE-II and HLA epitope mismatch algorithm (HLA-EMMA) were used to calculate eplet mismatch (EpMM) loads, T-cell epitope mismatch (TEpMM) loads and solvent accessible amino acid mismatch (SAMM) loads, respectively. Multivariate analyses showed that HLA-EMMA was the only tool with a significant association between the total score for all HLA loci and dnDSA production [odds ratio (OR) 1.021, 95% confidence interval (CI) 1.003–1.042, $p = .0225$] though this increased risk was marginal. The majority of dnDSA were directed against HLA-DQ and patients with higher HLA-DQ TEpMM loads (OR = 1.008, CI = 1.002–1.014, $p = .007$), and HLA-DR+DQ SAMM loads (OR = 1.035, CI = 1.010–1.064, $p = .0077$) were most at risk of producing dnDSA. We also showed that patients with a risk epitope within the HLA molecule encoded for by HLA-DQA1*05 + HLA-DQB1*02/03:01 were significantly more likely to produce dnDSA. The use of HLA epitope-based matching tools could be used for cardiothoracic transplant risk stratification to enable early intervention and monitoring of patients at increased risk of producing dnDSA.

KEYWORDS
antibodies, cardiothoracic, epitopes, HLA, immunology, transplantation

## 1 | INTRODUCTION

Cardiothoracic transplantation can offer life-saving treatment for patients with end-stage heart and lung disease. However, long-term survival is still relatively low with a median survival of 12.5 years for heart transplantation and median survival for lung or heart-lung transplantation is close to half that of heart transplantation (Chambers et al., 2019; Khush et al., 2019). Due to the short acceptable cold ischaemia time, limited donor pool and the clinical urgency of cardiothoracic (heart and/or lung) patients, human leukocyte antigen (HLA) matching is not currently considered in the allocation of hearts and lungs, and clinicians instead rely on high dose immunosuppression to reduce the risk of rejection (Ansari et al., 2014). However, long-term immunosuppression treatment significantly increases the risk of infection and malignancy which are common causes of mortality in cardiothoracic transplant recipients (Frohlich et al., 2012; Gallagher et al., 2010).

In the United Kingdom, the allocation of deceased donor hearts and lungs is primarily based on clinical urgency. Other factors such as donor and recipient size matching and ABO blood group compatibility are also considered. Immunological risk assessment at the time of transplant is based solely on the level of pre-transplant donor-directed HLA antibodies (Walton et al., 2016). The strength of HLA antibody binding is commonly determined by solid phase Luminex single antigen beads and is expressed as mean fluorescence intensity (MFI) (BTS, 2014). However, the MFI cut-offs are not strongly evidence based and were instead born out of necessity as the clinical significance of antibodies detected by the sensitive Luminex single antigen bead assay was unclear (Tait, 2016). This application of such cut-off values is controversial as Luminex single antigen bead MFI results are semi-quantitative at best and were not designed by the bead manufacturers to be quantitative, despite being treated as such (H. C. Sullivan et al., 2017).

It is widely accepted that donor-specific HLA antibodies (DSA) are strongly associated with rejection, allograft failure and increased mortality in heart and lung transplant recipients (Le Pavec et al., 2016; Lobo et al., 2013; Njue & Chih, 2019; Smith et al., 2011). Without consideration of HLA matching, current guidelines only assess the immediate post-transplant risk and do not assess the long-term risk of rejection associated with allorecognition of non-self HLA and production of de novo DSA (dnDSA). Mismatched HLA is recognized immunologically at the epitope level rather than recognition of the entire antigen. Some HLA proteins share epitopes, making some HLA proteins more similar than others at an epitope level (Claas & Heidt, 2017). This concept underpins in silico HLA epitope-based matching tools which use algorithms to predict immunologically recognisable B-cell or T-cell mismatched HLA epitopes. HLA epitopes are not equally immunogenic, but an increased number of epitope mismatches theoretically increases the likelihood of immunogenic epitopes being present (Hönger et al., 2020).

HLAMatchmaker is an algorithm that determines patches of antibody accessible amino acid residues at polymorphic positions (eplets) that are present on donor HLA but absent on recipient HLA which are therefore mismatched. Overall, this predicts the mismatch load of potentially immunogenic eplets that may be recognized by B cells (Duquesnoy, 2006). Similarly, the HLA epitope mismatch algorithm (HLA-EMMA) predicts B-cell epitopes at the amino acid level by calculating the number of solvent accessible amino acid mismatches at polymorphic positions based upon the molecule's crystal structure (Kramer et al., 2020).

An alternate approach is the Predicted Indirectly ReCognisable HLA Epitopes (PIRCHE) algorithm which predicts T-cell epitopes. The PIRCHE-II module calculates the number of mismatched donor HLA peptides that can be presented by HLA-class II to CD4+ T cells by indirect recognition, leading to subsequent B-cell activation. This pathway plays a key role in the production of dnDSA. As CD4+ T cells have an important role in both the activation of CD8+ cytotoxic T cells and B cells, the calculated PIRCHE-II mismatch load may have a role in predicting both antibody-mediated and cell-mediated rejection (Geneugelijk et al., 2018).

Consensus guidelines do not currently recommend regular DSA monitoring for standard-risk patients who have no history of donor-directed antibody pre-transplant as it is not deemed cost-effective, and post-transplant DSA testing is only carried out when clinically indicated. Regular HLA antibody screening for the first year post-transplant is only recommended for cardiothoracic transplant patients above standard risk (BTS, 2014; Tait et al., 2013). HLA epitope-based matching tools could be used to identify patients at increased risk of dnDSA production who would be suitable for routine DSA monitoring to enable early intervention for antibody-mediated rejection.

The relationship between HLA epitope mismatch loads and poorer survival outcomes including increased dnDSA development is well-supported in renal transplantation (Daniëls et al., 2018; Geneugelijk et al., 2018; Lachmann et al., 2017). However, this is still at a population level, and individualized risk assessment based on the HLA epitope mismatch load is still in its early days. It has been shown that patients with increased HLA-DR/DQ eplet mismatch (EpMM) loads are at increased risk of dnDSA development, antibody-mediated rejection, T cell-mediated rejection and graft loss in renal transplant recipients (Wiebe et al., 2019). Another study found that renal transplant patients with higher HLA-DR/DQ EpMM loads were less able to tolerate low tacrolimus levels (Wiebe et al., 2017). This demonstrates that epitope matching tools could be used to facilitate personalized post-transplant treatment and monitoring.

In heart and lung transplantation, the relationship between the epitope mismatch load and poorer post-transplant outcomes is less clear as far fewer studies have been carried out and some reports are contradictory. In this study, we aimed to determine if the risk of dnDSA development can be determined based on HLA epitope mismatch loads. In the future, this could be used to inform post-transplant risk stratification for heart and lung transplantation. We hypothesize that higher HLA epitope mismatch loads will be associated with increased production of dnDSA.

## 2 | MATERIALS AND METHODS

### 2.1 | Patient population

This retrospective study included all patients who received a heart and/or lung transplant at Wythenshawe Hospital between 2015 and 2020 and were tested for donor-specific HLA antibodies post-transplant (N = 79). Patients with donor-directed antibody that was detected pre-transplant were excluded as this antibody is not de novo (N = 3). Patients were also excluded if follow-up time was under 2 weeks as this period may be insufficient for dnDSA production (N = 3). Follow-up time was defined as the time from transplant to the latest serum sample received. As serum samples are not routinely sent to the laboratory from post-transplant cardiothoracic patients, time since transplant was also included. Overall, a total of 73 donor-recipient pairs were included in this study.

In line with the transplant centre's standard practice, patients are expected to have received a standard immunosuppression protocol (ciclosporin, azathioprine for lung transplant patients or mycophenolate mofetil for heart transplant patients and Prednisolone) with rabbit anti-thymocyte globulin and Methylprednisolone induction therapy. Adaptations to this standard protocol are made in high-risk transplants and in response to evidence of rejection. Such adaptations can include the use of tacrolimus. Due to the coronavirus disease 2019 pandemic, ethical review was not possible for this study. As a result, clinical data could not be accessed, so this study only included retrospective analysis of anonymized in-house data. Hence, individualized immunosuppression treatment protocols and clinical outcomes such as rejection and mortality could not be included in the analyses.

The primary endpoint of this study was the production of dnDSA as detected by Luminex single antigen beads.

### 2.2 | HLA antibody testing and HLA typing

HLA antibody testing of ethylenediaminetetraacetic acid-treated serum samples was performed using the Labscreen® Single-antigen Bead assay (OneLambda, CA, USA). A modified protocol was followed, as described by Liwski et al. (2017). An MFI of 2000 was used as the cut-off for positivity for newly defined HLA antibodies. This was not treated as a clear cut-off as factors such as known sensitization, control reactions, reactions to self-antigens, previous results, previous sensitization and the reaction pattern were considered in analysis of HLA antibody testing results. This analysis strategy is supported by consistent satisfactory performance in external quality assurance schemes.

Variable resolution of HLA typing data was available for donors and recipients, depending on the techniques routinely used when the patient was registered on the national transplant list and when the transplant took place (Table 1). High-resolution HLA types were obtained by TruSight HLA v2 Next Generation Sequencing (Illumina, CA, USA and CareDX, PA, USA) following the assay manufacturer's protocol. Intermediate resolution HLA types were obtained by reverse polymerase chain reaction (PCR) and sequence-specific oligonucleotides (SSO) (OneLambda). A modified protocol was used for SSO HLA typing where all volumes used were 50% of those in the manufacturer's instructions. Some donor types were obtained by LinkSeq real-time PCR (OneLambda). The test manufacturer's protocol was followed. In all cases, the highest resolution type available was used. Even with high-resolution techniques, some ambiguities do exist. Where only low to intermediate resolution HLA typing data was available, HaploStats (http://www.haplostats.org) was used to predict of four-digit resolution HLA types.

### 2.3 | HLA epitope-based matching tools

HLAMatchmaker (http://www.epitopes.net) was used to calculate the number of mismatched antibody accessible epitopes containing polymorphic amino acid residues on the HLA molecule surface, designated as eplets, between each donor and recipient pair. A mismatched eplet is classed as an eplet present in the donor's HLA but not the recipient's HLA. The HLA-ABC eplet matching programme v4.0 was used to determine the EpMM load at HLA-A, B and C for each donor-recipient pair. The HLA-DRDQDP eplet matching programme v3.1 was used to calculate to EpMM load at HLA-DRB345, DRB1, DQA1, DQB1, DPA1 and DPB1 for each donor-recipient pair. Both antibody-verified and non-antibody-verified eplets were included in the analysis.

The PIRCHE-II algorithm (https://www.pirche.com) was used for prediction of the number of mismatched processed donor HLA peptides that could be presented by recipient HLA class II molecules to CD4+ T cells for each donor-recipient pair. PIRCHE-II scores or T cell epitope mismatch (TEpMM) loads were calculated for each locus as well as a total score that reflects the contribution of all HLA loci.

HLA-EMMA v1.04 (https://hla-emma.com) was used to determine the number of solvent accessible polymorphic amino acid mismatches on the surface of HLA molecules, according to their crystal structure. The solvent accessible amino acid mismatch (SAMM) load was calculated as a total for all HLA loci as well as for each HLA locus. For HLA class II, the SAMM load was calculated for the entire molecule including both the alpha and beta chains. The default settings were used.

Novel HLA alleles could not be recognized by the algorithms used (N = 1). In this case, the field for the novel allele was left blank.

### 2.4 | Statistical analysis

Patient population statistics were summarized as mean and SD for parametrically distributed continuous variables or median and interquartile range (IQR) in the case of non-parametric distribution. The D'Agostino and Pearson test was used to test numerical data for normality. To investigate differences between the dnDSA negative and dnDSA groups, Fisher's exact test was used for categorical variables. For continuous variables, unpaired *t*-tests were used for normally distributed data, and the Mann–Whitney *U* test was used for skewed data. Multiple logistic regression models were used to investigate the

**TABLE 1** Patient characteristics ($n = 73$)

| Population characteristics | DSA negative (N = 50) | DSA positive (N = 23) | *p*-Value |
|---|---|---|---|
| Median recipient age, years (IQR) | 53 (42.25–61.00) | 54 (41.00–59.00) | .845 |
| Recipient sex, *n* | | | |
| Male | 29 (58.0%) | 11 (47.8%) | .457 |
| Female | 21 (42.0%) | 12 (52.2%) | |
| Donor sex | | | |
| Male, *n* | 27 (54.0%) | 8 (34.8%) | .141 |
| Female, *n* | 23 (46.0%) | 15 (65.2%) | |
| Median donor age, years (IQR) | 40 (24.75–52.25) | 49 (25.00–57.00) | .152 |
| Organ transplanted | | | |
| Heart, *n* | 15 (30.0%) | 1 (4.3%) | |
| Double lung, *n* | 30 (60.0%) | 18 (78.2%) | – |
| Single lung, *n* | 5 (10.0%) | 2 (8.7%) | |
| Heart lung, *n* | 0 (0.0%) | 2 (8.7%) | |
| Mean time since Tpx, days (SD) | 1038 (416.4) | 848.1 (343.2) | .061 |
| Median follow-up time, days (IQR) | 423.5 (121.8–993.0) | 475 (298.0–638.0) | .452 |
| Follow-up time, *n* | | | |
| 0–3 months | 8 (16%) | 0 (0%) | |
| 3–6 months | 6 (12%) | 3 (13%) | |
| 6–12 months | 8 (16%) | 6 (26%) | – |
| 1–2 years | 10 (20%) | 11 (48%) | |
| 2–6 years | 18 (36%) | 3 (13%) | |
| Time from Tpx to first dnDSA detected, *n* | | | |
| 0–3 months | | 13 (57%) | |
| 3–6 months | | 2 (9%) | |
| 6–12 months | N/A | 4 (17%) | – |
| 1–2 years | | 3 (13%) | |
| 2–6 years | | 1 (4%) | |
| Recipient HLA typing method | | | |
| NGS, *n* | 21 (42.0%) | 11 (47.8%) | – |
| PCR-SSO, *n* | 29 (58.0%) | 12 (52.2%) | |
| Donor HLA typing method | | | |
| NGS, *n* | 13 (26.0%) | 10 (43.5%) | – |
| PCR-SSO, *n* | 31 (62.0%) | 10 (43.5%) | |
| RT-PCR, *n* | 6 (12.0%) | 3 (13.0%) | |
| Pre-transplant sensitization status | | | |
| HLA antibody negative | 36 (72%) | 12 (52%) | .117 |
| HLA antibody positive | 14 (28%) | 11 (48%) | |
| Median total HLA mismatches at | | | |
| HLA-A, B, C, DRB1, DQB1, DPB1, *n* (IQR) | 8 (6–9) | 8 (7–9) | .348 |
| Median HLAMatchmaker total score, (IQR) | 38 (30.00–51.50) | 45 (32.25–62.00) | .066 |
| Median PIRCHE-II total score, (IQR) | 399 (296.5–566.5) | 498.5 (381.3–713.0) | .021* |
| Median HLA-EMMA total score, (IQR) | 45 (32.25–62.00) | 77.0 (47.50–111.80) | .023* |

*Note*: Data is summarized as median (interquartile range, IQR), mean (standard deviation, SD) or *n* (%). *p*-Values < .05 are indicated by an asterisk.
Abbreviations: dnDSA, de novo donor-specific antibody; HLA-EMMA, human leucocyte antigen epitope mismatch algorithm; NGS, next generation sequencing; RT-PCR, real-time polymerase chain reaction; SSO, sequence-specific oligonucleotides; Tpx, transplant.

relationship between HLAMatchmaker, PIRCHE-II and HLA-EMMA scores or 'mismatch loads' and the presence or absence of dnDSA post-transplant. Initially, logistic regression models for the total score for all HLA were performed as well as models for HLA class I (HLA-A, B, C) and HLA class II (HLA-DR, DQ, DP) for each algorithm. If significant findings were observed, further exploration of individual loci was carried out. Non-significant data were not explored further to reduce the likelihood of incidental findings. Time since transplant and recipient sex were adjusted for in the analyses, and odds ratios (OR) and 95% confidence intervals (CI) were calculated. A $p$-value < .05 was considered to be statistically significant. The Hosmer–Lemeshow test was used to assess goodness of fit of the logistic regression models. A $p$-value < .05 describes poor fit. GraphPad Prism 8.4 was used for statistical analysis (GraphPad Software, CA, USA).

## 3 | RESULTS

### 3.1 | Patient population

Of 79 patients, three patients had donor-directed HLA antibodies that were identified pre-transplant or in the day of transplant serum sample. As these were not de novo, these patients were excluded from the study. Three further patients were excluded as the only serum samples tested for HLA antibodies were taken <14 days post-transplant. As the production of de novo antibodies takes approximately 10–14 days, these patients could not be classified as having been tested for dnDSA (Murphy et al., 2012). A total of 73 donor-recipient pairs were included in this retrospective study, 50 of whom were DSA negative and 23 were DSA positive (Table 1). From this point, all DSA were considered to be de novo. The majority of patients were double lung transplant recipients ($n$ = 48), followed by heart recipients ($n$ = 16), single lung ($n$ = 7) and heart lung ($n$ = 2) recipients. Four-digit high-resolution HLA typing was available for 43.8% of all recipients and 31.5% of all donors. The median score calculated by each HLA epitope-based matching tool was greater in the DSA positive group than the DSA negative group though overlap of the interquartile ranges was observed in each case. No statistically significant difference was observed between the dnDSA negative and dnDSA positive groups for donor and recipient age and sex, time since transplant or follow-up time. There was also no significant difference in the total number of HLA antigen mismatches or the HLA-Matchmaker total score between both groups (Table 1). However, a significant difference was observed between the PIRCHE-II total scores ($p$ = .021) and the HLA-EMMA total scores ($p$ = .023).

### 3.2 | De novo DSA production

Of the 73 recipients included in this study, 23 developed dnDSA (31.5%). A total of 62 dnDSA were defined by Luminex single antigen beads, the largest proportion of which were directed against HLA-DQB1 antigens (38.7%) (Figure 1). Of the dnDSA positive patients, 21 had dnDSA directed against HLA-DQA1 and/or DQB1 antigens (91.3%) with only two patients having no HLA-DQ dnDSA. In both of these patients, their dnDSA was directed against HLA-B antigens. Seven patients developed only one dnDSA (30.4%), and the remaining 16 had more than one dnDSA (69.6%). HLA-DR was the second most common DSA following HLA-DQ.

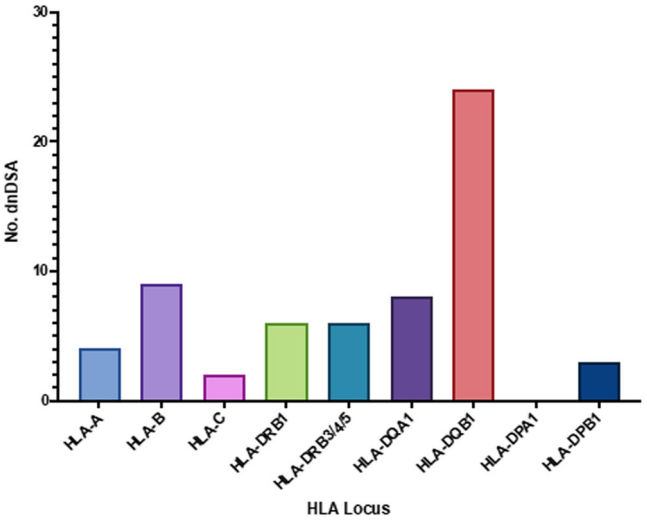



**FIGURE 1** The greatest proportion of de novo donor-specific antibody (dnDSA) is directed against human leukocyte antigen (HLA)-DQB1. Shown are the total number of dnDSA directed against each HLA locus from all DSA positive patients. Of 23 DSA positive patients, 82.6% had at least one HLA-DQB1 dnDSA

### 3.3 | The association of dnDSA development and HLA epitope mismatch load

In an adjusted logistic regression for recipient sex and time since transplant, no significant association was observed between the number of HLA antigen mismatches and dnDSA development (Figure 2a). HLA-matchmaker EpMM was not significantly associated with dnDSA development when all HLA loci were included in the analysis (OR = 1.008, C = 0.9879–1.030, $p$ = .4330) or when HLA class I (OR = 0.9953, CI = 0.9506–1.026, $p$ = .7718) or HLA class II loci (OR = 1.020, CI = 0.9917–1.058, $p$ = .2063) were investigated alone (Figure 2b).

The total PIRCHE-II score for all HLA loci was not significantly associated with dnDSA production (OR = 1.002, CI = 0.9998–1.005, $p$ = .0748). However, when HLA class I and class II loci were considered separately, HLA class II TEpMM load was significantly associated with dnDSA development (OR = 1.004, CI = 1.001–1.008, $p$ = .0271) (Figure 2c).

HLA-EMMA was the only tool that reached significance for the association of SAMM load of all HLA loci with dnDSA development (OR = 1.021, CI = 1.003–1.042, $p$ = .0255). When the SAMM loads for HLA class I and class II loci were analysed separately, only HLA class II showed statistical significance (OR = 1.031, CI = 1.008–1.058, $p$ = .0123) (Figure 2d).

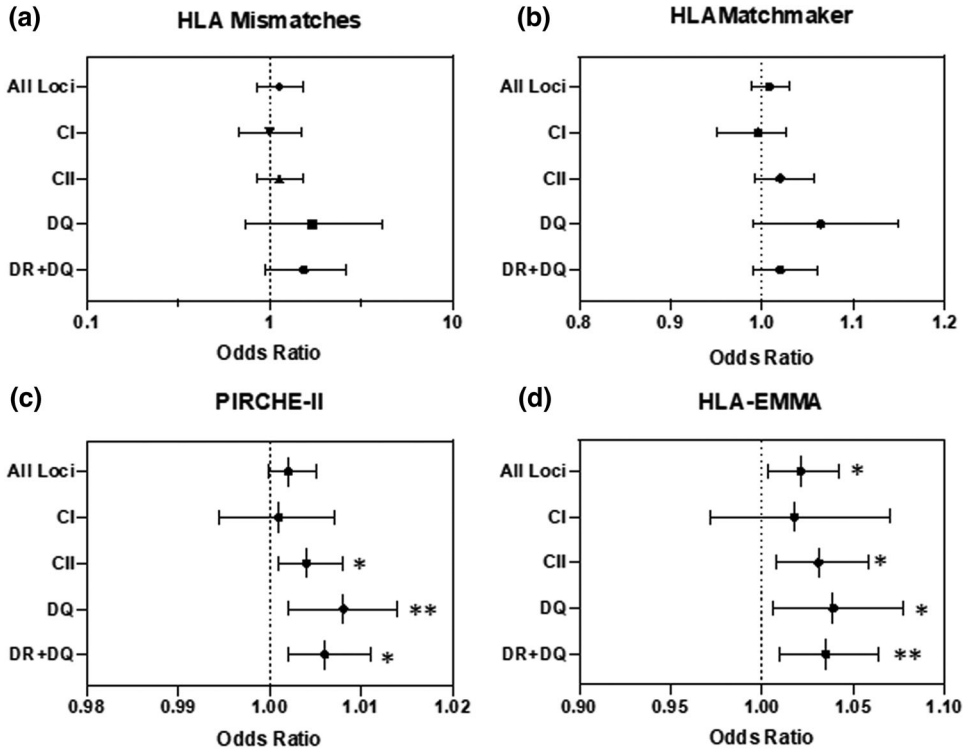



**FIGURE 2** Odds ratios (OR) of de novo donor-specific antibody (dnDSA) development according to the number of human leukocyte antigen (HLA) antigen mismatches (a) and HLA epitope/eplet mismatch loads calculated by HLAMatchmaker (b), PIRCHE-II (c) or HLA epitope mismatch algorithm (HLA-EMMA) (d). Points on the graph represent adjusted OR determined by multiple logistic regression analyses, adjusted for recipient sex and time since transplant. The axis range is smaller for PIRCHE-II as the scores generated by this tool are much greater values than HLAMatchmaker and HLA-EMMA. As a result, the effect of each individual point on dnDSA development is smaller. Similarly, the axis range is much greater for HLA antigen mismatches as the scores are much smaller. Thus, the effect of each individual point is greater. Bars represent 95% confidence intervals. Statistical significance is indicated by an asterisk (* = $p < .05$, ** = $p < .01$)

As only HLA class II loci reached clinical significance using PIRCHE-II and HLA-EMMA, individual class II loci were examined further. The greatest proportion of dnDSA in this cohort were directed against HLA-DQ, so the association between PIRCHE-II and HLA-EMMA HLA-DQ (HLA-DQA1 and HLA-DQB1) mismatch scores with dnDSA was investigated. The PIRCHE-II TEpMM load for HLA-DQ was more strongly associated with dnDSA development (OR = 1.008, CI = 1.002–1.014, $p$ = .007) (HL = 6.608, $p$ = .579) than when all class II loci were considered. With an OR of 1.008, the odds of dnDSA development increases 0.8% per HLA-DQ TEpMM. Though this is a small increase per point, the HLA-DQ TEpMM load in this cohort ranged from 0 to 393. Therefore, a patient with an HLA-DQ TEpMM load of 300 would be 160% more likely to produce dnDSA than a patient with an HLA-DQ TEpMM load of 100. An increased risk of dnDSA development was also observed for higher HLA-DQ amino SAMM, calculated by HLA-EMMA (OR = 1.039, CI = 1.006–1.077, $p$ = .0237).

As the second most common HLA class II locus for dnDSA was HLA-DR, the cumulative effect of HLA-DR and HLA-DQ epitopes was examined. The odds of dnDSA development were significantly associated with the TEpMM load (OR = 1.006, CI = 1.002–1.011, $p$ = .0114). A greater level of significance was reached when considering HLA-DR+DQ SAMM load (OR = 1.035, CI = 1.010–1.064, $p$ = .0077) (HL = 6.258, $p$ = .618) than HLA-DQ SAMM load alone. This OR reflects a 3.5% increase in the odds of dnDSA development per HLA-DR+DQ SAMM.

## 3.4 | HLA epitope mismatch loads and dnDSA development in lung transplantation

In this study, only one of 15 heart transplant recipients produced dnDSA. Due to this, no statistical tests could be performed to investigate the relationship between dnDSA development and HLA epitope mismatch loads in heart transplantation alone. The same analyses performed on the entire cohort were performed again with the lung transplant recipients only ($n$ = 55). A significant association between the PIRCHE-II TEpMM PIRCHE-II load for HLA-DQ (OR = 1.008, CI = 1.002–1.016, $p$ = .0197) and DR+DQ (OR = 1.0086, CI = 1.001–1.012, $p$ = .0353) was observed. No other significant associations with PIRCHE-II TEpMM loads were found. The HLA-EMMA SAMM load for DR+DQ was also significantly associated with dnDSA development (OR = 1.031, CI = 1.003–1.063, $p$ = .0358). As with the complete cardiothoracic cohort, no significant association was observed between individual HLA class II loci HLAMatchmaker EpMM loads or total EpMM load and dnDSA development.

### 3.5 | Risk epitope mismatches and dnDSA production

The most common dnDSA in this study was directed against HLA-DQ7 ($n = 7$). This was followed by HLA-DQA1*05 ($n = 6$) and HLA-DQ2 ($n = 5$). McCaughan et al. (2018) describe a high-risk structural epitope mismatch found in the DQ antigen coded for by DQA1*05 + DQB1*02 and DQA1*05 + DQB1*03:01 which conferred a significantly increased risk of dnDSA production in heart and lung transplant recipients (McCaughan et al., 2018). In our cohort, significantly more patients with DQA1*05 + DQB1*02/DQB1*03:01 mismatched transplants produced dnDSA than those without this mismatch (OR = 4.74, CI = 1.32–14.99, $p = .012$).

## 4 | DISCUSSION

In this retrospective single-centre study, we investigated if the HLA epitope mismatch loads calculated by HLAMatchmaker, PIRCHE-II and HLA-EMMA are associated with the production of dnDSA in cardiothoracic transplantation. In this study cohort, if dnDSA was produced, it was targeted primarily against HLA-DQ. This suggests that HLA-DQ may be more immunogenic than other HLA loci, and as a result HLA-DQ epitope mismatches were investigated further.

Despite the association between HLAMatchmaker EpMM load and poor transplant outcomes being well-supported in renal patients, our study found no significant association between EpMM load and dnDSA production. Studies investigating the relationship between EpMM load and outcomes such as graft loss and dnDSA development in heart and lung transplant patients have shown contradictory findings. Thus, HLA-Matchmaker may be less suitable to predict poor transplant outcomes in heart and lung transplant patients.

Previous studies using HLAMatchmaker, PIRCHE-II or HLA-EMMA have found that patients with greater HLA-DR and/or HLA-DQ epitope mismatch loads have increased risk of poor transplant outcomes including antibody-mediated rejection, dnDSA production, chronic lung allograft dysfunction, cardiac allograft vasculopathy and graft loss in heart and lung transplant patients (Nilsson et al., 2019; Osorio-Jaramillo et al., 2020; Walton et al., 2016; Wu et al., 2020; Zhang et al., 2020). However, published studies are contradictory, and some investigations have observed no relationship between HLA epitope mismatches and poor transplant outcomes (McCaughan et al., 2018; P. M. Sullivan et al., 2015). A major limitation of many published studies is the lack of high-resolution typing data available, and the requirement for imputation can result in inaccuracies (Engen et al., 2021).

PIRCHE-II and HLA-EMMA are much less studied in cardiothoracic transplantation than HLA matchmaker. We found that the total scores for both PIRCHE-II and HLA-EMMA were significantly different between dnDSA negative and dnDSA positive groups. This significant difference was not observed with the number of HLA mismatches at HLA-A, B and DR. This suggests that epitope-based matching by PIRCHE-II and HLA-EMMA may be a better tool to predict dnDSA development than traditional HLA matching. Furthermore, our results suggest that patients with higher HLA-DQ TEpMM and HLA-DQ+DR SAMM loads are most at risk of dnDSA production. HLA-EMMA may be the best overall tool to stratify the risk of dnDSA development in cardiothoracic transplant patients as this was the only algorithm in our study that demonstrated an association between epitope mismatch load and dnDSA development for all HLA loci though the greatest significance was observed with HLA-DR+DQ SAMM load.

In this study, only 6% of the heart transplant recipients produced dnDSA, whereas 36% of lung transplant recipients produced dnDSA and both heart lung transplant recipients produced dnDSA. This supports the idea that the immunogenicity of lung transplantation is greater than heart transplantation due to the constant immune challenge associated with environmental exposure through respiration (Jiang & Nicolls, 2014). As the immunological risk of both heart and lung transplantation is currently assessed in the same way, both heart and lung transplant recipients were analysed together. Additional analysis of lung transplant recipients alone found that patients with higher HLA-DQ TEpMM and HLA-DQ+DR SAMM loads are most at risk of dnDSA production. However, if HLA epitope-based immunological risk stratification were to be implemented in the future, it may be more appropriate to assess heart and lung transplant recipients differently in line with the differing immunogenicity of the transplants.

To our knowledge, only one study has investigated the effect of amino acid mismatches calculated by HLA-EMMA on heart transplant outcomes, and no studies are published for lung transplant patients. Osorio-Jaramillo et al. (2020) found inferior graft survival and increased risk of rejection in patients with a higher HLA-DR amino acid mismatch load. HLA-A+B amino acid mismatch load had no impact on graft survival but increased rejection risk. However, they only had low-resolution typing available at HLA-A, B and DR, so no other loci were investigated. Our study builds upon this using intermediate and high-resolution HLA typing data to support the idea that higher SAMM load can predict an increased risk of poor heart and lung transplant outcomes including dnDSA development.

The relationship between the number of HLA epitope mismatches and the risk of dnDSA development may be influenced by the fact that not all HLA epitope mismatches are equally immunogenic (Kramer et al., 2017). An alternative approach is to identify HLA epitopes that confer an increased risk of dnDSA development when mismatched. Our study supports the findings of McCaughan et al. (2018) and shows that the high-risk epitope encoded for by DQA1*05 + DQB1*02/DQB1*03:01 was associated with a 4.7-fold increase in the odds of dnDSA production in our cohort. As these HLA alleles are relatively common, it would not be suitable to veto all cardiothoracic transplants with this mismatch. However, these patients may benefit from increased immunosuppression to reduce the risk of dnDSA production.

This study has several limitations. The follow-up time is variable within the cohort, and patients with a shorter follow-up time may have produced dnDSA but have not been tested, resulting in their incorrect categorization in this study. Serum samples are often only sent for post-transplant cardiothoracic patients when clinically indicated such

as a possible rejection episode or an observed decline in graft function. Due to this, we decided to include only patients with a long follow-up time as this may skew the data towards patients with poorer outcomes. Another major limitation in this study is the lack of information of immunosuppression and clinical data as incidence of rejection, and mortality is unknown which may confound the results. In addition, 56% of recipients and 68% of donors did not have high-resolution typing available, so a high-resolution HLA type was imputed. Further studies are required using a larger cohort with high-resolution HLA typing available for all donor and recipient pairs, longer follow-up and more clinical information to confirm the findings of this study.

Overall, this study supports the idea that HLA epitope-based matching tools can be used to predict the risk of dnDSA development in cardiothoracic transplant recipients. If our findings are supported by further studies, classification of donor-recipient pairs with a high number of HLA-DQ TEpMM or HLA-DR+DQ SAMM may allow for identification of patients for increased post-transplant monitoring. Patients who receive transplants from donors with mismatched 'high-risk epitopes' may also be suitable for increased monitoring. Similarly, it may be possible to predict patients with low risk of dnDSA development who may benefit from reduced immunosuppression regimes. Thus, the risk of poor outcomes associated with immunosuppression such as malignancy and infection could be reduced in patients where this is suitable. This would not be required to be performed pre-transplant and instead could be calculated in the days following transplantation. As a result, time-critical factors such as cold-ischaemia time would not be affected. HLA epitope-based matching tools such as HLA-EMMA could be used to predict long-term risk and allow for a personalized therapeutic approach which could potentially improve overall survival of cardiothoracic transplant patients in the future.

## ACKNOWLEDGEMENTS

All authors are thankful to the Manchester Transplantation Laboratory for collecting and providing the data. We would also like to thank Professor Jamie Kirkham for his advice on statistical analysis.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ORCID

*Marcus Lowe* https://orcid.org/0000-0002-0799-1674
*Kay Poulton* https://orcid.org/0000-0002-5608-8107

## REFERENCES

Ansari, D., Bućin, D., & Nilsson, J. (2014). Human leukocyte antigen matching in heart transplantation: Systematic review and meta-analysis. *Transplant International*, *27*(8), 793–804. https://doi.org/10.1111/tri.12335

BTS. (2014). *The detection & characterisation of clinically relevant antibodies in allotransplantation*. https://bts.org.uk/guidelines-standards/

Chambers, D. C., Cherikh, W. S., Harhay, M. O., Hayes, D. Jr, Hsich, E., Khush, K. K., Meiser, B., Potena, L., Rossano, J. W., Toll, A. E., Singh, T. P., Sadavarte, A., Zuckermann, A., Stehlik, J., & International Society for Heart and Lung Transplantation. (2019). The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult lung and heart-lung transplantation Report-2019; Focus theme: Donor and recipient size match. *The Journal of Heart and Lung Transplantation: The Official Publication of the International Society for Heart Transplantation*, *38*(10), 1042–1055. https://doi.org/10.1016/j.healun.2019.08.001

Claas, F. H. J., & Heidt, S. (2017). Epitope-based HLA matching: A useful strategy with many shortcomings to overcome. *Transplantation*, *101*(8), 1744–1745. https://doi.org/10.1097/TP.0000000000001705

Daniëls, L., Naesens, M., Bosmans, J. L., Abramowicz, D., Nagler, E., Van Laecke, S., Peeters, P., Kuypers, D., & Emonds, M. P. (2018). The clinical significance of epitope mismatch load in kidney transplantation: A multicentre study. *Transplant Immunology*, *50*, 55–59. https://doi.org/10.1016/j.trim.2018.06.006

Duquesnoy, R. J. (2006). A structurally based approach to determine HLA compatibility at the humoral immune level. *Human Immunology*, *67*(11), 847–862. https://doi.org/10.1016/j.humimm.2006.08.001

Engen, R. M., Jedraszko, A. M., Conciatori, M. A., & Tambur, A. R. (2021). Substituting imputation of HLA antigens for high-resolution HLA typing: Evaluation of a multiethnic population and implications for clinical decision making in transplantation. *American Journal of Transplantation*, *21*(1), 344–352. https://doi.org/10.1111/ajt.16070

Fröhlich, G. M., Rufibach, K., Enseleit, F., Wolfrum, M., von Babo, M., Frank, M., Berli, R., Hermann, M., Holzmeister, J., Wilhelm, M., Falk, V., Noll, G., Lüscher, T. F., & Ruschitzka, F. (2012). Statins and the risk of cancer after heart transplantation. *Circulation*, *126*(4), 440–447. https://doi.org/10.1161/CIRCULATIONAHA.111.081059

Gallagher, M. P., Kelly, P. J., Jardine, M., Perkovic, V., Cass, A., Craig, J. C., Eris, J., & Webster, A. C. (2010). Long-term cancer risk of immunosuppressive regimens after kidney transplantation. *Journal of the American Society of Nephrology*, *21*(5), 852–858. https://doi.org/10.1681/ASN.2009101043

Geneugelijk, K., Niemann, M., Drylewicz, J., van Zuilen, A. D., Joosten, I., Allebes, W. A., van der Meer, A., Hilbrands, L. B., Baas, M. C., Hack, C. E., van Reekum, F. E., Verhaar, M. C., Kamburova, E. G., Bots, M. L., Seelen, M., Sanders, J. S., Hepkema, B. G., Lambeck, A. J., Bungener, L. B., … Spierings, E. (2018). PIRCHE-II is related to graft failure after kidney transplantation. *Frontiers in Immunology*, *9*, 321. https://doi.org/10.3389/fimmu.2018.00321

Hönger, G., Niemann, M., Schawalder, L., Jones, J., van Heck, M. R., van de Pasch, L. A. L., Vendelbosch, S., Rozemuller, E. H., Hösli, I., Blümel, S., & Schaub, S. (2020). Toward defining the immunogenicity of HLA epitopes: Impact of HLA class I eplets on antibody formation during pregnancy. *HLA*, *96*(5), 589–600. https://doi.org/10.1111/tan.14054

Jiang, X., & Nicolls, M. R. (2014). Working toward immune tolerance in lung transplantation. *The Journal of clinical investigation*, *124*(3), 967–970. https://doi.org/10.1172/JCI74701

Khush, K. K., Cherikh, W. S., Chambers, D. C., Harhay, M. O., Hayes, D. Jr, Hsich, E., Meiser, B., Potena, L., Robinson, A., Rossano, J. W., Sadavarte, A., Singh, T. P., Zuckermann, A., Stehlik, J., & International Society for Heart and Lung Transplantation (2019). The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult heart transplantation report - 2019; focus theme: Donor and recipient size match. *The Journal of Heart And Lung Transplantation: The Official Publication of the International Society for Heart Transplantation*, *38*(10), 1056–1066. https://doi.org/10.1016/j.healun.2019.08.004

Kramer, C. S. M., Roelen, D. L., Heidt, S., & Claas, F. (2017). Defining the immunogenicity and antigenicity of HLA epitopes is crucial for optimal epitope matching in clinical renal transplantation. *HLA*, *90*(1), 5–16. https://doi.org/10.1111/tan.13038

Kramer, C. S. M., Koster, J., Haasnoot, G. W., Roelen, D. L., Claas, F. H. J., & Heidt, S. (2020). HLA-EMMA: A user-friendly tool to analyse HLA class I and class II compatibility on the amino acid level. *HLA*, *96*(1), 43–51. https://doi.org/10.1111/tan.13883

Lachmann, N., Niemann, M., Reinke, P., Budde, K., Schmidt, D., Halleck, F., Pruß, A., Schönemann, C., Spierings, E., & Staeck, O. (2017). Donor-recipient matching based on predicted indirectly recognizable HLA

Epitopes independently predicts the incidence of de novo donor-specific HLA antibodies following renal transplantation. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *17*(12), 3076–3086. https://doi.org/10.1111/ajt.14393

Le Pavec, J., Suberbielle, C., Lamrani, L., Feuillet, S., Savale, L., Dorfmüller, P., Stephan, F., Mussot, S., Mercier, O., & Fadel, E. (2016). De-novo donor-specific anti-HLA antibodies 30 days after lung transplantation are associated with a worse outcome. *The Journal of Heart And Lung Transplantation: The Official Publication of the International Society for Heart Transplantation*, *35*(9), 1067–1077. https://doi.org/10.1016/j.healun.2016.05.020

Liwski, R. S., Greenshields, A. L., Murphey, C., Bray, R. A., & Gebel, H. M. (2017). It's about time: The development and validation of a rapid optimized single antigen bead (ROB) assay protocol for LABScreen. *Human Immunology*, *78*(7–8), 489–499. https://doi.org/10.1016/j.humimm.2017.05.001

Lobo, L. J., Aris, R. M., Schmitz, J., & Neuringer, I. P. (2013). Donor-specific antibodies are associated with antibody-mediated rejection, acute cellular rejection, bronchiolitis obliterans syndrome, and cystic fibrosis after lung transplantation. *The Journal of Heart and Lung Transplantation*, *32*(1), 70–77. https://doi.org/10.1016/j.healun.2012.10.007

McCaughan, J. A., Battle, R. K., Singh, S., Tikkanen, J. M., Moayedi, Y., Ross, H. J., Singer, L. G., Keshavjee, S., & Tinckam, K. J. (2018). Identification of risk epitope mismatches associated with de novo donor-specific HLA antibody development in cardiothoracic transplantation. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *18*(12), 2924–2933. https://doi.org/10.1111/ajt.14951

Murphy, K., & M. (2012). *Janeway's immunobiology* (8th ed.). Garland Science.

Nilsson, J., Ansari, D., Ohlsson, M., Höglund, P., Liedberg, A. S., Smith, J. G., Nugues, P., & Andersson, B. (2019). Human leukocyte antigen-based risk stratification in heart transplant recipients-implications for targeted surveillance. *Journal of the American Heart Association*, *8*(15), e011124. https://doi.org/10.1161/JAHA.118.011124

Njue, F., & Chih, S. (2019). When to intervene for donor-specific antibody after heart transplantation. *Current Opinion in Organ Transplantation*, *24*(3), 271–278. https://doi.org/10.1097/MOT.0000000000000634

Osorio-Jaramillo, E., Haasnoot, G. W., Kaider, A., Schaefer, A. K., Haberl, T., Goekler, J., Angleitner, P., Moayedifar, R., Zuckermann, A., Fischer, G. F., Laufer, G., Claas, F., & Aliabadi-Zuckermann, A. Z. (2020). Molecular-level HLA mismatch is associated with rejection and worsened graft survival in heart transplant recipients - a retrospective study. *Transplant International: Official Journal of the European Society for Organ Transplantation*, *33*(9), 1078–1088. https://doi.org/10.1111/tri.13657

Smith, J. D., Banner, N. R., Hamour, I. M., Ozawa, M., Goh, A., Robinson, D., Terasaki, P. I., & Rose, M. L. (2011). De novo donor HLA-specific antibodies after heart transplantation are an independent predictor of poor patient survival. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *11*(2), 312–319. https://doi.org/10.1111/j.1600-6143.2010.03383.x

Sullivan, H. C., Liwski, R. S., Bray, R. A., & Gebel, H. M. (2017). The road to HLA antibody evaluation: Do not rely on MFI. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *17*(6), 1455–1461. https://doi.org/10.1111/ajt.14229

Sullivan, P. M., Warner, P., Kemna, M. S., Albers, E. L., Law, S. P., Weiss, N. S., & Law, Y. M. (2015). HLA molecular epitope mismatching and long-term graft loss in pediatric heart transplant recipients. *The Journal of Heart and Lung Transplantation: The Official Publication of the International Society for Heart Transplantation*, *34*(7), 950–957. https://doi.org/10.1016/j.healun.2014.12.017

Tait, B. D. (2016). Detection of HLA antibodies in organ transplant recipients—Triumphs and challenges of the solid phase bead assay. *Frontiers in Immunology*, *7*, 570–570. https://doi.org/10.3389/fimmu.2016.00570

Tait, B. D., Süsal, C., Gebel, H. M., Nickerson, P. W., Zachary, A. A., Claas, F. H., Reed, E. F., Bray, R. A., Campbell, P., Chapman, J. R., Coates, P. T., Colvin, R. B., Cozzi, E., Doxiadis, I. I., Fuggle, S. V., Gill, J., Glotz, D., Lachmann, N., Mohanakumar, T., … Opelz, G. (2013). Consensus guidelines on the testing and clinical management issues associated with HLA and non-HLA antibodies in transplantation. *Transplantation*, *95*(1), 19–47. https://doi.org/10.1097/TP.0b013e31827a19cc https://doi.org/10.1097/TP.0b013e31827a19cc

Walton, D. C., Hiho, S. J., Cantwell, L. S., Diviney, M. B., Wright, S. T., Snell, G. I., Paraskeva, M. A., & Westall, G. P. (2016). HLA matching at the eplet level protects against chronic lung allograft dysfunction. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *16*(9), 2695–2703. https://doi.org/10.1111/ajt.13798

Wiebe, C., Kosmoliaptsis, V., Pochinco, D., Gibson, I. W., Ho, J., Birk, P. E., Goldberg, A., Karpinski, M., Shaw, J., Rush, N. D., & Nickerson, P. W. (2019). HLA-DR/DQ molecular mismatch: A prognostic biomarker for primary alloimmunity. *American Journal of Transplantation*, *19*(6), 1708–1719. https://doi.org/10.1111/ajt.15177

Wiebe, C., Rush, D. N., Nevins, T. E., Birk, P. E., Blydt-Hansen, T., Gibson, I. W., Goldberg, A., Ho, J., Karpinski, M., Pochinco, D., Sharma, A., Storsley, L., Matas, A. J., & Nickerson, P. W. (2017). Class II eplet mismatch modulates tacrolimus trough levels required to prevent donor-specific antibody development. *Journal of the American Society of Nephrology*, *28*(11), 3353. https://doi.org/10.1681/ASN.2017030287

Wu, R., Weston, M., Lunz, J. G., & Lopez-Cepero, M. (2020). A mismatch made in medicine: potential implications of HLA DQ epitope mismatching following heart transplantation. *The Journal of Heart and Lung Transplantation*, *39*(4), S167. https://doi.org/10.1016/j.healun.2020.01.1125

Zhang, X., Kransdorf, E., Levine, R., Patel, J. K., & Kobashigawa, J. A. (2020). HLA-DQ mismatches stimulate de novo donor specific antibodies in heart transplant recipients. *Human Immunology*, *81*(7), 330–336. https://doi.org/10.1016/j.humimm.2020.04.003

**How to cite this article:** Bedford, A., Jervis, S., Worthington, J., Lowe, M., & Poulton, K. (2022). Human leukocyte antigen epitope mismatch loads and the development of de novo donor-specific antibodies in cardiothoracic organ transplantation. *International Journal of Immunogenetics*, *49*, 30–38. https://doi.org/10.1111/iji.12563