Implementing computational techniques to investigate immune-mediated adverse drug reactions

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Gurpreet Singh Ghattaoraya

January 2018

Declaration

I hereby declare that this thesis is the result of my own work and has not been published for any other degree.

……

Gurpreet Singh Ghattaoraya

Acknowledgements

I would like to thank Dr Ana Alfirevic, Professor Andy Jones and Professor Rumona Dickson for supervising the work that went into my PhD thesis. I would also like to convey my appreciation for the guidance I have received and time invested by them. They have been exceptionally patient with me, extremely encouraging and not afraid to provide a telling off when needed. I would also like to extend my gratitude to Dr Lorna Halley for her guidance and mentorship.

I must also express my thanks to my colleagues at the Wolfson Centre for Personalised Medicine and the post-genomic bioinformatics group. Dr Eduardo Santos was instrumental in organising the data extraction team for the literature review carried out in this thesis and conducting some of the statistical analysis. I like to particularly thank Dr Philippe Marlot, Dr Karen Tse, Alex Penrose and Samantha Jones for their friendship during my time in Liverpool and hope that it will extend beyond the project.

Finally, I like to give my biggest thanks to my mother and brother for their continued support throughout my academic studies and being a much-needed pillar to lean against during some tough times.

Table of contents

Table of abbreviations

Abbreviation	Definition
ADR	Adverse drug reaction
AFND	Allele Frequency Net Database
AIA	Aspirin-intolerant asthma
AUI	Aspirin-induced urticaria
API	Application program interface
ASP	Active server pages
cADR	Cutaneous adverse drug reactions
CDC	Complement-dependent cytotoxicity
CDS	Clinical decision support
CPIC	Clinical Pharmacogenetics Implementation Consortium
CSS	Cascading style sheets
CSV	Comma separated value
DILI	Drug induced liver injury
DIHS	Drug-induced hypersensitivity syndrome
DISI	Drug-induced skin reaction
DRESS	Drug rash with eosinophilia and systemic symptoms
EMA	European Medicines Agency
EMBASE	Excerpta Medica database
EMBL	European molecular biology laboratory
FDA	(United States) Food and Drug Administration
GUI	Graphical user interface
HC-SC	Health Canada / Santé Canada
HLA	Human leukocyte antigen
HLA-ADR	The HLA adverse drug reaction database (part of AFND)

HSS	Hypersensitivity syndrome
HTML	HyperText Markup Language
KIR	Killer-cell immunoglobulin-like receptors
MeSH	Medical Subject Headings
MHC	Major histocompatibility complex
MPE	Maculopapular exanthema
MSA	Multiple sequence alignment
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NPV	Negative predictive value
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PCR	Polymerase chain reaction
PHP	PHP: Hypertext Preprocessor (recursive acronym)
PMDA	(Japanese) Pharmaceuticals and Medical Devices Agency
PMID	PubMed Identifier
PPV	Positive predictive value
RDBMS	Relational database management systems
RFLP	Restriction length fragment polymorphism
SBT	Sequence based typing
SJS	Stevens–Johnson syndrome
SNP	Single nucleotide polymorphism
SQL	Structured query language
SSO	Sequence specific oligonucleotide
SSP	Sequence-specific primer
TEN	Toxic epidermal necrolysis

List of Tables

List of Figures

Publications/Posters

Publications

Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, Teles e Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR, and Middleton D. **2015**. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Research*, **43**: D784–788

Ghattaoraya GS, Dundar Y, González-Galarza FF, Maia MH, Santos EJ, da Silva AL, McCabe A, Middleton D, Alfirevic A, Dickson R and Jones AR. **2016** A web resource for mining HLA associations with adverse drug reactions: HLA-ADR *Database (Oxford)* **2016**: baw069.

Ghattaoraya GS, Middleton D, Santos EJM, Dickson R, Jones AR, and Alfirevic A. **2017** Human leucocyte antigen–adverse drug reaction associations: from a perspective of ethnicity. *International Journal of Immunogenetics* **44**: pp 7-26.

Posters

Poster presentation at the International Society of Pharmacovigilance, 17th annual meeting 2017 Liverpool, United Kingdom 15-18 October 2017.

Ghattaoraya GS, Alfirevic A, Dickson R, Jones AR, O'Hara A, Pirmohamed M. **2017** A clinical decision support tool for genetic HLA testing to predict, pre-empt and diagnose adverse drug reactions. *Drug Safety* **40**: pp 937

Conference talks

Conference talk at Deep Learning in Precision Medicine Workshop - Riva del Garda, Italy 23 September 2016. (Held in conjunction with The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases) **Talk:** Opportunities to Implement Deep Learning within the Field of Immune-Mediated Adverse Drug Reactions.

Abstract

The highly polymorphic human leukocyte antigens are an important player in the immune response against pathogens. However, the variability can also cause problems to patients, clinicians and drug manufacturers due to the phenomenon of immunologically based adverse drug reactions (ADRs) where the interaction between the drug and antigen presenting proteins can interfere with the body's ability to recognise host cells leading the immune system to mount a response against the host. The underlying mechanisms by which this is brought about are not fully understood. Over recent years a number of studies have reported HLA alleles are associated with adverse drug reactions where this number is growing. This has made research into the field increasingly challenging as relevant data has become spread across the literature.

To address this issue, a centralised and searchable web database (HLA-ADR) was constructed which was populated with data extracted from a literature review investigating HLA alleles that have been associated with ADRs. The HLA-ADR database was then used to investigate a question that arose when analysing the data; there was an apparent trend whereby it seems that patients of Asian descent were more prone to ADRs compared to other populations. The analysis in this thesis found no evidence to support this hypothesis and instead revealed potential reporting bias and issues with the poor replication of findings in the literature.

Finally, HISTO SPOT, a new HLA typing method that aims to type patients with high turnaround and at low cost where it highlights patients carrying one or more of a panel of HLA alleles known to be associated with immune mediated ADRs was investigated. The method is able to accurately type patients with 100% concordance with currently implemented HLA sequence based typing. Alongside this, a clinical decision support tool was developed to provide clinicians with a means to interpret the HISTO SPOT results by providing expert knowledge.

In conclusion, the bioinformatical tools developed in this thesis provide a potential framework by which to aid research and clinical decision making. The tools will hopefully advance the field and provide a stepping stone towards the goal of personalised medicine.

Chapter 1. General Introduction

1.1. Adverse drug reactions

One of the biggest issues faced by clinicians and pharmaceutical companies is the risk that patients might experience adverse drug reactions (ADRs) upon exposure to a drug treatment. In the United Kingdom, between 6 and 12% of hospital admissions are attributed to ADRs. [1, 2] Approximately 3-5% of approved drugs in Europe and North America (approved between 2002-2011 and 1990-2009 respectively) were withdrawn from the market for safety reasons with hepatotoxicity being the second most commonly cited reason for both Europe and North America. It is worth noting that hepatotoxicity is not always caused by immune mediated ADRs. Cardiovascular events was the most common reason for drug withdrawal. [3, 4]

1.1.1. Definition and classification of adverse drug reactions

An ADR is defined by the World Health Organisation as 'a response to a drug that is noxious, unintended and occurs at doses normally used as part of standard treatment'. [5] ADRs have been categorised into six subdivisions by Rawlins and Thompson based on the differences of their general responses to the drug. [6] These classifications are defined in Table 1.1. The main focus of this thesis revolves around Type B reactions whilst the mainstream definition for Type B reactions describes these responses as unpredictable. There have been some that argue that due to the implication of particular human leukocyte antigen alleles as risk factors (described in section 1.1.4) the 'unpredictable' description should be removed from the definition. [7, 8]

1.1.2. Clinical manifestations of ADRs

Immune mediated ADRs have been reported to induce a wide variety of phenotypes which can affect any organ with the most common being skin and liver. The severity of cutaneous reactions varies greatly from mild skin rashes to Stevens–Johnson syndrome / toxic epidermal necrolysis (SJS/TEN) – symptoms include but are not limited to extreme blistering and skin detachment and possible mortality. [9]

Table 1.1: Classification of ADRs by Rawlins and Thompson

Type	Definition
A	'Augmented' - The most common variant, also known as 'on target' ADRs. These are the result of the drug's known pharmacology and toxic effects which are part of the desired pharmacological action.
B	'Bizarre' – Also known as 'off target' or 'idiosyncratic'. These reactions have no relation to the known pharmacological effects of the drug and are not dose dependent. These reactions are often referred to as immune mediated and are thought to be unpredictable.
C	'Continuing' – Reactions that are the related to the cumulative dose of the drug.
D	'Delayed' – Usually related to the dose and are reactions that occur after a significant amount of time has passed (from drug administration).
E	'End-of-use' – Reactions that observed after drug treatment has stopped (withdrawal).
F	'Failure' – usually caused by drug-drug interactions.

Maculopapular exanthema (MPE) is the most common type of cutaneous ADR accounting for more than 90% of drug induced ADRs and the time of onset is usually after 1-2 weeks after the start of treatment. [10] SJS/TEN is the most severe type of cutaneous skin reaction and has an approximate mortality rate of 30% SJS and TEN are considered to be the same condition with the difference defined as the extent of skin detachment (< 10% for SJS; 10-30% for SJS-TEN overlap and >30% for TEN). [11]

As mentioned at the start of this chapter, drug induced liver injury (DILI) is one of the leading causes of drug withdrawal. It is a probable site for adverse drug reactions as the organ is key for drug metabolism where many of the drug's metabolites are formed. DILI has two major forms where the alanine aminotransferase (ALT) and alkaline phosphatase (ALP) levels change dramatically relative to the upper limit of normal (ULN). Hepatocellular liver injury is the most common form of DILI and is characterised by a dramatic increase of ALT (≥ 3 times ULN) and the ALT/ALP ratio

(≥5 times ULN) - the time of onset is typically 1-3 months. The other major form, cholestatic DILI, is characterised by ALP ≥2 times ULN and ALT/ALP ratio of ≤2 times ULN. [10, 12] It has been reported that hepatocellular DILI is exhibited in a higher portion of women and tends to occur in younger patients compared to cholestatic DILI where older age and higher male : female ratio has been reported. [13, 14]

Table 1.2 shows a list of well-known ADR phenotypes that have been reported in the literature. The table also demonstrates that different patients on the same drug may exhibit different ADR phenotypes.

Table 1.2: Examples of different drug induced ADR phenotypes

Drug class	Drug	ADR phenotype
Antiretroviral	abacavir	drug-induced hypersensitivity syndrome (DIHS) [15]
	nevirapine	DILI [16]
Anticonvulsant	carbamazepine	SJS/TEN [17] DIHS [18, 19]
	phenytoin	SJS/TEN [20]
	lamotrigine	maculopapular exanthema (MPE) [21] SJS/TEN [22]
Antigout	allopurinol	drug-induced skin reaction (DISI) [23]
Antibiotic	flucloxacillin	DILI [24]
	amoxicillin-clavulanate	DILI [25]
Antirheumatics and NSAIDs	aspirin	aspirin-intolerant asthma (AIA) [26] aspirin-induced urticaria (AIU) [27]
Anticancer	lapatinib	DILI [28]

1.1.2.1. Alternative names for ADR phenotypes

The multitude of ADR phenotypes may also have interchangeable terms used throughout the literature. This thesis may use some of the terms interchangeably however, the term in the left column of Table 1.3 will be the preferred term. These alternative terms were identified using medical subject headings (MeSH). The MeSH

terms are controlled by the United States' National Library of Medicine and acts as a thesaurus for interchangeably used medical terms. [29]

Table 1.3: Interchangeable terms referring to certain ADR phenotypes

Term used in this thesis	Alternative term(s)
drug-induced hypersensitivity syndrome (DIHS)	Hypersensitivity syndrome (HSS) drug rash with eosinophilia and systemic symptoms (DRESS)
drug-induced skin reaction (DISI)	cutaneous adverse drug reactions (cADR) severe cutaneous adverse drug reaction (SCAR)
maculopapular exanthema (MPE)	mild maculopapular eruption (MPE) exanthema skin rash
Stevens-Johnson syndrome (SJS) / toxic epidermal necrolysis (TEN)	erythema multiforme exudativum erythema multiforme major bullosum
drug induced liver injury (DILI)	hepatotoxicity
gout	hyperuricemia

1.1.3. Current understanding of immune mediated adverse drug reactions

Type B reactions are regarded as being immune mediated that whilst rare (20% of all ADRs), can potentially cause severe morbidity and possibly death. [30, 31] In many cases, human leukocyte antigens have been implicated as risk factors for ADR events. [19, 24, 27, 32-36] The exact mechanisms by which the drug interacts with the HLA system is not fully understood however, probable mechanisms of action have been proposed.

All of these mechanisms involve interference between the HLA antigen and T-cell interaction that enables the immune system to identify self and non-self cells.

1.1.3.1. Hapten hypothesis

A hapten is a small molecule (< 1,000 Daltons) that on its own is too small to become antigenic (elicit the production of antibodies) but can covalently bind to a large carrier (e.g. a protein) whereby the modified carrier induces an immune response via T-cell activation. Different haptens can be produced depending on the hapten agent and modified carrier. For example, derivatives of penicillin were shown to bind covalently to lysine residues of serum albumin and that anti-penicillin T-cells are present in the blood of patients hypersensitive to penicillin. [10, 37] This is a possible explanation as to why the different drug-MHC allele combinations may induce different ADR phenotypes.

In the hapten hypothesis model (see Figure 1.1), the drug (or a metabolite) covalently bonds with a protein. The drug-protein hybrid is taken up by a cell and hydrolysed and the peptide fragment containing the bound drug is processed for antigen presenting on the cell surface via the MHC molecules. The modified antigen is recognised by the interacting T-cell as non-self, which then triggers a cascade to induce an immune response. A study by Ortmann *et al*., 1992 demonstrated that hapten modified antigens were able to induce T-cell responses. [38]

1.1.3.2. Pharmacological interaction concept

The pharmacological interaction (p-i) concept was proposed as an alternative hypothesis as it was noted that clones of T-cells obtained from ADR patients would proliferate in the presence of the drug that caused the ADR but outside of any metabolic activity that would generate the drug's reactive metabolite. [39] The theory proposed (see Figure 1.2) that the drug binds noncovalently and reversibly to the T-cell receptors or to the antigen itself, which would interfere with the normal antigen presentation and T-cell interaction and therefore the T-cell would trigger the cascade leading to an immune response as if it had encountered a non-self antigen. [40, 41]

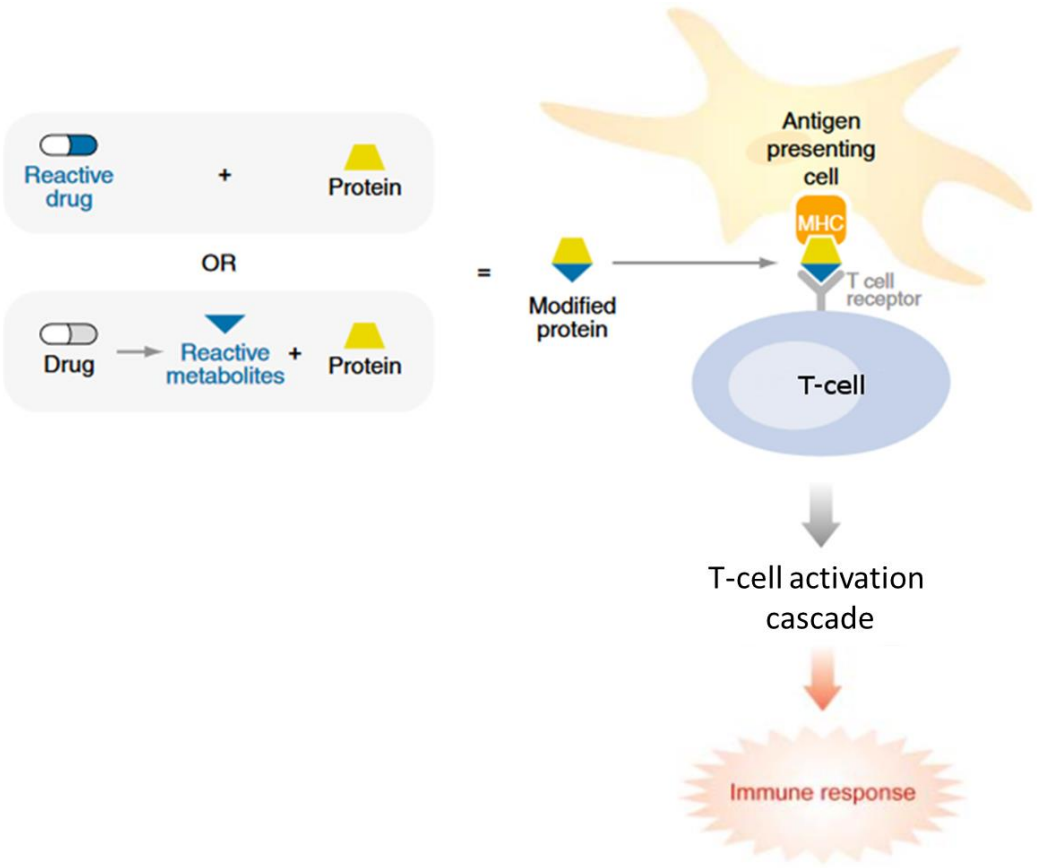


Figure 1.1: Model of hapten hypothesis

Adapted from Uetrecht, 2007 [42]

1.1.3.3. Danger hypothesis

The danger hypothesis was proposed as an alternative as it argued that one signal was insufficient for T-cell activation – i.e. it would need at least one more signal in addition to the altered interaction between antigen and T-cell to trigger the immune response. [43] An individual expresses certain proteins much later in life which are required for development. These proteins are ones that the body would not have previously been exposed to – e.g. hormones released during puberty or pregnancy. The individual's immune system would be unfamiliar with antigens produced from the breakdown of these time/event specific proteins and yet they do not induce an immune response. The theory goes on to add that cellular damage (e.g. caused by a pathogen) would release a so called "danger" signal that combined with the non-self antigen signal will trigger the immune response. Without the danger signal, the result is immune tolerance (see Figure 1.3 showing both the danger and tolerance pathways).

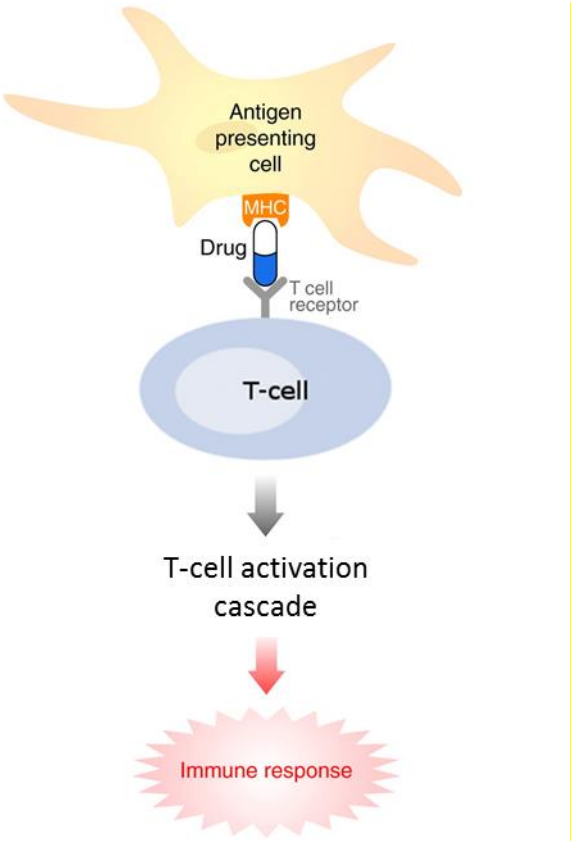


Figure 1.2: Model of p-i concept

Adapted from Uetrecht, 2007 [42]

Despite investigations, the identification of these damage signals has been difficult although suggestions have been offered including: heat shock proteins, interferon-α, interleukin-1 β, uric acid. [44, 45] Although the danger hypothesis is its own separate model, it can hypothetically be additive to the other models discussed in this section.

1.1.3.4. Altered peptide repertoire hypothesis

The altered peptide repertoire model was developed from the discovery of the mechanism of abacavir induced hypersensitivity. Illing *et al*. demonstrated using protein crystallography that unmodified abacavir binds noncovalently inside the F-pocket of the antigen binding cleft of the HLA-B*57:01 MHC molecule. [46] There is exquisite specificity in this interaction that the group was also able to show that this binding does not occur with the HLA-B*57:03 protein which only differs from the first allele by the substitution of two amino acid residues. The result of this binding alters the shape of the antigen binding cleft and therefore the antigen repertoire of the MHC molecule. This means that a different peptide fragment is instead presented

resulting in a T-cell recognising the antigen as non-self and triggering an immune reaction.

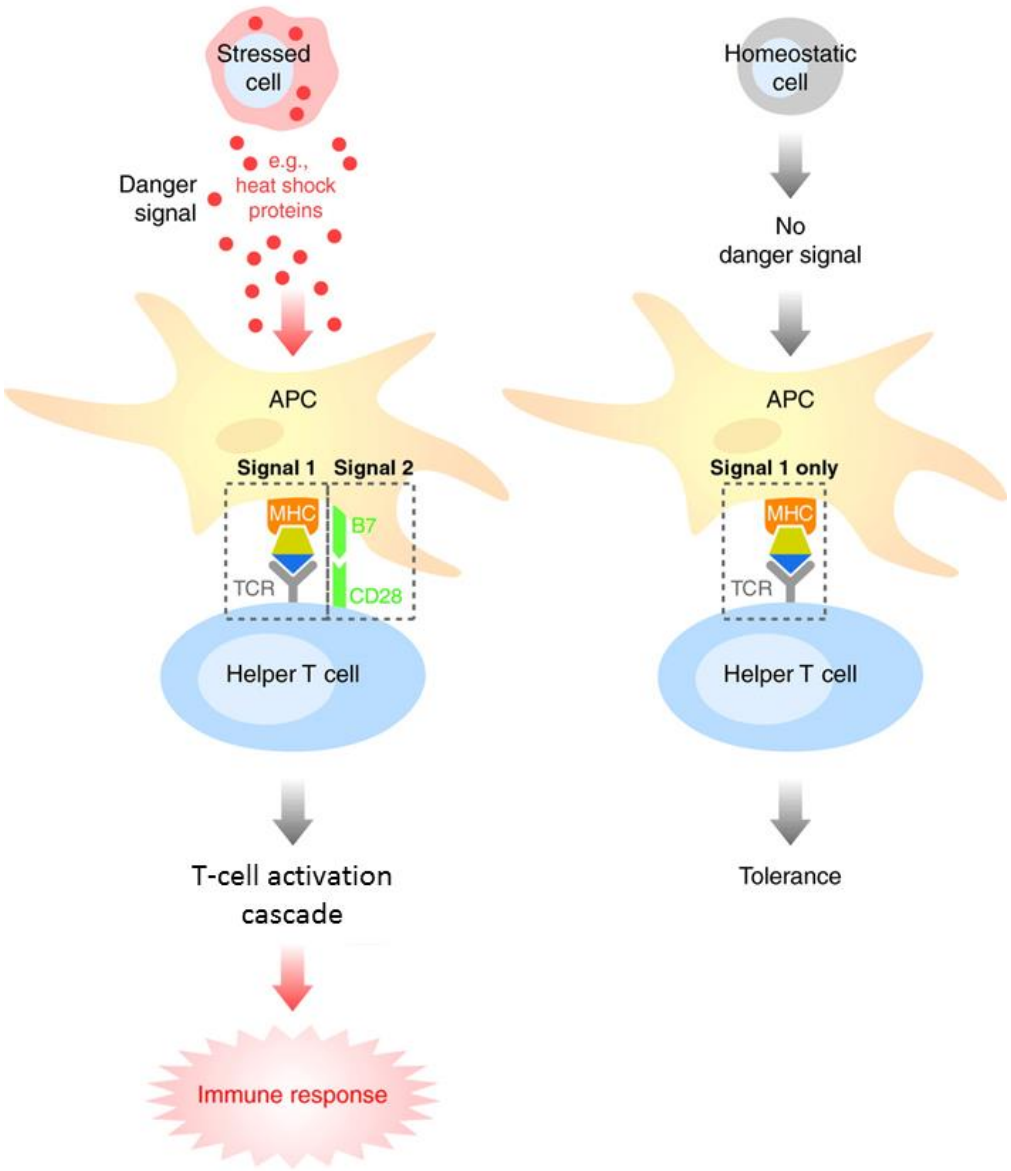


Figure 1.3: Model of danger hypothesis

Adapted from Uetrecht, 2007 [42]

The specificity was further demonstrated when Hughes *et al*. showed that populations with higher carriage of the HLA-B*57:02, HLA-B*58:01 and the above-mentioned HLA-B*57:03 alleles had a lower incidence rate of abacavir hypersensitivity compared to populations where HLA-B*57:01 is higher. [47]

Although the exact model of abacavir hypersensitivity has been demonstrated, the positive predictive value for patients carrying HLA-B*57:01, being treated with

abacavir and developing an immune mediated ADR is 47.9% indicating that other factors are likely involved. [48]

1.1.4. Current clinical guidelines and future perspectives

As the evidence supporting immune-mediated ADRs is increasing, the recommendations for drug usage to clinicians has altered to reflect this knowledge. The warning labels for abacavir have been changed in the regions regulated by the United States Food and Drugs Administration (FDA), European Medicines Agency (EMA) and the Canadian HC-SC bodies to recommend screening for the carriage of HLA-B*57:01 in patients about to undergo abacavir treatment. [49, 50] Genetic screening for HLA-B*15:02 in patients of Asian ancestry is also recommended when considering carbamazepine and the structurally related compounds: oxcarbazepine and phenytoin. [51-53] Other drug allele combinations associated with immune mediated ADRs, have not been recommended for routine genetic screening prior to the start of the treatment. [54]

Monoclonal antibodies that supress the immune response have been used to treat autoimmune diseases for a number of years. [55] However, advances in immunotherapy, allow the generation of antibodies that respond against tumour cells, thereby allowing the individual to mount an immune response against the tumour cells. [56] This has become an important tool in cancer therapy. Very recently, evidence has emerged that a patient's HLA class I genotype can influence how T-cells recognise tumour peptides and how it responds to checkpoint inhibitor immunotherapies. The findings indicate that maximal Class I HLA heterozygosity improved patient survival chances compared to HLA homozygosity for at least one HLA locus. In addition, somatic loss of heterozygosity was associated with poor outcome in two melanoma cohorts. [57] It is likely that increased research into this phenomenon will play an important avenue in the field of oncology. Heterozygosity and homozygosity describe the degree of similarity between alleles for a locus within the genome. As explained in section 1.2.2, the HLA region is very polymorphic. "Maximal heterozygosity" as used earlier describes that the loci of the Class I HLA genes have a very low degree of similarity meaning that the two alleles of the diploid loci confer different protein products.

1.2. Human leukocyte antigens

Human leukocyte antigens (HLAs) are part of the major histocompatibility complex (MHC) region. This region is approximately 4,000 kilobases in length and contains a group of >200 related genes located on the short arm of chromosome 6 (specifically Chr 6p21). The genes within the MHC region are subdivided into three classes based on functionally related but distinct proteins. [58] The focus of this thesis will around classical HLA class I and class II genes. A diagrammatic representation of the MHC region in the context of the chromosome can be viewed in Figure 1.4 below).

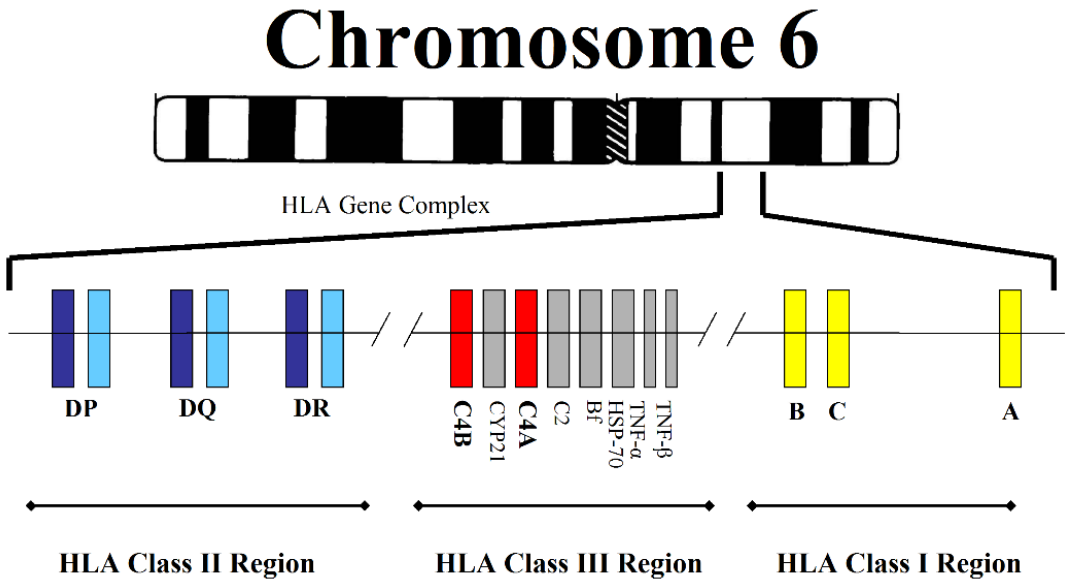


Figure 1.4: Diagram of MHC region located on chromosome 6

Adapted from Westover et al., 2011 [59]

1.2.1. Primary function of classical HLA genes

MHC class I (also HLA class I) genes/proteins present antigens that are generated from the proteolysis of cytosolic proteins, which are then bound to the membrane of the cell. The host organism's own proteins are cleaved and bound to the presenting HLA protein. The antigenic peptide is presented to cytotoxic T-cells (CD8+) which will recognise the peptide fragment (and therefore the cell as a whole) as an antigen belonging to a host cell due to previous exposure to the same antigenic peptide. In the case of bacterial or viral infections, proteins from the invading cell are processed/broken down and exported to the surface of the human cell. In these cases, the presenting antigen is recognised as non-self, inducing cytotoxic T-cells to kill the

cell that is presenting the foreign peptide. Class I proteins are represented in yellow in Figure 1.4. Additionally, see section 1.2.2 for list of genes belonging to this class.

Class II HLA proteins present extra-cellular peptides to T-lymphocytes, which in turn induce the production of T-helper cells (CD4+). For antigens recognised as foreign, the T-helper cells will initiate B-cells to produce antibodies that are specific to that antigen whilst for antigens recognised as host, suppressor T-cells prevent the production of antibodies that are specific to host antigens. Class II proteins are encoded in pairs of genes (producing the α and β chains) and are represented in Figure 1.4 as pale blue and dark blue). Both class I and class II molecules are involved with the adaptive immunity.

Conversely, Class III MHC genes are associated with the complement system in relation to immunity and carry out a variety of different functions including the inflammation response as well as the tumour-necrosis factor alpha (TNF) gene, lymphotoxin alpha (LTA), heat shock proteins and also contain an assortment of non-immune related genes.

Non-classical HLA molecules differ from their classical counterparts by their relative lack of allele diversity. The role of these is not as well defined however, there have been suggestions that these non-classical HLA molecules play a role in mediating the activation of natural killer cells. [60, 61] The classical genes are made up of HLA-A, HLA-B and HLA C for MHC class I loci whilst for class II the classical alleles are HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1 and HLA-DRB1. The other genes that are featured in Table 1.4 and Table 1.5 (see section 1.2.2) are non-classical.

1.2.2. Highly polymorphic characteristic of HLA genes

As of October 2017, there are 17,331 known HLA alleles (see Table 1.4 and Table 1.5), which makes these genes one of the most polymorphic within the human genome. [62] The high levels of polymorphisms, particularly with the classical loci, means that probability that an individual is homozygous for each HLA gene is extremely low with the exception of consanguineous parentage. This high level of MHC heterozygosity maximises the adaptive immunity against infections and disease. [58] Pathogens have

evolved to try and evade the immune system by constantly altering the sequence of their antigen binding peptides with the primary aim of fooling the immune cells into recognising them as host cells.

The extreme diversity found within human population (seen even despite the out of Africa migratory bottleneck limiting genetic diversity in general) has been built from the infectious agents encountered during the course of human evolution and aims to prevent a single pathogen from displaying antigens which are structurally similar to human host antigens for all individuals. This provides a greater chance that at least a subset of the human population will be immune to the emergence of a new pathogen capable of spreading rapidly. In addition to this, the role of self-recognition by the HLA genes is also an important factor in terms of organ transplantation. For a successful transplant to occur, the donor organ needs to present antigens which can be recognised as self by the recipient's immune system.

1.2.3. Structure of HLA molecules

The high levels of polymorphisms generally occur in the antigen presenting region of the molecule allowing for variation in the presented peptide antigen – for MHC class I, this region (also referred to as the α1 and α2 domains) is encoded by exons 2 and 3 of the gene. The other regions of the protein molecule must be conserved in order to maintain their function. Exon 1 encodes the leader peptide (or signal peptide), exon 4 encodes the α3 domain, exon 5 encodes the transmembrane domain and the cytoplasmic tail being encoded by exons 6 and 7. [63] For MHC class II molecules, the final peptide is encoded by a pair of genes which produce the α and β subunits. The exons that produce the domains for each subunit are listed in Table 1.6 below. [64] A diagram showing the arrangement of the MHC class I and class II molecules is shown in Figure 1.5. Both MHC class I and class II polypeptide molecules are heterodimeric proteins. The class I molecule is comprised of a single α chain that is paired with a β_2-microglobulin protein (β_2m). β_2m is not involved with antigen presentation. The bulges in the α1 and α2 domains represent regions that are genetically conserved to allow for T-cell interaction. Similarly in class II, the bulges for the α1 and β_1 domains are also conserved for the same reason.

Table 1.4: Number of class I HLA alleles, protein products and null alleles that are currently known (as of October 2017) for each gene

HLA Class I						
Gene	**A**	**B**	**C**	**E**	**F**	**G**
Alleles	3,997	4,859	3,605	26	26	56
Proteins	2,792	3,518	2,497	8	5	18
Nulls	186	147	131	1	0	2

Table 1.5: Number of class II HLA alleles, protein products and null alleles that are currently known (as of October 2017) for each gene

HLA Class II												
Gene	***DRA***	***DRB***	***DQA1***	***DQB1***	***DPA1***	***DPA2***	***DPB1***	***DPB2***	***DMA***	***DMB***	***DOA***	***DOB***
Alleles	7	2,395	92	1,152	56	5	942	6	7	13	12	13
Proteins	2	1,751	35	779	26	0	655	0	4	7	3	5
Nulls	0	66	3	31	0	0	22	0	0	0	1	0

Table 1.6: HLA class II exons that encode each of the peptide domains

Peptide domain	α chain	β chain
Leader peptide	Exon 1	Exon 1
Antigen presenting domain	Exon 2 & Exon 3	Exon 2 & Exon 3
Transmembrane domain	Exon 4	Exon 4
Cytoplasmic tail		Exon 5

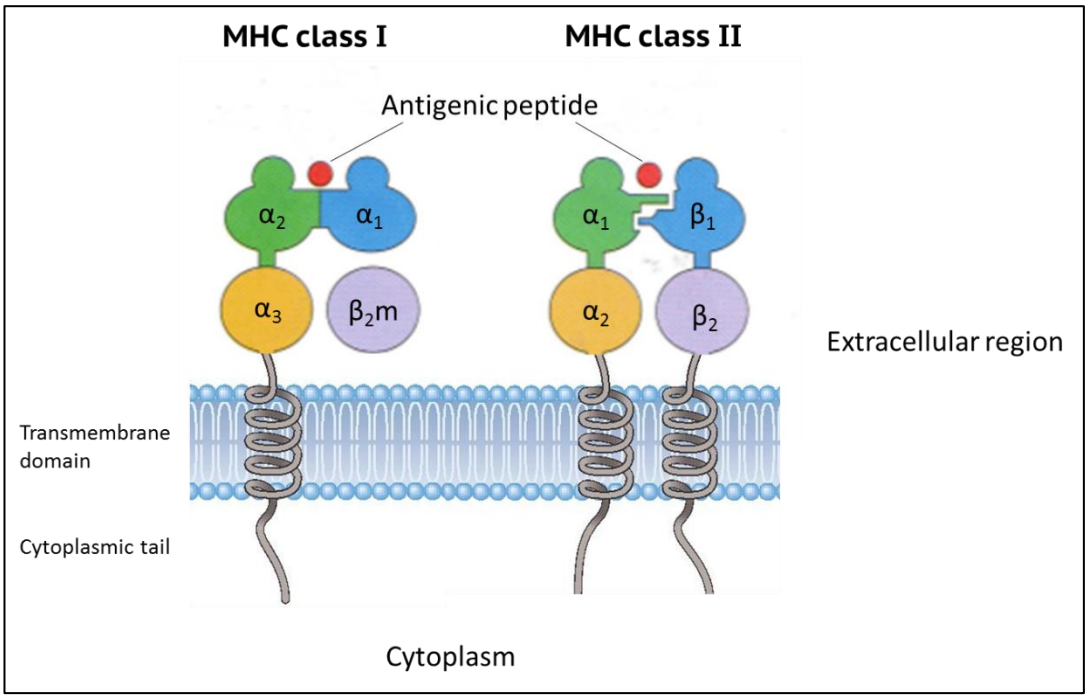


Figure 1.5: MHC molecule structure for class I and class II proteins.

Modified from Lyczak, 2004 [58] and The-Crankshaft Publishing, 2012 [65]

1.2.4. Variation in peptide binding

The MHC/HLA molecules are not limited to presenting a single peptide fragment, instead the HLA molecule is capable of presenting several distinct antigens to the T-cells. The HLA protein typically accommodates a peptide fragment of 9 amino acid residues in length. [58] For class I MHC molecules, the N and C termini of the fragment are used for anchoring to the MHC molecule and the middle bulges slightly. With class II HLA molecules the fragments are not anchored and so longer fragments are possible although 9 residues still appears to be the norm. Peptide fragments of length between 8 to 15 residues have been observed with increased bulging in the middle. [66] There

have been several computational models built to predict binding affinity of peptide fragments to the MHC molecules. [67-74] Many of these work by analysing the peptide binding groove of the MHC molecule and statistically determining the most likely residues that can be accommodated within that particular area of 3D space. Aspects including, but not limited to residue charge, residue size, bond types (e.g. van der Waals, hydrogen etc.) are taken into account with the software prediction algorithms.

1.2.5. Nomenclature system of HLA alleles

This immensely high allelic variation has led to the nomenclature for the HLA gene products to be reformed on multiple occasions. The original nomenclature system, from the 1960s, was based on serotype - utilising antigens raised as a result of the rejection of transplanted organs. There were originally 15 identified antigens HL-A1 to HL-A15. It was later observed that this set could be divided into two groups as it was only possible to identify 2 members of these groups in any individual. As a result, the new A series comprised of the original HL-A1, 2, 3, 9, 10, 11 types and the newly formed series B list contained HL- A5, A7, A8, A12, A13, A14, A15 which were renamed as B5, B7, B8, B12, B13, B14 and B15 respectively. The example of B1 and others were never used as a serotype identifiers to allow for the transition from the original system to this version. Thereby, explaining the lack of a B1 or conversely, an A7 serotype, which also prevails in the current nomenclature system. Over the years, additional groups were identified.

From the 1970s onwards, methods for serotyping began to be implemented across the world for transplantation; however, in certain cases the serotype did not provide enough resolution to give reliable prediction. For example, in some cases patients who were typed and verified to accept organs from a B5 donor could still reject the transplant. Gradually, techniques were improved allowing for the discovery of new serotypes which could specify a smaller, subset of antigens with a higher degree of precision meant that broad antigen groups such as B5 were sub-divided into split antigens (B51 and B52) however; rejection was still a significant risk.

The advent of molecular genetics led to the formal discovery of the C series and also the D group which later became the class II gene sub-family – which is why all of the

class II genes start with a D prefix. With molecular sequencing, the nomenclature system was changed to so that 4 numbers followed the gene letter. The new names were based on the serotype information (e.g. B*5101, B5102 and B5103) are all distinct protein sequences within the B51 serotype. [75] This was later reformed into the current system (through intermediary steps) so that a colon would separate a level of resolution to a maximum of 4 levels, referred to as fields (e.g. A*01:01:01:01). [76] The first level/field denotes the allele group, which is based on the historical serotype (A1) although this does not always apply. The second field allows for differentiation between distinct proteins by the detection of non-synonymous mutations. The 3rd field reports synonymous mutations i.e. HLA-A*01:01:01 produces the same protein as HLA-A*01:01:02. The final level shows the differences in non-coding regions of the gene (see Figure 1.6).

Occasionally a suffix letter is also included which provides additional information. Null alleles are given the suffix 'N' to report this status. Other suffixes include 'L' for low surface expression (relative to normal levels), 'S' to indicate alleles that are expressed as secreted soluble molecules and 'Q' if the expression of the allele is questionable. [77]

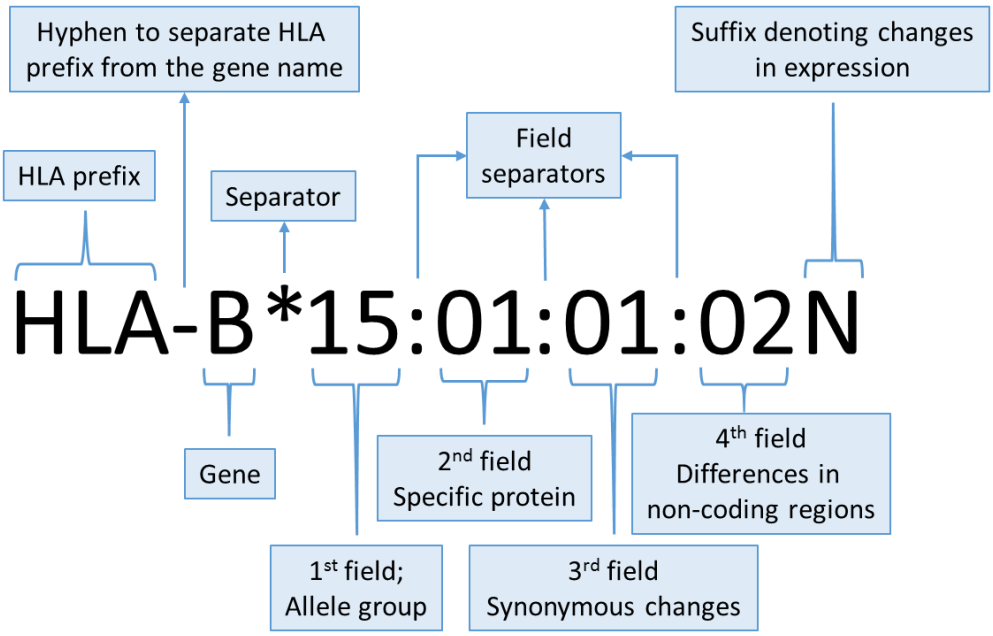


Figure 1.6: Nomenclature system for HLA alleles

The HLA informatics group maintains an internationally recognised IMGT/HLA database, [78] which provides the official repository for HLA sequences for the World Health Organization Nomenclature Committee. For new HLA alleles to be submitted into the IMGT/HLA database, the minimum requirement is to provide the genetic sequences for the exons that encode the antigen presenting domain. Although, complete gene sequence data is preferred.

1.3. HLA typing methods

1.3.1. Serological assay typing

The first method that was employed for classifying MHC/HLA proteins looked at the observed response of the cell samples against specific antibodies that were able to distinguish these proteins based on epitope expression. These antibodies are able to distinguish HLA protein down to gene/loci and the allele group (explained in 1.2.5) meaning that relative to the standardised nomenclature system, HLA proteins can be classified to 1st field resolution using this method. [58]

The tests are performed on patient's lymphocytes using a method known as a complement-dependent cytotoxicity (CDC) test. This involves exposing the patient lymphocytes to a panel of anti-HLA antibodies that are known to react against particular HLA types. These anti-HLA antibodies are usually derived from cell lines or donors where the HLA type is known and are arranged in a plate (or typing tray) within individual wells. The lymphocytes of the individual to be typed is then introduced and distributed across the wells where the cross-reactivity reactions are evaluated based on the uptake of a reporter dye – usually trypan blue or eosin red. The cross-reactivity of the anti-HLA antibodies against the HLA molecules (antibody-antigen complexes) will cause the cells to become permeable and therefore take up the dye and appear that reporter dye's colour when observed visually. The percentage of lysed/stained cells is then measured against a scoring system (see Table 1.7) where a score of >6 (high cytotoxicity) is classed as a positive hit. The antibody belonging to these wells where high cytotoxicity occurred is read as the patient's HLA type. Due to the subjective nature of the interpretation, tests require at least 2 independent technicians to agree before the final HLA type is recorded. [79]

Table 1.7: Interpretation of complement-dependent cytotoxicity (CDC) results

% of lysed (stained) cells	Score	Interpretation
	0	unreadable
0-10	1	Negative
11-20	2	Doubtful negative
21-50	4	Weak positive
51-80	6	Positive
81-100	8	Strong positive

1.3.2. Sequence-specific primer assay

The HLA-SSP typing method works by using (as the name suggests) primers that are complementary to specific HLA alleles or group of similar alleles, i.e. 1st field and 2nd field resolution respectively. These primers are designed so that the polymorphisms in the allele's genetic sequence is identified at the 3′ terminus of the primer. This means that perfect matches to the primer (i.e. the specific allele containing the sequence) will be extended by the *Taq* DNA polymerase during the PCR step. Once the HLA-SSPs are created, they are distributed within microtiter plate where they are then stored frozen or freeze dried and stored at approx. 4˚C. For HLA typing an individual's DNA sample, the sample's extracted DNA is mixed with dNTPs and *Taq* DNA polymerase which is then aliquoted into the microtiter trays that contain the SSP sets and the DNA is amplified by PCR. Modern primer sets have been designed so that all of these primer sets require the same PCR conditions. The amplified (or not) PCR product can be then detected by electrophoresis through agarose gel and intercalating with ethidium bromide to allow the presence of DNA (i.e. the PCR product) to be revealed under exposure of ultra violet light. HLA alleles that correlate to these amplified primers are assigned to the sample's HLA typing status. [80]

1.3.3. Sequence-specific oligonucleotide assay

HLA-SSO typing involves the extracted and PCR amplified DNA samples to be exposed to DNA primers that are known to be specific to all alleles for a particular HLA locus. These SSO typing primers are focused to specific exons: primers targeting exons 2 and 3 for HLA class I genes and exon 2 for HLA class II genes (explained in

section 1.2.3 as the region that confers functional variability). The amplified DNA is then blotted onto a positively charged nylon membrane, one for each oligonucleotide probe, allowing for hybridisation of complementary DNA-oligonucleotide sequences to occur. Stringency washes remove any unbound oligonucleotide probes. The SSO probes are designed such that the allele specific polymorphism is located at the middle of the oligonucleotide sequence, if there are partially bound probes then there is an increased chance of the probe being washed away by the stringent buffer. Hybridised molecules can then be detected using chemiluminescent or colour-metric reactions where pattern of the positive reactions can be used to determine the HLA type. [80]

1.3.4. Typing with SNP tags

SNP tagging is used commonly to type the HLA status of patients and involves amplifying sample DNA with DNA probes that are specific to an allele's single point polymorphism that is known to be in complete (or very high) linkage with the allele and detecting these variations with for example, fluorescently labelled probes or restriction length fragment polymorphism (RFLP) where the amplified DNA containing the different alleles of the SNP can (or can't) be digested with a restriction enzyme. The SNP tag acts as a proxy marker to detect for the presence of an allele of interest.

SNP tagging methods are commonly used as a rapid and cost-effective method for sequencing which detect for the presence or absence of individual and specific alleles – e.g. detect for the presence or absence of HLA-B*57:01. For the SNP tagging methods, the SNP's location can be found either inside or outside the gene/allele of interest. If the SNP is found outside, then the allele must be shown to be in complete (or very strong) linkage disequilibrium with the tagged allele. Not all SNPs can be designed to be within the gene due to the above mentioned difficulties in designing probes. A paper published by de Bakker *et al*. lists SNPs that can be used as a proxy for many common HLA alleles. [81]

1.3.5. Sequencing based typing

1.3.5.1. Sanger sequencing

Sanger sequencing is a technique where the DNA sample that is to be sequenced (after being amplified) is mixed with a DNA primer (tailored for the region to be sequenced) and a DNA polymerase enzyme and a mixture of deoxynucleosidetriphosphates (dNTPs) and modified di-deoxynucleosidetriphosphates (ddNTPs). The modification of the ddNTPs render the nucleotide chain to be unable to be extended by the DNA polymerase enzyme and the ddNTPs are also labelled (fluorescent or radiolabelled). The sample DNA is separated into four individual sequencing reactions/lanes (A, C, G, T) which also contain dNTPs, one of the four ddNTPs at a 100-fold lower concentration relative to the dNTPs and the DNA primer that is bound to the sample DNA. The procedure allows several copies of the same region to be made which are of different length due to the terminating ddNTP. The fragments are then separated by size (e.g. gel electrophoresis) and the radio/fluorescent labels are measured. The differences in fragment size allow the sequence to be built using the terminating ddNTP labelled base residue. In the case of HLA, the Sanger sequencing procedure is performed using the same exons as SSO typing (exons 2 and 3 for class I HLA genes and exon 2 for class II). [82]

1.3.5.2. Next generation sequencing

The technology for next generation sequencing (NGS) has come a long way in recent years to overcome some of the challenges faced with typing HLA using this method (see section 1.3.6 below). Many different NGS technologies exist and the process differs from all other sequencing methods discussed in this section as the method looks to sequence across the entire gene (including all exons, the untranslated terminals, introns, and also some upstream and downstream regions beyond the gene itself). This means that the process can be described as "ultra-high resolution" where fourth field resolution can be achieved (see section 1.2.5 on HLA nomenclature) i.e. resolution of polymorphisms in intergenic regions, as well as exonic.

The NGS process works by sequencing the PCR amplified DNA across long regions where the DNA primers, that enable DNA extensions, overlap so multi-fold coverage of the sequence is obtained. The overlapping fragment sequences/scaffolds allow the

entire gene to be sequenced using the NGS machines included software as the overlapping sequences are pieced together to form a single contiguous sequence.

1.3.6. Comparison of the typing methods

Serological typing methods typically do not offer typing resolution beyond the first field. Even then, there have been reports of some HLA class I specificities which have been called inaccurately where in one study 18.5% of HLA-A and 25.4% of HLA-B typing showed discrepancies with some of the causes were reported as antigen mis-assignment or missed antigens. [83] This study was conducted using a black patient cohort meaning that serological testing needs to take into account patient ethnicity. Another study reported a mis-assignment of HLA-B antigens by serological methods to be 22.5%. [84] Although serological HLA typing has been largely supplanted by DNA based typing methods, the process is still sometimes used in conjunction with DNA based typing methods as a means of determining whether the allele is expressed on the cell surface; many null alleles have been identified this way. [85]

SSO typing is very proficient at processing a large number of samples in a single run. This makes it ideal for HLA typing laboratories, which need to perform these tests routinely where, with the inclusion of control samples, approx. 90 samples can be run in a single microtitre tray. The system is also adaptable and additional membranes containing new oligonucleotides can be added to detect new alleles or resolve ambiguities. The most significant disadvantage to the SSO based typing is that it is unable to distinguish between *cis* and *trans* polymorphism whilst SSP typing implements a system of using both forward and reverse primers allowing it to detect *cis* and *trans* polymorphisms that are located between the forward and reverse primer binding sites allowing it to resolve some of the ambiguities that would disadvantage SSO typing allele calls. Current SSO typing methods take approximately 2 days from the PCR step before the full result of the typing can be assigned. Whilst not quick, the above mentioned high throughput enables efficient typing of many samples in a cost effective manner. The SSO primers are also designed to select from a non-variable part of the gene meaning that as long as the polymorphism does not exist in this region and is covered in the region covered by SSO typing, SSO typing is better at detecting new, previously unknown alleles compared to SSP as it will appear as a new pattern

that does not correspond to the reference of known alleles. Both SSO and SSP typing methods are only able to detect polymorphisms that occur within exon 2 (and in some cases, exon 3). It is generally accepted that only polymorphisms in this region confer peptide binding specificity [86] - i.e. functional changes which is the region that is covered with the hypothesised ADR models detailed in Chapter 1.1.3.

Although SNP tagging provides a very quick and cheap method of HLA typing, it is impractical to type for a large range of alleles due to the need for a labelled probe/restriction enzyme for each allele (Chapter 1 discusses the thousands of known alleles). It can however, test for the presence/absence of specific alleles which means that in very particular cases (e.g. testing for HLA-B*57:01 prior to abacavir treatment) the method could be used to detect for that particular allele – the method has been recommended for pharmacogenetic screening (see example featured in Chapter 2.2.3), however this would require different SNPs in populations of different ancestry. Additionally, an absence result won't tell the user what the allele actually is so negative results only offer very limited HLA typing information depending on how much information is desired. This limitation may still provide adequate information in some circumstances. For example, in the context of ADRs, a clinician may only need to know if a patient carries a risk allele, if not then the allele is insignificant information.

Whilst these limitations exist in current SNP tagging methods, advances have been made whereby HLA alleles can be imputed using multiple SNP positions – for example HIBAG. [87] SNP imputing algorithms that were designed previous to this are limited as they require extremely large ethnic controlled cohorts (>1,000) as training data in order to make predictions with sufficient accuracy. These forms of training data sets are not readily on hand resulting in increased difficulty in accurately calling alleles that are rare in these populations. [88] With newer imputing algorithms, such as the above mentioned HIBAG, the limitation of controlling for sample ethnicity can be overcome due to the use of multiple data points to analyse, greatly increase the call accuracy. However, further analysis and scrutiny of the newer approach method is required before it can be readily implemented.

Sanger sequencing is a more established method of sequence based typing compared to NGS. It is able to obtain the sequence over exon 2 (and also exon 3 for class I genes) and therefore providing more information, in terms of raw data points – i.e. more of the gene sequence, compared to SSO, SSP and SNP tagging methods. However, the above mentioned ambiguities brought about by limiting sequencing to exons 2 and 3 apply with Sanger sequencing. It has been reported that this ambiguity can affect around 66% of HLA-A, 71% of HLA-B and 58% of HLA-C allele calls using traditional Sanger sequencing although the study also reported on an in-house Sanger sequencing technique with 4.4%, 4.4% and 0% ambiguity for HLA-A, HLA-B and HLA-C respectively – note that these ambiguities do not necessarily translate to an incorrect call as SNP tags and other known polymorphisms in high linkage along with ethnicity/population data can be factored in to statistically infer the most probable allele. [82]

Although the genomic HLA region is highly polymorphic, it also shows a high degree of sequence similarity across the different loci which makes it very challenging to design primers/probes to amplify/detect the alleles. Consequently, sequence based typing methods of HLA alleles to high resolution is expensive, time consuming and laborious procedure. This issue is less problematic in the rest of the genome where the technology is widely accessible. This meant that at the turn of the millennium, although DNA based methods such as SSO and SSP typing were available, serological typing was still the primary typing method and it took many years after that before DNA based typing methods became the *de facto* method. However, more recently, the technology for next generation sequencing has improved to overcome many of the issues faced. Previously, due to the high sequence similarity in some areas of the MHC region, it was difficult to form overlapping contiguous sequences using earlier forms of NGS technology. However, with newer technological advances, massive parallelisation, high throughput and longer reads can be achieved at more affordable costs allowing for not only sequencing of entire individual HLA genes, but the possibility of phasing the entire MHC region. [89] The ability to phase the MHC region allows for haplotype information to be obtained and the ability to analyse the non-coding regions of the gene that could affect expression. The other methods which limit themselves to just exons 2 and 3 (e.g. SSO, SSP etc) have some inherent ambiguity with the final called allele, for example, the combined exons 2 and 3 sequence for

HLA-A*02:01:01 is identical to 21 other HLA-A*02 alleles. [82] The biggest drawback for all forms of NGS methods is that the turnaround time is still relatively long which negatively impacts its suitability in clinical scenarios where expedient results are required.

1.4. General database design

Over recent years, as computational power has increased and become more widespread, the usage within biology has similarly expanded. There are many aspects to employing computational approaches to solving biological problems, too many to detail in this thesis. This thesis makes particular use of websites and databases - which will be discussed in this section.

1.4.1. Relational databases

Databases are used to store and allow for easy retrieval of information in a centralised manner, this makes them ideal for websites which require a collection of large amounts of data. Relational Database Management Systems (RDBMS) are one of the most common ways that, as indicated in the name, a database is managed by system administrators. On top of the basic functionality described above (easy storage and retrieval), an RDBMS also provides some levels of security (i.e. permissions and user access), data integrity and other administrative actions (e.g. performance monitoring, data backup and recovery). The security aspect is of great importance as it allows system administrators with user credentials that grant them full read, write and modify access of not just the data within but also for the database schema (an example is described in Chapter 3.2.2) whilst also restricting end users to only view data within the confines of their permitted access rights, in most cases the end user will only be able to view information from a select set of the overall data. These restrictions based on user access provides the database administrators with the ability to effectively maintain and manage the database whilst preventing end users from accidently or maliciously altering the data or accessing restricted parts of the data.

Relational databases contain one or more tables to house the data. Each row of the table represents one record with each column representing a field. Table 1.8 below shows a simple example based on a fictitious student course registration database

displaying students and the subjects they are taking along with their teachers. In this example, each student is a record within the table which consists of several fields to represent a component of the record, in this case, the student's name, subjects/courses they are enrolled on, the classroom where they are taught and the staff member who teaches the subject. This system of records and fields allows for efficient storage of information that needs to be kept in a rigid and organised structure. Efficiency can be further increased using a concept known as "database normalisation" to spread the data across multiple tables in order to reduce data redundancy and ensure data that has dependencies are stored logically (see 0).

Table 1.8: Fictitious example of a basic database table structure using a class enrolment scenario. Columns represent a field or aspect and rows represent a record.

student_id	first_name	last_name	courses	room	teacher
58001	Laurits	Anderson	Biology, Computing	103, 105	Woodward, James
58002	Amy	Jones	Biology, Maths	103, 104	Woodward, Shaw
58003	Manpreet	Kaur	Biology, Chemistry	103, 102	Woodward
58004	John	Smith	Computing, Maths	105, 104	James, Shaw

1.4.2. Database table normalisation

The process of database normalisation aims to decompose tables in order to reduce data redundancy that can occur between records. There are a minimum of 3 levels of normalisations that are typically employed, 1st normal form, 2nd normal form and 3rd normal form (1NF, 2NF and 3NF respectively), higher levels do exist but are not typically implemented unless using extremely complex relationships. [90] Although database normalisation initially changes the data layout to a more complex structure (compared to a single 2 dimensional table of simply rows and columns), from a computational point of view, database normalisation allows for easier data maintenance and updating in situations with larger and/or complex databases containing more information. [91] Using the above Table 1.8 as an example, there are some unnecessary data redundancy so this section will describe each forms of normalisation.

The first normal form requires that each column only contain one attribute where all entries in that column must be of the same type and that each row must be unique. [92] With regards to Table 1.8, the rule that each column only contain one attribute has been violated, whilst the other aspects of the first normal form are true – student IDs column only contains their unique identifying number, the first name and surname columns only contain "name" information for the student, the courses column only contains the subjects the students are enrolled on with their rooms column containing the room numbers and the teachers' names in the teacher column. In order for the table to meet the requirements of the 1st normal form the individual items within the courses, room and teachers column to have separate records/rows as shown in Table 1.9.

The 2nd normal form (2NF) requires that the database conforms to 1NF and that each column can only exist in the table where it is either the primary key or the column is dependent of the entire primary key for its existence; a primary key is a column or columns which can uniquely identify that particular row. In this context, the "student_id" column represents a primary key for the individual however, classroom number and the teacher is dependent on the course/subject, not the individual's student ID which therefore violates 2NF. In order to conform to 2NF, the table will need to be separated; one table will describe the students whilst the other depicts the course characteristics. This reduces data redundancy as the teacher and room information are not needlessly duplicated for each student. It also makes input errors or mistypes when entering the information easier to identify as the mistake will affect all records associated with that key rather than a specific record as editing a single record is more efficient than editing the same mistake/type for multiple records (see Table 1.10).

A junction table (c) is required to show the relationship between the students and courses because the enrolment information was lost in the process to achieve 2nd normal form. In comparison to Table 1.9 the data redundancy has been reduced where the student details have not been duplicated on a per enrolled course basis.

Table 1.9: Showing the information displayed in Table 1.8 which conforms to the standards of the 1st normal form (1NF; for database normalisation).

student_id	first_name	last_name	course	room	teacher
58001	Laurits	Anderson	Biology	103	Woodward
58001	Laurits	Anderson	Computing	105	James
58002	Amy	Jones	Biology	103	Woodward
58002	Amy	Jones	Maths	104	Shaw
58003	Manpreet	Kaur	Biology	103	Woodward
58003	Manpreet	Kaur	Chemistry	102	Woodward
58004	John	Smith	Computing	105	James
58004	John	Smith	Maths	104	Shaw

Table 1.10: a, b and c Showing the information from Table 1.8 which conforms to both 1st normal form and 2nd normal form (1NF and 2NF respectively)

a)

student_id	first_name	last_name
58001	Laurits	Anderson
58002	Amy	Jones
58003	Manpreet	Kaur
58004	John	Smith

b)

student_id	course
58001	Biology
58001	Computing
58002	Biology
58002	Maths
58003	Biology
58003	Chemistry
58004	Computing
58004	Maths

c)

course	room	teacher
Chemistry	102	Woodward
Biology	103	Woodward
Maths	104	Shaw
Computing	105	James

The next stage in the process of database normalisation is 3rd normal form (3NF) which requires, in addition to conforming to 1NF and 2NF, that each of the columns in a table must be either the primary key or dependent only on the contents of the primary key and no other column. In the previous example in Table 1.10 the conditions for 3NF were violated as the information for the teacher in Table 1.10b was dependent on the course information but did not require the room number for its existence. Therefore, the course and teacher information required its own table.

Although higher forms of normalisation exist to take the process beyond 3NF, the other forms are rarely implemented unless the information has very complex relationships. The database described in this chapter was only decomposed to conform to the standards of 3NF. A common mnemonic that is used to remember the order of database normalisation and over simplistic description to the process is a phrase whereby the non-key fields/columns in a table must be "dependent on the key (1NF),

the whole key (2NF) and nothing but the key (3NF), so help me Codd" where Codd refers to the author who proposed the process of database normalisation. [93, 94]

It is also worth noting that, decomposing the data such that it is spread over multiple tables must then rely on relationships between keys and non-key fields in order to maintain the integrity of the information. Using the example in Table 1.11, although the student and course information are kept in separate tables, the junction Table 1.11c provides a link to relate the other 3 tables together such that all of the information can be accessed when using the correct query (see Chapter 1.4.3.2).

1.4.3. Database considerations

1.4.3.1. Available relational database management systems (RDBMS)

There are many developers that supply RDBMSs. Accurate numbers for database management systems that are based on usage are difficult to obtain. However, DB-Engines provides rankings based on web metrics and lists the major providers for the various database management systems. [95] Table 1.12 shows the top 20 database management systems based on DB-Engines' scoring system. By far the most prevalent type of database model used is the relational database management system which has been described previously in Chapter 1.4.1. For this chapter, an RDBMS was chosen as the strict, logical structure requirements enforces good data handling practice, the data lends itself quite well for this logically enforced model and the prevalence of RDBMS models provides the database, that will underlie the clinical decision support tool, with great flexibility in terms of deployment as a greater number of hosting servers will be able to support RDBMS allowing for greater choice for server deployment. Specifically, the database was built using MySQL as this provides the greatest level of flexibility for RDBMSs as the language itself is directly supported by Oracle. The differences between Oracle's RDBMS and MySQL is that the latter is the community supported/open source implementation based on Oracle. Additional RDBMSs such as MariaDB (ranked 18th in Table 1.12) are based upon, as a fork of, MySQL which allows MariaDB to support code written in MySQL, however, this cross-compatibility could eventually be lost depending on the direction of the fork/divergence. Overall, writing SQL queries with MySQL provides the greatest level of flexibility overall for database deployment.

Table 1.11: a, b, c and d Showing the information from Table 1.8 but now conforms to 1st normal form, 2nd normal form and 3rd normal form.

a)

student_id	first_name	last_name
58001	Laurits	Anderson
58002	Amy	Jones
58003	Manpreet	Kaur
58004	John	Smith

b)

course	room
Chemistry	102
Biology	103
Maths	104
Computing	105

c)

student_id	course
58001	Biology
58001	Computing
58002	Biology
58002	Maths
58003	Biology
58003	Chemistry
58004	Computing
58004	Maths

d)

course	teacher
Chemistry	Woodward
Biology	Woodward
Maths	Shaw
Computing	James

Table 1.12: DB-Engines Ranking of database management systems based on their popularity.

Rank	Database Management System	Database Model	DB-Engines Score
1	Oracle	RDBMS	1374.88
2	MySQL	RDBMS	1349.11
3	Microsoft SQL Server	RDBMS	1226
4	PostgreSQL	RDBMS	369.44
5	MongoDB	Document store	332.77
6	DB2	RDBMS	191.25
7	Microsoft Access	RDBMS	126.13
8	Cassandra	Wide column store	124.12
9	Redis	Key-value store	121.51
10	Elasticsearch	Search engine	115.98
11	SQLite	RDBMS	113.86
12	Teradata	RDBMS	78.37
13	SAP Adaptive Server	RDBMS	66.91
14	Solr	Search engine	66.02
15	HBase	Wide column store	63.62
16	Splunk	Search engine	60.3
17	FileMaker	RDBMS	58.65
18	MariaDB	RDBMS	54.36
19	SAP HANA	RDBMS	47.94
20	Hive	RDBMS	46.2

Note: The method for calculating DB-Engines' score is listed in full on their website (https://db-engines.com/en/ranking_definition) but in general, it is based on internet/social media metrics and job advertisements. The scores are current as of July 2017.

1.4.3.2. Structured Query Language (SQL)

The SQL based RDBMS also enables for the retrieval of data from the database with specific commands to refine the search to specifics that the user desires. [96] This is achieved with the use of HTML elements embedded into webpages such as dropdowns/radio buttons etc. whereby users can chose from the available options. Once submitted, the webserver can generate the SQL code based on the selection. Effectively, the SQL code is already partially complete and just requires the user selected options to fill in the missing section. For example, using a simple scenario involving fruit as described in Table 1.13, the webserver already has the SQL code "SELECT column_fruit FROM table_fruits WHERE column_fruit = '<<user option>>';". Once the user selects an option the code is then completed and then executed; example "SELECT column FROM table_fruits WHERE column_fruit = 'gooseberry';". In this scenario the record for gooseberry is retrieved.

Multiple entries can be retrieved as well; continuing with the fruits example, a user can search for all the berries if they wanted to by using "SELECT column_fruit FROM table_fruits WHERE fruit_type = 'berry';" which will return banana, gooseberry and blackcurrant.

Table 1.13: Hypothetical table called "table_fruits" which is made of the columns (fields) of the identification number, the fruit names and the botanical category that fruit belongs to.

id_fruit	column_fruit	fruit_type
1	apple	pome
2	pear	pome
3	banana	berry
4	gooseberry	berry
5	blackcurrant	berry
6	blackberry	aggregate fruit

1.5. Web development languages

The server-side programming languages are used to generate interactive webpages, which allow the users to interact with the data stored on the database. The webpages act both to limit the users' access to the database by removing an interface with direct database access, this is due to the possibility of holding sensitive information such as user login credentials etc.). The webpages also act as a means of a friendly graphical user interface (GUI) so that the information can be provided in an aesthetically pleasing manner.

It is difficult to obtain the full usage statistics for the various server-side programming languages available as this is not usually declared or automatically collected. However, the W^{3}Techs website aims to specifically collect metrics regarding the server-side programming languages used by websites. [97] The methods by which they obtain the data is described within their FAQ page. An adapted bar chart showing the W^{3}Techs data is shown in Figure 1.7. The usage statistics show that the PHP programming language dominates the market share and therefore would be highly likely to be installed by hosting server providers. Similarly to SQL based RDBMS solutions described earlier, developing a website with PHP served webpages will allow the website the greatest amount of flexibility for deployment.

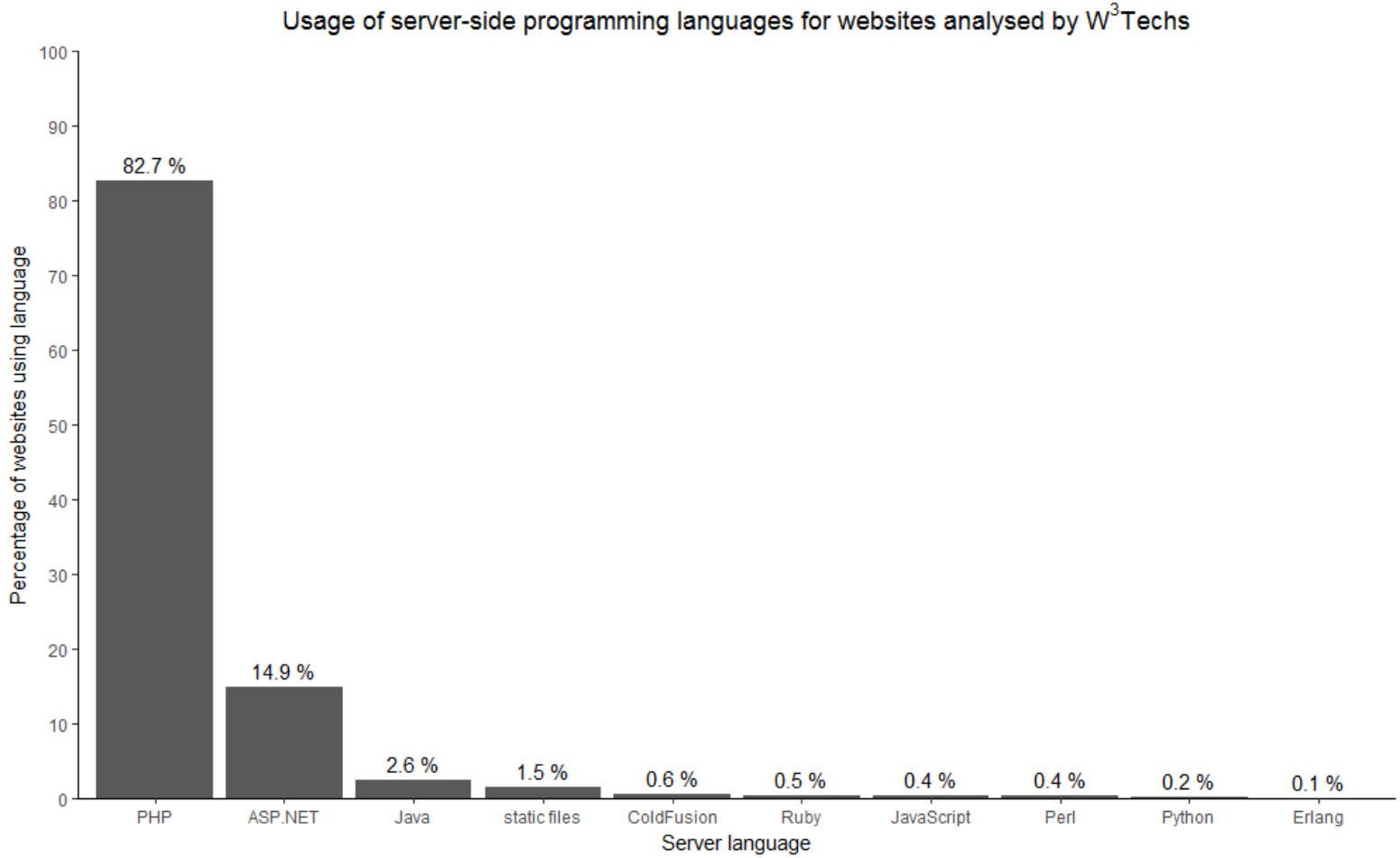


Figure 1.7: Bar chart showing usage statistics of server-side programming languages that are used to serve webpages across the web. [97]

Not all available languages are represented here.

Note that websites may use more than one language to generate its webpages.

1.6. Aim of thesis

The fields of pharmacogenomics and immune-mediated adverse drug reactions has been expanding at an immense rate. New HLA allele-drug combinations are being reported every year meaning it has become ever more important for clinicians and pharmaceutical companies to keep up to date with this growing field for reasons of patient health and the development of new medicines. It is therefore clear that the importance of research into ADRs has also developed alongside the growth in the literature.

However, much of this research is dispersed amongst the published articles which makes the task of keeping track of this field ever more difficult. This thesis aims to address this by:

1. Review the literature and collect information to gather an overall picture for the field of adverse drug reactions that have been associated with HLA allele carriage.

2. Exploring computational methods that will enable researchers to interact with the collected data and enable new methods of conducting analyses for investigating immune mediated adverse drug reactions.

3. Validate a new method of HLA typing that is designed to test patient DNA for the carriage of one or more HLA alleles in a panel of preselected HLA alleles that are known to cause adverse drug reactions

4. In addition to the aim 3, develop a clinical decision support tool that will aid clinicians with interpreting the results from the HLA typing apparatus.

Chapter 2. A systematic review to examine the association between immune-mediated adverse drug reactions and patient HLA genotype

2.1. Introduction

Due to the great impact that ADRs can have on patient health as well the financial investments made by the pharmaceutical companies into drug development, research into this field of pharmacogenetics is vitally important to patients, clinicians and the pharmaceutical companies. However, one of the main obstacles that these groups face is that much of the research is dispersed amongst the literature. Several literature reviews regarding ADRs have been published covering specific drugs, drug classes or a broad range of drugs. [32, 33, 98] However, the field of pharmacogenetics is rapidly evolving with new ADR-inducing drugs becoming reported on a continuous basis. This presents great challenges for researchers attempting to keep up to date with the field and obtain an overall view of the field.

2.1.1. Current resources available for researchers

Researchers have several methods for obtaining published literature regarding ADRs and three major resources have been listed below.

HuGE Navigator is a web-based tool which contains a database of published PubMed abstracts of studies relating to genetic associations with diseases. [99] The studies within the database are manually curated from the NCBI Medline/PubMed. This tool has the ability to allow for searches using MeSH terms (Medical Subject Headings available through- http://www.nlm.nih.gov/mesh/meshhome.html) which standardise alternative names of diseases giving them specific keywords. This allows many related terms to be searched using only one search term.

MEDLINE is a web resource containing one of the largest repositories of citations and title/abstracts from peer reviewed published biomedical articles and is hosted by the (United States) National Library of Medicine. [100] PubMed is the web-browser based interface to retrieve abstracts from MEDLINE. Similarly, the Excerpta Medica

database (EMBASE) also offers bibliographic records with citations for biomedical and pharmacological peer-reviewed articles. [101]

2.1.2. Aim and objective for systematic review

The primary aim of this systematic review was to summarise the relevant data that examine the relationship of adverse drug reactions and patient HLA genetics. Specifically, what genetic factors of the HLA gene family are involved with immune-mediated adverse drug reactions?

2.2. Materials and methods

2.2.1. Literature search

To answer the question laid out in the aim for the systematic review, the first step was to retrieve peer-reviewed publications containing the case-control studies. This involved searches using the three major tools (see Table 2.1) that allow users to retrieve published studies involving immune-mediated ADRs.

Table 2.1: Summary of the literature resources used to retrieve studies for this systematic review.

The table also includes the search terms that were entered to create the search.

Literature resource	Search terms used
PubMed / Medline	Main search (papers published between the year 2000 until April 2014) – see Search strategy 2.2.6.1 MEDLINE Supplementary search (papers published up to April 2014) - "<<drug>> AND (hypersensitivity OR pharmacogenetics OR HLA)" Follow-up search (April 2014 – August 2015) and a second covering (August 2015 – October 2017) – see section 2.2.5 Semi-automated literature mining
EMBASE	See Appendix Search Strategy 2.2.6.2 EMBASE
HuGE Navigator	"Drug Toxicity AND HLA"

The date range for the literature search started with the year 2000. This date was chosen based on an initial pilot search that was conducted using only the results from HuGE Navigator search (this was a curated set of data that has no initial start date). When looking at this data, it was noted that studies which featured high resolution typing (a requirement of the literature search that is described in greater detail in section 2.2.3) were published after 2000. Articles published before this date featured low resolution HLA typing. This was the basis as to why the main literature search used the year 2000 as the start for the date range. April 2014 was the selected end date as this was the time the search was conducted therefore, this provided the ability to select the end date that would provide the latest information available at the time of the search.

With HuGE Navigator, ADRs come under the MeSH term "drug toxicity". The tool also includes the alternative terms: "Erythema Multiforme", "Drug Eruptions", "Poisoning by medicaments NOS" and "Toxic Epidermal Necrolysis". For this review, the search was limited to just HLA genes so as to remove other genes implicated with ADRs, for example, cytochrome P450 genes. Therefore, the term "Drug Toxicity and HLA" was entered into the search.

For both MEDLINE and EMBASE, a search strategy for each resource was developed (see section 0) and implemented. The search was constructed using keywords related to HLAs and drug hypersensitivity. The search also included a list of drugs which were featured in recent publications reviewing drug hypersensitivity. [31, 33, 102, 103] The list can be found on each search strategy (section 0 - identical in both) under entry 9. The list of drugs is designed to cast a wider net for the search. It allows to the search to include these drugs (prior knowledge) but does not function to restrict the search to these drugs. The list was also used to perform a supplementary search within PubMed. Here, the search term followed the format: "<<drug>> AND (hypersensitivity OR pharmacogenetics OR HLA)" where <<drug>> represents one of the entries from the drugs list e.g. "Abacavir AND (hypersensitivity OR pharmacogenetics OR HLA)" was then followed by "Allopurinol AND (hypersensitivity OR pharmacogenetics OR HLA)" until searches using all the drugs in the list were covered. This was achieved through using a computer script written in the Python computing language and used the e-utils module which allows for access and interaction with the Medline

application program interface (API). [104] The script would then collect the PMIDs from the results of each search and place the PMIDs into a basic text (.txt) file.

The results from all of these searches were collected as a list of PubMed Identifiers (PMID) and then the individual PMID lists (HuGE Navigator, MEDLINE, PubMed and EMBASE) were merged into one super-list in which duplicates were removed, thereby representing published articles found by the literature search that will be considered for review.

2.2.2. Literature screening

Once the list of publications (as represented by their PMIDs) was composed, it was assessed to determine if it contained publications which featured data relevant to the systematic review. This process was divided into two stages of screening with the details of eligibility summarised in Table 2.2. For the first stage, the titles and abstracts were reviewed to identify potential studies of relevance and the second stage involved examining the entire study manuscript to fully assess whether the paper fulfilled the eligibility criteria. A second reviewer (Ana Alfirevic) also assessed these selected studies with discrepancies being resolved through discussion.

2.2.3. Eligibility criteria

For this review, only case-controlled studies or randomised controlled trials that investigated ADRs in patients were considered. These investigations were required to provide statistical evidence (e.g. number of study participants, p-values, and odds ratios with 95% confidence intervals) so that statistical analyses could be undertaken by users of the data (eventually made available in Chapters 3 and 4). Studies were also limited to those where the participants' HLAs were genotyped to determine specific proteins (high resolution typing sometimes referred to as "4-digit resolution"). This was decided because, with the example of abacavir, it has been demonstrated that HLA proteins which are part of the same antigenic group and only differ by a few amino acids do not stimulate the same adverse response. [46] This means that protein specific resolution (> 4-digits) is required to distinguish between HLA alleles that are associated with ADRs and those that are not. Consequently, serotype or antigenic resolution alleles do not contribute much meaningful information. However, studies

were included in cases where they provided a mixture of high and low resolution typing.

Table 2.2: Inclusion criteria for the systematic review.

Studies were considered for data extraction if they included the following criteria.

Population	Patients with a hypersensitivity reaction to drugs administered as part of standard treatment
Study design	Retrospective and prospective case-controlled studies, randomised controlled trials
Statistical evidence	Provided statistical evidence (e.g. number of cases and controls, p-values, odds ratio with 95% confidence intervals) to determine the strength of the association.
HLA typing	Investigations where patient and control HLA status was determined to the protein level (high resolution typing sometimes referred to as "4-digit resolution"). Studies where the HLA allele were genotyped within the gene itself, i.e. not using proxy SNPs.
Study date	The formal systematic review was performed covering studies with the publication year between 2000 and the time of the search (April 2014).

In addition to this, studies which are able to genotype the patient-control cohorts at high resolution with high confidence were considered for inclusion. As a result, studies which implemented a SNP-based proxy/tagging method (as described in Chapter 1.3.4) were excluded. These SNP tags which are found outside of the relevant gene, elsewhere on the chromosome, have been shown to be limited when looking at different populations because the method relies on the tagged SNP to be in perfect linkage disequilibrium with the allele of interest. For example, hypersensitivity to abacavir has been linked to carriage of the HLA-B*57:01 allele. [15, 16, 105-110] A study conducted using a Mexican population reported that the SNP rs2395029 was in absolute linkage with HLA-B*57:01. [111] However, this is not observed across all populations, a study which investigated the rs2395029 SNP in an Italian population reported that the SNP and carriage of the HLA-B*57:01 allele were not in absolute linkage ($r^2 = 0.89$) and recommended against using SNP tag alone as a proxy marker

for HLA-B*57:01. [112, 113] A study by de Bakker *et al*. investigated SNP tags for HLA alleles in 4 different populations: Yoruba people from Ibadan, Nigeria (YRI); Utah residents of European ancestry (CEU); Han Chinese individuals from Beijing, China (CHB); and Japanese residents from Tokyo, Japan (JPT). [81] The study reported that different SNP tags would be required for the same allele in different populations. For example; the tagged SNP for HLA-A*31:01 (for carbamazepine hypersensitivity) would require: rs1061235 in Caucasians; rs3823318 & rs1061235 in Chinese and rs1061235 & rs1150739 in Japanese populations. The multiple tag SNPs required to represent a single allele would limit the clinical relevance and would also restrict the ability to perform meta-analysis with the data when comparing different populations. Therefore, studies which typed patients using tagging SNPs were excluded.

2.2.4. Data extraction

Data were extracted from the studies brought forward by the literature search of the systematic review. This was performed by 3 reviewers, with GG extracting data from 50% of the papers identified for extraction and the remaining studies being mined for data by MHM (25%) and AS (25%). This was then checked by 2 reviewers, GG independently verified the data extracted by MHM and AS (GG did not verify own data extraction) whilst ES verified the data extraction performed by GG. A list of the types of data that were extracted from the selected studies are shown in Table 2.3Table 2.3.

Data were extracted to a spreadsheet document file where as much of the data was recorded as it appeared in the original study. Certain circumstances required interpreting the data in order to facilitate meta-analysis. For example, ADRs may use alternative terms (refer to section 0), these terms were standardised based on expert medical knowledge of Dr Ana Alfirevic and is described in Table 1.3 . Additionally, certain papers would report a significance of association (p-value) simply as "significant" or "not-significant". In these cases the studies p-value threshold was used and the value was reported relative to the significance threshold (for statistically significant results after and correcting for multiple comparisons; "0.05" with a note stating it is less than the threshold or if the association is not-statistically significant, then the column was left blank and a note was included to explain this). As the

eligibility criteria required the number of cases and controls, the type of controls were divided into two distinct groups (recorded separately within the spreadsheet document) based on whether the controls were patients also exposed to the drug as part of their treatment or if these were population controls which were not patients being treated (i.e. not exposed to the drug) for disease/condition that the patient group were.

Table 2.3: Types of data that were extracted from the studies identified during the literature search of the systematic review.

Study details	Bibliographic information
	PubMed ID – to provide link back to original manuscript
	Corresponding authors and email
	Disease (that patients were initially treated for)
	Drug administered
Study cohort demographic	Country
	City & province/state
	Group sizes (patients, controls)
	Ethnicity
Association data	Allele
	Induced phenotypes (ADRs)
	Frequency of allele within each sample group
	Odds ratio (with confidence intervals)
	P-values
	Statistical method used for analysis (including corrections for multiple comparisons)

2.2.5. Semi-automated literature mining

In order to speed up the process for future data collection, a computer script written in the Python scripting language was developed. The script utilised the e-utils module in order for the computer to interact with the Medline/PubMed database. The searches still required manual input to perform, and just repeated the procedure described in

2.2.1 Literature search (for Medline) with the dates adjusted to capture studies with a publication date starting from the point of the previous search to the time of the search. A copy of the script can be found in the accompanied CD / USB flash drive under the chapter 2 folder as "semi_automated_literature_miner.py" (can be opened in notepad if the reader does not have a specialised program for viewing script files). The file has been modified to include the date range for the last performed search (August 2015 to October 2017). This can be found on lines 73-78 of the script.

The script was developed to automatically perform the first stage of screening (assessment of relevance for titles and abstracts). Using the list of unique PMIDs obtained from the search, the titles and abstracts were downloaded and scanned by the script to search for HLA alleles using regular expressions. The script was designed to compensate for the changes of the nomenclature system over the years i.e. the script understands the current nomenclature and also older formats of the nomenclature (as described in Chapter 1.2.5). If the title and/or abstract contained a reference to a named HLA allele then the PMID is placed into a new list (in a file called "filtered PMIDs") as potential studies to be considered during the second stage of screening. The script also searched for terms relating to ADRs – a list of these terms is found in Table 2.4Table 2.4 below. Additionally, many studies are present in Medline however, only contain the title and lack the abstract. In such cases the script scans the title and places any PMIDs for studies which contain references to HLA alleles and ADRs into a "forwarded titles" file, and rejected studies lacking an abstract were placed into a "rejected titles" list. Studies which contained a title and abstract and that were rejected by the script were placed into a "rejected" list. The second stage of screening and the data extraction was performed manually in the same manner as described previously. This manual review during the second stage of screening was required as the script is not able to understand the context of the title/abstract in the same way a human with the relevant expertise could achieve.

Table 2.4: List of ADR terms that were searched for by the computer script designed to perform the semi-automated literature search.

Search term
cutaneous adverse drug reaction
cADR
skin rash
mild maculopapular eruption
MPE
drug rash with eosinophilia and systemic symptoms
DRESS
Stevens-Johnson syndrome
Stevens-Johnsons syndrome
SJS
toxic epidermal necrolysis
TEN
severe cutaneous adverse drug reaction
SCAR
erythema multiforme major
EMM
drug-induced hypersensitivity syndrome
DIHS
drug-induced
erythema exudativum multiforme
EEM
exanthema
hepatotoxicity
drug induced skin injury
DISI
drug induced liver injury
DILI
drug induced kidney injury
DIKI
immune mediated
drug toxicity
adverse drug reaction
drug reactions
drug hypersensitivity
drug eruption

2.2.6. Search strategy

2.2.6.1. MEDLINE

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) <1946 to Present>

Search Strategy:

---

1 (drug induced or drug-induced).ti,ab. (27831)

2 (adverse drug reaction$ or adverse reaction$).ti,ab. (30786)

3 exp Drug Hypersensitivity Syndrome/ or exp Drug Hypersensitivity/ (37955)

4 hypersensitivit$.ti,ab. (48656)

5 exp Pharmacogenetics/ (9386)

6 exp HLA Antigens/ (64053)

7 (human leukocyte antigen$ or HLA).ti,ab. (81597)

8 (pharmacogenetic$ or genetic$).ti,ab. (670906)

9 (Abacavir or Allopurinol or Aminopenicillin or Aspirin or Carbamazepine or Clozapine or Co-amoxiclav or Co-trimoxazole or d-Penicillamine or Diclofenac or Feprazone or Flucloxacillin or Gold Sodium or Thiomalate or Hydralazine or Lapatinib or Levamisole or Lumiracoxib or Methazolamide or Nevirapine or Oxcarbazepine or Oxicam or Phenytoin or Sulphamethoxazole or Ticlopidine or Trichloroethylene or Ximelagatran).af. (146055)

10 or/1-4 (133860)

11 or/5-8 (758141)

12 10 and 11 (5967)

13 9 and 12 (772)

14 (animals not (humans and animals)).sh. (3860455)

15 13 not 14 (759)

16 limit 15 to english language (696)

****

2.2.6.2. EMBASE

Database: Embase <1974 to 2014 June 20>

Search Strategy:

---

1 (drug induced or drug-induced).ti,ab. (35493)

2 (adverse drug reaction$ or adverse reaction$).ti,ab. (44767)

3 exp drug hypersensitivity/ (49292)

4 hypersensitivit$.ti,ab. (59577)

5 exp pharmacogenetics/ (21416)

6 exp HLA antigen/ (80485)

7 (human leukocyte antigen$ or HLA).ti,ab. (103869)

8 (pharmacogenetic$ or genetic$).ti,ab. (781699)

9 (Abacavir or Allopurinol or Aminopenicillin or Aspirin or Carbamazepine or Clozapine or Co-amoxiclav or Co-trimoxazole or d-Penicillamine or Diclofenac or Feprazone or Flucloxacillin or Gold Sodium or Thiomalate or Hydralazine or Lapatinib or Levamisole or Lumiracoxib or Methazolamide or Nevirapine or Oxcarbazepine or Oxicam or Phenytoin or Sulphamethoxazole or Ticlopidine or Trichloroethylene or Ximelagatran).af. (336136)

10 or/1-4 (174284)

11 or/5-8 (902379)

12 10 and 11 (8682)

13 9 and 12 (1899)

14 limit 13 to (human and english language) (1596)

****

2.2.6.3. List of keywords used in titles scan (for PubMed entries without abstracts) performed by computer script

- cutaneous adverse drug reaction / cADR
- skin rash
- mild maculopapular eruption / MPE
- drug rash with eosinophilia and systemic symptoms / DRESS
- Stevens-Johnson syndrome / Stevens-Johnsons syndrome / SJS
- toxic epidermal necrolysis / TEN
- severe cutaneous adverse drug reaction / SCAR
- erythema multiforme major / EMM
- drug-induced hypersensitivity syndrome / DIHS
- drug-induced
- drug induced skin injury / DISI
- drug induced liver injury / DILI
- drug induced kidney injury / DIKI
- erythema exudativum multiforme / EEM
- exanthema
- hepatotoxicity
- immune mediated
- drug toxicity
- adverse drug reaction
- drug reactions
- drug hypersensitivity
- drug eruption

2.3. Results

2.3.1. Data

The PRISMA flow diagram (Figure 2.1) shows the number of studies that were identified along with the source during each stage of the systematic review and literature search performed by the semi-automated literature search computer script. From the initial literature search (covering studies published between 2000-August 2014), a total of 7,978 studies from Medline, HuGE Navigator and EMBASE were identified after duplicates were removed. From the 1st stage of screening, where the titles and abstracts of the studies were examined, 236 records were deemed to be potentially relevant, meaning 7,742 studies were excluded. The 236 studies brought forward for the 2nd stage screening where their full manuscript was analysed to determine eligibility for data extraction. Of these, 74 studies were included for the systematic review (see thesis bibliography - Studies featured in Systematic Review) and the remaining 155 studies were excluded. The follow-up data collection performed in August 2015 utilised the semi-automated literature mining computer script. The literature search, covering studies published between April 2014 and August 2015, yielded 4,969 articles for analysis. The script classified 64 studies as being potentially relevant where upon manual analysis of the full manuscripts 11 studies were included. Therefore, a total of 85 studies were included for data extraction (74 from the initial systematic review and 11 from the semi-automated literature mining script).

All 85 included articles were case-control studies (there were no randomised control trials) and reported ADRs with a total of 30 drug therapies (25 of which were named drugs), with carbamazepine being the most frequently represented (see Table 2.5). There were 1,009 ADR-allele comparisons, of which 190 were associations that showed statistical significance, which equates to 73 unique associations (compensating for that fact that, for example, there are several independent studies that reported that HLA-B*15:02 is associated with carbamazepine hypersensitivity). The associations covered 259 distinct alleles (214 high resolution alleles + 45 serotype/antigens) which represents 7 of the HLA family genes.

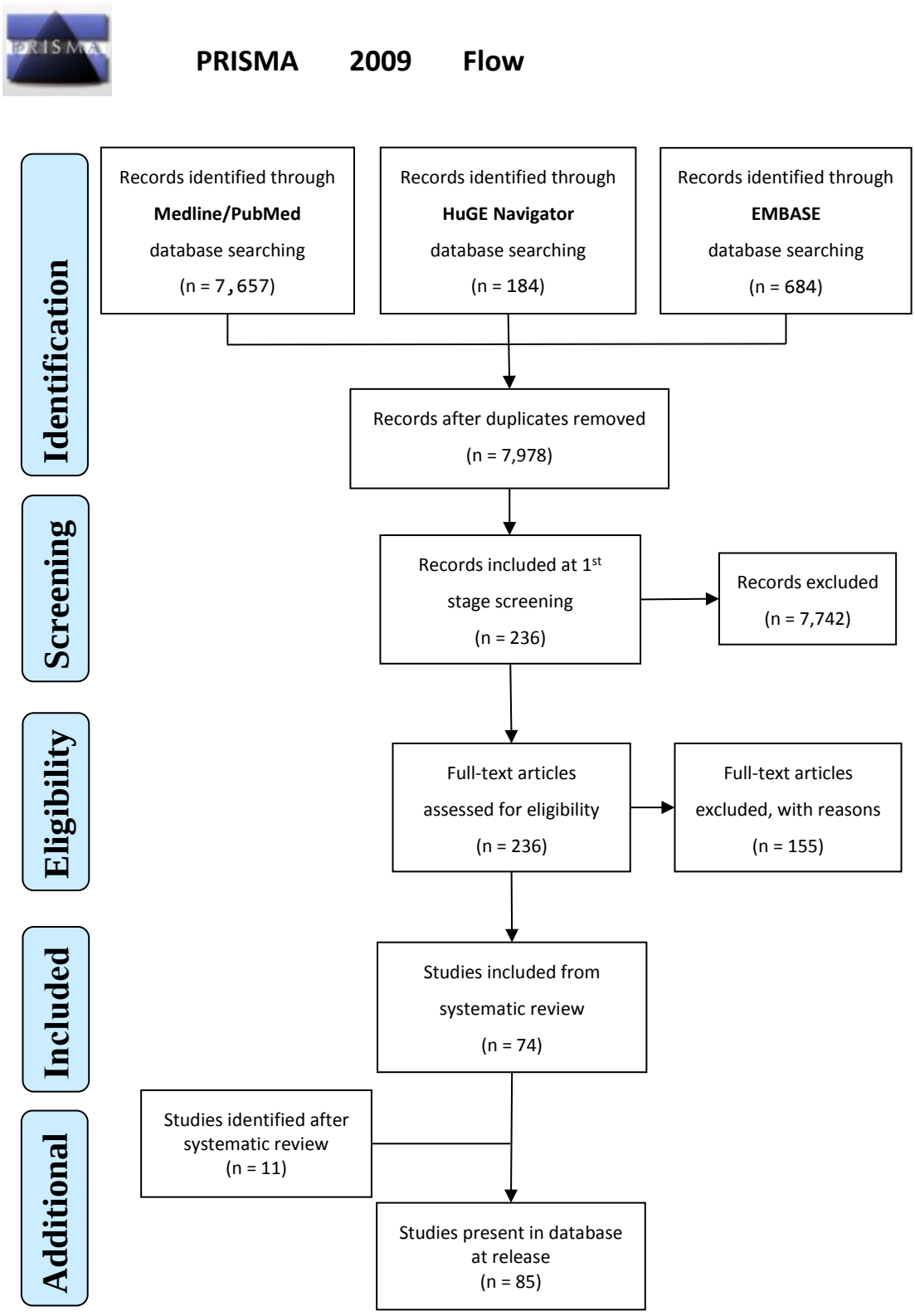


Figure 2.1: PRISMA flow diagram showing the number of studies being considered at each stage of the systematic review along with the additional studies identified via the semi-automated literature mining script.

Table 2.5: List of drugs that are in current use captured from the systematic review and semi-automated literature search as well as the number of studies which feature the drug. Some studies investigated more than one drug so there is some overlap.

Drugs	Number of studies identified from the systematic review featuring drug
abacavir	8
allopurinol	10
amoxicillin-clavulanate	2
aspirin	3
bucillamine	1
carbamazepine	29
clozapine	2
co-trimoxazole	1
dapsone	1
flucloxacillin	1
lamotrigine	7
lapatinib	2
lumiracoxib	1
methazolamide	1
nevirapine	4
oxcarbazepine	3
paracetamol	1
phenobarbital	2
phenytoin	4
sulfamethoxazole	2
sulfasalazine	1
ticlopidine	1
valproic acid	1
[1]antiepileptic drugs	2
[2]antituberculosis drugs	1
[3]EGF receptor inhibitors	1
[4]NSAID and 'multi-ingredient cold medication'	2
[5]oxicam NSAIDs	1
[6]statins	1

Notes:

Entries with superscript numbers represent studies where the statistics relating to the study participants were not separated by drug. Below are the list of the drugs that were featured in the studies.

1 - Carbamazepine, lamotrigine, oxcarbazepine, phenytoin, phenobarbital, topiramate and valproate.

2 - Combination regimen including isoniazid, rifampin, ethambutol, and pyrazinamide.

3 - cetuximab, panitumumab, erlotinib and gefitinib

4 - acetaminophen, analgin, aspirin, brufen, diclofenac, diclofenac & paracetamol, dipyrone, dipyrone & diclofenac, fenton (cold medicine), ibuprofen, mefenamic acid, nimesimide, nimesulide, paracetamol, piroxicam, tylenol (cold medicine), cold remedies (unknown), medicine for cold (NSAID), medicine for cold (detail unknown) and NSAIDs (drug name unknown)

5 - Not specified beyond "oxicam-NSAIDs"

6 - Not specified beyond statins.

2.4. Discussion

2.4.1. Analysis of data

The data collected from the systematic review suggests that a greater number of alleles that have been associated with ADRs are likely to be found in people of Asian descent (discussed in greater detail in Chapter 4). One potential explanation is the possibility of case-control study bias. Approximately 65% (56 out of 85) of the included studies were conducted in Asia and 20% (17 of 85) were conducted in Europe. Therefore, it may be possible that the high number of observed ADR-allele associations found in patients of Asian origin are likely to be a result of this case-control study bias. Furthermore, it is also worth noting that the drugs carbamazepine, lamotrigine, allopurinol, oxcarbazepine and phenytoin are all used in the treatment of epilepsy, additionally, phenytoin, carbamazepine and oxcarbazepine are antiepileptic drugs with aromatic structures that have ADRs associated with HLA-B*15:02. It is likely a similar biochemical mechanism could be involved with these ADRs. These similar drugs with similar mechanisms could be artificially increasing the count in the results,

making Asians seem more susceptible to ADRs compared to other population groups. This may be further compounded by the relative lack of data available from the other regions, particularly, South America (1 out of 85) and Africa (2 out of 85) which further perpetuates the impression that Asians are more susceptible to ADRs.

Another potential reason could be that the drugs were tested in a European population during the drug's clinical trials. Many of the pharmaceutical companies are based in Europe or North America where the population (and therefore the trial participants) is largely of European descent. It is likely that the drug clinical trials were conducted in the host nations, meaning that any ADRs (although rare) that were observed during the trials or shortly after release, during a period of active drug-safety vigilance, may have led to drug withdrawal. However, this hypothesis is difficult to verify as information regarding the location of the original drug trials is difficult to acquire, particularly for drugs that have been on the market for several years/decades.

Another issue faced with the data collected from this systematic review was that there was inconsistency in the data reporting. Poor replication with some alleles that are reported to be associated in some studies may be reported as not-significant in others. Typically, this was due to these studies using relatively small patient-control cohorts so information regarding individuals had a relatively larger impact on the final results. This issue is very difficult to resolve as ADRs are very rare, therefore the studies were constrained by the number of patients available to them, thereby limiting the statistical power of the study. The file generated in Chapter 6 highlights the rarity of ADRs by listing examples of ADR prevalence for patients receiving each of the included drugs (see USB/CD media "chapter_6_decision_support_table.xlsx").

It was also noted during the analysis of the literature that in terms of study designs (see Table 2.2), retrospective studies made up the majority of the studies collected. These retrospective case-control study designs are suitable for rare outcomes (as is the case for immune-mediated ADRs) as the cohort sizes are usually smaller compared to prospective cohort studies. However, the collection of data for retrospective studies are observational (occurs afterwards by examining medical records in the case of immune-mediated ADRs) meaning that it can sometimes be difficult to go back and collect additional data. This limitation introduces issues with data quality. However,

retrospective studies are often the only option as prospective cohort studies require recruitment of large numbers of patients in order to determine rare ADRs (an incidence rate of 1 in 1,000 to 1 in 10,000 individuals exposed to culprit medications experience ADRs) and are often prohibitively expensive to carry out. This is a likely explanation as to why a large proportion of the studies brought forward in the literature review were of a retrospective design.

2.4.2. Script validation

To assess the performance of the script, the resultant lists ("filtered PMIDs", "forwarded titles" and "rejected") were analysed. The first two of the analyses were performed using two previously conducted literature reviews. [98, 114] In the first analysis, the studies in each of these reviews were analysed to see if they were determined to be relevant by the script. One of the included studies in each of the reviews was not part of the validation test as it featured unpublished data (the reviewers had contacted the original authors of the studies to obtain the data). In the first review (Yip *et al*., 2012), all but one of the 21 studies contained an HLA reference in the title/abstract. The PMIDs of these studies were therefore found in the "filtered PMIDs" list. [98] The remaining paper did not feature an abstract but did contain a key-phrase and so was placed into the "forwarded titles" list. With regards to the other review (Cornejo Castro, 2014), 9 of the 12 included studies were found in the "filtered PMIDs" list and 1 paper was found in the "forwarded titles" file (without abstract but contained a reference to an HLA allele in the title). [114] The remaining two files were rejected by the script. Upon further analysis of the full articles, these two papers would have been rejected at the 2nd stage screening since the HLA alleles were only typed to serotype level resolution (often referred to as 2 digit resolution). This would be considered too low (by the criteria of the protocol) and would not be included for this review.

For the second validation test, a sample of 200 PMIDs from the "rejected" list were analysed manually to determine if they were indeed studies that would be rejected under the conditions of this literature review protocol by the reviewer (i.e. true negatives). Upon manual review, all 200 studies were determined to be those that

would not be included in this review and therefore, were correctly placed into the "rejected" list.

After the systematic review was performed, a third validation test was conducted using the 74 studies that were included in the systematic review. These 74 studies were used as a reference to compare how the semi-automated script compared to the manually performed review – this assumes that the manually performed systematic review was free from human error, although this is the current gold standard when performing systematic reviews. Only the 74 original studies were included for the comparison as the additional 11 studies were identified using the script so performing the analysis with the full list of the 85 studies (specifically, with the additional 11 articles) would be redundant. Of the 74 studies found manually, 61 were placed into the semi-automated script's "filtered PMIDs" list, 6 were placed into the "forwarded titles" list, 4 were found in the "rejected titles" list and the remaining 3 studies were rejected by the script. This means that 91% (67 out of 74) of the studies were determined to be potentially relevant by both the manual method and by the script. This assumes that the same 67 studies would be included upon manual review of the "filtered PMIDs" and "forwarded titles" lists.

Of the 4 studies that the script classified as "rejected titles" (i.e. the script could only scan the title of the studies), two studies investigated aspirin-intolerant asthma which was not part of the list of ADRs that were being scanned for by the script (refer back to Table 2.4Table 2.4). Adjusting the script to include aspirin-intolerant asthma (AIA) and the related aspirin-induced urticaria (AIU) would move one of the studies into the "forwarded titles" list as the script requires both a reference to an HLA allele (in this case the title includes "DPB1*0301") and an ADR phenotype (that the script recognises). [26] The second study would still be rejected as the title does not include a reference to the associated HLA allele. [115] Similarly, the other two studies [116, 117]that were rejected by the script (through a title only scan) also did not contain references to HLA alleles. [116, 117] Adjusting the script to compensate for this would greatly increase the number of studies that the script would place into the "forwarded titles" list and would require the user to manually review these studies during the 2nd stage manual review. This would offer little benefit compared to performing an entirely manual 1st stage screening of the literature.

The three remaining studies, which include a Medline/PubMed entry title and an abstract that the script can scan and that were rejected by the script, also faced similar issues. [34, 118, 119] The first of these studies (Mammen *et al*., 2012) investigated anti-hydroxymethylglutaryl-coenzyme A reductase induced myopathy that was associated with HLA-DRB1*11:01. [119] The script did not include "myopathy" as a recognised ADR phenotype (Table 2.4) and so was rejected (there were no other references to a generalised term for ADRs – e.g. "adverse drug reaction" or "ADR" etc.), this can be compensated for by including myopathy in the script. The other 2 studies did not refer to the HLA allele using a standardised reference to the HLA nomenclature due to the way they handled the asterisk symbol. Amstutz *et al*., 2013 [34] omitted the asterisk and simply used a space ("HLA-A 31:01") whilst Kindmark *et al*., 2008 [118] placed the asterisk between brackets, such as "DRB1(*)07", although this study would still have been rejected as the title and abstract implies that the study only typed the patients and controls' HLA-DRB1 gene to antigen level resolution (2-digits). Incidentally, during the manual systematic review, one of the reviewers (GG) initially rejected this study during the first stage review (analysis of the title and abstract) for the lack of allele level resolution with regards to patient/control typing. The second reviewer (AA) was already familiar with this study and recommended reassessment, the first reviewer evaluated the full manuscript and agreed for the study to be included as a small part of the investigation did involve typing the patients to allelic level (4-digit) resolution. Therefore it is worth recognising that human error is also an issue during the review process.

2.4.3. Discussion of script validation

Overall, the manually performed first stage systematic review took approximately four weeks to assess the titles and abstracts of 7,978 to identify 236 for analysis in the 2nd stage of the review. In comparison, there were 459 studies in the "filtered PMIDs" list and 344 studies in the "forwarded titles" list. This is a total of 803 studies which the script identified as potentially relevant and took approximately 1 hour to compile against the same 7,978 titles and abstracts. Potentially, 93% or 69 out of 74 studies (67 studies that were included by the script and an additional 2 studies that were rejected by the script but could be compensated for) can be identified using this script.

Whilst not perfect, it does offer >90% accuracy in identifying the relevant studies in a fraction of time allowing the user to conduct a literature review comparatively quicker by automating the first stage of the literature screening, and would be useful in scenarios where the time is a more crucial factor than absolute accuracy.

The accuracy is greatly improved when used to investigate individual ADRs, the ability of the script to correctly identify all the studies that were part of literature reviews focusing on individual drugs – for the first validation test: Yip *et al*., 2012 and Cornejo Castro, 2014 focused on carbamazepine and nevirapine hypersensitivity reactions respectively. In these cases the accuracy was greater as ADRs involving these drugs are fairly well studied so the script could be programmed with all the alternative names for the ADRs allowing for a more comprehensive search. This means that if the user knows that they are looking for then it can be argued that the script functions very well in its task. When the script was utilised for a generalised search of all ADRs, the script could not take into account ADRs that the user is unaware of (as was the case with anti-hydroxymethylglutaryl-coenzyme A reductase induced myopathy). This remains the main limiting factor for using any keyword based automated search.

It would therefore be recommended that the automated script be executed regularly (e.g. every 1-2 years) and after a few cycles (suggested as every 5 years) that a manual review be performed in parallel and therefore allows any studies missed by the semi-automated script a second chance to be picked up for potential inclusion and also allow for additional code flaws to be picked up and amended into the code.

2.4.4. Effectiveness of review

Interpretation of the data collected from the systematic review largely follows what is already known in the literature a few high-profile examples include: HLA-B*57:01 and abacavir hypersensitivity; HLA-A*31:01 and HLA-B*15:02 for carbamazepine hypersensitivity etc. However, there is some loss of granularity in the information, namely with regards to allelic linkage and haplotypes. For example, the case of carbamazepine hypersensitivity, 29 studies were identified that investigated ADRs in association with this drug and from these, 22 HLA alleles were reported to be

significantly associated (i.e. a p-value < 0.05) when statistically analysed. However, the HLA-A*31:01 and HLA-B*15:02 alleles were reported in 11 and 19 studies respectively (with some overlap) whereas the remaining alleles were reported to be significant in 3 or fewer studies. Many of these 20 alleles, when investigated, were found to be in linkage with either HLA-A*31:01 and HLA-B*15:02 indicating that these other 20 alleles are likely found to be statistically significant due to their co-occurring relationship with HLA-A*31:01 or HLA-B*15:02. For carbamazepine, this information was readily provided in the studies reporting these alleles and the contextual information was lost in this review when collecting the raw data from these studies. Therefore, careful interpretation of the collected data is required.

It is very likely that this review was able to identify the majority of available data given that all of the studies present in two previous literature reviews (regarding carbamazepine and nevirapine) were identified in this review. The high concordance with the two prior literature reviews provides a reasonable degree of confidence that all of the relevant studies were identified which in turn provides confidence that a suitable amount of data was collected to allow the review question to be adequately answered. Therefore, there is reasonable assurance to further utilise the data in other aspects, such as Chapter 3 – "The HLA-ADR web-database: a centralised location to store and provide data relating to immune-mediated adverse drug reactions".

2.4.5. Potential improvements

A key area for potential improvement is the ability for the semi-automated literature mining script to overcome the limitation of keyword searches that limit its ability to pick up associations with previously unreported drugs and clinical phenotypes. One method would be to review the titles and abstracts of interest (the studies that were included in the literature review) and analyse the language, beyond the specific references to drugs, HLA alleles and ADR phenotypes used within the text to determine if there is a common pattern with the included studies that could be used as a marker to weight the title/abstract it in favour of inclusion. This methodology has the potential to allow the identification of published literature which features HLA mediated ADRs where the user has not entered specific search terms thereby possibly overcoming the limitations of keyword based searches.

Chapter 3. The HLA-ADR web-database: a centralised location to store and provide data relating to immune-mediated adverse drug reactions

3.1. Introduction

The previous chapter describes that the field of immune mediated ADRs is growing rapidly with new data released at frequent intervals. It also shows that due to the high growth rate of the field, if an investigation regarding ADRs needs to be conducted, then the researcher/research group would need to perform an in-depth literature search or systematic review for each study. This process is highly inefficient if the investigations are exploring similar questions with significant overlap, as multiple groups could independently conduct the same search. It is therefore important that a centralised database be constructed where users can view relevant studies from which to initiate their investigation, and enabling dynamic linkage to other accessible information about drugs and immune genes/alleles.

This chapter describes the creation of such a database built upon the information collected in Chapter 2 regarding ADRs that have been associated with the carriage of particular HLA alleles. As this database is built using HLA-ADR data and that the website that overlays the database utilises tools from the Allele Frequency Net Database (AFND), http://www.allelefrequencies.net/ it was decided to construct the web database as a module within AFND. [120, 121]

3.1.1. AFND Website

The Allele Frequency Net Database (AFND) was initially developed to store allele, haplotype and genotype frequencies of several immune-related genes including HLA, Killer-cell immunoglobulin-like receptors (KIR), cytokines and Human MHC class I chain related (MIC) genes from diverse worldwide populations although the main focus is on genes. [120, 121] The population data sets are comprised of healthy, unrelated individuals (for example, bone marrow registries). As of March 2017, there are HLA allele, haplotype and genotype data from 1,086 populations present in the AFND representing 10,447,550 individuals. The website is built upon a Microsoft SQL

Server with active server pages (ASP classic) delivering the webpages, the various database and web development languages are described in Chapter 0. New data from researchers can be added via the web-based framework that provides an online submission system which is then manually curated by a process of data validation to enforce strict accuracy with the results.

3.2. Design of the HLA-ADR database

3.2.1. Database implementation

Since the aim of the website is to create a centralised location where the data collected from the systematic review can be retrieved, a web database was created. Firstly, a relational database was constructed in order to store the information. Although many relational database management systems (RDBMS) are available (as described in Chapter 1.4.3.1), the HLA-ADR database utilises Microsoft SQL Server as its RDBMS. SQL Server is the system that was implemented with the main AFND website and therefore, chosen in order to ease integration into the AFND. This setup also allows for future interaction/crossover and joint maintenance with the main database underlying AFND.

3.2.2. Database schema

The schema for HLA-ADR database is shown in Figure 3.1 below which shows the individual tables as boxes where the connecting lines demonstrate the relationships that are enforced by the RDBMS. The database tables have been normalised following the guidelines of normalisation as laid out in section 0. However, this was not done to fully conform to 3NF as there would be diminished returns with regards to the potential benefits that normalising the database would bring. For example, referring to Figure 3.1 the HLA-ADR database table "tbl_association_data", the information for the ADR frequency in patients (stored in the column "patient_adr_frequency"), this information is dependent on the key (1NF) and the whole key (2NF) but it is not dependent on the other fields for its existence, e.g. the type of statistical correction only applies to the p-value, which therefore violates the requirements for 3NF. However, the information only retains contextual meaning along with the other information in the table therefore, it still makes sense to keep these information

together as although it is violating the strict sense of 3NF, it is in contextual 3NF. If the table was decomposed to separate these fields so that it does conform to the strict definition of 3NF, this would require additionally complex joins with the query. It was therefore decided that contextual 3NF would be the optimal structure for this database schema. The rest of the database follows this contextual 3NF principle

This contextual 3NF can be seen in Figure 3.1; where each box represents a table within the database with the table name in the first row. The first column displays the column names for the data contained in the table with the second column presenting the data type e.g. “int” refers to integer values; “nvarchar(45)” means that the field contains text (character limit within the brackets); “float” refers to floating point number (non-integer). Connecting lines represent relationships. In this schema, all relationships are one-to-many (infinity symbol) to implement database normalisation as described in 1.4.2 Database table normalisation. For example, for the relationship between tbl_drugs and tbl_study_details means that one record entry within the tbl_drugs table can be referenced by many entries within the tbl_study_details table.

3.2.3. Data curation

The initial data set was collected in the systematic review as described in Chapter 2. However, the database itself is designed to accept new data as it is received and will expand as more HLA-ADR association studies are published. Updates to the database can be performed via Python scripts which were developed with the primary aim to speed up and automate the process of adding additional and new data into the database. [104] Additionally, very little training or programming knowledge is actually required to run these finalised scripts and therefore, the scripts can be run to update the database by curators without knowledge of coding.

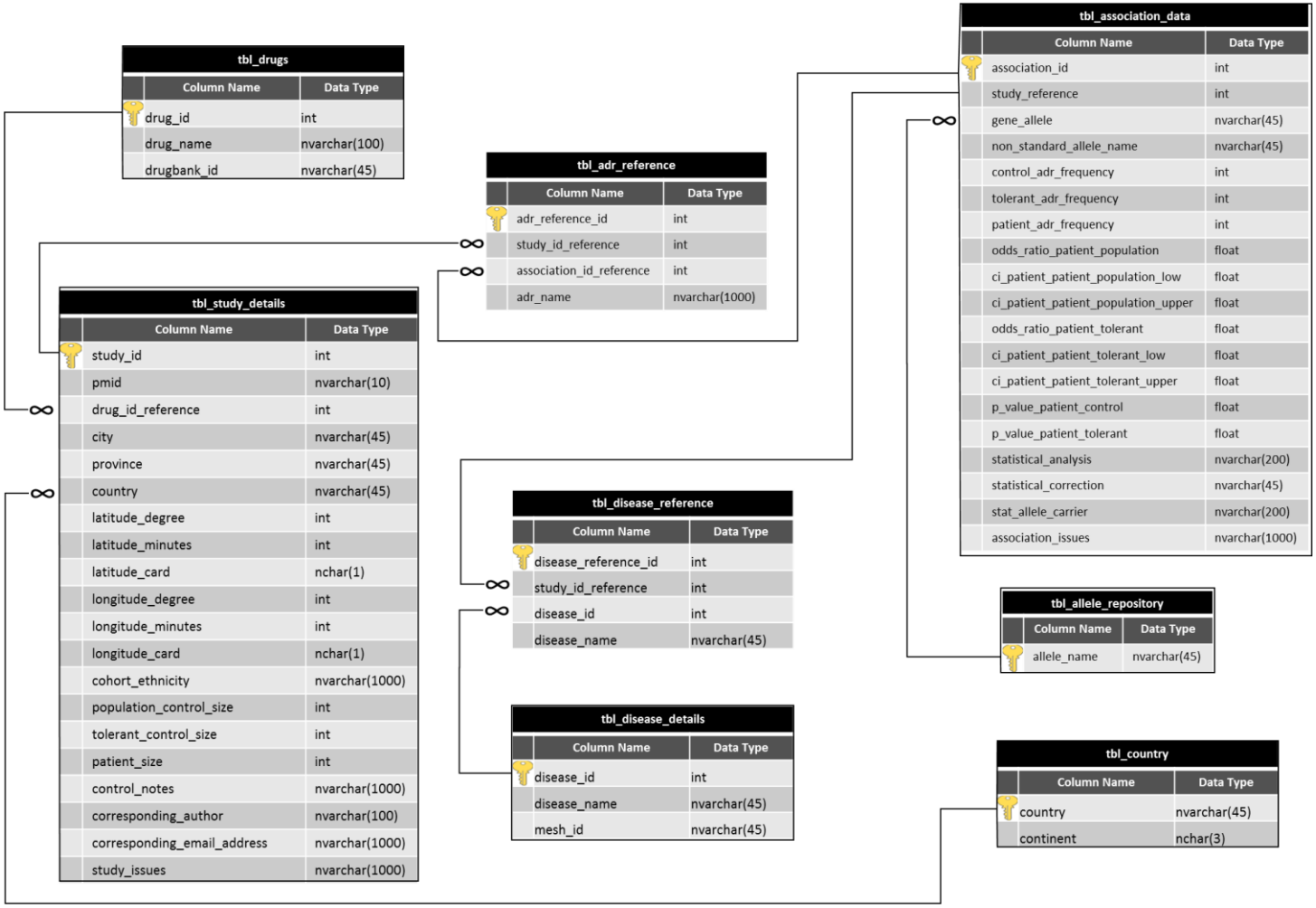


Figure 3.1: Database Schema for the HLA-ADR database.

The update scripts work by reading in the data from the spreadsheet document that is used to store and record the data extracted from the systematic review as described in Chapter 2.2.4 and converts the information to generate the SQL statements that are sent to the server hosting the database. These SQL statements provide instructions to add the new data entries. The scripts can be found in the accompanying USB/CD media for this thesis in the Chapter 3 folder. The first file, "HLA-ADR_create_source_files.py" reformats the data so that the spreadsheet document is converted into a series of text files that is in the same format that as the database schema. The second script "HLA-ADR_update_database.py" will read these text files and insert the information into the database.

3.2.4. Website organisation

Webpages were designed to allow users to easily access and visualise the HLA-ADR data in a user-friendly manner. These webpages were integrated into the main AFND site with the HLA-ADR database directly available with the URL: http://www.allelefrequencies.net/hla-adr which links to the HLA-ADR home page. All of the webpages are generated using Microsoft's classic Active Server Pages (ASP) with VBScript (based on Visual Basic) as the underlying language, the same as the main AFND webserver. The webserver is able to serve out HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) thereby allowing the retrieved data to be viewed in most common web browsers.

3.3. Results

3.3.1.1. Homepage

The home page acts as a gateway for the HLA-ADR website and allows users to access the other pages. The homepage can be accessed using the URL: http://www.allelefrequencies.net/hla-adr/default.asp (screenshot of the webpage is shown below in Figure 3.2). Users are presented with a brief introduction describing the purpose of the HLA-ADR database as well as hyperlinks to the Query page, Authors submission page, ADR report page and the bibliography page which are

detailed below. The users are also presented with basic information pertaining to the underlying data, including the unique number of: a) HLA alleles in the database, b) drugs which have been investigated for ADRs, c) the total number of association records (all reported alleles, both significant and non-significant associations across all studies) and d) the total number of studies from which the database is built upon. The information pertaining to this underlying data is acquired by directly querying the database meaning that the information is always kept up to date.

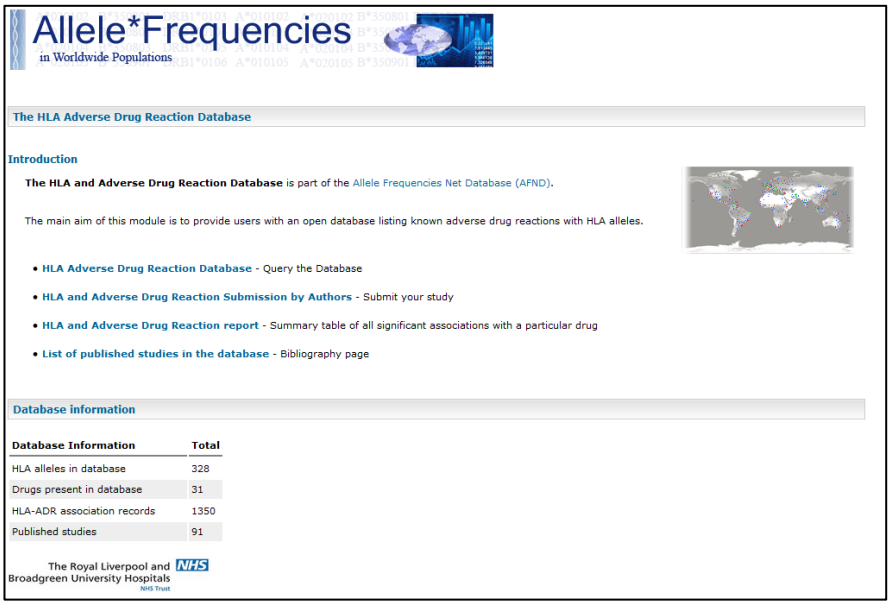


Figure 3.2: Screenshot showing the HLA-ADR website's homepage.

Here the users are presenting with a brief introduction for the HLA-ADR database, links to the other webpages that are a part of HLA-ADR and also some information pertaining to the data stored in the underlying database.

3.3.1.2. Query page

The query page is the primary method by which users can access the information stored in the HLA-ADR database. The user is able to choose and limit the type of data they receive from the search by using the dropdown filters as shown in Figure 3.3 below. These dropdown filters are divided into three sets: in the first set, the user has the option to choose a HLA gene, a specific allele or from what the site refers to as "non-standard" alleles (e.g. serotype or antigens). These dropdown options within the first set are mutually exclusive and therefore only one of these options can be chosen. If a user selects from two or more options, then an error message returns indicating

that there is a conflict with the chosen options and instructs the user to select a maximum of one option. However, options from the first set are not mutually exclusive with the other sets and can be selected in combination with options from the second and third sets.

The second set of options provide the user to select additional parameters including the specific drug investigated in the study, patient-control cohort ethnicity, strength of the association (based on the p-value), the country/geographic region where the study was conducted and the condition for which the patients were being treated (e.g. epilepsy). Unlike with the first set, the options within the second set are not mutually exclusive and can be used in any combination with each other (except when the country and geographic region options conflict). The third set only contains one dropdown bar where the user may select the order in which the results are displayed, using one of 3 options: the user can specify that the results from the same study, drug or allele are grouped together when the website returns the results of the query. Only one of these three options can be selected, with the website defaulting to "study".

In instances where potential conflicts could occur, e.g. selecting for the "HLA-A" gene but the "HLA-B*15:02" allele or contradictory country and geographic region options, an error is provided which states the nature of the conflict, e.g. "Error, there is a conflict: You have selected values for country and geographic region which do not make sense. Please use only one of these parameters". This provides the user with a meaningful error message to help them formulate a query that can be executed. There are exceptions to this inbuilt error check feature however, there is no inbuilt check to prevent contradictory errors between the options selected for the drug and patient disease. This was intentionally left as the relationship between drug and disease can potentially be a many-to-many relationship i.e. one drug can be used to treat multiple diseases and additionally, one disease can be treated by multiple drugs (e.g. antiepileptic drugs). This many-to-many relationship limits the benefit for enforcing a check for contradictions and therefore, it has been left for the users to understand and deal with any such potential conflicts.

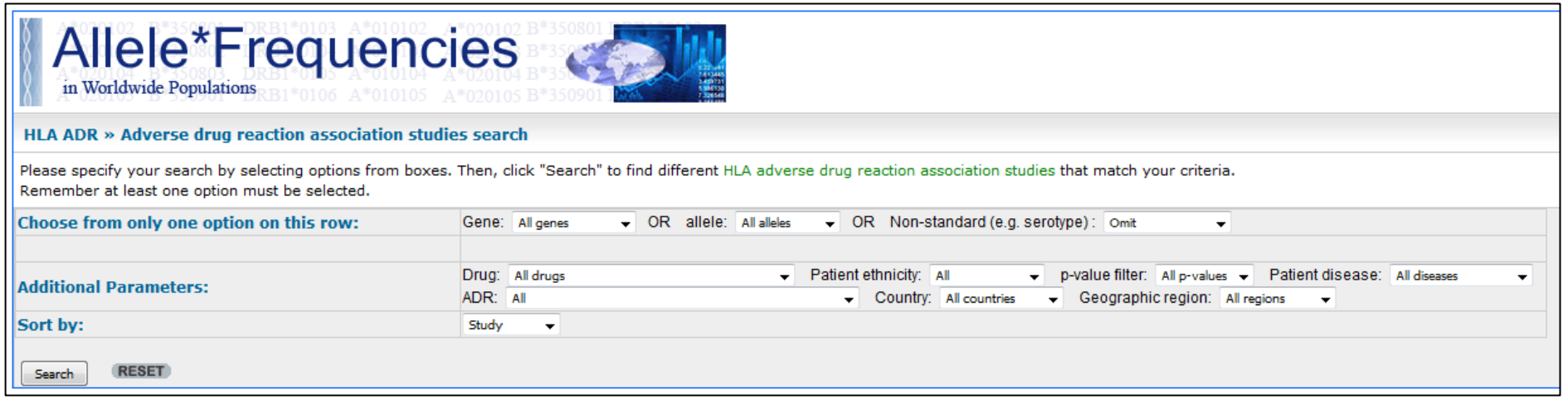
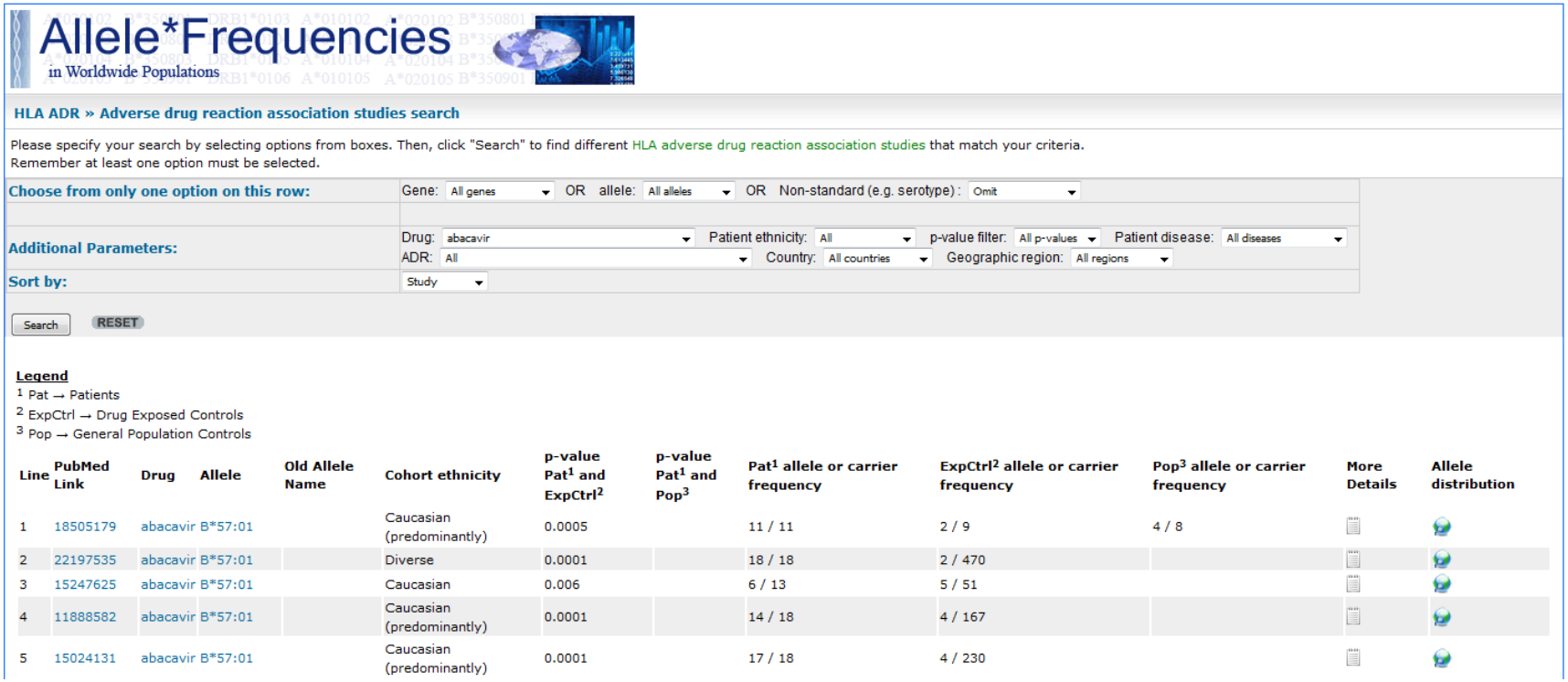


Figure 3.3: Screenshot showing the menu options on the HLA-ADR query page.

The options allow the user to select association records for particular genes/alleles, drugs, patient ethnicity, level of significance (p-value thresholds) a particular patient disease, the expressed ADR phenotype and/or the country or geographic region (continent) where the study was conducted. The user can also sort the results to group together results which are related by the reporting study, the drug inducing the ADR or grouping alleles together

Once the search options are chosen, the webserver utilises these options and generates the SQL query needed to retrieve the requested data from the HLA-ADR database, and then executes the query. The data that is retrieved from the query search is processed by the webserver to display the results in the table format that the webpage is designed to display, see example query search in Figure 3.4 which shows an example of a search using "abacavir" as the only criteria with all other options set to default (i.e. no specifications on HLA gene/allele, ethnicity, p-value thresholds, patient disease, adverse reaction phenotype or location of study). In the example search shown in Figure 3.4, the HLA-ADR database returns nine results which have been sourced from eight studies, all the studies have reported a HLA-B*57:01 association with one of these studies (Rauch et al., 2008) also reporting HLA-DRB1*07:01 to be statistically significant however, the investigation also reported that HLA-DRB1*07:01 was in high linkage with HLA-B*57:01. The returned results from the query shows information pertaining to each of the records as described in Table 3.1.

Not all of the information for the record stored in the database is displayed on the results of the query page. Displaying all of the record data would result in the user having to scroll along the horizontal axis in order to see all of the information. This was deemed to be a sub-optimal visual design, so the More Details page was created (see section 3.3.1.3 More Details page) where users can view the entire record, with only part of the entire record being displayed on the query page.



Line	PubMed Link	Drug	Allele	Old Allele Name	Cohort ethnicity	p-value Pat[1] and ExpCtrl[2]	p-value Pat[1] and Pop[3]	Pat[1] allele or carrier frequency	ExpCtrl[2] allele or carrier frequency	Pop[3] allele or carrier frequency	More Details	Allele distribution
1	18505179	abacavir	B*57:01		Caucasian (predominantly)	0.0005		11 / 11	2 / 9	4 / 8		
2	22197535	abacavir	B*57:01		Diverse	0.0001		18 / 18	2 / 470			
3	15247625	abacavir	B*57:01		Caucasian	0.006		6 / 13	5 / 51			
4	11888582	abacavir	B*57:01		Caucasian (predominantly)	0.0001		14 / 18	4 / 167			
5	15024131	abacavir	B*57:01		Caucasian (predominantly)	0.0001		17 / 18	4 / 230			

Figure 3.4: Screenshot of an example query performed on the HLA-ADR website query page. In this example, the only option chosen was to search for data pertaining to the drug abacavir. For all other options, the default settings were kept which does not apply any other restrictions, e.g. data for all gene/allele associations are displayed, no country restrictions are applied, non-significant and significant associations are displayed etc. The only subset of data that are omitted are those pertaining to non-standard alleles (i.e. serotype/antigens).

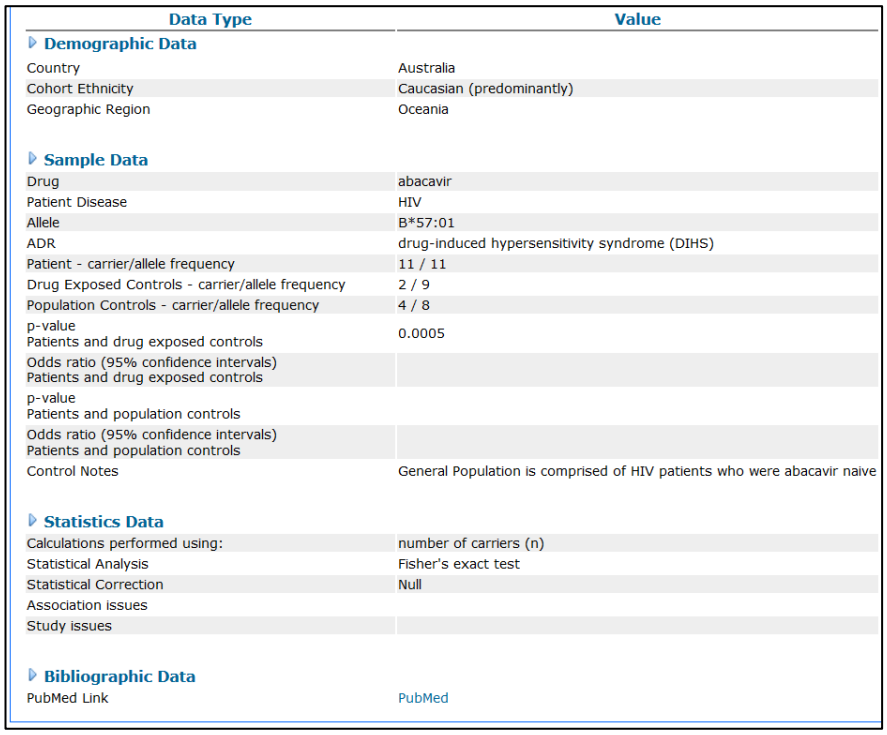
Table 3.1: Description of the information returned when a user performs a search using the HLA-ADR database's query page.

The description column provides information about the types of data that the query page shows.

Record information	Description
PubMed Link	Hyperlink that directs to the reporting article's PubMed entry
Drug	Drug that is reported in the association as well as link to the HLA-ADR database's report page (see section 3.3.1.4).
Allele	The allele that has been tested as part of the association, can include both statistically significant and non-significant associations. Also contains a hyperlink to the main Allele Frequencies Net Database's allele frequency search detailing the frequencies for that particular allele in various populations.
Cohort Ethnicity	Information about the ethnicity makeup for the study's cohort.
p-value (Pat[1] and ExpCtrl[2])	The p-value for the association as reported by the investigating study (comparing patients against the drug exposed controls)
p-value (Pat[1] and Pop[3])	The p-value for the association as reported by the investigating study (comparing the patients against the population controls)
Pat[1] allele or carrier frequency	The proportion of patients carrying the tested allele (n) or proportion of tested alleles in the patient group against total number of alleles (2n)
ExpCtrl[2] allele or carrier frequency	The proportion of drug exposed controls carrying the tested allele (n) or proportion of tested alleles in the drug exposed control group against total number of alleles (2n)
Pop[3] allele or carrier frequency	The proportion of population controls carrying the tested allele (n) or proportion of tested alleles in the population control group against total number of alleles (2n)
More Details	Hyperlink to view the complete information for that particular record (see section 3.3.1.3)
Allele distribution	Hyperlink to view the Allele Frequencies Net Database's allele frequency distribution page relating to the allele associated with the record.

3.3.1.3. More Details page

The More Details page displays the complete information that has been recorded with the association into the database. Here, the record's demographic data (the country and geographic region of where the study was conducted and the patient cohort's ethnicity makeup), information about the sample including the drug, the phenotype that the patients were being treated for, the allele that was analysed as part of this record, the ADR phenotype, the proportion of carriers/alleles (for the patients, drug exposed controls and population controls) and the study calculated p-values and odds ratios can be seen. The page also displays information about the statistics performed by the study investigators such as the type of test performed, whether the statistical calculations were performed using the number of carriers (n) or number of alleles (2n), whether any statistical corrections was applied as well as any issues/notes that are worth highlighting to the user. The page also provides a hyperlink to the PubMed entry for the original article that reported the association. For an example, see Figure 3.5.



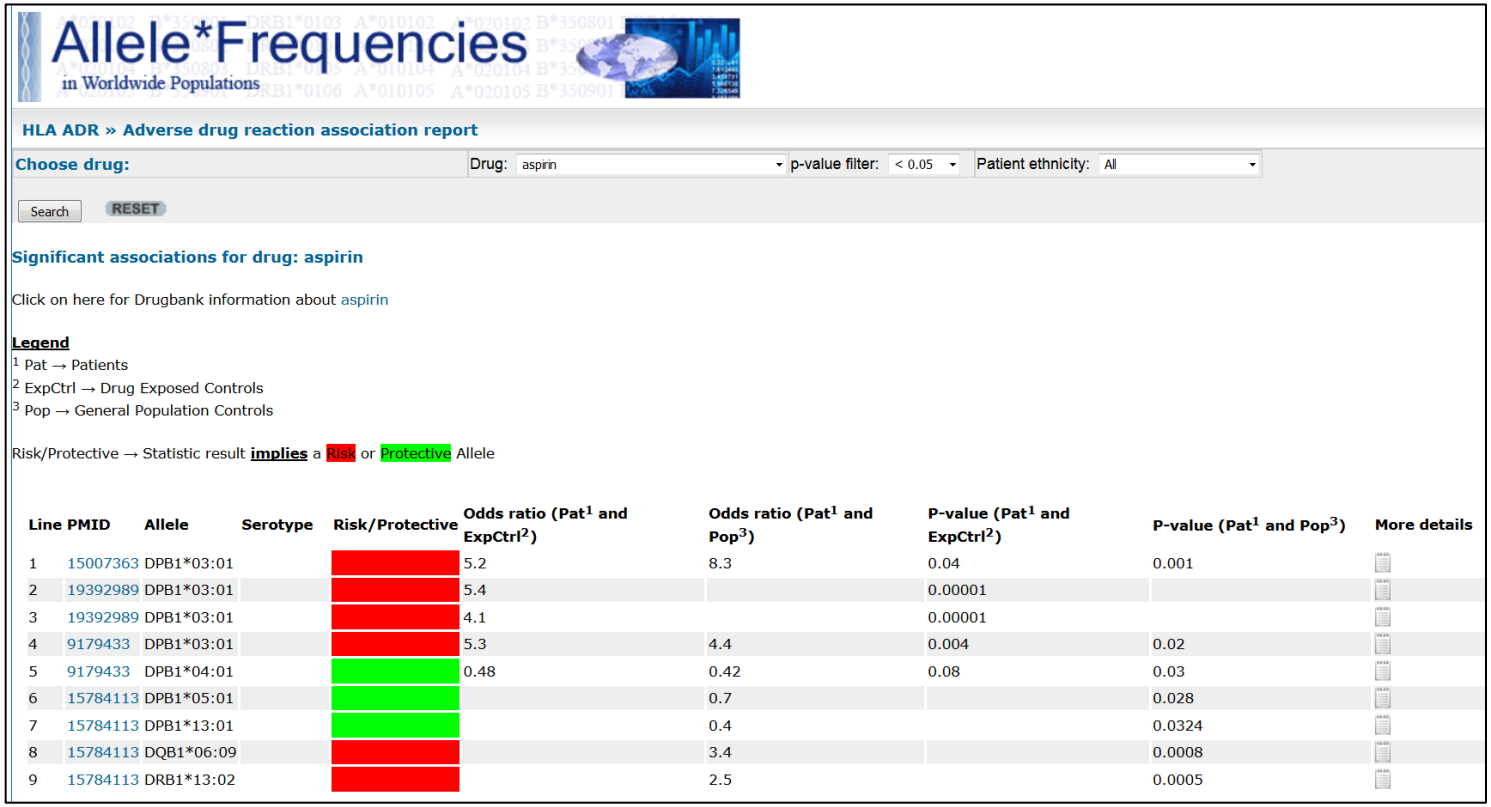
Data Type	Value
Demographic Data	
Country	Australia
Cohort Ethnicity	Caucasian (predominantly)
Geographic Region	Oceania
Sample Data	
Drug	abacavir
Patient Disease	HIV
Allele	B*57:01
ADR	drug-induced hypersensitivity syndrome (DIHS)
Patient - carrier/allele frequency	11 / 11
Drug Exposed Controls - carrier/allele frequency	2 / 9
Population Controls - carrier/allele frequency	4 / 8
p-value Patients and drug exposed controls	0.0005
Odds ratio (95% confidence intervals) Patients and drug exposed controls	
p-value Patients and population controls	
Odds ratio (95% confidence intervals) Patients and population controls	
Control Notes	General Population is comprised of HIV patients who were abacavir naive
Statistics Data	
Calculations performed using:	number of carriers (n)
Statistical Analysis	Fisher's exact test
Statistical Correction	Null
Association issues	
Study issues	
Bibliographic Data	
PubMed Link	PubMed

Figure 3.5: Example screenshot of the More Details page of the HLA-ADR website. The record displayed in this figure is associated with the first entry that was shown in Figure 3.4 (an association between HLA-B*57:01 and abacavir-induced hypersensitivity syndrome). [105]

3.3.1.4. ADR Report page

The HLA-ADR database also provides an ADR report page which provides a summary of all significant associations with a particular drug as well as additional external tools from which users can conduct further research. The user selects a particular drug and the webpage returns all HLA-ADR records which are statistically significant for the selected drug. The user has the option to include additional filters where they may select a particular significance threshold (the default is a p-value of ≤ 0.05; with ≤ 0.01 and ≤ 0.001 also available) and/or limit the results based on the patient-control cohort's ethnic background. The returned summary of an example search, for the drug abacavir is shown in Figure 3.6. The returned summary table of the results page provides the user with a link to the original reporting article's PubMed entry, the allele(s) that have been associated with the ADR, a risk/protective colour designation, the odds ratios and p-values as well as a link to the More Details page where users can view the entire record. The reports page also provides additional information gathered from external resources which may also be of benefit for the user. The additional information is provided on the reports page in HTML frame objects as shown in the figures described below. The frame objects allow the external data to be shown without the user having to manually search and refer to the external providers. Firstly, for each allele listed in the summary table at the top of the page, a worldwide map showing the allele frequency distribution is displayed (see Figure 3.7) which is generated by querying the allele from the AFND Allele Frequency Distribution page and the results are returned in an embedded window within the reports page.

Secondly, the reports page will display HLA haplotypes/alleles that have been shown to co-occur with each of the alleles listed in the summary table (see Figure 3.8). The haplotype searches are performed using the main AFND website where HLA-ADR performs the search with each of the unique alleles identified in the summary section at the top of the reports page. A window for each of the haplotype searches is then provided in the reports page so that users may see if the allele has a high co-occurrence with another allele that was in the reports page summary. The summary table at the top of the page can list out associations with alleles that are found in haplotypes. This feature allows users to determine any co-occurring alleles.



Line	PMID	Allele	Serotype	Risk/Protective	Odds ratio (Pat[1] and ExpCtrl[2])	Odds ratio (Pat[1] and Pop[3])	P-value (Pat[1] and ExpCtrl[2])	P-value (Pat[1] and Pop[3])	More details
1	15007363	DPB1*03:01			5.2	8.3	0.04	0.001	
2	19392989	DPB1*03:01			5.4		0.00001		
3	19392989	DPB1*03:01			4.1		0.00001		
4	9179433	DPB1*03:01			5.3	4.4	0.004	0.02	
5	9179433	DPB1*04:01			0.48	0.42	0.08	0.03	
6	15784113	DPB1*05:01				0.7		0.028	
7	15784113	DPB1*13:01				0.4		0.0324	
8	15784113	DQB1*06:09				3.4		0.0008	
9	15784113	DRB1*13:02				2.5		0.0005	

Figure 3.6: Screenshot of an example search performed on the HLA-ADR reports page.

In this example, aspirin has been selected with the p-value and patient ethnicity filters kept to their default ($p \leq 0.05$ and all ethnicities respectively).

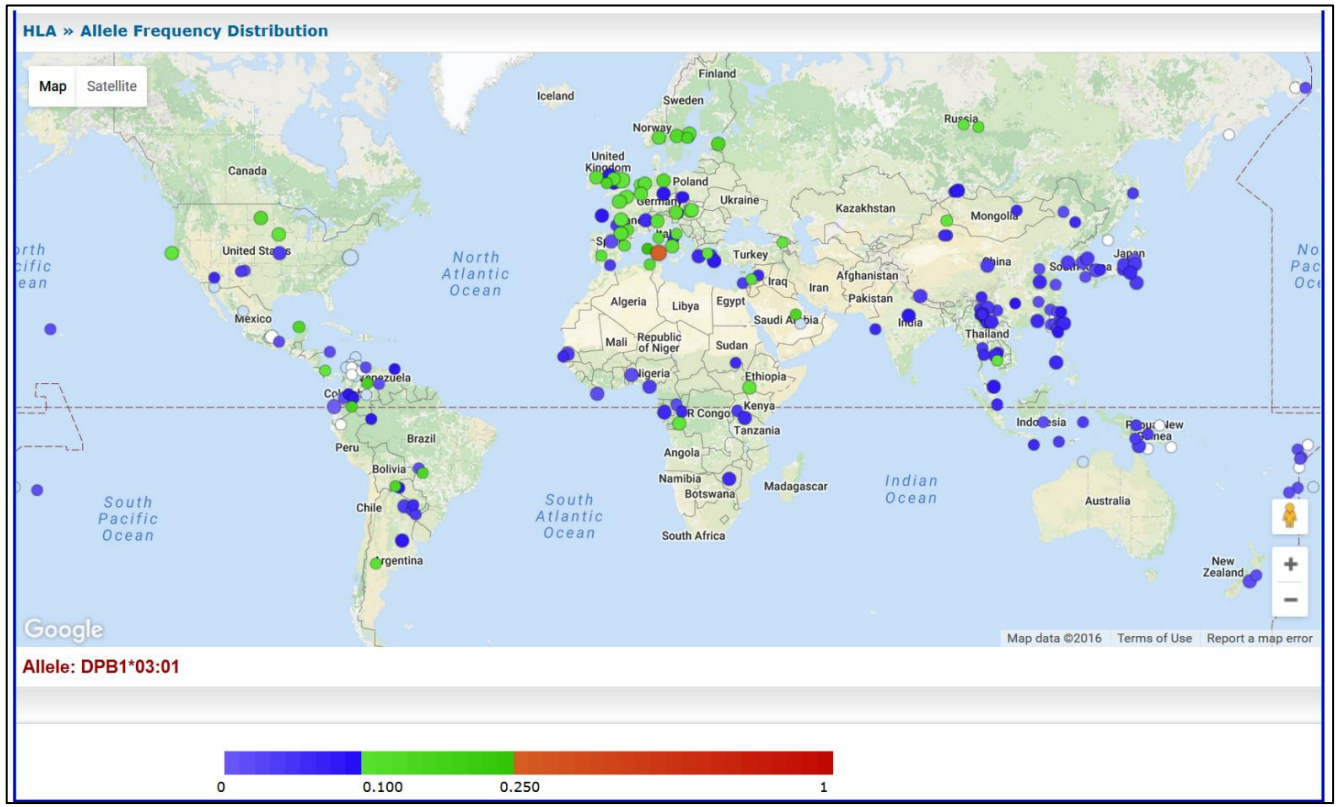


Figure 3.7: Screenshot of the HLA-ADR reports page's Allele Frequency Distribution map.

This is generated by querying each allele on the main AFND Allele Frequency Distribution tool and returning a window of the results into the reports page. Each circle on the map represents a population which is then coloured based on the frequency of the allele relative to the population represented by the circle.

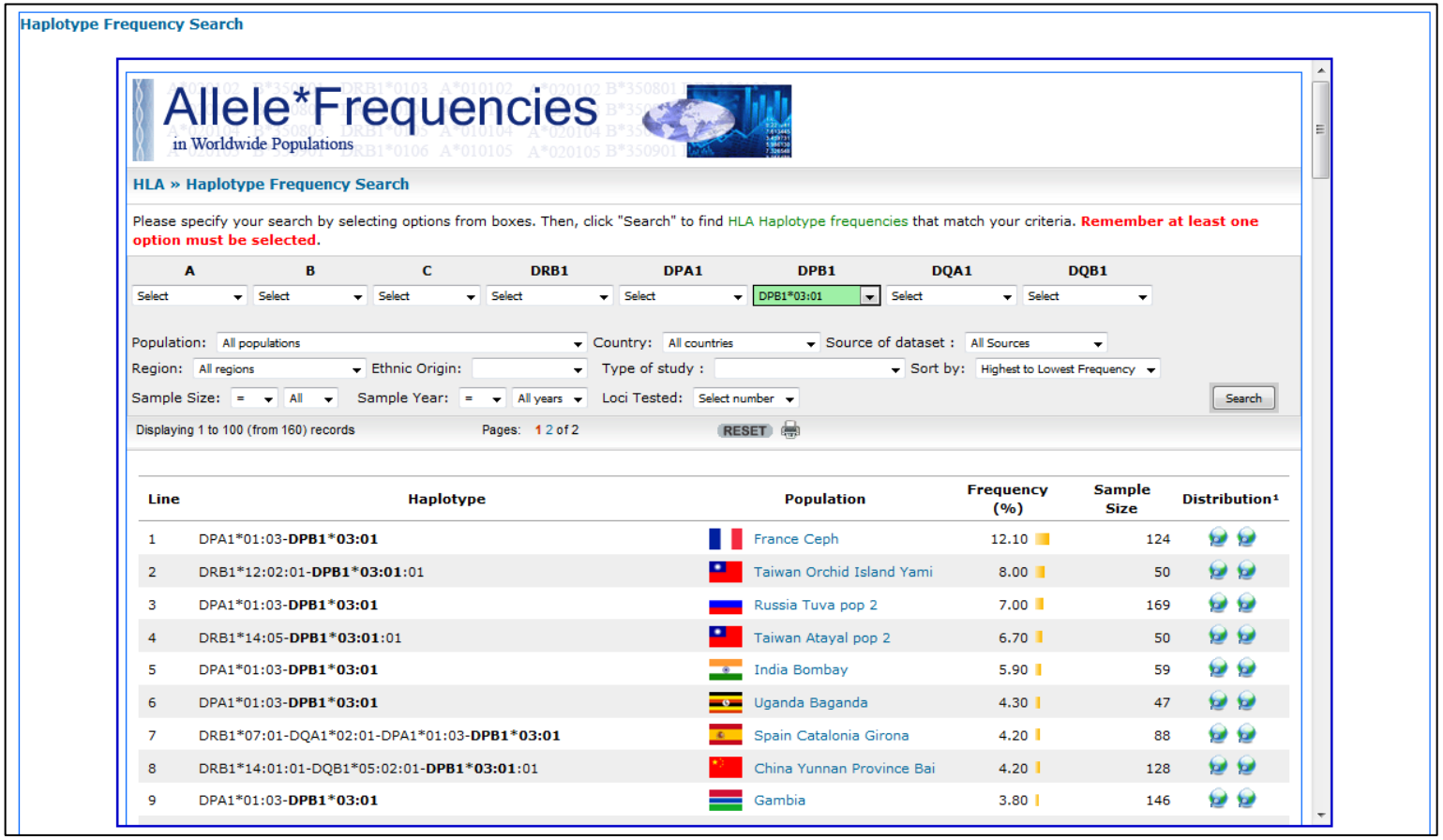


Figure 3.8: Screenshot of the HLA-ADR reports page's Haplotype frequency search section.

This example screenshot continues on from Figure 3.6 and Figure 3.7 with a search for aspirin. The screenshot shows the returned AFND haplotype search for HLA-DPB1*03:01

In addition to the resources from the main AFND website, the HLA-ADR reports page also provided links to other tools that are provided by the EMBL-EBI's IMGT/HLA resource. [78] The first of these IMGT/HLA tools are multiple sequence alignments (MSA), see Figure 3.9, which are performed using the alleles that are featured in the summary table at the top of the page. The alleles are grouped by each gene/loci where the allele(s) that appear in the summary table along with the 10 most common alleles for that loci are passed to the IMGT/HLA MSA. The 10 most common alleles for the gene were determined by reviewing the allele frequencies for each gene using the Allele Frequencies Net Database and are not specific to a certain population (i.e. this list of 10 alleles is based on worldwide frequencies).

3.3.1.5. Author's submission page

A systematic review is an effective method for capturing studies which are relevant to a task, however, it may not capture all available studies due to limitations such as only studies published up to the moment that the systematic review was conducted and are available in the chosen publications database such as MEDLINE and EMBASE can be analysed. Whilst the HLA-ADR website intends to carryout periodic systematic reviews every six months, users of the website can request certain papers to be included manually. To achieve this, an authors' submission page was developed.

The author's submission page provides users with a tutorial .pdf document to download which provides instructions to fill in the spreadsheet document that is also available to download on the author's submission page. The layout of the spreadsheet document is similar to spreadsheet that is used to collect data (as described in Chapter 2.2.4) which therefore makes the transfer of new data from the author's submission spreadsheet relatively straightforward as the data on the author submitted document can be copied over to the master document where it can then be entered into the database. Once the author fills the spreadsheet form with their data, they can then upload the document on the author's submission page where the HLA-ADR administrators are notified of the submission by an automated email.

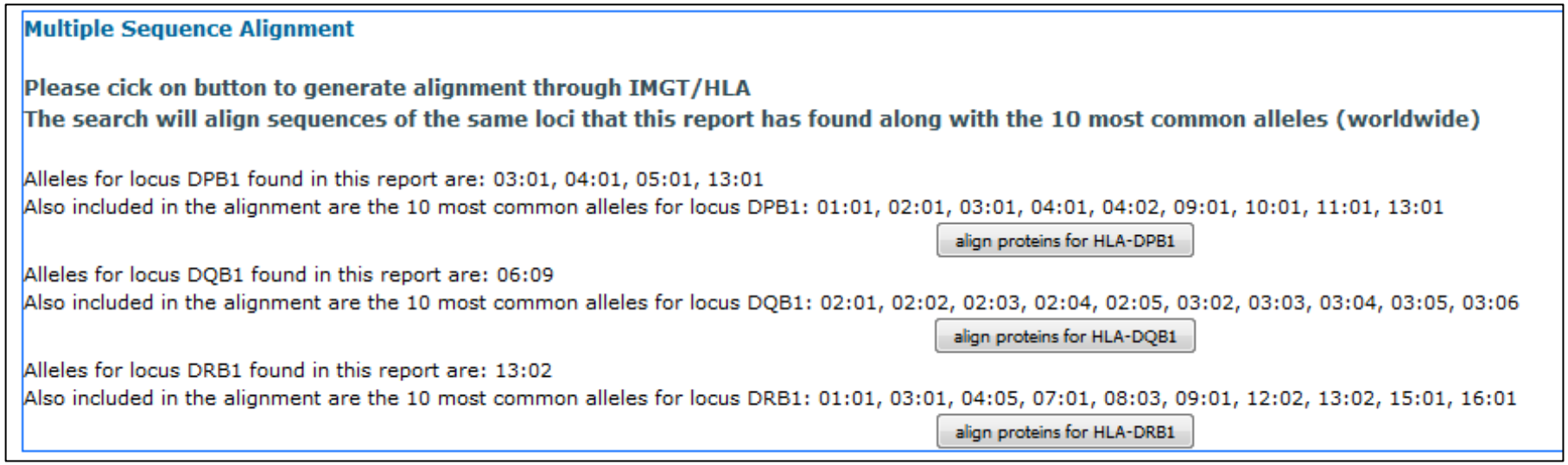
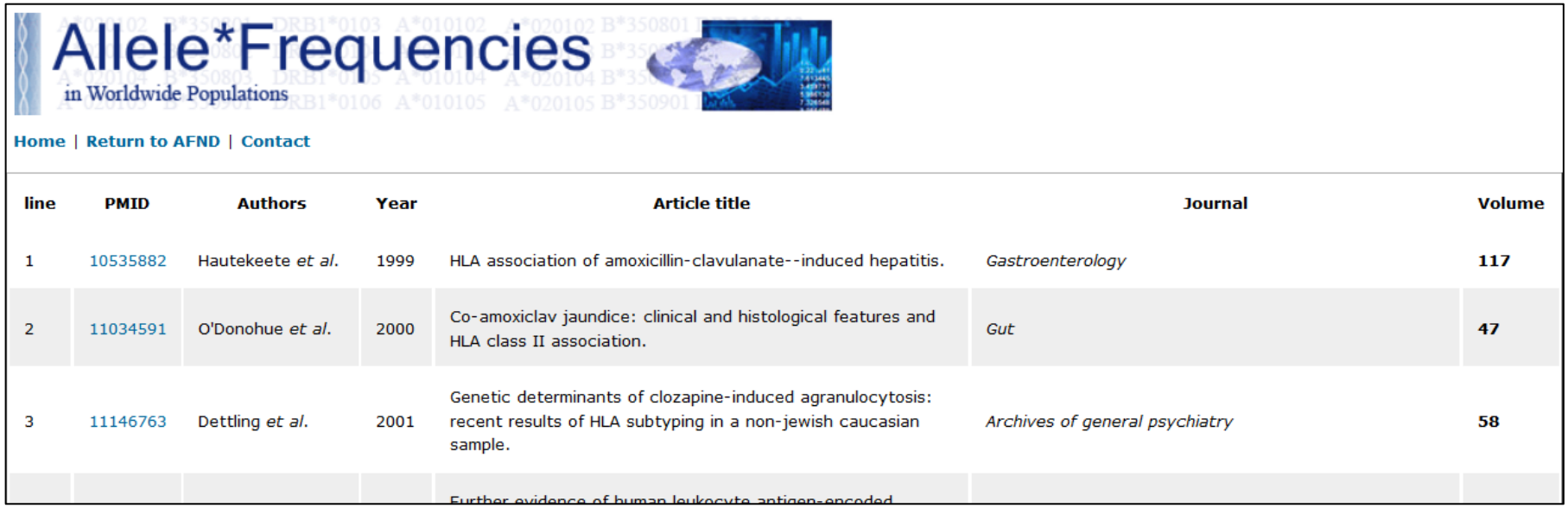


Figure 3.9: Screenshot of the HLA-ADR reports page's multiple sequence alignment (MSA) section.

This example screenshot continues on from Figure 3.6, Figure 3.7 and Figure 3.8 with a search for aspirin. The page shows an MSA for each loci of the alleles that were in the summary table (as featured in Figure 3.6). A button is provided which links to IMGT/HLA MSA page where the MSA uses the alleles identified in the summary table and from the list of the 10 most common alleles for that gene



line	PMID	Authors	Year	Article title	Journal	Volume
1	10535882	Hautekeete *et al.*	1999	HLA association of amoxicillin-clavulanate--induced hepatitis.	*Gastroenterology*	**117**
2	11034591	O'Donohue *et al.*	2000	Co-amoxiclav jaundice: clinical and histological features and HLA class II association.	*Gut*	**47**
3	11146763	Dettling *et al.*	2001	Genetic determinants of clozapine-induced agranulocytosis: recent results of HLA subtyping in a non-jewish caucasian sample.	*Archives of general psychiatry*	**58**
				Further evidence of human leukocyte antigen-encoded		

Figure 3.10: Screenshot of the first three entries of the HLA-ADR bibliography page.

From left to right, the table displays a line number so that the users can see the total number of entries in the database at a glance; the PMID number which acts as a link so the user can be directed to the study's abstract hosted on PubMed; the year; title; the publishing journal and the volume number for the study.

3.3.1.6. Bibliography page

The HLA-ADR website also provides users with a table displaying the studies from which the HLA-ADR database has collected its data from (see Figure 3.10 above). For each study the user is able to view the line number, the PubMed ID (PMID) which also provides a hyperlink to the abstract page for the study hosted on the NCBI PubMed page as well as the year of publication, title, publishing journal and journal volume associated with the publication of the study.

3.4. Discussion

The ADR reports page aims to provide researchers with information from multiple, internal and external resources in one report thereby reducing the time taken to conduct investigations. The risk/protective colour designation in the table at the top of the report is provided as some drug-allele associations that are statistically significant may actually show protective markers where more of the control group carry the allele as compared to the patient group. From a clinical perspective, the apparent protective biomarkers need to be treated differently from risk markers since clinicians can administer the drug to carriers of the apparent protective alleles similarly to patients carrying alleles that are statistically not-significant for ADR associations. Therefore, this distinction between risk and apparent protective alleles is valuable information with regards to the report.

Certain trends in the results as reported in Chapter 2.4.1, the presented hypothesis of reporting bias - the geographic distributions of investigations reporting ADRs, led to the inclusion of the maps displaying the allele frequency distributions so that users could investigate if a reported allele was specific to a certain region(s) of the world. This was aimed to address such conjecture and is investigated in 0.

In addition to these maps, the haplotypes frequency search frame objects were specifically included to allow users to investigate instances where studies report multiple alleles which form haplotypes as significant associations although this is discussed in detail in Chapter 4. As a demonstration of concept, an example search is provided here to illustrate that two alleles, HLA-B*57:01 and HLA-DRB1*07:01, reported by Daly *et al*. as significant, however the study reports that the two alleles

are in linkage whereby the study also demonstrated that HLA-DRB1*07:01 was not independent of the linkage with HLA-B*57:01. [24] The HLA-ADR reports page's Haplotype Frequency search also supports the co-occurrence of these two alleles (see Table 3.2). This haplotype search was conducted using the AFND haplotype frequency search tool, after a filter is applied to only display results where the population sample size is greater than 1,000. Ten out of the top 22 hits (ordered by percentage frequency) were of haplotypes which contained both alleles, although none of the hits exclusively contained just HLA-B*57:01 and HLA-DRB1*07:01 thereby making an exact quantification difficult. It is worth noting that the study by Daly *et al*. [24] reports HLA-B*57:01 as the primary allele of interest to bring forward for further investigation. The co-occurrence of alleles to be significantly associated with certain ADRs is reported in the literature (such as Daly *et al*.) [24] however, it is of potential interest to see if this occurs across many independent studies, this is further investigated in Chapter 4.

The tools that have been built into the HLA-ADR are in themselves useful and powerful to enable users to conduct additional research into immune-mediated ADRs, however the practical utility for these tools is only of value as long as the data is kept up to date. As mentioned above, the author submission page was developed to assist this process. Researchers can submit their own data, both published and unpublished or contact the AFND/HLA-ADR administrators to request data be included. It was decided that unpublished data would be useful as negative results (i.e. an investigation that did not find a significant association) would have greater difficulty in acquiring publication of their results, but the data itself is of great importance as it can be made available for meta-analysis or possibly refute the findings of other studies. The HLA-ADR website and the associated publication (see section regarding publications before the thesis abstract) specifically mention the intention to capture unpublished data. The author's submission page also allows for researchers who feel that their study was not included in the database to be incorporated, this was done specifically to address scenarios where the initial data collection conducted in Chapter 2 was incomplete or not thorough enough to capture all relevant studies. This situation did in fact arise during the HLA-ADR's associated publication as the reviewers indicated specific studies that were not included. The mentioned studies were subsequently examined and were included into the HLA-ADR database.

Table 3.2: Listing the top 22 haplotypes when HLA-B*57:01 is selected in the AFND Haplotype frequency search tool.

Only population sample size > 1,000 is shown. The two alleles have been highlighted in bold and underlined.

Line	Haplotype	Population	Frequency (%)	Sample Size
1	A*01:01-**B*57:01**-DRB1*13:05	Israel Poland Jews	2.47	13,871
2	A*01:01-**B*57:01**-DRB1*13:05	Israel USA Jews	2.40	6,058
3	A*01:01-**B*57:01**-DRB1*13:05	Israel Argentina Jews	2.18	4,307
4	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**-DQB1*03:03	USA NMDP South Asian Indian	2.16	185,391
5	A*01:01-**B*57:01**-DRB1*13:05	Israel Ashkenazi Jews	2.00	4,625
6	**B*57:01**-C*06:02	USA Asian pop 2	1.99	1,772
7	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**-DQB1*03:03	USA NMDP Southeast Asian	1.49	27,978
8	A*01:01-**B*57:01**	USA Asian pop 2	1.41	1,772
9	A*01:01-**B*57:01**-DRB1*13:05	Israel USSR Jews	1.37	45,681
10	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**	Germany DKMS - Austria minority	1.33	1,698
11	A*01:01-**B*57:01**-**DRB1*07:01**	Israel Libya Jews	1.31	3,739
12	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**	Germany DKMS - United Kingdom minority	1.20	1,043
13	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**	Germany DKMS - Netherlands minority	1.14	1,374
14	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**	Germany DKMS - France minority	1.07	1,406
15	A*01:01-**B*57:01**-DRB1*04:03	Israel Bukhara Jews	1.04	2,317
16	A*01:01-**B*57:01**-DRB1*13:03	Israel Bukhara Jews	0.93	2,317
17	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**-DQB1*02:01	USA NMDP South Asian Indian	0.92	185,391
18	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**-DQB1*02:01	USA NMDP Vietnamese	0.87	43,540
19	**B*57:01**-C*06:02	USA Hispanic pop 2	0.85	1,999
20	A*01:01-**B*57:01**-DRB1*13:03	Israel Iraq Jews	0.85	13,270
21	A*02:01-**B*57:01**-C*06:02-**DRB1*07:01**	Poland DKMS	0.84	20,653
22	A*01:01-**B*57:01**-C*06:02-**DRB1*07:01**-DQB1*03:03	USA NMDP European Caucasian	0.83	1,242,890

Provisions have also been made whereby database administrators do not need to be knowledgeable and proficient in both immune-mediated ADRs and database management in order to keep the database up to date. A researcher into immune-mediated ADRs who can understand the data can maintain the spreadsheet document (mentioned in 2.2.4) relatively easily and then execute the specially designed Python script which will automatically read in the data from the spreadsheet and then update the database accordingly. This reduces the need to hire a database administrator who is proficient in both computation and biological fields or hire two individuals from each field in order to maintain the database. For security reasons, the Python script should only be accessible to members of the database administration/curation team.

It is envisaged that the database will primarily be used to conduct meta-analyses by researchers. HLA-ADR can allow these researchers to perform a quick search about their drug(s)/ADR phenotype etc. of choice and have access to the data from the published literature and save them from performing an extensive literature search themselves. The tools created for the HLA-ADR website and the links to external resources will enable these meta-analyses to be performed in a relatively straightforward and streamlined approach whereby HLA-ADR is either the singularly, or one of few, utilities that are employed or even act as a starting point from which to conduct further informatics investigations.

Although the HLA-ADR website has been live for only a relatively short period of time, there are already a few examples demonstrating the utility of the website for researchers. There are two non-self-citing studies that are accessible and have cited HLA-ADR (see 3.6 for the HLA-ADR publication details). [122, 123] Chapter 4 also demonstrates an example of the database being used to support meta-analysis and Chapter 6 showcases the data being used with a clinical perspective. The ultimate goal for a bioinformatics resource is to aid several studies.

3.5. Conclusion

Investigating immune-mediated adverse drug reactions requires the lengthy process of conducting deep literature reviews in order to collect the necessary data for the researchers' investigation. Over the years, data pertaining to HLA associated ADRs

has grown and will likely continue to grow as more drugs are developed and enter the market. It is therefore important that the data is digitised and centralised to enable clinicians and researchers to be able to access all of the available information in a flexible manner without having to conduct a systematic review each time. Therefore, it is proposed that this web based HLA-ADR database which provides flexible access to relevant data will be of great benefit to the research and clinical communities.

Chapter 4. Investigating observed patterns between reported HLA associated adverse drug reactions and patient ethnicity

4.1. Introduction

Following on from the previous chapters, a hypothesis that was posed from the analysis of the data from Chapter 2 was the possibility of case-control publication bias whereby many of the studies relating to immune-mediated ADRs were published from a limited number of countries – predominantly in Asia and Europe. This was proposed as a possible reason as to why certain ethnicities are reported to be at higher risk of experiencing ADRs when exposed to a drug (as well as the possibility that the observations are indicative of a genuine difference in response between the population groups).

The relatively high number of studies from Asia has influenced drug labelling advice from the regulatory drug bodies (see Table 4.1) where certain antiepileptic drugs are contraindicated in patients of Asian origin. The table shows that to date, the United States Food and Drug Administration (FDA) and the Clinical Pharmacogenetics Implementation Consortium (CPIC) have issued recommendations for the pre-treatment genetic screening for four drugs including carbamazepine and phenytoin. The labels other drugs featured in this thesis did not include any advice with regards to screening patients for risk HLA alleles.

As a result of this reported pattern showing differing responses to drugs based on patient ethnicity, this chapter includes an investigation carried out using the data collected in Chapter 2 and the toolset developed in Chapter 3 to see if this data set could show any statistical evidence to support these observations that ethnicity can influence risk profiles for a particular ADR. The aim of this chapter is to summarise the data present in the HLA-ADR database with particular focus on the variability based on patient ethnicity in the reporting of HLA associated ADRs. The findings of this chapter have been published and the paper serves a condensed version of this chapter. [124]

Table 4.1: FDA and CPIC drug labelling advice related to HLA associated adverse drug reactions

Drug	FDA advice	CPIC advice
Abacavir	"Prior to initiating therapy with abacavir, screening for the HLA-B*5701 allele is recommended." [49]	"HLA-B*57:01 screening should be performed in all abacavir-naive individuals before initiation of abacavir-containing therapy." [125]
Allopurinol	No information available pertaining to HLA mediated ADRs.	No direct recommendation regarding pre-screening of patients for HLA-B*58:01. However, for patients known to carry HLA-B*58:01: "allopurinol is contraindicated." [126]
Carbamazepine	"Patients with ancestry in genetically at-risk populations [Asia] should be screened for the presence of HLA-B*1502 prior to initiating treatment with Tegretol [carbamazepine]. Patients testing positive for the allele should not be treated with Tegretol unless the benefit clearly outweighs the risk." [51]	Directly quotes FDA - "'patients with ancestry in at-risk populations should be screened for the presence of HLA-B*15:02 allele prior to starting carbamazepine'. Individuals at highest risk are those of Han Chinese descent, followed by those in Vietnam, Cambodia, the Reunion Islands, Thailand, India (specifically Hindus), Malaysia, and Hong Kong." [127]
Phenytoin	"Consideration should be given to avoiding phenytoin as an alternative for carbamazepine in patients positive for HLA-B*1502. The use of HLA-B*1502 genotyping has important limitations and must never substitute for appropriate clinical vigilance and patient management." [53]	Directly quotes FDA - "'Consideration should be given to avoiding phenytoin as an alternative for carbamazepine in patients positive for *HLA-B*15:02*' due to the increased risk of SJS/TEN in patients of Asian ancestry." [128]

4.2. Methods

The alleles that were reported to be statistically significantly associated (defined as $p < 0.05$) from the data within the HLA-ADR database were brought forward for analysis in this investigation. The alleles were grouped by the drug, e.g. for the drug lapatinib, the alleles HLA-DRB1*07:01, HLA-DQA1*02:01 and DQB1*02:02 were reported as significantly associated and therefore analysed together. This was repeated for each drug featured in the results section. Alleles were permitted to be part of multiple groups e.g. HLA-B*15:02 was part of the groups relating to carbamazepine, phenytoin, lamotrigine etc. Next the frequencies for each allele were obtained using the Allele Frequency Net Database (AFND) allele frequency search tool using what AFND labelled as 'gold standard' population data. These 'gold standard' populations are defined as information for the population that is available at two fields or greater for the HLA alleles. [129] Wherever possible, these frequencies were obtained where the sample size for each population was greater than or equal to 1,000. In situations where populations with sample sizes above or equal to 1,000 were not available, the size of the populations are provided in the text. All of the allele prevalences that were collected for this investigation are quoted as the frequency of the allele within the population as such, the carrier frequencies would be approximately double the allele frequency. Individuals that are homozygous across the entire MHC region are extremely rare except in situations of consanguineous parentage. Therefore, doubling the allele frequency to obtain the carrier frequency provides a good approximation.

In addition to the allelic frequency data, the frequencies for HLA haplotypes with each set of the above mentioned grouped alleles were obtained from the AFND haplotype search tool. As with the allele frequencies, sample sizes greater than 1,000 were preferred for these haplotype frequencies however, this was not always possible. For all instances of referring to haplotype data, the sample sizes have been included in the results section. For this investigation, when referring to the studies individually, the populations are described using the country name or geographic location as used in the original study. However, when conducting the meta-analysis, the populations are categorised into three major demographic groups: African, Asian and European. It is also worth noting that populations represented within AFND are labelled based on the ancestry of that group and these are reflected in this investigation, for example

populations from the United States where the members of the population sample are of Chinese origin are classed as Asian in this meta-analysis. This classification based on ancestry allows the analysis to explore the relationship of ADRs with ethnicity without having to compensate for population nationality. The primary advantage for this simplification of populations into three major groups provides the analysis with a higher statistical power, however this is done at the expense of some granularity of the information.

For the analysis of all of the data, the information was summarised as a table of descriptive statistics with the Mann-Whitney U test (a nonparametric comparison of the medians) applied to compare the data pertaining to Asian and Europeans to see if certain alleles could be described as having higher frequencies in certain populations - the descriptive statistics and the Mann-Whitney U test were performed by EJM Santos, co-author of the study that published these findings. To expand on this, a Kruskal-Wallis test was applied to perform a three-way comparison of the three ethnic super-groups and Bonferroni correction for multiple comparisons was applied (The Mann-Whitney U test only makes a pairwise comparisons). The Kruskal-Wallis test was performed using the R statistical software package (Version 3.2.2 - R Foundation for Statistical Computing, 2015, UK https://cran.r-project.org). The Kruskal-Wallis test was performed by the author.

For the data relating to carbamazepine, there was a sufficient number of alleles to perform a Spearman's correlation test to investigate the co-occurrence of the alleles reported to be associated with carbamazepine hypersensitivity. The statistical analyses were performed independently by two researchers; EJM Santos performed the Spearman's correlation and the analysis using the AFND haplotype searches was performed by the author. The AFND haplotype data are recorded from population samples, i.e. obtained directly from sample sources. The two analyses were compared to measure relative concordance.

For the analysis of dapsone, there is reference to the geographic region of Oceania. For the context of this thesis, Oceania is defined by the countries listed in Table 4.2.

Table 4.2: Countries classified as 'Oceania' in the HLA-ADR database

Country
American Samoa
Australia
Christmas Island
Cocos Islands
Cook Islands
East Timor
Federal States of Micronesia
Fiji
French Polynesia
Guam
Indonesia
Kiribati
Marshall Islands
Nauru
New Caledonia
New Zealand
Niue
Norfolk Island
Northern Mariana Islands
Palau
Papua New Guinea
Pitcairn Islands
Samoa
Solomon Islands
Tokelau
Tonga
Tuvalu
Vanuatu
Wallis and Futuna Islands

4.3. Results

All frequencies referenced in this results section were obtained from the Allele Frequencies Net Database [120, 121] (AFND) using the following URLs: the frequencies of individual alleles the URL was http://www.allelefrequencies.net/hla6006a.asp and the frequencies for HLA haplotypes were obtained using the URL http://www.allelefrequencies.net/hla6003a.asp. These URLs were accessed June 2016.

4.3.1. Antigout drugs

4.3.1.1. Allopurinol

Allopurinol is prescribed to patients with gout/hyperuricaemia. Several studies have been published which have reported allopurinol hypersensitivity to be associated with carriage of HLA-B*58:01 across many populations including Caucasian, [130, 131] Han Chinese, [23, 132, 133] Japanese, [133] Korean [133-135] and Thai [136] patients. Another study also reported the association with HLA-B*58:01 in a patient-control cohort which has been marked as 'diverse' as the study did not separate the cohort by ethnicity but did state that the cohort consisted of individuals of South Asian, South American, European and African ancestries.

As well as HLA-B*58:01, the studies referred to above also reported associations that were statistically significant between allopurinol-induced hypersensitivity and HLA-A*33:03, HLA-C*03:02 and HLA-DRB1*03:01. The HLA-A*33:03 allele has been reported to be associated with allopurinol induced hypersensitivity in patients of Caucasian, Han Chinese and Korean origin. HLA-B*58:01, HLA-A*33:03 and HLA-C*03:02 were found to be significantly associated in Han Chinese and Korean patients. [133, 135] The AFND HLA haplotype frequency search (URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016) shows that these three alleles co-occur with the haplotype relatively commonly in East Asian populations where the frequency is as high as 8% of the population. In comparison, this haplotype frequency is < 2% in European populations where the frequency of the individual alleles is also < 2%. The frequency of the haplotype is not available for

African populations. The frequencies for the individual alleles are between 5-6% for HLA-A*33:03, 4% for HLA-B*58:01 and 2% for HLA-C*03:02.

The class II allele HLA-DRB1*03:01 has been found with relatively high co-occurrence with HLA-A*33:03 and HLA-B*58:01 in east and south-east Asian populations, specifically, as high as 6.6% co-occurrence for all three alleles in Taiwan – from AFND Haplotype Frequency tool (accessed June 2016). Furthermore, the haplotype containing all four of the alleles (HLA-A*33:03, HLA-B*58:01, HLA-C*03:02 and HLA-DRB1*03:01) is reported to have 4.4% frequency in south-west China. The haplotype data indicates that the simplest explanation is that not all four alleles are independently causative (or are a risk factor), but that the alleles gained statistical significance due to high linkage disequilibrium with the actual risk factor allele. Given that the association for HLA-B*58:01 carriage and allopurinol induced hypersensitivity was replicated by multiple studies and across many ethnic groups, HLA-B*58:01 is the strongest candidate as the true risk factor for this ADR.

4.3.2. Antituberculosis drugs

Tuberculosis (TB) is treated using a combination regimen of multiple drugs (isoniazid, rifampicin, ethambutol and pyrazinamide). It is consequently difficult to establish the causative agent for anti-TB induced ADRs. A study investigating anti-TB drug-induced hypersensitivity syndrome (DIHS) in Korean patients who were receiving the combination regimen reported HLA-C*04:01 as having statistically significant association, $p = 0.02$, OR = 6.9 (95% CI: 2.2-20.11), based on HLA-C*04:01 allele carriage of 7/14 hypersensitive patients, 21 of 166 drug-tolerant controls and 62/485 population controls. [137] AFND allele frequency search tool reports that this allele is found across the world in many populations including 22.8% in a population of African ancestry, 19.3% in a Turkish population, 17.5% in Latin Americans and 13.6% in South Asians. All of the populations referred to above were of sample size > 1,000.

HLA-associated hepatotoxicity has also been reported in an Indian patient cohort. [138] The reporting study presented HLA-DQB1*02:01 as a statistically significant risk factor for drug induced liver injury (DILI) where 25/47 DILI patients and 84/256 drug-exposed controls carried the allele: $p \leq 0.01$ (OR = 2.2 and 95% CI: 1.19-4.15).

Conversely, the study also reported that carriers of HLA-DQA1*01:02 were statistically less likely to experience an ADR based on 3/47 DILI patients and 68/251 drug exposed controls carrying the allele: $p < 0.001$ (OR = 0.2 and 95% CI: 0.05-0.61). Within the Indian population, HLA-DQB1*02:01 is found at a frequency of 19.9% (sample sizes were limited to between 58 and 202). In other populations, HLA-DQB1*02:01 is observed at frequencies of 59% in Sardinians, 22.8% in a US population of European ancestry, 22.3% in population of African origin and 21.8% in Latin Americans. In all cases of these referenced populations, sample sizes > 1,000.

4.3.3. Antibiotics

4.3.3.1. Dapsone

The antibiotic, dapsone, is used to treat leprosy and has been associated with hypersensitivity reactions with Han Chinese patients. [139] The group reporting the association presented that carriage of HLA-B*13:01 and HLA-C*03:04 crossed the threshold of statistical significance, $p = 2.04 \times 10^{-16}$; $odds\ ratio = 21.67$ (95% CI: $1.41 - 45.12$) and $p = 1.84 \times 10^{-14}$; $Odds\ ratio = 13.43$ (95% CI: $6.91 - 26.11$) respectively. However, when the study adjusted for the fact that these two alleles are in linkage, only HLA-B*13:01 remained significantly associated – adjusted $p = 5.16 \times 10^{-5}$; $odds\ ratio = 17.92$ (95% CI values were not provided in the publication).

HLA-B*13:01 is found predominantly in the region of East Asia, South-east Asia and Oceania (as defined by AFND and provided in Table 4.2 in this chapter's Methods section). The AFND allele frequency tool reports that HLA-B*13:01 is found at a frequency of, 6.4% in a Han Chinese population, 3.9% in Vietnamese and 2.8% in the Philippines for samples sizes > 1,000. Higher frequencies are reported in populations with smaller sample sizes, approx. 100, in the populations originating from Papua New Guinea, Taiwan and Australia (Aborigine) where the allele frequencies were between 25-28%.

4.3.3.2. Amoxicillin-clavulanate

Amoxicillin is a broad-spectrum antibiotic and can be prescribed as on its own or combined with clavulanic acid where it is known as either co-amoxiclav or amoxicillin-clavulanate. Adverse reactions with this drug have been reported to be associated with HLA-DRB1*15:01, HLA-DRB5*01:01, HLA-DQB1*06:02, HLA-A*02:01, and HLA-A*30:02 (see Table 4.3). In all four of these studies, the patient/control cohorts were of European ancestry, specifically Belgium, Scotland, mixed European (north-west European and Spanish subpopulations) and Spain. Of these listed alleles, HLA-DRB1*15:01 and HLA-DQB1*06:02 are of particular note as they were reported by all four studies for the former allele and three out of four studies for the latter. The AFND haplotype search tool reported that co-occurrence of the two alleles is > 10% in many diverse populations including Russian, Indian, European, East Asian and Australian Aborigine. The two alleles are also found together at relatively lower frequencies in many other populations.

Table 4.3: List of alleles that have been reported to be associated with amoxicillin-clavulanate/co-amoxiclav induced adverse drug reactions along with the reporting study and the origin of the studies' cohorts.

Allele	Reporting study	Study cohort subpopulation
HLA-A*02:01	Lucena *et al*., 2011 [140]	mixed European
HLA-A*30:02	Stephens *et al*., 2013 [141]	Spain
HLA-DRB1*15:01	Hautekeete *et al*., 1999 [25]	Belgium
	O'Donohue *et al*., 2000 [142]	Scotland
	Lucena *et al*., 2011	mixed European
	Stephens *et al*., 2013	Spain
HLA-DRB5*01:01	Hautekeete et al., 1999	Belgium
HLA-DQB1*06:02	Hautekeete et al., 1999	Belgium
	Lucena *et al*., 2011	mixed European
	Stephens *et al*., 2013	Spain

4.3.3.3. Flucloxacillin

The β-lactam antibiotic, flucloxacillin is typically used to treat gram-positive bacterial infections. The HLA-B*57:01 allele has been associated with flucloxacillin-induced DILI. [24] The study also reported that HLA-DRB1*07:01 was also statistically significant, however, the focus of the report's conclusion was to HLA-B*57:01 as this allele had the stronger association ($p = 8.97 \times 10^{-19}$ for HLA-B*57:01 compared with $p = 1.9 \times 10^{-5}$ for HLA-DRB1*07:01). Based on the strength of the associations, HLA-B*57:01 is the more likely candidate to be classed as the true risk marker for flucloxacillin-induced hepatotoxicity.

Whilst this study was conducted in the United Kingdom with a patient cohort of European ancestry, both of these alleles are found in many diverse populations. HLA-B*57:01 carriage has been reported by AFND to be in populations of Indian, European or East Asian ancestry at frequencies over 1%. HLA-DRB1*07:01 is found at frequencies greater than 10% in Indian, European, North African, East Asian, Native American and Latin American populations. The study also reported that the two alleles co-occur frequently as part of the haplotype which is shown in Table 4.4 below and the frequencies for the haplotype which contain all 5 of these alleles were obtained from AFND to provide frequency data for the co-occurrence are listed in Table 4.5 which show the diversity of the haplotype.

Table 4.4: List of HLA alleles that are reported by Daly *et al*. that form a commonly occurring haplotype with the two alleles that the study reported to be statistically significantly associated with flucloxacillin-induced hepatotoxicity (HLA-B*57:01 and HLA-DRB1*07:01).[24]

HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1
A*01:01	B*57:01	C*06:02	DRB1*07:01	DQB1*03:03

Table 4.5: Haplotype frequencies obtained from Allele Frequencies Net Database for the haplotype containing the alleles from Table 4.4 (HLA-A*01:01–HLA-B*57:01–HLA-C*06:02–HLA-DRB1*07:01–HLA-DQB1*03:03).
URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016.

Population	Haplotype occurrence	Sample size
North African (Tunisia)	9%	100
South Asian	2.16%	185,391
South-east Asian	1.49%	27,878
European	0.83%	1,242,890

4.3.3.4. Sulfamethoxazole

Sulfamethoxazole is a broad spectrum antibiotic and has been reported to induce SJS/TEN in some patients. A study by Kongpan *et al*. reported that HLA-B*15:02, HLA-C*06:02 and HLA-C*08:01 were statistically significantly associated with these drug-induced hypersensitivity reactions in a Thai patients. [143] HLA-B*15:02 is found at high frequency in East Asian and Southeast Asians and the populations showing the highest carriage rate are shown in Table 4.6. The allele frequencies listed in this table refer to populations with a sample size greater than 1,000. There were additional, smaller populations which showed a larger representation for the HLA-B*15:02 allele. There are five subpopulations from various provinces of China (Han Chinese) that have HLA-B*15:02 allele frequencies between 14.8-35.8% where the sample sizes ranged from 108-153.

HLA-C*06:02 is found in many populations and can be found at frequencies (where sample sizes are >1000) of 13.9% in South Asians, between 8 and 11% in many European populations, between 3 and 5% in East Asians and around 8% in populations of recent African ancestry. HLA-C*08:01 is particularly common in East and South-East Asian populations where it can be found at frequencies of 16.4% in Vietnamese, 10% Chinese and 7% in Japanese and Korean populations. It has also been reported in North American (Mexican, Native Alaska and Native American) population at a frequency between 1 and 3%.

Table 4.6: Populations which have the highest frequencies of HLA-B*15:02 carriage. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016 and the website was set to show populations where the sample size was greater than 1,000

Population	HLA-B*15:02 Frequency	Sample size
Vietnamese (USA)	13.8%	43,540
Chinese (Hong Kong)	9.38%	7,595
Chinese (USA)	6.47%	99,672
Filipino (USA)	4.21%	50,614

4.3.4. Antiepileptic drugs

4.3.4.1. Carbamazepine

The anticonvulsant, carbamazepine is employed in the treatment of epilepsy, trigeminal neuralgia and bipolar disorders. It is one of the most studied drugs in terms of HLA associated ADRs and is represented in the HLA-ADR database with over 27 studies investigating carbamazepine induced hypersensitivity reactions. From these studies, 22 HLA alleles, at 2 field resolution, have been reported to be statistically significantly associated (where $p < 0.05$) for carbamazepine induced ADRs – these alleles are shown in Figure 4.1. Of particular note are the HLA-A*31:01 and HLA-B*15:02 alleles which are reported as significant in 11 and 19 papers respectively. Also, shown in Figure 4.1 are that the remaining 20 alleles have not been well replicated as they were reported in three or fewer studies. For example, a study by Li *et al*. reported the allele HLA-A*02:01 to be statistically significant associated ($p = 0.033$) with carbamazepine induced hypersensitivity in Han Chinese patients. [21] However, statistical significance was not achieved in other studies in which the allele was part of the patient-control cohort; i.e. these studies observed HLA-A*02:01 in their respective cohorts but carriage of the allele was not found in a statistically significantly higher proportion of the hypersensitive patients compared to the drug exposed controls. This lack of statistical significance for HLA-A*02:01 replication includes studies investigating carbamazepine induced ADRs in other Han Chinese patients [144, 145] or Japanese and Korean patients. [18, 146]

For studies reporting HLA-B*15:02 associations with carbamazepine hypersensitivity reactions, 15 of the 19 reporting studies were conducted with a patient-control cohort of Han Asian origin, specifically: Han Chinese (n = 10), Thai (n = 3), Malay (n = 1) and Indian (n = 1) with the remaining four investigations reported each of their associations based on ethnically diverse study cohorts. Referring back to Table 4.6 which shows the populations where HLA-B*15:02 allele frequency is at its highest, this is predominantly in the East and South-East Asian populations, in fact the first entry in AFND allele frequency search for a non-East or South-East Asian population is an entry for Caucasians (Italy) where the allele was found at 0.1% in a sample of 1,089. This low allelic frequency in non-Asians is the most probable reason as to why no associations have been reported between HLA-B*15*02 and carbamazepine induced SJS/TEN in Caucasians. [147, 148]

In contrast, HLA-A*31:01 is a globally occurring allele at a low frequency throughout the populations which are shown in Table 4.7. The literature reflects the universality of the allele where of the 11 of the reported significant associations for HLA-A*31:01 carbamazepine hypersensitivity, two studies were conducted with a Japanese patient cohort, [18, 117] two of the associations were Korean, [146] three used a Han Chinese cohort, [144, 149, 150] two were performed with Europeans [19, 149] and two associations were found using an ethnically diverse patient-control cohort. [34] Note that some studies reported the association twice (as part of the systematic review in Chapter 2, the patient/control groups were separated by ethnicity or ADR phenotype so this section may report them separately – e.g. above mention of two Korean associations but only one cited study).

A Spearman's correlation analysis (Table 4.8) was performed on the alleles featured in Figure 4.1 (i.e. alleles present in HLA-ADR that have been associated with carbamazepine ADR). This analysis is a statistical inference of the likelihood that two alleles are likely to co-occur. Only alleles which are statistically significant ($p < 0.05$) as reported by the literature sources contained within the HLA–ADR database and where more patients carried the allele compared to the control group – that is positive statistical significance indicating risk were included for the analysis. Alleles where one or more study reported the allele to have an apparent protective effect – statistical significance where more controls carried the allele compared to the hypersensitive

patients were excluded. The table shows that only two alleles: HLA-A*31:01 and HLA-B*39:02 did not display evidence of linkage disequilibrium with the other alleles associated with carbamazepine induced ADRs which suggests that these alleles may be independently associated with the carbamazepine hypersensitivity. For the remaining alleles, a more complex model of linkage disequilibrium exists with varying degrees in the strength of the association. For example, HLA-A*02:01 showed weak linkage with HLA-DRB1*07:01 ($0.05 < p > 0.01$), whilst HLA-DRB1*12:02 seems to be in strong linkage with four of the other alleles associated with carbamazepine hypersensitivity (HLA-B*15:02, HLA-B*58:01, HLA-C*03:02 and HLA-C*08:01). This correlates with the observed haplotype frequency from AFND whereby HLA-B*15:02, HLA-C*08:01 and HLA-DRB1*12:02 are known to form a haplotype which is common in Asians. The highest frequencies for this haplotype are found in 5.6% of Vietnamese people (n = 43 540) and also 2.5% in a Chinese population (n = 99 672).

Table 4.7: Populations where the HLA-A*31:01 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-A*31:01 frequency	Sample size
Japanese (Japan)	8.49%	24,582
Native American (USA)	7.37	35,791
Korean (USA)	5.35%	77,584
Latin American (USA)	5.31%	261,235
South Asian – Indian (USA)	3.31%	185,391
Caucasian (Netherlands)	3.30%	1,305

Table 4.8: Nonparametric Spearman's correlation analysis showing the statistically inferred measure of linkage disequilibria between the HLA alleles reported to be associated with carbamazepine induced hypersensitivity reactions.

This analysis was performed across all the ethnicities/populations. Higher positive correlation values signify that the alleles are more likely to appear together. Data obtained from URL Allele Frequencies Net Database: URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Allele	A*02:01	A*31:01	B*15:02	B*15:11	B*39:02	B*58:01	C*03:02	C*08:01	DRB1*07:01	DRB1*12:02
A*02:01										
A*31:01	0.221									
B*15:02	-0.684 [b]	-0.303								
B*15:11	-0.333	-0.067	0.071							
B*39:02	0.371	-0.200	-0.400	0.500						
B*58:01	-0.785 [b]	-0.396 [a]	0.662 [b]	0.190	-0.257					
C*03:02	-0.812 [b]	-0.215	**0.601** [a]	0.310	-0.800	**0.847** [b]				
C*08:01	-0.793 [b]	-0.004	**0.709** [b]	0.000	0.600	**0.667** [b]	0.714 [b]			
DRB1*07:01	**0.343** [a]	0.041	-0.200	-0.233	-1.000 [b]	-0.468 [b]	-0.557 [b]	-0.596 [b]		
DRB1*12:02	-0.776 [b]	0.026	**0.754** [b]	0.107	0.000	**0.713** [b]	**0.749** [b]	**0.881** [b]	-0.592 [b]	
DRB1*14:05	-0.200	-0.147	-0.143	**0.786** [a]	0.500	0.050	0.170	0.571	-0.564	0.627 [a]

In bold are the significant positive correlations (nonparametric Spearman's correlation).

[a] Correlation is significant at the 0.05 level (two-tailed);

[b] Correlation is significant at the 0.01 level (two-tailed)

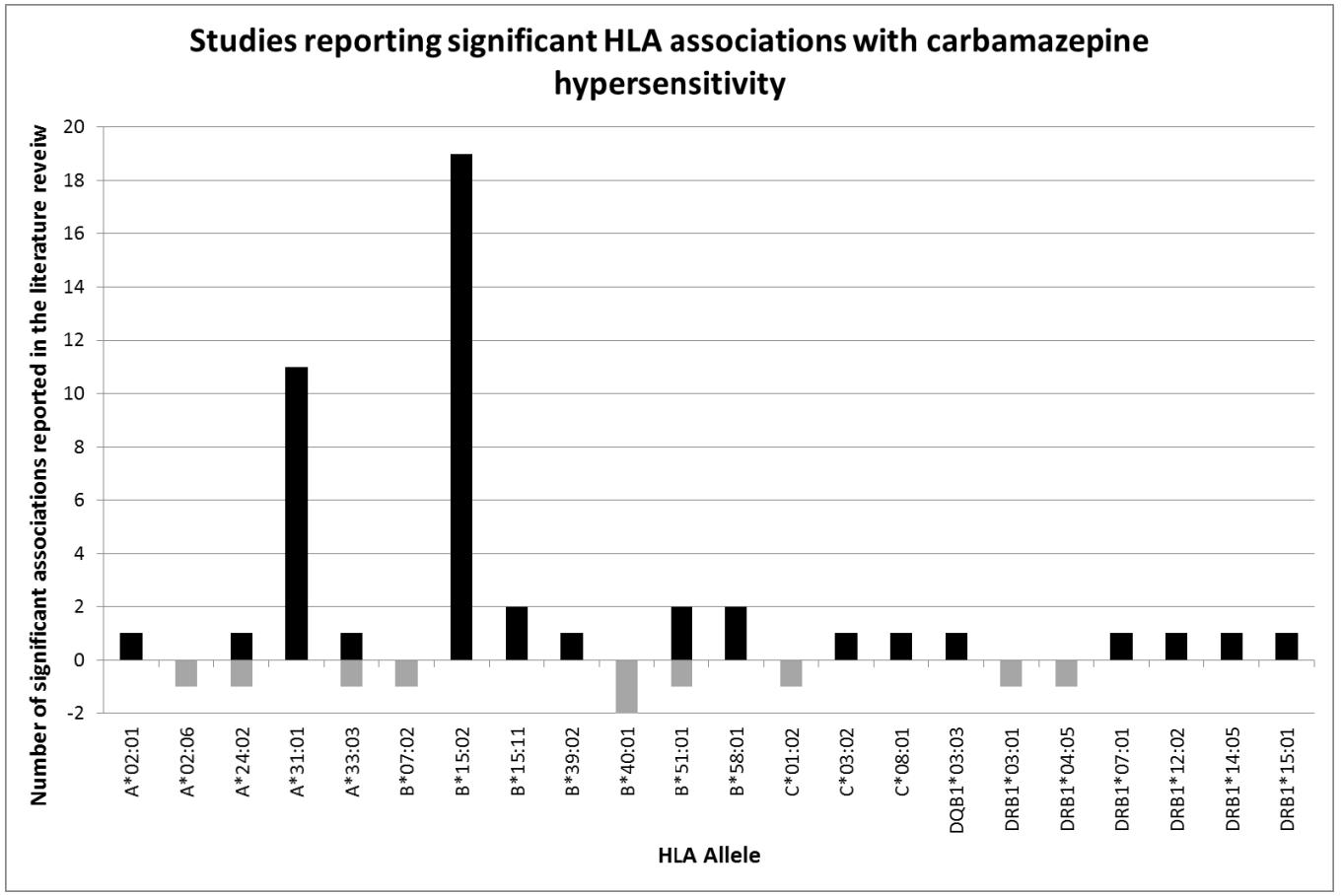


Figure 4.1: HLA alleles which have been reported in the HLA–ADR database to be significantly associated with carbamazepine-induced adverse drug reactions.

The black bars (positive *y*-axis) in Figure 4.1 represent studies that report risk alleles, that is with $p < 0.05$ and where allele was found in proportionally more hypersensitive patients as compared to the control groups(s). The grey bars (negative *y*-axis) represent studies where the reported allele is suggestive to be protective, that is with $p < 0.05$ and where fewer patients exhibiting hypersensitive reactions to carbamazepine carried the allele compared to the control group(s). Note there is some overlap with the studies – where studies have reported multiple alleles as significant. In these cases, the graph shows these as separate entries.

4.3.4.2. Oxcarbazepine

Oxcarbazepine is a structural derivative of carbamazepine and is also used as an anticonvulsant. Several HLA-B alleles have been implicated with oxcarbazepine induced hypersensitivity reactions (see Table 4.9). It is worth noting that all three studies listed in the table were conducted using Han Chinese patient cohort. However, none of the studies demonstrated a statistically significant association between the hypersensitive patients and the drug exposed controls. The significant differences are with the hypersensitive patients and the population controls (non-drug exposed). It is also worth noting that none of the studies was able to replicate an association from one of the other studies.

Table 4.9: HLA-B*15:02 alleles that have been reported to be associated with oxcarbazepine induced hypersensitivity along with the reference to the reporting study.

HLA-B allele	Reporting study	Cohort ethnicity in study
HLA-B*13:02	He *et al*., 2012 [151]	Han Chinese
HLA-B*15:27		
HLA-B*15:19		
HLA-B*27:09		
HLA-B*48:04		
HLA-B*15:02	Hu *et al*., 2011 [152]	
HLA-B*38:02	Lv *et al*., 2013 [153]	

4.3.4.3. Phenytoin

Phenytoin is another drug used in the treatment of epilepsy. Of the studies included in this review, three reported a significant association between phenytoin-induced hypersensitivity and HLA-B*15:02 and HLA-B*15:13 carriage (see Table 4.10). Whilst the table does include other alleles reported by Hung *et al*., it is worth noting that these associations are of borderline significance: $p = 0.0154$; $p = 0.0281$; $p = 0.0128$ for HLA-B*13:01, HLA-C*08:01 and HLA-DRB1*16:02 respectively. Another study by Manuyakorn *et al*. investigating phenytoin-induced hypersensitivity in Thai patients was unable to find statistical significance for the association with HLA-B*15:02. [154] However, all five investigating studies were conducted with relatively small patient cohorts ranging between 4 and 26 and therefore it can be argued that these studies are underpowered. Recently, a meta-analysis by Bloch *et al*. using the pooled HLA-B*15:02 data from four of the five listed studies (the Chang *et al* investigation was published after Bloch *et al*) reported a combined odds ratio of 3.48 ($95\%\ CI: 1.75 - 6.91$) which supports the association between HLA-B*15:02 and phenytoin-induced SJS/TEN. [33]

Table 4.10: Alleles reported to be significantly associated with phenytoin induced adverse drug reactions along with patient-control cohort ethnicity and the size of the hypersensitive/patient group size.

HLA allele	Reporting study	Hypersensitive patient group size	Cohort ethnicity in study
HLA-B*15:13	Chang *et al*., 2017 [155]	13	Malay
HLA-B*15:02	Locharernkul *et al*., 2008 [156]	4	Thai
	Cheung *et al*., 2013 [17]	15	Han Chinese
	Hung *et al*., 2010 [22]	26	
HLA-B*13:01			
HLA-C*08:01			
HLA-DRB1*16:02			

4.3.4.4. Lamotrigine

Two meta-analyses have reported the carriage of HLA-B*15:02 to be a significant risk factor for patients being treated with the anticonvulsant lamotrigine. [22, 33] The two studies needed to pool data from primary studies to achieve sufficient statistical power (see Table 4.11). The number of lamotrigine-induced SJS patients was very small, despite this, Hung et al. was able to report borderline significance for the HLA-B*15:02 association ($p = 0.03$ odds ratio $= 7.62$ (95% CI: $1.4 - 42.4$) and Bloch et al. reported a pooled odds ratio of 3.59 (95% CI: $1.15 - 11.22$).

In addition to SJS, lamotrigine-induced maculopapular exanthema (MPE) has been reported as an observed ADR phenotype in patients by Li *et al*., 2013. [21] The study reported statistical significance between the ADR and the carriage of HLA-A*30:01 ($p = 0.013$) and HLA-B*13:02 ($p = 0.013$). The haplotype frequency of these two alleles is listed below in Table 4.12for populations where the co-occurrence is proportionally at its highest; in East Asians and the Middle East (Israel). AFND does report that the haplotype can be found at lower frequencies (below 1%) in other populations including Europeans and South Asians. The study also reported a decreased occurrence of HLA-A*33:03 in lamotrigine-induced maculopapular exanthema patients compared with the control group ($p = 0.048$).

Table 4.11: List of studies from which the meta-analyses (table headings) used to conduct their respective investigations.
The meta-analyses pooled data from the listed studies in order to acquire a sufficiently powered dataset.

Bloch *et al*., 2014	**Hung *et al*., 2010**
An et al., 2010 [157]	Man et al., 2007 [20]
Hung et al., 2010 [22]	Self (own primary data) [22]
Shi et al., 2011 [158]	
Cheung et al., 2013 [17]	

Table 4.12: Populations which show the highest frequencies for haplotypes containing the HLA-A*30:01 and HLA-B*13:02 alleles along with the sample sizes from which the frequencies were obtained.
The values were obtained from the Allele Frequencies Net Database haplotype search tool (URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016) and were limited to sample sizes above 1,000 to reduce the impact of small sample sizes skewing the frequencies.

Population (country)	Haplotype frequency A*31:01 – B*13:02	Sample size
Han Chinese (China)	8.87%	3,238
Bukhara Jews (Israel)	5.38%	2,317
Han Chinese (Germany)	4.52%	1,282
Korean (USA)	2.26%	77,584
Korean (South Korea)	2.19%	4,128
Han Chinese (USA)	2.03%	99,672
Kavkazi Jews (Israel)	1.98%	2,840

4.3.5. Antiretrovirals

4.3.5.1. Abacavir

The nucleoside reverse transcriptase inhibitor, abacavir, is used in the treatment of HIV. To date, immune mediated ADRs with abacavir have only been reported to be statistically significant to the HLA-B*57:01 and HLA-DRB1*07:01 alleles (see Table 4.13). The latter allele has only reported in one study which also stated that the association was stronger with HLA-B*57:01 ($p < 0.001$ compared with $p = 0.01$).

Although the studies described in this section investigated abacavir ADRs in predominantly Caucasian patient-control cohorts, the HLA-B*57:01 allele is found across many ethnic populations which is reflected by the studies reporting the associations in non-Caucasian individuals as well. Examples to demonstrate the wide distribution of the allele is shown in Table 4.14below.

Table 4.13: List of studies present in the HLA-ADR database that have reported abacavir induced adverse drug reactions along with the study cohort ethnicity. Each instance of the term "predominantly" is defined immediately below the table.

Allele	Reporting study	Cohort ethnicity in study
HLA-B*57:01	Mallal *et al*., 2002 [107]	Caucasian (predominantly) [a]
	Hughes *et al*., 2004 [15]	Caucasian
	Martin *et al*., 2004 [16]	Caucasian (predominantly) [b]
	Stekler *et al*., 2006 [110]	Caucasian (predominantly) [c]
	Rodríguez-Nóvoa *et al*., 2007 [109]	Caucasian
	Almeida *et al*., 2008 [105]	Caucasian (predominantly) [d]
	Berka *et al*., 2012 [106]	Diverse
HLA-DRB1*07:01	Rauch *et al*., 2008 [108]	Caucasian (predominantly) [e]

[a] 18/18 abacavir ADR patients; 147/167 drug exposed controls reported as Caucasian in the study. [b] 18/18 abacavir ADR patients; 196/230 drug exposed controls reported as of European descent in the study. [c] 7/9 abacavir ADR patients; 39/41 drug exposed controls reported as Caucasian in the study. [d] 11/11 abacavir ADR patients; 7/9 drug exposed controls; 8/8 drug naïve controls reported as Caucasian in the study. [e] 138/149 of possible abacavir ADR patients; 1,451/1,728 drug exposed controls reported as Caucasian in the study - Although only 27 ADR patients were brought forward for HLA typing, the ethnicity of these 27 patients is not described.

Table 4.14: Populations where the HLA-B*57:01 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population ethnicity (country)	HLA-B*57:01 frequency	Sample size
South Asian – Indian (USA)	6.76%	185,391
South-east Asian (USA)	4.79%	27,978
Ashkenazi Jews (Israel)	4.51%	4,625
Caucasian/European (USA)	4.50%	1,245
Black/African (USA)	0.71%	416,581

4.3.5.2. Nevirapine

The non-nucleoside reverse transcriptase inhibitor, nevirapine has been associated with hypersensitivity reactions, see Table 4.15. No common allele has been reported as statistically significantly associated with nevirapine hypersensitivity. However, HLA-C*04:01 and HLA-B*35:05 are in linkage that is common to people of East and South-East Asian origin. AFND data for the co-occurrence of these two alleles is limited; it is found at a rate of 0.97% in a USA based Asian population (sample size 1,772), 0.13% in a USA based Hispanic population (sample size = 1,999) and 0.10% a USA based Black African population (sample size = 2,411). For haplotypes which contain these two alleles along with other HLA alleles the data indicates that the co-occurrence rate is higher (see Table 4.16Table 4.16), particularly so for East and South-east Asians.

Whilst Phillips *et al*. [159] reported many alleles as statistically significant, the p-values were not corrected for multiple comparisons. The study performed a multivariate analysis which specified that HLA-B*58:01 and HLA-DRB1*01:02 show an increased risk for nevirapine induced ADRs (only these alleles have been listed in Table 4.15). The co-occurrence for these two alleles is relatively low; AFND only reports frequencies of less than 0.05% in any population included in the database. The Phillips *et al*. multivariate analysis also indicated patients carrying HLA-B*15:01, HLA-B*58:02 and HLA-C*02:10 have a decreased risk to nevirapine hypersensitivity. Contrary to the Chantarangsu *et al*. [36] report Phillips *et al*. did not find statistically significant evidence to support HLA-B*35:05 as a risk factor in their ethnically diverse patient cohort ($p = 0.81$).

Table 4.15: List of studies present in the HLA-ADR database that have reported nevirapine induced adverse drug reactions along with the study cohort ethnicity.

Allele	Reporting Study	Study cohort ethnicity
HLA-C*04:01	Carr et al., 2013 [35]	Black African (Malawi)
HLA-B*35:05	Chantarangsu et al., 2009 [36]	Thai (Thailand)
HLA-B*58:01	Phillips et al., 2013 [159]	Diverse
HLA-DRB1*01:02		

Table 4.16: Frequencies for haplotypes that include HLA-B*35:05 and HLA-C*04:01 (bold) obtained from Allele Frequencies Net Database (AFND).

The AFND contains limited data for haplotype information for just these two alleles so this table contains haplotypes that also include alleles from other genes. The population ethnicity is also included as well as the sample size from which the frequencies were obtained (AFND restrictions were applied to only display frequencies where the sample sizes were greater than 1,000). The table is ordered descending for the haplotype frequencies with the final entry in the table being that of the first non-Asian entry i.e. the non-Asian population where the haplotype has been reported to be at its highest frequency. URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016.

Haplotype	Population ethnicity (country)	Haplotype Frequency	Sample Size
A*24:07-**B*35:05-C*04:01**-DRB1*12:02-DRB3*03:01-DQB1*03:01	Filipino (USA)	3.98%	50,614
A*24:07-**B*35:05-C*04:01**-DRB1*12:02-DRB3*03:01-DQB1*03:01	Vietnamese (USA)	1.18%	43,540
A*24:07-**B*35:05-C*04:01**-DRB1*15:02-DRB5*01:01-DQB1*05:02	Filipino (USA)	0.73%	50,614
A*24:07-**B*35:05-C*04:01**-DRB1*12:02-DRB3*03:01-DQB1*03:01	South-East Asian (USA)	0.31%	27,978
A*24:07-**B*35:05-C*04:01**-DRB1*12:02	Chinese (Hong Kong)	0.14%	7,595
A*68:01-**B*35:05-C*04:01**-DRB1*04:11-DQB1*03:02	European (Italy minority in Germany)	0.04%	1,159

4.3.6. Antirheumatics and NSAIDs

4.3.6.1. Aspirin

Aspirin is a widely used nonsteroidal anti-inflammatory drug (NSAID) and antirheumatic, antipyretic and analgesic agent that is available over the counter without the need of a prescription. In addition to this, low dose aspirin is often used to reduce the likelihood of cardiovascular disease and to prevent pre-eclampsia in pregnancy.

There are three studies within the HLA-ADR database which have investigated the association between HLA carriage in patients and aspirin-induced urticaria (AIU) and aspirin-intolerant asthma (AIA); shown in Table 4.17 below. All three studies were conducted with Korean case-control cohorts based in South Korea. HLA-DPB1*03:01 was found by two of the three studies. Interestingly, the allele did appear in the patient cohort for the Kim *et al*. study but the association was not significant.

The HLA-DPB1*03:01 allele is not specific to people of Korean descent (or even East Asia) and can actually be found in greater frequencies in Europe and North America (see Table 4.18) where the population with the highest frequency of HLA-DPB1*03:01 carriage is 48.5% which is remarkably high for a HLA allele. The three alleles reported by Kim *et al.* [27] were presented as a haplotype in the study. The AFND haplotype search tool does not contain information pertaining to these three alleles grouped on their own. However, haplotypes containing these three alleles along with others are included in Table 4.19 which shows that although the haplotype is not exclusive to Koreans, it is found at relatively high frequencies (1-2%) in Koreans compared to other populations.

Table 4.17: List of studies present in the HLA-ADR database that have reported HLA associations with aspirin-induced urticaria (AIU) and aspirin-intolerant asthma (AIA).

Studies featured in the table were conducted using a Korean patient-control cohort.

HLA allele	Study	ADR phenotype	Study cohort ethnicity
HLA-DPB1*03:01	Lee *et al*., 2009 [115]	AIU AIA	Korean
	Choi *et al*., 2004 [26]	AIA	
HLA-DRB1*13:02	Kim *et al*., 2005 [27]	AIU	
HLA-DQB1*06:09			
HLA-DPB1*02:01			

Table 4.18: Populations present in the Allele Frequencies Net Database which show relatively high frequencies for HLA-DPB1*03:01 carriage.

Information for populations where sample size is greater than 1,000 was limited so this restriction was relaxed to show the frequencies for samples > 100. Data obtained from Allele Frequencies Net Database - URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-DPB1*03:01 frequency	Sample size
Caucasian – Sardinian (Italy)	48.50%	1,129
Caucasian – NW Slavic (Russia)	16.0%	200
Native American (USA)	13.3%	101
Caucasian (UK)	12.93%	537
Korean (South Korea)	3.70%	324

Table 4.19: Frequencies for haplotypes that include HLA-DRB1*13:02, HLA-DQB1*06:09 and HLA-DPB1*02:01(bold) obtained from Allele Frequencies Net Database (AFND).

The AFND contains limited data for haplotype information for just these three alleles so this table contains haplotypes that also include alleles from other genes. The population ethnicity is also included as well as the sample size from which the frequencies were obtained. URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016.

Haplotype	Population (country)	Haplotype frequency	Sample size
DRB1*13:02-DQA1*01:02-**DQB1*06:09-DPB1*02:01**	Korean (South Korea)	2.10%	149
DRB1*13:02-DQA1*01:02-**DQB1*06:09-DPB1*02:01**	Korean (South Korea)	1.20%	207
A*32:01-B*07:02-C*07:02-**DRB1*13:02**-DQA1*01:02-**DQB1*06:09-DPB1*02:01**	African (Kenya)	0.50%	100
A*66:01-B*58:02-C*06:02-**DRB1*13:02**-DQA1*01:02-**DQB1*06:09-DPB1*02:01**	African (Kenya)	0.50%	100
A*33:03-B*58:01-C*03:02-**DRB1*13:02**-DQA1*01:02-**DQB1*06:09**-DPA1*01:03-**DPB1*02:01**	Japanese (Japan)	0.13%	3,078
A*24:02-B*58:01-C*03:02-**DRB1*13:02**-DQA1*01:02-**DQB1*06:09**-DPA1*01:03-**DPB1*02:01**	Japanese (Japan)	0.03%	3,078

4.3.6.2. Bucillamine

Another antirheumatic agent, bucillamine, has been reported to cause proteinuria in Japanese patients. [160] The reporting study found an association between bucillamine induced proteinuria and the alleles: HLA-DRB1*08:02 ($p = 1.96 \times 10^{-5}; OR = 25.17, 95\%\ CI: 7.98 - 79.38$) and HLA-DQB1*04:02 ($p = 2.69 \times 10^{-4}; OR = 10.35, 95\%\ CI: 2.99 - 26.83$). The study reported that due to high linkage between the two alleles, they were unable to reach the threshold for significance when attempting to separate the two alleles during analysis to determine which was the more probable allele for being the risk factor for bucillamine induced ADRs. The AFND haplotype search tool is able to support the high co-occurrence for the two alleles (Table 4.20). The table indicates that that allele is particularly prevalent in indigenous North American populations.

Table 4.20: Frequencies for haplotypes within the Allele Frequencies Net Database containing the HLA-DRB1*08:02 and HLA-DQB1*04:02 alleles.
The population ethnicity and the sample sizes are also included. URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016

Population (country)	Haplotype frequency DRB1*08:02 – DQB1*04:02	Sample size
Latin American (Mexico)	21.60%	103
Latin American (Mexico)	19.02%	234
Alaska Yupik (USA)	13.30%	252
Japanese (Japan)	10.00%	50
Hispanic (USA)	5.61%	1,999
Caucasian (Sweden)	3.00%	154
Japanese (Japan)	2.20%	371
Caucasian (Sweden)	1.20%	130
Asian[1] (USA)	0.53%	1,772
Caucasian (Italy)	0.16%	975

1. Presumed to refer to individuals of East Asian origin.

4.3.6.3. Lumiracoxib

One study reported that lumiracoxib, a nonsteroidal anti-inflammatory drug, can cause hepatotoxicity associated with a common HLA haplotype: [HLA-DRB1*15:01, HLA-DQB1*06:02, HLA-DRB5*01:01, HLA-DQA1*01:02]. [161] Interestingly, the study noted that this haplotype has also been associated with multiple sclerosis. AFND has limited information for HLA-DRB5 as this gene is not routinely typed. AFND only has one population group in which the individuals have a haplotype containing these alleles. However, they also contain HLA-DPB1 alleles and are listed below in Table 4.21Table 4.21.

It is also worth noting that the HLA-DRB1*15:01 allele has also been associated as a risk factor for amoxicillin-clavulanate as mentioned earlier in this manuscript. Amoxicillin-clavulanate and lumiracoxib are different classes of drugs, so it is unlikely that the two drugs interact with HLA-DRB1*15:01 protein using the same mechanisms.

4.3.6.4. Sulfasalazine

Inflammatory bowel diseases and rheumatoid arthritis can be treated with sulfasalazine. The drug has been reported to induce hypersensitivity reactions in Han Chinese patients carrying HLA-B*13:01, HLA-B*15:05 and/or HLA-B*39:01. [162] HLA-B*13:01 is a commonly occurring allele in the South-East Asian and Australasian region (see Table 4.22). Conversely, HLA-B*15:05 is a very low occurring allele, less than 1% for sample sizes greater than 1,000 with AFND reporting the East and South Asian populations showing the greatest allele frequency (see Table 4.23). HLA-B*39:01 has been reported in Japanese, Native Americans, Europeans and Han Chinese individuals (see Table 4.24Table 4.24). Greater frequencies are observed when sample size restrictions are relaxed with the highest frequency being reported in several Taiwanese populations: Saisiat ethnic group at 54.9% and Tsou ethnic group at 24.5%.

Table 4.21: Frequencies for haplotypes that contain the HLA-DRB1*15:01, HLA-DQB1*06:02, HLA-DRB5*01:01 and HLA-DQA1*01:02 obtained from Allele Frequencies Net Database (AFND).

The AFND haplotype search tool has limited information for this haplotype so the HLA-DPB1 alleles were also included. All entries in this table contain HLA-DRB1*15:01, HLA-DQB1*06:02, HLA-DRB5*01:01 and HLA-DQA1*01:02 and only differ by the HLA-DPB1 allele for which the frequencies are included. For all haplotypes, the population groups are actually the same group.

URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016

Haplotype					Frequency	Population	Sample Size
HLA-DRB1	HLA-DRB5	HLA-DQA1	HLA-DQB1	HLA-DPB1			
15:01	01:01	01:02	06:02	04:01	9.3%	Caucasian (USA)	220
				02:01	2.1%		
				04:02	1.8%		
				23:01	1.1%		

Table 4.22: Frequencies obtained from Allele Frequencies Net Database (AFND) for the HLA-B*13:01 allele and the populations where it is found at relatively high frequencies.

The first three entries in the table are for the top entries where the sample size was smaller than 1,000 and show that the allele is present in populations outside of East Asia. The bottom four entries are the top results from AFND where the sample size was greater than 1,000. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-B*13:01 frequency	Sample size
Melanesian (Papua New Guinea)	28.30%	131
East Asian (Taiwan)	28.00%	50
Aborigine (Australia)	27.00%	103
Han Chinese (Hong Kong)	7.94%	7,595
Han Chinese (Hong Kong)	7.75%	3,892
Han Chinese (USA)	6.35%	99,672
Vietnamese (USA)	3.94%	43,540

Table 4.23: Populations where the HLA-B*15:05 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-B*15:05 frequency	Sample size
South Asian/Indian (USA)	0.45%	185,391
South-East Asian (USA)	0.33%	27,978
"Asian" (USA)[1]	0.28%	1,772
Han Chinese (Germany)	0.12%	1,282
Japanese (Japan)	0.10%	1,018

1. Presumed to refer to East Asian

Table 4.24: Populations where the HLA-B*39:01 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-B*39:01 frequency	Sample size
Japanese (Japan)	3.40%	1,018
Native American (USA)	3.39%	35,791
Japanese (Japan)	3.32%	18,604
Japanese (USA)	3.03%	24,582
Caucasian (Croatia)	2.56%	4,000
Caucasian (Germany)[1]	2.24%	1,028
Han Chinese (USA)	1.93%	99,672

1. Minority Bosnia and Herzegovina population residing in Germany

4.3.7. Antipsychotics

4.3.7.1. Clozapine

Clozapine is an antipsychotic agent used to treat patients with treatment-resistant schizophrenia. Clozapine-induced agranulocytosis has been reported in two studies by the same group which likely used the same patient and control groups of European origin. [116, 163]The initial study limited the investigation to HLA-B, HLA-DRB and HLA-DQB typing, whilst the full investigation typed these loci and also the HLA-C, HLA-DPB as well as the HLA-DRB1, DRB3, DRB4 and DRB5 loci. The authors reported HLA-DQB1*05:02, DRB5*02, HLA-C*07 and HLA-DPB*04:01 were associated with clozapine-induced agranulocytosis. However, their reporting of HLA was variable as they employed old and current HLA nomenclature as well as a mix of high- and low-level resolution typing which limits the ability to incorporate these results within a meta-analysis as the allele and resolution data are not comparable with data from other studies.

The only allele reported as a risk factor for a clozapine-induced adverse response typed to two-field resolution was HLA-DQB1*05:02 ($p = 0.006$ and $OR = 15.4, 95\%\ CI: 1.72 - 138$). Table 4.25 shows the allele frequencies for HLA-DQB1*05:02 in populations where the allele is at its highest prevalence. The other significant associations cannot be determined to their specific allele product (two-field resolution). A recent GWAS also reported that a SNP that is in strong linkage with HLA-DQB1*05:02 is a risk allele for agranulocytosis to clozapine. [164] This study also reported that a SNP located in HLA-B played an important role, and the HLA-B*38, HLA-B*39 and HLA-B*67 alleles were offered as candidates to investigate further.

Table 4.25: Populations where the HLA-DQB1*05:02 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-DQB1*05:02 frequency	Sample size
Caucasian (Italy – Sardinia)	10.60%	1,129
"Asian" (USA)[1]	7.35%	1,772
Turkish (Germany)	6.63%	4,856
Vietnamese (USA)	5.77%	43,540
Black (USA)	2.73%	2,411

1. Presumed to refer to East Asians

4.3.8. Antithrombotic drugs

4.3.8.1. Ticlopidine

Ticlopidine is a drug which is used for preventing thrombotic stroke in patients who are unable to take aspirin, for example the above-mentioned AIU and AIA, or if aspirin has previously failed to adequately treat the patient for a stroke. Ticlopidine-induced hepatotoxicity has been reported in Japanese patients carrying HLA-A*33:03, HLA-B*44:03, HLA-C*14:03, HLA-DRB1*13:02 and HLA-DQB1*06:04. [165] The haplotype containing all these listed alleles is the second most commonly occurring in

the Japanese populations and was present in 10 out of the 14 hypersensitive individuals of the reporting study.

4.3.9. Anticoagulants

4.3.9.1. Ximelagatran

Ximelagatran is an anticoagulant which was withdrawn from distribution in 2006 due to reported elevation of liver enzyme levels and hepatic damage in patients. [166] A study by Kindmark *et al*. found an association between HLA-DRB1*07:01 and HLA-DQA1*02 with ximelagatran based DILI in a Northern European patient-control cohort. The study also reported that HLA-DQB1*02 was associated as 65% of the cases which carried the DRB1*07-DQA1*02 haplotype also possessed a DQB1*02 allele. [118] The AFND haplotype frequency search tool shows that this haplotype can be found in many geographic regions including Africa, Asia, Australia (Aborigine) Europe and North America; this is likely due to the HLA-DQA1 and HLA-DQB1 alleles only being resolved to the 1st field so is therefore extremely broad for the search tool.

4.3.10. Anticancer drugs

4.3.10.1. Lapatinib

Lapatinib is used in the treatment of solid tumours including breast cancer and lung cancer. Lapatinib induced hepatotoxicity has been reported in three HLA class II alleles: HLA-DQA1*02:01, HLA-DQB1*02:02 and HLA-DRB1*07:01. [28, 167] The two reporting studies were each conducted with ethnically diverse patients. The distribution of the three alleles follows this pattern as they all have a wide geographic distribution which is not specific to a particular region in the world or ethnic group as shown in Table 4.26. A recent investigation identified HLA-DRB1*07:01 to be strongly associated with lapatinib based DILI and did not find a similar association with the other alleles of the above mentioned haplotype. [168] It is also worth noting that HLA-DRB1*07:01 and serotypes that HLA-DQA1*02:01 and HLA-DQB1*02:02 belong to have also been associated with ximelagatran ADRs despite the dissimilar structures between the drugs. HLA-DRB1*07:01 is a commonly occurring allele not specific to a particular geographic region (see Table 4.27).

Table 4.26: Frequencies for haplotypes within the Allele Frequencies Net Database containing the HLA-DQA1*02:01, HLA-DQB1*02:02 and HLA-DRB1*07:01 alleles. The population ethnicity and the sample sizes are also included. URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016

Population (country)	Haplotype frequency	Sample size
North African (Morocco)	16.70%	98
Caucasian (USA)	11.08%	1,899
North African (Tunisia)	10.20%	100
Caucasian (USA)	8.90%	220
Korean (South Korea)	6.60%	467

Table 4.27: Populations where the HLA-DRB1*07:01 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-DRB1*07:01 frequency	Sample size
Morocco Jewish (Israel)	23.86%	36,718
Libya Jewish (Israel)	23.32%	3,739
Tunisia Jewish (Israel)	17.98%	9,070
Yemen Jewish (Israel)	17.84%	15,542
South Asian – Indian (USA)	16.95%	185,391
Caucasian (Germany)[1]	15.94%	1,176
Han Chinese (China)	15.80%	3,238

1. Portuguese minority residing in Germany

4.3.11. Ocular glaucoma drug

4.3.11.1. Methazolamide

The carbonic anhydrase inhibitor, methazolamide is used to treat glaucoma. The drug has been reported to be associated with SJS/TEN in Han Chinese patients. [169] The study reported the associations to be with the haplotype consisting of HLA-B*59:01 and HLA-C*01:02. When the study looked at the individual alleles, they reported that

the association with HLA-B*59:01 was statistically more significant ($p = 6.3 \times 10^{-7}; OR = 305.0; 95\%\ CI: 11.3 - 8259.9$) compared to HLA-C*01:02 ($p = 1.6 \times 10^{-2}; OR = 12.1; 95\%\ CI: 1.3 - 111.7$). The HLA-C*01:02 allele is commonly occurring in Chinese and Japanese populations (see Table 4.28). The allele is also found at lower frequencies in Latin American and Europeans (≤ 5.7% where sample sizes are greater than 1,000). The HLA-B*59:01 allele is found at much lower frequencies as shown in Table 4.29. Although the table itself lists East Asian populations, the allele is found at frequencies that approach zero in other populations such as Native American, South Asian, Caribbean Black, Latin Americans and Europeans. The difference in the significance values between the two alleles can be explained as HLA-B*59:01 was reported to be carried by none of the drug-exposed controls and HLA-C*01:02 was found in 11 of 30 controls. As such, the study reported that HLA-B*59:01 is the most likely candidate for the risk factor for methazolamide-induced SJS/TEN.

Table 4.28: Populations where the HLA-C*01:02 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-C*01:02 frequency	Sample size
Han Chinese (China)	20.17%	1,098
Han Chinese (USA)	19.15%	99,672
Han Chinese (Hong Kong)	18.99%	7,595
Japanese (Japan)	18.20%	1,018
Japanese (Japan)	17.60%	18,604
Korean (USA)	16.77%	77,584
Vietnamese (USA)	16.73%	43,540

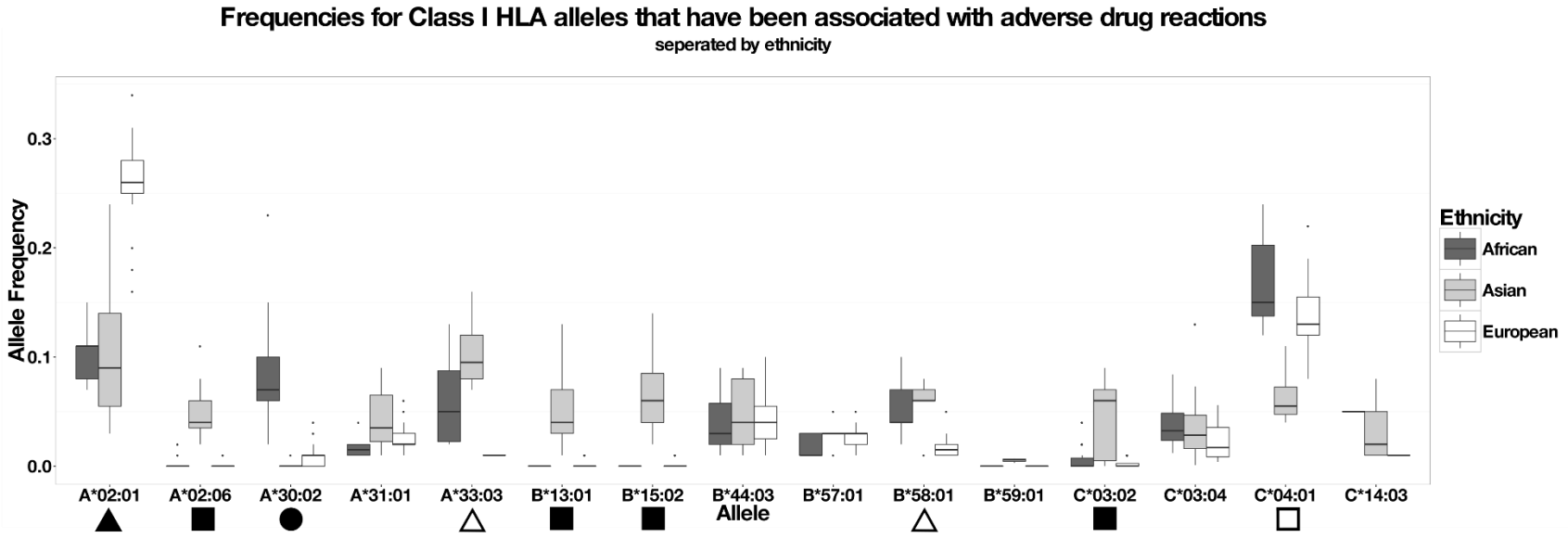
Table 4.29: Populations where the HLA-B*59:01 allele is found at relatively high frequencies where the sample size of the population was greater than 1,000. These frequencies were obtained from the Allele Frequencies Net Database. URL http://www.allelefrequencies.net/hla6006a.asp accessed June 2016.

Population (country)	HLA-B*59:01 frequency	Sample size
Japanese (USA)	2.24%	24,582
Japanese (Japan)	2.04%	18,604
Korean (South Korea)	1.85%	4,128
Korean (USA)	1.67%	77,584
Han Chinese (China)	0.30%	3,238
Han Chinese (USA)	0.05%	99,672

4.3.12. Analysis of all data

To investigate whether any observable trends can be determined in responses between different ethnic groups, the information was analysed by producing a table of descriptive statistics (see USB/CD media file "chapter_4_descriptive_statistics"), a Kruskal-Wallis test (Table 4.30) and boxplots (Figure 4.2) of the frequencies for the alleles which have been associated with ADRs. A limitation was applied to only included strong associations (where $p \leq 0.01$) and/or have been independently replicated in two or more studies were included for this analysis. In these analyses, the information has been grouped by the three major ethnic groups (as described in the methods section). Across the range of drugs featured in this chapter, there is an apparent trend whereby more of the alleles that have been associated with adverse events have higher allele frequencies in Asians. The results show that 6 of 33 HLA alleles which show significant associations ($p \leq 0.01$) have statistically higher allele prevalence in Asians compared to two in Africans and two in Europeans. This is particularly noticeable in patients prescribed with antiepileptic drugs where 43% of the alleles associated with carbamazepine hypersensitivity, 33% of the alleles associated with lamotrigine ADRs and 67% of alleles associated with phenytoin ADRS are found in higher frequencies in Asians. For these three drugs, the associations with alleles that show strong evidence/well replicated were either associated with Asians or were neutral (i.e. not specific to a particular ethnic group). None of the alleles can be classed as strongly associated with Europeans or Africans.

a)



▲ Allele significantly more common in Europeans (p > 0.05)

△ Allele significantly less common in Europeans (p > 0.05)

■ Allele significantly more common in Asians (p > 0.05)

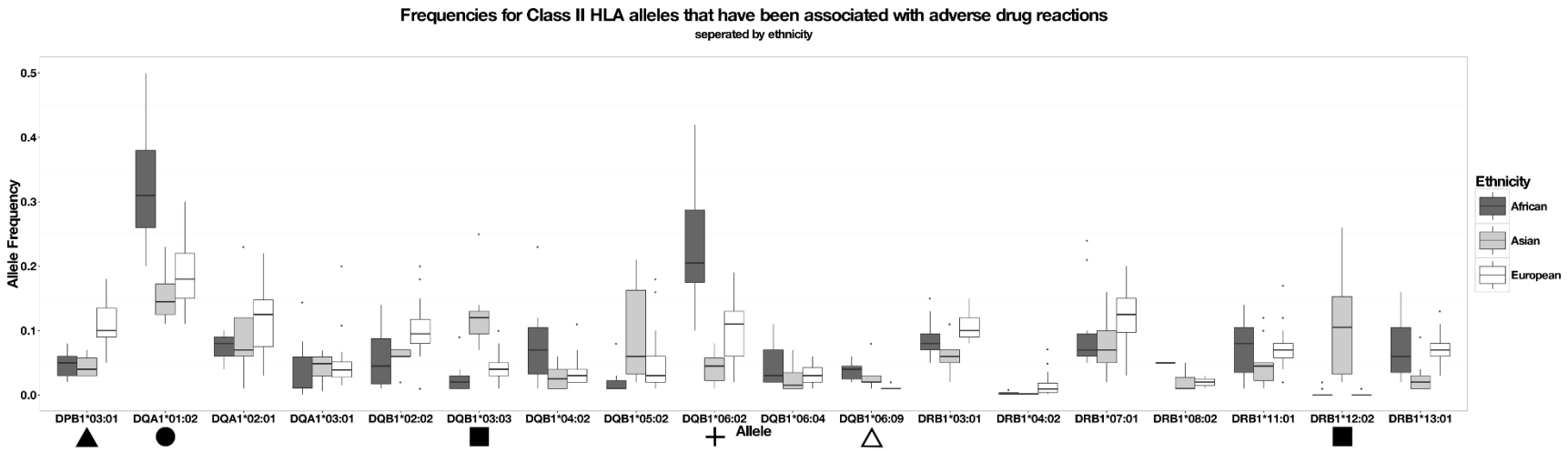
□ Allele significantly less common in Asians (p > 0.05)

● Allele significantly more common in Africans (p > 0.05)

+ Allele significantly different between all groups (p > 0.05)

Figure 4.2: a) Boxplots showing the frequencies for class I HLA alleles that have been reported in this chapter to have a strong association with adverse drug reactions ($p < 0.01$) and/or have been reported independently by two or more studies.

b)



▲ Allele significantly more common in Europeans (p > 0.05)

△ Allele significantly less common in Europeans (p > 0.05)

■ Allele significantly more common in Asians (p > 0.05)

□ Allele significantly less common in Asians (p > 0.05)

● Allele significantly more common in Africans (p > 0.05)

+ Allele significantly different between all groups (p > 0.05)

Figure 4.2: b) Boxplots showing the frequencies for class II HLA alleles that have been reported in this chapter to have a strong association with adverse drug reactions ($p < 0.01$) and/or have been reported independently by two or more studies.

Table 4.30: Kruskal–Wallis test displaying the level of significance for the differences in allele frequencies between the three major ethnic groups (Africans, Asians and Europeans).

The selected alleles are those that have been reported in this chapter to be significantly associated with an adverse drug reaction. The allele frequencies were obtained from Allele Frequencies Net Database (AFND) for various subpopulations that belong to the major ethnic groups. The Kruskal–Wallis test was applied using the statistical package r version 3.2.2 where Bonferroni correction for multiple comparisons was applied. Comparisons that are significant ($P < 0.05$) after correction for multiple testing have been highlighted in the boxplot shown in Figure 4.1.

Allele	Africans vs Asian	African vs European	Asian vs European
A*02:01	1	6.6×10^{-10}	8.25×10^{-10}
A*02:06	6.6×10^{-15}	1	6.6×10^{-15}
A*30:02	1.716×10^{-09}	5.61×10^{-08}	0.462
A*31:01	0.2838	0.1439	0.2319
A*33:03	0.3036	3.1×10^{-6}	9.9×10^{-10}
B*13:01	6.6×10^{-15}	0.62	6.6×10^{-15}
B*15:02	6.6×10^{-15}	0.36	6.6×10^{-15}
B*44:03	1	1	1
B*57:01	0.37	0.21	1
B*58:01	1	2.904×10^{-5}	3.0×10^{-5}
B*59:01	0.49	1	0.49
C*03:02	0.0363	1	0.00396
C*03:04	1	0.37	1
C*04:01	8.6×10^{-5}	0.14	0.001353
C*14:03	1	0.97	1
DPB1*03:01	1	2.2×10^{-5}	5.6×10^{-4}
DQA1*01:02	0.00495	0.0132	0.4328
DQA1*02:01	1	0.1	0.78
DQA1*03:01	0.93	0.82	1

Allele	Africans vs Asian	African vs European	Asian vs European
DQB1*02:01	1	0.48	0.17
DQB1*03:03	3.6×10^{-6}	0.075	6.9×10^{-5}
DQB1*04:02	0.065	0.255	0.888
DQB1*05:02	0.066	0.11	0.084
DQB1*06:02	2.871×10^{-7}	0.002343	0.0132
DQB1*06:04	0.15	1	0.39
DQB1*06:09	0.09387	7.6×10^{-5}	0.01617
DRB1*03:01	0.4835	1	0.0198
DRB1*04:02	1	0.429	0.132
DRB1*07:01	1	0.145	0.825
DRB1*08:02	0.54	0.86	1
DRB1*11:01	0.21	1	0.14
DRB1*12:02	6.6×10^{-15}	1	6.6×10^{-15}
DRB1*13:01	0.44979	1	0.01386

4.4. Discussion

Whilst from a general perspective, there is an abundance of data with regards to the reported ethnic response and ADRs, little can be said with absolute certainty. The primary reason is that there are inconsistencies in the way the data are reported. Some of the reported alleles are not well replicated and appear to be statistically significant in some studies whilst non-significant in others - the results section in this chapter highlight this where examples can be seen with carbamazepine and nevirapine. A likely reason for the inconsistency can be argued by the relatively small patient-control cohorts from which the studies are conducted. These small studies are prone to the effects where the data from a few patients can greatly influence the statistical analyses as a single patient can represent a larger proportion of the available data. The small patient-control cohort issue coupled with HLA allele linkage disequilibria as featured in this chapter also prevents the risk allele/biomarker from being identified. In these cases, only the haplotype can be reported as the risk factor (e.g. in the case of aspirin-induced urticaria and the haplotype containing the alleles: DRB1*13:02–DQB1*06:09–DPB1*02:01). [27] However, overcoming this issue of small patient-control cohort size is easier said than done since specific ADRs are rare and therefore

studies have difficulty in obtaining a sufficient number of ADR patients for an adequately powered study.

Additionally, there has been few studies published which have demonstrated how the drug-allele pairs interact (e.g. the report by Illing *et al*. showing how abacavir interacts with HLA-B*57:01) and the majority of the reported evidence still relies on association studies. [46, 170] This not only restricts the understanding for the biological mechanisms that cause ADRs, it also forces the majority of investigations to be reliant on these association studies.

Looking at specifics, the association between anti-tuberculosis drugs and the carriage for HLA-C*04:01 in terms of patient ADRs, the number of ADR patients (n = 14) is low and has not been replicated by other groups therefore, any statement with regards to the association cannot be made with confidence. In the case with nevirapine, the three studies featured in the results section were unable to replicate the findings of each other where an allele reported as significantly associated by one study but was not found to be significant in another study even when the allele was carried by members of the patient-control cohort. With these studies, Carr *et al*. reported HLA-C*04:01 as a risk factor however, Chantarangsu *et al*. reported the risk factor to be HLA-B*35:05 and found HLA-C*04:01 to be non-significant after statistical correction. [35, 36] Similarly, Phillips *et al*. found HLA-B*35:05 to be non-significant in their study. [159] With all this conflicting evidence, further studies are required to investigate the situation for nevirapine induced ADRs.

For the other antiretroviral abacavir, this review only focused on the allele frequencies for HLA-B*57:01 and not DRB1*07:01 as it has been demonstrated that abacavir directly interacts with the F-pocket located in the antigen-binding cleft of the HLA-B*57:01 allele and only one of the featured studies reported HLA-DRB1*07:01 as a co-occurring allele. [46] In this instance, the evidence is overwhelmingly in favour of HLA-B*57:01 as a sole risk factor for abacavir induced ADRs.

In other cases, it is often difficult to pinpoint the exact risk factor as due to the limited number of studies available for a particular drug-ADR, in some cases this may be only one study. In such reports, as exampled with ticlopidine, the risk factor can only be

resolved to a haplotype but not any of the specific alleles. The strong linkage between some alleles can create obstacles in research for determining the causative allele. The ideal way to overcome this is to conduct a comparable investigation in another population where the ADR is observed, but the haplotype is not common. However, such scenarios can be rare to achieve.

As showcased in Chapter 2 there are issues with publication bias where these types of studies are conducted and are limited with East Asia and Europe representing the majority of the data present in the HLA-ADR database, only a handful of publications are produced from other regions. This can limit the impact of declaring certain ADRs to be more prevalent in patients of particular ethnic groups. For example, to date, bucillamine induced ADRs has only been implicated in Japanese patients. However, the results of this review indicate that the associated haplotype is also found in other populations such as indigenous North Americans where very little information regarding immune-mediated ADRs is available. It is possible that these populations are at greater risk of experiencing bucillamine related ADR events compared to the baseline ADR incidence rate. Conversely, a situation where the haplotype is in lower frequency for a population but one of the alleles is present would present a good opportunity to conduct a study to isolate the true risk allele. For example, HLA-DQB1*04:02 is present in many populations of recent African origin but does not co-occur with HLA-DRB1*08:02 at a high rate (see Table 4.31Table 4.31). If a study is conducted using this population, it could help separate the risk allele from the co-occurring allele.

There are also different responses to similar class drugs. Both nevirapine and abacavir are used in the treatment of HIV however, different alleles have been implicated with their respective ADRs. However, the difference in response is likely due to the difference in chemical structure and method of reverse transcriptase inhibition (one is a nucleoside analogue reverse transcriptase inhibitor and the other is a non-nucleoside reverse transcriptase inhibitor). The chemical structure also likely explains why many of the antiepileptics have HLA-B*15:02 in common as an associated risk factor as their chemical structures are similar. This similarity in structure suggests a similar mechanism of action which triggers the hypersensitivity reactions with the possibility of cross-reactivity between the antiepileptic drugs. [33]

Table 4.31: Allele frequencies net database haplotype search where HLA-DQB1*04:02 was selected and “any” for the other HLA genes. The search was further restricted to patients of Black/African origin where the sample size was greater than 1,000. The table shows the other alleles that co-occur with HLA-DQB1*04:02 and the corresponding haplotype frequency is included. URL http://www.allelefrequencies.net/hla6003a.asp accessed June 2016.

Haplotype	Population	Haplotype frequency	Sample size
DRB1*03:02-**DQB1*04:02**	Black (USA)	5.94%	2,411
A*30:01-B*42:01-C*17:01-DRB1*03:02-DRB3*01:01-**DQB1*04:02**	Caribbean Black (USA)	1.60%	33,328
A*30:01-B*42:01-C*17:01-DRB1*03:02-**DQB1*04:02**	Black (USA)	1.52	2,411
A*30:01-B*42:01-C*17:01-DRB1*03:02-DRB3*01:01-**DQB1*04:02**	Black (USA)	1.48	416,581

The manifestation of hypersensitivity reactions to antiepileptic drugs has been reported to be approximately five times greater in patients who have previously exhibited ADRs with a different antiepileptic drug. However, the rate of cross-reactivity varied between different drugs. [171] The reporting study also recommended against prescribing phenytoin, carbamazepine and lamotrigine (two of which have ADRs associated with HLA-B*15:02) to patients with a history of antiepileptic-induced ADRs. All three of these drugs also have ADRs associated with HLA alleles that have higher frequencies in Asian populations. High cross-reactivity has been reported in Chinese patients that were prescribed carbamazepine, oxcarbazepine, phenytoin and lamotrigine. [172, 173] The precise mechanisms which cause cross-reactivity of antiepileptic drugs are not fully understood therefore, additional investigations which look to elucidate the mechanism of action would be required to confirm this hypothesis. Whilst the Kruskal–Wallis analysis of population control data does show a skewing towards Asians, more data are required to form robust conclusions. The observed skewing towards Asians is possible because many drugs particularly in the past were likely developed in Europe and/or North America and where clinical trials were conducted with participants of European descent.

Some of the associations for the drugs featured in this chapter have been reported with both class I and class II HLA alleles. However, there is limited haplotype data available for alleles covering both class I and class II alleles because most of the available data focus on each class type. Of the 91 studies that are featured in systematic review of Chapter 2, 56 studies focus only on class I HLA alleles, 18 studies focused solely with class II HLA alleles and only 17 studies typed patients across both of these classes. These restrictions pose as obstacles for a full understanding of the relationship that alleles in high linkage have in the context of ADRs. The solution to this issue would be to sequence/type patients across the entire MHC region allowing all of the HLA allele data to become available. Recently, there have been techniques that have been developed which aim to produce complete phasing of the MHC region but the technology is still in its infancy. [174]

The findings of this chapter also suggest that the different prescribing decisions may require revaluation for each drug–ADR combination and should take into account the

patient's ethnicity (a proxy to give an indication as to the likelihood of the patient carrying a risk allele). For example, whilst the analysis of all the data indicates that both the HLA-A*31:01 and HLA-B*15:02 alleles have been strongly associated with carbamazepine hypersensitivity, evidence is stronger to support pre-treatment genotyping of patients for HLA-B*15:02 in Asians before prescribing the drug. [175] Comparatively, HLA-A*31:01 is an allele that can be found throughout many populations across the world at fairly low frequencies.

Chapter 5. *Validation of a new HLA typing method based on a panel of select HLA alleles associated with ADRs*

5.1. Introduction

Historically, HLA typing of samples was performed using a serological based approach. However more recently, HLA typing is performed using one of three main molecular based methods. These are: sequence-specific primer assay (SSP), sequence-specific oligonucleotide (SSO) assay and sequence based typing (SBT) which are described in Chapter 1.3. One of the biggest challenges faced in HLA typing is the difficulty in acquiring the phase of the polymorphisms- that is to say due to the high level of heterozygosity and diversity in the MHC region, using the non-next generation sequencing techniques mentioned above across the entire HLA locus has poor reliability. [176, 177] This is further complicated as the multiple HLA genes within the region share some degree of sequence homology which means that attempting to overlap the sequence reads in an attempt to determine phase will be near impossible to obtain unambiguously. [178]

5.1.1. General principle of the HISTO MATCH typing method

The HISTO SPOT HLA typing methods that were tested as part of this project involved an SSO based system where the assay tests for six HLA gene loci. The six HLA loci are: HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1 and HLA-DQB1. The test principle of the procedure is shown below in Figure 5.1. The procedure involves extracting DNA from the patient sample which is then amplified via PCR (PCR protocols are shown in Methods section) – the amplification is locus specific (MHC/HLA region). The PCR plate is then placed within the MR SPOT machine along with the assay reagents and the reaction wells which contain the SSO probes that are fixed at the bottom. The SSO probes are designed to be specific to certain sequences for particular HLA alleles. Any amplified DNA that is bound to the SSO probes trigger an enzyme mediated colourimetric reaction which appear as dark spots at the bottom of the well. The pattern of dark spots is then interpreted by the HISTO MATCH software (included with the assay equipment) and if the pattern of probes correlates with that of one or more predetermined alleles then these are highlighted as

risk alleles to the user in the final report. In this context the risk alleles are defined in section 5.1.2.

This chapter features two variants of the MR SPOT kit, the first method which will be referred to as a "one well system" and is a test where the risk alleles for all five tested genes/loci are tested for in one well. The other variant of the kit is referred to as a "five well cartridge system" where the five HLA loci are each tested in separate wells contained within one cartridge (see Figure 5.2). The cartridges are actually composed of eight wells where the first five wells each correlate to an HLA loci whilst the remaining three are left unused but are left available for future expansions to the kit. With the one well system, the MR SPOT machine can accommodate 12×8 well strips (a total of 96 samples can be tested in a single run) whilst the five well cartridges system can hold a total of six cartridges so only six patients can be HLA typed in a single run.

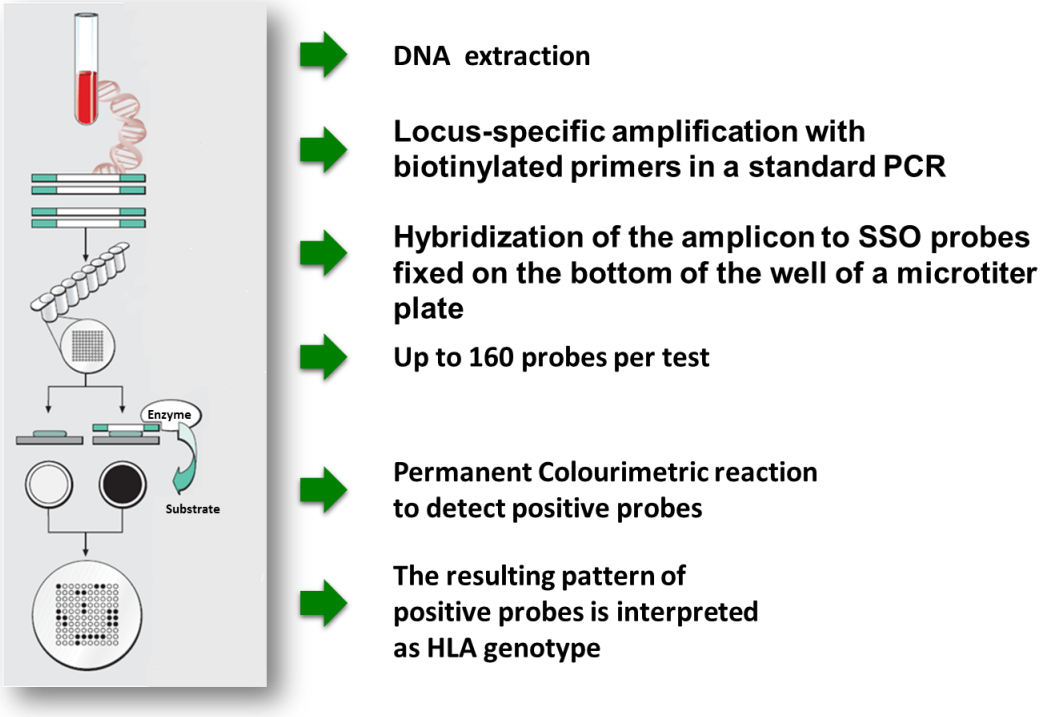


Figure 5.1: Test principle showing work path / procedure steps of the HISTO SPOT assay.

Sourced from MC Diagnostic

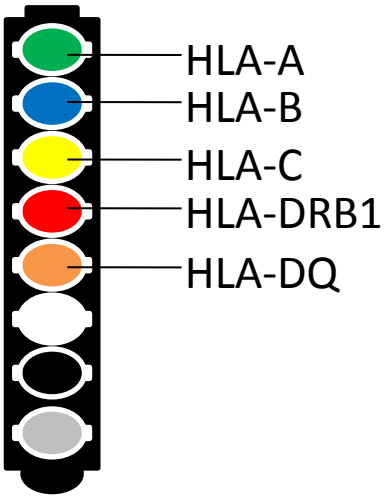


Figure 5.2: Diagrammatic representation to show the SSO probes distribution based on HLA loci.

5.1.2. Determination of panel of HLA alleles

The predetermined alleles are a panel of over 20 HLA alleles which have been reported to cause HLA mediated ADRs (see Figure 5.3). [179] The panel was designed around alleles that showed significant evidence to be associated with ADRs where the alleles were typed to at least 2-field resolution. As part of the data analysis performed in 0, the inclusion/exclusion of some of the alleles were re-evaluated and subsequently, the panel has since been modified – the details of the changes and reasons are discussed in that chapter.

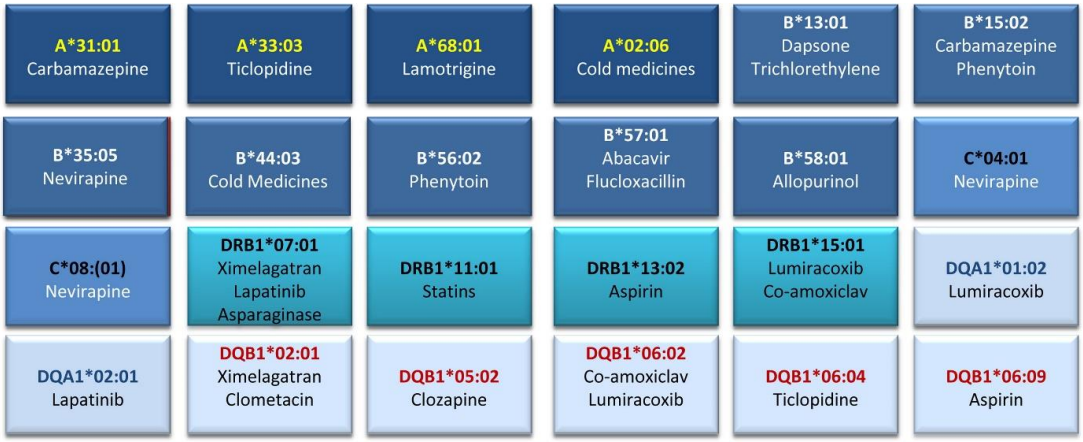


Figure 5.3: Panel of 24 alleles which were initially reported by the HISTO MATCH software.

From Pirmohamed *et al.*, 2015.

5.1.3. HISTO MATCH software

5.1.3.1. Worklist creation

The HISTO MATCH software is provided by the manufacturer to their customers who have purchased the lab equipment. It is designed to prepare the initial worklist that allows the test to run and also interpret the results (the colourmetric probes pattern) from the test assay. The user must first, as a requirement, create a worklist detailing the layout of the samples into the software. This can be achieved in two ways: first, the user can create a comma separated value (CSV) list whereby the samples are on separate lines and the sample ID and program type are separated by a comma. The sample ID is for the benefit of the user to refer back to the samples using their own ID system that is in place and in the program type allows the HISTO MATCH interpretation software gain context for the samples. In the one well system, the program type is "ADR" for all samples whilst in the five well cartridge system, the program type is the HLA gene locus being tested in the reaction assay (e.g. "A" or "B" etc). The order of the samples is important, relative to the PCR plate where the well designated as A1 is located at the top left corner, the next sample must refer to the well B1, then C1 etc – i.e. the order is top-down, left to right; this ensures that the software refers to the correct sample.

Once the CSV file is created, the file can be loaded into the HISTO MATCH software where the program will generate the XML worklist file that the MR SPOT equipment requires for the assay. The second method to generate an XML worklist is done within the MR SPOT software itself where in the absence of an XML document from HISTO MATCH, the machine will automatically generate a worklist and preload the sample IDs with 1 until the end number of samples that the assay can support (96 for the one well assay and 6 samples for the five well assay system) along with the appropriate program type. After this has been entered into the machine, the user can complete the steps laid out in the Methods section.

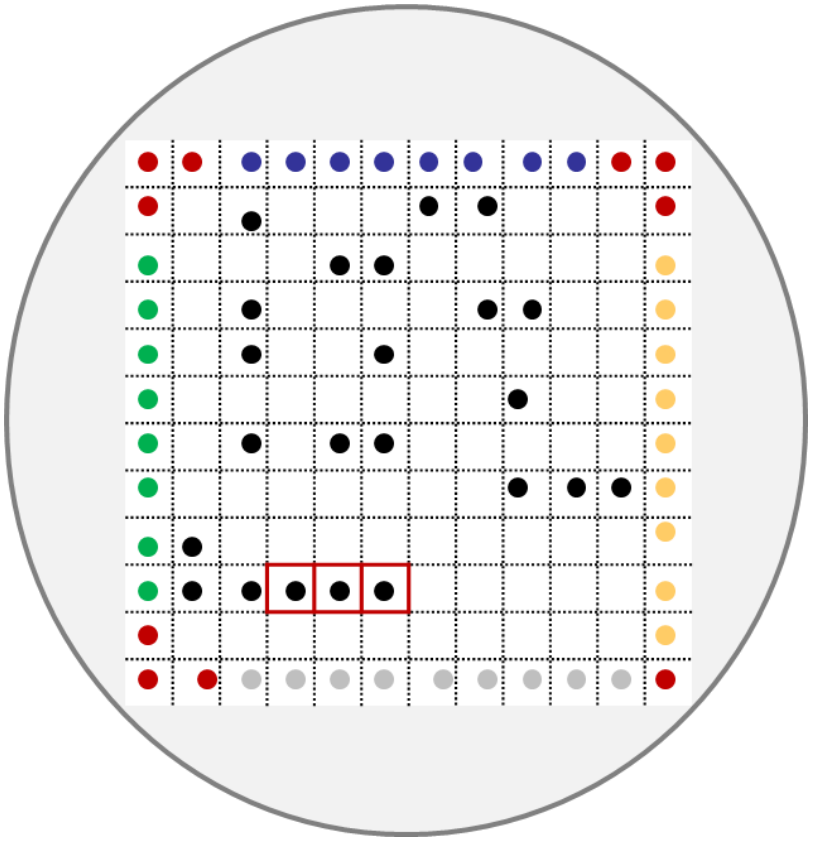
5.1.3.2. Results interpretation

When the reaction assay is complete, the user can export the XML file and a zip folder containing images of the reaction wells. The exported XML file is a modified file of the original XML worklist and contains the metadata associated with that specific run

including the date of the run, machine ID, the lot/batch number of the reagents used in the test, the batch number of the assay wells/cartridges and the status of completion for each step. This allows the user and/or manufacturer to pinpoint any issues that may arise if a test run should fail. The user then imports these XML files and zip folders back into the HISTO MATCH software where the results can be analysed for HLA typing. The software reads in the XML file to refer to the sample ID and program type so that it has context on how to interpret the colourmetric probes pattern (see Figure 5.4 for the premise of the spots and Figure 5.5 for premise of the software's gridding function). If the intensity of the colourmetric probe reaches the threshold that was defined by the manufacturer for that probe then that probe is interpreted as a positive hit (Figure 5.6). The pattern of positive and negative hits on the probes will determine the allele called.

5.1.4. Premise of project

The primary aim for the manufacturer is for the equipment to be deployed in clinical scenarios where clinicians need a relatively quick answer as to whether a patient carries an ADR risk allele for a particular drug they wish to administer to their patient. The new method aims to have a very quick turnaround time (approx. 48 hours from when the extracted DNA is processed) and be more cost effective compared to current methods. This would allow a clinician to predict, prevent and/or diagnose ADRs. For this thesis project, the equipment requires validation before being implemented in a clinical setting. The HISTO MATCH kit was tested using samples which have been previously typed using a currently accepted method and therefore the HLA alleles for the samples are already known. As part of the validation process, the samples were independently tested twice using two researchers to demonstrate if the kit is both accurate and consistent.



Control spots:

Indicates that master mix was added to the PCR run and also act as positional probes that enables the software to locate the picture

Kit identity

Lot identity

Batch identity

Check digits:

Control, if kit, lot and batch identity are correct. If not, something is wrong with the gridding

Positive specific probes

Positive controls:

Positive for all alleles and indicates that the PCR was successful

Basis for defining the reactivity of the other probes in the well

Figure 5.4: Diagram showing the colourmetric probes pattern within the well and what each of the spots represent.

Sourced from manufacturer.

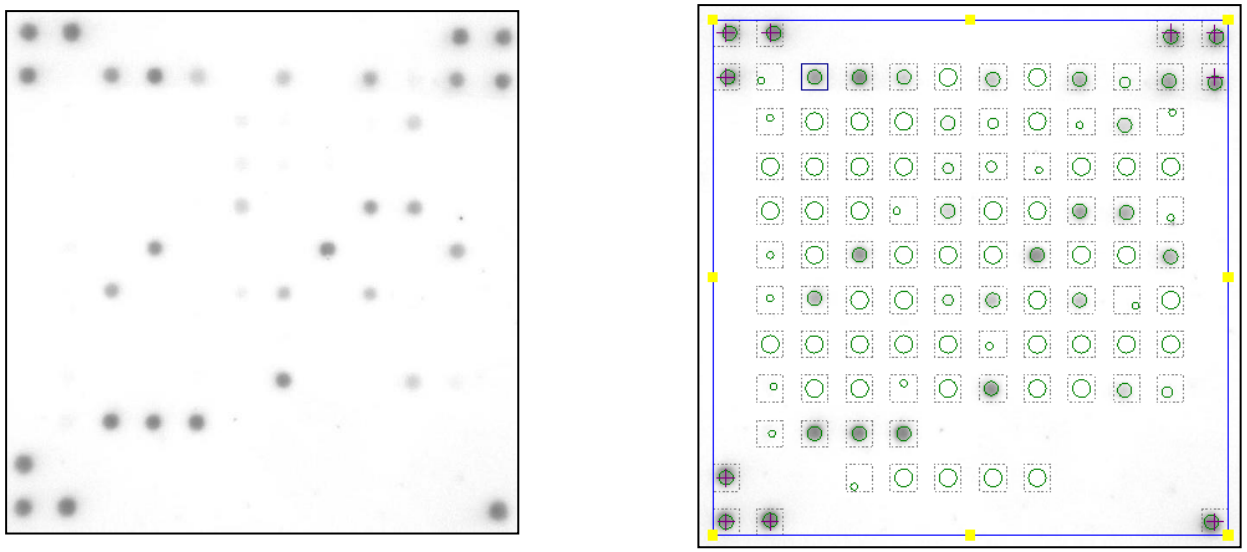


Figure 5.5: HISTO MATCH software's gridding function to align colourmetric assays against reference (Figure 5.4).

a) An example of the raw image file

b) Shows the grid overlay generated by the HISTO MATCH software. The software automatically performs the gridding, however it is advised that users manually check these and adjust if required.

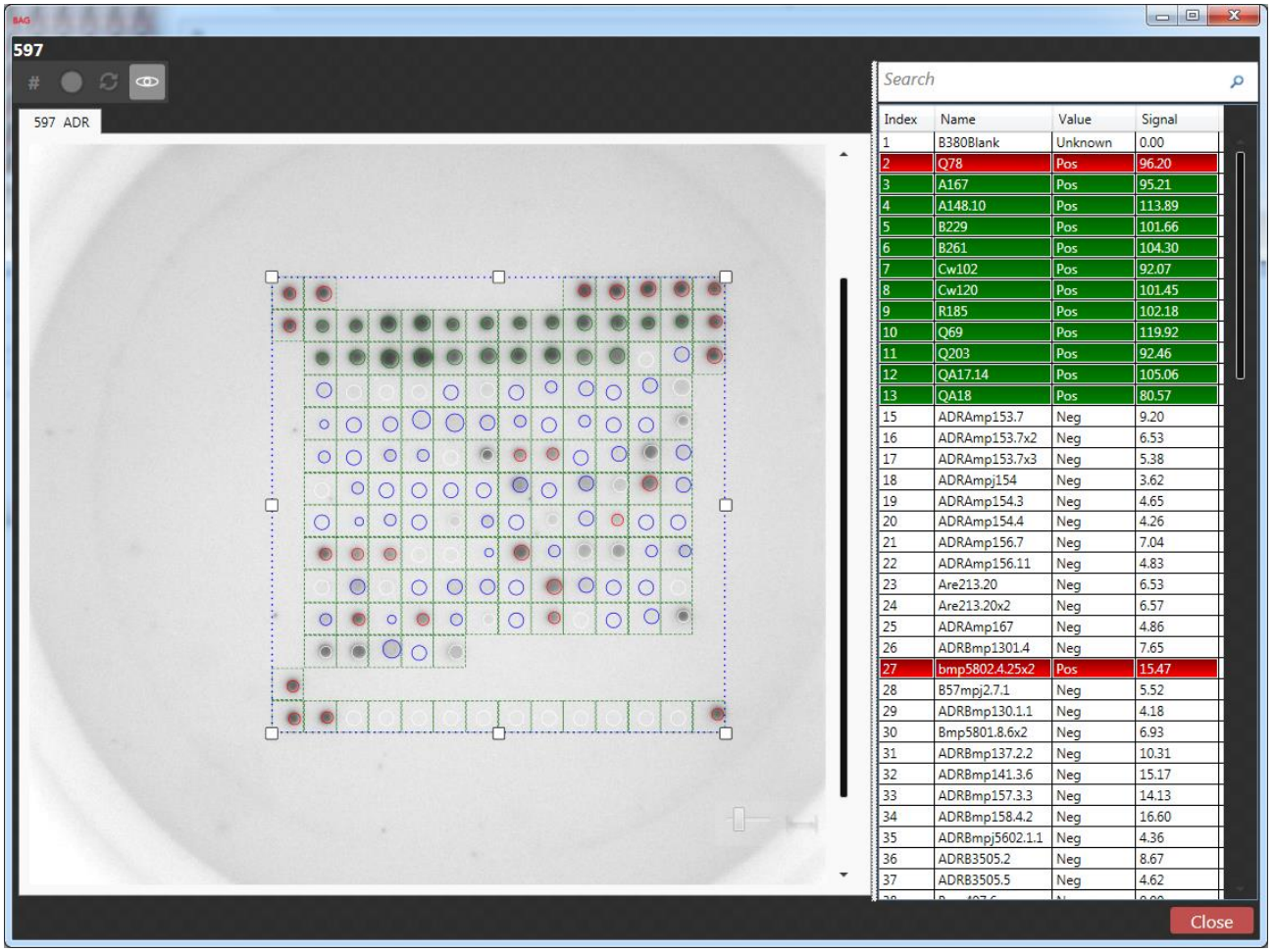


Figure 5.6: Screenshot of the HISTO MATCH software analysing the colourmetric probes to determine positive/negative hits for that probe.

In addition to validating the assay, the panel of HLA alleles was evaluated in the context of this project, allowed changes to be recommended where necessary. Additionally, this was the first time that the assay was tested using DNA extracted directly from human cells so there was an expectation that the machine would need some minor calibration to be able to adapt to the change – the data collected during initial runs was to be reviewed and used to formulate the required adjustments. As well as the equipment's primary function of informing clinicians of the carriage of risk alleles associated with ADRs, the project was also be complemented with a clinical decision support tool (0). The work carried out in that chapter is supportive for the development of the project in this chapter and vice versa.

5.2. Materials and Methods

Table 5.1: List of reagents/consumables that are used with the new HLA typing methods investigated in this chapter.

Protocol step	Item(s)	Supplier
PCR amplification	Mastermix Magnesium chloride ($MgCl_{2)}$ PCR plates (96 wells) PCR seal	MC Diagnostic St Asaph Wales United Kingdom
	dH_2O DNA	
Mr Spot Assay	Conjugate Block buffer Stringent buffer Tris-buffered saline (TBS) Substrate buffer Hybridisation buffer	MC Diagnostic St Asaph Wales United Kingdom

5.2.1. Researchers

The following people were involved in the project and their roles are defined below:

AA – Principal/supervising investigator

LH – One of the researchers who was conducting the analysis of the HISTO SPOT kit. (Departed the project mid-way though)

AO – researcher who took over upon the departure of LH

GG – other researcher who conducted the analysis of HISTO SPOT

5.2.2. Volunteers / DNA samples

The DNA samples were obtained from an archive generated from 1,000 healthy volunteers which were recruited from the North West region of England. The DNA samples were collected prior to this project and were collected in two stages with the demographic characteristics of the volunteers listed in the (see Table 5.2) below. In the first collection, 400 patients were recruited with 385 of those being brought forward – this thesis will refer to these as the first 400 based on the internal ID numbering system used. [180] An additional 600 volunteers were recruited in the second round. [181] 10mL of venous blood was used where the genomic DNA was extracted using Chemagic magnetic separation (Chemagen, Baesweiler, Germany). The samples were analysed using sequence based typing by Histogenetics laboratory (Histogenetics, New York, USA). The sequencing was performed over exon 2 for class II HLA genes and exons 2 and 3 for class I loci – the HLA loci covered were HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQB1 and HLA-DQA1. The DNA concentrations were determined using Nanodrop spectrophotometer and picogreen measurements and the samples were normalised to 50 ng/μL concentrations. The HLA typing information for these samples were performed prior to this project. [180, 181]

Table 5.2: Demographic information of volunteers' DNA samples.

Characteristic	Archive samples	
	1-400	401-1000
Median age (standard dev.)	29 (± 10 years)	31.5 (± 10.6 years)
Sex, female : male	0.64 : 0.36	0.66 : 0.34
Ethnic diversity		
Caucasian	77.4%	90.2%
South Asian	7.3%	2.8%
Chinese	3.9%	1.5%
Black	1.3%	1.2%
Other / Mixed	10.1%	3.8%

During initial assay runs to test and calibrate the HISTO SPOT equipment, samples with ID 1-1000 (the whole sample range) were used. However, several rounds of testing and equipment adjustments were required, more than originally envisioned, so samples with tube ID 1-400 were close to the point of depletion. Consequently, a decision was made to conserve the samples and not use samples 1-400 in the final results. Therefore, the results section only displays the allelic calls for samples in the range of 401-1000. The one well assay method shows the results for samples 401-804 – a decision was made to stop at this sample and not continue with samples 805-1,000 as it became apparently clear that the one well assay method would not come anywhere close to 100% accuracy for HLA typing. Priority was given to conserve the DNA for further analysis (both relating to this project and for other studies) – similarly as to why samples 1-400 were not included.

The HISTO MATCH software was adjusted with the data collected from the interpretation of samples 401-804. It was decided that the samples that were not included (805-1,000), could be used for the five well cartridge system as these samples are unknown to the HISTO MATCH software meaning that the software would be making the allele calls without the manufacturers programming the software to automatically recognise the samples' HISTO SPOT profiles. This is similar to the concept of training data in bioinformatics and statistical analysis which is used as a means to prevent overfitting of the data. Since the five well cartridge system requires

using five times the quantity of sample DNA, it was apparent that there were insufficient quantities of samples 805-901 therefore, the five well cartridge assays were performed using 96 samples from tube ID range 902-999.

5.2.3. One well assay method

The first stage of the procedure was to amplify the DNA in the samples that were provided in the project. The samples were previously diluted from the original stock solutions to have a standardised concentration of 50 ng/μL. The HISTO SPOT protocol requires that the DNA samples be at a concentration of 15 ng/μL. Therefore for this project, the DNA was diluted down to 15 ng/μL for the procedure using Tris-EDTA (TE) buffer. The PCR amplification reagents are prepared as shown below:

Table 5.3: Quantities of reagents required for mastermix solution for one well assay method.

Reagent	**Quantity**	
	Per Sample	**96 well plate (x110)**
Mastermix	10μL	1100μL
$MgCl_2$	5μL	550μL

In 2mL microcentrifuge tube:

- Add the mastermix and $MgCl_2$ together and vortex

In PCR plate:

- Add 15μL of mastermix and $MgCl_2$ solution into each well
- Add 5μL of DNA – one patient sample per well
- Cover plate using PCR seal
- Centrifuge PCR plate for 60 seconds at 500rpm
- Transfer to thermal cycler for amplification as described in Table 5.4

Table 5.4: PCR thermal cycler program for the one well assay method as provided by the manufacturer.

Program Step	Time (mm:ss)	Temperature	Number of cycles
1st Denature	03:00	95˚C	X1
Denature	00:15	95˚C	X35
Annealing	00:30	65˚C	
Extension	02:00	68˚C	
Final extension	10:00	68˚C	X1
Hold	Program end	22˚C	

5.2.4. Five well cartridge assay method

The procedure for the five well cartridge assay is similar to that of the one well method with the main difference being that five times as much sample DNA is required as the HLA loci have been separated into individual wells of the cartridge. The mastermix and $MgCl_2$ solution is mixed as below:

Table 5.5: Quantities of reagents required for mastermix solution for five well cartridge assay method.

Reagent	Quantity	
	Per Sample	6x five well cartridges (x35)
Mastermix	10µL	350µL
$MgCl_2$	5µL	175µL

In microcentrifuge tube:

- Add the mastermix and $MgCl_2$ together and vortex

In PCR plate:

- Add 15µL of mastermix and $MgCl_2$ solution into each well
- Add 5µL of DNA in each well of cartridge (25µL of DNA over 5 wells)
- Cover plate using PCR seal
- Centrifuge PCR plate for 60 seconds at 500rpm

Transfer to thermal cycler for amplification as described in Table 5.6

Table 5.6: PCR thermal cycler program for the five well cartridge assay method as provided by the manufacturer.

Program Step	Time (mm:ss)	Temperature	Number of cycles
1st Denature	02:00	96˚C	X1
Denature	00:15	96˚C	X10
Annealing + Extension	01:00	65˚C	
Denature	00:10	96˚C	X20
Annealing	00:50	61˚C	
Extension	00:30	72˚C	
Hold	∞	22˚C	

5.2.5. Loading of automated HLA-typing machine

After the PCR amplification step, the samples can be loaded onto the MR SPOT machine which will carry out the SSO probe based colourmetric reaction that allows the HISTO MATCH software to analyse the assay and call the patient's HLA allele status. The steps are described below.

- Warm the hybridisation buffer and stringent wash in 30˚C water bath.
- Load the XML work sheet onto the MR SPOT machine via USB flash drive
- Follow the on-screen instructions to prepare the machine, which are:
    - Clear the tips disposal and waste liquid troughs
    - Re-rack the 200µL and 1,000µL pipette tips
    - Fill the water bottle with distilled H_2O
    - Add the reaction strips/wells to the machine
    - Add the reagents to the labelled wells (see below)
    - Run the program
    - Once run is complete, export results to USB flash drive for interpretation on HISTO MATCH software.

The reagents are proportional to the number of samples being tested. The quantities are provided by the MR SPOT machine as the samples have been declared in the XML worklist so the machine is able to calculate the required reagent amounts.

5.2.6. Protocol for MR SPOT typing assay

Whilst the SSO assay is running, the MR SPOT instrument runs through the following steps for both the one well system and five well cartridge system (performed automatically without user intervention):

1. Add block buffer into typing wells and leave for 5 minutes
2. Denature the samples for 5 minutes
3. Aspirate the blocking buffer
4. Add hybridisation buffer to the samples
5. Transfer samples to corresponding typing wells
6. Hybridise at 50˚C for 15 minutes
7. Perform 3x washes with stringency buffer
8. Add conjugate to typing wells and incubate for 15 minutes
9. Perform 3x washes with buffer
10. Add substrate and incubate for 15 minutes
11. Wash with H_2O
12. Dry for 5 minutes
13. Internal camera takes images of wells' colourmetric probes

14. User transfers images to external PC with HISTO MATCH software for analysis

5.2.7. Comparison of HISTO SPOT assay and sequence based typing

This section refers to project researchers using the codes explained in 5.2.1. The final reports of the HISTO SPOT assay were collected in a spreadsheet document so that they could be compared with the results of the (reference) sequence based typing method. In order to minimise potential of bias, the two researchers (GG and AO) who performed the HISTO SPOT assay were not told of the HLA status for the samples until after the results were collected. Additionally, for reasons of confidentiality, only the supervising researcher (AA - who did not perform the HISTO SPOT assays) had access to the data that details the samples' HLA status as derived from the sequence based typing method. Researcher AA only provided data stripped of any potentially identifying information that was not relevant to the project. The sample ID number (with the study code stripped) containing only the sample's tube number and the HLA allele status for that sample were provided for the comparison. The results were then analysed and marked to determine the percentage of correct calling of the HLA alleles using the SBT derived alleles as the reference.

5.2.8. HISTO SPOT known ambiguities

The HISTO SPOT assay had ambiguities in calling HLA alleles that are known to the manufacturer but could not be resolved. For the one well assay method the ambiguities are as follows:

- HLA-A*33:03:01 could not be uniquely identified in the presence of HLA-A*31:01:02:01 allele

- HLA-B*13:01 was ambiguous with HLA-B*44:01

- HLA-B*58:01 could not be uniquely identified in the presence of HLA-B*58:02

- HLA-DRB1*15:01 identification is ambiguous with HLA-DRB1*15:03 and HLA-DRB1*15:06

- The HISTO MATCH software reported an ambiguity with HLA-A*68:01:01:01 and HLA-A*68:01:02:01 although this is not treated as an ambiguity in this thesis as both alleles produce identical proteins through synonymous polymorphisms.

- HLA-DQB1*02:01 and HLA-DQB1*02:02 can be ambiguous due to the method of typing. The original Histogenetics sequencing only sequenced class II alleles at exon 2 whilst the HISTO SPOT method types using exons 2 and 3 – the two alleles are identical in exon 2 so the Histogenetics typing method cannot distinguish between the two alleles. The HISTO MATCH software was subsequently modified so that, for this experiment, the allele calls for HLA-DQB1*02:02 are reported as ambiguous to allow the comparison using the Histogenetics typing can be made.

With the exception of the HLA-DQB1*02:01 and HLA-DQB1*02:02 issue, the other known ambiguities do not affect the five well cartridge assay method and is able to discriminate between the alleles.

5.3. Results

The HISTO MATCH software will only display, to the user, the risk alleles that are part of the 23 alleles that make up biomarker panel. Alleles from outside this panel are not included. The Histogenetics sequencing does include the calls for all the results, however, these are not relevant to this project so only the risk alleles are included in the results. The results were collected in a spreadsheet document and listed in groups of three. The first row of the group is the sample's HLA status as determined using the alleles called using Histogenetics SBT method. Next, is the alleles called using the HISTO SPOT biomarker panel carried out by researcher AO and the third row is the HISTO SPOT biomarker panel carried out by researcher GG. The results from HISTO

SPOT biomarker panel are colour coded depending on whether the allele called is correct relative to the sequence data obtained from the Histogenetics typing. A green cell within the table means that the allele(s) for that HLA loci called by the biomarker panel agrees with that of sequenced data, a red cell reflects the HISTO MATCH software assigning a value does not agree with the sequence data and yellow denotes an allele is one of the known ambiguities as described in the section above.

Samples shaded black are samples which were not utilised as part of this study with the reason being that these samples were already depleted before this investigation started. Table 5.7 shows an example of three results. The full table consisting of samples 401-804 – the entire run for the one well assay method can be found on the attached USB flash disk/compact disk that accompanies the thesis as the spreadsheet file "chapter_5_HISTO_SPOT_one_well.xlsx".

If a sample contained two risk alleles that were called for that HLA loci, then the colour code denoting the least favourable outcome was recorded to provide high level of stringency. For example in Table 5.7, for sample 447, the biomarker panel for both researchers called HLA-DQB1*03:03:02:01 which is a correct call as it matches with the sequence data. However, the HISTO MATCH software did not call the HLA-DQB1*02:01 allele – this is an allele that is known to be ambiguous so the cell could be shaded either green or yellow, since the ambiguous result is the least favourable, the cell is shaded yellow.

Table 5.8 shows the values that each of the risk alleles were called correctly. A correct call is described when the HISTO MATCH biomarker panel result agrees with the result for the sequence data obtained by Histogenetics' HLA typing. Table 5.9 shows the percentage of samples where the HLA calls agreed with the sequence data for the sample as a whole – e.g. in the examples shown in Table 5.7 sample 401 was called correctly across all HLA loci so this sample is registered as correct in Table 5.9Table 5.9 whilst for sample 405, the HISTO MATCH biomarker panel for researcher AO there were two alleles called that did not match with the sequence data, this is classed as incorrect in Table 5.9.

5.3.1. One well assay method

Table 5.7: Example of results from the one well HISTO SPOT assay.

Tester	Tube ID	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1	HLA-DQA1
Sequence data	401		57:01		07:01	03:03	02:01
biomarker panel AO	401		57:01:01		07:01:01:01	03:03:02:01	02:01:01:01
biomarker panel GG	401		57:01:01		07:01:01:01	03:03:02:01	02:01:01:01
Sequence data	405						
biomarker panel AO	405			08:01:01		03:03:02:01	
biomarker panel GG	405						
Sequence data	447		57:01		07:01	02:01/ 03:03	02:01
biomarker panel AO	447					03:03:02:01	02:01:01:01
biomarker panel GG	447				07:01:01:01	03:03:02:01	02:01:01:01

Legend

Green	HISTO SPOT assay called sample allele correctly (relative to sequenced based typing method)
Yellow	Known ambiguities as defined in section 5.2.7 (e.g. HLA-DQB1*02:01)
Red	HISTO SPOT assay called sample allele incorrectly

Table 5.8: a) Statistical results measuring the performance of the one well HISTO SPOT assay method relative to the sequence data for each allele. Results obtained from test carried out by researcher AO

Samples 401-804								
Allele	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
A*31:01	19	371	1	4	0.826	0.997	0.950	0.989
A*33:03	8	375	11	1	0.889	0.972	0.421	0.997
A*68:01	17	373	5	0	1.000	0.987	0.773	1.000
B*13:01	0	384	11	0	N/A	0.972	0.000	1.000
B*15:02	1	390	2	2	0.333	0.995	0.333	0.995
B*35:05	1	391	3	0	1.000	0.992	0.250	1.000
B*56:02	0	394	1	0	N/A	0.997	0.000	1.000
B*57:01	27	359	1	8	0.771	0.997	0.964	0.978
B*58:01	4	385	1	5	0.444	0.997	0.800	0.987
C*04:01	47	299	36	13	0.783	0.893	0.566	0.958
C*08:01	3	352	39	1	0.750	0.900	0.071	0.997
DRB1*07:01	84	291	2	18	0.824	0.993	0.977	0.942
DRB1*11:01	25	364	0	6	0.806	1.000	1.000	0.984
DRB1*13:02	11	371	5	8	0.579	0.987	0.688	0.979

DRB1*15:01	88	276	7	24	0.786	0.975	0.926	0.920
DQB1*02:01	111	195	4	85	0.566	0.980	0.965	0.696
DQB1*03:03	36	344	8	7	0.837	0.977	0.818	0.980
DQB1*05:02	7	372	13	3	0.700	0.966	0.350	0.992
DQB1*06:02	106	275	7	7	0.938	0.975	0.938	0.975
DQB1*06:04	9	376	7	3	0.750	0.982	0.563	0.992
DQB1*06:09	3	384	3	5	0.375	0.992	0.500	0.987
DQA1*01:02	133	248	2	12	0.917	0.992	0.985	0.954
DQA1*02:01	94	289	3	9	0.913	0.990	0.969	0.970

Table 5.8: b) Statistical results measuring the performance of the one well HISTO SPOT assay method relative to the sequence data for each allele. Results obtained from test carried out by researcher GG

Samples 401-804								
Allele	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
A*31:01	19	366	6	4	0.826	0.984	0.760	0.989
A*33:03	9	372	14	0	1.000	0.964	0.391	1.000
A*68:01	17	372	6	0	1.000	0.984	0.739	1.000
B*13:01	0	383	12	0	N/A	0.970	0.000	1.000

B*15:02	3	392	0	0	1.000	1.000	1.000	1.000
B*35:05	1	381	13	0	1.000	0.967	0.071	1.000
B*56:02	0	393	2	0	N/A	0.995	0.000	1.000
B*57:01	14	360	0	21	0.400	1.000	1.000	0.945
B*58:01	5	386	0	4	0.556	1.000	1.000	0.990
C*04:01	42	222	113	18	0.700	0.663	0.271	0.925
C*08:01	4	276	115	0	1.000	0.706	0.034	1.000
DRB1*07:01	79	291	2	23	0.775	0.993	0.975	0.927
DRB1*11:01	20	363	1	11	0.645	0.997	0.952	0.971
DRB1*13:02	8	368	8	11	0.421	0.979	0.500	0.971
DRB1*15:01	84	274	9	28	0.750	0.968	0.903	0.907
DQB1*02:01	74	195	4	122	0.378	0.980	0.949	0.615
DQB1*03:03	31	343	9	12	0.721	0.974	0.775	0.966
DQB1*05:02	5	373	12	5	0.500	0.969	0.294	0.987
DQB1*06:02	99	275	7	14	0.876	0.975	0.934	0.952
DQB1*06:04	1	368	15	11	0.083	0.961	0.063	0.971
DQB1*06:09	1	387	0	7	0.125	1.000	1.000	0.982
DQA1*01:02	123	248	2	22	0.848	0.992	0.984	0.919
DQA1*02:01	92	290	2	11	0.893	0.993	0.979	0.963

Table 5.9: Overall concordance value for the one well HISTO SPOT assay method relative to the sequence data across samples 401-804

Percentage of samples called correctly	
researcher AO	50.91
researcher GG	23.24

5.3.2. Five well cartridge assay method

Table 5.10 shows examples of the results from the five well cartridge method. This table uses a same colour code for the results in Table 5.7; green for results that agree with the Histogenetics sequenced data and black for samples that could not be assayed (insufficient sample quantity). For the known ambiguity for HLA-DQB1*02:01 (and 02:02), these were classed as correct calls if the biomarker panel HISTO MATCH software called HLA-DQB1*02:02 when the sequenced data value is HLA-DQB1*02:01. These instances were agreed to be more precise due to the nature of the typing method which have previously been validated as the manufacturer already provides services to type the HLA-DQB1 alleles using the technique that has been implemented in this five well cartridge assay.

The full table consisting of samples 902-999; the entire run for the five well cartridge assay method can be found on the attached USB flash disk/compact disk that accompanies the thesis as the spreadsheet file "chapter_5_HISTO_SPOT_five_well.xlsx". Table 5.11 shows the percentage that each of the risk alleles were called correctly. A correct call is described when the HISTO MATCH biomarker panel result agrees with the result for the sequence data obtained by Histogenetics' HLA typing. Table 5.12 shows the percentage of samples where the HLA calls agreed with the sequence data for the sample across all six loci.

Table 5.10: Example of results from the five well cartridge HISTO SPOT assay.

Tester	Tube ID	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1	HLA-DQA1
Sequenced data	944		57:01				
biomarker panel AO	944		57:01:01				
biomarker panel GG	944		57:01:01				
Sequenced data	945					02:01/ 03:03	
biomarker panel AO	945					02:02:01/ 03:03:02:01	
biomarker panel GG	945					02:02:01/ 03:03:02:01	

Legend

Green	HISTO SPOT assay called sample allele correctly (relative to sequenced based typing method)
Yellow	Known ambiguities as defined in section 5.2.7 (e.g. HLA-DQB1*02:01)
Red	HISTO SPOT assay called sample allele incorrectly
Black	Insufficient DNA sample to carry out HISTO SPOT typing

Table 5.11: a) Statistical results showing the performance of the five well cartridge HISTO SPOT assay method relative to the sequence data for each allele in the biomarker panel. Results obtained from test carried out by researcher AO.

Samples 901-999								
Allele	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
A*31:01	2	190	0	0	1	1	1	1
A*33:03	1	191	0	0	1	1	1	1
A*68:01	6	186	0	0	1	1	1	1
B*13:01	0	192	0	0	N/A	1	N/A	1
B*15:02	0	192	0	0	N/A	1	N/A	1
B*35:05	0	192	0	0	N/A	1	N/A	1
B*56:02	0	192	0	0	N/A	1	N/A	1
B*57:01	8	184	0	0	1	1	1	1
B*58:01	3	189	0	0	1	1	1	1
C*04:01	19	173	0	0	1	1	1	1
C*08:01	1	191	0	0	1	1	1	1
DRB1*07:01	27	157	0	0	1	1	1	1
DRB1*11:01	8	178	0	0	1	1	1	1
DRB1*13:02	6	180	0	0	1	1	1	1
DRB1*15:01	33	153	0	0	1	1	1	1
DQB1*02:01	24	158	0	0	1	1	1	1
DQB1*03:03	12	170	0	0	1	1	1	1
DQB1*05:02	2	180	0	0	1	1	1	1

DQB1*06:02	30	152	0	0	1	1	1	1
DQB1*06:04	3	179	0	0	1	1	1	1
DQB1*06:09	3	179	0	0	1	1	1	1
DQA1*01:02	39	143	0	0	1	1	1	1
DQA1*02:01	27	155	0	0	1	1	1	1

Table 5.11: b) Statistical results showing the performance of the five well cartridge HISTO SPOT assay method relative to the sequence data for each allele in the biomarker panel. Results obtained from test carried out by researcher GG.

Samples 901-999								
Allele	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
A*31:01	2	188	0	0	1	1	1	1
A*33:03	1	189	0	0	1	1	1	1
A*68:01	6	184	0	0	1	1	1	1
B*13:01	0	190	0	0	N/A	1	N/A	1
B*15:02	0	190	0	0	N/A	1	N/A	1
B*35:05	0	190	0	0	N/A	1	N/A	1
B*56:02	0	190	0	0	N/A	1	N/A	1
B*57:01	8	182	0	0	1	1	1	1
B*58:01	3	187	0	0	1	1	1	1

C*04:01	18	150	0	0	1	1	1	1
C*08:01	1	167	0	0	1	1	1	1
DRB1*07:01	27	157	0	0	1	1	1	1
DRB1*11:01	8	178	0	0	1	1	1	1
DRB1*13:02	6	180	0	0	1	1	1	1
DRB1*15:01	33	153	0	0	1	1	1	1
DQB1*02:01	24	158	0	0	1	1	1	1
DQB1*03:03	12	170	0	0	1	1	1	1
DQB1*05:02	2	180	0	0	1	1	1	1
DQB1*06:02	30	152	0	0	1	1	1	1
DQB1*06:04	3	179	0	0	1	1	1	1
DQB1*06:09	3	179	0	0	1	1	1	1
DQA1*01:02	39	143	0	0	1	1	1	1
DQA1*02:01	27	155	0	0	1	1	1	1

Table 5.12: Overall concordance value for the five well cartridge HISTO SPOT assay method relative to the sequence data across samples 902-999.

Percentage of samples called correctly	
researcher AO	100.00
researcher GG	100.00

5.4. Discussion

5.4.1. Limitations of the HISTO SPOT assay

The one well assay method proved to be unreliable when tested with samples that were extracted from human volunteers, where the system called 51% and 23% (performed by researcher AO and researcher GG respectively) of the volunteers samples' HLA status correctly relative to the sequence based typing method used as the reference. This was the final result after several rounds of calibration and improvements were conducted with the method. During the calibration stages, the performance was very poor – but adjustments were expected to be performed as this was the first live test with unseen human samples. Researchers AO, GG and LH (who was heavily involved during the calibration stages) collaborated with the manufacturer to discuss improvements during which several factors were discovered to be influential. Firstly, it was noted that relative to a standard PCR program, there was an extremely long DNA polymerase extension cycle. This was done by the manufacturer as the HLA region is notoriously difficult region for PCR techniques and so the multiple probes covering the many HLA loci were extremely sensitive to slight perturbations. Consequently, the PCR machine used by LH and GG was adjusted to decrease the ramp rate (the rate of temperature change). This showed noticeable improvement but performance was still poor (results not included). A different PCR thermal cycler machine that matched with the manufacturer's recommended model was later used which again showed improvement.

After the changes to the PCR program, there were several rounds of HISTO MATCH probe thresholds as certain alleles showed extremely poor reliability as the concordance percentage against the reference was very low – alleles HLA-DQB1*06:04 (61.9% for AO and 5.0% for GG for the final run for the single well assay) and HLA-DQB1*06:09 (33.3% for AO and 14.3% for GG). When analysed in the full results, the biomarker panel would often give false positives for these alleles which indicated that there were possibly some issues with the threshold of the HISTO MATCH software when measuring the probe/spot relating to these alleles. The manufacturer also raised concern with the method with which the DNA was extracted from the volunteers' blood – they have only validated the technique using Qiagen DNA extraction columns (Qiagen, Manchester, United Kingdom). There were also

problems with gathering consistent results between the two researchers where the allele calls from the assay performed by researcher GG was considerably lower than that of researcher AO. Observation of GG's laboratory technique (e.g. pipetting) did not reveal any poor technique or other obvious signs of human error. The results from Table 5.8 did not show constancy – some alleles AO performed better and other alleles GG performed better. The most likely explanation appears to be statistical chance.

After the several rounds of calibration and very poor final results, it was agreed that many of the issues that were subsequently controlled for were unrealistic to expect when making full product deployment as many labs use different methods for DNA extraction and different PCR thermal cyclers so limiting to one approved method and equipment machines was unrealistic to expect in practice. The inability to gather consistent results (differences between the results when carried out by researchers AO and GG) meant that the reliability could not be trusted. Consequently, it was decided that the probes for the HLA loci be separated into different wells as part of a cartridge system – the downside to this approach was the reduction in turnaround as the only six cartridges could be accommodated by the HISTO SPOT automated machine. However, the chief investigator of the project has extensive clinical expertise in a hospital scenario around ADRs and informed the group that based on his experience, it was unlikely that a clinical laboratory would be required to test that many patients in a day (personal correspondence with Pirmohamed M., 2017).

The five well cartridge system allowed more probes per locus to be included in each reaction well allowing for probe redundancy that greatly aids in robustness and reliability. This method involved adapting reaction wells that the company already supplies commercially. The method proved to be reliable where both AO and GG obtained 100% concordance with the reference sequence data when testing 96 samples that were previously unknown to the technique (in either one well or five well cartridge assays).

5.4.2. Cost-effectiveness for the HISTO SPOT assay

It has been difficult to obtain costs/prices for the current procedures as the various laboratories supplying the service often negotiate the prices for each individual

customer based on several factors. However, the NHS Blood and Transplant Service is charged approximately £90 for a single test for HLA-A*31:01 and personal communications with the members of this project during progress meetings suggests that the cost to type a patient's DNA sample for an individual HLA loci somewhere between £50 and £200. [182] The new HISTO MATCH process aims to test for the carriage of the >20 HLA alleles across all five of the tested HLA loci at a lower price (still to be set). That is to say HISTO MATCH will type >20 alleles at a cost lower than or competitive with an individual HLA loci using currently available methods.

The benefits of a clinical decision support tool will vary depending on the specific drug-allele combination. A review describing the cost effectiveness of the HLA panel from Figure 5.3 reported that there is evidence to support the cost effectiveness for the testing of HLA-B*57:01 (abacavir screening), HLA-A*31:01 and HLA-B*15:02 (prior to carbamazepine treatment) and HLA-B*58:01 (prior to allopurinol treatment). [183] These drug-allele combinations already have relevant warnings pertaining to HLA mediated ADRs – most likely due to the relatively high amount of studies published for these specific drugs/alleles. Given that the screening test using the HISTO SPOT kit costs less than or competitive with testing for an individual allele using already available services (> £50) then the potential benefits particularly for the alleles mentioned in this paragraph are clear.

The cost-benefit of the other alleles in the panel varies depending on the scenario, some of these cost-benefit analysis are implemented in the clinical decision support tool in 0. For example, the benefit to screening patients before the treatment of flucloxacillin is very low. Despite the data in 0 showing that the negative predictive value approaches 1.00, the positive predictive value is very low (0.001) meaning that it is not cost-beneficial to screen patients beforehand, but in the (very rare) event that a patient experiences DILI whilst on the drug, a negative result for HLA-B*57:01, using HISTO SPOT, with the patient experiencing liver injury may help in excluding flucloxacillin as a cause. Additionally some alleles do not show strong enough evidence to warrant the discontinuation of the drug if the patient carries the risk allele or the alternative treatment (if any) does not meet favourable cost-benefit ratio. In such cases, the clinician and patient can be informed of the potential risk and be monitored more closely. This has already been recommended for carbamazepine in patients

carrying HLA-A*31:01. [184] Given the mixed utility for pre-emptive screening of patients prior to a particular treatment, it is highly recommended that the cost-effectiveness also be included in the clinical decision of drug administrating, as such the clinical decision support tool in 0 must be mindful of the value for each drug-allele combination.

The patient's HLA information can be included into the patient record allowing it to be reviewed for future use if the patient needs to be treated with the same or different drug. For example, the cost-effectiveness is difficult to predict. However, since the information is already on the patient's health record, if pre-emptive screening is not cost-beneficial but the HLA status is already on file then the clinician can take advantage of this when devising the treatment for the patient.

This also links directly with time savings that HISTO-SPOT offers compared to current typing methods. The typical turnaround time for current typing methods is around two weeks which is often unacceptable for clinicians wanting to make a decision regarding a particular treatment/drug they wish to prescribe. The HISTO SPOT methodology offers a typical turnaround of 24 to 48 hours to produce a report describing the sample's HLA typing profile. The faster reporting of HLA type is patently advantageous to the quicker informed clinical decision making. The time savings are further improved using the previous paragraph's example of the information already on patient health records due to previous testing for another of the >20 HLA alleles.

5.4.3. Future work

The five well cartridge system has one further benefit over the one well system, due to the added probes in each well, the resulting robustness allows the system to adapt to changes in the biomarker panel. If in the future a new allele needs to be added to the biomarker panel, it is likely that the existing probes are already present to allow the allele to be identified meaning only an update to the HISTO MATCH software is required – the probes cover a relatively large area of exons 2 and 3 of each loci. This means that the five well cartridge system is very flexible and adaptable to future discoveries.

The spare wells in the 5-well cartridge (3 unused wells) can also be used to cover additional loci such as the cytochrome P450 gene superfamily. These are non-HLA genes/alleles that have been implicated in other forms of ADRs that are not covered in this thesis. However as an example, the CYP2C19 biomarker was shown to be cost-effective for pharmacogenetic screening. [183] meaning that there is potential that the currently unused wells can be utilised by detecting other biomarkers beyond the HLA gene family.

The HISTO SPOT assay's major limitation is that it relies on a SSO based methodology for HLA typing. The assay is only able to type using probes designed to detect polymorphisms in exons 2 and 3. As discussed in Chapter 1.3.6 there is a possibility that for each of the alleles in the panel, there are other alleles from the same gene that have identical sequences inside exons 2 and 3. To HISTO SPOT, such alleles would appear identical. To investigate this, it would be worthwhile to see which alleles share sequence similarity with the risk alleles from the panel. The frequencies for these alleles can be compared within a particular population (e.g. the United Kingdom).

The HISTO MATCH software in its current form (at the time writing) requires updating. One key change needed is that the final report is not user friendly. The HISTO MATCH software produces a spreadsheet document where each loci is reported separately for each patient meaning that the user needs some processing before being interpreted. For this project an in house Python script was written to format the information so that it reflected the sample (the format observed in the results). This has been communicated to the manufacturer and will be included in a future update. Additionally for certain circumstances e.g. if a sample is reported to carry HLA-DRB1*06:02:01 and HLA-DRB1*06:09:01, the end report shows this as "DRB1*06:02 /09:01" which is a common practice in HLA research field but not in medicine (the end user) so has the risk of creating confusion in clinicians who are not well versed in HLA nomenclature practices/common shorthand.

There is also a desire by the researcher team and manufacturer for the HISTO MATCH software directly form links with the clinical decision support software that is described about in the next chapter. This would allow the end user to receive the

HISTO SPOT report and be taken directly to the relevant page on the clinical decision support tool. This can be achieved with little effort as the clinical decision support tool is built upon open HTML standards. This would allow the user to seamlessly transition from the patient's HLA status to the support tool's recommendation allowing for streamline clinical decision making.

5.5. Conclusion

The ADR biomarker panel aims to create an assay able to identify 23 HLA alleles which will aid clinicians in predicting, preventing and/or diagnosing serious immune-mediated ADRs. Initially, the project looked to develop an SSO based typing assay with all 23 HLA alleles, covering both class I and class II HLA alleles in a single test well that could be processed in an automated machine where the SSO probe profile can be interpreted by the manufacturers HISTO MATCH software to identify the 23 alleles at 2 field resolution or higher. However, after considerable assay development, it was determined that the assay required five wells to achieve sufficient robustness that can be deemed reliable enough to be deployed in a clinical setting. The accompanying HISTO MATCH software which produces the final report only informs the user of the alleles that make up the biomarker panel – the other, non-risk alleles are superfluous to the end user clinician and therefore not reported. The cost-effectiveness for the biomarker panel varies depending on the specific drug-allele combination but the wider literature reports on the demand and benefit for clinicians to be able to pre-empt, prevent and diagnose patients for immune-mediated ADRs. When this assay is combined with the clinical decision support tool featured in the next chapter, the system has great potential to aid with clinical decision making for drugs that are known to cause immune mediated ADRs.

Chapter 6. The implementation of a clinical decision support tool to aid clinicians with prescribing medication known to cause immune mediated adverse drug reactions

6.1. Introduction

As mentioned in Chapter 2, there is a great amount of literature for clinicians and researchers to keep track of. So whilst the results from Mr Spot are informative, they may require interpretation of the results depending on the clinician's background knowledge of immune-mediated ADRs. Therefore, a means to support this clinical decision is needed to allow clinicians to make an informed decision when it comes to the best treatment they can provide to their patients. Therefore, to complement Mr Spot (Chapter 5), a clinical decision support tool was developed to help achieve this goal of equipping clinicians with relevant knowledge as to whether a particular drug they wish to prescribe carries above baseline risks of developing ADRs.

6.1.1. Software specification

To aid with the development of any software solution, it is recommended that the specifications be laid out so that it is clear to the developer and people who commission any piece of software as to what the final product will be able to perform. Whilst this section is not a formal software specification, it provides the necessary information to describe the aims of the clinical decision support tool.

The tool is required to be easily accessible, for this reason a website was chosen as opposed to installing it as standalone, installable software. This method of deployment has advantages over packing the tool in a standalone program including a lower barrier to entry/access. With the aim of the software to aid clinicians, it is important that the clinicians have relative ease of access, a standalone program can be accessible in offline scenarios (e.g. non-networked computers) whereas with a website, the client computer must be connected to the web in order to access the information. However, some IT management systems specifically restrict the installation of additional software, as a general security precaution, which would potentially limit the outreach

of the software. Since all that is required for the website is a web browser which are either already installed with the computer's operating system or a common item to be installed by IT management systems (or requested from them), a web based approach has the broadest reach to as many clinicians as possible. An additional benefit to web browsers are the underlying information can be stored on a database which can be kept up to date for all users and limits the possibility that users neglect or are otherwise prevented form updating standalone software packages so end up using outdated information.

In reference to keeping the information up to date, the clinical decision support tool requires that the database be relatively easy to update. This is to allow a team of immune-mediated ADR experts who understand the data to be able to update the information without requiring knowledge of database administration. Similar to the setup described in Chapter 3, a spreadsheet document serves as the primary collection point for the information. Spreadsheets also allow for easy manipulation of the data so that information from different sources can be reformatted to be included in a uniform structure. Once the spreadsheet data is set and ready for deployment, a computer script can be written to extract the data and convert it to the relevant database language code which can then be executed to update the database from which the website will draw its information from.

The database can also be changed in response to the literature, new data can be added and current data can also be edited or deleted if required. This is achieved by first making the relevant additions/edits with the spreadsheet document and then applying the computer script which proceeds to empty the tables in the database and then fill in the information from the spreadsheet. Although, this methods is not as quick to execute as the individual tasks (e.g. only inserting new data as a opposed to emptying/truncating the entire database's tables and then inserting all of the information from the spreadsheet document back in), employing only one method for database changes simplifies the process for administration team and reduces the need for a dedicated, formally trained database manager. The downside to this approach is the time to conduct a database update increases as the number of SQL commands is always the maximum needed to update the database rather than the minimum.

6.1.2. Functional requirements

The clinical decision support tool had specific requirements in terms of functionality which are listed here. A basic username and password credentials system will be required by all users to access the tool (this includes an ability to create new user credentials). The tool will then present users with a list of both the drugs and HLA alleles that are present in the database from which to conduct a search. The selection must also allow the user to select multiple alleles as humans are diploid and multiple HLA genes from those that have been included in the underlying data. Once a user has selected a drug-allele combination, the user will then be presented with the information pertaining to the ADR(s) associated with that particular drug-allele combination. This presented information must draw from the provided data (which will be developed by ADR/clinical experts). The reports section of the tool will also provide the current guidelines from a selection of drug regulatory bodies (e.g. FDA, EMA etc. guidelines) along with hyperlinks to the studies that have reported the association. The tool will also perform calculations to determine certain clinically applied statistical values including the sensitivity/specificity, negative/positive predictive values, pre and post-test probabilities and the number of screenings needed to prevent 1 ADR case.

6.2. Materials and methods

6.2.1. Underlying dataset

The initial dataset that has generated the panel of HLA alleles that have been associated with ADRs was pre-selected and sourced from a previously published article (see Figure 6.1). [179] However, it is worth noting that the study itself describes the figure as not an "exhaustive list". So whilst the figure was used as the initial underlying database build, the database will eventually be built using a curated set of published literature, - the manuscript for this dataset is in development for publication. Since there is considerable overlap with the data from the HLA-ADR database, data from HLA-ADR was used to aid in the construction of the clinical decision support tool as the clinical decision support tool was in development in parallel with the data collection. The data that will eventually go into the clinical decision support tool has been sourced independently of HLA-ADR but will effectively appear to be a curated subset of the HLA-ADR data. The HLA-ADR was used as a source for relevant

literature for the clinical decision support tool. A spreadsheet document was used to collect the information and a Python script was developed to transfer the information from the spreadsheet document to the RDBMS – similar to the process in Chapter 3.

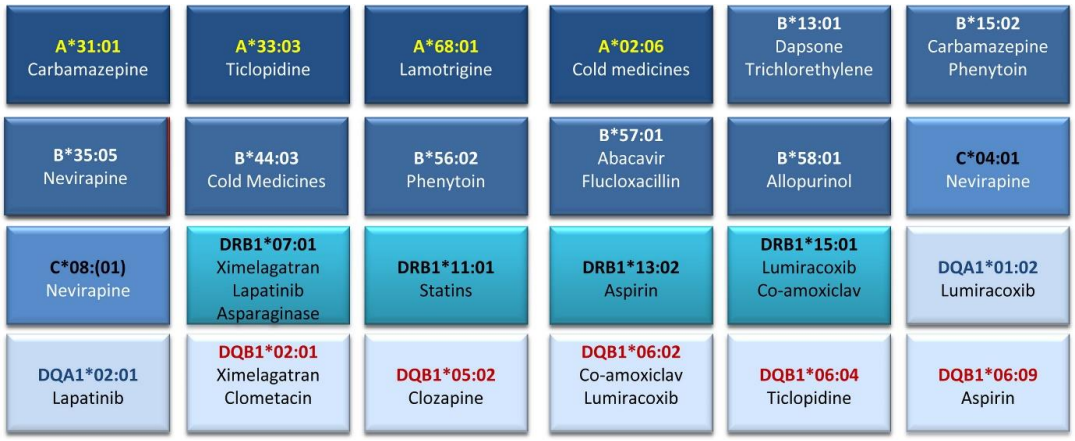


Figure 6.1: Panel of 24 HLA alleles and corresponding drugs that were used as the basis for the underlying ADR data for the clinical decision support tool.

6.2.2. Calculations from the underlying data

The data present in the underlying database was analysed from the original spreadsheet document. The spreadsheet contains the base information (the information sourced from the original study reporting the drug-allele associated ADR and the prevalence of the ADR). From this information calculations are performed to calculate that this chapter has described as the clinical statistics (see Figure 6.2). The website actually performs these calculations as it retrieves the information, however, these calculations are also performed by the spreadsheet document so that the data could be analysed in this project to aid with the eventual clinical recommendations which the clinical decision support tool displays, see USB/CD media file "chapter_6_decision_support_table.xlsx". A full view of this table can be found in the accompanying USB/CD media in the file "chapter_6_decision_support_table.xlsx". The condensed version of the table displays the base information collected from the original study along with the number of screenings needed to prevent one ADR case but excludes the other values calculated using the formulae below.

Figure 6.2: Formulae used to calculate the clinical statistics which are utilised in the clinical decision support tool.

$$sensitivity = \frac{cases\ carrying\ allele}{cases\ carrying\ allele + cases\ not\ carrying\ allele}$$

$$specificity = \frac{controls\ not\ carrying\ allele}{controls\ not\ carrying\ allele + controls\ carrying\ allele}$$

$$positive\ likelihood\ ratio\ (LR^{+}) = \frac{sensitivity}{1 - specificity}$$

$$negative\ likelihood\ ratio\ (LR^{-}) = \frac{1 - sensitivity}{specificity}$$

$$pre\ test\ odds = \frac{ADR\ prevalence}{1 - ADR\ prevalence}$$

$$post\ test\ probability^{+} = \frac{pre\ test\ odds \times LR^{+}}{((pre\ test\ odds \times LR^{+}) + 1)}$$

$$post\ test\ probability^{-} = \frac{pre\ test\ odds \times LR^{-}}{((pre\ test\ odds \times LR^{-}) + 1)}$$

$$\begin{aligned} absolute\ risk\ reduction \\ &= number\ of\ ADRs\ per\ 100{,}000\ patients \\ &- individuals\ in\ population\ without\ allele\ (per\ 100{,}000) \\ &\times post\ test\ probability^{-}) \end{aligned}$$

$$screenings\ needed\ to\ prevent\ one\ ADR\ case = \frac{100{,}000}{absolute\ risk\ reduction}$$

6.2.3. Implementation

6.2.3.1. Database schema

The database for the clinical decision support tool followed a contextual 3rd normal form as described in Chapter 3.2.2 whereby the information in the database was decomposed into tables that always followed 1NF and 2NF but only followed the conditions of 3NF if it made contextual sense and only required a reasonable/limited number of table joins. The diagram schema is shown in Figure 6.3. The database itself was developed within the MariaDB RDBMS which allows it to be compatible with other MySQL based languages when it comes to rolling the web database for full/public deployment.

6.2.3.2. Website construction

The webpages that act as the interface for the user to access the information contained in the database were constructed using PHP 5.4 due to the greater market support for the language as mentioned in Chapter 1..5 – the language itself is more than capable of providing the desired functionality for the website include inbuilt support for accessing the commonly available RBDMS including MySQL. As well as PHP, additional functions are provided using JavaScript which is able to provide client-side functions (PHP is only server-side). These client-side only functions include the site's main page's ability to only allow the user to select a maximum of two alleles per HLA gene and for the site's results page to collapse the information card into a condensed version which can then be expanded - the user first sees the clinical decision support tool's recommendation and then can choose to expand the cards to view the supporting information such as the drug body guidelines, calculated statistics and links to the supporting evidence.

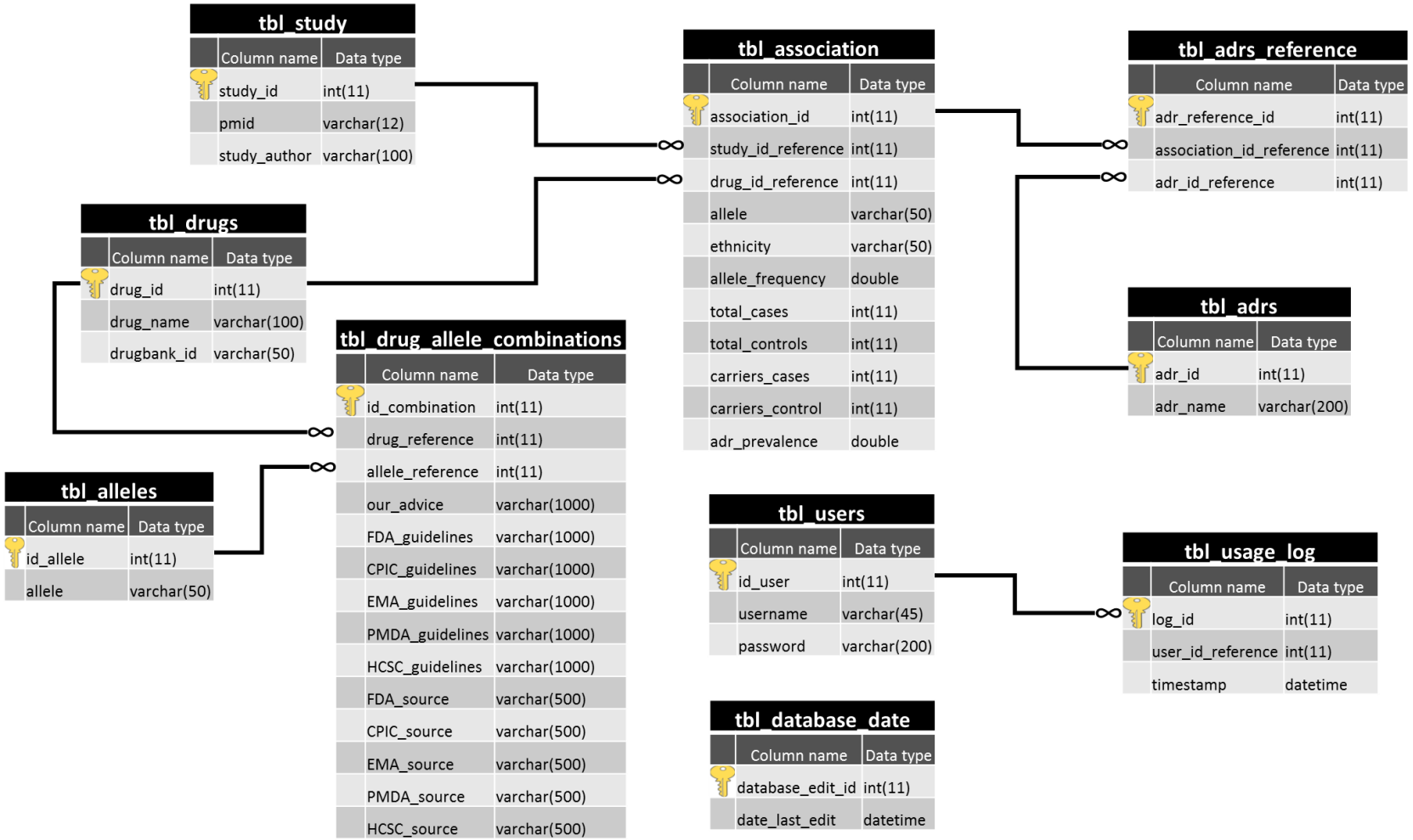


Figure 6.3: Database schema diagram for the clinical decision support tool.

The colour scheme was also carefully selected. Namely, the base colour scheme was kept with a simple white background and black text/objects as the general scheme. This was to provide a very minimalistic and clean look where the main focus of the site was the selection buttons and information cards which are the only areas of the site that are coloured. These selection buttons required a colour scheme which were distinguishable for people with common colour perception deficiencies (sometimes referred as "colour blindness"). As such, colours that are close to each other on the colour spectrum were avoided (e.g. blue and green on the same page were avoided). For the drug selection page, the blue colour was chosen for drug selection buttons and orange was chosen for the allele selection buttons. Green was chosen for the information cards.

6.2.4. Consultation for feedback

At the time of writing this thesis, the clinical decision support tool has not yet been fully deployed. One key task before deployment that is planned is the stage of external feedback. Currently, the tool has only been reviewed internally by those who have been directly involved with the project. MC Diagnostics (manufacturers of the Mr Spot machines) have made arrangements with a group of researchers to test the Mr Spot machine and have been provided with user credentials to access the clinical decision support tool. Feedback has not been provided as of yet. Additional testing has also been planned to gain feedback from clinicians. The feedback consultation is intended to provide a constructive critique for both the user experience (e.g. the user interface) as well as the information displayed to the end user. The feedback will be assessed by the members involved with development to decide what will be implemented.

As part of the initial feedback by the development team, the first design draft that was presented (shown in section 6.3.1), the drug-allele selection page the default HTML radio button layout was described as confusing, clunky and uninspiring. To address this, a layout similar to that of the figure (shown in Figure 6.1) from the previously published article depicting the 24 plus alleles that have been associated with HLA mediated ADRs was then used as a basis to design the buttons that eventually appeared in the final version – although the drugs and alleles were separated to allow for a

customisable search. Additionally, for the recommendation page, the table layout also received similarly negative remarks with regards to the user experience.

Only two people with colour perception deficiencies have viewed the website, both of whom have red-green colour vision deficiency. They both reported that the blue and orange colour scheme was easily distinguishable allowing the drug selection buttons to be discerned from the allele selection buttons.

6.3. Results

6.3.1. Initial design drafts

Before the website was constructed, rough design sketches were made for the project members to view, critique and vote upon. These sketches specifically relate to the graphical user interface (GUI) that end users will eventually interact with. The website design sketches are shown below. In these design sketches, the grey boxes represent the web elements that the user will be able to see and the black text next to these boxes provide a very short description to describe the type of HTML element that will be needed to achieve the web element.

6.3.1.1. Design drafts for drug-allele selection page

Two design drafts were created for the drug-allele selection page; a version where the user must select their drug and/or allele options using a dropdown menu (Figure 6.4) and the second version of the page where the user must select their options using radio buttons/checkboxes (Figure 6.5).

6.3.1.2. Design drafts for recommendation page

For the recommendation page, the information is proposed to be displayed in tables with the two supplied variants only differing in that the first variant (Figure 6.6) will display the alleles horizontally with the information below and the second variant displaying the same information in a vertical format. The second variant supplies the same information as the first but the tables are constructed in a vertically rendered alleles approach (Figure 6.7).

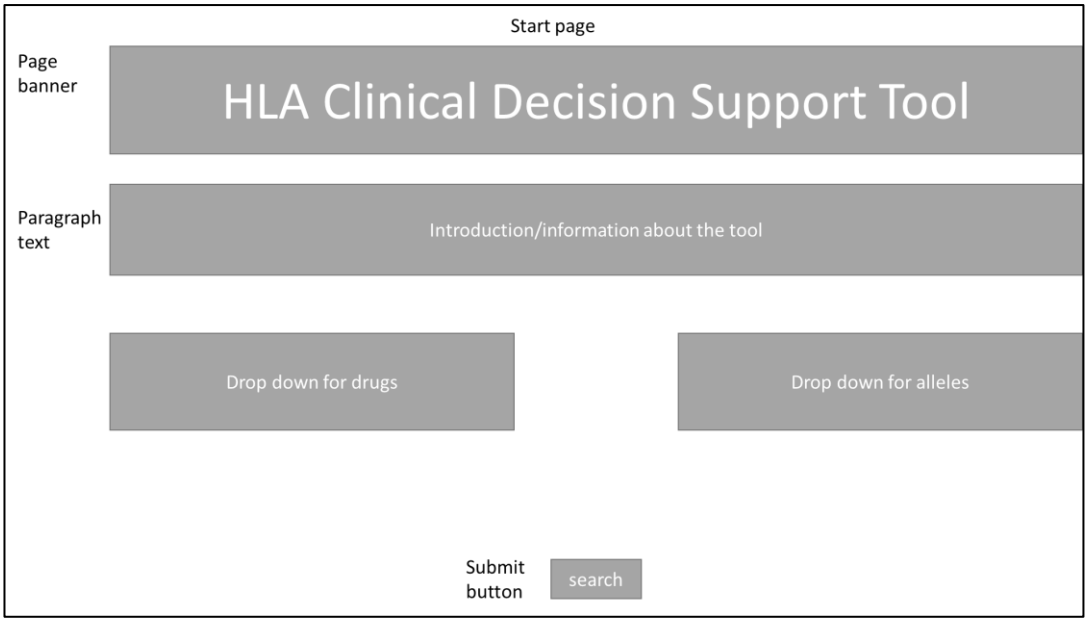


Figure 6.4: Design draft of drug-allele selection page number 1 – dropdown version.

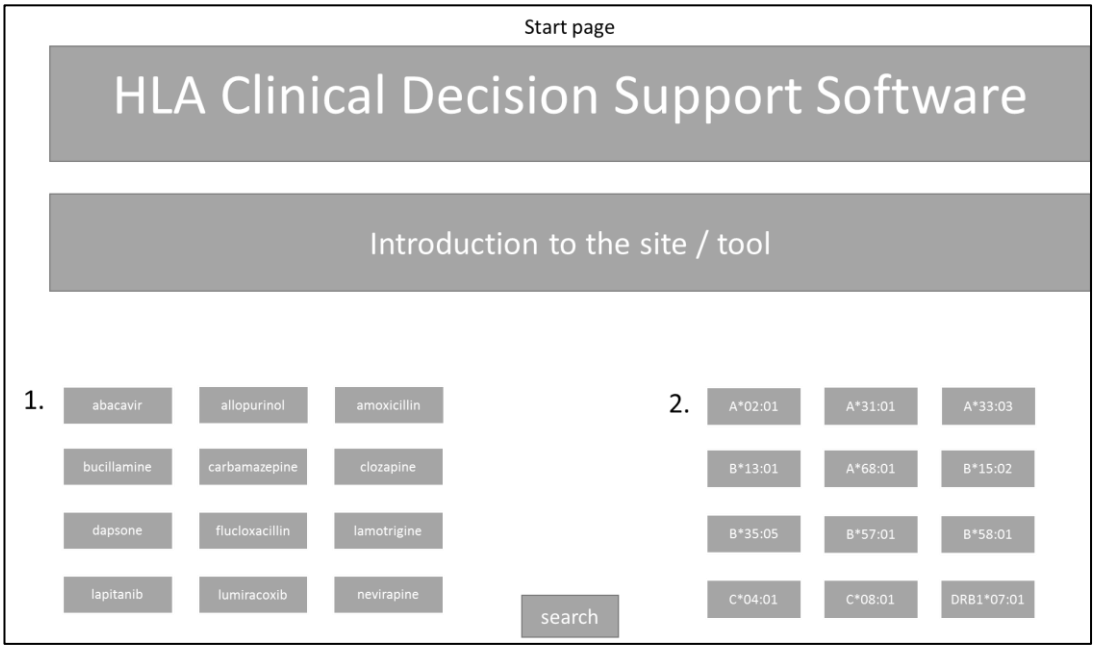


Figure 6.5: Design draft of drug-allele selection page number 2 – radio buttons version.

1. Drug selection radio buttons
2. Allele selection radio buttons

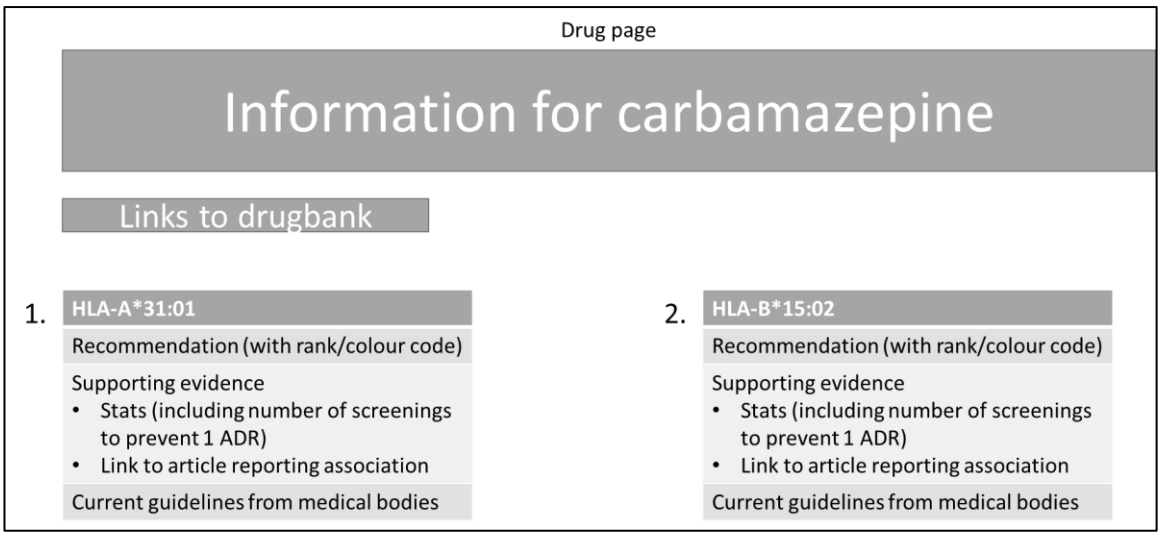


Figure 6.6: Design draft of recommendation page number 1 – horizontal table headers version in drug view.

Example search using carbamazepine option where 1 and 2 show the two returned alleles along with the relevant information.

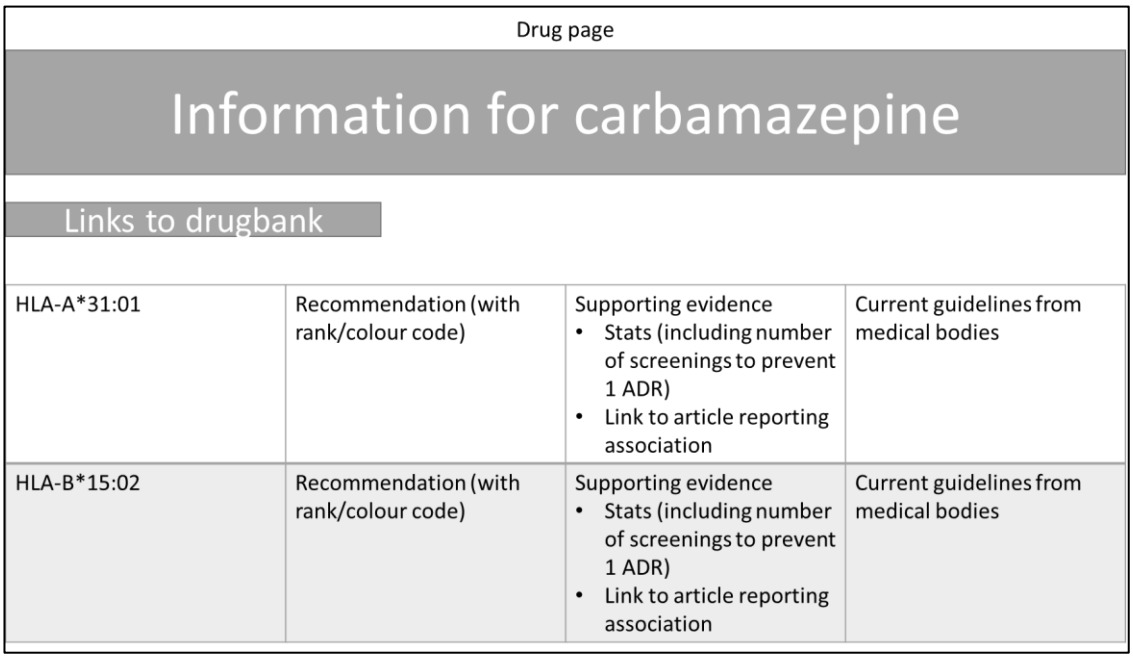


Figure 6.7: Design draft of recommendation page number 2 – vertical table headers version in drug view.

6.3.2. Skeletal design of website

From the proposed design drafts, drug-allele selection page variation 2 and recommendation page variation 1 were chosen to be created for the clinical decision support tool's beta version. These were constructed as described in the materials and methods section of this chapter and are shown in Figure 6.8 and Figure 6.9 respectively. As mentioned in section 6.2.4, regarding feedback, the layouts were modified and improved and shown later in section 6.3.3 – the version current as of writing this thesis.

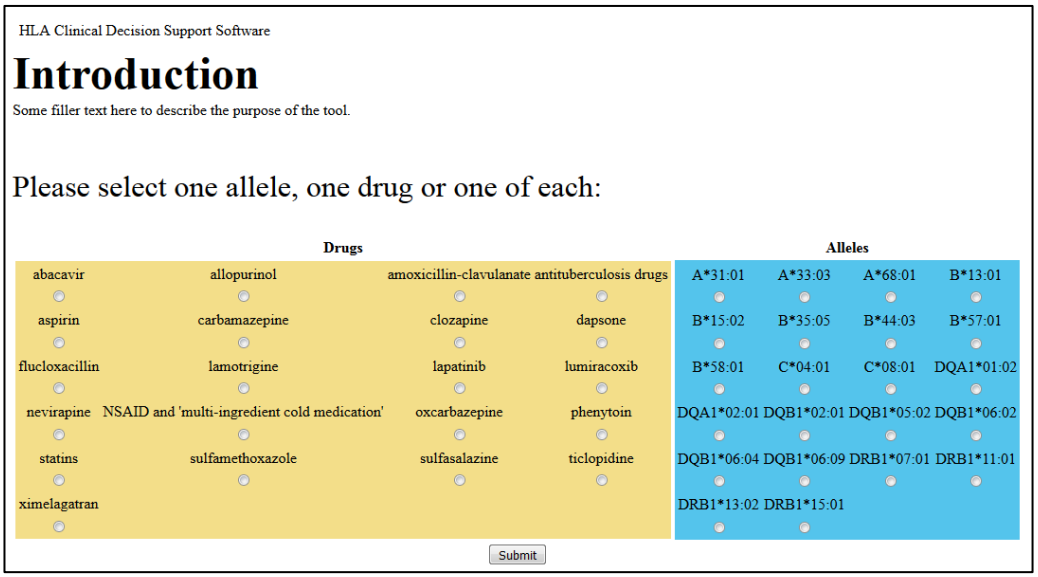


Figure 6.8: Drug-allele selection page – version 1 with radio buttons.

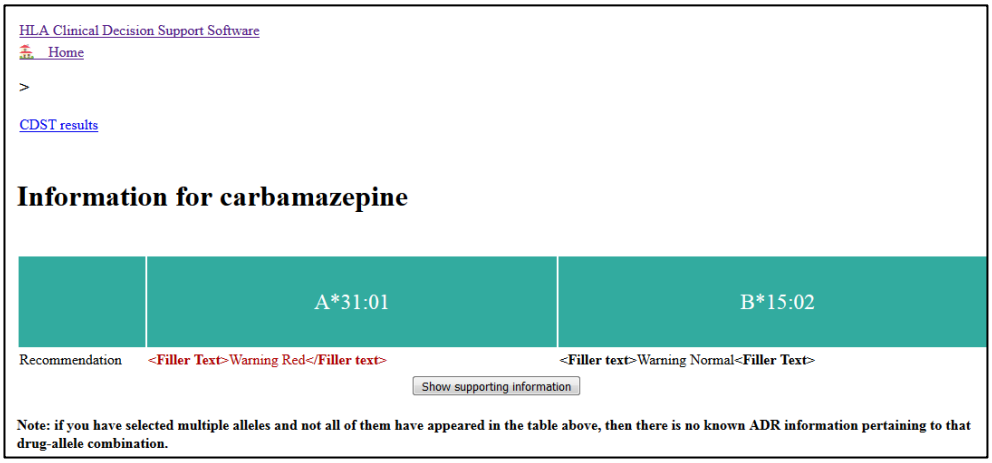


Figure 6.9: Recommendation page – version 1 results rendered in table with horizontally aligned alleles.

6.3.3. Website organisation

The webpages were designed to allow users to easily access and visualise the clinical decision support tool data in a user-friendly manner. The support tool is available with the URL: http://pgb.liv.ac.uk/~gurpreet/register_form which links to the support tool registration page where users can gain credentials to access the tool. Note that this is a link to the development server and is not designed for mainstream/high web traffic use scenarios. All of the webpages are generated using PHP which serves out HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) thereby allowing the retrieved data to be viewed in most common web browsers. The CSS style sheets have been enhanced using CSS Bootstrap which enables a responsive web design whereby the webpage to adjust itself based on the screen dimensions (measured in pixels) and display the Bootstrap objects appropriate for the screen, i.e. the layout of the website will be different when viewed with a large screen desktop computer vs a laptop screen vs a smartphone screen etc. [185]

6.3.3.1. Drug-allele selection page

After logging into the site the user is presented with the drug-allele selection page (Figure 6.10) which from a web development point of perspective acts as the home/index page. Here the user can select the allele(s) of interest which in the projected use case scenario would be the results from the Mr Spot machine. On the website, depicted in the figures below, the selection buttons for the alleles are coloured orange. The user can also select a drug (blue selection buttons) – the use case scenario for this is for a clinician to select the drug that they wish to prescribe to their patient.

Figure 6.10 shows the drug-allele selection page as rendered on a large desktop screen. The website is capable of being viewed on other devices where the website reorganises itself appropriately due to the above mentioned responsive design features of CSS Bootstrap. The responsive design of the website can be viewed in Figure 6.11 whereby the drug and allele selection buttons have arranged themselves to better fit the smaller dimensions of the screen that was used to view it. Instead of a $4 \times n$ grid for the drug selection buttons and a $4 \times n$ grid for the allele selection buttons, each set of buttons are displayed in a $2 \times n$ grid. On a smartphone, the selection buttons will also display

in a $2 \times n$ grid but the allele buttons will be displayed below the drug selection buttons.

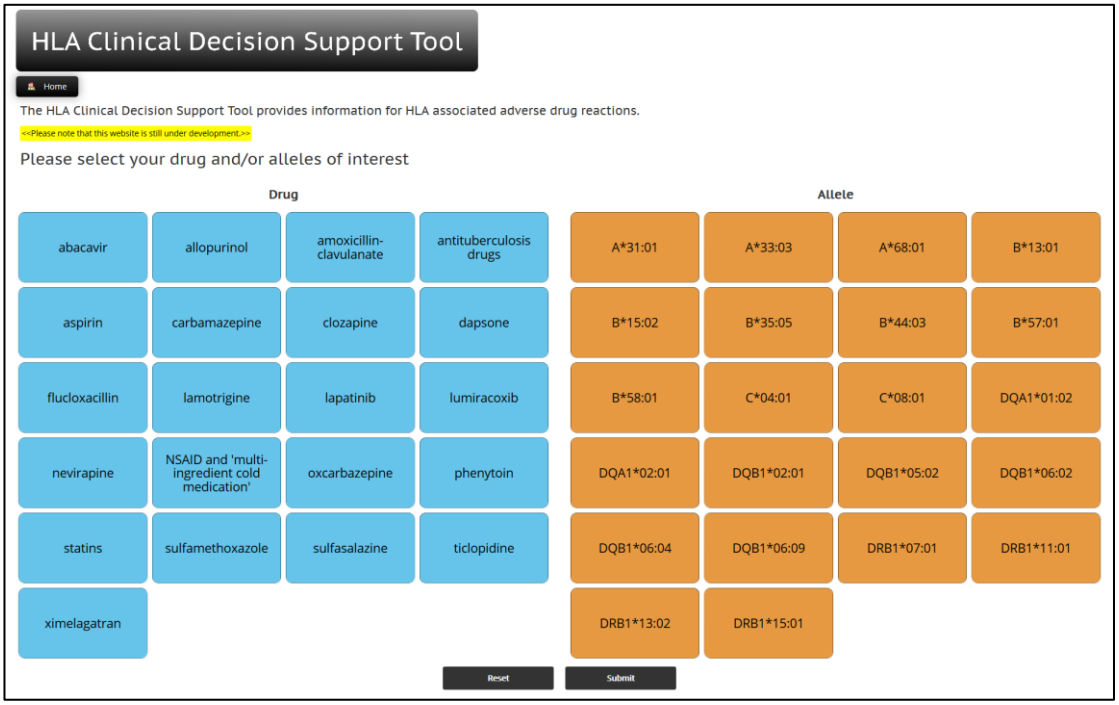


Figure 6.10: Screenshot of the clinical decision support tool's home page or drug-allele selection page.

6.3.3.2. Recommendation page

Once the user selects their drug and/or allele(s) of choice, they will then be presented with information cards that are relevant to the search criteria. In the example screenshot shown in Figure 6.12 below, where "carbamazepine" was selected for the drug option and "HLA-B*15:02" for the allele option, the recommendation page has rendered an information card where the relevant information stored in the underlying database has been retrieved in a similar manner as described in Chapter 3. The viewing format of these information cards varies depending on the options selected and is described later in section 6.3.3.3.

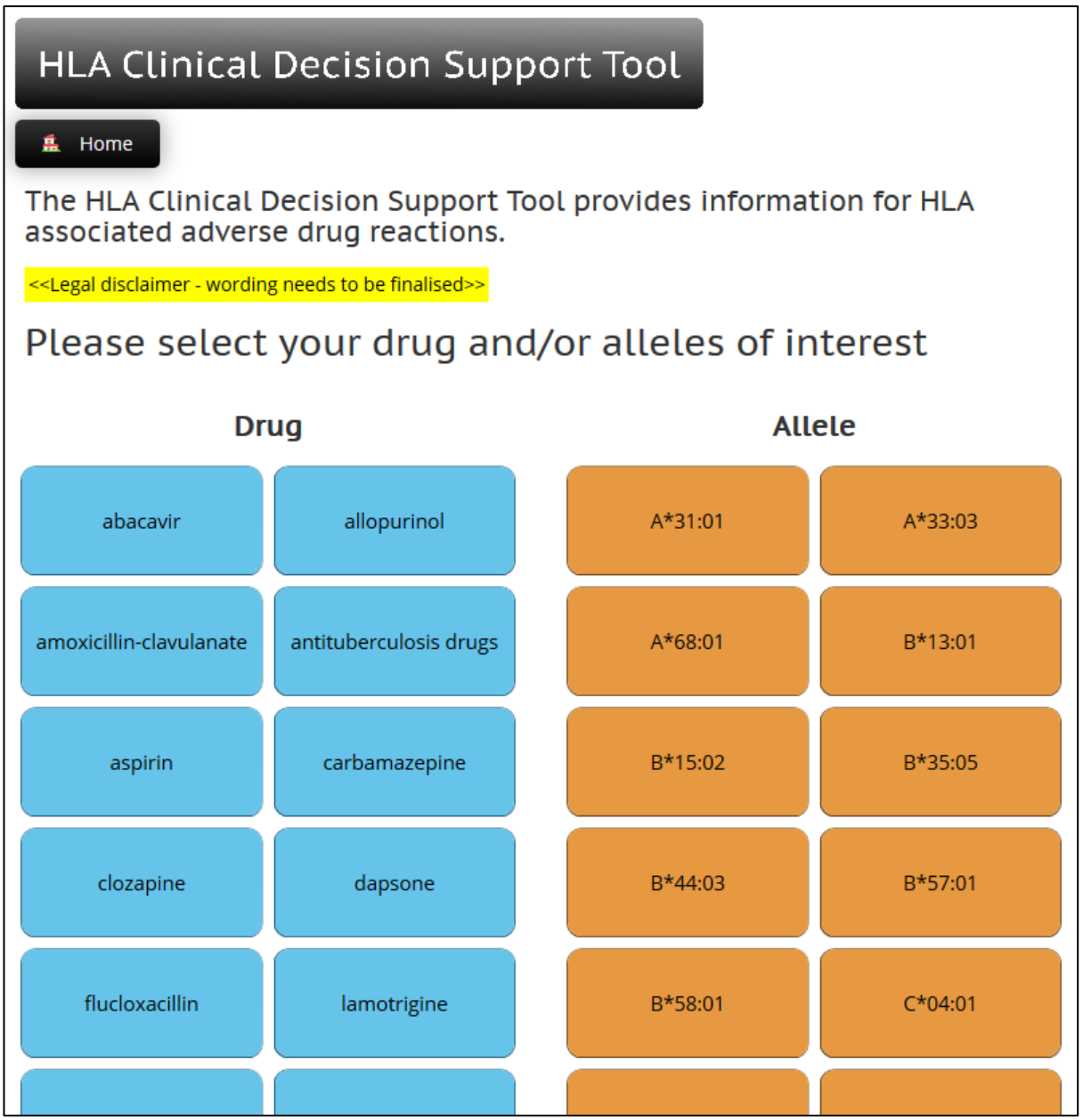


Figure 6.11: Screenshot of the clinical decision support tool's drug-allele selection page as viewed on a smaller screen compared to that of Figure 6.10.
Here, the drugs and allele selection buttons are arranged in a grid that is two buttons wide (as opposed to four buttons wide seen in Figure 6.10).

The information card is collapsed by default to only show the clinical decision support tool recommendation as depicted in Figure 6.12. If the user requires, the additional information can be viewed by clicking on the "show supporting information" button located at the bottom of the information card. Once clicked the card will expand and the user will also be able to view the available drug-allele guidelines which have been

set out by the regulatory drug bodies including: the USA's Food and Drug Administration (FDA); Clinical Pharmacogenetics Implementation Consortium (CPIC); the European Union's European Medicines Agency (EMA); the Japanese Pharmaceuticals and Medical Devices Agency (PMDA) and Health Canada / Santé Canada (HC-SC) – hyperlinks to the original source page are also included.

The expanded information card also displays the reported risk ADR phenotypes (e.g. SJS/TEN or DILI etc.) for that drug-allele combination; clinical statistics calculation for the number of screenings required to prevent one ADR case and hyperlinks to the studies that have reported the ADR. An example of the expanded information card using the example search where the drug is carbamazepine and the allele is HLA-B*15:02 is shown in Figure 6.13.

6.3.3.3. Recommendation page display mode

The website also changes its display mode depending on what the user selected on the drug-allele selection page. This causes a subtle change with the displaying of the information cards. If a user selects only a drug option or a drug and allele(s) option then the recommendation page will present the information in "drug" display mode. Alternatively, if the user only selects an allele(s) then the information cards are presented in "allele" display mode.

In drug display mode, the cards are formatted around the drug. The information cards are separated by allele based on either the user selection or if no allele was selected, the alleles that have known associations with the user selected drug. Figure 6.12above is an example of the results page presenting the information in the drug display mode. If no allele is selected then the example figure would show information cards for both the HLA-A*31:01 and HLA-B*15:02 alleles that have been associated with carbamazepine hypersensitivity. Figure 6.14 shows the slight variation in the allele display mode. In this view, the recommendations are separated by allele. This two display mode feature is a requirement to display information based on user input as the information cannot be displayed uniformly if the user selects a drug; a drug and allele or just an allele.

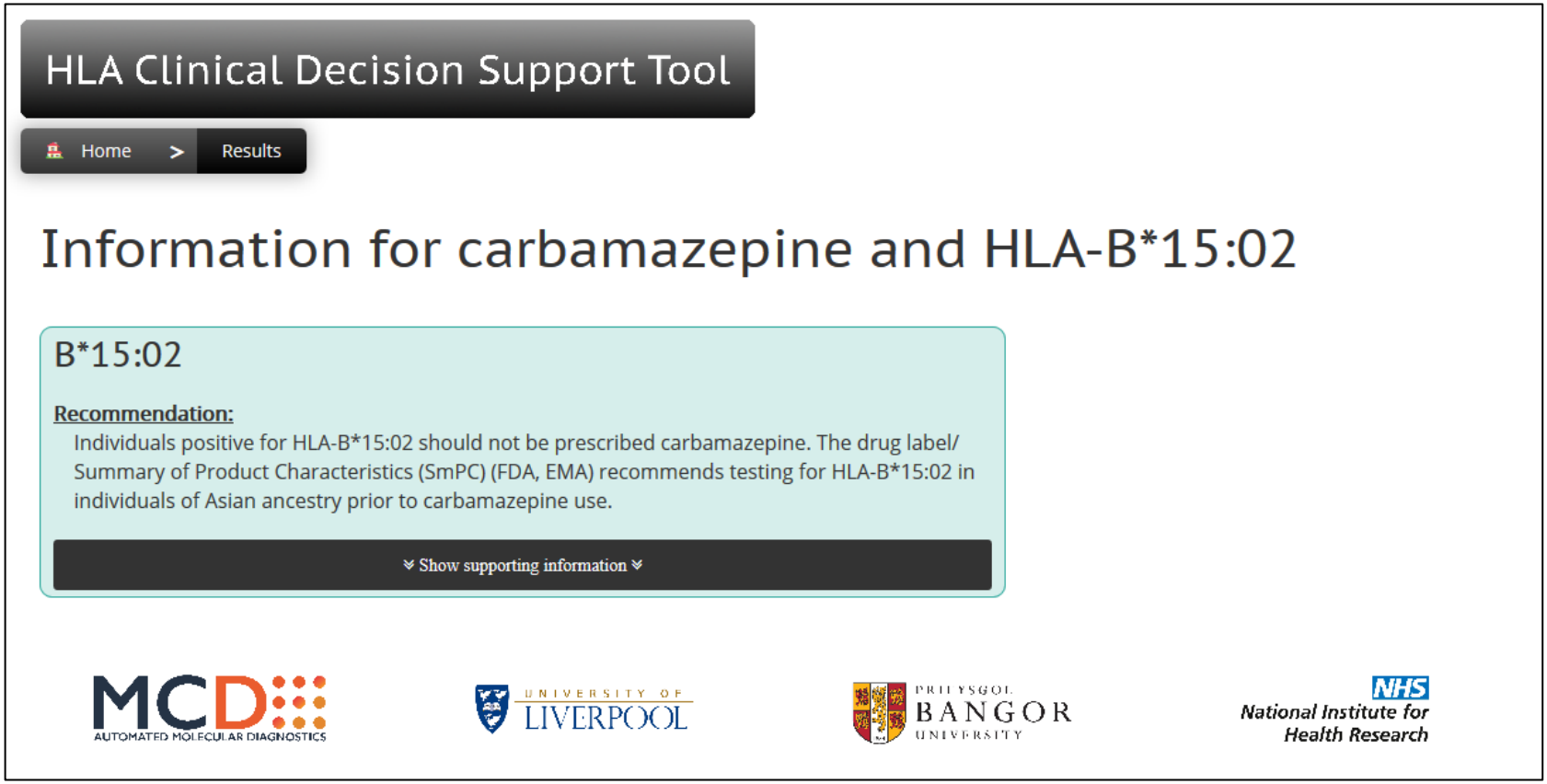


Figure 6.12: Screenshot of the clinical decision support tool's results page or recommendation page.

In this screenshot an example search for "carbamazepine" and "HLA-B*15:02" were chosen for the drug and allele selection respectively in the drug-allele selection page which has rendered the page display mode to the "drug" view.

Figure 6.13: Screenshot of the recommendation page with the information card fully expanded (example selection of “carbamazepine” and “HLA-B*15:02”.

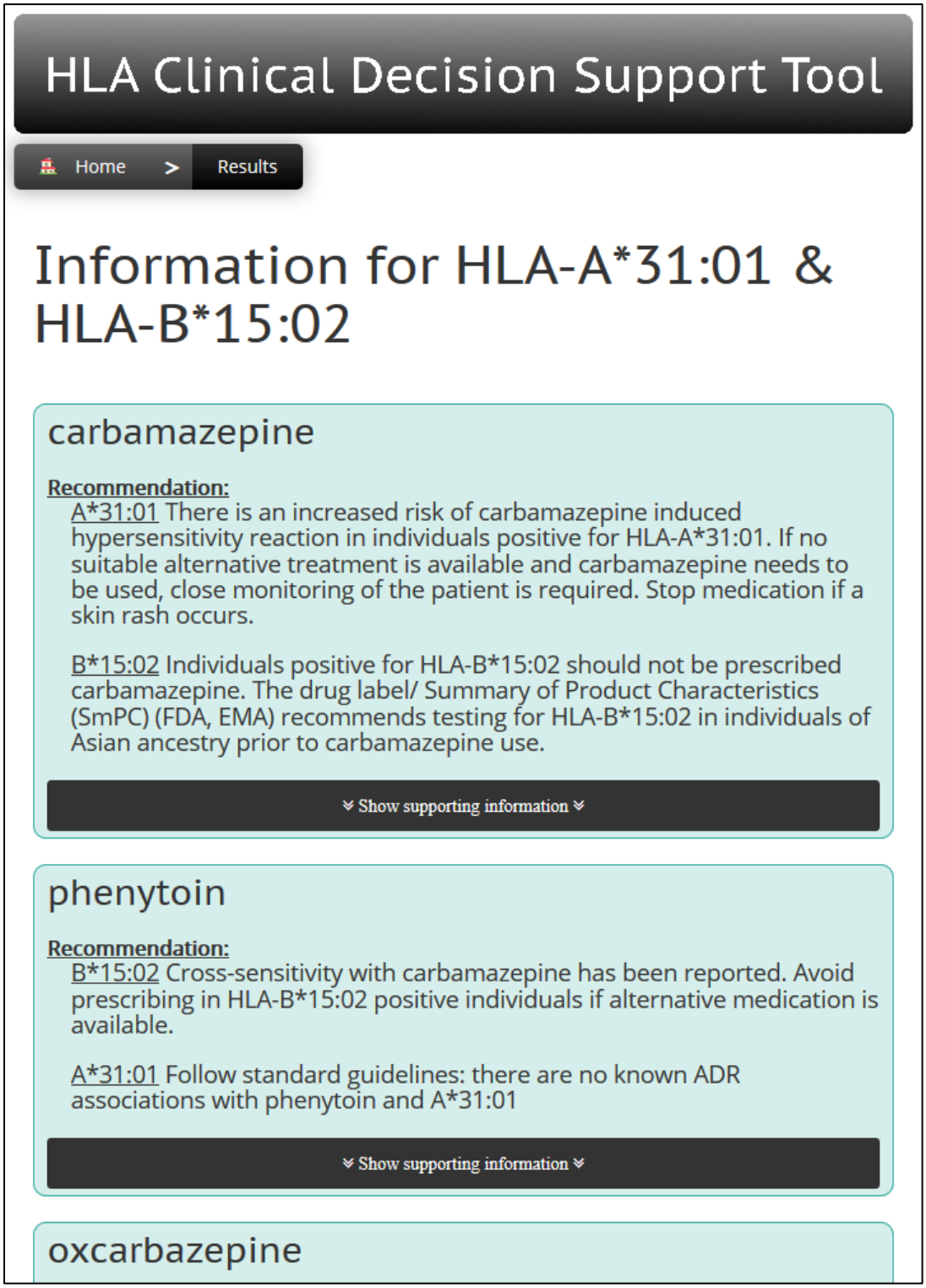


Figure 6.14: Screenshot of the clinical decision support tool's or recommendation page.

In this screenshot an example search of "HLA-A*31:01" and "HLA-B*15:02" were chosen for the allele selection respectively in the drug-allele selection page which has rendered the page display mode to the "allele" view.

6.4. Discussion

6.4.1. Effectiveness and limitations of presenting relevant information

Whilst the process of provided feedback by external testers has not been carried out, initial feedback from the members of the internal project team about the presentation of relevant information has been positive. Of particular note was the aesthetic layout which made the presentation of the information easier/intuitive for reading compared to the earlier builds of the website.

In addition to the aesthetics, feedback was provided for the content contained in the information card; the level of information provided by the recommendation, the drug body guidelines and references to the reporting literature, with hyperlinks to the PubMed/MEDLINE abstract were received positively. There were comments that the provided clinical statistics were limited – the tool currently only shows the range for the numbers needed to screen to prevent one ADR. The website itself does calculate the other clinical statistics from the data as described in section 6.2.2 however, displaying all the information would make the information card far too large and would potentially bombard the user with too much data. Creating an additional page that displays the entirety of the clinical statistics calculations would resolve this issue – this solution has been noted as a future update.

One major limitation that was noted when analysing the underlying dataset was the quality of some of the data. It was observed that in some instances the sensitivity, specificity, positive/negative predictive values were a perfect 1.00 and that the most likely reason behind this was the fact that the results were skewed/heavily influenced by small sample sizes. Table 6.1 shows such examples of small sample skewing which lead to clinical stats artificially indicating high usefulness of the tests. In the first row, with carbamazepine, all five patients carried the risk HLA-B*15:02 allele with only 10% of the controls doing so. Whilst the association itself has been repeatedly demonstrated the phenomenon has greater impact with associations with fewer studies supporting the link – this is demonstrated in the second row where the association with lamotrigine and the HLA-A*68:01 allele for SJS/TEN. The risk allele was only found in a small percentage of patients and none of the controls. As a result of this, the

eventual recommendation that is provided by the clinical decision support tool reads as "There may be an increased risk of lamotrigine-induced hypersensitivity syndrome in individuals positive for HLA-A*68:01, but at this stage no recommendation can be made regards screening." Situations such as this greatly highlights the need to manually curate the data rather than creating an algorithm to come up with the suggestion. The algorithm was initially considered with the design of the clinical decision support tool but this was rejected due to other limitations discussed later in section 6.4.2.2. The example of HLA-A*68:01 and lamotrigine is not as widely reported as the example with HLA-B*15:02 and carbamazepine induced hypersensitivity so the interpretation of results become less reliable as fewer studies are available for review into the overall analysis. The statistics derived from other studies reporting the HLA-B*15:02 and carbamazepine can be used to provide context as to the impact of the study shown in the example table, however the same is not applicable for lamotrigine and HLA-A*68:01.

Another limitation for this clinical decision support tool will be that the recommendations will need to be population specific. As discussed in 0, some alleles are more prevalent in specific populations and as such the recommendations will need to reflect this. This information was taken into account when devising the current recommendations e.g. the recommendation relating to carbamazepine only recommends screening patients for the HLA-B*15:02 allele if they are of Asian ancestry. For this reason it is therefore imperative that future updates to the database continue to factor in population differences.

Table 6.1: Examples of data within the clinical decision support tool where the sensitivity, specificity and positive/negative predictive values (PPV/NPV) were calculated to be 1.00.

Reference	Drug	Risk HLA Allele	ADR	Patient Ethnicity	Patients with risk allele	Controls with risk allele	ADR prevalence	Sensitivity	Specificity	PPV	NPV
Chong et al. 2014	carbamazepine	B*15:02	SJS/TEN	Diverse	5 / 5	1 / 10	0.0001	1	0.9	0.001	1
Kazeem et al. 2009	lamotrigine	A*68:01	SJS/TEN, DIHS	Caucasian	4 / 44	0 / 86	0.001	0.091	1	1	0.999

Since the clinical decision support tool aims to provide interpretation of the results from Mr Spot, the cost effectiveness from a clinical perspective for the clinical decision support tool is directly tied to the cost effectiveness of the Mr Spot assay test and is therefore already discussed in Chapter 5. However, the overall effectiveness, is a little more difficult to discuss without the external feedback/testing that is pending. Whilst not being able to fully address this point of effectiveness of the tool, a small feasibility study which tested a medication decision support system with a panel consisting of HLA-B*58:01 and 32 non HLA genes implicated with drug-induced ADRs reported successful outcome of the study goals. [186] However, this study does not go into great detail as to the feedback relating to the decision support system but the study does report relatively high usage of the software indicating that the clinical utility was high although the full context of the usage is not discussed in the study so whilst interpretation of this point is positive on a surface level, it requires more detail for conclusive interpretation.

Additionally, individual tests that are already recommended in primary care (e.g. screening for HLA-B*57:01 prior to abacavir treatment) indicates that uptake and desire for such screening/testing has potential and that the relative cost benefits and expansion of available alleles will likely further drive demand. Since the regulatory drug bodies are relatively slow to enforce alterations to the drug labels, high scrutiny of the available data is likely to be of benefit in order to reduce falsely demonising a treatment option, this can leave clinicians from missing out on relevant knowledge for other drug-allele combinations. This is showcased in that only a few of the drug-allele combinations known to cause ADRs have led to label modifications to include warnings (For FDA controlled drug labels, only abacavir, carbamazepine, lapatinib, oxcarbazepine and phenytoin have such warnings and this is the regulatory body with the most amount of enforced label warnings relating to HLA mediated ADRs). The database for the clinical decision support tool can be updated quicker as the data is analysed by experts in clinical pharmacogenetics who are already familiar with the field and are actively looking for newly published literature affording them with up to date information.

6.4.2. Future work

6.4.2.1. Database maintenance

The clinical decision support tool was specifically designed from the ground up for expansion/updating in mind. The database was deliberately over normalised relative to the data that was going in at construction but the computational benefits of this normalisation process would be seen once more data is inputted. The website's drug-allele selection page also renders the selection buttons by pulling the data from the database rather than being manually constructed. Individually created buttons would be faster to load for the end user, however, each new drug and allele would need to be manually created as they are added. The method that is employed by the website actually allows the clinical decision support tool's development team to only change the database and the website will automatically reflect these changes. This system makes sure that the website is easier to maintain as specific changes can be reflected by all aspects of the database with relatively little input by the development team. New drug-allele associations can be added which are already expected. The chief investigator of this project has indicated that since the time of the website's construction, new drug-allele associations have been published or are pending publication and will need to be considered for inclusion.

The system was also designed so that the clinical experts can comfortably add any such new data (or make alterations to existing data) using a familiar data storage package (spreadsheets) and pass on the file to the database administrators where they can execute a pre-existing script to commit the changes to the MySQL database. This allows the database to be maintained with greater flexibility as it does not require an expert in both HLA mediated ADRs and database management in order to make the changes as these tasks have been deliberately separated. An example of this has already been realised, since the original proposal consisting of 24 alleles two alleles were removed as further evaluation of the data indicated that there was little evidence and clinical utility for these particular alleles to be included. The removal of the alleles from the spreadsheet database was the only manual input required where it cascaded and was automatically reflected in the removal from the database and the website.

6.4.2.2. Role of artificial intelligence

Earlier in this discussion, it was briefly mentioned about the initial prospect of an algorithm to formulate the recommendations as part of the clinical decision support tool. However, at this current time there are limitations regarding the introduction that would render this consideration to be rejected. The largest issues for artificial intelligence are with public trust and the legality, in 2017 there were news reports of a partnership between the NHS and Google's DeepMind project that intended to identify patients that were at risk of developing kidney disease – the deal was subsequently regarded as illegal. [187] As well as the legal issues, the news article featured a quote stating about the public mistrust with sharing data with Google. A lot of the problems faced in this project can possibly be attributed to the concept of allowing artificial intelligence to diagnose patients being a very new technology. As such there is no point of reference/comparison to assess the potential impact from all aspects (societal, ethical, legal liability as well as the direct clinical impact). Due to the relative infancy of implementing artificial intelligence that has direct public interaction beyond low impact functions; Google's DeepMind was previously involved in a project for playing recreational games so was removed of major societal, ethical and legal liability hurdles; more time is needed to allow artificial intelligence to develop and incorporate itself in other, low impact aspects of public interaction before it can be accepted for use in clinical practice. If such a time is reached, then the concept of developing algorithms in clinical scenarios can be revisited as it would make a very interesting avenue to pursue.

Chapter 7. Final discussion

Adverse drug reactions (ADRs) contribute to between 6 and 12% of all hospital admissions in the United Kingdom. [1, 2] ADRs also lead to the withdrawal of 3-5% of all drugs that were newly approved between 2002-2011 and 1990-2009 in Europe and North America respectively where hepatotoxicity was the second most commonly reported reason. [3, 4] This thesis has focused on "Type B" ADRs also referred to as immune-mediated ADRs where the carriage of certain human leucocyte antigens (HLA) have been associated as a risk factor. [19, 24, 27, 34-36] These reactions represent approximately 20% of all ADRs. [31] The clinical manifestations of these ADRs can range from simple skin rashes which can be tolerated but can also lead to severe morbidity and possible even death for the patient. [9, 11] As a result, ADRs represent a major challenge for clinicians and pharmaceutical companies. The mechanisms by which and ADRs occur is not fully understood although the leading hypotheses involve the drug (or metabolite) interfering with how the T-cell interacts with antigen (HLA) molecule and therefore disrupts how the immune system recognises between self and non-self cells. [10, 38, 40, 43, 44, 46] In the case of abacavir, while there is published evidence that the drug interacts noncovalently with HLA-B*57:01, it has also been demonstrated that abacavir can form hapten type interactions with human serum albumin. [46, 188] Abacavir's ability to form both covalent and noncovalent interactions indicates that there is potential that the mechanism of action for ADRs may not follow one exclusive pathway and instead are complementary.

7.1.1. Evaluation of literature review

The findings from Chapter 2 shows a significant proportion of the published literature, 65%, emanated from studies conducted using a patient-control cohort of Asian ancestry. The chapter discusses the possibility of publication bias as there is a noticeable shortage of studies conducted using patients of South American or African origin (1 and 2 respectively out of 85 studies). In addition to this, a hypothesis was proposed that since the major pharmaceutical companies are based in Europe and/or North America, then the drugs' clinical trials were conducted on cohorts of predominantly European origin. Admittedly, this is difficult to test and/or verify due

to the information regarding the locations of the original drug trials being difficult to locate.

In addition to the literature review performed in Chapter 2, a semi-automated literature mining script was developed to speed up the process of performing future data collections from the literature. The dramatic decrease in time taken to collect data allows for regular updates for the HLA-ADR database that was developed for Chapter 3 and the database for the clinical decision support tool of Chapter 6. The script implemented a keyword search of the titles and abstracts of searches performed in Medline/Pubmed. The performance of the script was compared against searches performed manually within the chapter as well as against the literature reviews of colleagues who performed similar searches where the premise significantly overlapped. [98, 114] It is worth reiterating that this script would only substitute the first stage of the review process (analysis of the title and abstracts only) and is not capable of performing the second stage of the review process (analysis of the complete article). The analysis of the script's performance revealed that with knowledge of the specific ADR (drug, allele and ADR phenotypes) that the script is capable of highlighting virtually all relevant studies. The script's general performance was measured to be >90% in terms of forwarding relevant studies for the user to review manually for the second stage of the review. Therefore, it was determined that in circumstances where absolute accuracy was not required relative to time taken to perform the review, then the script offers a very appealing alternative to manual review. For the literature search in Chapter 2, the first stage of the manual review took four weeks to complete vs one hour for the semi-automated literature mining script. Whilst the semi-automated literature mining script shows great promise for utility, there is still a case to be argued for performing literature reviews manually. This is the current arrangement for the HLA-ADR tool (Chapter 3) is for the semi-automated literature mining script be executed regularly and a manual review on a semi-regular basis (currently set to be 1 year and 3-5 years respectively).

7.1.2. Evaluation of HLA-ADR

A web database (HLA-ADR) was created to allow relatively easy access to the data collected in Chapter 2. [189] The overall aim of this website is to reduce the time and

effort required to investigate adverse drug reactions by researchers. The website also links out to information from the Allele Frequency Net Database (AFND) where the website allows users to access stored allele, haplotype and genotype frequencies of several immune related genes including HLAs. [120, 121] As a result of this overlap it was decided to make the HLA-ADR database as a module within AFND.

The Query page allows users to interact with the information held in the database by forming customised queries using the dropdown menus provided on the page. The ADR reports page provides a summary of all significant associations for a particular drug and also links out to external resources (e.g. the main AFND website and IMGT/HLA site). [78] The site also enables researchers to submit their own data via the authors' submission page.

The centralised nature of the web database allows it to achieve its primary aim of allowing researchers accessing information (regarding immune mediated ADRs) that is spread across the literature. Whilst the website has been active for a relatively short period of time its use has been reported in very recent (non-self-citing) literature and the web database was used to aid in the investigations carried out in Chapter 4 of this thesis and as the basis for the underlying data in Chapter 6. [122, 123] The ultimate goal for any bioinformatical tool is to aid in the research of many studies.

In addition to the direct usage of HLA-ADR to assist other researchers with their investigations into immune-mediated ADRs, the HLA-ADR database has been referenced as an available tool to aid in the study of ADRs. [190] The review makes reference to another database that provides data for HLA allele associations with ADRs – HLADR. [191] This database has a larger available dataset to choose from as the authors chose to implement less stringent inclusion criteria (e.g. 1st field resolution typing). The database also shows information pertaining to the drug, HLA allele, ADR phenotype, ethnicity of patient-control cohort etc. HLADR provides additional information from that of HLA-ADR as it is capable of calculating the odds ratios, sensitivity, specificity, PPV and NPV. However, compared to HLA-ADR the ability to customise a search is limited. A user may narrow down the drug or HLA allele or ADR phenotype, but not a combination of two or more of these. There is also no evidence that the database has been updated since release (three years from the point

of this thesis) despite the authors' promise of updates every three months. Like HLA-ADR this database also has two citations. [192, 193]

7.1.3. Ethnicity review

The analysis of Chapter 4 also showed an unexpectedly high number of studies that originated in East Asia, the HLA-ADR web database was utilised to investigate the possibility that Asians are at greater risk of developing immune-mediated ADRs. The hypothesis is reported in individual instances where for example, carbamazepine hypersensitivity is reported to be more prevalent in patients of Asian origin to the point where the major drug/medicines regulatory bodies recommend pre-emptive screening of HLA-B*15:02 for patients of Asians decent (as mentioned at the start of this chapter). [155, 175, 194]

In this analysis, only alleles that showed a statistically significant association ($p < 0.05$) were selected and the alleles were analysed on a per drug basis. The frequencies of each allele was obtained using AFND. The haplotype information was also recorded to determine reported alleles were likely to co-occur along with the population/ethnic information. Analysis of this data indicated that where a particular drug had more than one reported HLA allele that has been associated as a risk factor, the alleles in question were likely to be in linkage in the reporting studies' populations groups. For example, in the case of allopurinol, the alleles HLA-B*58:01, HLA-A*33:03, HLA-C*03:02 and DRB1*03:01 have been reported to be risk factors. Of these alleles: HLA-B*58:01, HLA-A*33:03 and HLA-C*03:02 were found to be significantly associated in Han Chinese and Korean patients and the analysis conducted in Chapter 4 demonstrated that these three alleles co-occur at a relatively high frequency in East Asian populations (up to 8%). [133, 135] The analysis also revealed that the co-occurrence of HLA-A*33:03, HLA-B*58:01 and HLA-DRB1*03:01 can be found at relatively high frequency (up to 6.6%) in a Taiwanese population and that all four alleles can be found as a haplotype in south-west China at a frequency of 4.4%. [120, 124] This pattern of co-occurring risk factor alleles was found in many other drugs within Chapter 4 indicating that some of the co-occurring alleles were reported as risk factors due to the linkage disequilibrium existing between the allele in question and the true risk factor allele.

This phenomenon of co-occurring false positive alleles can be seen most evidently in carbamazepine where 11 alleles have been reported to be associated with carbamazepine induced hypersensitivity. After performing a Spearman's correlation analysis, only 2 of the 11 alleles (HLA-A*31:01 and HLA-B*39:02) were found to be statistically independent of any linkage disequilibrium of the other alleles. This was manually confirmed using AFND haplotype search. It is likely that in the case of the 9 out of 11 alleles showing linkage disequilibrium, the true risk factor is likely to be HLA-B*15:02 as this allele was reported in 19 independent studies with the other alleles from this group being reported only a maximum of twice. [17, 34, 144, 145, 149, 150, 156, 172, 194-205]

Additionally, an observation was made that the literature showed an over representation of studies conducted in patients of Asian or European origin (based on the findings of Chapter 2). The Kruskal-Wallis test performed in the chapter indicates that of the 33 HLA alleles that were analysed (where the reported drug-allele ADR association has $p \leq 0.01$ and/or independently replicated in two or more studies), only six alleles could be described as showing a higher allele frequency in Asian populations compared to two in African and two in Europeans. Of particular interest was the antiepileptic drugs where 43%, 33% and 67% of the alleles associated with carbamazepine, lamotrigine and phenytoin respectively can be found in higher frequencies in Asians. For these drugs, none of the remaining alleles could be only be described as "neutral" i.e. no significant association with European or African population.

This analysis shows inconclusive evidence to support the hypothesis that Asians are at greater risk for ADRs as the majority (23 out of 33) HLA alleles that show strong evidence of immune mediated ADR associations are classed as "neutral" – where there is no population that shows significantly greater prevalence for the risk factor HLA allele. However there are 10 drug-allele combinations where the patient's ethnicity may need to be taken into account for clinical decision making (as is already the case for HLA-B*15:02 and carbamazepine and other related antiepileptic drugs). This indicates that the general literature should not focus on patient ethnicity when investigating the wider picture for immune mediated ADRs even though it cannot be

ignored in individual examples of drug-allele combinations. This is evident in that whilst HLA-B*15:02 has a large focus with regards to patients of Asian origin, the allele has been reported in the relatively smaller Spanish Romani population. [206]

The analysis in this chapter again highlighted the inconstancy in reporting immune mediated ADR associations as many drug-allele associations were not well replicated and some alleles appearing to be significant in one study and nonsignificant in another (e.g. HLA-C*04:01 is reported to be significant and nonsignificant in different studies [35, 36] and similarly HLA-B*35:05 [36, 159] also for nevirapine). This limitation in the data makes it difficult to say with confidence as to whether an allele is a true risk factor. The relatively small sized patient-control cohorts found in many of the studies likely contributed as each individual patient/alleles had a larger impact on the final result and thereby disproportionately increasing the significance for a certain drug-allele association. Due to the relative rarity of ADRs (see USB/CD media file "chapter_6_decision_support_table.xlsx" for ADR prevalence values) the problem of the patient-control cohort sizes may be impractical to overcome.

7.1.4. HISTO SPOT HLA typing and the clinical decision support system

The medicines/drug regulatory bodies for the United States, Canada and Europe only recommend screening for HLA-B*57:01 in patients who are about to receive abacavir [49, 50] and for HLA-B*15:02 in patients of Asian ancestry when prescribing carbamazepine, oxcarbazepine or phenytoin. [51, 52] The other drug allele combinations that have been discussed in this thesis do not have any recommendations for pre-emptive screening; e.g. for allopurinol and HLA-B*58:01 – the European medicines agency does not advocate screening but if the patient is already known to be a carrier for HLA-B*58:01 then it does recommend the additional risk be taken into account when making the clinical decision. [52] There has been evidence recently published that states pre-emptive screening for HLA-A*31:01 (for carbamazepine) and HLA-B*58:01 for allopurinol would be a cost effective policy to implement in addition to the drug-allele combinations previously mentioned in this paragraph. [182, 183]

Conversely, the example of flucloxacillin induced DILI in relation to HLA-B*57:01 carriage, the cost-effectiveness for pre-emptive screening is weaker. Although, the negative predictive value (NPV) is 0.9999 (calculated using Daly et al., 2009 data for the clinical decision support tool in Chapter 6), the positive predictive value (PPV) is 0.0011. [24] This difference in NPV and PPV indicates that patients who experience DILI are likely to carry the HLA-B*57:01 allele, however this still only represents a small fraction of flucloxacillin taking patients that carry HLA-B*57:01. In other words, carriage of HLA-B*57:01 is alone not enough to predict DILI. This only goes to highlight that the exact mechanism for ADRs is not fully understood and more research is still needed.

Computerised clinical decision support systems (CDS) have been implemented in many forms across many aspects of health care that are integrated into electronic health records. For example, there have been CDS systems that have been trialled for management of familial cancer risk, identifying patients at risk of developing diabetic foot complications, prescription management for patients with reduced renal function etc. [207-209] These examples show that there is both a need and demand for such tools. Multiple reviews have presented the benefits that CDS systems in the context of clinical decision making, although they also recommend exercising caution as CDS systems are still in their relative infancy and has not yet met widespread adoption.[210-214]

The clinical decision support system that was developed in this thesis was designed to complement the new HISTO SPOT HLA typing kit (Chapters 6 and 5 respectively). The combination of the two technologies aims to aid clinicians in predicting, pre-empting, preventing and diagnosing immune mediated ADRs. The clinical decision support tool reports current guidelines from various drugs and medicines regulatory bodies (United States FDA, Canada's HC-SC, the Japanese PMDA, the European Medicines Agency and the Clinical Pharmacogenetics Implementation Consortium. However, as discussed earlier in this chapter, these guidelines are not comprehensive. HISTO SPOT and the associated clinical decision support tool looks to expand on this information by providing its own recommendation based for a relatively larger panel of drugs and alleles. The information provided by the clinical decision support tool is relatively simplistic (deliberately so) to enable only the necessary information to be

conveyed. The provided recommendations are only designed to enable clinicians to make an informed decision. Given that, as discussed earlier in the chapter, the HLA carriage status alone is not enough to induce an ADR, it is important that the clinical decision support tool not be used for decision making and that the clinician assesses other relevant clinical factors for each patient.

Whilst the HISTO SPOT HLA typing assay and the associated clinical decision support tool have not yet been released, the growth in technologies for HLA typing, as shown in the various methods discussed in the general introduction (Chapter 1.3), there is clearly a desire for cost effective methods with a very fast turnaround. This assay has the capability to achieve this as the 48 hour turnaround, accurate HLA calling and with a single test for over 20 alleles is cheaper than currently available HLA typing methods places it in a good position to achieve this goal of clinical adoption.

7.2. Future directions

7.2.1. Future directions from this thesis

The semi-automated literature mining script has a lot of potential to be expanded upon. The benefits for time saving whilst still maintaining a relatively high level of performance accuracy has demonstrated its value. One area that the script can develop on is the ability to identify adverse drug reactions that have not been pre-programmed in. For example, the script was initially not able to identify anti-hydroxymethylglutaryl-coenzyme A reductase induced myopathy. An aspect of deep learning in the form of natural language processing is an area of interest to help identify such papers by analysing the text of the title/abstract and noting the similarity with other included studies – i.e. the way a drug induced ADR is referenced in general rather than exact keyword based criteria. [215] However, the implementation of natural language processing and other forms of deep learning in clinical data is still in its infancy, considerably more research is still required.

It is also imperative that the HLA-ADR database continue to be maintained. More HLA associations with ADRs are being reported in the literature meaning the database needs to be updated to include this new data for the tool to continue to be of use to

researchers. This will also apply to the clinical decision support tool once that is publically released. There are also conversations taking place to expand the clinical decision support tool to include some of the currently hidden clinical calculations such as the sensitivity, specificity, PPV and NPV etc. A method is required to display this data without overwhelming the user with large amounts of information.

Another area of future development is to investigate the return of the single well HISTO SPOT assay. If it is possible to perform the assay within a single well, this will greatly increase the turnaround and capacity for the test. However, as evidenced by the performance issues seen during this thesis, there will be a great many challenges to overcome which may ultimately become impractical due to the rapid advancement of competing techniques (e.g. next generation sequencing) or if/when complete genomic data becomes part of patient electronic health records. [216]

7.2.2. The future of the pharmacogenomics field

The literature review in this thesis highlights the need to independently replicate findings to strengthen the validity of a specific drug-allele combination. The examples of HLA-A*31:01 and HLA-*15:02 where there are multiple confirming studies was able to resolve issues of the linkage disequilibrium phenomenon The issue comes with other associations that lack replicated findings – for example nevirapine where reported associations are found to be inconclusive in other studies. This would be greatly aided with the inclusion of trans-ethnic studies as this can also address the issue of haplotypes/co-occurring alleles as evidenced in Chapter 4.

Additionally, the advances in next-generation sequencing (NGS) typing that utilises massive parallelisation so that a single sequence is cloned and millions of strands are simultaneously sequenced will provide four-field level resolution (all coding and non-coding regions of the gene). [89] This will likely prove to be more informative compared to current mainstream usage of sequence-specific oligonucleotide and sequence-specific primer typing methods which only sequence exons 2 and 3 for class I genes and exon 2 for class II genes as this is where most of the known functional variability for HLA genes resides.

With NGS advances, it will be possible to obtain phasing of the whole gene and/or MHC region and will facilitate a broader understanding and further investigations into the relationship between HLA alleles in phased haplotypes and patient ethnicity. The advances will also allow for HLA sequencing to be cost-effective for larger volunteer cohorts – with more available data the field as a whole will benefit.

As well as improvements into the collection of the genetic data, it is clear that simply carrying a particular HLA allele is not enough to trigger an ADR. It will be valuable to continue researching into the other components that contribute to ADRs. A logical place to investigate is to look at performing a multivariate analysis with other clinical factors such as the age, gender, ethnicity, blood protein levels, protein levels from the ADR affected organ, red/white blood cell counts, co-morbidities etc. This will likely be a huge undertaking in terms of collaborative organisation, patient recruitment, ethical approval and patient sample collection and analysis, especially in the recruitment of enough patients and controls that will be sufficient to achieve statistical power. However, such an analysis could reveal other co-risk factors/biomarkers and thereby increase understanding of ADRs.

7.3. Concluding remarks

The era of personalised medicine is rapidly approaching, already patient's genomic information is now being taken into account when it comes to clinical decision making. However, the number of genetic biomarkers that are recommended for clinical screening is still very limited. It is clear that as the field expands, additional HLA alleles will become biomarkers for consideration when making clinical decisions. The bioinformatical and diagnostic resources discussed in this thesis provide a framework to potentially aid in this endeavour.

The HLA-ADR database provides a centralised searchable resource for information pertaining to immune-mediated ADRs to enable researchers to conduct investigations. The ethnicity review of Chapter 4 provides an example of such usage of HLA-ADR and highlights some interesting findings in relation to the reporting of co-occurring alleles in the literature as well as the role that population diversity has in relation to

HLA allele carriage in the context of ADRs. The HISTO SPOT and the associated clinical decision support tool aims to give clinicians a means to arm themselves with accessible expert knowledge and allow for an informed decision be taken. The work undertaken for this thesis encompasses many disciplines, starting with literature/data synthesis which has been used for the development of computational tools that will aid further research. Additionally, lab based experimentation supplemented by a clinical decision support tool will be implemented into clinical practice.

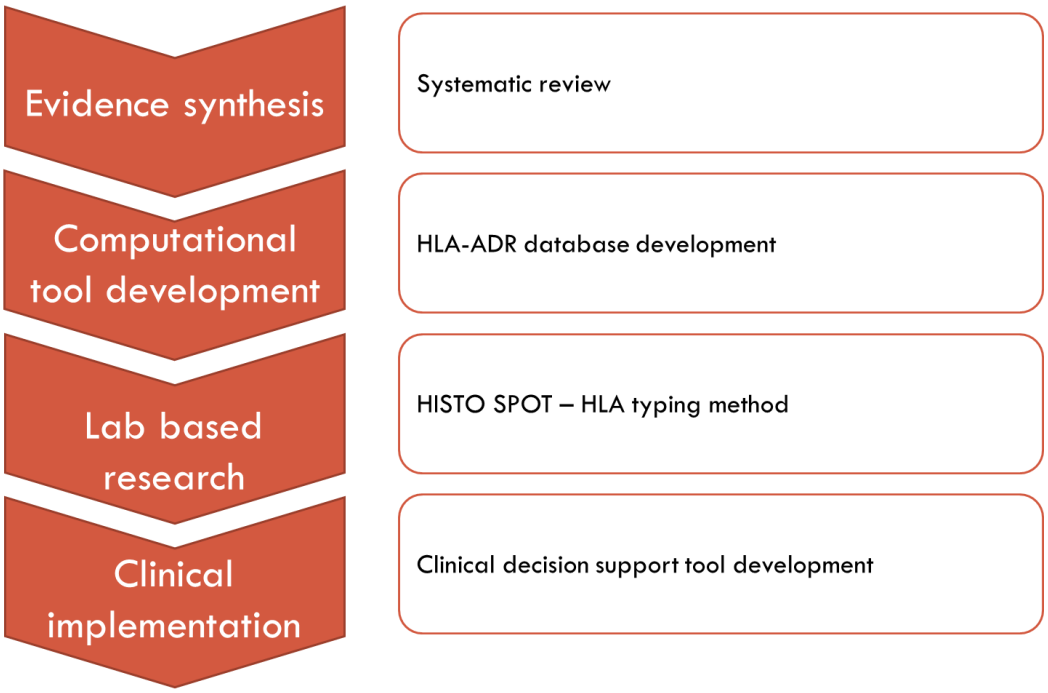


Figure 7.1: Overview of the work undertaken during this PhD thesis

This thesis also emphasises that further research is needed to investigate a multifactorial problem of immune mediated ADRs where only one piece of the puzzle is known (the genetic component). Even then, that single piece is not entirely in focus as the difficulties highlighted in this thesis with poor replication of studies further highlights the need for a better understanding which will hopefully improve drug safety and lead clinical practice into an era of personalised medicine.

Bibliography

[1] Kongkaew C, Hann M, Mandal J, Williams SD, Metcalfe D, Noyce PR, and Ashcroft DM. **2013**. Risk factors for hospital admissions associated with adverse drug events. *Pharmacotherapy*, **33**: 827–837.

[2] Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, Farrar K, Park BK, and Breckenridge AM. **2004**. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18,820 patients. *British Medical Journal (Clinical Research Edition)*, **329**: 15–19.

[3] Lexchin Joel. **2014**. How safe are new drugs? Market withdrawal of drugs approved in Canada between 1990 and 2009. *Open Medicine*, **8**(1): e14.

[4] McNaughton Rhian, Huet Gwenaël, and Shakir Saad. **2014**. An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. *BMJ open*, **4**(1): e004221.

[5] World Health Organization . **1972**. International drug monitoring: the role of national centres, report of a WHO meeting [held in Geneva from 20 to 25 September 1971]. *World Health Organization technical report series*.

[6] Rawlins MD and Thompson JW. *Textbook of adverse drug reactions*, chapter Pathogenesis of adverse drug reactions, page 10. Oxford University Press, Oxford, 1977.

[7] Phillips Elizabeth J. **2016**. Classifying ADRs–does dose matter? *British journal of clinical pharmacology*, **81**: 10–12.

[8] Aronson Jeffrey K and Ferner Robin E. **2016**. The law of mass action and the pharmacological concentration-effect curve: resolving the paradox of apparently non-dose-related adverse drug reactions. *British journal of clinical pharmacology*, **81**: 56–61.

[9] Roujeau Jean-Claude, Kelly Judith P, Naldi Luigi, Rzany Berthold, Stern Robert S, Anderson Theresa, Auquier Ariane, Bastuji-Garin Sylvie, Correia

Osvaldo, Locati Francesco, and others . **1995**. Medication use and the risk of Stevens–Johnson syndrome or toxic epidermal necrolysis. *New England Journal of Medicine*, **333**(24): 1600–1608.

[10] Uetrecht Jack and Naisbitt Dean J. **2013**. Idiosyncratic adverse drug reactions: current concepts. *Pharmacological reviews*, **65**(2): 779–808.

[11] Downey Andrew, Jackson Chris, Harun Nadia, and Cooper Alan. **2012**. Toxic epidermal necrolysis: review of pathogenesis and management. *Journal of the American Academy of Dermatology*, **66**(6): 995–1003.

[12] Katarey Dev and Verma Sumita. **2016**. Drug-induced liver injury. *Clinical Medicine*, **16**(Suppl 6): s104–s109.

[13] Chalasani Naga, Fontana Robert J, Bonkovsky Herbert L, Watkins Paul B, Davern Timothy, Serrano Jose, Yang Hongqiu, Rochon James, (DILIN Drug Induced Liver Injury Network, and others . **2008**. Causes, clinical features, and outcomes from a prospective study of drug-induced liver injury in the United States. *Gastroenterology*, **135**(6): 1924–1934.

[14] Lucena M Isabel, Andrade Raúl J, Kaplowitz Neil, Garcá-Cortes Miren, Fernández M Carmen, Romero-Gomez Manuel, Bruguera Miguel, Hallal Hacibe, Robles-Diaz Mercedes, Rodriguez-González Jose F, and others . **2009**. Phenotypic characterization of idiosyncratic drug-induced liver injury: The influence of age and sex. *Hepatology*, **49**(6): 2001–2009.

[15] Hughes DA, Vilar FJ, Ward CC, Alfirevic A, Park BK, and Pirmohamed M. **2004**. Cost-effectiveness analysis of HLA B*5701 genotyping in preventing abacavir hypersensitivity. *Pharmacogenetics*, **14**: 335–342.

[16] Martin AM, Nolan D, Gaudieri S, Almeida CA, Nolan R, James I, Carvalho F, Phillips E, Christiansen FT, Purcell AW, McCluskey J, and Mallal S. **2004**. Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 4180–4185.

[17] Cheung YK, Cheng SH, Chan EJ, Lo SV, Ng MH, and Kwan P. **2013**. HLA-B alleles associated with severe cutaneous reactions to antiepileptic drugs in Han Chinese. *Epilepsia*, **54**: 1307–1314.

[18] Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, Shirakata Y, Ikezawa Z, Iijima M, Shiohara T, Hashimoto K, Kamatani N, and Nakamura Y. **2011**. Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. *Human molecular genetics*, **20**: 1034–1041.

[19] McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, Sills GJ, Marson T, Jia X, de PI, Bakker, Chinthapalli K, Molokhia M, Johnson MR, O'Connor GD, Chaila E, Alhusaini S, Shianna KV, Radtke RA, Heinzen EL, Walley N, Pandolfo M, Pichler W, Park BK, Depondt C, Sisodiya SM, Goldstein DB, Deloukas P, Delanty N, Cavalleri GL, and Pirmohamed M. **2011**. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *The New England journal of medicine*, **364**: 1134–1143.

[20] Man CB, Kwan P, Baum L, Yu E, Lau KM, Cheng AS, and Ng MH. **2007**. Association between HLA-B*1502 allele and antiepileptic drug-induced cutaneous reactions in Han Chinese. *Epilepsia*, **48**: 1015–1018.

[21] Li LJ, Hu FY, Wu XT, An DM, Yan B, and Zhou D. **2013**. Predictive markers for carbamazepine and lamotrigine-induced maculopapular exanthema in Han Chinese. *Epilepsy research*, **106**: 296–300.

[22] Hung SI, Chung WH, Liu ZS, Chen CH, Hsih MS, Hui RC, Chu CY, and Chen YT. **2010**. Common risk allele in aromatic antiepileptic-drug induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Han Chinese. *Pharmacogenomics*, **11**: 349–356.

[23] Chiu ML, Hu M, Ng MH, Yeung CK, Chan JC, Chang MM, Cheng SH, Li L, and Tomlinson B. **2012**. Association between HLA-B*58:01 allele and severe cutaneous adverse reactions with allopurinol in Han Chinese in Hong Kong. *The British journal of dermatology*, **167**: 44–49.

[24] Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A, Daly MJ, Goldstein DB, John S, Nelson MR, Graham J, Park BK, Dillon JF, Bernal W, Cordell HJ, Pirmohamed M, Aithal GP, and Day CP. **2009**. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nature genetics*, **41**: 816–819.

[25] Hautekeete M L, Horsmans Y, Van Waeyenberge C, Demanet C, Henrion J, Verbist L, Brenard R, Sempoux C, Michielsen P P, Yap P S, Rahier J, and Geubel A P. **1999**. HLA association of amoxicillin-clavulanate–induced hepatitis. *Gastroenterology*, **117**: 1181–1186.

[26] Choi JH, Lee KW, Oh HB, Lee KJ, Suh YJ, Park CS, and Park HS. **2004**. HLA association in aspirin-intolerant asthma: DPB1*0301 as a strong marker in a Korean population. *The Journal of allergy and clinical immunology*, **113**: 562–564.

[27] Kim SH, Choi JH, Lee KW, Kim SH, Shin ES, Oh HB, Suh CH, Nahm DH, and Park HS. **2005**. The human leucocyte antigen-DRB1*1302-DQB1*0609-DPB1*0201 haplotype may be a strong genetic marker for aspirin-induced urticaria. *Clinical and experimental allergy : journal of the British Society for Allergy*, **35**: 339–344.

[28] Schaid DJ, Spraggs CF, McDonnell SK, Parham LR, Cox CJ, Ejlertsen B, Finkelstein DM, Rappold E, Curran J, Cardon LR, and Goss PE. **2014**. Prospective validation of HLA-DRB1*07:01 allele carriage as a predictive risk factor for lapatinib-induced liver injury. *Journal of clinical oncology : official journal of the American Society of*, **32**: 2296–2303.

[29] UK Medicines & Healthcare products Regulatory Agency. **2014**. Guidance on adverse drug reactions. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403098/Guidance_on_adverse_drug_reactions.pdf. Accessed on: 2017-12-18.

[30] Karlin E and Phillips E. **2014**. Genotyping for severe drug hypersensitivity. *Current Allergy And Asthma Reports*, **14**: 418.

[31] Rive CM, Bourke J, and Phillips EJ. **2013**. Testing for drug hypersensitivity syndromes. *The Clinical Biochemist. Reviews / Australian Association Of Clinical Biochemists*, **34**: 15–38.

[32] Kaniwa N and Saito Y. **2013**. Pharmacogenomics of severe cutaneous adverse reactions and drug-induced liver injury. *Journal of Human Genetics*, **58**: 317–326.

[33] Bloch KM, Sills GJ, Pirmohamed M, and Alfirevic A. **2014**. Pharmacogenetics of antiepileptic drug-induced hypersensitivity. *Pharmacogenomics*, **15**: 857–868.

[34] Amstutz U, Ross CJ, Castro-Pastrana LI, Rieder MJ, Shear NH, Hayden MR, and Carleton BC. **2013**. HLA-A 31:01 and HLA-B 15:02 as genetic markers for carbamazepine hypersensitivity in children. *Clinical pharmacology and therapeutics*, **94**: 142–149.

[35] Carr DF, Chaponda M, Jorgensen AL, Castro EC, van JJ, Oosterhout, Khoo SH, Lalloo DG, Heyderman RS, Alfirevic A, and Pirmohamed M. **2013**. Association of human leukocyte antigen alleles and nevirapine hypersensitivity in a Malawian HIV-infected population. *Clinical infectious diseases : an official publication of the Infectious Diseases*, **56**: 1330–1339.

[36] Chantarangsu S, Mushiroda T, Mahasirimongkol S, Kiertiburanakul S, Sungkanuparph S, Manosuthi W, Tantisiriwat W, Charoenyingwattana A, Sura T, Chantratita W, and Nakamura Y. **2009**. HLA-B*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in HIV-infected Thai patients. *Pharmacogenetics and genomics*, **19**: 139–146.

[37] Schneider CH and De Weck AL. **1965**. A New Chemical Aspect of Penicillin Allergy: The Direct Reaction of Penicillin with É›-Amino-Groups. *Nature*, **208**(5005): 57–59.

[38] Ortmann B, Martin S, Von Bonin A, Schiltz E, Hoschützky H, and Weltzien HU. **1992**. Synthetic peptides anchor T cell-specific TNP epitopes to MHC antigens. *The Journal of Immunology*, **148**(5): 1445–1450.

[39] Zanni Martin P, von Greyerz Salome, Schnyder Benno, Brander Karl A, Frutig Karin, Hari Yvonne, Valitutti Salvatore, and Pichler Werner J. **1998**. HLA-restricted, processing-and metabolism-independent pathway of drug recognition by human alpha beta T lymphocytes. *Journal of Clinical Investigation*, **102**(8): 1591.

[40] Pichler Werner J. **2002**. Pharmacological interaction of drugs with antigen-specific immune receptors: the pi concept. *Current opinion in allergy and clinical immunology*, **2**(4): 301–305.

[41] Pichler Werner J, Adam Jacqueline, Watkins Stephen, Wuillemin Natascha, Yun James, and Yerly Daniel. **2015**. Drug hypersensitivity: how drugs stimulate T cells via pharmacological interaction with immune receptors. *International archives of allergy and immunology*, **168**(1): 13–24.

[42] Uetrecht Jack. **2007**. Idiosyncratic drug reactions: current understanding. *Annu. Rev. Pharmacol. Toxicol.*, **47**: 513–539.

[43] Matzinger Polly. **1994**. Tolerance, danger, and the extended family. *Annual review of immunology*, **12**(1): 991–1045.

[44] Pradeu Thomas and Cooper Edwin. **2012**. The danger theory: 20 years later. *Frontiers in Immunology*, **3**: 287.

[45] Yun James, Cai Fenfen, Lee Frederick J., and Pichler Werner J. **2016**. T-cell-mediated drug hypersensitivity: immune mechanisms and their clinical relevance. *Asia Pacific allergy*, **6**: 77–89.

[46] Illing PT, Vivian JP, Dudek NL, Kostenko L, Chen Z, Bharadwaj M, Miles JJ, Kjer-Nielsen L, Gras S, Williamson NA, Burrows SR, Purcell AW, Rossjohn J, and McCluskey J. **2012**. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature*, **486**: 554–558.

[47] Hughes Arlene R., Mosteller Michael, Bansal Aruna T., Davies Kirstie, Haneline Stephen A., Lai Eric H., Nangle Keith, Scott Trevor, Spreen William R., Warren Liling L., and Roses Allen D. **2004**. Association of genetic

variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. *Pharmacogenomics*, **5**: 203–11.

[48] Mallal Simon, Phillips Elizabeth, Carosi Giampiero, Molina Jean-Michel, Workman Cassy, Tomazic Janez, Jagel-Guedes Eva, Rugina Sorin, Kozyrev Oleg, Cid Juan Flores, Hay Phillip, Nolan David, Hughes Sara, Hughes Arlene, Ryan Susanna, Fitch Nicholas, Thorborn Daren, and Benbow Alastair. **2008**. HLA-B*5701 screening for hypersensitivity to abacavir. *The New England journal of medicine*, **358**: 568–79.

[49] US Food and Drug Administration. **2017**. Medication guide - abacavir. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/020977s032,020978s035lbl.pdf#page=28. Accessed on: 2017-08-30.

[50] European Medicines Agency. **2016**. Trizivir / abacavir. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000338/human_med_001109.jsp&mid=WC0b01ac058001d124. Accessed on: 2017-10-12.

[51] US Food and Drug Administration. **2015**. Carbamazepine label. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2015/016608s097,018281s045,018927s038,020234s026lbl.pdf. Accessed on: 2017-08-19.

[52] European Medicines Agency. **2012**. PhVWP monthly report on safety concerns, guidelines and general matters. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/07/WC500130391.pdf. Accessed on: 2017-05-17.

[53] US Food and Drug Administration. **2016**. Phenytoin label. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/008762s057s058lbl.pdf. Accessed on: 2017-12-12.

[54] Clare Kathleen E., Miller Michael H., and Dillon John F. **2017**. Genetic Factors Influencing Drug-Induced Liver Injury: Do They Have a Role in Prevention and Diagnosis? *Current hepatology reports*, **16**: 258–264.

[55] von Ahsen Nicolas and Chan Andrew. **2012**. Therapeutic monitoring of immunotherapies in autoimmune diseases. *Current pharmaceutical design*, **18**: 4550–4555.

[56] Khalil Danny N, Smith Eric L, Brentjens Renier J, and Wolchok Jedd D. **May 2016**. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nature reviews. Clinical oncology*, **13**: 273–290.

[57] Chowell Diego, Morris Luc G T, Grigg Claud M, Weber Jeffrey K, Samstein Robert M, Makarov Vladimir, Kuo Fengshen, Kendall Sviatoslav M, Requena David, Riaz Nadeem, Greenbaum Benjamin, Carroll James, Garon Edward, Hyman David M, Zehir Ahmet, Solit David, Berger Michael, Zhou Ruhong, Rizvi Naiyer A, and Chan Timothy A. **February 2018**. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science (New York, N.Y.)*, **359**: 582–587.

[58] Lyczak J. *Immunology, Infection and Immunity*, chapter 11: The major histocompatibility complex, pages 261–282. ASM Press, 2004.

[59] Westover Jonna B., Sweeten Thayne L., Benson Michael, Bray-Ward Patricia, and Torres Anthony R. Immune Dysfunction in Autism Spectrum Disorder. In Eapen Valsamma, editor, *Autism - A Neurodevelopmental Journey from Genes to Behaviour*, chapter 18. InTech, Rijeka, 2011.

[60] Halenius Anne, Gerke Carolin, and Hengel Hartmut. **2015**. Classical and non-classical MHC I molecule manipulation by human cytomegalovirus: so many targetsâ€"but how many arrows in the quiver? *Cellular & molecular immunology*, **12**(2): 139–153.

[61] Kelly Adrian and Trowsdale John. **2017**. Introduction: MHC/KIR and governance of specificity. *Immunogenetics*, **69**(8-9): 481–488.

[62] EBI - IMGT/HLA. **2017**. HLA Statistics. Available from: https://www.ebi.ac.uk/ipd/imgt/hla/stats.html. Accessed on: 2017-12-20.

[63] National Center for Biotechnology Information. **2017**. NCBI - Gene - HLA Class I - HLA B. Available from: https://www.ncbi.nlm.nih.gov/gene/3106. Accessed on: 2017-12-20.

[64] National Center for Biotechnology Information. **2017**. NCBI - Gene - HLA Class I - HLA*DRB1. Available from: https://www.ncbi.nlm.nih.gov/gene/3123. Accessed on: 2017-12-20.

[65] The-Crankshaft Publishing. **2012**. Adaptive Immunity: Histocompatibility Antigens and Immune Response Genes Part 1. Available from: http://what-when-how.com/acp-medicine/adaptive-immunity-histocompatibility-antigens-and-immune-response-genes-part-1/. Accessed on: 2017-12-20.

[66] Meydan Cem, Otu Hasan H, and Sezerman Osman Ugur. **2013**. Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. *BMC bioinformatics*, **14**(2): S13.

[67] Brusic Vladimir, Bajic Vladimir B., and Petrovsky Nikolai. **2004**. Computational methods for prediction of T-cell epitopes–a framework for modelling, testing, and applications. *Methods (San Diego, Calif.)*, **34**: 436–43.

[68] Jones E. Yvonne, Fugger Lars, Strominger Jack L., and Siebold Christian. **2006**. MHC class II proteins and disease: a structural perspective. *Nature reviews. Immunology*, **6**: 271–82.

[69] Rapin Nicolas, Hoof Ilka, Lund Ole, and Nielsen Morten. **2008**. MHC motif viewer. *Immunogenetics*, **60**: 759–65.

[70] Speir J. A., Stevens J., Joly E., Butcher G. W., and Wilson I. A. **2001**. Two different, highly exposed, bulged structures for an unusually long peptide bound to rat MHC class I RT1-Aa. *Immunity*, **14**: 81–92.

[71] Trost Brett, Bickis Mik, and Kusalik Anthony. **2007**. Strength in numbers: achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools. *Immunome research*, **3**: 5.

[72] Wang Peng, Sidney John, Dow Courtney, Mothe Bianca, Sette Alessandro, and Peters Bjoern. **2008**. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS computational biology*, **4**: e1000048.

[73] Yassai Maryam, Afsari Amin, Garlie Jason, and Gorski Jack. **2002**. C-terminal anchoring of a peptide to class II MHC via the P10 residue is compatible with a peptide bulge. *Journal of immunology (Baltimore, Md. : 1950)*, **168**: 1281–5.

[74] Chaves Francisco A, Richards Katherine A, Torelli Andrew, Wedekind Joseph, and Sant Andrea J. **2006**. Peptide-binding motifs for the I-Ad MHC class II molecule: alternate pH-dependent binding behavior. *Biochemistry*, **45**(20): 6426–6433.

[75] Erlich H. **2012**. HLA DNA typing: past, present, and future. *HLA*, **80**(1): 1–11.

[76] Marsh SG. *The HLA Complex in Biology and Medicine A Resource Book*, chapter HLA Nomenclature, pages 79–85. Jaypee Brothers Medical Publishers, 1 edition, 2010.

[77] HLA Informatics Group. **2017**. Nomenclature for Factors of the HLA System. Available from: http://hla.alleles.org/nomenclature/naming.html. Accessed on: 2017-12-21.

[78] Robinson James, Halliwell Jason A, Hayhurst James D, Flicek Paul, Parham Peter, and Marsh Steven G E. **2015**. The IPD and IMGT/HLA database: allele variant databases. *Nucleic acids research*, **43**: D423–D431.

[79] Buckingham L. *Molecular Analysis of the MHC*, chapter DNA-Bassed Tissue Typing, pages 427–430. FA Davis Company, Philadelphia, 2 edition, 2012.

[80] Dunckley Heather. **2012**. HLA typing by SSO and SSP methods. *Methods in molecular biology (Clifton, N.J.)*, **882**: 9–25.

[81] de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, and Rioux JD. **2006**. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature Genetics*, **38**: 1166–1172.

[82] Voorter Christina E M, Palusci Fausto, and Tilanus Marcel G J. **2014**. Sequence-based typing of HLA: an improved group-specific full-length gene sequencing approach. *Methods in molecular biology (Clifton, N.J.)*, **1109**: 101–114.

[83] Mytilineos J, Lempert M, Scherer S, Schwarz V, and Opelz G. **1998**. Comparison of serological and DNA PCR-SSP typing results for HLA-A and HLA-B in 421 Black individuals: a Collaborative Transplant Study report. *Human immunology*, **59**: 512–517.

[84] BozÃ³n M V, Delgado J C, Selvakumar A, Clavijo O P, Salazar M, Ohashi M, Alosco S M, Russell J, Yu N, Dupont B, and Yunis E J. **1997**. Error rate for HLA-B antigen assignment by serology: implications for proficiency testing and utilization of DNA-based typing methods. *Tissue antigens*, **50**: 387–394.

[85] Howell W M, Carter V, and Clark B. **2010**. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *Journal of clinical pathology*, **63**: 387–390.

[86] EBI - IMGT/HLA. **2017**. Ambiguous Allele Combinations. Available from: https://www.ebi.ac.uk/ipd/imgt/hla/ambig.html. Accessed on: 2017-11-20.

[87] Zheng X, Shen J, Cox C, Wakefield J C, Ehm M G, Nelson M R, and Weir B S. **2014**. HIBAG–HLA genotype imputation with attribute bagging. *The pharmacogenomics journal*, **14**: 192–200.

[88] Nunes Kelly, Zheng Xiuwen, Torres Margareth, Moraes Maria Elisa, Piovezan Bruno Z, Pontes Gerlandia N, Kimura Lilian, Carnavalli Juliana E P, Mingroni Netto Regina C, and Meyer Diogo. **2016**. HLA imputation in an admixed population: An assessment of the 1000 Genomes data as a training set. *Human immunology*, **77**: 307–312.

[89] Gabriel C, FÃ¼rst D, FaÃ© I, Wenda S, Zollikofer C, Mytilineos J, and Fischer G F. **2014**. HLA typing by next-generation sequencing - getting closer to reality. *Tissue antigens*, **83**: 65–75.

[90] W3 Schools. **2017**. Database Normalization. Available from: https://www.w3schools.in/dbms/database-normalization/. Accessed on: 2017-10-18.

[91] Hillyer Mike. **2003**. An introduction to database normalization. *MySQL AB*, **2003**.

[92] Microsoft. **2017**. Description of the database normalization basics. Available from: https://support.microsoft.com/en-gb/help/283878/description-of-the-database-normalization-basics. Accessed on: 2017-12-06.

[93] Codd Edgar F. **1970**. A relational model of data for large shared data banks. *Communications of the ACM*, **13**(6): 377–387.

[94] Codd Edgar F. **1972**. Further normalization of the data base relational model. *Data base systems*, pages 33–64.

[95] DB-Engines. **2017**. DB-Engines Ranking. Available from: https://db-engines.com/en/ranking. Accessed on: 2017-07-27.

[96] W3 Schools. **2017**. Introduction to SQL. Available from: https://www.w3schools.com/sql/sql_intro.asp. Accessed on: 2017-12-18.

[97] W^{3}Techs. **2017 July**. Usage of server-side programming languages for websites (survey from July 2017). Available from: https://w3techs.com/technologies/overview/programming_language/all. Accessed on: 2017-07-31.

[98] Yip VL, Marson AG, Jorgensen AL, Pirmohamed M, and Alfirevic A. **2012**. HLA genotype and carbamazepine-induced cutaneous adverse drug reactions: a systematic review. *Clinical Pharmacology And Therapeutics*, **92**: 757–765.

[99] Yu W, Gwinn M, Clyne M, Yesupriya A, and Khoury MJ. **2008**. A navigator for human genome epidemiology. *Nature Genetics*, **40**: 124–125.

[100] National Center for Biotechnology Information. **No date**. MEDLINE/PubMed. Available from: http://www.ncbi.nlm.nih.gov/pubmed/. Accessed on: 03/12/2015.

[101] Elsevier. **2015**. Embase. Available from: https://www.elsevier.com/solutions/embase-biomedical-research. Accessed on: 03/12/2015.

[102] Daly AK. **2012**. Using genome-wide association studies to identify genes important in serious adverse drug reactions. *Annual Review Of Pharmacology And Toxicology*, **52**: 21–35.

[103] de Graaff LC, van Schaik RH, and van Gelder T. **2013**. A clinical approach to pharmacogenetics. *The Netherlands Journal Of Medicine*, **71**: 145–152.

[104] Python Software Foundation. **2014**. Python Language Reference, version 3.4. Available from: http://www.python.org. Accessed on: 2014-07-29.

[105] Almeida CA, Martin AM, Nolan D, Lucas A, Cameron PU, James I, Phillips E, and Mallal S. **2008**. Cytokine profiling in abacavir hypersensitivity patients. *Antiviral therapy*, **13**: 281–288.

[106] Berka N, Gill JM, Liacini A, O'Bryan T, and Khan FM. **2012**. Human leukocyte antigen (HLA) and pharmacogenetics: screening for HLA-B*57:01 among human immunodeficiency virus-positive patients from southern Alberta. *Human immunology*, **73**: 164–167.

[107] Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, Sayer D, Castley A, Mamotte C, Maxwell D, James I, and Christiansen FT. **2002**. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and

hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*, **359**: 727–732.

[108] Rauch A, Nolan D, Thurnheer C, Fux CA, Cavassini M, Chave JP, Opravil M, Phillips E, Mallal S, and Furrer H. **2008**. Refining abacavir hypersensitivity diagnoses using a structured clinical assessment and genetic testing in the Swiss HIV Cohort Study. *Antiviral therapy*, **13**: 1019–1028.

[109] Rodriguez-Novoa S, Garcia-Gasco P, Blanco F, Gonzalez-Pardo G, Castellares C, Moreno V, Jimenez-Nacher I, Gonzalez-Lahoz J, and Soriano V. **2007**. Value of the HLA-B*5701 allele to predict abacavir hypersensitivity in Spaniards. *AIDS research and human retroviruses*, **23**: 1374–1376.

[110] Stekler J, Maenza J, Stevens C, Holte S, Malhotra U, McElrath MJ, Corey L, and Collier AC. **2006**. Abacavir hypersensitivity reaction in primary HIV infection. *AIDS (London, England)*, **20**: 1269–1274.

[111] Sanchez-Giron Francisco, Villegas-Torres Beatriz, Jaramillo-Villafuerte Karla, Silva-Zolezzi Irma, Fernandez-Lopez Juan Carlos, Jimenez-Sanchez Gerardo, and Carnevale Alessandra. **2011**. Association of the genetic marker for abacavir hypersensitivity HLA-B*5701 with HCP5 rs2395029 in Mexican Mestizos. *Pharmacogenomics*, **12**(6): 809–814.

[112] Melis R, Lewis T, Millson A, Lyon E, McMillin GA, Slev PR, and Swensen J. **2012**. Copy number variation and incomplete linkage disequilibrium interfere with the HCP5 genotyping assay for abacavir hypersensitivity. *Genetic Testing And Molecular Biomarkers*, **16**: 1111–1114.

[113] Badulli C, Sestini R, Sbarsi I, Baroncelli M, Pizzochero C, Martinetti M, and Porfirio B. **2012**. Tag SNPs of the ancestral haplotype 57.1 do not substitute HLA-B*57:01 typing for eligibility to abacavir treatment in the Italian population. *Pharmacogenomics*, **13**: 247–249.

[114] Cornejo Castro EM. *Characterisation of molecular mechanisms involved in nevirapine- induced hypersensitivity*. PhD thesis, University of Liverpool, 2014.

[115] Lee HY, Lee JW, Lee KW, Park MH, and Park HS. **2009**. The HLA allele marker for differentiating ASA hypersensitivity phenotypes. *Allergy*, **64**: 1385–1387.

[116] Dettling M, Cascorbi I, Roots I, and Mueller-Oerlinghausen B. **2001**. Genetic determinants of clozapine-induced agranulocytosis: recent results of HLA subtyping in a non-jewish caucasian sample. *Archives of general psychiatry*, **58**: 93–94.

[117] Kashiwagi M, Aihara M, Takahashi Y, Yamazaki E, Yamane Y, Song Y, Muramatsu M, and Ikezawa Z. **2008**. Human leukocyte antigen genotypes in carbamazepine-induced severe cutaneous adverse drug response in Japanese patients. *The Journal of dermatology*, **35**: 683–685.

[118] Kindmark A, Jawaid A, Harbron CG, Barratt BJ, Bengtsson OF, Andersson TB, Carlsson S, Cederbrant KE, Gibson NJ, Armstrong M, Lagerstrom-Fermer ME, Dellsen A, Brown EM, Thornton M, Dukes C, Jenkins SC, Firth MA, Harrod GO, Pinel TH, Billing-Clason SM, Cardon LR, and March RE. **2008**. Genome-wide pharmacogenetic investigation of a hepatic adverse event without linical signs of immunopathology suggests an underlying immune pathogenesis. *The pharmacogenomics journal*, **8**: 186–195.

[119] Mammen AL, Gaudet D, Brisson D, Christopher-Stine L, Lloyd TE, Leffell MS, and Zachary AA. **2012**. Increased frequency of DRB1*11:01 in anti-hydroxymethylglutaryl-coenzyme A reductase-associated autoimmune myopathy. *Arthritis care & research*, **64**: 1233–1237.

[120] Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, Teles e Silva AL, Ghattaoraya GS, Alfirevic A, Jones AR, and Middleton D. **2015**. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Research*, **43**: D784–788.

[121] Gonzalez-Galarza Faviel F, Christmas Stephen, Middleton Derek, and Jones Andrew R. **2011**. Allele frequency net: a database and online repository for

immune gene frequencies in worldwide populations. *Nucleic acids research*, **39**: D913–D919.

[122] Fan WL, Shiao MS, Hui RCY, Su Shih-Chi, Wang CW, Chang YC, and Chung WH. **2017**. HLA Association with Drug-Induced Adverse Reactions. *Journal of Immunology Research*, **2017**: 1–10.

[123] Maggio Nicola, Firer Maria, Zaid Huda, Bederovsky Yana, Aboukaoud Mohammed, Gandelman-Marton Revital, Noyman Iris, Ekstein Dana, Blatt Ilan, Marom Eli, and others . **2017**. Causative Drugs of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in Israel. *The Journal of Clinical Pharmacology*.

[124] Ghattaoraya G S, Middleton D, Santos E J M, Dickson R, Jones A R, and Alfirevic A. **2017**. Human leucocyte antigen-adverse drug reaction associations: from a perspective of ethnicity. *International journal of immunogenetics*, **44**: 7–26.

[125] Martin M A, Klein T E, Dong B J, Pirmohamed M, Haas D W, Kroetz D L, and Consortium Clinical Pharmacogenetics Implementation. **April 2012**. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. *Clinical pharmacology and therapeutics*, **91**: 734–738.

[126] Hershfield M S, Callaghan J T, Tassaneeyakul W, Mushiroda T, Thorn C F, Klein T E, and Lee M T M. **February 2013**. Clinical Pharmacogenetics Implementation Consortium guidelines for human leukocyte antigen-B genotype and allopurinol dosing. *Clinical pharmacology and therapeutics*, **93**: 153–158.

[127] Leckband S G, Kelsoe J R, Dunnenberger H M, George A L, Tran E, Berger R, MÃ¼ller D J, Whirl-Carrillo M, Caudle K E, Pirmohamed M, and Consortium Clinical Pharmacogenetics Implementation. **September 2013**. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. *Clinical pharmacology and therapeutics*, **94**: 324–328.

[128] Caudle K E, Rettie A E, Whirl-Carrillo M, Smith L H, Mintzer S, Lee M T M, Klein T E, Callaghan J T, and Consortium Clinical Pharmacogenetics Implementation. **November 2014**. Clinical pharmacogenetics implementation consortium guidelines for CYP2C9 and HLA-B genotypes and phenytoin dosing. *Clinical pharmacology and therapeutics*, **96**: 542–548.

[129] Dos Santos Eduardo Jose Melos, McCabe Antony, Gonzalez-Galarza Faviel F, Jones Andrew R, and Middleton Derek. **2016**. Allele Frequencies Net Database: Improvements for storage of individual genotypes and analysis of existing data. *Human immunology*, **77**: 238–248.

[130] Cristallo AF, Schroeder J, Citterio A, Santori G, Ferrioli GM, Rossi U, Bertani G, Cassano S, Gottardi P, Ceschini N, Barocci F, Ribizzi G, Cutrupi V, Cairoli R, Rapisarda V, Pastorello EA, and Barocci S. **2011**. A study of HLA class I and class II 4-digit allele level in Stevens-Johnson syndrome and toxic epidermal necrolysis. *International journal of immunogenetics*, **38**: 303–309.

[131] Goncalo M, Coutinho I, Teixeira V, Gameiro AR, Brites MM, Nunes R, and Martinho A. **2013**. HLA-B*58:01 is a risk factor for allopurinol-induced DRESS and Stevens-Johnson syndrome/toxic epidermal necrolysis in a Portuguese population. *The British journal of dermatology*, **169**: 660–665.

[132] Cao ZH, Wei ZY, Zhu QY, Zhang JY, Yang L, Qin SY, Shao LY, Zhang YT, Xuan JK, Li QL, Xu JH, Xu F, Ma L, Huang HY, Xing QH, and Luo XQ. **2012**. HLA-B*58:01 allele is associated with augmented risk for both mild and severe cutaneous adverse reactions induced by allopurinol in Han Chinese. *Pharmacogenomics*, **13**: 1193–1201.

[133] Hung SI, Chung WH, Liou LB, Chu CC, Lin M, Huang HP, Lin YL, Lan JL, Yang LC, Hong HS, Chen MJ, Lai PC, Wu MS, Chu CY, Wang KH, Chen CH, Fann CS, Wu JY, and Chen YT. **2005**. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proceedings of the National Academy of Sciences of the United States of America*, **102**: 4134–4139.

[134] Kaniwa N, Saito Y, Aihara M, Matsunaga K, Tohkin M, Kurose K, Sawada J, Furuya H, Takahashi Y, Muramatsu M, Kinoshita S, Abe M, Ikeda H, Kashiwagi M, Song Y, Ueta M, Sotozono C, Ikezawa Z, and Hasegawa R. **2008**. HLA-B locus in Japanese patients with anti-epileptics and allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis. *Pharmacogenomics*, **9**: 1617–1622.

[135] Niihara H, Kaneko S, Ito T, Sugamori T, Takahashi N, Kohno K, and Morita E. **2013**. HLA-B*58:01 strongly associates with allopurinol-induced adverse drug reactions in a Japanese sample population. *Journal of dermatological science*, **71**: 150–152.

[136] Kang HR, Jee YK, Kim YS, Lee CH, Jung JW, Kim SH, Park HW, Chang YS, Jang IJ, Cho SH, Min KU, Kim SH, and Lee KW. **2011**. Positive and negative associations of HLA class I alleles with allopurinol-induced SCARs in Koreans. *Pharmacogenetics and genomics*, **21**: 303–307.

[137] Kim SH, Lee SK, Kim SH, Park HW, Chang YS, Lee KW, and Jee YK. **2013**. Antituberculosis drug-induced hypersensitivity syndrome and its association with human leukocyte antigen. *Tuberculosis (Edinburgh, Scotland)*, **93**: 270–274.

[138] Sharma Surendra K, Balamurugan Arumugam, Saha Pradip Kumar, Pandey Ravindra M, and Mehra Narinder K. **2002**. Evaluation of clinical and immunogenetic risk factors for the development of hepatotoxicity during antituberculosis treatment. *American journal of respiratory and critical care medicine*, **166**: 916–919.

[139] Zhang FR, Liu H, Irwanto A, Fu XA, Li Y, Yu GQ, Yu YX, Chen MF, Low HQ, Li JH, Bao FF, Foo JN, Bei JX, Jia XM, Liu J, Liany H, Wang N, Niu GY, Wang ZZ, Shi BQ, Tian HQ, Liu HX, Ma SS, Zhou Y, You JB, Yang Q, Wang C, Chu TS, Liu DC, Yu XL, Sun YH, Ning Y, Wei ZH, Chen SL, Chen XC, Zhang ZX, Liu YX, Pulit SL, Wu WB, Zheng ZY, Yang RD, Long H, Liu ZS, Wang JQ, Li M, Zhang LH, Wang H, Wang LM, Xiao P, Li JL, Huang ZM, Huang JX, Li Z, Liu J, Xiong L, Yang J, Wang XD, Yu DB, Lu XM, Zhou GZ, Yan LB, Shen JP, Zhang GC, Zeng YX, de Bakker PI, Chen SM,

and Liu JJ. **2013**. HLA-B*13:01 and the dapsone hypersensitivity syndrome. *The New England Journal Of Medicine*, **369**: 1620–1628.

[140] Lucena M Isabel, Molokhia Mariam, Shen Yufeng, Urban Thomas J, Aithal Guruprasad P, Andrade RaÃºl J, Day Christopher P, Ruiz-Cabello Francisco, Donaldson Peter T, Stephens Camilla, Pirmohamed Munir, Romero-Gomez Manuel, Navarro Jose Maria, Fontana Robert J, Miller Michael, Groome Max, Bondon-Guitton Emmanuelle, Conforti Anita, Stricker Bruno H C, Carvajal Alfonso, Ibanez Luisa, Yue Qun-Ying, Eichelbaum Michel, Floratos Aris, Pe'er Itsik, Daly Mark J, Goldstein David B, Dillon John F, Nelson Matthew R, Watkins Paul B, Daly Ann K, Registry Spanish DILI, EUDRAGENE , DILIN , DILIGEN , and SAEC International. **2011**. Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. *Gastroenterology*, **141**: 338–347.

[141] Stephens C, Lopez-Nevot MA, Ruiz-Cabello F, Ulzurrun E, Soriano G, Romero-Gomez M, Moreno-Casares A, Lucena MI, and Andrade RJ. **2013**. HLA alleles influence the clinical signature of amoxicillin-clavulanate hepatotoxicity. *PloS one*, **8**.

[142] O'Donohue J, Oien KA, Donaldson P, Underhill J, Clare M, MacSween RN, and Mills PR. **2000**. Co-amoxiclav jaundice: clinical and histological features and HLA class II association. *Gut*, **47**: 717–720.

[143] Kongpan Thachanan, Mahasirimongkol Surakameth, Konyoung Parinya, Kanjanawart Sirimas, Chumworathayi Pansu, Wichukchinda Nuanjun, Kidkeukarun Runglak, Preechakul Suphanlinee, Khunarkornsiri Usanee, Bamrungram Warawut, Supharatwattanakun Butsaban, Mootsikapun Piroon, Kwangsukstid Supanida, Denjanta Sukanda, Vannaprasaht Suda, Rungapiromnan Watcharee, Suwankesawong Wimon, Tassaneeyakul Wongwiwat, and Tassaneeyakul Wichittra. **2015**. Candidate HLA genes for prediction of co-trimoxazole-induced severe cutaneous reactions. *Pharmacogenetics and genomics*, **25**: 402–411.

[144] Hsiao YH, Hui RC, Wu T, Chang WC, Hsih MS, Yang CH, Ho HC, Chang YG, Chen MJ, Lin JY, Chen DP, Chang PY, Wu TL, Hung SI, and Chung

WH. **2014**. Genotype-phenotype association between HLA and carbamazepine-induced hypersensitivity reactions: strength and clinical correlations. *Journal of dermatological science*, **73**: 101–109.

[145] Shi YW, Min FL, Qin B, Zou X, Liu XR, Gao MM, Wang Q, Zhou JQ, and Liao WP. **2012**. Association between HLA and Stevens-Johnson syndrome induced by carbamazepine in Southern Han Chinese: genetic markers besides B*1502? *Basic & clinical pharmacology & toxicology*, **111**: 58–64.

[146] Kim SH, Lee KW, Song WJ, Kim SH, Jee YK, Lee SM, Kang HR, Park HW, Cho SH, Park SH, Min KU, and Chang YS. **2011**. Carbamazepine-induced severe cutaneous adverse reactions and HLA genotypes in Koreans. *Epilepsy research*, **97**: 190–197.

[147] Alfirevic A, Jorgensen AL, Williamson PR, Chadwick DW, Park BK, and Pirmohamed M. **2006**. HLA-B locus in Caucasian patients with carbamazepine hypersensitivity. *Pharmacogenomics*, **7**: 813–818.

[148] Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, Halevy S, Naldi L, Bouwes-Bavinck JN, Sidoroff A, de C, Toma, Schumacher M, Roujeau JC, Hovnanian A, and Mockenhaupt M. **2008**. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenetics and genomics*, **18**: 99–107.

[149] Genin E, Chen DP, Hung SI, Sekula P, Schumacher M, Chang PY, Tsai SH, Wu TL, Bellon T, Tamouza R, Fortier C, Toubert A, Charron D, Hovnanian A, Wolkenstein P, Chung WH, Mockenhaupt M, and Roujeau JC. **2014**. HLA-A*31:01 and different types of carbamazepine-induced severe cutaneous adverse reactions: an international study and meta-analysis. *The pharmacogenomics journal*, **14**: 281–288.

[150] Hung SI, Chung WH, Jee SH, Chen WC, Chang YT, Lee WR, Hu SL, Wu MT, Chen GS, Wong TW, Hsiao PF, Chen WH, Shih HY, Fang WH, Wei CY, Lou YH, Huang YL, Lin JJ, and Chen YT. **2006**. Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenetics and genomics*, **16**: 297–306.

[151] He N, Min FL, Shi YW, Guo J, Liu XR, Li BM, Zhou JH, Ou YM, Liao JX, and Liao WP. **2012**. Cutaneous reactions induced by oxcarbazepine in Southern Han Chinese: incidence, features, risk factors and relation to HLA-B alleles. *Seizure : the journal of the British Epilepsy Association*, **21**: 614–618.

[152] Hu FY, Wu XT, An DM, Yan B, Stefan H, and Zhou D. **2011**. Pilot association study of oxcarbazepine-induced mild cutaneous adverse reactions with HLA-B*1502 allele in Chinese Han population. *Seizure : the journal of the British Epilepsy Association*, **20**: 160–162.

[153] Lv YD, Min FL, Liao WP, He N, Zeng T, Ma DH, and Shi YW. **2013**. The association between oxcarbazepine-induced maculopapular eruption and HLA-B alleles in a northern Han Chinese population. *BMC neurology*, **13**: 75.

[154] Manuyakorn W, Siripool K, Kamchaisatian W, Pakakasama S, Visudtibhan A, Vilaiyuk S, Rujirawat T, and Benjaponpitak S. **2013**. Phenobarbital-induced severe cutaneous adverse drug reactions are associated with YP2C19*2 in Thai children. *Pediatric allergy and immunology : official publication of the European Society*, **24**: 299–303.

[155] Chang C-C, Ng C-C, Too C-L, Choon S-E, Lee C-K, Chung W-H, Hussein S H, Lim K-S, and Murad S. **2017**. Association of HLA-B*15:13 and HLA-B*15:02 with phenytoin-induced severe cutaneous adverse reactions in a Malay population. *The pharmacogenomics journal*, **17**: 170–173.

[156] Locharernkul C, Loplumlert J, Limotai C, Korkij W, Desudchit T, Tongkobpetch S, Kangwanshiratada O, Hirankarn N, Suphapeetiporn K, and Shotelersuk V. **2008**. Carbamazepine and phenytoin induced Stevens-Johnson syndrome is associated with HLA-B*1502 allele in Thai population. *Epilepsia*, **49**: 2087–2091.

[157] An DM, Wu XT, Hu FY, Yan B, Stefan H, and Zhou D. **2011**. Association study of lamotrigine-induced cutaneous adverse reactions and HLA-B*1502 in a Han Chinese population. *Epilepsy Research*, **92**: 226–230.

[158] Shi YW, Min FL, Liu XR, Zan LX, Gao MM, Yu MJ, and Liao WP. **2011**. Hla-B alleles and lamotrigine-induced cutaneous adverse drug reactions in the Han Chinese population. *Basic &Amp; Clinical Pharmacology &Amp; Toxicology*, **109**: 42–46.

[159] Phillips E, Bartlett JA, Sanne I, Lederman MM, Hinkle J, Rousseau F, Dunn D, Pavlos R, James I, Mallal SA, and Haas DW. **2013**. Associations between HLA-DRB1*0102, HLA-B*5801, and hepatotoxicity during initiation of nevirapine-containing regimens in South Africa. *Journal of acquired immune deficiency syndromes (1999)*, **62**.

[160] Furukawa H, Oka S, Shimada K, Sugii S, Hashimoto A, Komiya A, Fukui N, Miyashita T, Migita K, Suda A, Nagaoka S, Tsuchiya N, and Tohma S. **2014**. HLA-DRB1*08:02 Is Associated with Bucillamine-Induced Proteinuria in Japanese Rheumatoid Arthritis Patients. *Biomarker insights*, **9**: 23–28.

[161] Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, Klickstein L, Wright TM, Meyer J, and Paulding CA. **2010**. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nature genetics*, **42**: 711–714.

[162] Yang Fanping, Gu Bo, Zhang Lirong, Xuan Jiekun, Luo Heng, Zhou Peng, Zhu Qinyuan, Yan Sijia, Chen Sheng-an, Cao Zhihao, Xu Jinhua, Xing Qinghe, and Luo Xiaoqun. **2014**. HLA-B*13:01 is associated with salazosulfapyridine-induced drug rash with eosinophilia and systemic symptoms in Chinese Han population. *Pharmacogenomics*, **15**: 1461–1469.

[163] Dettling M, Schaub RT, Mueller-Oerlinghausen B, Roots I, and Cascorbi I. **2001**. Further evidence of human leukocyte antigen-encoded susceptibility to clozapine-induced agranulocytosis independent of ancestry. *Pharmacogenetics*, **11**: 135–141.

[164] Goldstein JI, Jarskog LF, Hilliard C, Alfirevic A, Duncan L, Fourches D, Huang H, Lek M, Neale BM, Ripke S, Shianna K, Szatkiewicz JP, Tropsha A, van den Oord EJ, Cascorbi I, Dettling M, Gazit E, Goff DC, Holden AL, Kelly DL, Malhotra AK, Nielsen J, Pirmohamed M, Rujescu D, Werge T, Levy DL, Josiassen RC, Kennedy JL, Lieberman JA, Daly MJ, and Sullivan PF. **2014**.

Clozapine-induced agranulocytosis is associated with rare HLA-DQB1 and HLA-B alleles. *Nature Communications*, **5**: 4757.

[165] Hirata K, Takagi H, Yamamoto M, Matsumoto T, Nishiya T, Mori K, Shimizu S, Masumoto H, and Okutani Y. **2008**. Ticlopidine-induced hepatotoxicity is associated with specific human leukocyte antigen genomic subtypes in Japanese patients: a preliminary case-control study. *The pharmacogenomics journal*, **8**: 29–33.

[166] EMA . Press release AstraZeneca withdraws its application for Ximelagatran 36-mg film-coated tablets. press release, 2006.

[167] Spraggs CF, Budde LR, Briley LP, Bing N, Cox CJ, King KS, Whittaker JC, Mooser VE, Preston AJ, Stein SH, and Cardon LR. **2011**. HLA-DQA1*02:01 is a major risk factor for lapatinib-induced hepatotoxicity in women with advanced breast cancer. *Journal of clinical oncology : official journal of the American Society of*, **29**: 667–673.

[168] Parham L R, Briley L P, Li L, Shen J, Newcombe P J, King K S, Slater A J, Dilthey A, Iqbal Z, McVean G, Cox C J, Nelson M R, and Spraggs C F. **2016**. Comprehensive genome-wide evaluation of lapatinib-induced liver injury yields a single genetic signal centered on known risk allele HLA-DRB1*07:01. *The pharmacogenomics journal*, **16**: 180–185.

[169] Yang F, Xuan J, Chen J, Zhong H, Luo H, Zhou P, Sun X, He L, Chen S, Cao Z, Luo X, and Xing Q. **2016**. HLA-B*59:01: a marker for Stevens-Johnson syndrome/toxic epidermal necrolysis caused by methazolamide in Han Chinese. *The pharmacogenomics journal*, **16**: 83–87.

[170] Alfirevic Ana. **2017**. Patient ethnicity and the risk of immune-mediated adverse drug reactions. *Pharmacogenomics*, **18**: 1375–1378.

[171] Arif H, Buchsbaum R, Weintraub D, Koyfman S, Salas-Humara C, Bazil CW, Resor SR, and Hirsch LJ. **2007**. Comparison and predictors of rash associated with 15 antiepileptic drugs. *Neurology*, **68**: 1701–1709.

[172] Wang Q, Zhou JQ, Zhou LM, Chen ZY, Fang ZY, Chen SD, Yang LB, Cai XD, Dai QL, Hong H, and Wang HX. **2011**. Association between HLA-B*1502 allele and carbamazepine-induced severe cutaneous adverse reactions in Han people of southern China mainland. *Seizure*, **20**(6): 446–448.

[173] Wang XQ, Lang SY, Shi XB, Tian HJ, Wang RF, and Yang F. **2011**. Cross-reactivity of skin rashes with current antiepileptic drugs in Chinese population. *Seizure*, **19**: 562–566.

[174] Selvaraj S, Schmitt AD, Dixon JR, and Ren B. **2015**. Complete haplotype phasing of the MHC and KIR loci with targeted HaploSeq. *Bmc Genomics*, **16**: 900.

[175] Chen Pei, Lin Juei-Jueng, Lu Chin-Song, Ong Cheung-Ter, Hsieh Peiyuan F, Yang Chih-Chao, Tai Chih-Ta, Wu Shey-Lin, Lu Cheng-Hsien, Hsu Yung-Chu, Yu Hsiang-Yu, Ro Long-Sun, Lu Chung-Ta, Chu Chun-Che, Tsai Jing-Jane, Su Yu-Hsiang, Lan Sheng-Hsing, Sung Sheng-Feng, Lin Shu-Yi, Chuang Hui-Ping, Huang Li-Chen, Chen Ying-Ju, Tsai Pei-Joung, Liao Hung-Ting, Lin Yu-Hsuan, Chen Chien-Hsiun, Chung Wen-Hung, Hung Shuen-Iu, Wu Jer-Yuarn, Chang Chi-Feng, Chen Luke, Chen Yuan-Tsong, Shen Chen-Yang, and Consortium Taiwan SJS. **2011**. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. *The New England journal of medicine*, **364**: 1126–1133.

[176] Major Endre, RigÃ³ Krisztina, Hague Tim, BÃ©rces Attila, and Juhos Szilveszter. **2013**. HLA typing from 1000 genomes whole genome and whole exome illumina data. *PloS one*, **8**: e78410.

[177] Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, Walker R, Hsiao T, McLaughlin L, D'Arcy M, Gai X, Goodridge D, Sayer D, and Monos D. **October 2010**. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. *Human immunology*, **71**: 1033–1042.

[178] Mayor Neema P, Robinson James, McWhinnie Alasdair J M, Ranade Swati, Eng Kevin, Midwinter William, Bultitude Will P, Chin Chen-Shan, Bowman

Brett, Marks Patrick, Braund Henny, Madrigal J Alejandro, Latham Katy, and Marsh Steven G E. **2015**. HLA Typing for the Next Generation. *PloS one*, **10**: e0127153.

[179] Pirmohamed Munir, Ostrov David A, and Park B Kevin. **2015**. New genetic findings lead the way to a better understanding of fundamental mechanisms of drug hypersensitivity. *The Journal of allergy and clinical immunology*, **136**: 236–244.

[180] Alfirevic Ana, Gonzalez-Galarza Faviel, Bell Catherine, Martinsson Klara, Platt Vivien, Bretland Giovanna, Evely Jane, Lichtenfels Maike, Cederbrant Karin, French Neil, Naisbitt Dean, Park B Kevin, Jones Andrew R, and Pirmohamed Munir. **2012**. In silico analysis of HLA associations with drug-induced liver injury: use of a HLA-genotyped DNA archive from healthy volunteers. *Genome medicine*, **4**: 51.

[181] Faulkner Lee, Gibson Andrew, Sullivan Andrew, Tailor Arun, Usui Toru, Alfirevic Ana, Pirmohamed Munir, Naisbitt Dean J, and Kevin Park B. **2016**. Detection of Primary T Cell Responses to Drugs and Chemicals in HLA-Typed Volunteers: Implications for the Prediction of Drug Immunogenicity. *Toxicological sciences : an official journal of the Society of Toxicology*, **154**: 416–429.

[182] Plumpton Catrin O, Yip Vincent L M, Alfirevic Ana, Marson Anthony G, Pirmohamed Munir, and Hughes Dyfrig A. **2015**. Cost-effectiveness of screening for HLA-A*31:01 prior to initiation of carbamazepine in epilepsy. *Epilepsia*, **56**: 556–563.

[183] Plumpton Catrin O, Roberts Daniel, Pirmohamed Munir, and Hughes Dyfrig A. **2016**. A systematic review of economic evaluations of pharmacogenetic testing for prevention of adverse drug reactions. *Pharmacoeconomics*, **34**(8): 771–793.

[184] Alfirevic Ana and Pirmohamed Munir. **2017**. Genomics of Adverse Drug Reactions. *Trends in Pharmacological Sciences*, **38**: 100–109.

[185] Otto, M and Thornton, J. **2017**. Bootstrap. Available from: https://getbootstrap.com/. Accessed on: 2017-08-22.

[186] Dawes Martin, Aloise Martin N, Ang J Sidney, Cullis Pieter, Dawes Diana, Fraser Robert, Liknaitzky Gideon, Paterson Andrea, Stanley Paul, Suarez-Gonzalez Adriana, and Katzov-Eckert Hagit. **2016**. Introducing pharmacogenetic testing with clinical decision support into primary care: a feasibility study. *CMAJ open*, **4**: E528–E534.

[187] BBC News. **2017**. Google NHS deal rebuked again by DeepMind panel. Available from: http://www.bbc.co.uk/news/technology-40497020. Accessed on: 2017-08-31.

[188] Meng Xiaoli, Lawrenson Alexandre S, Berry Neil G, Maggs James L, French Neil S, Back David J, Khoo Saye H, Naisbitt Dean J, and Park B Kevin. **2014**. Abacavir forms novel cross-linking abacavir protein adducts in patients. *Chemical research in toxicology*, **27**: 524–535.

[189] Ghattaoraya Gurpreet S, Dundar Yenal, GonzÃ¡lez-Galarza Faviel F, Maia Maria Helena Thomaz, Santos Eduardo JosÃ© Melo, da Silva AndrÃ©a Luciana Soares, McCabe Antony, Middleton Derek, Alfirevic Ana, Dickson Rumona, and Jones Andrew R. **2016**. A web resource for mining HLA associations with adverse drug reactions: HLA-ADR. *Database : the journal of biological databases and curation*, **2016**.

[190] Fricke-Galindo Ingrid, LLerena AdriÃ¡n, and LÃ³pez-LÃ³pez Marisol. **2017**. An update on HLA alleles associated with adverse drug reactions. *Drug metabolism and personalized therapy*, **32**: 73–87.

[191] Du Tingting, Yang Lun, Luo Heng, Zhou Peng, Mei Hu, Xuan Jiekun, Xing Qinghe, Ning Baitang, Mendrick Donna L, and Shi Leming. **2015**. HLADR: a database system for enhancing the discovery of biomarkers for predicting human leukocyte antigen-mediated idiosyncratic adverse drug reactions. *Biomarkers in medicine*, **9**: 1079–1093.

[192] Van Den Driessche George and Fourches Denis. **2017**. Adverse drug reactions triggered by the common HLA-B*57:01 variant: a molecular docking study. *Journal of cheminformatics*, **9**: 13.

[193] BÃ¶hm Ruwen and Cascorbi Ingolf. **2016**. Pharmacogenetics and Predictive Testing of Drug Hypersensitivity Reactions. *Frontiers in pharmacology*, **7**: 396.

[194] He Xiao-Jing, Jian Ling-Yan, He Xiao-Lin, Wu Yan, Xu Yuan-Yuan, Sun Xiao-Jie, Miao Li-Yan, and Zhao Li-Mei. **2013**. Association between the HLA-B*15:02 allele and carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis in Han individuals of northeastern China. *Pharmacological reports : PR*, **65**: 1256–62.

[195] Chong Kok Wee, Chan Derrick W. S., Cheung Yin Bun, Ching Leng Kee, Hie Szu Liang, Thomas Terrence, Ling Simon, and Tan Ene Choo. **2014**. Association of carbamazepine-induced severe cutaneous drug reactions and HLA-B*1502 allele status, and dose and treatment duration in paediatric neurology patients in Singapore. *Archives of disease in childhood*, **99**: 581–4.

[196] Chung Wen-Hung, Hung Shuen-Iu, Hong Hong-Shang, Hsih Mo-Song, Yang Li-Cheng, Ho Hsin-Chun, Wu Jer-Yuarn, and Chen Yuan-Tsong. **2004**. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature*, **428**: 486.

[197] Khor Amy Hui-Ping, Lim Kheng-Seang, Tan Chong-Tin, Wong Su-Ming, and Ng Ching-Ching. **2014**. HLA-B*15:02 association with carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in an Indian population: a pooled-data analysis and meta-analysis. *Epilepsia*, **55**: e120–4.

[198] Kulkantrakorn Kongkiat, Tassaneeyakul Wichittra, Tiamkao Somsak, Jantararoungtong Thawinee, Prabmechai Napat, Vannaprasaht Suda, Chumworathayi Pansu, Chen Pei, and Sritipsukho Paskorn. **2012**. HLA-B*1502 strongly predicts carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Thai patients with neuropathic pain. *Pain practice : the official journal of World Institute of Pain*, **12**: 202–8.

[199] Mehta Timir Y., Prajapati Laxman M., Mittal Bharti, Joshi Chaitanya G., Sheth Jayesh J., Patel Dinesh B., Dave Dinkar M., and Goyal Ramesh K. **2009**. Association of HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome among Indians. *Indian journal of dermatology, venereology and leprology*, **75**: 579–82.

[200] Tassaneeyakul Wichittra, Tiamkao Somsak, Jantararoungtong Thawinee, Chen Pei, Lin Shu-Yi, Chen Wei-Hsuan, Konyoung Parinya, Khunarkornsiri Usanee, Auvichayapat Narong, Pavakul Kasemsin, Kulkantrakorn Kongkiat, Choonhakarn Charoen, Phonhiamhan Siranun, Piyatrakul Namfon, Aungaree Thiti, Pongpakdee Sunsanee, and Yodnopaglaw Praphan. **2010**. Association between HLA-B*1502 and carbamazepine-induced severe cutaneous adverse drug reactions in a Thai population. *Epilepsia*, **51**: 926–30.

[201] Then Sue-Mian, Rani Zam Zureena Mohd, Raymond Azman Ali, Ratnaningrum Safrina, and Jamal Rahman. **2011**. Frequency of the HLA-B*1502 allele contributing to carbamazepine-induced hypersensitivity reactions in a cohort of Malaysian epilepsy patients. *Asian Pacific journal of allergy and immunology*, **29**: 290–3.

[202] Toh D. S. L., Tan L. L., Aw D. C. W., Pang S. M., Lim S. H., Thirumoorthy T., Lee H. Y., Tay Y. K., Tan S. K., Vasudevan A., Lateef A., Chong Y. Y., Chan Y. C., Loke C., Chan C. L., Koay E. S. C., Ren E. C., Lee E. J. D., and Sung C. **2014**. Building pharmacogenetics into a pharmacovigilance program in Singapore: using serious skin rash as a pilot study. *The pharmacogenomics journal*, **14**: 316–21.

[203] Wu X. T., Hu F. Y., An D. M., Yan B., Jiang X., Kwan P., Stefan H., and Zhou D. **2010**. Association between carbamazepine-induced cutaneous adverse drug reactions and the HLA-B*1502 allele among patients in central China. *Epilepsy & behavior : E&B*, **19**: 405–8.

[204] Zhang Yan, Wang Jin, Zhao Li-Mei, Peng Wei, Shen Guo-Qing, Xue Ling, Zheng Xiao-Xian, He Xiao-Jing, Gong Chun-Yan, and Miao Li-Yan. **2011**. Strong association between HLA-B*1502 and carbamazepine-induced

Stevens-Johnson syndrome and toxic epidermal necrolysis in mainland Han Chinese patients. *European journal of clinical pharmacology*, **67**: 885–7.

[205] Chang Choong-Chor, Too Chun-Lai, Murad Shahnaz, and Hussein Suraiya Hani. **2011**. Association of HLA-B*1502 allele with carbamazepine-induced toxic epidermal necrolysis and Stevens-Johnson syndrome in the multi-ethnic Malaysian population. *International journal of dermatology*, **50**(2): 221–224.

[206] BellÃ³n Teresa, RamÃ-rez Elena, Borobia Alberto M, Lerma Victoria, Moreno-Hidalgo Miguel A, Laosa Olga, Aramburu JosÃ© A, GonzÃ¡lez-Herrada Carlos, and de Abajo Francisco J. **2016**. The HLA-B*15:02 allele in a Spanish Romani patient with carbamazepine-induced Stevens-Johnson syndrome. *Pharmacogenomics*, **17**: 541–545.

[207] Emery J, Morris H, Goodchild R, Fanshawe T, Prevost A T, Bobrow M, and Kinmonth A L. **2007**. The GRAIDS Trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *British journal of cancer*, **97**: 486–493.

[208] Schoen Deborah E, Glance David G, and Thompson Sandra C. **2015**. Clinical decision support software for diabetic foot risk stratification: development and formative evaluation. *Journal of foot and ankle research*, **8**: 73.

[209] Shemeikka Tero, Bastholm-Rahmner Pia, Elinder Carl-Gustaf, VÃ©g AnikÃ³, TÃ¶rnqvist Elisabeth, Cornelius Birgitta, and Korkmaz Seher. **2015**. A health record integrated clinical decision support system to support prescriptions of pharmaceutical drugs in patients with reduced renal function: design, development and proof of concept. *International journal of medical informatics*, **84**: 387–395.

[210] Marasinghe Keshini Madara. **2015**. Computerised clinical decision support systems to improve medication safety in long-term care homes: a systematic review. *BMJ open*, **5**: e006539.

[211] Syrowatka Ania, KrÃ¶mker DÃ¶rthe, Meguerditchian Ari N, and Tamblyn Robyn. **2016**. Features of Computer-Based Decision Aids: Systematic Review, Thematic Synthesis, and Meta-Analyses. *Journal of medical Internet research*, **18**: e20.

[212] Bright Tiffani J., Wong Anthony, Dhurjati Ravi, Bristow Erin, Bastian Lori, Coeytaux Remy R., Samsa Gregory, Hasselblad Vic, Williams John W., Musty Michael D., Wing Liz, Kendrick Amy S., Sanders Gillian D., and Lobach David. **2012**. Effect of clinical decision-support systems: a systematic review. *Annals of internal medicine*, **157**: 29–43.

[213] Garg Amit X., Adhikari Neill K. J., McDonald Heather, Rosas-Arellano M. Patricia, Devereaux P. J., Beyene Joseph, Sam Justina, and Haynes R. Brian. **2005**. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*, **293**: 1223–38.

[214] Roshanov Pavel S., Misra Shikha, Gerstein Hertzel C., Garg Amit X., Sebaldt Rolf J., Mackay Jean A., Weise-Kelly Lorraine, Navarro Tamara, Wilczynski Nancy L., and Haynes R. Brian. **2011**. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. *Implementation science : IS*, **6**: 92.

[215] Nadkarni Prakash M, Ohno-Machado Lucila, and Chapman Wendy W. **2011**. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, **18**(5): 544–551.

[216] Kannry Joseph L and Williams Marc S. Integration of genomics into the electronic health record: mapping terra incognita, 2013.

Studies featured in Systematic Review

Alfirevic A, Jorgensen AL, Williamson PR, Chadwick DW, Park BK, and Pirmohamed M. **2006**. HLA-B locus in Caucasian patients with carbamazepine hypersensitivity. *Pharmacogenomics*, **7**: 813–818.

Alfirevic A, Vilar FJ, Alsbou M, Jawaid A, Thomson W, Ollier WE, Bowman CE, Delrieu O, Park BK, and Pirmohamed M. **2009**. TNF, LTA, HSPA1L and HLA-DR gene polymorphisms in HIV-positive patients with hypersensitivity to cotrimoxazole. *Pharmacogenomics*, **10**: 531–540.

An DM, Wu XT, Hu FY, Yan B, Stefan H, and Zhou D. **2010**. Association study of lamotrigine-induced cutaneous adverse reactions and HLA-B*1502 in a Han Chinese population. *Epilepsy research*, **92**: 226–230.

Cao ZH, Wei ZY, Zhu QY, Zhang JY, Yang L, Qin SY, Shao LY, Zhang YT, Xuan JK, Li QL, Xu JH, Xu F, Ma L, Huang HY, Xing QH, and Luo XQ. **2012**. HLA-B*58:01 allele is associated with augmented risk for both mild and severe cutaneous adverse reactions induced by allopurinol in Han Chinese. *Pharmacogenomics*, **13**: 1193–1201.

Carr DF, Chaponda M, Jorgensen AL, Castro EC, van JJ, Oosterhout, Khoo SH, Lalloo DG, Heyderman RS, Alfirevic A, and Pirmohamed M. **2013**. Association of human leukocyte antigen alleles and nevirapine hypersensitivity in a Malawian HIV-infected population. *Clinical infectious diseases : an official publication of the Infectious Diseases*, **56**: 1330–1339.

Chang CC, Too CL, Murad S, and Hussein SH. **2011**. Association of HLA-B*1502 allele with carbamazepine-induced toxic epidermal necrolysis and Stevens-Johnson syndrome in the multi-ethnic Malaysian population. *International journal of dermatology*, **50**: 221–224.

Chantarangsu S, Mushiroda T, Mahasirimongkol S, Kiertiburanakul S, Sungkanuparph S, Manosuthi W, Tantisiriwat W, Charoenyingwattana A, Sura T, Chantratita W, and Nakamura Y. **2009**. HLA-B*3505 allele is a strong predictor for nevirapine-induced skin adverse drug reactions in HIV-infected Thai patients. *Pharmacogenetics and genomics*, **19**: 139–146.

Almeida CA, Martin AM, Nolan D, Lucas A, Cameron PU, James I, Phillips E, and Mallal S. **2008**. Cytokine profiling in abacavir hypersensitivity patients. *Antiviral therapy*, **13**: 281–288.

Berka N, Gill JM, Liacini A, O'Bryan T, and Khan FM. **2012**. Human leukocyte antigen (HLA) and pharmacogenetics: screening for HLA-B*57:01 among human immunodeficiency virus-positive patients from southern Alberta. *Human immunology*, **73**: 164–167.

Amstutz U, Ross CJ, Castro-Pastrana LI, Rieder MJ, Shear NH, Hayden MR, and Carleton BC. **2013**. HLA-A 31:01 and HLA-B 15:02 as genetic markers for carbamazepine hypersensitivity in children. *Clinical pharmacology and therapeutics*, **94**: 142–149.

Cheung YK, Cheng SH, Chan EJ, Lo SV, Ng MH, and Kwan P. **2013**. HLA-B alleles associated with severe cutaneous reactions to antiepileptic drugs in Han Chinese. *Epilepsia*, **54**: 1307–1314.

Chiu ML, Hu M, Ng MH, Yeung CK, Chan JC, Chang MM, Cheng SH, Li L, and Tomlinson B. **2012**. Association between HLA-B*58:01 allele and severe cutaneous adverse reactions with allopurinol in Han Chinese in Hong Kong. *The British journal of dermatology*, **167**: 44–49.

Choi JH, Lee KW, Oh HB, Lee KJ, Suh YJ, Park CS, and Park HS. **2004**. HLA association in aspirin-intolerant asthma: DPB1*0301 as a strong marker in a Korean population. *The Journal of allergy and clinical immunology*, **113**: 562–564.

Chong KW, Chan DW, Cheung YB, Ching LK, Hie SL, Thomas T, Ling S, and Tan EC. **2014**. Association of carbamazepine-induced severe cutaneous drug reactions and HLA-B*1502 allele status, and dose and treatment duration in paediatric neurology patients in Singapore. *Archives of disease in childhood*, **99**: 581–584.

Chung WH, Hung SI, Hong HS, Hsih MS, Yang LC, Ho HC, Wu JY, and Chen YT. **2004**. Medical genetics: a marker for Stevens-Johnson syndrome. *Nature*, **428**: 486.

Cristallo AF, Schroeder J, Citterio A, Santori G, Ferrioli GM, Rossi U, Bertani G, Cassano S, Gottardi P, Ceschini N, Barocci F, Ribizzi G, Cutrupi V, Cairoli R, Rapisarda V, Pastorello EA, and Barocci S. **2011**. A study of HLA class I and class II 4-digit allele level in Stevens-Johnson syndrome and toxic epidermal necrolysis. *International journal of immunogenetics*, **38**: 303–309.

Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A, Daly MJ, Goldstein DB, John S, Nelson MR, Graham J, Park BK, Dillon JF, Bernal W, Cordell HJ, Pirmohamed M, Aithal GP, and Day CP. **2009**. HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nature genetics*, **41**: 816–819.

Dettling M, Cascorbi I, Roots I, and Mueller-Oerlinghausen B. **2001**. Genetic determinants of clozapine-induced agranulocytosis: recent results of HLA subtyping in a non-jewish caucasian sample. *Archives of general psychiatry*, **58**: 93–94.

Dettling M, Schaub RT, Mueller-Oerlinghausen B, Roots I, and Cascorbi I. **2001**. Further evidence of human leukocyte antigen-encoded susceptibility to clozapine-induced agranulocytosis independent of ancestry. *Pharmacogenetics*, **11**: 135–141.

Fuerst D, Parmar S, Schumann C, Rudiger S, Boeck S, Heinemann V, Kaechele V, Stiebel K, Paul T, Seufferlein T, Mytilineos J, and Stingl JC. **2012**. HLA polymorphisms influence the development of skin rash arising from treatment with EGF receptor inhibitors. *Pharmacogenomics*, **13**: 1469–1476.

Furukawa H, Oka S, Shimada K, Sugii S, Hashimoto A, Komiya A, Fukui N, Miyashita T, Migita K, Suda A, Nagaoka S, Tsuchiya N, and Tohma S. **2014**. HLA-DRB1*08:02 Is Associated with Bucillamine-Induced Proteinuria in Japanese Rheumatoid Arthritis Patients. *Biomarker insights*, **9**: 23–28.

Goncalo M, Coutinho I, Teixeira V, Gameiro AR, Brites MM, Nunes R, and Martinho A. **2013**. HLA-B*58:01 is a risk factor for allopurinol-induced DRESS and Stevens-Johnson syndrome/toxic epidermal necrolysis in a Portuguese population. *The British journal of dermatology*, **169**: 660–665.

He N, Min FL, Shi YW, Guo J, Liu XR, Li BM, Zhou JH, Ou YM, Liao JX, and Liao WP. **2012**. Cutaneous reactions induced by oxcarbazepine in Southern Han Chinese: incidence, features, risk factors and relation to HLA-B alleles. *Seizure : the journal of the British Epilepsy Association*, **21**: 614–618.

He XJ, Jian LY, He XL, Wu Y, Xu YY, Sun XJ, Miao LY, and Zhao LM. **2013**. Association between the HLA-B*15:02 allele and carbamazepine-induced tevens-Johnson syndrome/toxic epidermal necrolysis in Han individuals of ortheastern China. *Pharmacological reports : PR*, **65**: 1256–1262.

Hirata K, Takagi H, Yamamoto M, Matsumoto T, Nishiya T, Mori K, Shimizu S, Masumoto H, and Okutani Y. **2008**. Ticlopidine-induced hepatotoxicity is associated with specific human leukocyte antigen genomic subtypes in Japanese patients: a preliminary case-control study. *The pharmacogenomics journal*, **8**: 29–33.

Hsiao YH, Hui RC, Wu T, Chang WC, Hsih MS, Yang CH, Ho HC, Chang YG, Chen MJ, Lin JY, Chen DP, Chang PY, Wu TL, Hung SI, and Chung WH. **2014**. Genotype-phenotype association between HLA and carbamazepine-induced hypersensitivity reactions: strength and clinical correlations. *Journal of dermatological science*, **73**: 101–109.

Hu FY, Wu XT, An DM, Yan B, Stefan H, and Zhou D. **2011**. Pilot association study of oxcarbazepine-induced mild cutaneous adverse reactions with HLA-B*1502 allele in Chinese Han population. *Seizure : the journal of the British Epilepsy Association*, **20**: 160–162.

Hung SI, Chung WH, Jee SH, Chen WC, Chang YT, Lee WR, Hu SL, Wu MT, Chen GS, Wong TW, Hsiao PF, Chen WH, Shih HY, Fang WH, Wei CY, Lou YH, Huang YL, Lin JJ, and Chen YT. **2006**. Genetic susceptibility to carbamazepine-induced cutaneous adverse drug reactions. *Pharmacogenetics and genomics*, **16**: 297–306.

Hung SI, Chung WH, Liou LB, Chu CC, Lin M, Huang HP, Lin YL, Lan JL, Yang LC, Hong HS, Chen MJ, Lai PC, Wu MS, Chu CY, Wang KH, Chen CH, Fann CS, Wu JY, and Chen YT. **2005**. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proceedings of the National Academy of Sciences of the United States of America*, **102**: 4134–4139.

Hughes DA, Vilar FJ, Ward CC, Alfirevic A, Park BK, and Pirmohamed M. **2004**. Cost-effectiveness analysis of HLA B*5701 genotyping in preventing abacavir hypersensitivity. *Pharmacogenetics*, **14**: 335–342.

Hung SI, Chung WH, Liu ZS, Chen CH, Hsih MS, Hui RC, Chu CY, and Chen YT. **2010**. Common risk allele in aromatic antiepileptic-drug induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Han Chinese. *Pharmacogenomics*, **11**: 349–356.

Kang HR, Jee YK, Kim YS, Lee CH, Jung JW, Kim SH, Park HW, Chang YS, Jang IJ, Cho SH, Min KU, Kim SH, and Lee KW. **2011**. Positive and negative associations of HLA class I alleles with allopurinol-induced SCARs in Koreans. *Pharmacogenetics and genomics*, **21**: 303–307.

Kaniwa N, Saito Y, Aihara M, Matsunaga K, Tohkin M, Kurose K, Furuya H, Takahashi Y, Muramatsu M, Kinoshita S, Abe M, Ikeda H, Kashiwagi M, Song Y, Ueta M, Sotozono C, Ikezawa Z, and Hasegawa R. **2010**. HLA-B*1511 is a risk factor for carbamazepine-induced Stevens-Johnson syndrome nd toxic epidermal necrolysis in Japanese patients. *Epilepsia*, **51**: 2461–2465.

Kaniwa N, Saito Y, Aihara M, Matsunaga K, Tohkin M, Kurose K, Sawada J, Furuya H, Takahashi Y, Muramatsu M, Kinoshita S, Abe M, Ikeda H, Kashiwagi M, Song Y, Ueta M, Sotozono C, Ikezawa Z, and Hasegawa R. **2008**. HLA-B locus in Japanese patients with anti-epileptics and allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis. *Pharmacogenomics*, **9**: 1617–1622.

Kashiwagi M, Aihara M, Takahashi Y, Yamazaki E, Yamane Y, Song Y, Muramatsu M, and Ikezawa Z. **2008**. Human leukocyte antigen genotypes in carbamazepine-induced severe cutaneous adverse drug response in Japanese patients. *The Journal of dermatology*, **35**: 683–685.

Kazeem GR, Cox C, Aponte J, Messenheimer J, Brazell C, Nelsen AC, Nelson MR, and Foot E. **2009**. High-resolution HLA genotyping and severe cutaneous adverse reactions in lamotrigine-treated patients. *Pharmacogenetics and genomics*, **19**: 661–665.

Kim SH, Choi JH, Lee KW, Kim SH, Shin ES, Oh HB, Suh CH, Nahm DH, and Park HS. **2005**. The human leucocyte antigen-DRB1*1302-DQB1*0609-DPB1*0201 haplotype may be a strong genetic marker for aspirin-induced urticaria. *Clinical and experimental allergy : journal of the British Society for Allergy*, **35**: 339–344.

Kim SH, Lee KW, Song WJ, Kim SH, Jee YK, Lee SM, Kang HR, Park HW, Cho SH, Park SH, Min KU, and Chang YS. **2011**. Carbamazepine-induced severe cutaneous adverse reactions and HLA genotypes in Koreans. *Epilepsy research*, **97**: 190–197.

Kim SH, Lee SK, Kim SH, Park HW, Chang YS, Lee KW, and Jee YK. **2013**. Antituberculosis drug-induced hypersensitivity syndrome and its association with human leukocyte antigen. *Tuberculosis (Edinburgh, Scotland)*, **93**: 270–274.

Kindmark A, Jawaid A, Harbron CG, Barratt BJ, Bengtsson OF, Andersson TB, Carlsson S, Cederbrant KE, Gibson NJ, Armstrong M, Lagerstrom-Fermer ME, Dellsen A, Brown EM, Thornton M, Dukes C, Jenkins SC, Firth MA, Harrod GO, Pinel TH, Billing-Clason SM, Cardon LR, and March RE. **2008**. Genome-wide pharmacogenetic investigation of a hepatic adverse event without linical signs of immunopathology suggests an underlying immune pathogenesis. *The pharmacogenomics journal*, **8**: 186–195.

Kulkantrakorn K, Tassaneeyakul W, Tiamkao S, Jantararoungtong T, Prabmechai N, Vannaprasaht S, Chumworathayi P, Chen P, and Sritipsukho P. **2012**. HLA-B*1502 strongly predicts carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Thai patients with neuropathic pain. *Pain practice : the official journal of World Institute of Pain*, **12**: 202–208.

Lee HY, Lee JW, Lee KW, Park MH, and Park HS. **2009**. The HLA allele marker for differentiating ASA hypersensitivity phenotypes. *Allergy*, **64**: 1385–1387.

Li LJ, Hu FY, Wu XT, An DM, Yan B, and Zhou D. **2013**. Predictive markers for carbamazepine and lamotrigine-induced maculopapular exanthema in Han Chinese. *Epilepsy research*, **106**: 296–300.

Littera R, Carcassi C, Masala A, Piano P, Serra P, Ortu F, Corso N, Casula B, La G, Nasa, Contu L, and Manconi PE. **2006**. HLA-dependent hypersensitivity to nevirapine in Sardinian HIV patients. *AIDS (London, England)*, **20**: 1621–1626.

Locharernkul C, Loplumlert J, Limotai C, Korkij W, Desudchit T, Tongkobpetch S, Kangwanshiratada O, Hirankarn N, Suphapeetiporn K, and Shotelersuk V. **2008**. Carbamazepine and phenytoin induced Stevens-Johnson syndrome is associated with HLA-B*1502 allele in Thai population. *Epilepsia*, **49**: 2087–2091.

Lonjou C, Borot N, Sekula P, Ledger N, Thomas L, Halevy S, Naldi L, Bouwes-Bavinck JN, Sidoroff A, de C, Toma, Schumacher M, Roujeau JC, Hovnanian A, and Mockenhaupt M. **2008**. A European study of HLA-B in Stevens-Johnson syndrome and toxic epidermal necrolysis related to five high-risk drugs. *Pharmacogenetics and genomics*, **18**: 99–107.

Lv YD, Min FL, Liao WP, He N, Zeng T, Ma DH, and Shi YW. **2013**. The association between oxcarbazepine-induced maculopapular eruption and HLA-B alleles in a northern Han Chinese population. *BMC neurology*, **13**: 75.

Man CB, Kwan P, Baum L, Yu E, Lau KM, Cheng AS, and Ng MH. **2007**. Association between HLA-B*1502 allele and antiepileptic drug-induced cutaneous reactions in Han Chinese. *Epilepsia*, **48**: 1015–1018.

Mallal S, Nolan D, Witt C, Masel G, Martin AM, Moore C, Sayer D, Castley A, Mamotte C, Maxwell D, James I, and Christiansen FT. **2002**. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet*, **359**: 727–732.

Mammen AL, Gaudet D, Brisson D, Christopher-Stine L, Lloyd TE, Leffell MS, and Zachary AA. **2012**. Increased frequency of DRB1*11:01 in anti-hydroxymethylglutaryl-coenzyme A reductase-associated autoimmune myopathy. *Arthritis care & research*, **64**: 1233–1237.

Manuyakorn W, Siripool K, Kamchaisatian W, Pakakasama S, Visudtibhan A, Vilaiyuk S, Rujirawat T, and Benjaponpitak S. **2013**. Phenobarbital-induced severe cutaneous adverse drug reactions are associated with YP2C19*2 in Thai children. *Pediatric allergy and immunology : official publication of the European Society*, **24**: 299–303.

McCormack M, Alfirevic A, Bourgeois S, Farrell JJ, Kasperaviciute D, Carrington M, Sills GJ, Marson T, Jia X, de PI, Bakker, Chinthapalli K, Molokhia M, Johnson MR, O'Connor GD, Chaila E, Alhusaini S, Shianna KV, Radtke RA, Heinzen EL, Walley N, Pandolfo M, Pichler W, Park BK, Depondt C, Sisodiya SM, Goldstein DB, Deloukas P, Delanty N, Cavalleri GL, and Pirmohamed M. **2011**. HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *The New England journal of medicine*, **364**: 1134–1143.

Mehta TY, Prajapati LM, Mittal B, Joshi CG, Sheth JJ, Patel DB, Dave DM, and Goyal RK. **2009**. Association of HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome among Indians. *Indian journal of dermatology, venereology and leprology*, **75**: 579–582.

Niihara H, Kakamu T, Fujita Y, Kaneko S, and Morita E. **2012**. HLA-A31 strongly associates with carbamazepine-induced adverse drug reactions but not with carbamazepine-induced lymphocyte proliferation in a Japanese population. *The Journal of dermatology*, **39**: 594–601.

Niihara H, Kaneko S, Ito T, Sugamori T, Takahashi N, Kohno K, and Morita E. **2013**. HLA-B*58:01 strongly associates with allopurinol-induced adverse drug reactions n a Japanese sample population. *Journal of dermatological science*, **71**: 150–152.

O'Donohue J, Oien KA, Donaldson P, Underhill J, Clare M, MacSween RN, and Mills PR. **2000**. Co-amoxiclav jaundice: clinical and histological features and HLA class II association. *Gut*, **47**: 717–720.

Ozeki T, Mushiroda T, Yowang A, Takahashi A, Kubo M, Shirakata Y, Ikezawa Z, Iijima M, Shiohara T, Hashimoto K, Kamatani N, and Nakamura Y. **2011**. Genome-wide association study identifies HLA-A*3101 allele as a genetic risk factor for carbamazepine-induced cutaneous adverse drug reactions in Japanese population. *Human molecular genetics*, **20**: 1034–1041.

Phillips E, Bartlett JA, Sanne I, Lederman MM, Hinkle J, Rousseau F, Dunn D, Pavlos R, James I, Mallal SA, and Haas DW. **2013**. Associations between HLA-DRB1*0102, HLA-B*5801, and hepatotoxicity during initiation of nevirapine-containing regimens in South Africa. *Journal of acquired immune deficiency syndromes (1999)*, **62**.

Martin AM, Nolan D, Gaudieri S, Almeida CA, Nolan R, James I, Carvalho F, Phillips E, Christiansen FT, Purcell AW, McCluskey J, and Mallal S. **2004**. Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 4180–4185.

Rauch A, Nolan D, Thurnheer C, Fux CA, Cavassini M, Chave JP, Opravil M, Phillips E, Mallal S, and Furrer H. **2008**. Refining abacavir hypersensitivity diagnoses using a structured clinical assessment and genetic testing in the Swiss HIV Cohort Study. *Antiviral therapy*, **13**: 1019–1028.

Shi YW, Min FL, Liu XR, Zan LX, Gao MM, Yu MJ, and Liao WP. **2011**. Hla-B alleles and lamotrigine-induced cutaneous adverse drug reactions in the Han Chinese population. *Basic & clinical pharmacology & toxicology*, **109**: 42–46.

Shi YW, Min FL, Qin B, Zou X, Liu XR, Gao MM, Wang Q, Zhou JQ, and Liao WP. **2012**. Association between HLA and Stevens-Johnson syndrome induced by carbamazepine in Southern Han Chinese: genetic markers besides B*1502? *Basic & clinical pharmacology & toxicology*, **111**: 58–64.

Singer JB, Lewitzky S, Leroy E, Yang F, Zhao X, Klickstein L, Wright TM, Meyer J, and Paulding CA. **2010**. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nature genetics*, **42**: 711–714.

Spraggs CF, Budde LR, Briley LP, Bing N, Cox CJ, King KS, Whittaker JC, Mooser VE, Preston AJ, Stein SH, and Cardon LR. **2011**. HLA-DQA1*02:01 is a major risk factor for lapatinib-induced hepatotoxicity in women with advanced breast cancer. *Journal of clinical oncology : official journal of the American Society of*, **29**: 667–673.

Stephens C, Lopez-Nevot MA, Ruiz-Cabello F, Ulzurrun E, Soriano G, Romero-Gomez M, Moreno-Casares A, Lucena MI, and Andrade RJ. **2013**. HLA alleles influence the clinical signature of amoxicillin-clavulanate hepatotoxicity. *PloS one*, **8**.

Tassaneeyakul W, Jantararoungtong T, Chen P, Lin PY, Tiamkao S, Khunarkornsiri U, Chucherd P, Konyoung P, Vannaprasaht S, Choonhakarn C, Pisuttimarn P, Sangviroon A, and Tassaneeyakul W. **2009**. Strong association between HLA-B*5801 and allopurinol-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in a Thai population. *Pharmacogenetics and genomics*, **19**: 704–709.

Tassaneeyakul W, Tiamkao S, Jantararoungtong T, Chen P, Lin SY, Chen WH, Konyoung P, Khunarkornsiri U, Auvichayapat N, Pavakul K, Kulkantrakorn K, Choonhakarn C, Phonhiamhan S, Piyatrakul N, Aungaree T, Pongpakdee S, and Yodnopaglaw P. **2010**. Association between HLA-B*1502 and carbamazepine-induced severe cutaneous adverse drug reactions in a Thai population. *Epilepsia*, **51**: 926–930.

Then SM, Rani ZZ, Raymond AA, Ratnaningrum S, and Jamal R. **2011**. Frequency of the HLA-B*1502 allele contributing to carbamazepine-induced ypersensitivity reactions in a cohort of Malaysian epilepsy patients. *Asian Pacific journal of allergy and immunology / launched by the Allergy and*, **29**: 290–293.

Rodriguez-Novoa S, Garcia-Gasco P, Blanco F, Gonzalez-Pardo G, Castellares C, Moreno V, Jimenez-Nacher I, Gonzalez-Lahoz J, and Soriano V. **2007**. Value of the HLA-B*5701 allele to predict abacavir hypersensitivity in Spaniards. *AIDS research and human retroviruses*, **23**: 1374–1376.

Stekler J, Maenza J, Stevens C, Holte S, Malhotra U, McElrath MJ, Corey L, and Collier AC. **2006**. Abacavir hypersensitivity reaction in primary HIV infection. *AIDS (London, England)*, **20**: 1269–1274.

Wang Q, Zhou JQ, Zhou LM, Chen ZY, Fang ZY, Chen SD, Yang LB, Cai XD, Dai QL, Hong H, and Wang HX. **2011**. Association between HLA-B*1502 allele and carbamazepine-induced severe cutaneous adverse reactions in Han people of southern China mainland. *Seizure: the journal of the British Epilepsy Association*, **20**: 446–448.

Wu XT, Hu FY, An DM, Yan B, Jiang X, Kwan P, Stefan H, and Zhou D. **2010**. Association between carbamazepine-induced cutaneous adverse drug reactions and the HLA-B*1502 allele among patients in central China. *Epilepsy & behavior : E&B*, **19**: 405–408.

Zhang FR, Liu H, Irwanto A, Fu XA, Li Y, Yu GQ, Yu YX, Chen MF, Low HQ, Li JH, Bao FF, Foo JN, Bei JX, Jia XM, Liu J, Liany H, Wang N, Niu GY, Wang ZZ, Shi BQ, Tian HQ, Liu HX, Ma SS, Zhou Y, You JB, Yang Q, Wang C, Chu TS, Liu DC, Yu XL, Sun YH, Ning Y, Wei ZH, Chen SL, Chen XC, Zhang ZX, Liu YX, Pulit SL, Wu WB, Zheng ZY, Yang RD, Long H, Liu ZS, Wang JQ, Li M, Zhang LH, Wang H, Wang LM, Xiao P, Li JL, Huang ZM, Huang JX, Li Z, Liu J, Xiong L, Yang J, Wang XD, Yu DB, Lu XM, Zhou GZ, Yan LB, Shen JP, Zhang GC, Zeng YX, de PI, Bakker, Chen SM, and Liu JJ. **2013**. HLA-B*13:01 and the dapsone hypersensitivity syndrome. *The New England journal of medicine*, **369**: 1620–1628.

Zhang Y, Wang J, Zhao LM, Peng W, Shen GQ, Xue L, Zheng XX, He XJ, Gong CY, and Miao LY. **2011**. Strong association between HLA-B*1502 and carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in mainland Han Chinese patients. *European journal of clinical pharmacology*, **67**: 885–887.