Chapter 2

HLA-DQA1*05 is associated with immunogenicity to anti-TNF therapy

Parts of this chapter were published in Gastroenterology [165]. This project is a collaboration of more than 400 principal investigators, research nurses, technicians, and other professions who have contributed to it. My contribution was to co-design and to run the majority of the genetic analyses, with the exception of genotype imputation, HLA to amino-acid translation, and HLA typed versus imputed concordance. I am grateful for the opportunity to be involved in this study and would like to acknowledge the contributions of other co-authors.

2.1 Introduction

Biological therapies, commonly known as biologics, are typically large and complex proteins manufactured in, or derived from, living sources. Biologics have transformed the management of immune-mediated diseases and are are starting to find their application in the treatment of HIV and cancer. In 2017, biologics accounted for a global expenditure in excess of $100 billion [25]. Compared to the traditional small molecule drugs, biologics offer greater specificity, resulting in better effectiveness and less off-target effects during the treatment [138].

Globally, one of the biggest challenges to the wider adoption of biologics is cost. Just in 2017/2018, the National Health Service (UK) spent £400 mil on a single biologic therapy – adalimumab, which is was prescribed to 46,000 patients. Generally, for the immune-mediated disorders, a course of treatment with biologics far exceeds the cost of an alternative treatment with small-molecule therapies. The high costs are explained by several factors. Like with most modern therapies, the cost of development for novel a drug 'from lab bench to the market' can often be in the range of billions of dollars. Expectedly, the pharmaceutical companies use the time during which the drug is protected by the patents to maximise the profit in order to cover the development costs, subsidise the development of future therapies, and to make a profit. Such a costing strategy is not unique to biologics, but it is exacerbated by the cost and complexity of the development and production of the generic version of the biologics – biosimilars.

Blackstone et al. [25] estimate that it costs $1 million to $4 million to bring a generic for a small-molecule drug to the market. In contrast, it takes 7 to 8 years to develop a biosimilar, at a cost of between $100 million and $250 million. Sadly, even the arrival of the first competitive biosimilar does not always bring down the pricing of the treatment. Dave et al. [46] estimate that, on average, the price of the first generic drug is 87% of that of the brand-name drug and 60% once four generic manufacturers have entered the market. As of March 2019, there are ten adalimumab biosimilars, produced by six manufacturers, approved for use in the European Union. According to the NHS estimates, the arrival of the adalimumab biosimilars will allow savings of 'at least £150 million per year by 2021'. The cost of the treatment must not limit the prescription of the biologic therapies to the patients who will benefit from them, but there is a real urgency in understanding who exactly those patients are.

The high cost of treatment is not the the only issue facing the biologic treatments. The current generation of biologics rely on invasive injections because of their poor bioavailability via the oral route [8]. Some biologics, like infliximab, have to be delivered intravenously (IV) every few weeks, putting an additional burden on the patients' quality of life. Others, like adalimumab, can be used via an autoinjector pen. Future biologics, currently in development, might be delivered orally (e.g., OPRX-106 that has reached the phase 2 trial for UC).

Most importantly, just like the majority of other therapies, biologics are not always effective. Considering that biologic therapies are often used for patients refractory to other treatments, it is important to study the risk factors that influence this. In the chapter below, I describe results from the PANTS study. Our work has uncovered the first robust genetic association for development of antibodies against anti-TNF which are know to have a substantial negative impact on treatment efficacy.

Anti-tumour necrosis factor (anti-TNF) therapies are the most widely used biologics for treating immune-mediated diseases. Anti-tumour necrosis factor (anti-TNF) therapy has been effective in the treatment of UC and CD, as well as a variety of other immune-mediated complex traits, including rheumatoid arthritis and refractory asthma. Anti-TNF therapies interferes with the action of TNF, a proinflammatory cytokine that is involved in the innate immune response. In 2018, anti-TNF therapies accounted for a global expenditure in excess of $23.5 billion [86].

For Crohn's disease patients, anti-TNF therapy has been associated with mucosal healing, improved quality of life, and reduced hospitalisation and surgery rates [17]. Despite the general efficacy of the treatment, 10–30% of patients do not respond to anti-TNF and a further 23–46% of patients lose response over time [156]. As with other biologic therapies, repeated administration of anti-TNF often induces the formation of anti-drug antibodies (immunogenicity) leading to treatment failure [96].

Immunogenicity is more common in patients treated with infliximab (a murine-human chimeric monoclonal antibody) than adalimumab (a fully human monoclonal antibody) and is a major cause of low anti-TNF drug level, infusion reactions, and non-remission in patients with Crohn's disease [96, 183]. Combination immunomodulator therapy reduces the risk of immunogenicity to both adalimumab and infliximab, and for infliximab, improves treatment outcomes [96, 41, 140]. Despite these benefits, many patients are still treated with anti-TNF monotherapy because of concerns about the increased risk of adverse drug

reactions, opportunistic infections, and malignancies associated with combination therapy with immunomodulators [54, 109, 139].

The ability to identify patients at increased risk of immunogenicity would direct the choice of anti-TNF treatment and the use of preventative strategies, including combination therapy with immunomodulators. However, our understanding of the cellular and molecular mechanisms underpinning immunogenicity to biologics is limited. Retrospective small-scale studies have suggested variants in *FCGR3A* [158], CD96 [11] and *HLA-DRB1* [23, 112] increase susceptibility to immunogenicity to anti-TNF therapy. These associations either did not achieve genome-wide significance [23, 112, 158] or are yet to be independently replicated [158, 11]. Both studies reporting the *HLA-DRB1* only looked at a small subset of HLA alleles, which is problematic given the long-range linkage disequilibrium within the human leukocyte antigen (HLA) region. The CD96 was uncovered in a small discovery cohort (N=62, OR = 20.2, 95% CI, 5.57–73.27, P = 1.88×10^{-9}), yet has a strikingly different effect size in a larger self-replication cohort (N=88, OR = 1.16, 95% CI, 1.09–1.23, P = 0.044), violating the homogeneity of odds assumption for replication. Here, I report the first genetic locus robustly associated with immunogenicity to anti-TNF therapies.

2.2 Methods

2.2.1 PANTS study: patient recruitment and phenotyping

The Personalising Anti-TNF Therapy in Crohn's disease (PANTS) study is a UK-wide, multi-centre, prospective observational cohort reporting the treatment failure rates of the anti-TNF drugs infliximab (originator, Remicade [Merck Sharp & Dohme, UK] and biosimilar, CT-P13 [Celltrion, South Korea]), and adalimumab (Humira [Abbvie, USA]) in 1,610 anti-TNF-naive patients with Crohn's disease [96].

The South West Research Ethics committee approved the study (REC reference: 12 / SW / 0323) in January, 2013. Patients were included after providing informed, written consent. The protocol is available online (www.ibdresearch.co.uk).

At inclusion, subjects were aged 6 years or over and had active luminal Crohn's disease involving the colon and/or small intestine. Choice of anti-TNF drug and use of concomitant

immunomodulator therapy was at the discretion of the treating physician as part of usual care. Patients were initially studied for 12 months or until drug withdrawal. In the first year, study visits were scheduled at first dose, post-induction (weeks 12–14), weeks 30, 54, and at treatment failure. For infliximab-treated patients, additional visits occurred at each infusion. After 12 months, patients were invited to continue follow-up for a further two years. Drug persistence was defined as the duration of time from initiation of anti-TNF therapy to exit from the study due to treatment failure. Patients who exited the study for other reasons, declined to participate in the two-year extension, or were lost to follow-up were censored at the time of last drug dose or study visit.

At each visit, serum infliximab or adalimumab drug and anti-drug antibody levels were analysed using total antibody enzyme linked immunosorbent assays [145]. The total antibody, unlike the more commonly reported free antibody assay, includes a drug-antibody disassociation step that allows the assessment of anti-drug antibodies in the presence of drug. In this study, immunogenicity was defined as an anti-drug antibody concentration of $\geq$10 AU/ml, irrespective of drug level, at one or more time points.

We assembled an independent cohort to replicate significant findings from the discovery cohort. This comprised 107 Crohn's disease, 64 ulcerative colitis, and 7 IBD type-unclassified patients all with cross-sectional drug and antibody levels measured as part of routine clinical practice. The samples were genotyped using either the Illumina CoreExome array (N=164) [49] or the Affymetrix 500k array (N=14) [190]. Quality control and imputation methods were the same as in the discovery cohort (see 2.2.3).

2.2.2 Measurement of drug and anti-drug antibody levels

Antibody and drug level measurements were carried out by our collaborators at the University of Exeter.

Serum infliximab and adalimumab drug levels were analyzed on the Dynex (Chantilly Virginia, USA) DS2 automated Enzyme-Linked ImmunoSorbent Assay (ELISA) platform, using the Immundiagnostik (Immundiagnostik AG, Bensheim, Germany) IDKmonitor® drug (K9655 infliximab drug level and K9657 adalimumab drug level) and total antibody ELISA assays (K9654 infliximab total anti-drug antibody and K9651 adalimumab total anti-drug antibody). These assays allow quantitative determination of free infliximab and adalimumab

using a sandwich ELISA technique. The IDK monitor infliximab drug level assay has a measuring range of 0.8–45 mg/L, with an intra-assay CV of <9.7% and an inter-assay coefficient of variation (CV) of <11.0%. The IDK monitor adalimumab drug level assay has a measuring range of 0.8–45 mg/L, with an intra-assay CV of <2.6% and an inter-assay CV of <13.0%. Positive anti-drug antibody status was defined in line with the manufacturer's recommendations as a concentration $\geq$10 AU/mL, irrespective of drug level. All assays, including drug and antibody levels, were tested for stability.

In independent experiments, our collaborators confirmed that this cut-off corresponds to the 98th percentile of the anti-drug antibody titre distribution in more than 500 drug-naïve controls.

2.2.3 Genotyping, quality control, and imputation

DNA was extracted from pre-treatment blood samples from 1,524 individuals in the PANTS cohort and genotyping undertaken using the Illumina CoreExome microarray (522,049 markers), with genotype calls made using optiCall [167]. Pre-imputation quality control (QC) of samples and single nucleotide polymorphisms (SNPs) was performed as previously described [49]. SNP genotypes were imputed via the Sanger Imputation Service using the Haplotype Reference Consortium (HRC) panel as a part of a larger GWAS study [120].

Following imputation, I ran additional variant and sample QC procedures. I calculated the first 15 principal components (PC) for the 1,323 samples that passed the QC criteria (Figure 2.1a). The Tracy-Widom test showed that only the first principal component explained a significant proportion of the variance. In addition, I confirmed the European ancestry of our cohort by performing principal component analyses with 2,504 samples from the 1000 Genomes Project (1KGP) [1] (Figure 2.1). Following standard practices, I discarded poorly imputed SNPs, defined as having an information score <0.4. I removed SNPs significantly deviated from Hardy-Weinberg equilibrium ($P<1\times10^{-10}$), with a call-rate $\leq$0.95 or minor allele frequency $\leq$1%, leaving 7,578,947 variants. PCA and genotype filtering were performed using the Hail framework (version 0.1).

I excluded individuals of non-European ancestry (identified using principal component analysis), one individual from each related pair (defined as a pi-hat >0.1875, halfway between the expected pi-hat for third- and second-degree relatives [7]), and those with an outlying

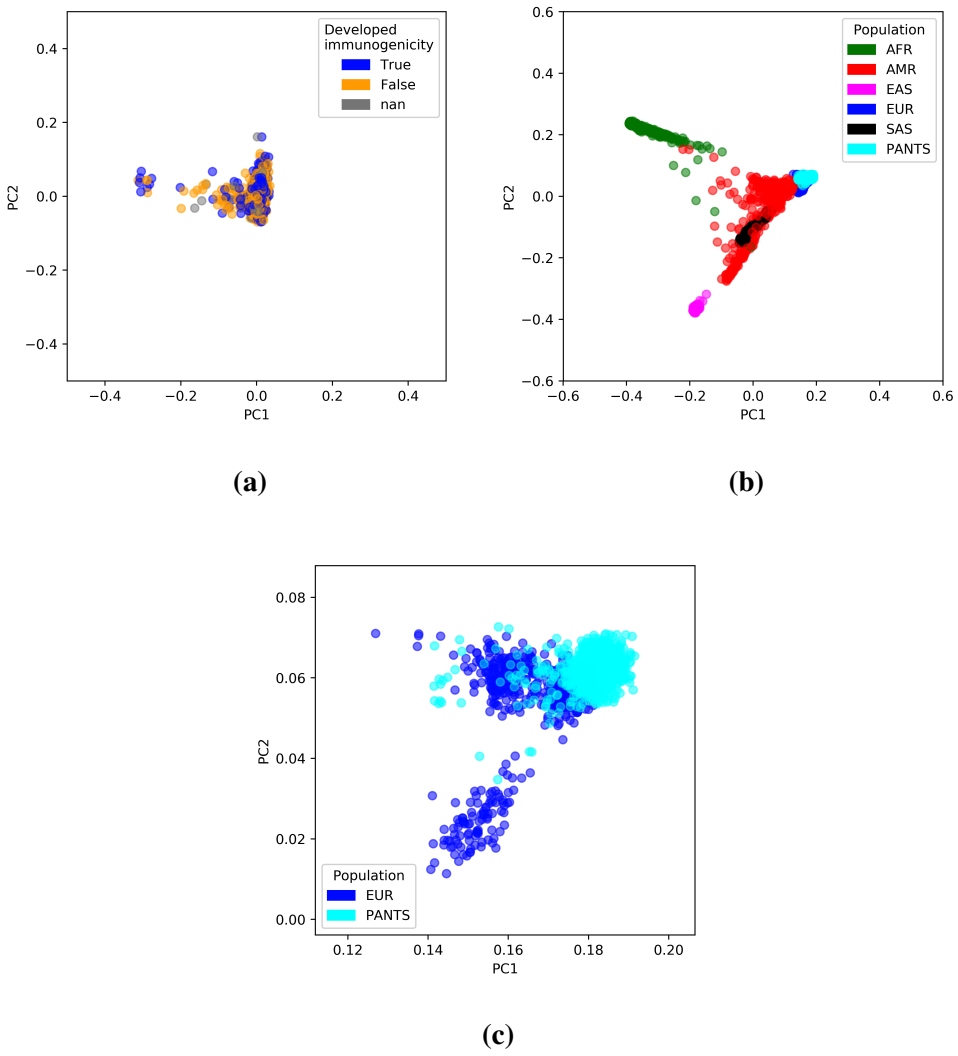


Figure 2.1 Principal component analysis (PCA) on imputed data from 1,323 individuals in the PANTS study. (A) PCA on the PANTS cohort alone. Blue dots show individuals who develop immunogenicity and orange dots show individuals who did not. PC1 was not associated with immunogenicity status (P=0.99). (B, C) PCA analysis of the 1,323 individuals from the PANTS cohort together with the 2,504 individuals from the 1000 Genomes Project (1KGP) cohort. (B) PCA confirmed that PANTS samples clustered together with the individuals of European ancestry from the 1KGP. (C) A detailed view of PANTS samples plotted together with European samples from the 1KGP.

number of missing or heterozygous genotypes. 1,323 individuals remained in the study

following quality control, of which 1,240 had drug and antibody level data available (Figure 2.2).

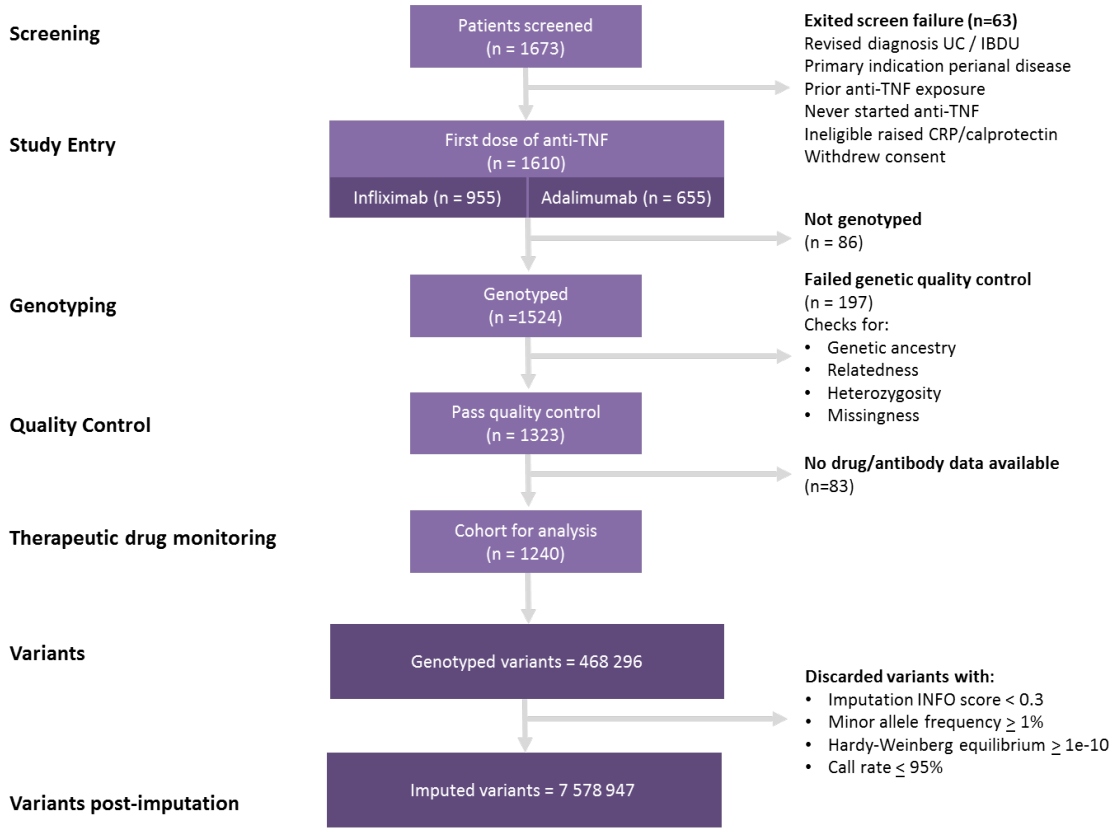


Figure 2.2 Flowchart describing the cohort used for the time to immunogenicity genetic analysis Abbreviations: UC = ulcerative colitis, IBDU: inflammatory bowel disease unclassified, anti-TNF: anti-tumor necrosis factor, CRP: C-reactive protein, INFO: information content metric

HLA imputation was carried out using the HIBAG package [198] in R, using pre-fit classifiers trained specifically for the CoreExome genotyping microarray on individuals of European ancestry. HLA types were imputed at 2-, and 4-digit resolution for the following loci: *HLA-A, HLA-C, HLA-B, HLA-DRB1, HLA-DQA1, HLA-DQB1,* and *HLA-DPB1*. In addition, my colleague obtained amino acid sequences for all the imputed HLA alleles, using the IPD-IMGT/HLA database [155]. Following the recommended best practices for HIBAG, HLA-allele and amino acid calls with posterior probability <0.5 were set to missing for the given individual. To confirm our imputation and genetic association our collaborators carried out long-read sequencing of the HLA (Histogenetics, New York, USA).

Variable	Level	infliximab (n = 742)	adalimumab (n = 498)	p
Sex (male)		353 (47.6%)	234 (47.0%)	0.862
Age (years)		31.3 (21.2 - 46.0)	37.6 (28.7 - 50.3)	1.1×10^{-11}
Disease duration (years)		2.6 (0.7 - 9.8)	3.1 (0.8 - 11.5)	0.033
Age at diagnosis (years)		24.4 (16.4 - 35.9)	29.5 (21.8 - 41.8)	1.7×10^{-11}
Montreal location	**L1**	207 (28.2%)	157 (32.0%)	0.334
	L2	190 (25.9%)	108 (22.0%)	
	L3	332 (45.2%)	223 (45.4%)	
	L4	6 (0.8%)	3 (0.6%)	
Montreal L4		78 (10.6%)	23 (4.7%)	1.8×10^{-4}
Montreal behaviour	**B1**	455 (61.7%)	277 (56.3%)	1.9×10^{-5}
	B2	196 (26.6%)	184 (37.4%)	
	B3	86 (11.7%)	31 (6.3%)	
Perianal		104 (100.0%)	40 (100.0%)	
Immunomodulator	**azathioprine**	339 (45.7%)	196 (39.4%)	0.009
	mercaptopurine	57 (7.7%)	32 (6.4%)	
	methotrexate	50 (6.7%)	24 (4.8%)	
	tacrolimus	2 (0.3%)	0 (0.0%)	
	none	294 (39.6%)	246 (49.4%)	
Steroids		220 (29.6%)	133 (26.7%)	0.275
Past resectional surgery		170 (22.9%)	122 (24.5%)	0.539
HBI		6.0 (3.0 - 9.0)	5.0 (3.0 - 8.0)	0.224
sPCDAI		25.0 (10.0 - 50.0)		
BMI		23.1 (20.0 - 27.2)	24.6 (21.6 - 28.4)	6.5×10^{-7}
Hemoglobin (g/L)		126.0 (115.8 - 136.0)	132.0 (122.0 - 141.0)	2.2×10^{-10}
White cell count ($\times 10^9$/L)		8.0 (6.2 - 10.3)	7.9 (6.3 - 9.8)	0.324
Platelet count ($\times 10^9$/L)		343.0 (281.0 - 412.8)	312.0 (256.0 - 386.0)	3.4×10^{-6}
Albumin (g/L)		39.0 (34.0 - 42.0)	39.0 (36.0 - 43.0)	0.051
CRP (mg/L)		8.0 (3.0 - 21.0)	6.0 (2.0 - 13.0)	5.1×10^{-5}
Fecal calprotectin (μg/g)		402 (149 - 855)	292 (130 - 603)	1.4×10^{-4}

Table 2.1 Baseline demographic and clinical characteristics of the 1240 individuals from the PANTS cohort. P values were calculated using Fisher's exact or Mann Whitney U tests.

2.2.4 Statistical and genome-wide association analyses

Rates of immunogenicity were estimated using the Kaplan-Meier method. Clinical outcomes and genetic association tests with time to anti-drug antibody development were performed using multivariable Cox proportional hazards regression: sex, drug type (infliximab or adalimumab), immunomodulator use, and the first within-sample principal component, were included as covariates (Table 2.2). Patients who did not develop immunogenicity during the study were censored at the point of last observation. Post-hoc sensitivity analyses were undertaken to test our genetic findings with immunogenicity, firstly, at progressively higher antibody thresholds; secondly, to simulate a free-antibody assay and thirdly, excluding patients with a single anti-drug antibody level >10 AU/ml and subsequent negative anti-drug antibodies <10 AU/ml.

Covariate	**HR**	**95% CIs**		**P**
Drug (0 – infliximab; 1 – adalimumab)	3.27	2.67	4.02	2.82×10^{-34}
Immunomodulators (0 – yes; 1 – no)	2.41	2.02	2.87	1.84×10^{-22}
HLA-DQA1*05 (dominant)	1.90	1.60	2.25	5.88×10^{-13}
Sex (0 – female; 1 – male)	0.93	0.78	1.11	0.41
PC1 (continuous)	0.33	0.03	3.60	0.38

Table 2.2 Covariates used in the final model

The Akaike information criterion (AIC) was used to compare non-nested models to assess if the mode of inheritance was dominant or additive, and to determine whether HLA allele group, specific HLA alleles, or amino acid sequence best-explained the association. The fixed effects Q statistic was used to perform tests of heterogeneity of effect; this test is an extension of Cochran's Q-test and examines whether the observed effect size variability is larger than expected by chance. Interaction tests of the differential effects of drug type (infliximab versus adalimumab and Remicade versus CT-P13) and combination therapy (immunomodulator vs no immunomodulator) conditional on the genotype were performed. Mann-Whitney U tests were used to compare serum levels of anti-drug antibodies at week 54 stratified by anti-TNF drug and immunomodulator use.

2.3 Results

Within the first 12 months, 44% of patients developed anti-drug antibodies (95% CI, 0.41 to 0.48), and 62% of patients did so within 36 months (95% CI, 0.57 to 0.67). After correcting for immunomodulator use, the rate of immunogenicity was greater in patients treated with infliximab (N=742) than adalimumab (N=498) (hazard ratio (HR), 3.21; 95% CI, 2.61 to 3.95; P=1.18×10^{-28}). In a model including drug-type as a covariate, rates of immunogenicity were greater in patients treated with anti-TNF monotherapy (N=544) compared to combination therapy with immunomodulators (N=696), (HR, 2.30; 95% CI, 1.94 to 2.75; P<6.10×10^{-21}).

2.3.1 A locus within the HLA region is associated with time to immunogenicity

The time-to-event analysis identified a genome-wide significant association on chromosome 6 with time to development of immunogenicity, with the most associated SNP, rs2097432 (b38_pos: 6:32622994; HR, 1.70; 95% CI, 1.48 to 1.94; P=4.24×10^{-13}), falling within the major histocompatibility complex (MHC) region (Figures 2.5, 2.3, 2.4). I replicated this association in our independent cohort of 178 patients with IBD (HR, 1.69; 95% CI, 1.26 to 2.28; P=8.80×10^{-4}). A variant on chromosome 11, rs12721026 (b38_pos: 11:116835452; HR, 0.46; 95% CI, 0.33 to 0.63; P=4.76×10^{-8}) also reached genome-wide significance in our discovery analysis, though the association was not replicated in our independent cohort (HR, 0.85; 95% CI, 0.49 to 1.44; P=0.51).

2.3.2 Fine-mapping of the signal in the HLA region

At the HLA allele group level (2-digit resolution), only HLA-DQA1*05 achieved genome-wide significance (HR, 1.90; 95% CI, 1.60 to 2.25; P=5.88×10^{-13}) (Figure 2.6). At the specific allele level (4-digit resolution), no single allele reached genome-wide significance. The two most common HLA-DQA1*05 subtype alleles, HLA-DQA1*05:01 (HR, 1.57; 95% CI, 1.33 to 1.85; P=4.24×10^{-7}) and HLA-DQA1*05:05 (HR, 1.48; 95% CI, 1.24 to 1.78; P=5.54×10^{-5}), had similar effects on time to immunogenicity and a model containing these two 4-digit alleles was virtually indistinguishable from a model including only HLA-DQA1*05 (AIC_{05}=6659.07 versus $AIC_{05:01\&05:05}$=6659.50). I did not identify any

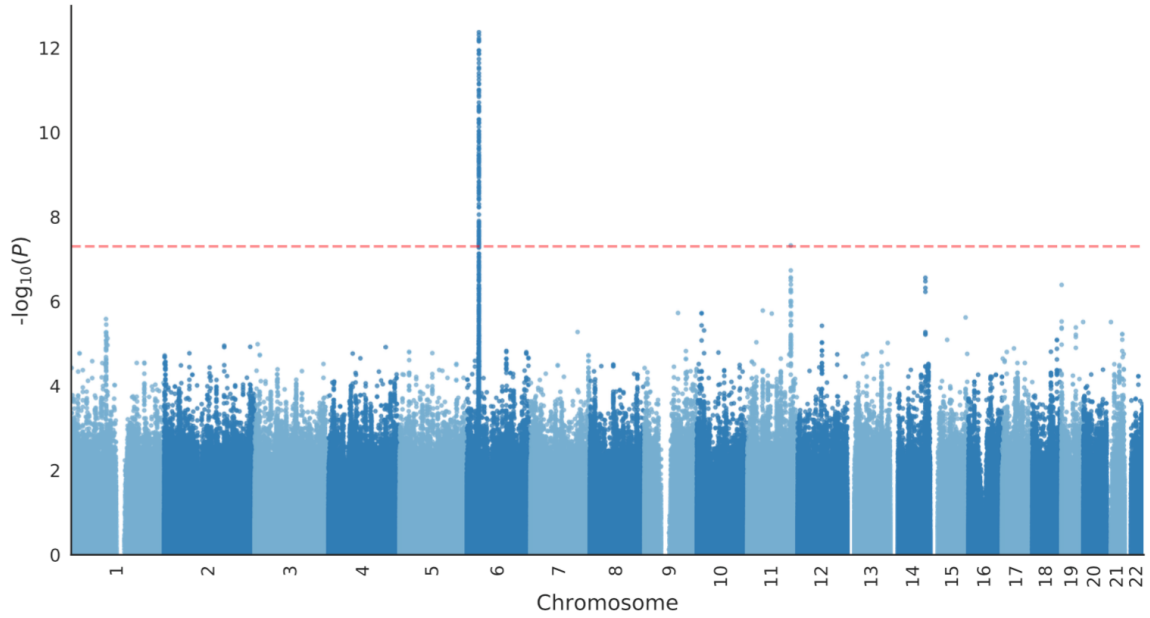


Figure 2.3 Manhattan plot for Cox proportional hazards model analysis of time to immunogenicity

amino acids that better fit the data than HLA-DQA1*05. Our collaborators observed >99% concordance between imputed and sequenced HLA genotypes at HLA-DQA1: amongst the 1,272 overlapping samples, only one sample was discordant between HIBAG (homozygous HLA-DQA1*05) and sequenced HLA (no copies of HLA-DQA1*05).

To formally assess the inheritance pattern of HLA-DQA1*05 mediated immunogenicity, I compared the fit of additive and dominant models and found that the dominant model gave a better fit (AIC_{DOM}=6652.12 vs AIC_{ADD}=6659.07), and stronger association signal for HLA-DQA1*05 (HR, 1.90; 95% CI, 1.60 to 2.25; P=5.88×10^{-13}) (Figure 2.7 and Table 2.3). I also looked for non-additive effects across all other HLA alleles, but the model assuming a dominant effect for HLA-DQA1*05 remained the best fit to the data, based on both the AIC and BIC. The HLA-DQA1*05 association was confirmed in our replication cohort (HR, 2.00; 95% CI, 1.35 to 2.98; P=6.60×10^{-4}), again with a better fit for the dominant model (AIC_{DOM}=942.51 vs AIC_{ADD}=944.81). After conditioning on HLA-DQA1*05 I did not identify any secondary signals of association with time to immunogenicity within the MHC region (Figure 2.8).

Sensitivity analyses showed that the effect size of HLA-DQA1*05 carriage on immunogenicity was similar across subgroups (Figures 2.9a and 2.9b): firstly, the association

Genetic effect included in the model	HR	95% CIs		P	AIC	BIC
AA-107-ile (dominant)	1.9	1.6	2.26	4.00×10^{-13}	6649.45	6675.04
AA-175-lys (dominant)	1.9	1.6	2.26	4.00×10^{-13}	6649.45	6675.04
DQA1*05 (dominant)	1.9	1.6	2.25	5.88×10^{-13}	6652.12	6677.73
AA-107-ile (additive)	1.59	1.4	1.8	1.50×10^{-11}	6656.55	6682.14
AA-175-lys (additive)	1.59	1.4	1.8	1.50×10^{-11}	6656.55	6682.14
DQA1*05 (additive)	1.58	1.39	1.8	1.94×10^{-11}	6659.07	6684.67
DQA1*05:01 and DQA1*05:05 (additive)	1.61	1.36	1.9	1.19×10^{-7}	6659.5	6690.2
DRB1*03 (additive)	1.58	1.34	1.85	2.43×10^{-7}	6661.3	6686.89
DRB1*03 (dominant)	1.67	1.38	2.03	1.48×10^{-7}	6662.46	6688.06
rs2097432 (additive)	1.69	1.48	1.94	4.24×10^{-13}	6674.24	6699.85
rs2097432 (dominant)	1.92	1.61	2.28	4.55×10^{-13}	6674.34	6699.95
DQA1*05:01 (additive)	1.57	1.33	1.85	4.24×10^{-7}	6676.74	6702.33
DQA1*05:01 (dominant)	1.67	1.38	2.03	5.49×10^{-7}	6677.26	6702.85
DQA1*05:05 (dominant)	1.67	1.27	1.9	3.39×10^{-5}	6684.97	6710.56
DQA1*05:05 (additive)	1.48	1.24	1.78	5.54×10^{-5}	6685.84	6711.43
"Null" model with no genotype parameter	–	–	–	–	6725.43	6745.92

Table 2.3 Comparison between different models for the observed effect in the MHC region. Models are sorted by Akaike information criterion (AIC) score (column 6), which is a measure of model fit to the data. All models included immunomodulator status, drug type, sex and PC1 as covariates (see Table 2.2 for details). HRs, CIs and p-values are calculated using SurvivalGWAS_SV and AIC is calculated using R surv package, as described in Methods. The effect assumed for the main genetic effect (additive or dominant) is shown in parentheses in the first column. Hazard ratios with 95% Confidence Intervals for the genetic effect are shown in columns 2–4, and the respective p-values in column 5. The last column contains the BIC score for each model. The ranking of the models is identical whether AIC or BIC is used.

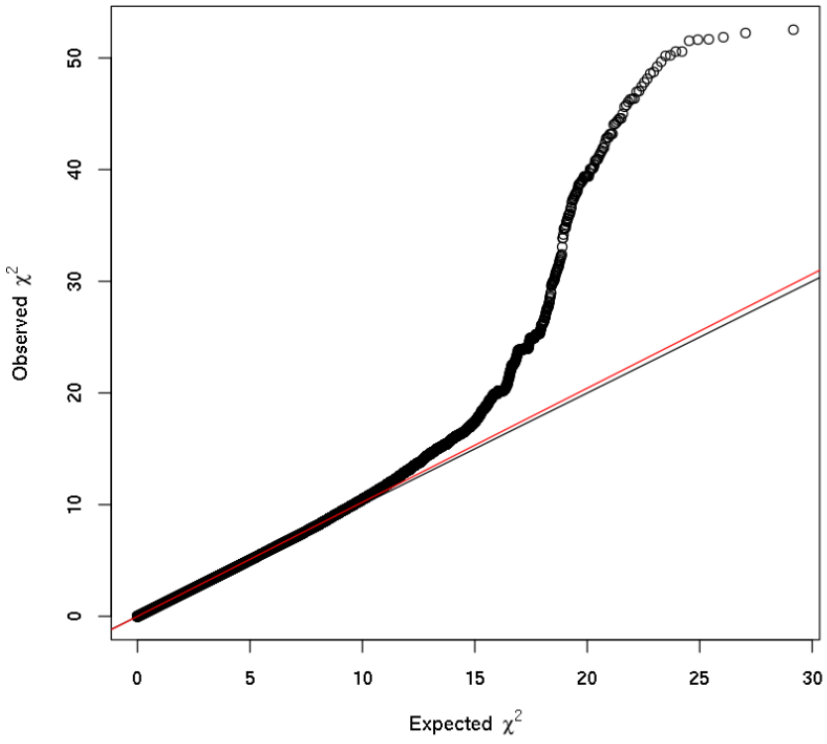


Figure 2.4 Quantile-quantile plot for Cox proportional hazards model analysis of time to immunogenicity. Using a regression model implemented in the GenABEL package in R, I estimated the inflation factor λ to be 1.02 (SE=2×10^{-5}), suggesting a good fit to the uniform distribution. See 2.2 for the covariates used in this model.

remained significant even when the threshold for defining immunogenicity was increased from >10AU/mL to >200 AU/mL. Secondly, when I simulated a drug-sensitive instead of a drug-tolerant assay, where immunogenicity was defined as an anti-drug antibody titer $\geq$10 AU/ml without detectable drug (HR, 1.57; 95% CI, 1.23–2.01; P = 3.66×10^{-4}). Thirdly, when I removed patients with a one-off transient anti-drug antibody level $\geq$10 AU/ml (HR, 1.94; 95% CI, 1.62-2.32; P=8.46×10^{-13}).

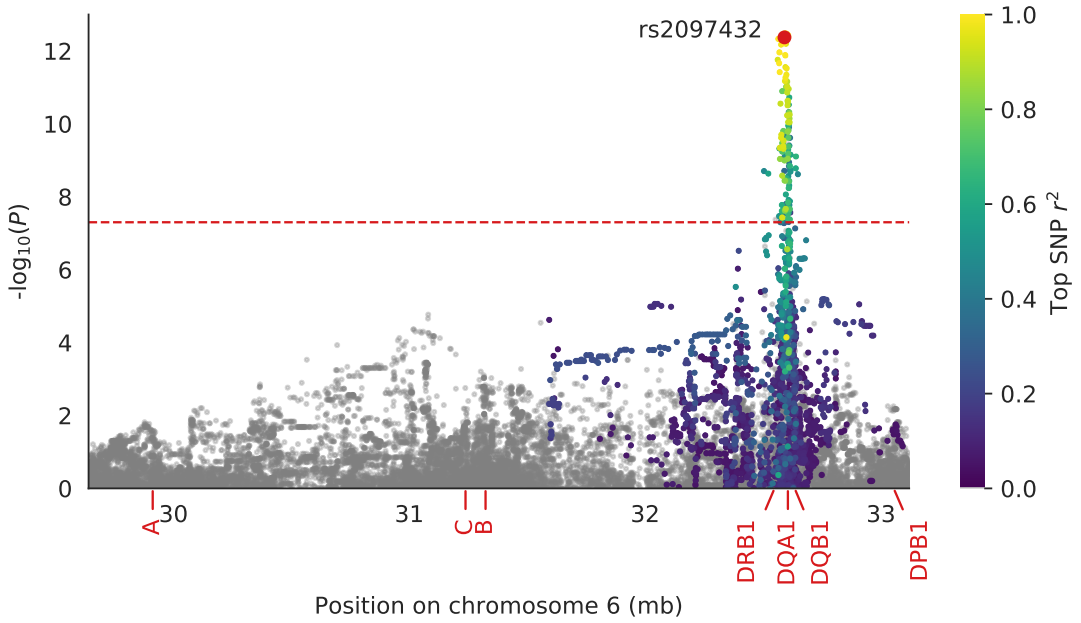


Figure 2.5 Regional plot of the association results with the MHC region on chromosome 6. Midpoint positions of the HLA alleles across the MHC region are shown in red on the x-axis. SNPs that passed the genome-wide significance threshold (P=5×10^{-8}) are shown above the red horizontal dashed line with the most significant SNP in red. SNPs correlated with the lead SNP (r^2>0.05) are colour-coded from purple to yellow. Pairwise genotype correlation (r^2) between SNPs was calculated using genotype data from the non-Finnish European population of the 1000 Genomes Project.

2.3.3 The effect of HLA-DQA1*05 across drug and treatment regimes

While immunogenicity rates were lower with adalimumab-treated compared to infliximab-treated patients, I did not detect a significant difference in the effect of HLA-DQA1*05 on the immunogenicity rate for these two drugs (HR, 1.89; 95% CI, 1.32–2.70 in adalimumab-, HR, 1.92; 95% CI, 1.57–2.33 in infliximab-treated patients; P_{het}=0.91) (Fig. 2.10). I also found no significant evidence for heterogeneity of effect of HLA-DQA1*05 on immunogenicity between patients treated with the infliximab originator, Remicade, and its biosimilar CT-P13 (P_{het}=0.23) (Figure 2.11). Likewise, I did not detect any significant heterogeneity of effect of HLA-DQA1*05 carriage on immunogenicity for individuals on monotherapy (HR, 1.75; 95% CI, 1.37–2.22) versus combination therapy (HR, 2.01; 95% CI, 1.57–2.58) with immunomodulators (P_{het}=0.14). In addition, I did not identify any significant

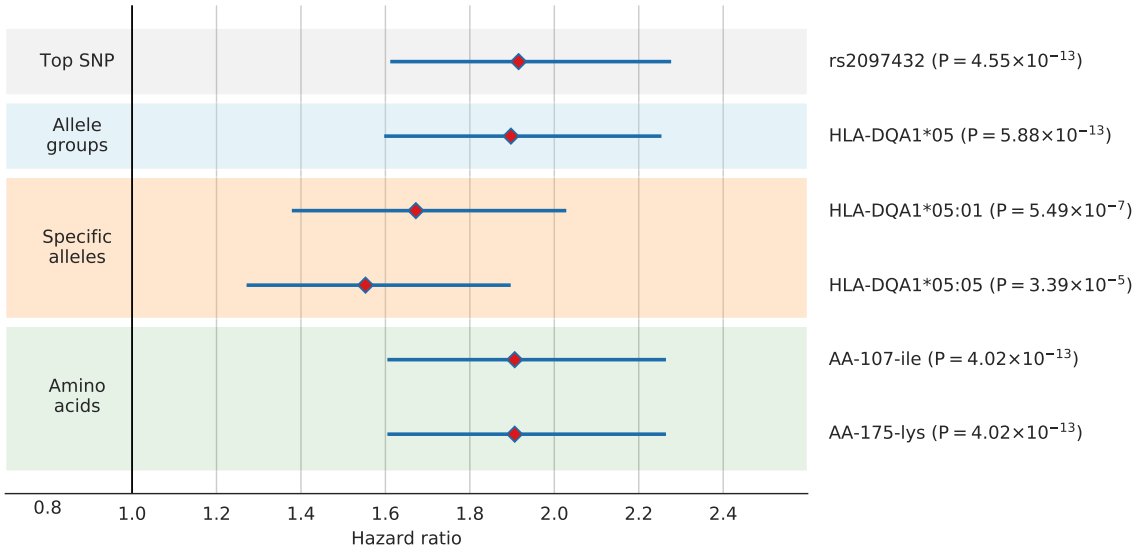


Figure 2.6 Effect sizes of the most strongly associated SNP, HLA alleles, and amino acids of time to immunogenicity. Blue lines represent 95% CIs. Association test P-values are shown in parentheses.

interactions between HLA-DQA1*05 and the clinical covariates (drug type: P=0.83; mono- vs combination therapy: P=0.71; Remicade vs CT-P13: P=0.59).

The highest rates of immunogenicity, 92% at 1 year, were observed in patients treated with infliximab monotherapy who carried HLA-DQA1*05 (Figure 2.13a). Conversely, the lowest rates of immunogenicity, 10% at 1 year, were observed in patients treated with adalimumab combination therapy who did not carry HLA-DQA1*05 (Figure 2.13b). Our final model, which includes HLA-DQA1*05 status, sex, drug, and immunomodulator usage, explained 18% of the variance in immunogenicity to anti-TNF in our cohort.

Having demonstrated that HLA-DQA1*05 was associated with time to immunogenicity we sought associations with anti-drug antibody titers after 1 year of treatment and subsequent non-persistence on drug. Carriage of HLA-DQA1*05 was associated with higher maximal anti-drug antibody titers ($P_{\text{infliximab}} = 8\times10^{-10}$; $P_{\text{adalimumab}} = 0.002$). I observed lower drug persistence rates to year 3 in patients treated with an anti-TNF drug without an immunomodulator (Figure 2.13); the optimal model ($AIC_{interaction} = 5937.19$ versus $AIC_{additive} = 5940.16$) here used the interaction between immunomodulator use and HLA-DQA1*05 (DQA1*05:

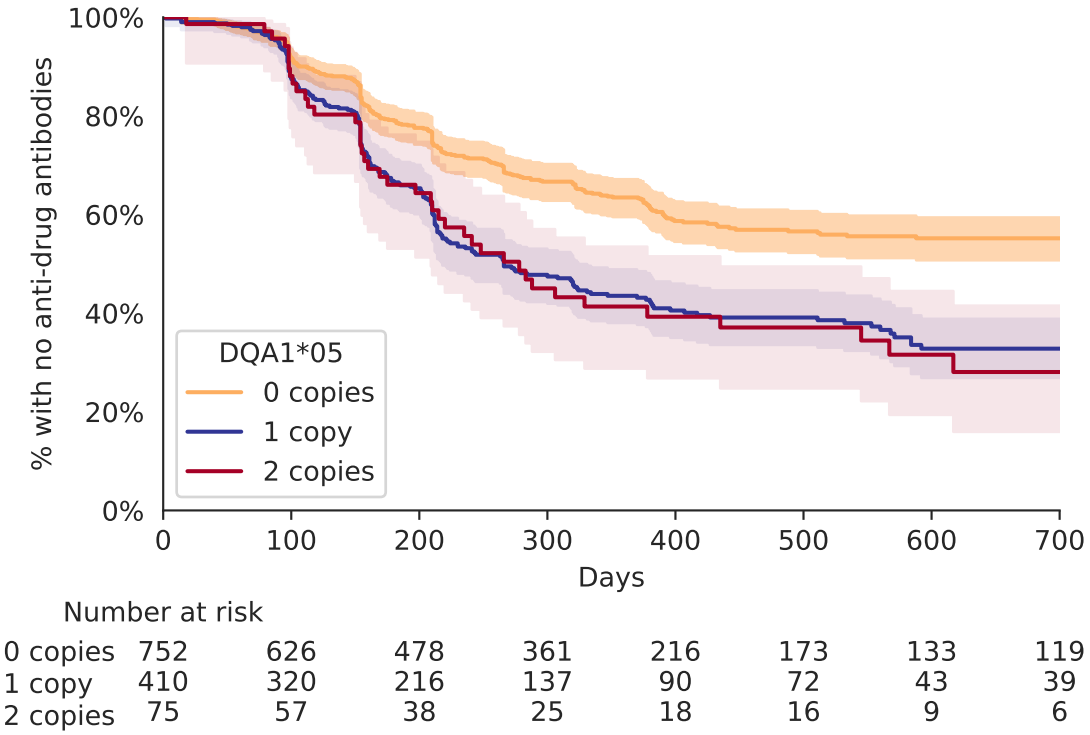


Figure 2.7 Kaplan–Meier estimator showing the rate of anti-drug antibody development, stratified by the number of HLA-DQA1*05 alleles carried. Orange, blue and red indicate 0, 1 and 2 copies of DQA1*05 allele, respectively. Carriers of one or two copies of the allele have a similar rate of immunogenicity development, and a dominant model is a better fit for the data than an additive model (AIC_{DOM}=6652.12 vs AIC_{ADD}=6659.07). X-axis truncated at 700 days, due to the low number of observations for longer time periods.

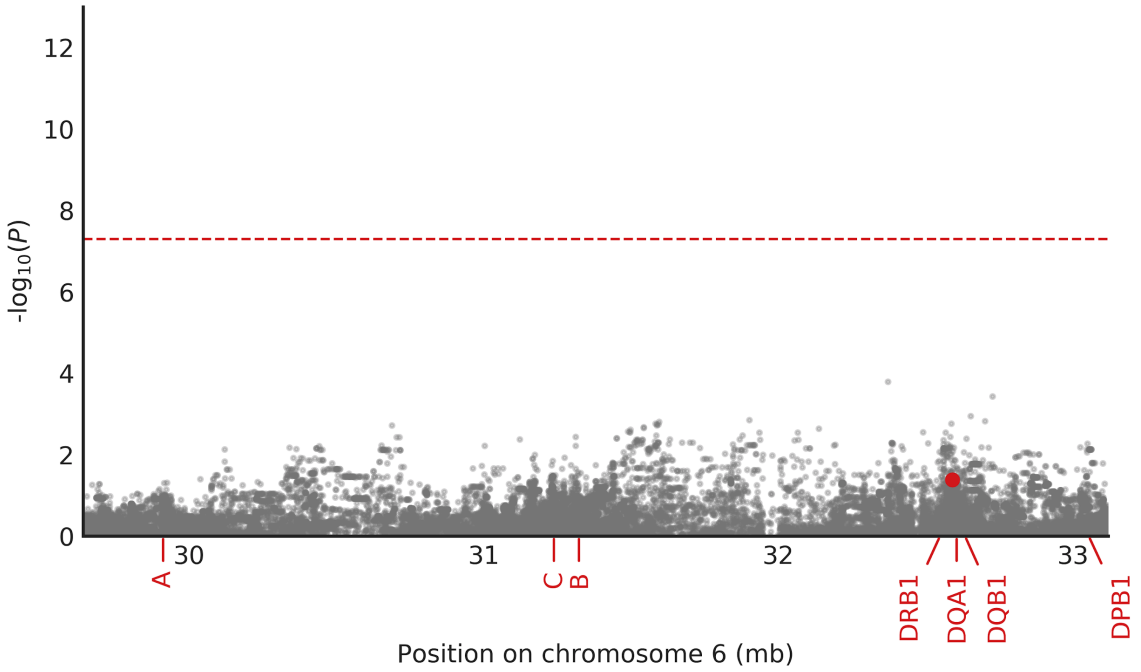
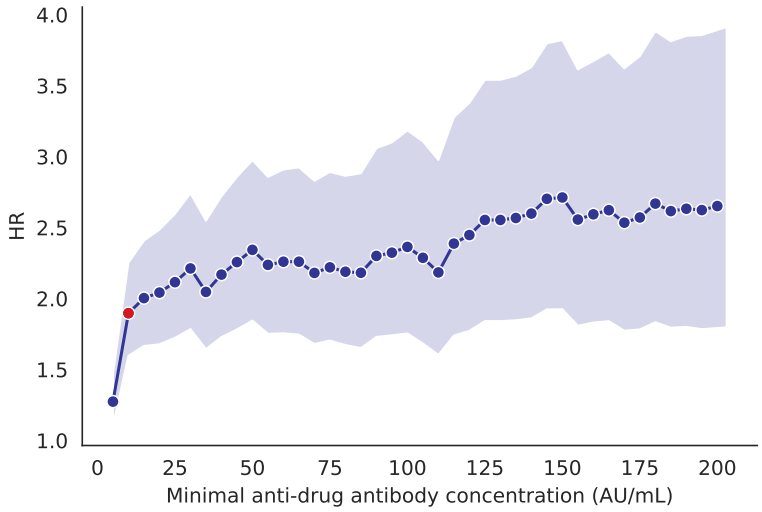


Figure 2.8 Residual association signal in the MHC region, after conditioning on HLA-DQA1*05. Midpoint positions of the HLA alleles across the MHC region are shown in red on the x-axis. The SNP most strongly associated with time to immunogenicity (rs2097432) during the initial analysis is marked with a red dot. No other SNPs passed the genome-wide significance threshold (red dashed line), suggesting that HLA-DQA1*05 explains the chromosome 6 signal (Figure 2.5).

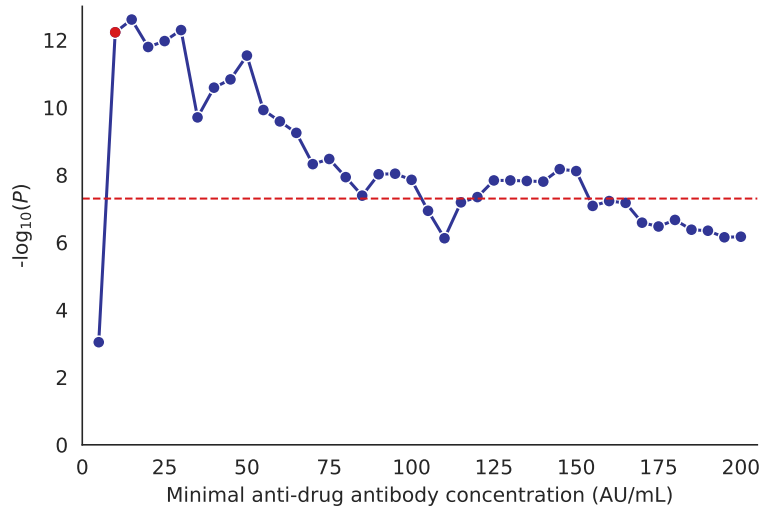
HR, 1.40; 95% CI 1.08–1.80; P=0.011, immunomodulator use: HR, 0.74; 95% CI, 0.58–0.94, P=0.014, interaction between DQA1*05 and immunomodulator use: HR, 0.65; 95% CI, 0.45-0.95; P=0.026).

2.4 Discussion

Immunogenicity to biologic therapies is a major concern for patients, regulatory authorities, and the pharmaceutical industry. I report the first genome-wide significant association with immunogenicity to anti-TNF therapy using the largest prospective cohort study of infliximab and adalimumab in Crohn's disease. I have demonstrated that carriage of one or more HLA-DQA1*05 alleles confers an almost two-fold risk of immunogenicity to anti-TNF therapy,



(a)



(b)

Figure 2.9 Sensitivity analysis of the (a) effect size and (b) significance of HLA-DQA1*05 association and time to immunogenicity. In the primary analyses, immunogenicity was defined as an anti-drug antibody concentration $\geq$10 AU/mL, irrespective of drug concentration (red dot). I repeated the time to immunogenicity analysis varying this definition from $\geq$5 to $\geq$200 AU/mL.

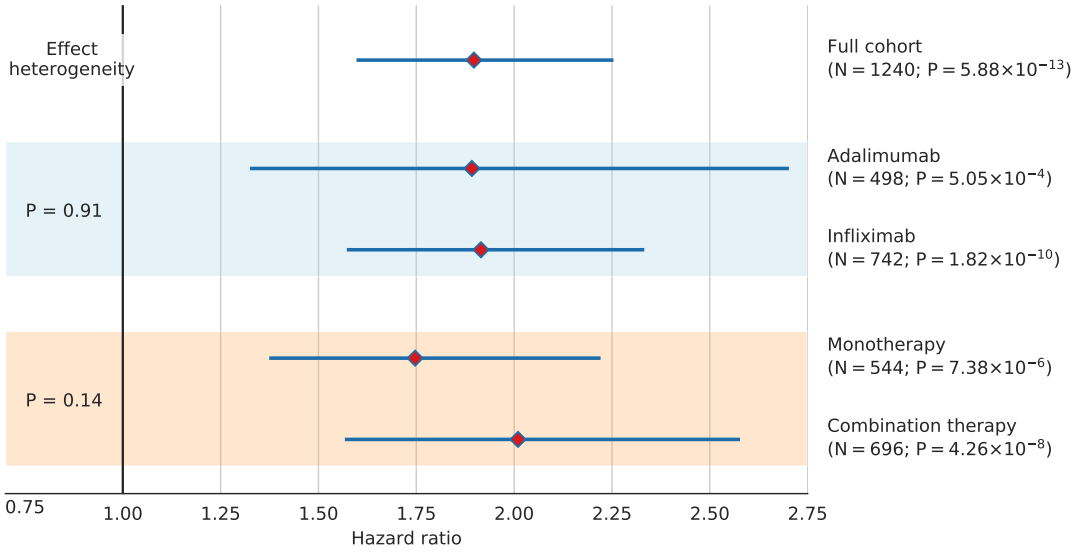


Figure 2.10 HLA-DQA1*05 has a consistent effect on immunogenicity in different patient subgroups. Blue lines represent 95% CIs. Association test P-values are shown in parentheses. The proportional hazard association analysis was repeated, separating the full cohort into subgroups by drug and therapy type. Estimated hazard ratios and standard errors between the pairings were compared using a heterogeneity of effects test (P>0.05), suggesting that the effect of HLA-DQA1*05 on immunogenicity is not affected by these clinical covariates.

irrespective of concomitant immunomodulator use or drug type (infliximab [Remicade or CT-P13], or adalimumab). Fine-mapping and confirmatory sequencing of the HLA identified that the specific alleles HLA-DQA1*05:01 and HLA-DQA1*05:05 mediated most of this risk. Carriage of HLA-DQA1*05 was associated with higher anti-drug antibody levels and lower drug persistence rates, although further studies are needed to more accurately quantify the relationship between HLA-DQA1*05 and drug persistence. An overview of the further genetic studies that can be carried out to better to elucidate the genetics of the anti-TNF immunogenicity and response is provided in Section 5.3 of the Discussion.

Arguably, based on these data and those presented in the PANTS clinical paper [96], all patients treated with an anti-TNF should ideally be prescribed an immunomodulator to lower the risk of immunogenicity [96]. We hypothesise that for patients who carry HLA-DQA1*05 in whom immunomodulators are contraindicated or not tolerated, clinicians might advise against the use of anti-TNF drugs, particularly infliximab. In contrast, patients who do

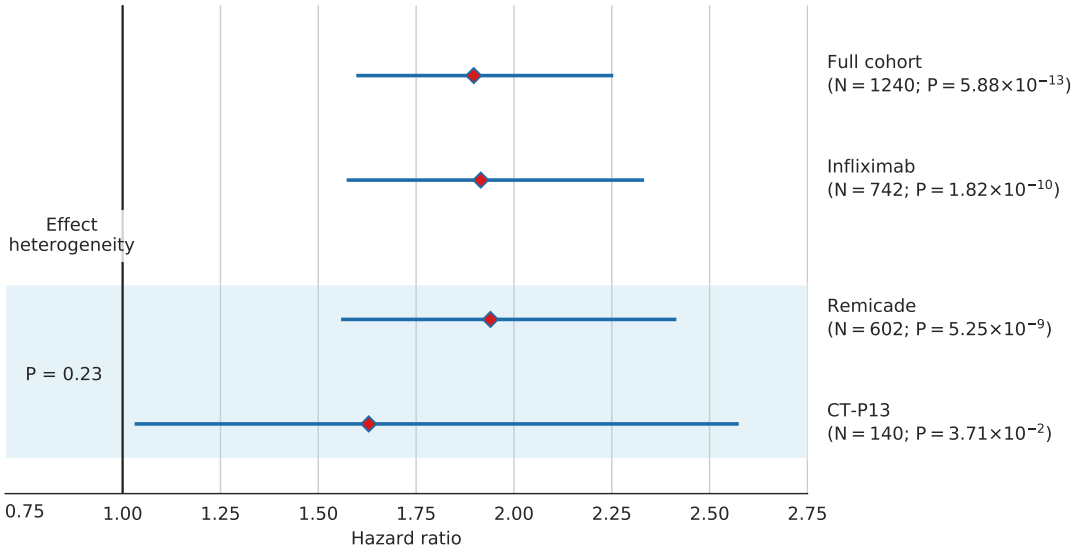
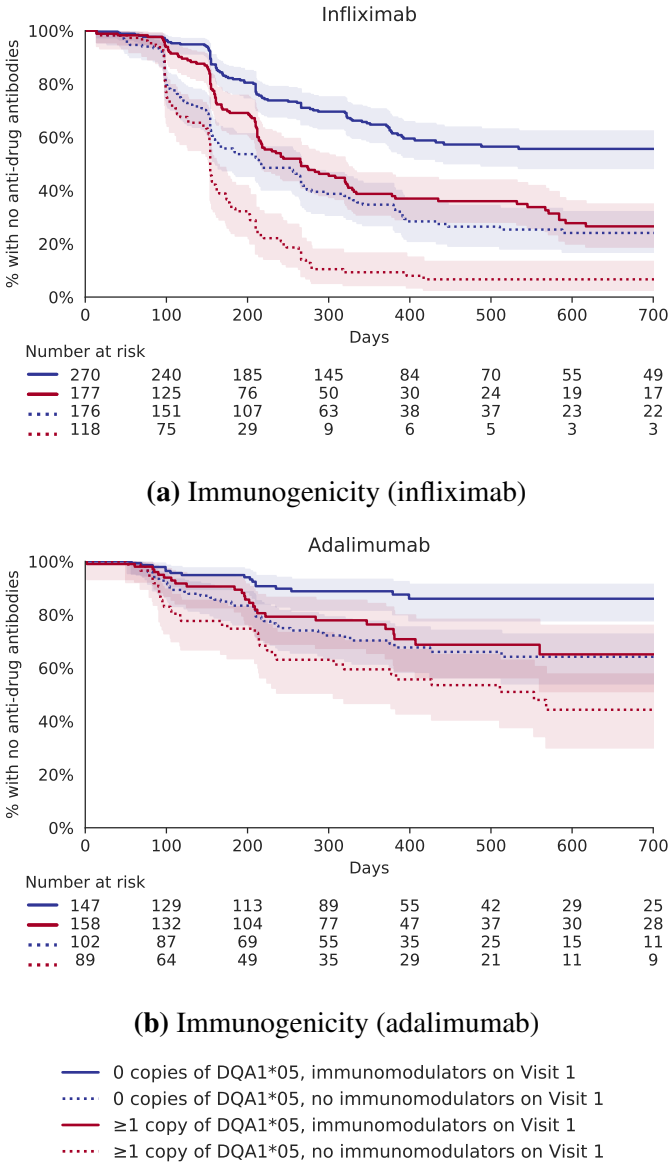


Figure 2.11 HLA-DQA1*05 has a consistent effect on immunogenicity between patients treated with the infliximab originator, Remicade, and its biosimilar CT-P13.

not carry HLA-DQA1*05 might be given the choice between adalimumab or infliximab combination therapy. Patients without the risk allele and a history of adverse drug reactions to thiopurines and/or methotrexate or who are at high risk of opportunistic infections might be spared the additional risks of combination therapy and treated with adalimumab monotherapy. A randomised controlled biomarker trial is required to explore these hypotheses and confirm whether HLA-DQA1*05 testing may help direct treatment choices in order to improve clinical outcomes.

The shared genetic association between HLA-DQA1*05 and immunogenicity to infliximab and adalimumab may explain the widely reported diminishing returns of switching between anti-TNF therapies at the time of loss of response [141, 161]. If the immunogenic effect of HLA-DQA1*05 extends to other therapeutic antibodies, then subjects who carry the variant may be candidates for non-antibody modality therapies such as small molecule drugs.

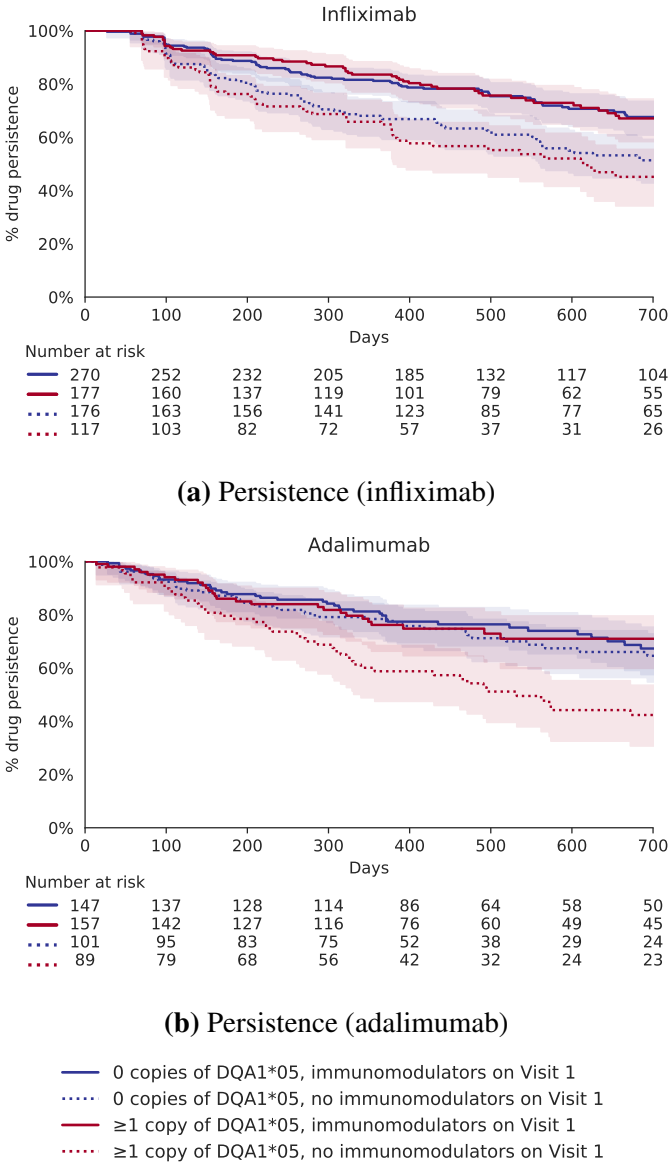
Allelic variation in the *HLA-DQA1* gene has been linked to aberrant adaptive immune responses. The HLA class II gene *HLA-DQA1* is expressed by antigen presenting cells and encodes the alpha chain of the *HLA-DQ* heterodimer that forms part of the antigen binding site where epitopes are presented to T-helper cells. Relevant to immunogenicity,



(a) Immunogenicity (infliximab)

(b) Immunogenicity (adalimumab)

Figure 2.12 Kaplan–Meier estimator showing the rate of anti-drug antibody development (A and B), stratified by carriage of HLA-DQA1*05 alleles and treatment regime. Dotted lines indicate patients undergoing anti-TNF monotherapy; solid lines indicate combination therapy with immunomodulators. Red indicates carriers of the HLA-DQA1*05 allele (1 or 2 copies); blue indicates non-carriers. For both drugs and treatment regimes, immunogenicity is higher for HLA-DQA1*05 carriers. The X-axis was truncated at 700 days due to the low number of observations.



(a) Persistence (infliximab)

(b) Persistence (adalimumab)

Figure 2.13 Kaplan–Meier estimator showing the rate of drug persistence (A and B), stratified by carriage of HLA-DQA1*05 alleles and treatment regime. Dotted lines indicate patients undergoing anti-TNF monotherapy; solid lines indicate combination therapy with immunomodulators. Red indicates carriers of the HLA-DQA1*05 allele (1 or 2 copies); blue indicates non-carriers. For both drugs and treatment regimes, drug persistence rates are higher for HLA-DQA1*05 carriers. The X-axis was truncated at 700 days due to the low number of observations.

carriage of HLA-DQA1*05 has been associated with coeliac disease and type 1 diabetes and protection against rheumatoid arthritis and pulmonary tuberculosis [122, 42, 182, 137]. Several hypotheses have been proposed, but exactly how specific HLA alleles contribute to disease pathogenesis or, in this case, increased immunogenicity, remains unknown.

HLA-DQ1A*05 may serve as a useful biomarker of immunogenicity risk and may impact how the next-generation of anti-TNF drugs are designed to minimise HLA-DQA1*05 mediated immunogenicity. Previous studies have shown that it is possible to map and eliminate potential immunogenic T cell epitopes with the aim of producing safer and more durable biologic drugs [48, 163]. However, caution needs to be exercised to ensure protein sequence modifications designed to reduce the risk of immunogenicity to patients carrying HLA-DQA1*05 do not put a different group of patients at risk.

Multiple assays are available to detect anti-drug antibodies and there is no universally accepted, validated threshold to diagnose immunogenicity. Our collaborators deliberately chose a total, or drug tolerant assay, that permits the measurement of anti-drug antibodies in the presence of drug, in order to minimise the number of false negative patients assigned to the control group. They then validated the manufacturer's positivity threshold in independent experiments in 500 drug-naïve controls and confirmed that the recommend cut-off of 10 AU/mL corresponds to the 99th percentile of the antidrug antibody titer distribution. In support of this threshold, our collaborators have recently demonstrated that even modestly elevated anti-drug antibodies levels (10–30 Au/ml) at weeks 14 and 54 of treatment are associated with lower drug levels at these time points, and non-remission at week 54 [96]. In addition, sensitivity analyses confirmed that the association and effect size between HLA-DQA1*05 and immunogenicity remained at progressively higher diagnostic thresholds for immunogenicity, when we simulated a free-assay, and when we removed patients with transient antibodies. Finally, HLA-DQA1*05 was associated with the quantitative trait of maximal anti-drug antibody titer.

Two important limitations of this study should be acknowledged. Firstly, we may have underestimated the contribution of HLA-DQA1*05 to immunogenicity because of the short duration of follow-up in patients who did not continue in the study beyond the first year. Secondly, because the study schedule was designed to minimise patients' inconvenience, there were fewer assessments for those treated with adalimumab than infliximab. As a result, we might have underestimated rates of immunogenicity amongst adalimumab-treated patients.

The genome-wide association study was limited to patients with Crohn's disease of European descent. Given that HLA-DQA1*05 is not associated with IBD risk [73] the percentage of carriers among our patients (39%) was similar to that reported in an independent British population cohort (38%) [71]. As such, I hypothesise that HLA-DQA1*05 will make a similar contribution to anti-TNF immunogenicity in other patient populations where the allele is not associated to disease susceptibility (e.g. ankylosing spondylitis). Due to the wide variation in the frequency of HLA-DQA1*05 across ethnic groups [71], further studies are required to assess the contribution of HLA-DQA1*05 to immunogenicity across populations. Whether HLA-DQA1*05 is also associated with immunogenicity to other biologic drugs also needs to be determined.

In this chapter, I report the first genome-wide significant association with immunogenicity to biologic drugs. Carriage of HLA-DQA1*05 almost doubles the rate of anti-TNF anti-drug antibody development, independent of immunomodulator use, for both infliximab and adalimumab. To minimise the risk of immunogenicity, pre-treatment genetic testing for HLA-DQA1*05 may help personalise the choice of anti-TNF and the need for combination therapy with an immunomodulator.