EXPLORING SEX-DIFFERENTIATED GENETIC EFFECTS IN JIA AND RESTRUCTURING T1DGC REFERENCE PANEL FOR COMPATIBILITY WITH HLA-TAPAS / MINIMAC4 PIPELINES

A dissertation submitted to The University of Manchester for the degree of Master of Science in the Faculty of Biology, Medicine and Health

2025

STUDENT ID: 11353031

School of Biological Sciences

CONTENTS

WORD COUNT: ~5700 words

ABSTRACT

The Human Leucocyte Antigen (HLA) is an important genetic factor in the development of autoimmune diseases such as Juvenile Idiopathic Arthritis (JIA). Its highly polymorphic nature, combined with dense linkage disequilibrium makes HLA research very challenging. This project contains two independent studies with distinct objectives focused on HLA research.

The first study investigated the sex-differentiated genetic effects in JIA. While there is substantial evidence of a sex difference in the clinical presentation and prevalence of JIA, the genetic basis for this difference remains poorly investigated. A sex-stratified analysis using logistic regression, followed by meta-analysis with GWAMA was conducted to assess sex-differentiated effect size differences (P_{diff}) and effect size heterogeneity (P_{het}), focusing on the extended HLA region. The analysis showed sex-differentiated effects in the HLA region, particularly involving HLA alleles, amino acid polymorphisms, and intragenic SNPs. The strongest signal was from an intragenic SNP in the *HLA-B* locus, which showed risk effect in males but protective effect in females. Furthermore, an amino acid polymorphism at the *HLA-DRB1* locus showed risk effect in both sexes; however, the effect was significantly larger in females than in males.

The second study focused on a compatibility issue that prevented using legacy-formatted reference panels with modern imputation tools. The Type 1 Diabetes Genetics Consortium (T1DGC) reference panel, while offering high quality imputation results and a strong representation of European ancestry, is not compatible with newer imputation tools, such as HLA-TAPAS and Minimac4. A methodology was developed to reformat the panel by standardizing variant encodings and converting them into a variant call format (VCF) that is compatible with modern imputation pipelines. The reformatted reference panel was validated for large dataset imputation and performed well.

DECLARATION

I declare that this project report is my original work unless referenced clearly to the contrary, and that no portion of the work referred to in the report has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

INTELLECTUAL PROPERTY STATEMENT

I. The author of this dissertation (including any appendices and/or schedules to this dissertation owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this dissertation, either in full or in extracts and whether in hard or electronic copy may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered into. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the dissertation, for example graphs and tables (“Reproductions”), which may be described in this dissertation, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialization of this dissertation, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Dissertation restriction declarations deposited in the University Library, and The University Library’s regulations.

ACKNOWLEDGEMENTS

I want to express my deepest gratitude to those who have supported me while completing this dissertation. I would like to thank my supervisor, Dr. John Bowes, for his guidance. His expertise and insights have greatly shaped the direction of my research. I am thankful to the Research IT team for using the Computational Shared Facility at the University of Manchester.

I am incredibly grateful to my family for their unwavering support, love, and understanding. Their belief in me made this possible.

Thank you all.

INTRODUCTION

This project report includes two related topics focusing on different research objectives: 1. To identify HLA loci with sex-differentiated effects in JIA 2. To reformat the T1DGC reference panel for use in HLA-TAPAS and Minimac4 imputation platforms. Accordingly, this report is structured.

JUVENILE IDIOPATHIC ARTHRITIS

Juvenile Idiopathic Arthritis (JIA) is an autoimmune disease characterized by synovial inflammation and joint swelling. It affects children under 16 years of age, and the extent and nature of inflammation range from mild oligoarticular involvement to systemic disease with multi-organ involvement. The International League of Association for Rheumatology (ILAR) has classified JIA into seven subtypes: oligoarthritis; rheumatoid factor (RF)-negative polyarthritis; RF-positive polyarthritis; juvenile psoriatic arthritis; enthesitis-related arthritis childhood spondyloarthropathy (ERA); systemic arthritis and undifferentiated arthritis (Petty et al. 2004). According to the National Rheumatoid Arthritis Society, approximately 12 000 children in the UK suffer from JIA, with an annual incidence of around 5.6 per 100 000 population (Costello et al. 2022). Most subtypes of JIA, including oligoarthritis and polyarthritis, are common in girls except ERA, which predominantly affects boys (Gmuca et al. 2017, Cattalini et al. 2019).

GENETIC SUSCEPTIBILITY OF JIA

It is well established that genetics play an important role in the development of JIA. Data from monozygotic twin studies show a 25-40% risk of inheritance between affected twins which is higher compared to 1% risk in the general unrelated population (Prahalad 2006). Furthermore, studies have shown that the siblings of JIA patients are 30 times more likely to develop the disease often at the same age as their affected sibling, further establishing that genetics drive disease onset rather than being solely driven by environmental factors (Glass and Giannini 1999). Recently, several genome-wide association studies (GWAS) have implicated multiple genetic variants, including single nucleotide polymorphisms (SNPs) within the major histocompatibility complex (MHC) region and several non-MHC genes to be associated with JIA (Prahalad and Glass 2008).

KEY GENETIC LOCI IN AUTOIMMUNE ARTHRITIS

Several genetic loci, including Human Leucocyte Antigen (HLA) and non-HLA loci, are implicated in autoimmune arthritis contributing through complex interactions with epigenetic and environmental factors (Kurkó et al. 2013).

Studies on adult-onset rheumatoid arthritis (RA) show that the *HLA-DRB1* gene, a class II gene in the MHC region, is the strongest risk factor, contributing nearly 30% to disease development (Deighton et al. 1989). Several *HLA-DRB1* alleles have been identified to be associated with RA risk. Certain alleles encode a ‘shared epitope’ (SE), a conserved sequence domain between positions 70-74 of the *HLA-DRB1* amino acid chain, which presents citrullinated peptides as antigen to CD4+ T cells initiating an autoimmune response. Activation of the immune system produces auto-

antibodies, including RF and anti-citrullinated protein antibodies (ACPA), which form the immune complexes and get deposited in the synovium (Sharma, Leung and Viatte 2024). Studies have shown that these SE alleles have shown considerable ethnic and racial variations. Predominant SE-positive alleles include *HLA-DRB1*04:01, DRB1*04:04, DRB1*01:01, DRB1*10:01, DRB1*04:05,* and *DRB1*14:02* across different populations (van Drongelen and Holoshitz 2017).

HLA GENES ASSOCIATED WITH JIA

In JIA, different subtypes are associated with specific HLA genes and alleles (Prahalad and Glass 2008). Among the different subtypes oligoarthritis, polyarticular RF-positive, polyarticular RF-negative, systemic and juvenile psoriatic forms share the strongest association with *HLA-DRB1* alleles (*HLA-DRB1*11, DRB1*04, DRB1*01, DRB1*08, DRB1*13, DRB1*03*) whereas *HLA*-B27 was shown to cause ERA. In addition to these, other HLA alleles such as *HLA-A*02:06, DQA1*03, DQA1*05, DQB1*04, DPB1*01, DPB1*03, DQB1*03, DQA1*04:01* and *DQB1*04:02* have also been implicated in various subtypes of JIA (La Bella et al. 2023, Yanagimachi et al. 2011). Specific *HLA-DRB1* alleles cause JIA by altering the peptide binding and dysregulating the immune system, triggering an autoimmune reaction like that of adult RA (La Bella et al. 2023). Interestingly, the polyarticular RF-positive JIA not only resembles the adult RA clinically but also shares an association with *HLA-DRB1*04:01* and *DRB1*04* and encodes the SE, causing similar pathogenesis (Hisa et al. 2017). In contrast, the oligoarticular, RF-negative subtypes and enthesitis-related forms are associated with distinct HLA-linked immune-pathogenic mechanisms that are not fully understood yet. On the other hand, based on the role of *HLA-B27* in spondyloarthropathies, scientists believe that in ERA, *HLA-B27* may involve CD8+ T-cell and IL-23/IL-17 activation, a mechanism similar to spondyloarthropathy (McMichael and Bowness 2002). Several other mechanisms have also been proposed, such as molecular mimicry, gut microbiome, and alternative antigen hypothesis, but are still in their early stages, and further research is required to understand them fully (Verwoerd et al. 2016, Rojas et al. 2018).

NON-HLA GENES ASSOCIATED WITH JIA

For several years, research on autoimmune arthritis primarily focused on HLA genes, particularly *HLA-DRB1*, due to its strong genetic association with RA. However, with the advancement of high-throughput sequencing technologies and GWAS, several non-HLA loci have also been identified as contributing to JIA susceptibility. In independent studies, at least 15 key loci have been consistently associated with JIA. Although these genes may not directly cause the disease, they regulate critical immune pathways, including T-cell activation and regulation (*PTPN22, PTPN2, IL2RA, IL2 STAT4, SH2B3 and CD80*), cytokine signaling and inflammatory regulation (TNFA, TNFAIP3, MIS), antigen presentation and immune modulation (*CLEC16A, VTCN1, SLC11A1*), tissue remodeling and joint inflammation (*WISP3, COG6, ZFP36L1*) (Nikopensius et al. 2021, Prahalad and Glass 2008, Hinks et al. 2013, McIntosh et al. 2017). Although several non-HLA-genetic loci were tested, only a few were significantly associated with JIA. Most of these associations appear population-

specific and have not been replicated across diverse study populations (Nikopensius et al. 2021). Furthermore, the diverse clinical subtypes of JIA posed problems, where a locus strongly associated with one subtype might have no role in another, and signals may be diluted when all disease subtypes were treated as a single entity. This genetic diversity has also made it challenging to assemble a large, well-defined cohort to generate robust genetic associations.

SEX-DIFFERENTIATED GENETIC EFFECTS IN JIA

Sex-differentiated genetic effects occur when a gene affects disease expression differently in males and females. The gender disparity among the clinical subtypes in JIA is in itself a strong indicator that sex plays a very significant role in the disease mechanism. The female dominant subtypes of JIA, such as oligoarticular, and polyarticular subtypes, are almost consistently associated with *HLA-DRB1* alleles, while the male dominant subtype, ERA, is associated with *HLA-B27* (La Bella et al. 2023). Like JIA, even in the adult RA, which shows a female-dominant disease pattern, there is a strong association with *HLA-DRB1*, suggesting it may exert its effects differently in females than in males. Supporting this, in one study, specific *HLA-DRB1* alleles were shown to exert different effects in males and females using mathematical models and penetrance studies (Meyer et al. 1996). While the shared epitope hypothesis partially explains the role of *HLA-DRB1* mediated disease mechanisms in autoimmune arthritis, such as RA, and specific subtypes of JIA, it does not account for all risk alleles, particularly in JIA. Some DRB1 alleles such as *DRB1*01:03, DRB1*13:01, DRB1*13:02 or DRB1*04:02* were also shown to exert a protective effect contrary to the recognized risk effect in RA (van Drongelen and Holoshitz 2017). Whether these differential effects also follow a sex-specific pattern remains to be investigated.

Non-HLA genes also exhibit sex-specific genetic effects in JIA. Genes such as *PTPN22*, which is involved in T and B cell signaling and proteasome-related genes such as *PSMA6, PSMC6* and *PSMA3* are significantly associated with females (Chiaroni-Clarke et al. 2015, Sjakste et al. 2014). Specific neutrophil-related genes and certain genes involved in cytokine signaling and lymphocyte activation pathways were identified to be associated with females more than males (Dodd and Menon 2022, Prada-Medina, Peron and Nakaya 2020). In contrast, a gene linked to CD4+ T cells has been associated explicitly with males (Chiaroni-Clarke, Munro and Ellis 2016).

In addition to direct effects of autosomal genes that affects both sexes, sex chromosomes, sex-linked genes, and sex hormones inherently contribute to sex-specific genetic effects. X chromosomes carry several immune regulatory genes (Bianchi et al. 2012). While females carry two X chromosomes, there is random inactivation of one X chromosome to prevent a doubling of gene dosage of X-linked genes (Bianchi et al. 2012). However, skewed X-chromosome inactivation (XCI) may result in the dysregulated activation of the X-linked immune regulatory genes, which may trigger an autoimmune response (Uz et al. 2009, Azzouz et al. 2011). In addition, specific X-linked genes such as *TLR, FOXP3* and *CD40L* can escape XCI, leading to excess immune activity predisposing females to autoimmune arthritis (Mousavi, Mahmoudi and Ghotloo 2020, Selmi et al.

2012). In addition, the female sex hormone estrogen modifies immune regulation through pro-inflammatory cytokines such as TNF-α and IL6, thereby increasing disease risk in females (Yang et al. 2023).

While the X-linked genes and female sex hormones cause immune activation and greater severity in females, Y-linked genes and the male hormone testosterone offer protective and immunosuppressive effects (Baillargeon et al. 2016). Mosaic loss of the Y chromosome in the elderly was shown to increase the severity of the disease in RA (Uchiyama et al. 2025, Serhal et al. 2020). Additionally, there is emerging evidence of Y-linked genes such as *KDM5D* and *UTY* modulating the immune system (Bhattacharya, Sadhukhan and Saraswathy 2024). Although the exact mechanism through which testosterone exerts this protective effect is still unclear, it is widely believed that an immune regulatory and inflammatory response may play a role (Mohamad et al. 2019). Further, the lack of testosterone in children with hypogonadotropic hypogonadism were shown to suffer from a more severe form of JIA (Diack et al.). However, it is important to note that most cases of JIA onset occur before puberty when the sex hormone levels are low, comparably similar between boys and girls.

To summarize, there is increasing evidence that sex, as a genetic factor, plays an important role in role in the pathogenesis of JIA. Despite this growing knowledge, genome-wide studies do generally not test for statistical gender-based heterogeneity. Sex-differentiated effects in genes responsible for disease progression may remain potentially obscured due to sex-aggregated analysis, which may mask the complex interplay between genetics and sex factors. Investigating sex-specific genetic effects in JIA is crucial for identifying risk factors, understanding disease pathogenesis and developing diagnostic and therapeutic strategies.

IMPUTATION IN HLA RESEARCH

HUMAN LEUCOCYTE ANTIGEN

HLAs are a group of genes located in the major histocompatibility complex (MHC) region in the short arm of chromosome 6 (6p21.3) that encode proteins crucial for the immune system, organ transplantation compatibility, and susceptibility to autoimmune diseases (Erlich and Apple 1998). These genes encode transmembrane glycoproteins expressed in almost all the nucleated cells in the body. MHC region extends for 4.2 million base pairs and contains over 200 genes divided into three major classes: MHC class I, II and III (Shiina et al. 2009).

POLYMORPHISMS IN THE HLA GENES

HLA polymorphism refers to the extraordinary variability of the HLA genes (Mungall et al. 2003). This extensive diversity arises from frequent mutations, recombination, and gene conversion-like events (Little and Parham 1999). Although it is highly beneficial for the immune system in recognizing a wider range of pathogens, this HLA diversity poses challenges for direct HLA genotyping and sequencing. As of 2025, the IPD-IMGT/HLA Database reports 42 214 HLA alleles, many of which differ only by fewer nucleotides. These high sequence similarities combined with dense linkage disequilibrium (LD) make it difficult for standard genotyping methods such as SNP microarray to capture the full extent of its variability. Even methods such as short sequence reads fail to accurately distinguish between alleles as they could differ even only by a single nucleotide (Larjo et al. 2017). In addition, the HLA region is also highly population-specific, varying significantly between ethnic groups (Arrieta-Bolaños, Hernández-Zaragoza and Barquera 2023).

HLA IMPUTATION

HLA imputation is a method for predicting HLA alleles using genetic markers such as SNPs that are typically genotyped through an SNP microarray. Direct laboratory-based HLA typing is a challenging and expensive as the MHC region is genetically complex and extensively polymorphic. Therefore, imputation tools are used, which use probabilistic prediction algorithms or machine learning approaches to predict the HLA alleles by comparing the patterns of SNPs to reference panel data, which is a collection of SNP and HLA data from multiple individuals that serve as a dictionary (Dilthey et al. 2011, Breiman 1996, Jia et al. 2013). Reference panels are built by phasing SNP genotype data alongside HLA alleles which is the core ingredient for imputation (Sakaue et al. 2023). In addition to HLA alleles, amino acid polymorphisms and SNPs within the HLA will also be imputed, if included in the reference panel. As HLA is population-specific and varies highly among ethnic groups, the imputation quality depends on the data quality in the reference panel and how well the reference panel represents the target population (Flanagan et al. 2024).

Over the years, several reference panels were developed for HLA imputation. Some panels aimed at multi-ethnic representation (Pan Asian), whereas others were population-specific and focused on a single ancestry (Jia et al. 2013). Although the multi-ancestry reference panel offers a large

sample size and genetic markers, they are less population-specific, making them less optimal for HLA studies.

T1DGC REFERENCE PANEL

The Type 1 Diabetes Genetics Consortium (T1DGC) reference panel was constructed by Jia et al. (2013) using genotype data of 5225 unrelated individuals of European ancestry from the Type 1 Diabetes Genetic Consortium study. These individuals were SNP-genotyped and HLA-typed (laboratory-based) at four-digit resolution for key HLA genes, including *HLA-A, B, C, DPA1, DPB1, DQA1, DQB1*, and *DRB1* relevant to autoimmune diseases, making it valuable for HLA imputation (Jia et al. 2013). Although the T1DGC remain one of the most well-characterized and robust reference panels for the European population, it was originally built for SNP2HLA (v1.0), which is now slow, memory inefficient and unable to handle large datasets. Modern imputation frameworks such as HLA-TAPAS allow faster and more efficient processing of large datasets (Roshyara et al. 2016). Despite having high-quality HLA data, the reference panel is not compatible with these newer pipelines in its current form. The second part of the project focuses on adapting the T1DGC panel for use in modern imputation pipelines.

AIMS AND OBJECTIVES

1. To perform sex-stratified association analysis and meta-analysis to investigate sex-differentiated genetic effects in Juvenile Idiopathic Arthritis (JIA) and Rheumatoid Arthritis (RA).
2. To reformat the existing T1DGC reference panel to ensure compatibility and functionality with the HLA-TAPAS and Minimac4-based imputation pipelines.

MATERIALS AND METHODS – SEX-DIFFERENTIATED GENETIC EFFECTS IN JIA

HLA-IMPUTED GENOTYPE DATASET FOR JIA

HLA imputed, and post quality-controlled (QC) dataset comprising individuals diagnosed with JIA (n= 3,496) and healthy individuals (n = 9,196), all of UK origin, was used for this study. The dataset is a subset derived from a previously published study (López-Isac et al. 2021).

HLA-IMPUTED GENOTYPE DATASET FOR RA

HLA imputed and post-QCed dataset comprising individuals diagnosed with RA (n = 2,406) and healthy individuals (n = 8,420), all of UK origin, was used for this study. The dataset is a subset derived from a previously published study (Eyre et al. 2012). HLA imputation and QC was done at the University of Manchester.

ASSOCIATION TESTING AND SEX-STRATIFIED META-ANALYSIS METHODS

Analysis was performed separately for JIA and RA. Primary association analysis was done using PLINK v1.9 (https://www.cog-genomics.org/plink/1.9/) using logistic regression, and top three principal components as covariates and summary statistics were generated separately for males and females. The logistic regression model used for standard association was:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1.Genotype + \beta_2.Covariate_1 + . .$$

Sex-stratified meta-analysis was done separately using the GWAMA package (Genome-Wide Association Meta-Analysis v.2.2.2; https://genomics.ut.ee/en/tools). Sex-stratified summary statistics were grouped and converted to GWAMA-compatible format using PLINK2GWAMA.pl script. A marker map file was created to include chromosome number and base pair positions. GWAMA calculated two statistics: P_{het}, which tests for heterogeneity in effect sizes between groups (males and females), and $P_{diff,}$ which tests whether the difference in effect sizes is statistically significant, independent of phenotype association.

For secondary sex-interaction analysis, logistic regression was done using an additive genetic model, and sex was included as interaction term (P_{int}). The logistic regression model used for sex interaction analysis was:

$$\mathbf{logit}(p) = \beta_0 + \beta_1.(Genotype) + \beta_2.(Sex) + \beta_3\,(Genotype\ x\ Sex) + . .$$

All GWAS association testing and GWAMA sex-stratified meta-analysis processes were automated using custom Bash scripts (https://github.com/jgurubalan/GWAS-Sex-Stratified).

DATA ANALYSIS AND VISUALIZATION

Data analysis was done using R (v4.4.1) and RStudio (Version 2024.12.1+563). Custom R scripts were used to format and process the data for analysis. Bonferroni correction was applied to

account for multiple testing and the calculated P-value was used to threshold the variants for significant association. The formula used for $P_{threshold}$ was:

$$P_{threshold} = \frac{0.05}{No\ of\ variants\ or\ SNPs}$$

Manhattan plots were generated using the CMplot package (v 4.5.1) in R to visualize the GWAS outputs. Forest plots were generated using the Metafor package (v2.0) (https://www.metaforproject.org/do-ku.php/metafor). Variants with the strongest P_{het} were identified and plotted to check for sex-differentiated effects. Effect sizes (odds ratio; OR) were converted to log (OR) for the forest plot.

MATERIALS AND METHODS – HLA IMPUTATION STUDY

RA DATASET

Genotyped and post-QCed dataset comprising individuals diagnosed with RA (n = 3,870) and healthy individuals (n = 8,430), all of UK origin, was obtained from a previously published study (Eyre et al. 2012).

HLA IMPUTATION TOOLS

HLA imputation was done using SNP2HLA module from HLA-TAPAS (HLA-Typing At Protein for Association Studies) pipeline (https://github.com/immunogenomics/HLA-TAPAS) (Luo et al. 2021) and Minimac4 (https://github.com/statgen/Minimac4) (Fuchsberger, Abecasis and Hinds 2015). Both tools infer classical HLA alleles, amino acid polymorphisms, and intragenic SNPs using reference panel-based probabilistic methods built on Hidden Markov Models (HMMs) (Sakaue et al. 2023).

T1DGC REFERENCE PANEL PREPARATION FOR IMPUTATION

Dr John Bowes provided the T1DGC reference panel data. It contained 8,961 genetic markers, including classical HLA alleles (2-digit and 4-digit resolution), amino acid polymorphisms, and intragenic SNPs. As the T1DGC reference panel was built on the GRCh36/hg18 reference genome, converting it to GRCh37 or GRCh38 is mandatory as the SNP2HLA in HLA-TAPAS uses updated Beagle 4.1. UCSC LiftOver tool was used to update the genomic positions of the bim and markers file to the correct genomic build. The reference panel files were in a legacy format where amino acid polymorphisms were represented using single-letter codes and HLA alleles from the same locus shared genomic positions, making them incompatible with newer software. The metadata files (bim/markers/FRQ.frq) were updated using AT-trick.py and redefineBPv1BH.py (scripts from HLA-TAPAS) to standardize allele encodings (reference and alternate allele for variants) and to assign unique base-pair position to all variants. Using the java application beagle2vcf.jar (https://faculty.washin-gton.edu/browning/beagle_utilities/beagle2vcf.jar), the original phased genotype file (phased.bgl) was converted to variant call format (VCF) using the updated metadata files.

The VCF file was compressed to an MSAV format using the Minimac4 pipeline for use in Minimac4. Figure. 1 provides an overview of the reformatting workflow.

TEST DATASET PREPROCESSING

To ensure compatibility with reference panels' genome build (GRCh37), UCSC LiftOver was used to convert the genome build of the study dataset from hg18 (GRCh36) to hg19 (GRCh37). No test dataset preprocessing was required for the HLA-TAPAS pipeline. For Minimac4, the study dataset was first phased using EAGLE (v2.4.1), then converted to VCF format using PLINK (v2.0). The final VCF file was indexed using tabix (v0.2.6) to ensure compatibility with Minimac4.

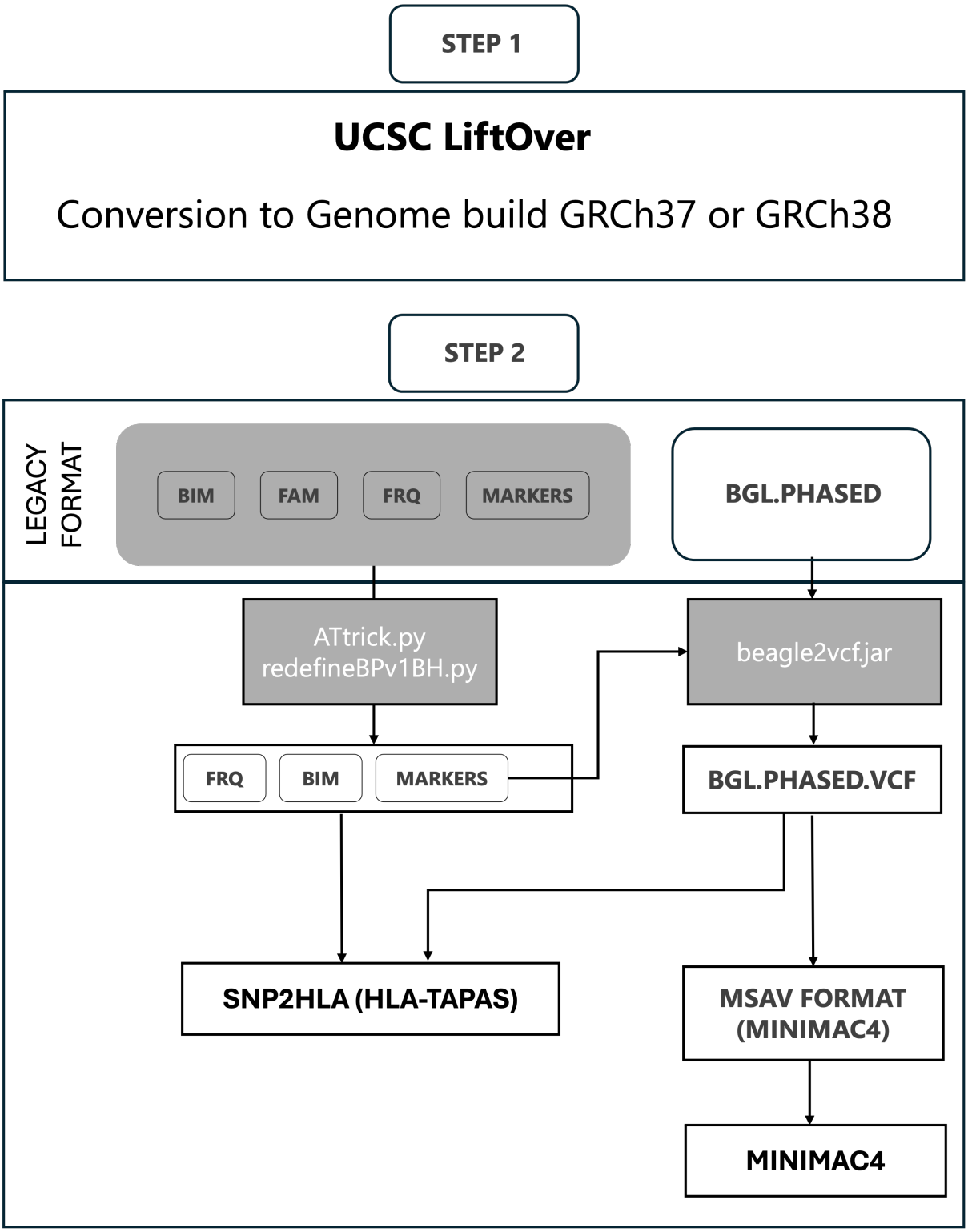

Figure 1: T1DGC Reference panel formatting workflow for SNP2HLA (HLA-TAPAS) and Minimac4.

EVAULATION OF IMPUTATION PERFORMANCE

Imputation performance was assessed using internal quality metrics. Key parameters include allelic R-squared (AR^2) values (SNP2HLA), R^2 (Minimac4) and dosage distribution inspection. For cross-platform consistency, imputed allele frequencies (AF) between SNP2HLA and Minimac4 were correlated. For cross-panel consistency, the test dataset was imputed using 1000G (publicly available reference panel) and AR^2 was correlated with T1DGC imputed outputs. All analyses were done using custom scripts in RStudio, plots were made using ggplot2 and correlation analysis was done using 'cor' function.

RESULTS - SEX-DIFFERENTIATED GENETIC EFFECTS IN JIA

ASSOCIATION ANALYSIS IN JIA

Association analysis was done on 3,494 JIA cases (males = 1123, females = 2269, unknown sex = 104) and 9,196 controls (males = 4,059, females = 5,137). 8,072 variants passed QC. Genomic inflation metrics (λGC) were not calculated as the HLA-imputed variants exhibited strong LD. Logistic regression on the sex-combined samples revealed 1049 significant variants (Bonferroni-corrected threshold P value < $6.2X10^{-6}$; results not shown).

SEX STRATIFIED META-ANALYSIS IN JIA

Sex-differentiated genetic effects were investigated by performing sex-stratified GWAS followed by a meta-analysis in GWAMA. P_{diff} was used to identify variants with significant differences in effect sizes between males and females ($P_{diff} < 6.2X10^{-6}$; Figure 2A). Subsequently, P_{het} was used to identify variants that showed significant heterogeneity between sexes ($P_{het} < 6.2X10^{-6}$; Figure 2B). A total of 199 variants showed significant gender heterogeneity of which 98 variants were in *HLA-B*, 14 in *DRB1*, 23 in *DQA1*, 23 in *DQB1*, 2 in *DPA1*, and 1 in *C* regions. The genotype-by-sex interaction analysis (P_{int}) results were largely consistent with the P_{het} findings, providing additional evidence for sex-differentiated effects. The strongest signal for P_{het} was from SNP_B_31432179_C ($P_{het} = 9.6\times10^{-23}$, $P_{diff} = 4.5X10^{-27}$, $P_{int} = 7.12X10^{-8}$). A list of the top significant variants is given in Table 1. To avoid over-representing single signals, only top signals from *DRB1* and *B* loci were prioritized as they have established associations with JIA. SNP_B_31432179_C and AA_DRB1_11_32659926_LE were chosen for downstream visualization using forest plot to illustrate the sex-differentiated effects.

Intragenic SNP, namely SNP_B_31432179_C in the sex-combined sample, was associated with an increased risk of JIA ($\log(OR_{sex\text{-}combined}) = 0.22$, 95% CI: 0.13 to 0.31). However, the sex-stratified analysis revealed divergent effects: in males, the variant conferred risk ($\log(OR_{male}) = 0.75$, 95% CI: 0.62 to 0.89), whereas in females, it showed a protective effect ($\log(OR_{female}) = -0.14$; 95% CI: -0.25 to -0.03), highlighting a clear sex-dimorphic effect. Similarly, amino acid polymorphism AA_DRB1_74_32659926_LE showed the strongest signal within the DRB1 region, with an overall sex-combined risk association ($\log(OR_{sex\text{-}combined}) = 0.71$, 95% CI: 0.62 to 0.80). The risk effect was significantly more in females ($\log(OR_{female}) = 0.88$, 95% CI: 0.77 to 0.99) when compared to males ($\log(OR_{male}) = 0.37$, 95% CI: 0.21 to 0.53) (Figure 2C).

ASSOCIATION ANALYSIS IN RA

Association analysis was done on 2,406 RA cases (males =645, females = 1,761) and 8,430 controls (males = 3,908, females = 4,442). 8,037 variants passed QC filters. GWAS was done on all individuals thresholded for Bonferroni correction ($P<6.2X10^{-6}$). Logistic regression on the sex-combined samples revealed1436 significant variants, including HLA alleles, amino acid polymorphisms and intragenic SNPs (results not shown).

SEX STRATIFIED META-ANALYSIS IN RA

P_{diff} was used to identify variants with significant differences in effect sizes between males and females (Figure. 3A). However, no variants showed significant gender heterogeneity based on P_{het} ($P_{het} < 6.2X10^{-6}$). As a result, no further downstream analysis was performed for sex-differentiated effects in the RA dataset (Figure 3B).

To summarize, sex-stratification revealed variants with significant genetic heterogeneity in JIA that were not detectable during routine sex-aggregated analysis. The strongest signal was identified in the *HLA-B* region which showed a sex dimorphic effect, while no comparable sex-differentiated effects were observed in RA. These findings emphasize the importance of sex-stratification in the genetic studies.

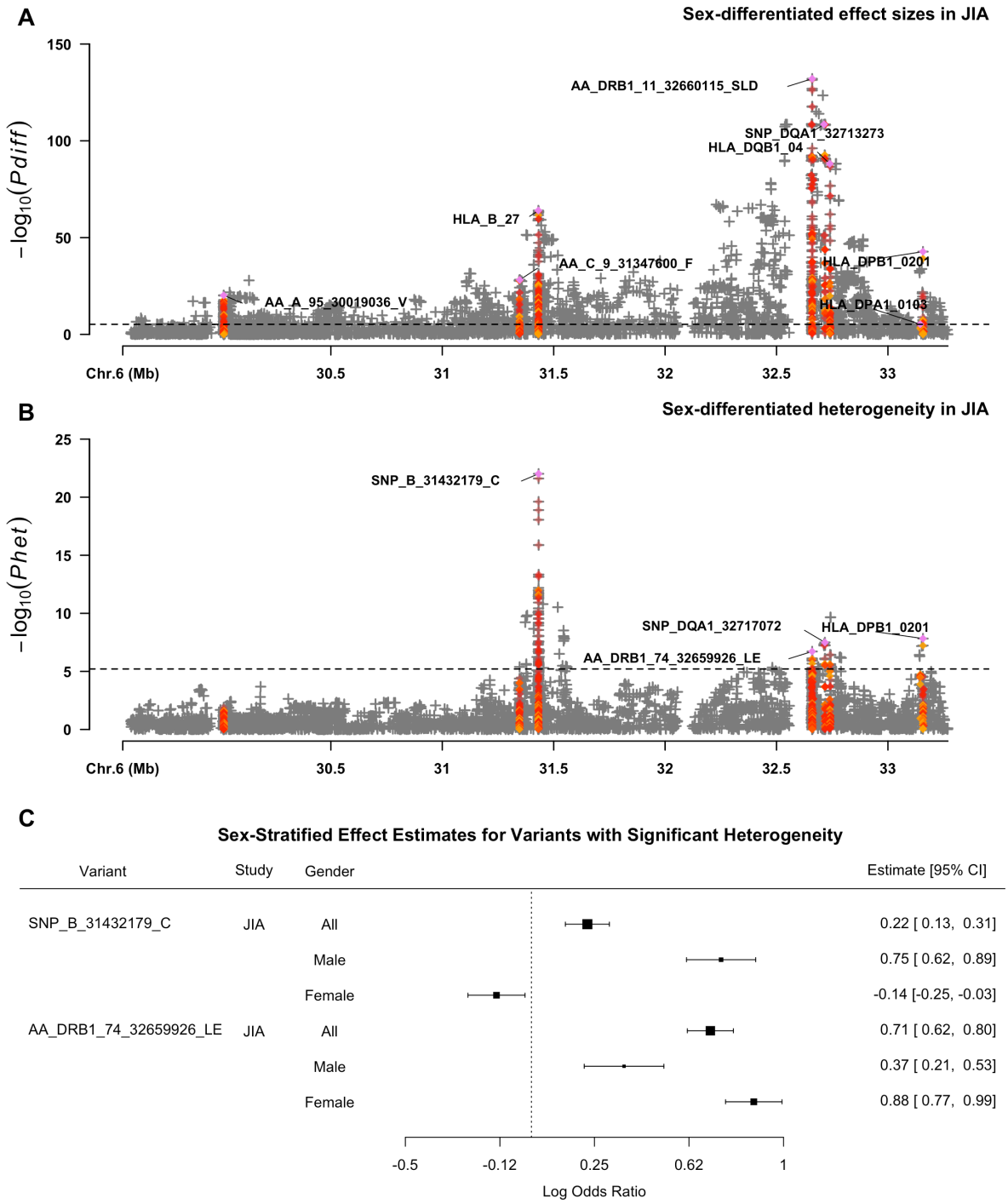

Figure 2: Sex-differentiated effects in JIA.

A – Manhattan plot shows -$\log_{10}(P_{diff})$ representing the sex-differentiated effect size for all variants across chromosome 6 in JIA. **B** – shows -log10(Phet) representing sex-differentiated heterogeneity for all variants. In both graphs, the horizontal dashed line (black) represents the Bonferroni correction threshold for significance ($P = 6.2X10^{-6}$) corresponding to -log10(P) ≈ 5.21. Data points for variants are represented by "+"; classical HLA alleles (orange); amino acid polymorphisms (red) and other variants (grey). Labelled data points represent top signals (significant) from individual HLA regions. **C** – Forest plot showing log(OR) for top significant variants with 95% CI. Vertical dashed line (black) represents 'logOR = 0' indicating no effect; 'logOR>0' indicates risk, and 'logOR<0' indicates protective effect. Variants whose CI do not cross suggest statistical significance.

Table 1: Top variants with significant sex-specific heterogeneity in JIA

VARIANT	EFFECT SIZE $P_{DIFF} < 6.26X10^{-6}$	SEX-COMBINED				MALE				FEMALE				HETEROGENEITY $P_{HET} < 6.26X10^{-6}$	SEX-INTERACTION $P_{INT} < 6.26X10^{-6}$
		logOR	CI 95%			logOR	CI 95%			logOR	CI 95%				
SNP_B_31432179_C	4.50×10^{-27}	0.22	0.13	to	0.31	0.75	0.62	to	0.89	-0.14	-0.25	to	-0.03	9.67×10^{-23}	7.12×10^{-8}
AA_B_97_31432180_SN	1.11×10^{-26}	0.22	0.13	to	0.31	0.75	0.61	to	0.89	-0.13	-0.25	to	-0.02	2.47×10^{-22}	9.9×10^{-8}
AA_B_97_31432180_SNV	1.06×10^{-21}	0.13	0.05	to	0.21	0.59	0.46	to	0.71	-0.17	-0.27	to	-0.07	2.43×10^{-20}	5.4×10^{-9}
AA_B_97_31432180_SWN	2.17×10^{-19}	0.08	0.00	to	0.16	0.54	0.41	to	0.67	-0.23	-0.33	to	-0.12	1.29×10^{-19}	2.0×10^{-7}
AA_B_97_31432180_RT	7.68×10^{-18}	0.02	-0.05	to	0.10	0.44	0.32	to	0.56	-0.25	-0.34	to	-0.15	8.83×10^{-19}	1.0×10^{-8}
AA_B_70_31432506_QK	4.70×10^{-21}	0.23	0.14	to	0.32	0.68	0.54	to	0.82	-0.08	-0.20	to	0.03	1.32×10^{-16}	9.3×10^{-15}
SNP_B_31432505	4.70×10^{-21}	0.23	0.14	to	0.32	0.68	0.54	to	0.82	-0.08	-0.20	to	0.03	1.32×10^{-16}	9.3×10^{-15}
AA_B_70_31432506_N	3.90×10^{-15}	0.12	0.04	to	0.20	0.49	0.36	to	0.61	-0.12	-0.22	to	-0.02	4.64×10^{-14}	$4.2\times\times 10^{-15}$
AA_B_69_31432509_T	3.90×10^{-15}	0.12	0.04	to	0.20	0.49	0.36	to	0.61	-0.12	-0.22	to	-0.02	4.64×10^{-14}	4.2×10^{-15}
SNP_B_31432510_T	3.90×10^{-15}	0.12	0.04	to	0.20	0.49	0.36	to	0.61	-0.12	-0.22	to	-0.02	4.64×10^{-14}	4.2×10^{-15}
AA_B_71_31432503	5.15×10^{-15}	0.12	0.05	to	0.20	0.48	0.36	to	0.60	-0.12	-0.22	to	-0.02	7.15×10^{-14}	6.9×10^{-15}
AA_B_69_31432509_A	5.15×10^{-15}	0.12	0.05	to	0.20	0.48	0.36	to	0.60	-0.12	-0.22	to	-0.02	7.15×10^{-14}	6.9×10^{-15}
SNP_B_31432504	5.15×10^{-15}	0.12	0.05	to	0.20	0.48	0.36	to	0.60	-0.12	-0.22	to	-0.02	7.15×10^{-14}	6.9×10^{-15}
SNP_B_31432510_C	5.15×10^{-15}	0.12	0.05	to	0.20	0.48	0.36	to	0.60	-0.12	-0.22	to	-0.02	7.15×10^{-14}	6.9×10^{-15}
SNP_B_31432180_CG	1.68×10^{-41}	0.56	0.47	to	0.65	0.95	0.81	to	1.10	0.26	0.13	to	0.38	6.46×10^{-13}	2.9×10^{-14}
HLA_B_2705	2.75×10^{-62}	0.88	0.77	to	1.00	1.33	1.17	to	1.50	0.51	0.35	to	0.66	1.02×10^{-12}	2.4×10^{-14}
AA_B_77_31432485_SN	2.83×10^{-52}	0.71	0.61	to	0.81	1.13	0.98	to	1.29	0.38	0.25	to	0.52	1.02×10^{-12}	1.5×10^{-14}
SNP_B_31432486	3.33×10^{-52}	0.71	0.61	to	0.81	1.12	0.97	to	1.28	0.39	0.25	to	0.52	2.14×10^{-12}	3.4×10^{-14}
HLA_B_27	1.01×10^{-64}	0.89	0.78	to	1.00	1.33	1.17	to	1.50	0.53	0.38	to	0.68	2.59×10^{-12}	7.4×10^{-14}
SNP_B_31430829	3.12×10^{-63}	0.88	0.77	to	1.00	1.32	1.16	to	1.49	0.52	0.37	to	0.67	3.50×10^{-12}	9.2×10^{-14}
SNP_B_31432179_A	7.08×10^{-64}	0.89	0.78	to	1.00	1.32	1.16	to	1.49	0.53	0.38	to	0.68	5.40×10^{-12}	1.4×10^{-13}
SNP_B_31432180_T	7.08×10^{-64}	0.89	0.78	to	1.00	1.32	1.16	to	1.49	0.53	0.38	to	0.68	5.40×10^{-12}	1.4×10^{-13}
HLA_DPB1_0201	1.83×10^{-43}	0.53	0.45	to	0.61	0.19	0.05	to	0.33	0.69	0.59	to	0.79	1.49×10^{-8}	2.7×10^{-8}
HLA_DPB1_02	6.29×10^{-40}	0.49	0.41	to	0.57	0.18	0.04	to	0.32	0.64	0.55	to	0.74	5.95×10^{-8}	1.7×10^{-7}
HLA_DRB1_08	3.57×10^{-92}	1.34	1.21	to	1.47	0.89	0.66	to	1.11	1.58	1.42	to	1.74	9.03×10^{-7}	1.7×10^{-6}
HLA_DRB1_1301	8.91×10^{-13}	0.34	0.23	to	0.46	0.10	-0.32	to	0.12	0.52	0.38	to	0.66	2.13×10^{-6}	2.4×10^{-6}
HLA_DQA1_04	2.35×10^{-93}	1.42	1.28	to	1.55	0.96	0.73	to	1.20	1.66	1.49	to	1.83	2.35×10^{-6}	3.7×10^{-6}
HLA_DQA1_0401	2.35×10^{-93}	1.42	1.28	to	1.55	0.96	0.73	to	1.20	1.66	1.49	to	1.83	2.35×10^{-6}	3.7×10^{-6}
HLA_DQB1_0603	1.75×10^{-12}	0.33	0.22	to	0.45	0.10	-0.31	to	0.11	0.51	0.37	to	0.65	2.35×10^{-6}	3.7×10^{-6}
HLA_DQA1_0103	1.42×10^{-11}	0.31	0.20	to	0.42	0.11	-0.33	to	0.10	0.49	0.35	to	0.62	2.85×10^{-6}	3.9×10^{-6}

P_{diff}: Effect size; P_{het}: Sex-based heterogeneity; P_{int}: Sex-interaction; OR: Odds ratio, log(OR) = 0: No effect, log(OR) > 0: risk, log(OR) < 0: protection; CI: Confidence Interval.

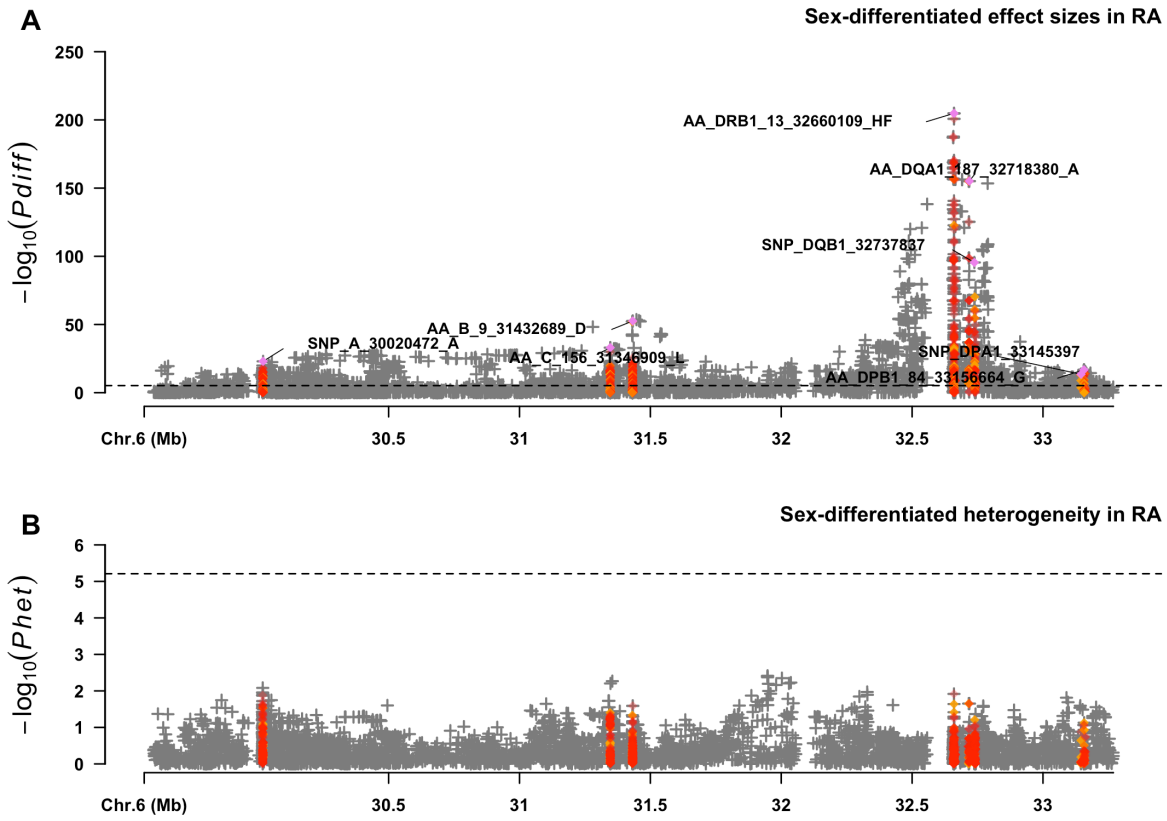

Figure 3: Sex-differentiated effects in RA.

A - Manhattan plot shows -$\log_{10}(P_{diff})$ representing the sex-differentiated effect size for all variants across chromosome 6 in RA. **B** – shows -log10(Phet) representing sex-differentiated heterogeneity for all variants. In both graphs, the horizontal dashed line (black) represents the Bonferroni correction threshold for significance ($P = 6.2X10^{-6}$) corresponding to -log10(P) ≈ 5.21. Data points for variants are represented by "+", classical HLA alleles (orange); amino acid polymorphisms (red) and other variants (grey). Labelled data points represent top signals (significant) from individual HLA regions.

RESULTS - HLA IMPUTATION STUDY

HLA IMPUTATION USING THE T1DGC REFERENCE PANEL

HLA imputation was performed using the T1DGC reference panel with two independent pipelines: SNP2HLA and Minimac4; both run on high-performance computing clusters. T1DGC panel enabled the imputation of 126 classical 2-digit and 298 classical 4-digit HLA alleles, 1276 amino acid polymorphisms, and 1390 intragenic SNPs. The study dataset comprised 12 300 individuals, including 3,870 RA cases, 8,430 controls and 7042 variants spanning the MHC region. An internal QC step within the imputation workflow filtered the genetic markers and samples with low quality, allele mismatch, missing genotypes, or poor mapping with the reference panel markers. All samples and variants passed the internal QC. Around 5099 genetic markers in the dataset matched with the reference panel. Between the study dataset and the reference panel, HLA-TAPAS retained 2,718 markers (~53%), while Minimac4 retained 2,578 (~50%) as anchors for inferring HLA alleles. The runtime for SNP2HLA was ~23 hours with a mandatory memory allotment of 24GB, while Minimac4 completed imputation in ~4 hours (no specific memory allotment needed).

IMPUTATION QUALITY ASSESSMENT

Assessment was done by evaluating the internal performance metrics. AR^2 (SNP2HLA) represented the squared correlation between the imputed and true allele dosages. R^2 (Minimac4) is an estimate of the squared correlation between imputed genotype dosages and their true genotypes. The mean AR^2 across all variants was 0.91, while the mean R^2 across all variants was 0.96, reflecting high-quality imputation in both pipelines. When variants were stratified by HLA alleles and amino acid polymorphisms, the imputation scores were consistent, with mean AR^2 at 0.5 and 0.8 respectively for SNP2HLA and mean R^2 at 0.78 and 0.907, respectively for Minimac4.

To verify the imputation performance across the allele frequency (AF) spectrum, AF was plotted against AR^2 for SNP2HLA and against R^2 for Minimac4. The scatter plot showed a general positive correlation between AF and imputation quality metric (AR^2 or R^2) across all variants, HLA alleles and amino acid polymorphisms, with moderate to higher frequency alleles exhibiting higher imputation scores (Figure 4 - Top and middle). In contrast, low-frequency (rare variants) and very high frequency alleles (extremely common) exhibited low imputation scores. This was as expected, as alleles with high frequency; almost all samples are homozygous, making it hard for the algorithm to predict. In contrast, alleles with low allele frequency are present in only a few individuals, which makes it harder for the imputation platform to learn its inheritance pattern.

Dosage distribution plots from SNP2HLA and Minimac4 showed distinct peaks at 0, 1 and 2, consistent with bi-allelic dosage profiles (Fig 4-bottom panel). This pattern indicates high-quality imputation and confident genotype dosage assignment across individuals.

CONSISTENCY BETWEEN SNP2HLA AND MINIMAC4 (CROSS-PLATFORM)

To evaluate the imputation consistency of the reformatted T1DGC reference panel across imputation tools, AFs were correlated between SNP2HLA and Minimac4 (Figure 5 top). Results showed a near-perfect correlation in the AFs between the two imputation platforms. Pearson correlation analysis estimated R of 1.0 across imputed HLA alleles and amino acid polymorphisms between the two platforms, indicating that the reformatted T1DGC reference panel was consistently and reliably interpreted across the platforms.

CONSISTENCY BETWEEN 1000G AND T1DGC PANELS (CROSS-PANEL)

Cross-panel consistency evaluation between 1000G and T1DGC showed a strong positive correlation between AR^2 derived from using both reference panels (Figure 5 bottom). Pearson correlation analysis estimated R of 0.823 for standard common variants and R of 0.972 for common HLA alleles between the two reference panels, indicating high consistency between imputation confidences across common variants between the two panels.

To summarize, the re-structured T1DGC reference panel integrated seamlessly with both SNP2HLA and Minimac4 imputation platforms. In imputing a dataset containing 12 300 samples and 5066 matched genetic markers, SNP2HLA required a longer runtime (~23 hours) and demanded larger memory allocation. At the same time, Minimac4 completed the task more efficiently (~4 hours) due to its batch-processing approach. Both platforms achieved high imputation performance.

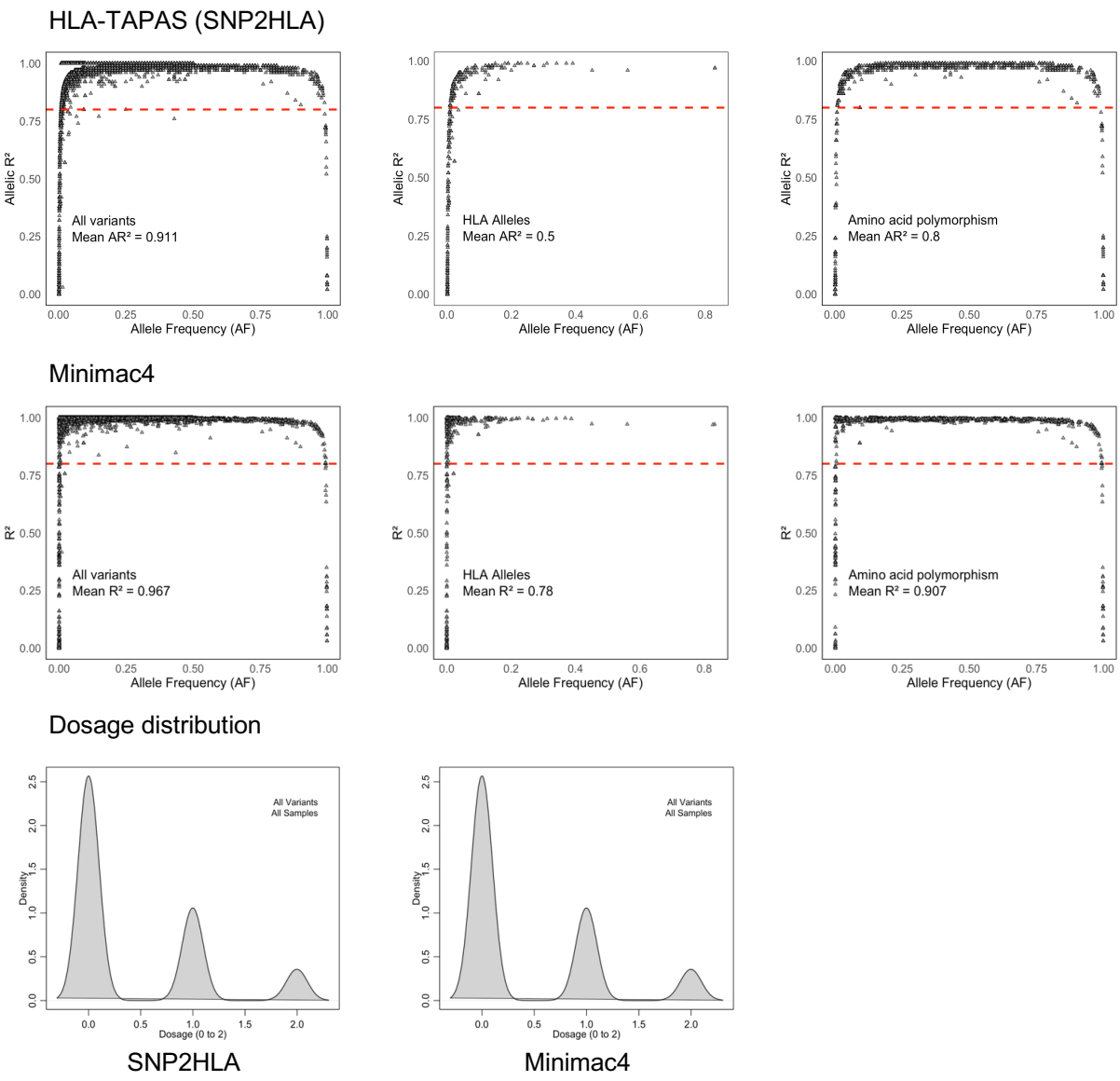

Figure 4: Imputation performance assessment between HLA-TAPAS (SNP2HLA) and Minimac4 using T1DGC reference panel.

Top - Correlation between AF and AR^2 (SNP2HLA) and between AF and R^2 (Minimac4) across all imputed variants, classical HLA alleles and amino acid polymorphisms. In the plot, each point represents a single imputed variant. Bottom - Dosage distribution plot with the dosage value between 0 and 2 on the x-axis and the density across overall imputation on the Y-axis plotted for all imputed variants in all samples for both SNP2HLA and Minimac4.

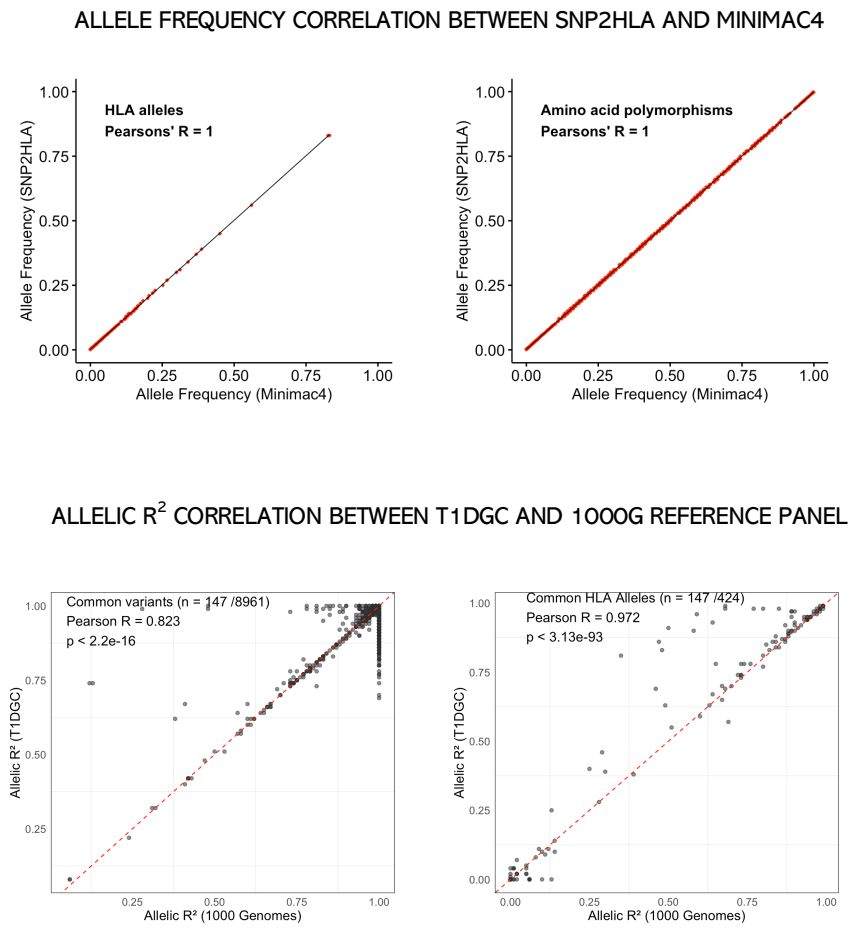

Figure 5: Assessment of cross-platform and cross-panel consistency.

Top - Correlation between allele frequencies (AFs) of HLA alleles and amino acid polymorphisms between Minimac4 and SNP2HLA (HLA-TAPAS). Data points (red) and trend line (black) shows positive correlation indicating high concordance between data points in both graphs. Bottom – Correlation between allelic-R^2 (imputation performance metric from SNP2HLA) for all common variants and common HLA alleles between 1000G and T1DGC reference panel. Data points (black) and trend line (red) shows positive correlation indicating high concordance in both graphs.

DISCUSSION

The project had two objectives. The first was to perform sex-stratified GWAS and meta-analysis in JIA and RA to identify genetic variants with differential effects in males and females.

Sex-differentiated effects in genetic variants were tested using a frequentist fixed-effect inverse-variance weighted meta-analysis using GWAMA by combining sex-stratified summary data from GWAS (Magi, Lindgren and Morris 2010). This approach allowed the identification of variants with statistically significant effect size differences (P_{diff}) and effect size heterogeneity (P_{het}) between sexes. Several genetic variants exhibited sex-differentiated genetic effects within the HLA loci in JIA (Table 1). The strongest signal was observed in an intragenic SNP at the *HLA-B* region which conferred risk in males but had a protective effect in females (SNP_B_31432179_C; $P_{het} = 9.6\times10^{-23}$; $P_{diff} = 4.5X10^{-27}$; $P_{int} = 7.12\times 10^{-8}$; $\log(OR_{male})= 0.75$; $\log(OR_{female}) = -0.14$; Figure 2C). Likewise, strongest sex-differentiated signal in the *HLA-DRB1* region was from amino acid polymorphism AA_DRB1_74_32659926_LE ($P_{het} = 1.93\times10^{-7}$; $P_{diff} = 5.57X10^{-57}$; $P_{int} = 4.4\times 10^{-6}$; $\log(OR_{male})= 1.44$; $\log(OR_{female}) = 2.41$) which conferred stronger risk in females compared to males. Without a sex-stratified analysis, this opposing or directionality effect would not have been identified. Only one study so far has reported sex-differentiated effects in HLA genes in JIA, which also reported similar observations (Tordoff et al. 2023).

The sex-differentiated effects observed in the present study correlates with clinical scenarios. Variant in the *HLA-B27* locus conferred disease risk in males while had a protective effect in females as described earlier. This finding was in concordance with previous observation where the *HLA-B27* was shown to be strongly associated with the male predominant JIA subtype (Fisher et al. 2021). Similarly, variant from the *DRB1* loci conferred greater risk in female which support the strong association of several *HLA-DRB1* alleles with female predominant subtypes of JIA such as oligoarthritis and polyarthritis (La Bella et al. 2023).

In contrast to the JIA study, no variant with significant gender heterogeneity was identified in adult-onset RA. A similar result was observed in a genome-wide SNP study in which the author reported that no statistically significant sex-specific effects in SNP were identified within the MHC (Zhuang and Morris 2009).

One of the significant limitations in the present study, is the unbalanced data distribution within the datasets, with twice as many females as males and twice as many controls as cases in both JIA and RA datasets. While this may reflect the epidemiology of the disease, it also reduces the statistical power for association testing (Zhou et al. 2018). Also, the project mainly aimed at demonstrating the existence of a sex-differentiated effect within the HLA alleles and not further downstream analyses were done. Therefore, the possibility of effect reversal (opposite directions of association between males and females) among the other alleles in loci described above is also possible. Fine mapping and conditional analyses will be required downstream to isolate the true

signals. Future work will involve stratification by disease subtypes, conditional and fine mapping analysis and functional validation to establish a biological mechanism.

The second aim of this study was to optimize the T1DGC reference panel for use in HLA-TAPAS and Minimac4 imputation platforms and to assess whether the integration was seamless and reproducible.

Using the method described, the T1DGC reference panel was reformatted to be compatible with the imputation pipelines. The imputation workflow was successfully implemented and validated for use in genome studies. The method used in this study enabled the use of a T1DGC reference panel on local machines and high-performance servers with full control over data security and flexibility.

Assessment of imputation quality metrics showed that the reformatted reference panel was highly effective and compatible with high-resolution HLA imputation using SNP2HLA and Minimac4 pipelines. Although the imputation quality was excellent for all imputed variants, there was slightly lower scores for classical HLA alleles (AR^2 = 0.5; R^2 = 0.7) and amino acid polymorphisms (AR^2 = 0.8; R^2 = 0.9). This was expected due to the HLA alleles' strong LD and complex multi-allelic nature within the MHC. Despite excellent cross-panel and cross-platform consistency with the T1DGC reference panel imputation, a key limitation is the imputation output was not externally validated with gold-standard true genotype data. Future work will include panel validation and testing with large population-scale datasets.

To conclude, older imputation frameworks such as SNP2HLA (older version) were constructed typically for small to medium-sized datasets. With the expansion in research methods, especially with larger datasets, tools such as the Michigan Imputation Server are the go-to platform for HLA imputation. However, these genotype imputation tools come with limitations and restrictions, especially for HLA imputation. Fixed reference panels restrict the ability to customize the imputation to a specific population ancestry or study design. Moreover, in countries such as the UK, where genomic data is considered a sovereign asset, uploading sensitive data to cloud-based servers outside national jurisdiction raises ethical and legal concerns (UK General Data Protection Regulation (UK GDPR), as supplemented by the Data Protection Act 2018). With the successful optimization of the T1DGC reference panel for SNP2HLA and Minimac4, it is now possible to perform high-resolution large-data genomic studies, especially within the rules of national genomic data governance. Moreover, this work also demonstrates that important sidelined reference panels that were considered obsolete due to their incompatibility with the modern imputation platforms can now be reintegrated effectively with precision into the current workflow.

REFERENCES

Arrieta-Bolaños, Esteban, Hernández-Zaragoza, Diana Iraíz, and Barquera, Rodrigo, 'An HLA Map of the World: A Comparison of HLA Frequencies in 200 Worldwide Populations Reveals Diverse Patterns for Class I and Class II', *Frontiers in Genetics*, 14 (2023) <https://www.frontiersin.orghttps://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.866407/full> [accessed 16 April 2025]

Azzouz, Doua F., Onat, Onur Emre, Balandraud, Nathalie, Kanaan, Sami B., Roudier, Jean, Ozcelik, Tayfun, et al., 'Skewed X Chromosome Inactivation in Rheumatoid Arthritis Women', *Annals of the Rheumatic Diseases*, 70/Suppl 2 (2011), A88–A88

Baillargeon, Jacques, Snih, Soham Al, Raji, Mukaila A., Urban, Randall J., Sharma, Gulshan, Sheffield-Moore, Melinda, et al., 'Hypogonadism and the Risk of Rheumatic Autoimmune Disease', *Clinical Rheumatology*, 35/12 (2016), 2983–87

Bhattacharya, Sombodhi, Sadhukhan, Debasmita, and Saraswathy, Radha, 'Role of Sex in Immune Response and Epigenetic Mechanisms', *Epigenetics & Chromatin*, 17/1 (2024), 1

Bianchi, Ilaria, Lleo, Ana, Gershwin, M. Eric, and Invernizzi, Pietro, 'The X Chromosome and Immune Associated Genes', *Journal of Autoimmunity*, 38/2–3 (2012), J187-192

Breiman, Leo, 'Bagging Predictors', *Machine Learning*, 24/2 (1996), 123–40

Cattalini, Marco, Soliani, Martina, Caparello, Maria Costanza, and Cimaz, Rolando, 'Sex Differences in Pediatric Rheumatology', *Clinical Reviews in Allergy & Immunology*, 56/3 (2019), 293–307

Chiaroni-Clarke, R. C., Li, Y. R., Munro, J. E., Chavez, R. A., Scurrah, K. J., Pezic, A., et al., 'The Association of PTPN22 Rs2476601 with Juvenile Idiopathic Arthritis Is Specific to Females', *Genes & Immunity*, 16/7 (2015), 495–98

Chiaroni-Clarke, Rachel C., Munro, Jane E., and Ellis, Justine A., 'Sex Bias in Paediatric Autoimmune Disease – Not Just about Sex Hormones?', *Journal of Autoimmunity*, 69 (2016), 12–23

Costello, Ruth, McDonagh, Janet, Hyrich, Kimme L., and Humphreys, Jenny H., 'Incidence and Prevalence of Juvenile Idiopathic Arthritis in the United Kingdom, 2000-2018: Results from the Clinical Practice Research Datalink', *Rheumatology (Oxford, England)*, 61/6 (2022), 2548–54

Deighton, C. M., Walker, D. J., Griffiths, I. D., and Roberts, D. F., 'The Contribution of HLA to Rheumatoid Arthritis', *Clinical Genetics*, 36/3 (1989), 178–82

Diack, Ngoné Diaba, Ndiaye, Nafy, Mbaye, Aminata, Kane, Baidy Sy, and Leye, Abdoulaye, 'Hypogonadotropic Hypogonadism and Juvenile Idiopathic Arthritis in an African Boy: What Is the Pathophysiological Link?', *Cureus*, 12/11, e11337

Dilthey, Alexander T., Moutsianas, Loukas, Leslie, Stephen, and McVean, Gil, 'HLA*IMP—an Integrated Framework for Imputing Classical HLA Alleles from SNP Genotypes', *Bioinformatics*, 27/7 (2011), 968–72

Dodd, Katherine C., and Menon, Madhvi, 'Sex Bias in Lymphocytes: Implications for Autoimmune Diseases', *Frontiers in Immunology*, 13 (2022), 945762

van Drongelen, Vincent, and Holoshitz, Joseph, 'Human Leukocyte Antigen–Disease Associations in Rheumatoid Arthritis', *Rheumatic Disease Clinics of North America*, Genomics in Rheumatic Diseases, 43/3 (2017), 363–76

Erlich, Henry, and Apple, Raymond, 'MHC Disease Associations', in *Encyclopedia of Immunology (Second Edition)*, ed. by Peter J. Delves (Oxford, 1998), 1690–1700 <https://www.sciencedirect.com/science/article/pii/B0122267656004412> [accessed 2 January 2025]

Eyre, Steve, Bowes, John, Diogo, Dorothée, Lee, Annette, Barton, Anne, Martin, Paul, et al., 'High-Density Genetic Mapping Identifies New Susceptibility Loci for Rheumatoid Arthritis', *Nature Genetics*, 44/12 (2012), 1336–40

Fisher, Corinne, Ciurtin, Coziana, Leandro, Maria, Sen, Debajit, and Wedderburn, Lucy R., 'Similarities and Differences Between Juvenile and Adult Spondyloarthropathies', *Frontiers in Medicine*, 8 (2021), 681621

Flanagan, Jack, Liu, Xiaoxi, Ortega-Reyes, David, Tomizuka, Kohei, Matoba, Nana, Akiyama, Masato, et al., 'Population-Specific Reference Panel Improves Imputation Quality for Genome-Wide Association Studies Conducted on the Japanese Population', *Communications Biology*, 7/1 (2024), 1–10

Fuchsberger, Christian, Abecasis, Gonçalo R., and Hinds, David A., 'Minimac2: Faster Genotype Imputation', *Bioinformatics*, 31/5 (2015), 782–84

Glass, D. N., and Giannini, E. H., 'Juvenile Rheumatoid Arthritis as a Complex Genetic Trait', *Arthritis and Rheumatism*, 42/11 (1999), 2261–68

Gmuca, Sabrina, Xiao, Rui, Brandon, Timothy G., Pagnini, Ilaria, Wright, Tracey B., Beukelman, Timothy, et al., 'Multicenter Inception Cohort of Enthesitis-Related Arthritis: Variation in Disease Characteristics and Treatment Approaches', *Arthritis Research & Therapy*, 19/1 (2017), 84

Hinks, Anne, Cobb, Joanna, Marion, Miranda C., Prahalad, Sampath, Sudman, Marc, Bowes, John, et al., 'Dense Genotyping of Immune-Related Disease Regions Identifies 14 New Susceptibility Loci for Juvenile Idiopathic Arthritis', *Nature Genetics*, 45/6 (2013), 664–69

Hisa, Kaori, Yanagimachi, Masakatsu D., Naruto, Takuya, Miyamae, Takako, Kikuchi, Masako, Hara, Rhoki, et al., 'PADI4 and the HLA-DRB1 Shared Epitope in Juvenile Idiopathic Arthritis', *PloS One*, 12/2 (2017), e0171961

Jia, Xiaoming, Han, Buhm, Onengut-Gumuscu, Suna, Chen, Wei-Min, Concannon, Patrick J., Rich, Stephen S., et al., 'Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens', *PLOS ONE*, 8/6 (2013), e64683

Kurkó, Júlia, Besenyei, Timea, Laki, Judit, Glant, Tibor T., Mikecz, Katalin, and Szekanecz, Zoltán, 'Genetics of Rheumatoid Arthritis — A Comprehensive Review', *Clinical Reviews in Allergy & Immunology*, 45/2 (2013), 170–79

La Bella, Saverio, Rinaldi, Marta, Di Ludovico, Armando, Di Donato, Giulia, Di Donato, Giulio, Salpietro, Vincenzo, et al., 'Genetic Background and Molecular Mechanisms of Juvenile Idiopathic Arthritis', *International Journal of Molecular Sciences*, 24/3 (2023), 1846

Larjo, Antti, Eveleigh, Robert, Kilpeläinen, Elina, Kwan, Tony, Pastinen, Tomi, Koskela, Satu, et al., 'Accuracy of Programs for the Determination of Human Leukocyte Antigen Alleles from Next-Generation Sequencing Data', *Frontiers in Immunology*, 8 (2017), 1815

Little, A. M., and Parham, P., 'Polymorphism and Evolution of HLA Class I and II Genes and Molecules', *Reviews in Immunogenetics*, 1/1 (1999), 105–23

López-Isac, Elena, Smith, Samantha L, Marion, Miranda C, Wood, Abigail, Sudman, Marc, Yarwood, Annie, et al., 'Combined Genetic Analysis of Juvenile Idiopathic Arthritis Clinical

Subtypes Identifies Novel Risk Loci, Target Genes and Key Regulatory Mechanisms', *Annals of the Rheumatic Diseases*, 80/3 (2021), 321–28

Luo, Yang, Kanai, Masahiro, Choi, Wanson, Li, Xinyi, Sakaue, Saori, Yamamoto, Kenichi, et al., 'A High-Resolution HLA Reference Panel Capturing Global Population Diversity Enables Multi-Ancestry Fine-Mapping in HIV Host Response', *Nature Genetics*, 53/10 (2021), 1504–16

Magi, Reedik, Lindgren, Cecilia M., and Morris, Andrew P., 'Meta-Analysis of Sex-Specific Genome-Wide Association Studies', *Genetic Epidemiology*, 34/8 (2010), 846–53

McIntosh, Laura A., Marion, Miranda C., Sudman, Marc, Comeau, Mary E., Becker, Mara L., Bohnsack, John F., et al., 'Genome-Wide Association Meta-Analysis Reveals Novel Juvenile Idiopathic Arthritis Susceptibility Loci', *Arthritis & Rheumatology (Hoboken, N.J.)*, 69/11 (2017), 2222–32

McMichael, Andrew, and Bowness, Paul, 'HLA-B27: Natural Function and Pathogenic Role in Spondyloarthritis', *Arthritis Research*, 4/Suppl 3 (2002), S153–58

Meyer, J. M., Han, J., Singh, R., and Moxley, G., 'Sex Influences on the Penetrance of HLA Shared-Epitope Genotypes for Rheumatoid Arthritis.', *American Journal of Human Genetics*, 58/2 (1996), 371–83

Mohamad, Nur-Vaizura, Wong ,Sok Kuan, Wan Hasan ,Wan Nuraini, Jolly ,James Jam, Nur-Farhana ,Mohd Fozi, Ima-Nirwana ,Soelaiman, et al., 'The Relationship between Circulating Testosterone and Inflammatory Cytokines in Men', *The Aging Male*, 22/2 (2019), 129–40

Mousavi, Mohammad Javad, Mahmoudi, Mahdi, and Ghotloo, Somayeh, 'Escape from X Chromosome Inactivation and Female Bias of Autoimmune Diseases', *Molecular Medicine*, 26/1 (2020), 127

Mungall, A. J., Palmer, S. A., Sims, S. K., Edwards, C. A., Ashurst, J. L., Wilming, L., et al., 'The DNA Sequence and Analysis of Human Chromosome 6', *Nature*, 425/6960 (2003), 805–11

Nikopensius, Tiit, Niibo, Priit, Haller, Toomas, Jagomägi, Triin, Voog-Oras, Ülle, Tõnisson, Neeme, et al., 'Association Analysis of Juvenile Idiopathic Arthritis Genetic Susceptibility Factors in Estonian Patients', *Clinical Rheumatology*, 40/10 (2021), 4157–65

Petty, Ross E., Southwood, Taunton R., Manners, Prudence, Baum, John, Glass, David N., Goldenberg, Jose, et al., 'International League of Associations for Rheumatology Classification of Juvenile Idiopathic Arthritis: Second Revision, Edmonton, 2001', *The Journal of Rheumatology*, 31/2 (2004), 390–92

Prada-Medina, Cesar Augusto, Peron, Jean Pierre Schatzmann, and Nakaya, Helder I., 'Immature Neutrophil Signature Associated with the Sexual Dimorphism of Systemic Juvenile Idiopathic Arthritis', *Journal of Leukocyte Biology*, 108/4 (2020), 1319–27

Prahalad, Sampath, 'Genetic Analysis of Juvenile Rheumatoid Arthritis: Approaches to Complex Traits', *Current Problems in Pediatric and Adolescent Health Care*, 36/3 (2006), 83–90

Prahalad, Sampath, and Glass, David N., 'A Comprehensive Review of the Genetics of Juvenile Idiopathic Arthritis', *Pediatric Rheumatology*, 6/1 (2008), 11

Rojas, Manuel, Restrepo-Jiménez, Paula, Monsalve, Diana M., Pacheco, Yovana, Acosta-Ampudia, Yeny, Ramírez-Santana, Carolina, et al., 'Molecular Mimicry and Autoimmunity',

Journal of Autoimmunity, Liver Autoimmunity: Paradigm versus Paradox and Breach of Tolerance, 95 (2018), 100–123

Roshyara, Nab Raj, Horn, Katrin, Kirsten, Holger, Ahnert, Peter, and Scholz, Markus, 'Comparing Performance of Modern Genotype Imputation Methods in Different Ethnicities', *Scientific Reports*, 6/1 (2016), 34386

Sakaue, Saori, Gurajala, Saisriram, Curtis, Michelle, Luo, Yang, Choi, Wanson, Ishigaki, Kazuyoshi, et al., 'Tutorial: A Statistical Genetics Guide to Identifying HLA Alleles Driving Complex Disease', *Nature Protocols*, 18/9 (2023), 2625–41

Selmi, Carlo, Brunetta, Enrico, Raimondo, Maria Gabriella, and Meroni, Pier Luigi, 'The X Chromosome and the Sex Ratio of Autoimmunity', *Autoimmunity Reviews*, 11/6–7 (2012), A531-537

Serhal, Lina, Lwin, May N., Holroyd, Christopher, and Edwards, Christopher J., 'Rheumatoid Arthritis in the Elderly: Characteristics and Treatment Considerations', *Autoimmunity Reviews*, 19/6 (2020), 102528

Sharma, Seema D., Leung, Shek H., and Viatte, Sebastien, 'Genetics of Rheumatoid Arthritis', *Best Practice & Research Clinical Rheumatology*, Genetics of Rheumatic Diseases, 38/4 (2024), 101968

Shiina, Takashi, Hosomichi, Kazuyoshi, Inoko, Hidetoshi, and Kulski, Jerzy K., 'The HLA Genomic Loci Map: Expression, Interaction, Diversity and Disease', *Journal of Human Genetics*, 54/1 (2009), 15–39

Sjakste, Tatjana, Paramonova, Natalia, Rumba-Rozenfelde, Ingrida, Trapina, Ilva, Sugoka, Olga, and Sjakste, Nikolajs, 'Juvenile Idiopathic Arthritis Subtype- and Sex-Specific Associations with Genetic Variants in the *PSMA6/PSMC6/PSMA3* Gene Cluster', *Pediatrics & Neonatology*, 55/5 (2014), 393–403

Tordoff, M., Smith, S., Morris, A., Eyre, S., Thomson, W., and Bowes, J., 'Sex Dimorphism Analysis in a Cohort of Jia Patients Reveals Differing Genetic Risk Factors for Females and Males', *Annals of the Rheumatic Diseases*, 82/Suppl 1 (2023), 128–128

Uchiyama, Shunsuke, Ishikawa, Yuki, Ikari, Katsunori, Honda, Suguru, Hikino, Keiko, Tanaka, Eiichi, et al., 'Mosaic Loss of Chromosome Y Characterises Late-Onset Rheumatoid Arthritis and Contrasting Associations of Polygenic Risk Score Based on Age at Onset', *Annals of the Rheumatic Diseases*, 2025 <https://www.sciencedirect.com/science/article/pii/S0003496725001840> [accessed 22 April 2025]

Uz, Elif, Mustafa, Chigdem, Topaloglu, Rezan, Bilginer, Yelda, Dursun, Ali, Kasapcopur, Ozgur, et al., 'Increased Frequency of Extremely Skewed X Chromosome Inactivation in Juvenile Idiopathic Arthritis', *Arthritis and Rheumatism*, 60/11 (2009), 3410–12

Verwoerd, Anouk, Ter Haar, Nienke M., de Roock, Sytze, Vastert, Sebastiaan J., and Bogaert, Debby, 'The Human Microbiome and Juvenile Idiopathic Arthritis', *Pediatric Rheumatology Online Journal*, 14 (2016), 55

Yanagimachi, Masakatsu, Miyamae, Takako, Naruto, Takuya, Hara, Takuma, Kikuchi, Masako, Hara, Ryoki, et al., 'Association of HLA-A*02:06 and HLA-DRB1*04:05 with Clinical Subtypes of Juvenile Idiopathic Arthritis', *Journal of Human Genetics*, 56/3 (2011), 196–99

Yang, Qianfan, Kennicott, Kameron, Zhu, Runqi, Kim, Jooyong, Wakefield, Hunter, Studener, Katelyn, et al., 'Sex Hormone Influence on Female-Biased Autoimmune Diseases Hints at Puberty as an Important Factor in Pathogenesis', *Frontiers in Pediatrics*, 11 (2023), 1051624

Zhou, Wei, Nielsen, Jonas B., Fritsche, Lars G., Dey, Rounak, Gabrielsen, Maiken E., Wolford, Brooke N., et al., 'Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies', *Nature Genetics*, 50/9 (2018), 1335–41

Zhuang, Joanna J., and Morris, Andrew P., 'Assessment of Sex-Specific Effects in a Genome-Wide Association Study of Rheumatoid Arthritis', *BMC Proceedings*, 3/7 (2009), S90