EXPLORING HLA: IMPROVING IMPUTATION AND INVESTIGATING SEX-DIFFERENTIATED GENETIC EFFECTS IN JUVENILE IDIOPATHIC ARTHRITIS

The University of Manchester
MANCHESTER 1824
Jayadevan Gurubalan
Contact:
jayadevan.gurubalan@postgrad.manchester.ac.uk

Introduction

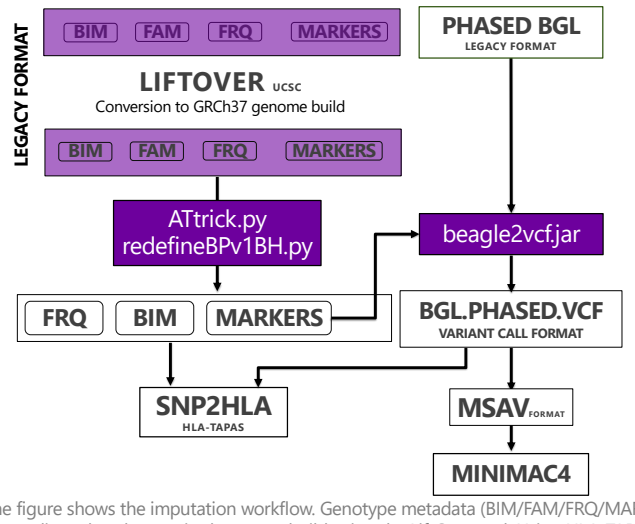
The Human Leucocyte Antigen (HLA) is an important genetic factor in developing autoimmune diseases such as Juvenile Idiopathic Arthritis (JIA). Its highly polymorphic nature and dense linkage disequilibrium make HLA research very challenging. This project contains two independent studies with distinct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

Objective 1: Reformatting T1DGC reference panel for compatibility with newer imputation tools such as SNP2HLA from HLA-TAPAS and Minimac4.

Objective 2: Investigate sex-differentiated genetic effects in JIA.

Methods

Objective 1: Protocol for reformatting the reference panel



The figure shows the imputation workflow. Genotype metadata (BIM/FAM/FRQ/MARKERS) were adjusted to the required genome build using the LiftOver tool. Using HLA-TAPAS scripts, ATtrick and redefineBPv1BH metadata files were updated. The Beagle2vcf tool was used to convert the legacy-formatted phased file to variant call format. For Minimac4, the VCF file was additionally converted to MSAV format using Minimac4 before imputation.

Objective 2: Sex-differentiated Analysis

Sex-stratified GWAS and GWAMA

The genotype dataset, which comprised 3,494 JIA patients and 9,196 controls containing single nucleotide polymorphisms (SNPs), imputed classical HLA alleles, and amino acid polymorphisms, was filtered by sex to create male-only and female-only datasets, which were analyzed separately for genome-wide association (GWAS) using logistic regression and three principal covariates. The logistic regression model used was:

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

Summary statistics from male and female GWAS were combined using genome-wide association meta-analysis (GWAMA) software to analyze the heterogeneity of SNP effects between sexes.

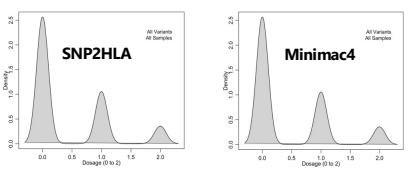
Sex-interaction analysis

Logistic regression analysis was done on sex-stratified data using an additive genetic model, and sex was included as interaction term. The logistic regression model used for sex interaction analysis was:

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

Result 1: Imputation using T1DGC reference panel

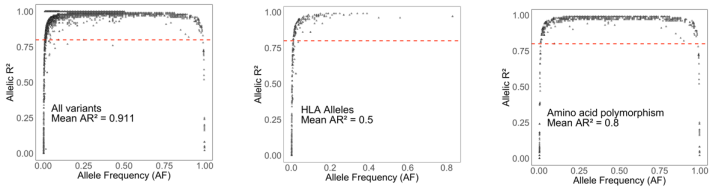
The dataset comprising 3,870 rheumatoid arthritis cases, 8,430 controls, and 7,042 genetic markers was imputed using SNP2HLA and Minimac4 with a T1DGC reference panel. The reformatted reference panel worked well with both imputation software with good imputation performance.



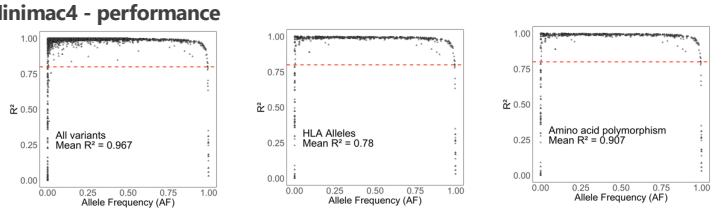
SNP2HLA	Minimac4
runtime ~ 23 hrs.	runtime ~ 4 hrs.
24GB (mandatory)	No memory allocation needed
2718 genetic markers used	2578 genetic markers used

Graphs show that most imputed markers showed discrete dosage peaks reflecting confident genotype calls in SNP2HLA and Minimac4.

SNP2HLA – performance



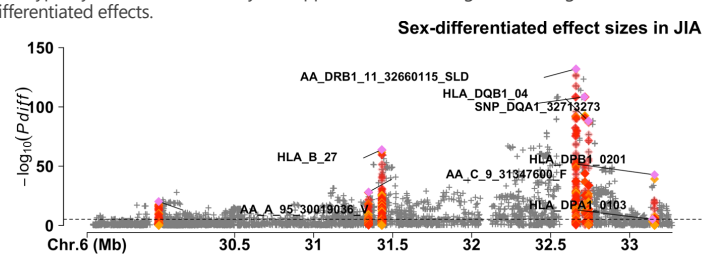
Minimac4 - performance

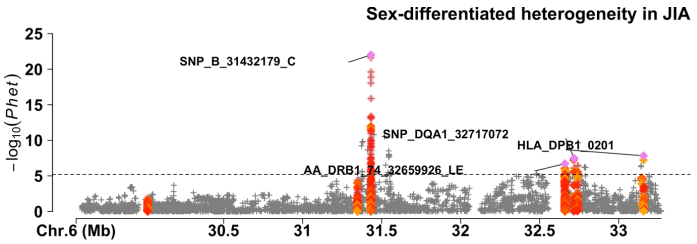


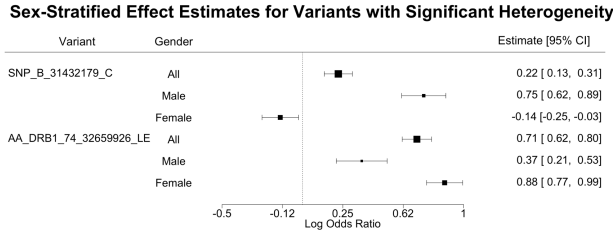
Plots show imputation score (AR^2 or R^2) across allele frequencies for variants including SNPs, HLA alleles and amino acid polymorphism. Data above the dashed line indicate high-quality variants.

Result 2: Sex-differentiated effects in JIA

One hundred ninety-nine variants showed sex-based heterogeneity, mainly in the HLA regions. Genotype-by-sex interaction analyses supported these findings, confirming robust sex-differentiated effects.







The upper and middle graph shows signals of variants with significant effect size differences (Pdiff) and heterogeneity (Phet) between sexes, respectively. The dashed horizontal line indicates a significant threshold ($P<6.2X10^{-6}$). Lower panel forest plot represents risk estimates Log Odds Ratio (LogOR) with 95% confidence interval for selected variants showing sex-specific risk differences. LogOR = 0: No effect; LogOR > 0: risk; LogOR < 0: protective effect.

Discussion & conclusion

This study demonstrates that using the methodology described above, it is now possible to reintegrate old sidelined reference panels built in legacy format to work with the modern imputation platforms. Secondly, in JIA, specific HLA alleles, amino acid polymorphisms and intragenic SNPs within the HLA region exhibit sex-dimorphic effects which are obscured during routine GWAS in which all samples (male and female combined) are analyzed together and can only be identified by performing sex-stratified analysis.