**Online data supplements**

**Figure S1.** (A) Imputation accuracy measured as average dosage R2 for 235,942 type 2 variants filtered for an information metric >0.4 and genotype probability call > 0.90. Data shown separately by chromosome and for different minor allele frequency (MAF) ranges, as indicated in the key. (B) Quantile-quantile plot of GWAS p-values.

**Figure S2**. Manhattan plots of GWAS results for the 4.46M high quality 1000G imputed SNP variants after conditioning on the top SNP rs9272622. Data are for analysis in FastLMM looking for association between SNPs and RHD.

**Figure S3.** Plots showing binding affinities for predicted epitopes of GAS M proteins recognised by HLA DQ-DB heterodimers. Epitope binding predictions were performed in NetMHCIIPan3.1. The y-axis shows the relative binding affinity (expressed as 1-log50,000 of the nM binding affinity) for heterodimers formed from risk (red, brown) and protective (blue) DQ\_DB haplotypes (see key); the x-axis indicates the amino acid sequence locations for mature proteins, also equivalent to the start position of overlapping 20mers (1-mer sliding window) in non-rheumatogenic GAS M4 (Accession number CAA33269) and M49 (Accession number AAA26868.1) sequences, GAS HSP70 (Accession number AAB39223.1) and GAS STRP1 (Accession number AAA26987.1) sequences. Horizontal lines indicate 500nM (upper) and 1000nM (lower) binding affinities.

**Table S1.** Basic demographic details (by gender, age at collection) for the 396 cases and 867 controls that passed all QC and were used in the GWAS analysis.

**Table S2.** Summary of experimentally confirmed published epitopes for (A) GAS M proteins for which there is evidence of cross-reaction with human heart-related proteins. (B) GAS M proteins for which there is evidence of non-cross-reactive T- and B-epitopes, and (C) other GAS proteins with evidence of cross-reactive epitopes.