

Supplemental Data

Identification of susceptibility loci for Takayasu arteritis through a large multi-ancestral genome-wide association study

Lourdes Ortiz-Fernández, Güher Saruhan-Direskeneli, Fatma Alibaz-Oner, Sema Kaymaz-Tahra, Patrick Coit, Xiufang Kong, Allan P. Kiprianos, Robert T. Maughan, Sibel Z. Aydin, Kenan Aksu, Gokhan Keser, Sevil Kamali, Murat Inanc, Jason Springer, Servet Akar, Fatos Onen, Nurullah Akkoc, Nader A. Khalidi, Curry Koenig, Omer Karadag, Sedat Kiraz, Lindsay Forbess, Carol A. Langford, Carol A. McAlear, Zeynep Ozbalkan, Sule Yavuz, Gozde Yildirim Çetin, Nilufer Alpaz-Kanitez, Sharon Chung, Askin Ates, Yasar Karaaslan, Kathleen McKinnon-Maksimowicz, Paul A. Monach, Hüseyin T.E. Ozer, Emire Seyahi, Izzet Fresko, Ayse Cefle, Philip Seo, Kenneth J. Warrington, Mehmet A. Ozturk, Steven R. Ytterberg, Veli Cobankara, Ahmet Mesut Onat, Nursen Duzgun, Muge Bicakcigil, Sibel P. Yentür, Lindsay Lally, Angelo A. Manfredi, Elena Baldissera, Eren Erken, Ayten Yazici, Bünyamin Kısacık, Timuçin Kaşifoğlu, Ediz Dalkilic, David Cuthbertson, Christian Pagnoux, Antoine Sreih, Guillermo Reales, Chris Wallace, Jonathan D. Wren, Deborah S. Cunninghame-Graham, Timothy J. Vyse, Ying Sun, Huiyong Chen, Peter C. Grayson, Enrico Tombetti, Lindi Jiang, Justin C. Mason, Peter A. Merkel, Haner Direskeneli, and Amr H. Sawalha

SUPPLEMENTARY MATERIAL

Supplementary Figures

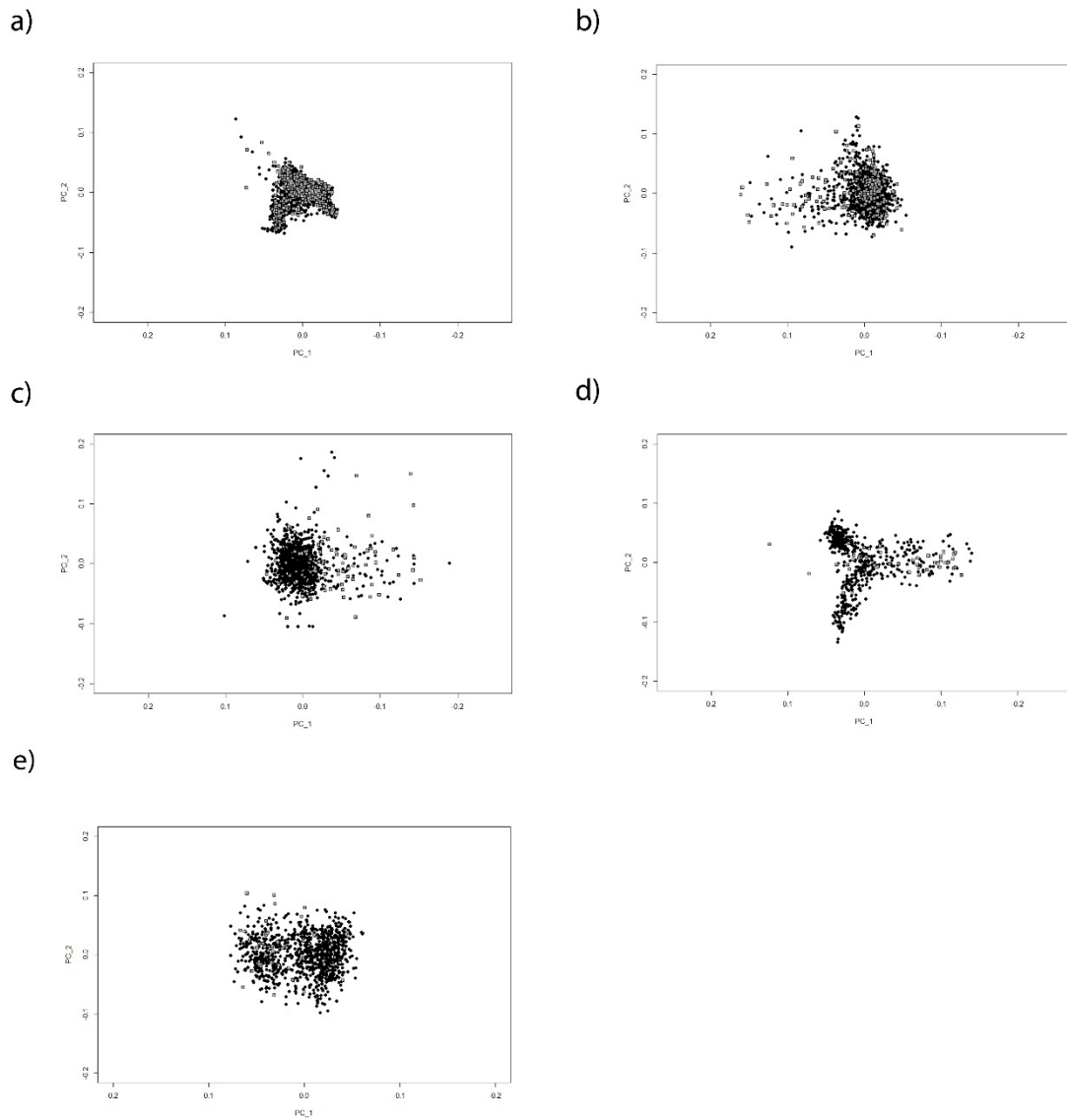


Figure S1. Principal component (PC) analysis illustrating PC1 and PC2 in each individual population: Turkish (a), Northern European descendant (b), Han Chinese (c), South Asian (d) and Italian (e). Cases and controls are shown as grey squares and black circles, respectively.

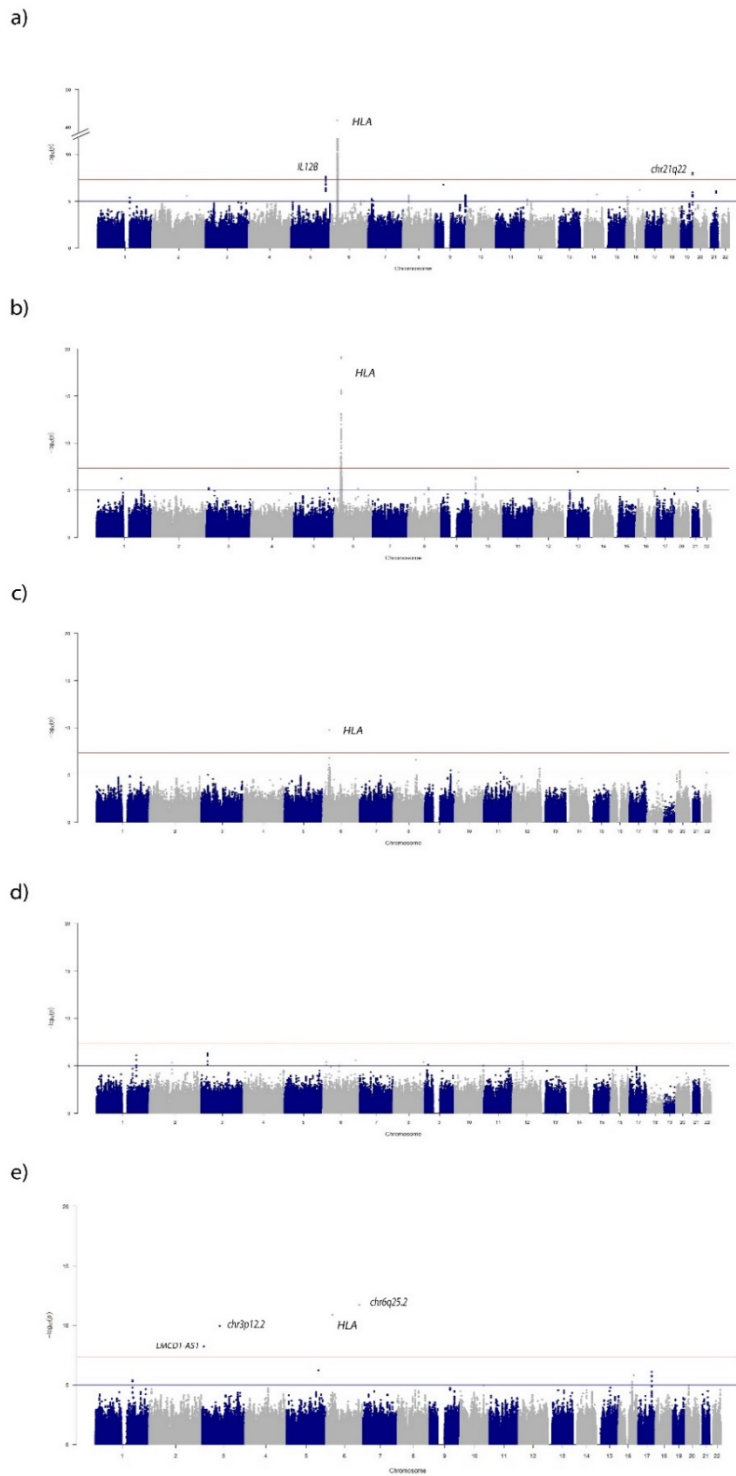


Figure S2. Manhattan plot representation of the GWAS results of each individual population: Turkish (a), Northern European descendant (b), Han Chinese (c), South Asian (d) and Italian (e). The $-\log_{10}$ p value for each genetic variant analyzed are plotted against their physical chromosomal position. The red and blue lines represent the genome-wide level of significance (p value 5×10^{-8}) and the suggestive level of significance (p value 5×10^{-5}), respectively.

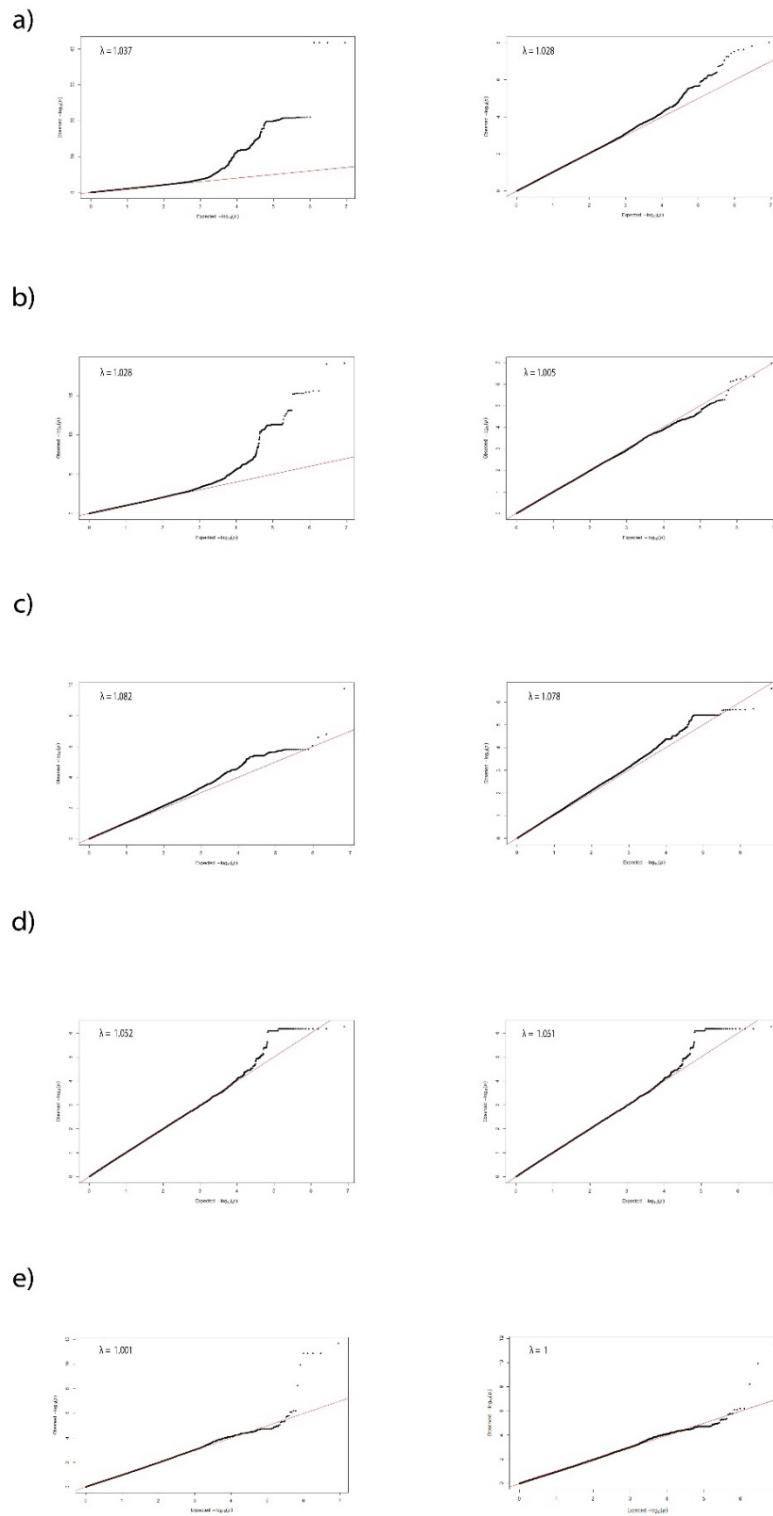
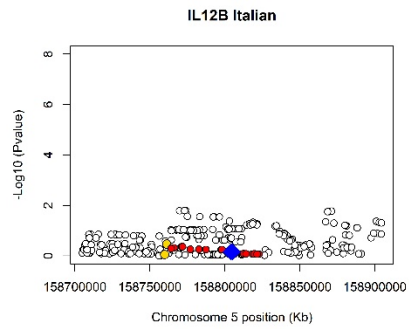
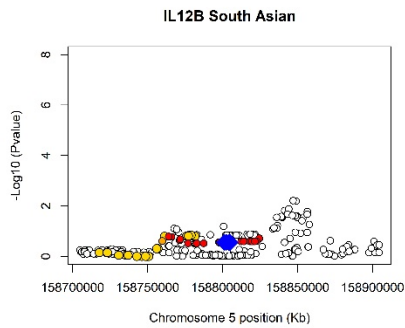
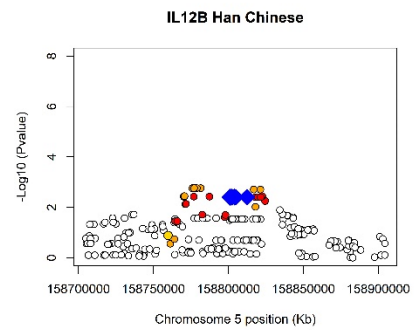
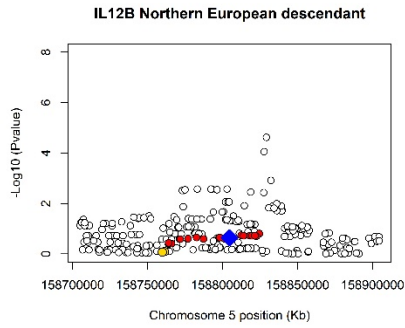
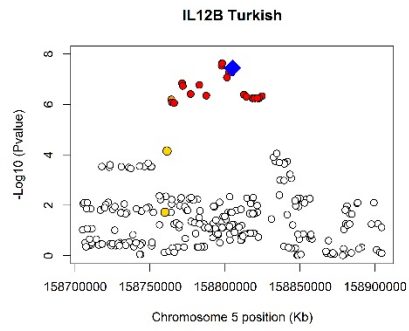
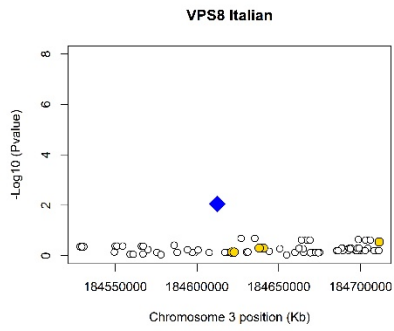
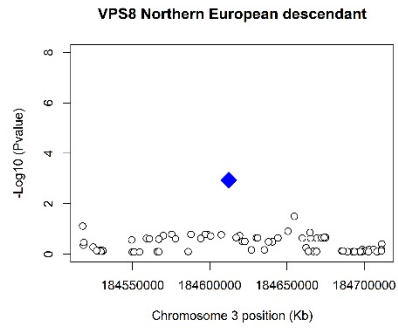
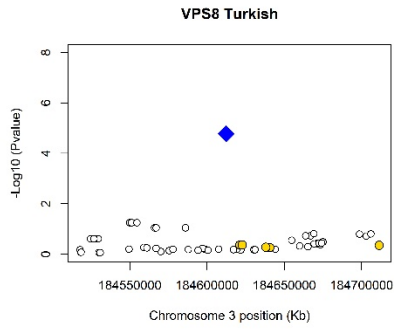
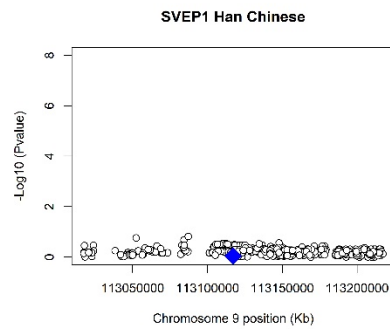
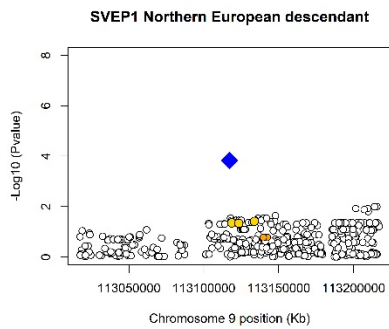
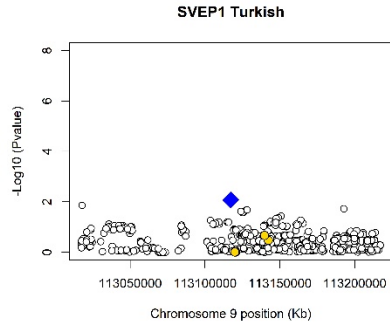
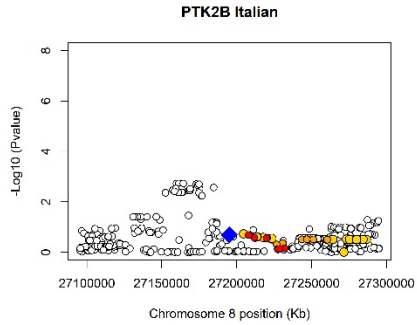
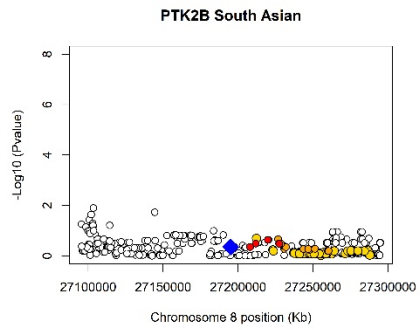
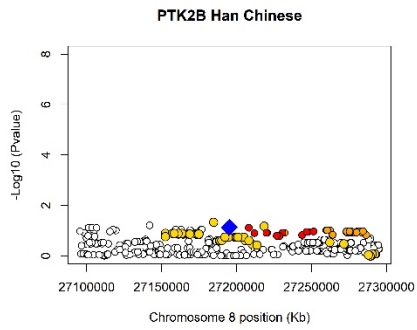
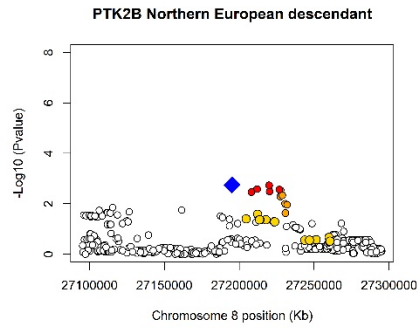
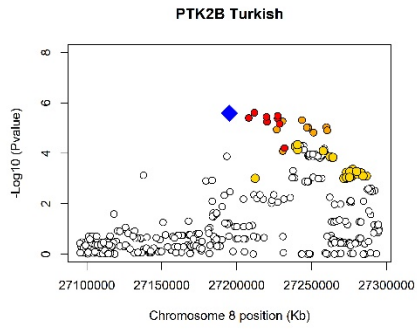
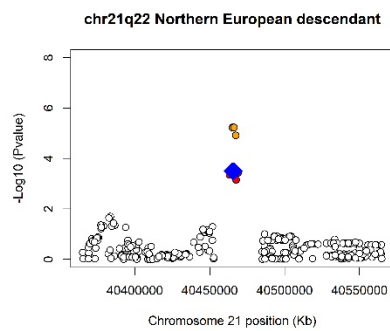
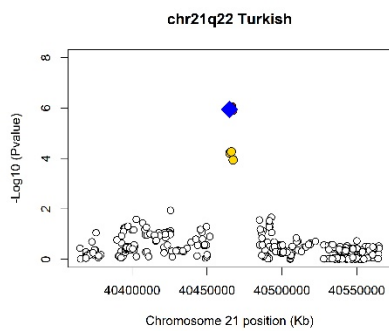
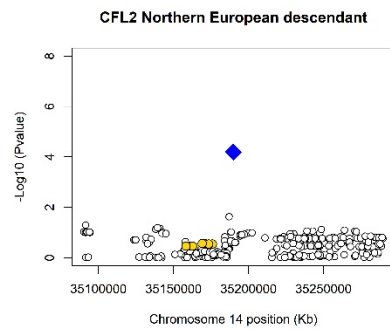
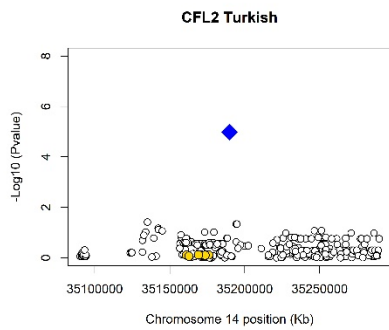
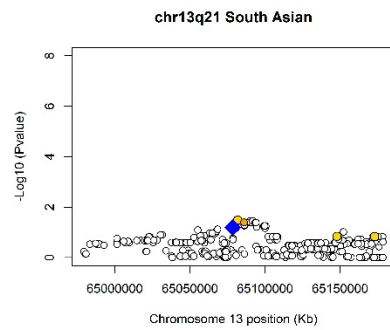
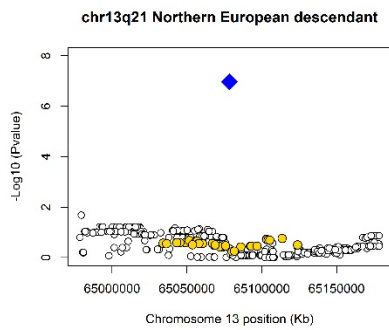
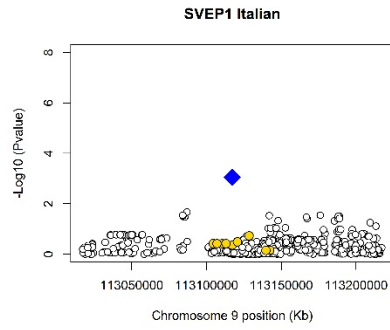
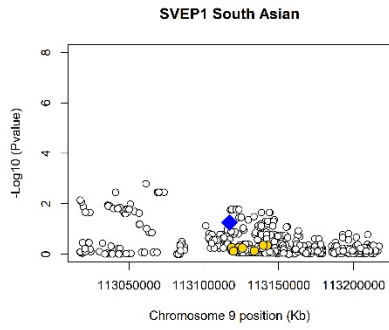


Figure S3. Quantile–quantile plots for the p values of each individual population: Turkish (a), Northern European descendant (b), Han Chinese (c), South Asian (d) and Italian (e). The x-axis indicates the expected distribution of $-\log_{10}(p \text{ values})$ and the y-axis indicates the observed distribution of $-\log_{10}(p \text{ values})$. Genomic inflation of the observed distribution is indicated by λ .







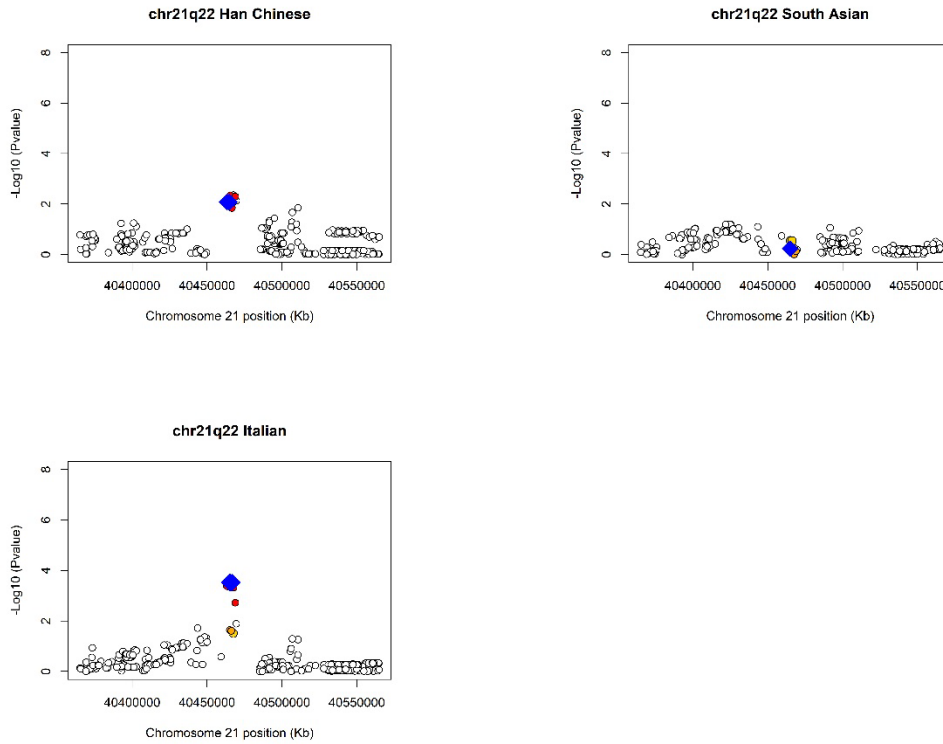


Figure S4. Manhattan-regional plots of the 7 GWAS-associated loci showing the results of the logistic regression analysis in each population. The $-\log_{10}$ p value for the genetic variants analyzed are plotted against their physical chromosomal position. Color represents the linkage disequilibrium (LD) with the index SNP (the most significant variant within each genomic region). The index variant and the ones in complete LD ($r^2 = 1$) are represented by a blue diamond. The rest of the variants are colored according to the following code: White: $r^2 < 0.4$; Yellow: $0.4 \leq r^2 < 0.6$; Orange: $0.6 \leq r^2 < 0.8$; Red: $r^2 \geq 0.8$.

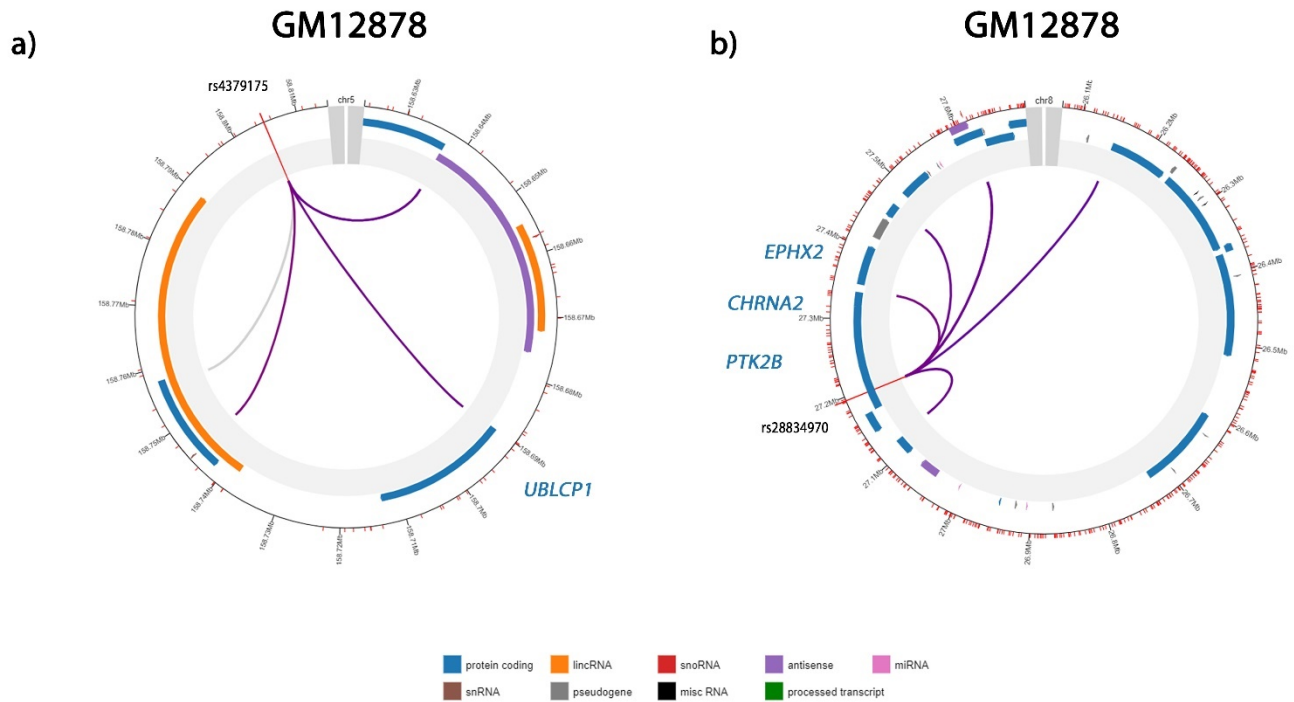


Figure S5. Circular view of the chromatic interaction between the most significant variants within *IL12B* (a) and *PTK2B* (b) and genes promoter in GM12878 cells obtained from Capture Hi-C plotter webtool. The color of the lines represents the confidence for the interactions with a range from blue to red. Grey represents those interactions that are below statistical threshold. Genes which have been also described as eQTLs are highlighted.

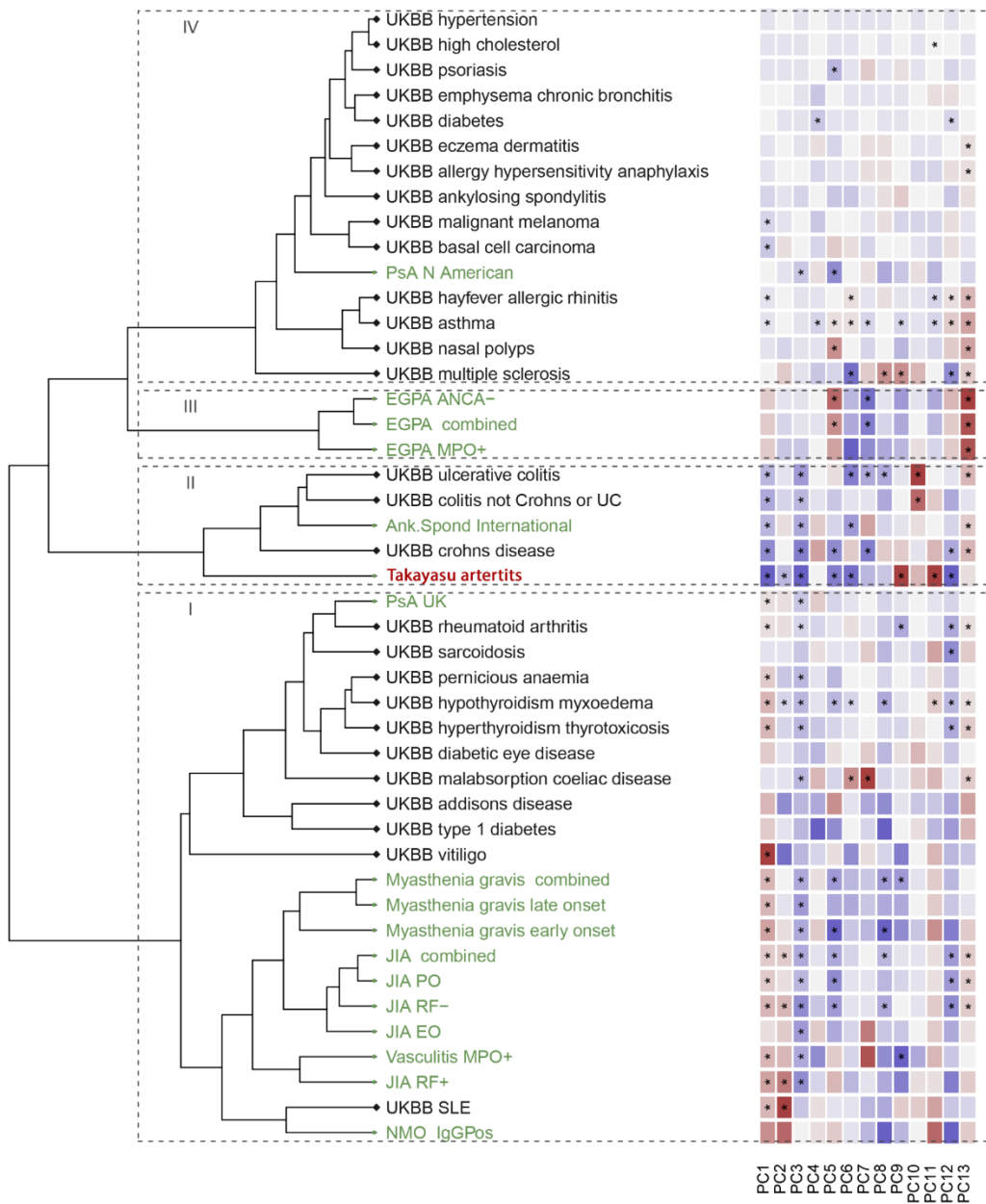


Figure S6. Hierarchical clustering of Takayasu arteritis with other immune-mediated diseases GWAS and self-reported disease traits in UK Biobank considered significantly different from control, shows that Takayasu arteritis clusters with inflammatory bowel disease and ankylosing spondylitis. Colored labels are used to distinguish, our Takayasu arteritis data (red), UKBB (grey) and other GWAS (green) datasets. Roman numerals indicate clusters. Heatmaps indicate delta values for each disease on each component PC1-PC13, with grey indicating 0 (no difference from control), and darker shades of blue or magenta showing departure from controls in one direction or the other. An overlaid * indicates delta was significantly non-zero (FDR < 0.01). Abbreviations: PsA, Psoriatic arthritis; EGPA, Eosinophilic granulomatosis with polyangiitis; ANCA, anti-neutrophil cytoplasmic antibody negative; MPO+, myeloperoxidase positive; Ank. Spond, ankylosing spondylitis; JIA, juvenile idiopathic arthritis; PO, persistent oligo; EO, extended oligo; RF+/-, rheumatoid factor positive/negative; NMO IgGPos, neuromyelitis optica IgG positive.

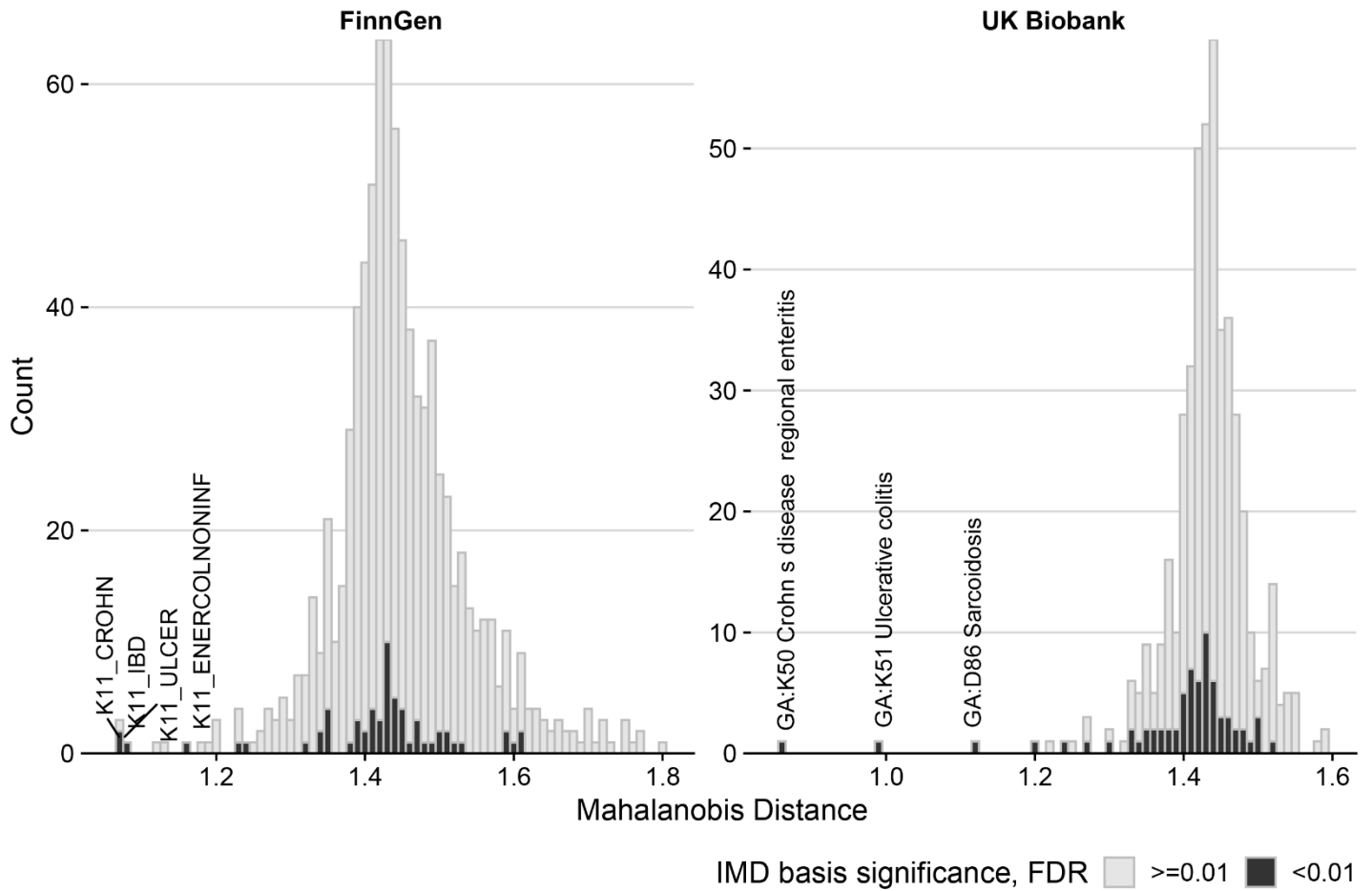


Figure S7. The Mahalanobis distances between Takayasu arteritis and 846 disease traits in FinnGen and 467 ICD codes in UK Biobank. Bars are shaded to highlight traits that were significantly different from control in the lower dimensional representation and are predominantly from the immune-mediated disease subsets of these broad collections of disease traits. The significant traits with the shortest distance to Takayasu arteritis are labelled by their trait name. K11_CROHN, K11_IBD, K11_ULCER and K11_ENERCOLNONINF represent, respectively, Crohn's disease, inflammatory bowel disease, ulcerative colitis and "Noninfective enteritis and colitis".