Analysis of 47 non-MHC Ankylosing Spondylitis Susceptibility Loci Reveals Shared Associated Variants across Caucasians and Han Chinese

Xuqi Zheng, Qiuxia Li, Xiaomin Li, Yanli Zhang, Xinyu Wu, Qiujing Wei, Shuangyan Cao, Mingcan Yang, Zhiming Lin, Zetao Liao, Jun Qi, Qing Lv, Ling Wang, Yi Li, Astrid Irwanto, Ching-Yu Cheng, Xiaoran Chai , Chiea Chuen Khor , Chew-Kiat Heng, Woon-Puay Koh, Jian-Min Yuan, Jinxin Bei, Furen Zhang, Xuejun Zhang, Yixin Zeng, Yan Shen, Jianjun Liu *, Jieruo Gu *

Key Indexing Terms: Ankylosing Spondylitis; Susceptibility loci; Han Chinese; Ethnic heterogeneity; Gene polymorphism

*From the Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-sen Universiy; Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR);the Jackson Laboratory,USA; the Singapore Eye Research Institute (SERI);the Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), the Health Systems and Services Research, Duke-NUS Medical School; the department of Ophthalmology, the department of Paediatrics,Yong Loo Lin School of Medicine;the Khoo Teck Puat - National University Children's Medical Institute; the Saw Swee Hock School of Public Health, National University of Singapore; the division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, the department of Epidemiology, Graduate School of Public Health, University of Pittsburgh; the State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center;the Shandong Provincial Institute of Dermatology and Venereology & Hospital for Skin Diseases, Shandong First Medical University & Shandong Academy of Medical Sciences; the National Clinical Key Project of Dermatology and Venereology;the Institute of Dermatology and Department of Dermatology, Huashan Hospital of Fudan University;the Institute of Dermatology and Department of Dermatology, No.1 Hospital, Anhui Medical University; the Key Laboratory of Dermatology, Ministry of Education (Anhui Medical University);the National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College.*

*Supported by the Science and Technology Planning Project of Guangdong Province,China (2016A020216013),the Guangdong Natural Science Funds for Distinguished Young Scholar (Grant No.2014A030306039), the Municipal Healthcare Joint-Innovation Major Project of Guangzhou and High-level personnel of special support program for Technology Innovative Talents and the Top Young of Guangdong Province (Grant No.2015TQ01R516),the National Supercomputer Center in Guangzhou and the Agency of Science, Technology and Research (A*STAR) of Singapore, the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002), the U.S. National Institutes of Health (Grant Numbers R01CA144034 and UM1 CA182876), the Singapore National Medical Research Council(Grant Number 1270/2010), the 5010 Project of Sun Yat-sen University(Grant No.2007023), Distinguished Young Scholar Candidates Programme for The Third Affiliated Hospital of Sun Yat-Sen University and Pearl River Nova Program of Guangzhou (Grant No. 201610010005).*

Conflict of interest: No.

Authors *X. Zheng, PhD, Q. Li, PhD, X. Li, PhD, Y. Zhang, PhD, X. Wu, PhD, Q Wei, BD, S. Cao, BD, M. Yang, PhD, Z. Lin, PhD, Z. Liao, PhD, J. Qi, PhD, Q. Lv, PhD, Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-sen Universiy; L. Wang,MD, Y. Li, PhD, A. Irwanto, PhD, Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR); Y. Li, PhD,The Jackson Laboratory, USA; C. Cheng, PhD,Singapore Eye Research Institute (SERI), Singapore National Eye Centre, Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Department of Ophthalmology, Yong Loo Lin School of Medicine; X. Chai, PhD,Singapore Eye Research Institute (SERI), Singapore National Eye Centre; C. Khor, PhD, Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR); C. Heng, PhD, Department of Paediatrics, Yong Loo Lin School of Medicine, Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore; W. Koh, PhD, Saw Swee Hock School of Public Health, National University of Singapore;Health Systems and Services Research, Duke-NUS Medical School Singapore; J. Yuan, PhD,Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh;* J. Bei, *PhD, State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China; F. Zhang, PhD, Shandong Provincial Institute of Dermatology and Venereology & Hospital for Skin Diseases, Shandong First Medical University & Shandong Academy of Medical Sciences; X. Zhang, PhD, Institute of Dermatology and Department of Dermatology, Huashan Hospital of Fudan University, Institute of Dermatology and Department of Dermatology at No.1 Hospital, Anhui Medical University,Key Laboratory of Dermatology, Ministry of Education (Anhui Medical University);* Y. Zeng, *PhD, State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China; Y. Shen, PhD, National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College;* J. Liu, *PhD, Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), The Third Affiliated Hospital of Sun Yat-sen Universiy; J Gu,MD, Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-sen Universiy.*

*Address Correspondence to Prof. J Gu, Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-sen Universiy, Guangzhou 510630, China; Email:gujieruo@163.com or Prof. J. Liu, Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore 138672; Email: liuj3@gis.a-star.edu.sg*

J.Liu and J. Gu jointly directed this work. X. Zheng and Q.Li contributed equally to this study.

Manuscript word count: 3151
Table count: 4

ABSTRACT

Objectives: We aimed to present a systematic evaluation of 47 non Major Histocompatibility Complex (MHC) Ankylosing Spondylitis (AS) susceptibility loci which have been initially discovered through Caucasian Genome-wide association studies (GWASs) in Han Chinese.

Methods: Totally 10,743 samples representing north and south Chinese in four datasets were obtained. After data quality control and imputation, meta-analysis results of 94,621 variants within 47 loci were extracted. Four *ERAP1* single-nucleotide polymorphisms (SNPs) and *HLA-B27* tag SNP rs13202464 were used for interaction analysis. Population-attributable risk percentages (PARPs) of AS-associated variants were compared. Functional annotation of AS-associated variants were conducted using HaploReg, RegulomeDB and rVarBase Database.

Results: We revealed 16 AS-associated variants with nominal evidence in Han Chinese, including rs10865331 (P=6.30×10^{-10}), rs10050860 (P=4.09×10^{-5}) and rs8070463 (P=1.03×10^{-4}). Potential susceptible SNPs within these 47 loci were also identified, such as rs13024541 (2p15), rs17401719 (5q15) and rs62074054 (17q21). Epistatic ineractions between three *ERAP1* SNPs (rs17401719, rs30187 and rs10050860) and *HLA-B27* were confirmed. Among the 16 AS-associated variants, rs30187 showed weaker risk effect while rs10050860 and rs12504282 seemed to attribute more risk in Han Chinese than Caucasians. Further genomic annotation pinpointed 35 candidate functional SNPs, especially in 2p15, *ERAP1* and *NPEPPS-TBKBP1* region.

Conclusions: Our results provided a detailed spectrum of all the reported non-MHC AS susceptibility loci in Han Chinese, which comprehensively exhibited the ethnic heterogeneity of AS susceptibility and highlighted that 2p15, *ERAP1* and *NPEPPS-TBKBP1* region may play a critical role in AS pathogenesis across diverse populations.

Abstract word count: 239

INTRODUCTION

Ankylosing spondylitis (AS) is a chronic inflammatory disease predominantly involving axial skeleton and entheseal insertion sites, with a prevalence of 0.2% in Asians[1-3]. The mechanism underlying AS remains to be elucidated, while genetic factors are known to play a prominent role in AS pathogenesis[4].

In the past decade, the understanding of AS genetic factors has been greatly improved by the application of genome-wide association studies (GWASs). The first genome-wide scan study identified two important AS susceptibility genes: *IL-23R* and *ERAP1*[5]. *ERAP1* association with AS was then found to be restricted in *HLA-B27* positive cases[6] . Subsequently, a variety of risk variants related to innate and adaptive immunity as well as inflammatory responses were identified through an Immunochip study[7]. Recently, a large-scale study using the Immunochip data with newly added controls revealed 17 novel AS susceptibility loci[8]. To date, over 40 non-MHC AS susceptibility loci have been identified at genome-wide significance in Caucasians, providing valuable insights into the underlysing disease mechanism[5-10].

Researches in Chinese population on the AS susceptibility loci were less informative. Our group performed the first AS GWAS in Han Chinese which analyzed 1,356,350 single-nucleotide polymorphisms (SNPs) in 1,837 cases and 4,231 controls in the discovery stage and confirmed the 2p15 association at a genome-wide level[11]. More recently, another genome-wide dataset by Immunochip was reported, comprising of 128,935 SNPs in 1,550 cases and 1,567 controls of East Asians. While suggestive associations were seen within several loci, including 2p15 and *ERAP1*, some known loci were not significant, suggesting potential ethnic heterogeneity of AS susceptibility across Caucasians and Han Chinese[7]. In addition, candidate gene studies have been conducted in Han Chinese, and a low-frequency SNP in *IL23R* as well as SNPs in or near *ERAP1*, *CARD9*, *MMP9* and *RUNX3* were found to be associated with AS[12-17]. However, few studies examined the frequency distributions between north and south Chinese, which may be a confounding factor[18]. Furthermore, most studies had small sample size, and the results were often not consistent.

In order to systematically evaluate all the known non-MHC AS susceptibility loci in Chinese population, we augmented the power of our previously published GWAS dataset by increasing the genome-wide coverage of genetic variants through imputation and including a big number of additional control samples and also investigated the frequencies of these risk variants in north and south Chinese individuals.

MATERIALS AND METHODS

Subjects. All cases met the 1984 modified New York criteria for AS, consisting of 968 individuals genotyped using HumanHap 610-Quad Bead chip and 997 individuals by OmniExpress Beadchip as previously reported[11]. Genotype data of 9,479 controls were accordingly derived from a series

of published GWAS, including 5 new datasets (Supplementary table 1) [18-25]. All healthy controls were screened through questionnaires to exclude those with diagnosed AS. All participants were self-reported as Chinese descent. This study was conducted according to the Declaration of Helsinki. This study was approved by the institutional review board at the Third Affiliated Hospital of Sun Yat-sen University (approval number:SYSU3-[2007]40) and informed consent was obtained from all individuals before blood samples were collected.

Study design and data analysis of GWAS datasets. We first performed quality control(QC) on the genotype data in 2 platforms separately. Samples in each platform were then divided into north and south datasets, using Principal Component Analysis (PCA), to better dissect the Chinese population genetic architecture. After imputation and association analysis in each dataset, the meta-analysis result got a broader coverage of variants. A flowchart is presented in Supplementary figure 1.

A primary QC in each dataset were performed in PLINK v1.07 to remove sample with call rate <0.96 or discrepancies between estimated and recorded gender, and exclude SNPs with call rate <0.90, with minor allele frequency (MAF)<0.01 or deviating from Hardy-Weinberg equilibrium (HWE) in controls (P <1.0×10^{-8}). After combining datasets in 2 platforms respectively, 451,984 SNPs and 532,418 SNPs passed stringent QC criteria of call rate <0.96, MAF<0.01,HWE deviation in controls (P <1.0×10^{-6}). Samples having heterozygosity >3 s.d. from the mean value of all individuals, or cryptic relatedness or being population outliers on PCA with 206 Hapmap samples were eliminated. PCA analyses were implemented with an in-house R script to remove outliners after dividing samples into north and south cohorts according to PCA plots (Supplementary figure 2). Phasing was then conducted using SHAPEIT version 2 in each cohort, followed by whole-genome imputation implemented by IMPUTE version 2, using the 1000 Genomes Project Phase 3 data (Nov 2014) as reference panel. After imputation, only SNPs with imputation certainty>=80%, MAF >=0.5% and without significant deviation from HWE in controls (P <1.0×10^{-6}) in all four cohorts were included. To achieve a reliable imputation quality, any SNP genotype with an imputed possibility <0.9 was classified as missing, and SNPs with missing rate >0.1 were excluded in subsequent association analysis. Finally, variants located within 500kb up/downstream of AS-associated SNPs within 47 reported non-MHC loci were included for the current study. The regional association plots for each locus were generated using LocusZoom (http://locuszoom.sph.umich.edu/).

Linkage Disequilibrium(LD) pattern and Interaction analysis. Five genotyped SNPs in or around *ERAP1* with a p-value less than 0.05 were extracted. And two imputed SNPs, a reported AS-associated SNP (rs10045403, $P = 9.65\times10^{-3}$, OR =1.14) and the most significant SNP in our dataset (rs17401719, $P = 4.21\times10^{-5}$, OR = 1.32), were also included using best-guess genotype. LD analysis of these 7 SNPs was then implemented in Haploview software 4.2, using the default algorithm Confidence Interval to define LD Blocks. After excluding SNPs in high/moderate LD, four SNPs (rs17401719, rs30187, rs10050860 and rs10045403) were tested for interaction with the genotyped *HLA-B27* tag SNP rs13202464 under a regression model in each cohort using R version 3.4.2. and the meta-analysis result was finally presented.

Statistical Analysis. Association analysis of both dosage-format genotyped SNPs and imputed SNPs was implemented using SNPTEST version 2.5.2 in each dataset, using the first one or two PCs as covariates to get a minimum λ_{GC} of genotyped variants. Fixed-effect meta-analysis was performed

using META version 1.5, where p-values from Cochran's Q statistics and the I2 heterogeneity were obtained. The quantile-quantile (Q-Q) plot and genomic inflation factor (λ_{GC}) were generated to observe the potential population stratification using R. The phenotypic variance explained by the input variants was computed as the pseudo R^2 in logistic regression model using an in-home R script. Statistical power calculation was implemented using the Genetic Association Study Power Calculator (http://csg.sph.umich.edu/abecasis/gas_power_calculator/index.html) based on the odds ratios reported in Caucasians with a two-sided type I error rate of 0.05, assuming the prevalence of AS in China is 0.2%. The population-attributable risk percentages (PARPs) was calculated using the formula RAF(OR-1)/[RAF(OR-1)+ 1]× 100%[26] .

Functional annotation. To pinpoint the potential causal variants, the 16 AS-associated SNPs and their proxy SNPs (r^2 >=0.80 in Asian population of 1000 Genome Project) were extracted from the HaploReg database version 4.1(HaploReg v4.1) [27]. Coding SNPs were then annotated through the ensemble database. For non-coding variants, the annotation strategy was as followed: (1) The regulatory consequence of all candidate SNPs were estimated based on the functionality score using RegulomeDB version 1.1 which integrates public datasets from GEO, the ENCODE project, and published literatures[28]. (2)We then searched through the HaploReg v4.1 to find out whether these potential functional SNPs overlap with promoter/enhancer histone modification, deoxyribonuclease (DNase) peaks or protein binding sites. (3)The rVarBase database (the 2.0 version of rSNPBase) providing experimentally validated regulatory information was used for chromatin state and overlapped rSNPs/rCNVs annotation[29].

RESULTS

After stringent QC and PCA analysis, the genome-wide genotype data were available in 1,860 AS cases and 8,883 controls, comprising of 4 datasets representing Chinese north and south population. Fixed-effect meta-analysis results of 5,822,845 non-MHC variants were obtained. The λ_{GC} of all these non-MHC variants was 1.033 (supplementary figure 3), suggesting that the population stratification was well controlled. Results of 94,621 variants within 47 reported non-MHC AS susceptibility loci in Caucasians were extracted for subsequent analyses.

Association analysis of reported AS loci in Han Chinese. Among 64 AS-associated SNPs within 47 non-MHC AS loci initially reported in Caucasians, 51 SNPs passed QC for association testing. Totally 16 SNPs with the same risk allele in 13 loci showed consistent association at nominal significance (Table 1, supplementary figure 4). The most significant SNP was rs10865331 (2p15,OR=1.26,P=6.30×10^{-10}) whose p-value reached the genome-wide significance level, which was followed by rs10050860 in *ERAP1* (5q15,OR=1.42, P=4.09×10^{-5}) and rs8070463 near *TBKBP1* (17q21,OR=1.16, P=1.03×10^{-4}). No statistically significant association (P <0.05) was seen in the other 35 reported SNPs, although it was noticable that all of them showed the same direction of association effect between Caucasian and Han Chinese (Supplementary Table 2).

We then calculated the power of our study for detecting the associations at these 51 available SNPs, based on their reported odds ratios of published Caucasian GWAS studies and allele frequency in Chinese population (as revealed by our current study). The power ranged from 46.0%

to 100.0% for the 16 SNPs with significant association and from 22.5% to 98.5% for the remaining SNPs (Table 1,Supplementary Table 2).

In addition to those reported variants, 94,570 additional SNPs within these loci were also tested for association with AS (Supplementary table 3a). Among them, SNP rs13024541 (2p15,P=3.12×10^{-10}), rs17401719 (5q15,P=4.21×10^{-5}) and rs62074054 (17q21, P=2.7×10^{-4}) had moderate to strong LD with the previously reported variants and showed more significant association evidence in our study (Supplementary table 3b). Additional suggestive associations were also detected within these loci and are independent of the reported variants. These potential novel associations were worth further investigation.

Overall, 16 variants were found to show significant association in Han Chinese. They accounted for 2.50% of phenotypic variance of AS in our datasets, with rs10865331 located in 2p15 (0.90%) and rs10050860 in *ERAP1* (0.30%) as the top two variants with highest contribution.

HLA-B27-ERAP1 interaction. We next performed the *HLA-B27-ERAP1* interaction analysis. Among seven SNPs around *ERAP1* region, SNP rs42398, rs30187 and rs27434 are in high LD (Supplementary figure 5), and as a result, 4 independent SNPs (rs17401719, rs30187, rs10050860 and rs10045403) were chosen to do the interaction analysis with rs13202464, the tag SNP for *HLA-B27*.

Strong evidence of interaction with rs13202464 was seen at rs30187 ($P\_interaction$ =7.08×10^{-4}, P_het=0.35) and rs10045403 ($P\_interaction$ =4.93×10^{-4}, P_het=0.97). The low-frequency variant, rs10050860, also exhibited interaction evidence with rs13202464 ($P\_interaction$ =9.8×10^{-3}, P_het=0.55). Consistent with the findings in Caucasians, the stratified analysis showed that all the three SNPs only showed significant association in *HLA-B27* positive cases and controls, but not in *HLA-B27* negative samples (Table 2).

Population risk comparison. As shown in table 3, the PARP values of four SNPs, including rs10510607, rs10440635, rs30187 and rs2297518, were significantly higher in Caucasians than Han Chinese. It was worth noting that rs30187 in *ERAP1* exhibited significantly higher odds ratio but lower allele frequency in Caucasians. Two variants, rs10050860 in *ERAP1* and rs12504282 in *ANTXR2*, attributed more risk in Han Chinese, due to their higher odds ratios as well as higher risk allele frequency distribution. Despite of differences in their risk allele frequencies and odds ratios, most results of these 16 variants were comparable.

Functional annotation of non-MHC variants. Finally, systematic genomic annotation of these 16 SNPs and 245 proxy SNPs (r^2>=0.8 in Asians) were conducted to explore their functional consequences. Totally 13 SNPs were missense or synonymous variants: rs6672420 (I18N) in *RUNX3*, rs6897932 (T244I) in *IL7R*, rs3742704 (I231L) in *GPR65*, rs2297518 (S608L) in *NOS2* and nine SNPs in *ERAP1*: rs27044 (Q730E), rs30187 (K528R), rs26653 (R127P), rs17482078 (R725Q), rs10050860 (D575N), rs2287987 (M349V), rs469783 (A637A), rs27529 (S453S), rs27434 (A356A). On the other hand, 22 SNPs within 7 loci exhibited multiple layers of ENCODE data, with all of them having expression quantitative trait loci (eQTLs) or DNaseI sensitivity Quantitative Trait Loci (dsQTLs) effects, intersecting with TF binding or DNase peak and displayed enhancer/transcription/active TSS chromatin state in blood (Table 4). For 2p15, rs13001372 was found to locate within a DNase I sensitivity quantitative trait locus, and may interact with various

proteins in lymphoblastoid cells. For other SNPs, they were found to affect binding and be linked to gene expression including *ERAP1*, *ACTA2* and *NPEPPS-TBKBP1*, highlighting their functional involvement in AS pathogenesis.

DISCUSSION

By augmenting our previously published GWAS dataset through increasing the genome-wide coverage of genetic variants after imputation and including a big number of additional control samples, we have done a comprehensive evaluation of the 47 non-MHC AS susceptibility loci in Chinese population. In addition to identify additional AS susceptibility loci in Chinese population, our study has revealed shared susceptibility loci between Han Chinese and Caucasian populations as well as potential population-specific loci, suggesting some extent of the ethnic heterogeity of AS genetic susceptibility between two populations.

Ethnic heterogeneity has always been an interesting topic in disease susceptibility studies. While most of common risk variants are shared across different populations, ethinc heterogeneity has been demonstrated and can be categorized into two types of heterogeneity: (i) different populations have different variants within shared loci (allelic heterogeneity); and (ii) ethnic specific loci, which only exhibits significant association evidence with specific population (locus heterogeneity) [23-25]. In this study, we have confirmed 16 susceptible variants in Chinese population and further showed their consistent association effects between Caucasians and Han Chinese (Table 1). Among them, only the 2p15 locus achieved the genome-wide significance, which was consistent with the Immunochip result of East Asians (OR=1.28,P=1.6×10^{-6}) [7]. It showed a strong effect as a common variant and has only been reported to associate with AS and IBD, indicating its vital role in AS pathogenesis besides *HLA-B* locus[26]. We also noticed that the association result of rs9901869 was in concordance with the Immunochip study (rs9901869: OR=1.18, P=2.0×10^{-3}, result of rs8070463 not available) [7]. Further multi-center meta-analysis may achieve the genome-wide significance of these suggestive variants, particularly in *ERAP1* and *TBKBP1* region, to better elucidate the AS susceptibility in Han Chinese. In addition, our study has also demonstrated that although the sample size of our current study provided sufficient statistical power for detection (>80%), some known loci, such as rs11190133 near *NKX2-3,* did not show any evidence of association in Chinese population (Supplementary table 2). Ethnic specific loci may help explain why Caucasians are more susceptible to AS. However, for the other variants showed the same effect direction across populations but with insufficient power, lack of evidence of statistical significance may due to small sample size. Furthermore, our study has also suggested some novel risk variants within the reported loci in Chinese population that are independent from the ones reported in Caucasians (Supplementary table 3a). Although the evidences at these novel variants are not statistically significant, these variants, such as rs650854 near *HHAT*, did show consistant association effect and frequency across 4 independent datasets. Futher studies with bigger sample size were needed to replicate these novel variants and confirmed the allelic heterogeneity of these loci across populations.

Meanwhile, we performed the *HLA-B27-ERAP1* interaction analysis. To date, their interaction study in Han Chinese is very few[27], and none included the reported *HLA-B27*-interacting SNPs (rs30187, rs10050860)[6]. With the largest number of Han Chinese individuals investigated so far, our results demonstrated strong evidence of the interaction and confirmed that the reported *ERAP1*

SNPs only showed association with AS in *HLA-B27* positive individuals (Table 2). Nevertheless, we noticed that the detection power of these SNPs in *HLA-B27* negative samples were 33.7%~57.0%, indicating that bigger sample size is still needed to confirm the lack of association in *HLA-B27* negative samples..

Besides genetic associations, those 16 variatns were used for PARP comparison. Most of the results were similar, while notable PARP differcences of rs30187 and rs10050860 were observed. Generally, higher RAF and/or OR get a higher PARP. SNP rs30187 has a lower RAF in Caucasians, but with a markedly higher OR, rs30187 attributed more risk in Caucasians. Conversely, the PARP of rs10050860 was higher in Han Chinese. This may explain part of the higher prevalence of AS in Caucasians, and whether this difference is related to the underlying mechanism is unknown.

Our function annotation analysis pinpointed 13 coding variants within the known risk loci. Of note, most of them located in gene *ERAP1*. As known, *ERAP1* encodes endoplasmic reticulum aminopeptidase 1 which trims peptides to appropriate lengths for MHC class I binding. And epistatic association between *ERAP1* and *HLA-C* was identified in psoriasis[28]. Previous study demonstrated that SNP rs30187 (R528K) can alter the expression levels of *B*27:05*-bound peptidomes and their structural features[29], while some nonsynonymous polymorphisms in *ERAP1* can influence its expression[30], accounting for its genetic association with AS. The non-coding annotation result highlighted that 2p15 and *NPEPPS-TBKBP1* region may also participate AS pathogenesis through regulatory mechanism. Recently sequencing on 2p15 was performed to identify causal variants but no definite coding variant was identified[31]. We found that the annotated proxy rs13001372 was a dsQTL in lymphoblastoid cell line (Table 4), and furthermore, the ChIP-seq data from RegulomeDB database indicated that this SNP can bind to transcription factor IRF4 which regulates CD8+ T cell differentiation, expansion and metabolism[32]. As potential influential contributors to phenotypic variation, how did this dsQTL get involved in AS pathogenesis remains unknown. On the other hand, proxies of rs9901869 had eQTL effect to *TBKBP1* and TBKBP1 is part of the network in TNF/NF-κB pathway which play pivotal role in AS. Recently, variants around *NPEPPS-TBKBP1-TBX21* region were found to influence T-bet expression and possibly IL-17[33]. Fine-mapping of this locus would help elucidate how these variants are engaged in AS pathogenesis.

As a limitation, with the exception of the 2p15, the association evidences were moderate for most of variants with nominal significance and did not survive after Bonferroni correction. And rare variants were not included in association testing, largely due to the limited power of our current study for detecting genetic association effects at rare variants. However, it is noticeable that vast majority of the variants analyzed in our current study showed consistent associations, suggesting the homogeneity nature of AS genetic susceptibility between these two populations.

In the future, independent large-scale studies in Han Chinese are needed. Adding the shared AS-associated variants into AS genetic predictive profiling besides HLA-B27 may help improve AS/Spondyloarthritis risk prediction, especially in HLA-B27 negative cases. It is also engaging to evaluate whether the combined genetic profiling could distinguish AS patients in chronic back pain cohorts. Meanwhile, how these susceptibility variants participate in the inflammation and bone formation mechanism underlying AS is worth further molecular investigation, which may help reveal the crucial pathway for targeted therapy. Last but not least, investigation on the susceptibility variants related to any specific manifestation of AS in Han Chinese, such as acute anterior uveitis[34], disease progression and severity[35] in addition to therapy response, will be of great interest and meaningful for clinical decision making.

In summary, the comprehensive association analyses of 47 non-MHC AS susceptibility loci were carried out in 10,743 Han Chinese individuals. We revealed 16 AS-associated variants and confirmed the *HLA-B27-ERAP1* interaction in Han Chinese. Our findings delineated the ethnic heterogeneity and homogeneity of genetic susceptibility across populations and highlighted that 2p15, *ERAP1* and *NPEPPS-TBKBP1* may play an essential role in AS pathogenesis.

ACKNOWLEDGEMENT

We thank all subjects for their participation in this study. We acknowledge E. Shyong Tai, Xueling Sim and Rajkumar Dorajoo for contributing genotype data of the Singapore Prospective Study Program (SP2) for the analysis.

SUPPLEMENTARY MATERIALS

Figure S1: Flowchart of study design and data analysis, Figure S2: Principal Component Analysis (PCA) plots of samples in two genotyped platforms and four cohorts of north and south China, Figure S3: Quantile-Quantile plot of observed P values and Manhattan plot for whole-genome associations, Figure S4: Regional association plots and linkage disequilibrium patterns of the 16 AS-associated SNPs in 13 reported non-MHC loci, Figure S5: The linkage disequilibrium haplotype blocks of 7 SNPs in *ERAP1* region; Table S1: Number of study subjects, Table S2: Association results of 64 reported SNPs with AS susceptibility across Europeans and Han Chinese, Table S3: Potential AS-associated non-MHC variants within reported loci.

REFERENCES

1. Taurog JD, Chhabra A, Colbert RA. Ankylosing Spondylitis and Axial Spondyloarthritis. New Engl J Med 2016;374:2563-74

2. Dean LE, Jones GT, MacDonald AG, Downham C, Sturrock RD, Macfarlane GJ. Global prevalence of ankylosing spondylitis. Rheumatology 2014;53:650-7

3. Liao ZT, Pan YF, Huang JL, Huang F, Chi WJ, Zhang KX, et al. An epidemiological survey of low back pain and axial spondyloarthritis in a Chinese Han population.Scand J Rheumatol 2009:1-5

4. Ranganathan V, Gracey E, Brown MA, Inman RD, Haroon N. Pathogenesis of ankylosing spondylitis - recent advances and future directions. Nat Rev Rheumatol 2017;13:359-67

5. Wellcome Trust Case Control C, Australo-Anglo-American Spondylitis C, Burton PR, Clayton DG, Cardon LR, Craddock N, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39:1329-37

6. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between *ERAP1* and *HLA-B27* in ankylosing spondylitis implicates peptide handling in the mechanism for *HLA-B27* in disease susceptibility. Nat Genet 2011;43:761-7

7. International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J, Pointon JP, Robinson PC, Karaderi T, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet 2013;45:730-8

8. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at

shared loci. Nat Genet 2016;48:510-8

9. Australo-Anglo-American Spondyloarthritis C, Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet 2010;42:123-7

10. Karaderi T, Keidel SM, Pointon JJ, Appleton LH, Brown MA, Evans DM, et al. Ankylosing spondylitis is associated with the anthrax toxin receptor 2 gene (ANTXR2). Ann Rheum Dis 2014;73:2054-8

11. Lin Z, Bei JX, Shen M, Li Q, Liao Z, Zhang Y, et al. A genome-wide association study in Han Chinese identifies new susceptibility loci for ankylosing spondylitis. Nat Genet 2011;44:73-7

12. Davidson SI, Liu Y, Danoy PA, Wu X, Thomas GP, Jiang L, et al. Association of STAT3 and TNFRSF1A with ankylosing spondylitis in Han Chinese. Ann Rheum Dis 2011;70:289-92

13. Davidson SI, Jiang L, Cortes A, Wu X, Glazov EA, Donskoi M, et al. Brief report: high-throughput sequencing of IL23R reveals a low-frequency, nonsynonymous single-nucleotide polymorphism that is associated with ankylosing spondylitis in a Han Chinese population. Arthritis Rheum 2013;65:1747-52

14. Li C, Lin Z, Xie Y, Guo Z, Huang J, Wei Q, et al. *ERAP1* is associated with ankylosing spondylitis in Han Chinese. J Rheumatol 2011;38:317-21

15. Ma X, Liu Y, Zhang H, Qiu R, Zhao H, Xin Q, et al. Evidence for genetic association of CARD9 and SNAPC4 with ankylosing spondylitis in a Chinese Han population. J Rheumatol 2014;41:318-24

16. Genetic Polymorphisms of the Matrix Metalloproteinase-3 (MMP-3) and Tissue Inhibitors of Matrix Metalloproteinases-1 (TIMP-1) Modulate the Development of Ankylosing Spondylitis. Ann Rheum Dis. 2009; 68:1781–1786

17. Lian Z, Chai W, Shi LL, Chen C, Liu J, Wang Y. Analysis of PPARGC1B, RUNX3 and *TBKBP1* Polymorphisms in Chinese Han Patients with Ankylosing Spondylitis: A Case-Control Study. Plos One 2013;8:e61527

18. Chen J, Zheng H, Bei JX, Sun L, Jia WH, Li T, et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. Am J Hum Genet 2009;85:775-85

19. Cole P, Macmahon B. Attributable risk percent in case-control studies. Br J Pre Soc Med 1971;25:242-4

20. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:D930-4

21. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012;22:1790-7

22. Guo L, Wang J. rSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks. Nucleic Acids Res 2018;46:D1111-D6

23. Kochi Y, Suzuki A, Yamada R, Yamamoto K. Ethnogenetic heterogeneity of rheumatoid arthritis-implications for pathogenesis. Nat Rev Rheumatol 2010;6:290-5

24. Kim-Howard X, Sun C, Molineros JE, Maiti AK, Chandru H, Adler A, et al. Allelic heterogeneity in NCF2 associated with systemic lupus erythematosus (SLE) susceptibility across four ethnic populations. Hum Mol Genet 2014;23:1656-68

25. Yin X, Low HQ, Wang L, Li Y, Ellinghaus E, Han J, et al. Genome-wide meta-analysis

identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. Nat Commun 2015;6:6916

26. Liu JZ, Sommeren Sv, Huang H, Ng SC, Alberts R. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2016;47:979-86

27. Zhang Z, Dai D, Yu K, Yuan F, Jin J, Ding L, et al. Association of *HLA-B27* and *ERAP1* with ankylosing spondylitis susceptibility in Beijing Han Chinese. Tissue Antigens 2014;83:324-9

28. Genetic Analysis of Psoriasis C, the Wellcome Trust Case Control C, Strange A, Capon F, Spencer CC, Knight J, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and *ERAP1*. Nat Genet 2010;42:985-90

29. Sanz-Bravo A, Campos J, Mazariegos MS, Lopez de Castro JA. Dominant role of the *ERAP1* polymorphism R528K in shaping the *HLA-B27* Peptidome through differential processing determined by multiple peptide residues. Arthritis Rheum 2015;67:692-701

30. Costantino F, Talpin A, Evnouchidou I, Kadi A, Leboime A, Said-Nahal R, et al. *ERAP1* Gene Expression Is Influenced by Nonsynonymous Polymorphisms Associated With Predisposition to Spondyloarthritis. Arthritis Rheum 2015;67:1525-34

31. Wang M, Xin L, Cai G, Zhang X, Yang X, Li X, et al. Pathogenic variants screening in seventeen candidate genes on 2p15 for association with ankylosing spondylitis in a Han Chinese population. Plos One 2017;12:e0177080

32. Vecellio M, Roberts AR, Cohen CJ, Cortes A, Knight JC, Bowness P, et al. The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IRF4 recruitment that alter gene expression. Ann Rheum Dis 2016;75:1534-40

33. Lau MC, Keith P, Costello ME, Bradbury LA, Hollis KA, Thomas R, et al. Genetic association of ankylosing spondylitis with TBX21 influences T-bet and pro-inflammatory cytokine expression in humans and SKG mice as a model of spondyloarthritis. Ann Rheum Dis 2017;76:261-9

34. Robinson PC, Claushuis TA, Cortes A, Martin TM, Evans DM, Leo P, et al. Genetic dissection of acute anterior uveitis reveals similarities and differences in associations observed with ankylosing spondylitis. Arthritis Rheum 2015; 67:140-51

35. Wong R-H, Wei JC-C, Huang C-H, Lee H-S, Chiou S-Y, Lin S-H, et al. Association of IL-12B genetic polymorphism with the susceptibility and disease severity of ankylosing spondylitis. J Rheumatol 2012;39:135-40

Tables

Table 1.Associations of 16 reported SNPs showing nominal significance with AS susceptibility in Han Chinese

Locus	SNP	Position (Build 37)	Nearby Gene(s)	A1/A2	Han Chinese								Power
					RAF(Case/control) in 4 cohorts				OR(95%CI)	*P*	*P_het*	I^2	
1p36	rs6600247	25305114	RUNX3	C/T	0.72/0.71	0.71/0.70	0.73/0.69	0.73/0.71	1.12(1.03~1.21)	7.08×10^{-3}	0.60	0.0	92.5%
2p15	rs10865331	62551472	Intergenic	A/G	0.45/0.41	0.60/0.52	0.49/0.42	0.59/0.53	1.26(1.17~1.35)	6.30×10^{-10}	0.08	55.1	100.0%
3p24	rs10510607	28286261	CMC1	C/T	0.59/0.51	0.58/0.56	0.55/0.55	0.57/0.54	1.10(1.02~1.18)	1.42×10^{-2}	0.27	22.7	94.9%
4q21	rs12504282	80927001	ANTXR2	T/C	0.92/0.91	0.92/0.91	0.92/0.91	0.94/0.91	1.18(1.03~1.34)	1.49×10^{-2}	0.87	0.0	46.0%
5p13	rs10440635	40490790	PTGER4	A/G	0.27/0.25	0.18/0.18	0.27/0.25	0.21/0.18	1.10(1.01~1.20)	3.26×10^{-2}	0.37	5.5	78.0%
5p13	rs11742270	35881443	IL7R	G/A	0.87/0.86	0.86/0.83	0.86/0.86	0.82/0.82	1.11(1.00~1.22)	4.63×10^{-2}	0.42	0.0	48.9%
5q15	rs30187	96124330	*ERAP1*	T/C	0.54/0.52	0.55/0.53	0.55/0.51	0.56/0.52	1.14(1.06~1.23)	3.27×10^{-4}	0.72	0.0	100.0%
5q15	rs10045403	96147733	*ERAP1*	A/G	0.85/0.81	0.84/0.83	0.84/0.82	0.86/0.83	1.14(1.03~1.26)	9.65×10^{-3}	0.67	0.0	91.4%
5q15	rs10050860	96122210	*ERAP1*	C/T	0.96/0.95	0.96/0.95	0.97/0.95	0.97/0.95	1.42(1.19~1.69)	$4.09\times10^{-5*}$	0.47	0.0	39.8%
6q15	rs639575	90991131	BACH2	T/A	0.53/0.50	0.50/0.49	0.53/0.49	0.50/0.50	1.10(1.02~1.18)	9.99×10^{-3}	0.48	0.0	54.0%
10q23	rs1800682	90749963	ACTA2	A/G	0.59/0.61	0.53/0.51	0.63/0.6	0.52/0.54	1.10(1.02~1.18)	1.24×10^{-2}	0.21	33.7	83.5%
14q31	rs11624293	88488821	GPR65	C/T	0.19/0.13	0.18/0.17	0.15/0.14	0.17/0.16	1.11(1.01~1.23)	3.82×10^{-2}	0.02	69.6	98.2%
17q11	rs2297518	26096597	NOS2	A/G	0.17/0.14	0.19/0.16	0.17/0.14	0.16/0.18	1.14(1.03~1.26)	1.13×10^{-2}	0.10	52.9	93.3%
17q21	rs9901869	45575206	*NPEPPS-TBKBP1*	A/G	0.67/0.64	0.63/0.62	0.67/0.62	0.64/0.60	1.15(1.07~1.24)	2.29×10^{-4}	0.48	0.0	92.8%
17q21	rs8070463	45768836	*NPEPPS-TBKBP1*	C/T	0.49/0.45	0.46/0.45	0.49/0.44	0.48/0.45	1.16(1.08~1.25)	1.03×10^{-4}	0.41	0.0	90.2%
21q22	rs2836883	40466744	Intergenic	G/A	0.83/0.77	0.87/0.84	0.80/0.79	0.87/0.85	1.16(1.06~1.28)	2.00×10^{-3}	0.56	0.0	92.5%

A1:risk allele,A2:non-risk allele. RAF: risk allele frequency. *P*_het: *p*-value for the heterogeneity test. Power: detection power in our dataset based on the reported odds ratios (ORs) of published GWAS in Europeans. *: Fisher's exact test.

Table 2. Stratified analysis of *ERAP1-HLA-B27* interaction.

ERAP1 SNPs	Allele	Strata	Additive Model			
			OR	SE	*P*	*P_het*
rs30187	C	rs13202464=AA[a]	1.11	0.10	2.79×10^{-1}	0.74
		rs13202464=GA/GG[b]	0.81	0.06	2.93×10^{-4}	0.42
rs10045403	G	rs13202464=AA[c]	1.11	0.13	4.14×10^{-1}	0.65
		rs13202464=GA/GG[d]	0.80	0.08	4.72×10^{-3}	0.57
rs10050860	T	rs13202464=AA[e]	1.12	0.23	6.54×10^{-1}	0.65
		rs13202464=GA/GG[f]	0.71	0.15	2.40×10^{-2}	0.98

a:sample size=207cases v.s. 7,546controls, b=1,653cases v.s. 1,337controls.

c:sample size=201cases v.s. 7,332controls,d=1,612cases v.s. 1,304controls.

e:sample size=207cases v.s. 7,546controls, f=1,653cases v.s. 1,337controls.

Table 3. Comparison of 16 shared associated non-MHC SNPs across European and Chinese

Locus	SNP	Nearby Gene(s)	Risk allele	RAF (CEU/CHN)	OR (CEU/CHN)	PARP% (CEU/CHN)
1p36	rs6600247	RUNX3	C	0.50/0.69	1.16/1.15	7.41%/9.42%
2p15	rs10865331	Intergenic	A	0.38/0.49	1.34/1.27	11.44%/11.60%
3p24	rs10510607	CMC1	C	0.83/0.54	1.15/1.14	11.07%/7.03%
4q21	rs12504282	ANTXR2	T	0.54/0.91	1.14/1.20	7.03%/15.38%
5p13	rs10440635	PTGER4	A	0.59/0.21	1.13/1.10	7.12%/2.10%
5p13	rs11742270	IL7R	G	0.73/0.84	1.11/1.14	7.43%/10.53%
5q15	rs30187	*ERAP1*	T	0.34/0.53	1.32/1.11	9.81%/5.47%
5q15	rs10045403	*ERAP1*	A	0.73/0.82	1.20/1.18	12.74%/12.92%
5q15	rs10050860	*ERAP1*	C	0.78/0.95	1.18/1.45	12.31%/30.01%
6q15	rs639575	BACH2	T	0.61/0.49	1.08/1.10	4.65%/4.69%
10q23	rs1800682	ACTA2	A	0.54/0.56	1.12/1.10	6.09%/5.22%
14q31	rs11624293	GPR65	C	0.09/0.15	1.23/1.11	1.96%/1.65%
17q11	rs2297518	NOS2	A	0.19/0.16	1.13/1.11	2.41%/1.72%
17q21	rs9901869	*NPEPPS-TBKBP1*	A	0.52/0.62	1.15/1.14	7.24%/7.95%
17q21	rs8070463	*NPEPPS-TBKBP1*	C	0.51/0.44	1.14/1.16	6.66%/6.64%
21q22	rs2836883	Intergenic	G	0.74/0.83	1.19/1.16	12.33%/11.71%

RAF: risk allele frequency in controls. PARPs: population-attributable risk percentages.

Table 4. Annotation information of twenty-two functional proxies of AS-associated variants.

SNP	Annotated gene	Proxies(r2 in ASN)	RegulomeDB		Haploreg v4.1				rVarBase			
			score	QTL	Histone marks*	DNase**	Proteins bound	Motifs changed	Chromatin state***	TF binding	rSNPs	mRNA abundance
rs10865331	-	rs13001372(0.80)	1f	dsQTL	12	16	9	0	37	0	6	-
rs10510607	CMC1	rs10510607(1)	1f	cis-eQTL	0	2	0	4	80	0	2	1
rs30187	*ERAP1*	rs27043(0.83)	1f	cis-eQTL	14	14	2	0	123	35	11	16
rs30187	*ERAP1*	rs469783(0.95)	1f	cis-eQTL	1	1	0	2	123	8	11	12
rs30187	*ERAP1*	rs27710(1.00)	1f	cis-eQTL	7	0	0	2	123	4	11	10
rs10050860	*ERAP1*	rs13170045(0.85)	1f	cis-eQTL	5	1	0	5	123	11	11	10
rs10050860	*ERAP1*	rs10050860(1.00)	1f	cis-eQTL	0	0	0	2	123	0	11	9
rs1800682	ACTA2	rs4934434(0.99)	1d	cis-eQTL	9	7	5	2	71	43	43	7
rs1800682	ACTA2	rs3740286(0.97)	1a	cis-eQTL	24	40	6	6	127	184	46	10
rs1800682	ACTA2	rs1324551(0.97)	1f	cis-eQTL	24	50	7	0	127	160	48	7
rs1800682	ACTA2	rs6586163(0.97)	1f	cis-eQTL	21	4	7	0	114	48	47	7
rs1800682	ACTA2	rs6586164(0.97)	1b	cis-eQTL	21	4	7	0	114	45	46	7
rs1800682	ACTA2	rs3824730(0.97)	1f	cis-eQTL	21	4	6	0	114	25	46	7
rs1800682	ACTA2,FAS	rs7097572(0.97)	1b	cis-eQTL	28	11	3	7	105	34	46	9
rs1800682	ACTA2,FAS	rs1926196(0.97)	1f	cis-eQTL	25	8	1	2	101	48	46	10
rs1800682	ACTA2	rs1926195(0.96)	1f	cis-eQTL	25	9	1	3	101	49	44	7
rs1800682	ACTA2	rs6586165(0.96)	1f	cis-eQTL	15	6	2	1	93	10	43	7
rs11624293	GALC	rs989275(0.91)	1f	cis-eQTL	2	0	0	0	7	0	93	2
rs11624293	GALC	rs989274(0.92)	1f	cis-eQTL	2	0	0	5	7	0	93	2
rs2297518	NOS2	rs2297518(1)	1f	cis-eQTL	1	0	0	2	27	0	4	3
rs9901869	*TBKBP1*	rs9904967(0.83)	1f	cis-eQTL	0	0	0	1	67	0	18	15
rs9901869	*TBKBP1*	rs9895509(0.85)	1f	cis-eQTL	1	2	0	0	106	0	17	16

Regulome DB score: 1d: eQTL + TF binding + any motif + DNase peak; 1f: eQTL + TF binding / DNase peak;

*: number of tissues with reported promoter/Enhancer histone marks evidence of the targeted SNP.

**: number of tissues with reported DNase evidence of the targeted SNP.

***:number of reported evidence about the chromatin state of the surrounding region, e.g. strong/weak transcription, enhancers, flanking active TSS.

-: No available data.

Abbreviations: eQTL:expression quantitative trait locus; dsQTL: DNase I sensitivity quantitative trait loci.rSNP: regulatory single nucleotide polymorphism; SNP: single nucleotide polymorphism.