

Prediction of Inflammatory Bowel Diseases using Genetic Risk Score in Asian populations

Shifteh Abedian (✉ s.abedian@umcg.nl)

University Medical Center Groningen

Sunny H Wong

Chinese University of Hong Kong

Suzanne Van Sommeren

University Medical Center Groningen

Atsushi Takahashi

RIKEN

Jae Hee Cheon

Yonsei University

Ajit Sood

Dayanand Medical college and Hospital

Homayoon Vahedi

Tehran University of Medical Sciences

Keiko Yamazaki

Nihon University

Won Ho Kim

Yonsei University

Thelma B K

Dehli University School of Medicine

Nasser E Daryani

Tehran University of Medical Sciences

Michiaki Kubo

Nihon University

Suk-Kyun Yang

University of Ulsan College of Medicine

Rupa Banerjee

Asian Institute of Gastroenterology

Reza Malekzadeh

Tehran University of Medical Sciences

Rinse K. Weersma

University Medical Center Groningen

Siew C. Ng

Chinese University of Hong Kong

Behrooz Z. Alizadeh

University Medical Center Groningen

Research Article

Keywords: Inflammatory bowel disease, Crohn's Disease, Ulcerative colitis, Genetic risk score, Explained disease susceptibility, Risk prediction, Risk estimate

Posted Date: February 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-153287/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background and Aims:

The incidence of Inflammatory bowel disease (IBD), including Crohn's disease (CD) and Ulcerative colitis (UC), is rising in Asian populations. We undertook a cross-population study to explore whether genetic risk scores (GRS) of IBD, CD and UC could explain their occurrence, and whether they can be used to predict disease occurrence in general populations from East Asia (EA) and Central Asia (CA).

Methods

We studied 9,698 subjects – 4,733 IBD patients (2,003 CD; 2,730 UC) and 4,965 matched controls – who had been genotyped using ImmunoChip. The subjects were from three East Asian (Japan, South Korea and China) and two Central Asian populations (India and Iran). We generated GRS for each population by combining information from up to 201 genome-wide significant IBD-associated variants to summarize the total load of genetic risk for each phenotype. We then estimated the explained variance and predictability of IBD using the GRS.

Results

IBD GRS could explain up to 4.40% and 4.14% of IBD variance at a significant level in East Asian and Central Asian populations, respectively, but, given a prevalence of 0.01% and 0.04% for IBD, these yield limited predictive probability. GRS for CD and UC separately proved less significant than GRS for IBD.

Conclusion

GRS alone can explain only a limited percentage of disease occurrence (< 5% of disease susceptibility) and cannot be used to predict IBD in Asian populations at this time. Our results highlight the significant missing heritability, which may be due to genetic epistasis, gene-environment interactions, or rare variants.

Introduction

Inflammatory bowel disease (IBD) is a chronic, debilitating disease that now affects 2.5 million people of European descent. However, its incidence and prevalence are rising in populations in the newly industrialized countries of Central and East Asia [1]. Non-European populations have distinct phenotypic characteristics for IBD, for instance, Asians with Crohn's disease (CD) and Ulcerative colitis (UC) have a lower proportion of family history and of extra intestinal manifestations compared with European populations. In CD, Asians show a male predominance, more stricturing disease and more perianal involvement than Europeans, whereas in UC, they report lower rates of extensive colitis and colectomy [2].

IBD occurs when the immune system responds inappropriately to gut microbiota in a genetically susceptible host [3]. Genome-wide association (GWAS) and deep-sequencing studies have identified over 240 genetic variants associated with IBD. These studies were largely conducted in individuals of European descent, with only a few in Asian and African-American populations [4, 5], but they provide biological insights into the disease mechanisms [6]. Apart from susceptibility to the disease, several studies have also investigated genetic loci affecting IBD sub-phenotypes, such as disease location and prognosis [7, 8]. In particular, a study of 29,838 patients identified genetic loci associated with disease location, but not with how the disease evolved over time [7]. Our earlier trans-ethnic association study, involving 9,846 subjects of non-European descent, identified 38 novel genetic loci and found genetic heterogeneity between populations [4]. It also demonstrated the importance of trans-ancestry genetic studies.

Recently, genetic risk scores (GRS) have been used to aggregate the contribution of multiple single nucleotide polymorphisms (SNPs) by combining genetic information and testing for improved performance in predicting disease incidence [9]. By using GRS, the genetic overlap and pleiotropic character of IBD sub-phenotypes was qualified [10] and supported a continuum of the disease that was better explained by three disease groups – ileal Crohn’s disease, colonic Crohn’s disease, and Ulcerative colitis – than by the current bipartite classification [7]. Furthermore, several studies have shown that information on multiple SNPs combined into a GRS was associated with complex diseases such as obesity, type 2 diabetes, and coronary heart disease [11- 13]. From another viewpoint, a better understanding of how GRS can be used to predict IBD could improve identification of high-risk individuals for whom preventive interventions could help avoid development of disease. However, studies so far on using GRS did not improve the risk prediction of IBD in the general population, nor in patients [14- 17], although in a European population GRS did yield more information on the genetic background of IBD than candidate SNP associations alone [7]. The validity of GRS in predicting IBD, CD and UC occurrence in Central and East Asian populations has not yet been reported.

In our current study, we investigated how GRS might explain susceptibility to IBD, as measured by the amount of variance explained by GRS for IBD. We also explored the predictive value of GRS for IBD across ancestrally diverse, non-European, general populations. We compared genetic association with IBD disease phenotypes (IBD, CD, or UC), and their predictability in general Asian populations.

Methods

Study Design Our current study built on an earlier trans-ethnic GWAS of 9,846 subjects of non-European descent, in which we identified 38 novel genetic loci for IBD and verified another 173 [3]. Now we went on to generate GRS for IBD, CD and UC phenotypes, and to estimate explained variance and predictability of IBD, CD and UC incidence in Asian populations of Japanese, Korean, Chinese, Indian and Iranian descent.

IBD diagnosis was determined by IBD specialists from reviews of case notes and clinical, radiological, pathological and endoscopic reports [3]. After quality control, we extracted detailed

information on the disease phenotypes for the five studied populations. The original study included patients from population-based registries, and from secondary- and tertiary medical referral centers at multiple locations [3] (Supplementary Text S1).

We now had data on 9,698 participants, including their gender, ethnicity, smoking status, family medical history, clinical and genetic data. We retrieved ImmunoChip array genotypes for 6,395 East Asian and 3,303 Central Asian patients and country-, age- and gender-matched controls. In our current analysis we used three East Asian populations, including 5,317 Japanese (CD 1,312, 723 UC, and 3,282 controls), 547 South Koreans (201 CD, 230 UC, and 114 controls) and 533 Chinese (155 CD, 143 UC, and 235 controls), and two Central Asian populations, including 2,413 Indians (184 CD, 1,237 UC, and 992 controls) and 890 Iranians (151 CD, 397 UC, and 342 controls) (Supplementary Figure S1). Detailed phenotype data were available for at least 74.5% of CD patients and 88.9% of UC patients in the three East Asian populations, and 71% of CD patients and 76% of UC patients of the two Central Asian populations. Data on age at diagnosis, family history, and smoking were available for 82.6%, 82.7% and 61.0% of patients in the East Asian populations, and 79%, 77% and 76% in the Central Asian populations. Data on age of disease onset, extra intestinal manifestations, and surgical history were also collected on 2,557 East Asian, 1,421 Indian, and 548 Iranian IBD patients. All 9,698 participants for the current study had been genotyped on the ImmunoChip array as part of the trans-ethnic IBD genetic consortium (IBDGC) initiative [3]. During quality control we removed individuals with more than 10% of their genotype missing. The genotyping methods and quality control have been explained elsewhere [3]. We selected genotype data on the 201 common IBD-associated SNPs discovered in our earlier study [3]. Genetic data were harmonized by filtering out the genetic variants that were missing in any of the five populations. At the missing SNPs, we first identified the proxy SNPs that were in the highest linkage disequilibrium (LD) of $r^2 > 0.99$ and determined those closest to the GWAS SNPs by using references panel from the 1000 Genomes project. Of the 201 IBD-associated SNPs, 19 had missing information in at least one of the three East Asian populations that could not be replaced by a proxy, yielding 182 common SNPs for final analysis. Of the 201 SNPs, we could not retrieve a proxy for seven in the Indian and Iranian populations, yielding 194 common SNPs.

Ethical considerations The protocol of described study is in line with the ethical guidelines of the 1975 Declaration of Helsinki as reflected in approval by the medical ethical review board of all involved cohorts in the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). All methods were performed in accordance with the relevant guidelines and regulations. The recruitment of study subjects, all methods and protocols were approved by the ethics committees or institutional review boards of all individual participating centers of involved countries in the IBDGC and this study containing: Institute of Medical Science, University of Tokyo, RIKEN Yokohama Institute of Japan, Yonsei University College of Medicine and Asan Medical Centre, Seoul, Korea, Chinese University of Hong Kong, China, Digestive Disease Research Institute, Tehran University of Medical Science, Iran and Department of Medicine, Dayanand Medical college and hospital, Ludhiana, India. Informed consent was obtained from all participants in this study.

Data Analysis Our final analysis was performed on the matched genotype and phenotype data of 4,733 IBD patients (2003 CD; 2,730 UC), and 4,965 country-, age- and gender-matched controls. Per SNP, the risk variant was used to build GRS, firstly to test the explained variance of IBD, CD and UC, and secondly to examine the disease predictability in general populations by implementing a previously applied systematic framework [6].

We fitted a number of genotype-phenotype models using mixed models for the East Asian populations to estimate independent risk and cross-checked these models across each of the three separate East Asian populations. The dataset of the East Asian populations was split into (1) a training set (including two out of the three East Asian populations) to build the model to calculate the odds ratio (OR), and (2) a test set (the third population) for evaluating and validating the predictive model built for the training set. The target population was excluded from the East Asian population and the association of each allele with the risk of the phenotype of interest was studied using the other two remaining populations. To calculate the independent risk for IBD, CD and UC per IBD SNP, we first combined the Korean and Chinese populations (OR_{KC}). Next, we combined the Japanese and Chinese populations (OR_{JC}) and finally, we combined the Japanese and Korean populations (OR_{JK}). We used an additive linear mixed model as implemented in the software package MMM (C-program for analyzing a linear mixed model) [18] to calculate the risk (OR) for each of the 182 common IBD SNPs (Figure 1).

The original risk alleles were defined as alleles associated with an increased risk of IBD, CD or UC in our original trans-ethnic meta-GWAS of IBD [3], these have been replicated in a follow-up study in Caucasians [6]. In brief, we included 201 top SNPs that were associated to IBD to form a genetic dataset and to build a genetic relatedness defined in an R matrix. The R matrix was calculated with the number of variants per phenotype and included as a random-effects component in the model to account for population stratification. The results of case-control association tests were presented as OR with associated p-values for the phenotype of interest. We evaluated the ORs per phenotype and any SNPs with an extreme OR were excluded. Finally, we included 176 IBD risk variants that had OR estimates across the three East Asian populations to build the GRS.

For the Indians and Iranians, we included data for each of the 194 IBD-associated variants. We defined the risk allele as the one obtained for the Caucasian population in the trans-ethnic meta-GWAS of IBD [3]. Likewise, we used the same SNPs and specific ORs estimated for the Caucasian population [3] to build the GRS and to test predictive models in the two CA populations.

Genetic Risk Score (GRS) We built a multi-locus GRS for each patient in the studied population by taking the frequency of a given risk allele per SNP from the controls for our target populations and multiplying it with the natural logarithm of its OR, as estimated in the above procedures. Unweighted GRS were built for each disease phenotype in the three East Asian populations (GRS IBD vs. controls, GRS CD vs. controls, and GRS UC vs. controls) and we thus arrived at nine GRS in 2,763 patients and 3,631 controls that utilized the allelic OR from the MMM model analyses. The models used two populations and the allele frequencies taken from controls of the third target population to account for the strength of the genetic

association in each allele in the target population. We shuffled the three East Asian populations into three settings. We calculated the combined independent risk for IBD, CD and UC in the Korean and Chinese populations (OR_{KC}), and then used these estimates to build GRS for 176 associated SNPs with the phenotype of interest for the Japanese population. Next, we applied the combined independent risk estimate for the Japanese and Chinese populations (OR_{JC}) to calculate the GRS per SNP for IBD, UC and CD in the Koreans. Finally, we combined the Japanese and Korean populations to calculate the independent risk estimate (OR_{JK}) per SNP (Supplementary Figure S2).

For the Indians and Iranians, we implemented the same procedures using ORs of the 194 associated IBD SNPs, as defined above. Genetic risk scores were calculated for each population using the R package “Mangrove” (See Web link).

Explained Variance and Predictive Analyses We estimated the explained variance (disease susceptibility) for IBD and its phenotypes’ risk alleles in the five populations. Mangrove holds the risk alleles, effect sizes (β values) and frequencies (f) for a set of genetic variants (i.e. 176 for East Asians or 194 for Central Asians) relevant to predicting a phenotype. It calculates the variance explained analytically, by converting the OR to liability scale units (i.e. the genetic risk variants included in the model) and adding them together (Figure 2). It gives the variance explained by the variants included in the model and plots the cumulative variance explained as the variants are added in one-at-a-time (in order of most to least variance explained). The distribution of predicted risks in patients was then compared to controls using the Wilcox rank sum test. Given the prevalence of IBD, CD and UC for each target population, we calculated the posterior probability of disease incidence for each phenotype in the target Asian populations.

Risk Prediction in Unrelated Individuals To predict IBD phenotypes from GRS, we considered a matrix of GRS with elements GRS_j for individual j , and a vector of standardized effect sizes $b \sim fb_{GRS}$. Next, the IBD phenotype predictions and probabilities of disease status as a function of GRS were calculated via a logistics link function as $disease\ status = (1 + e^{-(\mu_0 + (\text{meanGRS})^T \times GRS_j - (\text{meanGRS}))})^{-1}$ where mean GRS is a vector of the log odds ratios for GRS, μ_0 (baseline risk) is a function of K representing the prevalence of the IBD phenotype in the target population. Mean GRS was defined as $GRS_j = \sum_{ij} (2\beta_{ij} \times f_{ij})$ which is a normalizing constant accounting for f_i (i.e. the allele frequency), and the effect size β_i for the individual j for a given genetic variant (i.e. each of 176 variants for East Asians or 194 variants for Indo-Iranians). More details of the data modeling and related mathematical equations, are explained elsewhere (see Web link).

Results

Detailed clinical characteristics of the participants are shown in Supplementary table S1. Characteristics of the genetic variants included in the GRS containing mean ORs (\pm SE) and allele frequencies (AF) are presented in Table 1.

Explained Variance Table 1 presents the summary results of the combined ORs relative to the population average for all the participants (i.e. their GRS for IBD, CD and UC). The mean GRS (\pm Standard error (SE)) of IBD was 1.03 (\pm 0.93) in the Japanese, 1.18 (\pm 0.53) in the Koreans, and 1.09 (\pm 0.53) in the Chinese. The mean GRS (\pm SE) of IBD was 1.25 (\pm 55) in the Indians and 1.36 (\pm 58) in the Iranians (Figure 2). The histogram plots show the normal distribution of the predicted risks (ORs) on a log scale (Supplementary Figures S2-S6) for each IBD phenotype per population. The distributions of predicted risks in patients compared to controls show, as expected, that the predicted OR were significantly higher in patients for all the IBD phenotypes in every population (Supplementary Figures S2-S6). As shown in Figure 2, the cumulative variance explained by 176 IBD SNPs was 4.4% for Japanese, 1.5% for Koreans, and 1.34% for Chinese, while the cumulative variance explained by 194 IBD SNPs was 3.81% for Indians and 4.14% for Iranians.

Phenotype Prediction Given a prevalence (prior probability) of 0.08% for IBD in Japanese [1], we estimated a posterior probability of 8.8×10^{-4} after including the GRS in our model, this was not significantly different from the prior probability. This observation also held for the Koreans [19] and Chinese [20], with prior probabilities of 0.04% and 0.009%, respectively, for IBD, and posterior probabilities of 4.7×10^{-4} and 1.05×10^{-4} . However, our calculated GRS for CD and UC explained CD and UC to a lesser significance than for IBD in the Japanese, Koreans and Chinese and gave prior probabilities of 0.02%, 0.01% and 0.001%, respectively, for CD and of 0.06%, 0.03% and 0.008% for UC, respectively. The predictive probabilities yielded a negligible probability of 2.12×10^{-4} , 1.06×10^{-4} and 2.98×10^{-5} for CD and 6.18×10^{-4} , 3.73×10^{-4} and 7.13×10^{-5} for UC in the Japanese, Koreans and Chinese, respectively (Table 1).

Given a prevalence (prior probability) of 0.044% for IBD in Indians [1, 21], we estimated a posterior probability of 5.52×10^{-4} after including the GRS in our model, which was not significantly different from the prior probability. This observation also held for the Iranians [22], with a prior probability of 0.04% for IBD and a posterior probability of 5.4×10^{-5} . The calculated GRS for CD and UC explained CD and UC to a less significant extent than for IBD and gave prior probabilities of 0.002% and 0.005% for CD, and 0.044% and 0.035% for UC, in Indians and Iranians, respectively. The predictive probabilities yielded a negligible probability of 2.11×10^{-5} and 5.8×10^{-5} for CD, and 5.59×10^{-4} and 4.56×10^{-4} for UC, in the Indo-Iranian populations (Table 1 and Supplementary Figures S2-S6).

Discussion

IBD are chronic inflammatory diseases caused by an abnormal immune response towards microorganisms in genetically susceptible individuals. We aimed to understand the variance of IBD as explained by GRS for IBD, CD and UC in Asians. Across five populations, we showed that GRS could significantly explain up to 4.40% of IBD disease variance. The GRS, representing the cumulative effect of 176 risk alleles for IBD in East Asian populations and 194 risk alleles in Central Asians, could significantly explain 1.19 to 4.40% of the variance of IBD, CD and UC in East Asians, and 3.49 to 4.26% in Central Asians, but this yielded a negligible additive predictive probability for IBD, CD and UC in our populations.

The past few decades have witnessed a rapid rise in the incidence of IBD in newly industrialized countries, including those in East and Central Asia [1, 23]. Such increases have highlighted the importance of studying the disease in these geographical areas, particularly for genetic factors that may reveal additional population-specific risk loci. Although environmental triggers are important to causing the disease to develop, an underlying genetic susceptibility is also required [24]. It is known that the coefficient of heritability for siblings of IBD probands is 25 to 42 for CD, and 4 to 15 for UC, and that heritability estimates from pooled twin studies are 0.75 and 0.67, respectively [25- 27]. Furthermore, our earlier trans-ethnic GWAS reported ~220 variants for IBD, but we also highlighted significant genetic heterogeneity between European and non-European populations for the majority of IBD risk loci [3]. However, there are indeed population-specific loci for the disease. For example, a meta-analysis of Asian studies revealed that NOD2 and ATG16L1 were not associated with IBD in many Asian populations [28], whereas a GWAS in Ashkenazi Jewish CD patients identified five novel genetic loci that had not been found in non-Jewish Caucasian populations [29]. Since many genetic variants have small effect sizes and each variant accounts for only a small part of the disease heritability [4], it is now common to use GRS to overcome population-specific effects. GRS summarizes the overall genetic risk across the genome by aggregating information from multiple risk alleles, and this approach is robust to skewed effect sizes due to imperfect linkage or low allele frequencies [30]. Previously we demonstrated the role of GRS in representing the strong association of all known risk alleles for IBD with sub-phenotypes [6]. GRS has been shown to explain disease heritability and it helps to dissect genetic overlap between sub-phenotypes [31, 32].

We found that GRS could explain between 1.19–4.40% of IBD disease variance in Asian populations, and 10% in European populations [6] given that many more IBD-associated SNPs have been identified in European populations. Although the strong associations show an unequivocal genetic component in disease susceptibility in these populations, they only explain a small proportion of disease variance. These percentages are similar to those reported for other common diseases, such as diabetes mellitus (0.4%) [14, 15], coronary heart disease (2.2%) [33, 34], and breast cancer (0.6%) [35]. The disease variance percentages are comparable between CD and UC in our Asian populations, suggesting a similar contribution of the risk alleles to both diseases [36]. Likewise, despite much-anticipated interest, predicting the outcomes may not be achievable with the current data. The small proportion of variance explained reveals the presence of significant missing heritability, which may be due to genetic epistasis, gene-environment interaction, or the presence of unmapped genetic loci and/or rare variants [37- 39]. Importantly, the rapid rise of IBD incidence in Asia points to the importance of a changing environment, pinpointing the possibility of gene-environment interaction in its pathogenesis [40, 41]. In support of this hypothesis, several studies have shown that environmental factors, like smoking or gut microbes, can modify the risks conferred by the major genetic loci of STAT3 [42] and TNFSF15 [43]. The importance of gene-environment interactions in shaping disease susceptibility can best be studied in populations where the epidemiology is changing rapidly [44].

We found that IBD GRS showed little additive value in predicting IBD, CD or UC in the general Asian population. The low predictive proportions may also be attributable to our relatively small sample size

(the predictive ability of a polygenic score can be affected by the sample size [45]). This limitation reflects the need for much larger studies in non-European populations, they will likely yield new genetic loci for IBD. However, even with a relatively small sample size, we were able to replicate several European risk alleles in our Asian populations and showed that a composite weighted score from European risk alleles is still relevant. Another possibility for the ethnic differences is a different environmental burden to the disease. This may be relevant given the fast-evolving environment in Asian countries, which could itself be responsible for the rapidly increasing disease incidence in this region. We conclude there may be a lower genetic risk threshold for IBD in Asians.

The low predictability of GRS we found in Asian populations is not unique in IBD [46]. CD and UC GRS did not predict IBD phenotype or complications in Hispanics or in non-Hispanic Whites, except for indicating a younger age of onset in Hispanics and abdominal surgeries in CD, both with only weak significance. This study reported no relationship between colectomies for UC or predicting the number of IBD-related hospitalizations [46]. In a study on ischemic stroke, the combined GRS (cGRS) of 113 SNPs led to an increase of only 0.5% in predictive power when added to all co-variables [46]. This suggests there is no clinical advantage in constructing a multi-locus SNP panel for predicting stroke risk, even when extended to include variants acting on intermediate phenotypes such as hypertension or atrial fibrillation [46]. This study suggested that the gain in predictive power from adding GRS to gender alone is limited in stroke [46].

Furthermore, our current findings agree with studies on other conditions, including breast cancer [35], diabetes mellitus [14, 15], coronary heart disease [33, 34], and multiple sclerosis [47], that found limited improvement in risk prediction with using GRS. Two larger studies concluded that, at present, the discriminative power of Polygenic risk score (PRS) for schizophrenia is not sufficient to use in population screening to identify individuals at high risk and that PRS may never prove powerful enough for screening [48, 49]. However, PRS explains a substantial amount of the variance of schizophrenia in Europeans, probably more than any other traditional risk factor. Finally, genetic risk variants so far known to play a role in migraine, are not able to explain a comprehensive set of clinical characteristics of migraine severity [8]. Altogether, as we stated during the early development of genetic studies [50], these observations across different domains question the potential clinical utility of PRS in predicting complex diseases in general populations. As noted by us and others [50-52], the predictive utility of GRS for common diseases is likely to be very limited, especially considering the myriad factors of the exposome that also influence individual susceptibility [51].

Our current study has several methodological strengths, including replication of the GRS in independent case-control samples and validation in general populations. We used one of the largest and most accurate collections of population-based samples of IBD patients with genome-wide data available to date. We based this work on an earlier GWAS with enriched genetic data to capture IBD probability compared to Caucasian populations. The current study investigated the association between IBD and GRS in several Asian populations. We demonstrated the feasibility of applying GRS based on Caucasian risk alleles to several Asian populations. The method we used provided efficient, genome-wide coverage

for our Asian sample. However, in our total sample, GRS explained only 4.40% of the variance in predicting case-control status, most probably capturing the strong genetic signal from Europeans.

The results of our study must be interpreted in light of its limitations. Our study was not sufficiently powered to investigate the effects of GRS of IBD, CD, UC that might play a role, especially in disease prediction. Moreover, our study included fewer participants with IBD than Caucasian studies, which might have limited our statistical power to identify any association between GRS and the predictability of IBD phenotypes in Asian populations. Future studies with a balanced number of participants with IBD are needed to confirm our findings. Our estimate on the GRS' predictive value and explained variance was based on common SNPs that had been selected based on predefined criteria from recently published loci associated with IBD. Thus, we may have excluded many variants with an effect on IBD. We did not analyze GRS for IBD subtypes because this information was not available in all cohorts used for validation. Our participants were specifically of Asian descent, thus our results may not be generalizable to other ethnicities. To explore the clinical importance of the association of GRS with IBD using common variants, further research in the field should go beyond the association with case-control status. Alternative strategies for constructing GRS for IBD and for combining GRS with risk factor profiles and clinical information might eventually lead to better risk prediction. Future risk assessment for complex diseases such as IBD should include a much more careful consideration of gene-gene and/or gene-environmental interactions.

Conclusion

In conclusion, we found a multi-locus GRS derived from GWAS for established risk factors for IBD to be significantly associated with IBD risk in Asians. However, the power of the GRS in predicting IBD risk and hence its clinical usefulness was limited. Our current study shows that genetic findings based on trans-ethnic analyses are indeed applicable across Central and East Asian populations, but the association of GRS that was built upon combining the effect of genome-wide associated risk alleles for IBD is unlikely to provide a strong predictive probability of IBD, CD and UC in Central and East Asians. Taking these results into consideration means that any strategies to test common genetic variants for informing clinical decisions would need to be rigorously tested beforehand. Greater efforts will be required to use the available genetic information in an appropriate clinical context to optimize disease prevention and management.

Abbreviations

Inflammatory Bowel Disease: (IBD),

Crohn's Disease: (CD),

Ulcerative Colitis: (UC),

Genetic Risk Scores: (GRS),

East Asia: (EA),

Central Asia: (CA),

Standard error: (SE),

Polygenic Risk: (PRS),

Allele Frequencies (AF),

Genome-Wide Association Study: (GWAS),

Single Nucleotide Polymorphisms: (SNPs),

IBD Genetic Consortium: (IBDGC),

linkage Disequilibrium: (LD),

Explained Variance (EV),

Odds Ratio (OR).

Declarations

Acknowledgments We thank the IBD Genetics Consortium, all the individuals who contributed samples and data, and the physicians and nursing staff who helped with recruitment worldwide.

Authors' contributions Conceived and designed study: SA, SHW, SCN, BZA. Statistical analysis and interpretation of the data: SA, BZA. Contributed to writing of the manuscript: SA, SHW, SCN, BZA. Critically read the final version: SA, SHW, RKW, SCN, BZA. All authors read and approved the final manuscript before submission.

Availability of data and materials Not applicable.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing financial interests.

Web Link R programming courses and Knowledge based software can be found at <http://cran.rproject.org/web/packages/Mangrove/index.html>.MMM

References

1. Siew, N. G. C., et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet*. **390** (10114), 2769–2778 (2017).
2. Yun, Q. I. U. *et al.* Effects of combination therapy with immunomodulators on trough levels and antibodies against tumor necrosis factor antagonists in patients with inflammatory bowel disease: a meta-analysis. *Clinical Gastroenterology and Hepatology*. **15** (9), 1359–1372 e6. (2017).
3. KHOR, B., GARDET, A. & Ramnik, X. A. V. I. E. R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. **474** (7351), 307–317 (2011).
4. Jimmy, L. I. U. Z., et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*. **47** (9), 979 (2015).
5. DE LANGE, Katrina, M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature genetics*. **49** (2), 256 (2017).
6. JOSTINS *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. **491** (7422), 119–124 (2012).
7. CLEYNEN *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet*. **387** (10014), 156–167 (2016).
8. LEE, J. C. *et al.* Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nature genetics*. **49** (2), 262 (2017).
9. SONG, L. I. U. S. & Yiqing Building genetic scores to predict risk of complex diseases in humans: is it possible? *Diabetes*. **59** (11), 2729–2731 (2010).
10. ELLINGHAUS, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature genetics*. **48** (5), 510 (2016).
11. Lewis, K. U. L. L. E. R., MEILAHN, H. & Elaine, N. Risk factors for cardiovascular disease among women. *Current opinion in lipidology*. **7** (4), 203–208 (1996).
12. KULLO, Iftikhar, J. & Leslie, C. O. O. P. E. R. T. Early identification of cardiovascular risk using genomics and proteomics. *Nature Reviews Cardiology*. **7** (6), 309 (2010).
13. Simin, L. I. U. *et al.* A prospective study of inflammatory cytokines and diabetes mellitus in a multiethnic cohort of postmenopausal women. *Archives of internal medicine*. **167** (15), 1676–1685 (2007).
14. LYSSSENKO *et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *New England Journal of Medicine*. **359** (21), 2220–2232 (2008).
15. MEIGS, James B., et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *New England Journal of Medicine*, 2008, 359.21:2208–2219.
16. Nina, P. A. Y. N. T. E. R. P., et al. Association between a literature-based genetic risk score and cardiovascular events in women. *Jama*. **303** (7), 631–637 (2010).
17. DE SILVA, N., Maneka, G. & FRAYLING, Timothy, M. Novel biological insights emerging from genetic studies of type 2 diabetes and related metabolic traits. *Current opinion in lipidology*. **21** (1), 44–50

- (2010).
18. PIRINEN *et al.* Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *The Annals of Applied Statistics*. **7** (1), 369–390 (2013).
 19. KIM, E. S. & Won Ho. Inflammatory bowel disease in Korea: epidemiological, genomic, clinical, and therapeutic characteristics. *Gut and liver*. **4** (1), 1 (2010).
 20. LI *et al.* The disease burden and clinical characteristics of inflammatory bowel disease in the Chinese population: a systematic review and meta-analysis. *International journal of environmental research and public health*. **14.3**, 238 (2017).
 21. AHUJA, K. E. D. I. A. S. & Vineet Epidemiology of inflammatory bowel disease in India: the great shift east. *Inflammatory intestinal diseases*. **2** (2), 102–115 (2017).
 22. MALEKZADEH, Masoud, M. *et al.* Emerging epidemic of inflammatory bowel disease in a middle income country: a nation-wide study from Iran. *Archives of Iranian medicine*, 2016,1–14.
 23. Siew, N. G. C., *et al.* Incidence and phenotype of inflammatory bowel disease based on results from the Asia-pacific Crohn's and colitis epidemiology study. *Gastroenterology*. **145** (1), 158–165 (2013). e2
 24. VAN DER, S. L. O. O. T. *et al.* Inflammatory bowel diseases: review of known environmental protective and risk factors involved. *Inflammatory bowel diseases*. **23** (9), 1499–1509 (2017).
 25. Hannah, G. O. R. D. O. N. *et al.* Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies. *Inflammatory bowel diseases*. **21** (6), 1428–1434 (2015).
 26. Guo-Bo, C. H. E. N. *et al.* Estimation and partitioning of (co) heritability of inflammatory bowel disease from GWAS and immunochip data. *Human molecular genetics*. **23** (17), 4710–4720 (2014).
 27. TYSK, C. *et al.* Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut*. **29** (7), 990 (1988).
 28. PRIDEAUX *et al.* Inflammatory bowel disease in Asia: a systematic review. *Journal of gastroenterology and hepatology*. **27** (8), 1266–1280 (2012).
 29. KENNY, Eimear, E. *et al.* A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS genetics*, 2012, 8.3.
 30. HORNE, B. D. *et al.* Generating genetic risk scores from intermediate phenotypes for use in association studies of clinically significant endpoints. *Annals of human genetics*. **69** (2), 176–186 (2005).
 31. Rainer, M. A. L. I. K. *et al.* Shared genetic basis for migraine and ischemic stroke: a genome-wide analysis of common variants. *Neurology*. **84** (21), 2132–2145 (2015).
 32. Bendik, W. I. N. S. V. O. L. D. S., *et al.* Genetic analysis for a shared biological basis between migraine and coronary artery disease. *Neurology Genetics*. **1** (1), e10 (2015).
 33. RIPATTI *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *The Lancet*. **376** (9750), 1393–1400 (2010).

34. BRAUTBAR *et al.* A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis*. **223** (2), 421–426 (2012).
35. WACHOLDER *et al.* Performance of common genetic variants in breast-cancer risk models. *New England Journal of Medicine*. **362** (11), 986–993 (2010).
36. Sun-Gou, J. I. *et al.* Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nature genetics*. **49** (2), 269–273 (2017).
37. Hon-Cheong, S. O. *et al.* Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genetic epidemiology*. **35** (5), 310–317 (2011).
38. SO, Hon-Cheong, S. H. A. M. & Pak, C. A unifying framework for evaluating the predictive power of genetic variants based on the level of heritability explained. *PLoS genetics*, 2010, 6.12.
39. VERSTOCKT, B., SMITH, Kenneth, G. C. & LEE, James, C. Genome-wide association studies in Crohn's disease: Past, present and future. *Clinical & translational immunology*. **7** (1), e1001 (2018).
40. CIVELEK, M. & LUSIS, Aldons, J. Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*. **15** (1), 34–48 (2014).
41. HIRSCHHORN, Joel, N. & Mark, D. A. L. Y. J. Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics*. **6** (2), 95–108 (2005).
42. CHU, H. *et al.* Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science*, aad9948. 2016.
43. NAKAGOME *et al.* Confounding effects of microbiome on the susceptibility of TNFSF15 to Crohn's disease in the Ryukyu Islands. *Human genetics*. **136** (4), 387–397 (2017).
44. KO *et al.* Inflammatory bowel disease environmental risk factors: a population-based case-control study of Middle Eastern migration to Australia. *Clinical Gastroenterology and Hepatology*. **13** (8), 1453–1463 (2015). e1
45. DUDBRIDGE, Frank. Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, 2013, 9.3.
46. Rainer, M. A. L. I. K. *et al.* Multilocus genetic risk score associates with ischemic stroke in case-control and prospective cohort studies. *Stroke*. **45** (2), 394–402 (2014).
47. DE JAGER, Philip, L. *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol*. **8** (12), 1111–1119 (2009).
48. DICHGANS, M. Genetics of ischaemic stroke. *Lancet Neurol*. **6** (2), 149–161 (2007).
49. Lucia, H. I. N. D. O. R. F. F. *et al.* A., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 2009, 106.23: 9362–9367.
50. VAN DUIJN, Cornelia, M. & PORTA Miquel. Good prospects for genetic and molecular epidemiologic studies in the European Journal of Epidemiology. *European journal of epidemiology*. **18** (4), 285

(2003).

51. PISANU *et al.* A genetic risk score is differentially associated with migraine with and without aura. *Human genetics.* **136** (8), 999–1008 (2017).

52. VASSOS *et al.* An examination of polygenic score risk prediction in individuals with first-episode psychosis. *Biol. Psychiatry.* **81** (6), 470–477 (2017).

Table

Table 1. Association between GRS and clinical phenotype of IBD

	OR	EAF	Variance (%)	Explained	OR GRS	P value	Prevalence (%)	Predictive Probability
IBD								
Japan	1.10±0.14	0.50	4.40		1.03±0.93	0.006	0.08	8.80×10 ⁻⁴
Korea	1.06±0.09	0.49	1.51		1.18±0.53	0.000	0.04	4.76×10 ⁻⁴
China	1.07±0.10	0.48	1.34		1.09±0.53	0.000	0.01	1.05×10 ⁻⁴
India	1.09±0.14	0.51	3.81		1.25±0.66	0.000	0.046	5.52×10 ⁻⁴
Iran	1.09±0.09	0.51	4.14		1.36±0.55	0.000	0.04	5.45×10 ⁻⁴
CD								
Japan	1.11±0.14	0.47	3.63		1.00±0.91	0.028	0.02	2.12×10 ⁻⁴
Korea	1.06±0.10	0.48	1.02		1.06±0.43	0.005	0.01	1.06×10 ⁻⁴
China	1.08±0.15	0.47	1.19		1.10±0.66	0.003	0.00	2.98×10 ⁻⁵
India	1.10±0.13	0.50	3.5		1.05±0.81	0.000	0.002	2.11×10 ⁻⁵
Iran	1.10±0.14	0.51	4.26		1.16±0.43	0.000	0.005	5.82×10 ⁻⁵
UC								
Japan	1.09±0.12	0.48	3.33		1.03±0.81	0.040	0.06	6.18×10 ⁻⁴
Korea	1.14±0.20	0.51	1.28		1.24±1.24	0.003	0.03	3.73×10 ⁻⁴
China	1.07±0.13	0.51	1.21		1.02±0.55	0.029	0.01	7.13×10 ⁻⁵
India	1.08±0.10	0.51	3.49		1.27±0.60	0.000	0.044	5.59×10 ⁻⁴
Iran	1.08±0.13	0.51	3.65		1.30±0.53	0.000	0.035	4.56×10 ⁻⁴

OR: Odds ratio, SE: Standard Error, EAF: Effect allele frequency, GRS: Genetic risk score

*The OR column shows the mean± SE of the OR calculated by MMM software for three East Asian populations and the Caucasian OR for the Indo-Iranian populations.

Figures

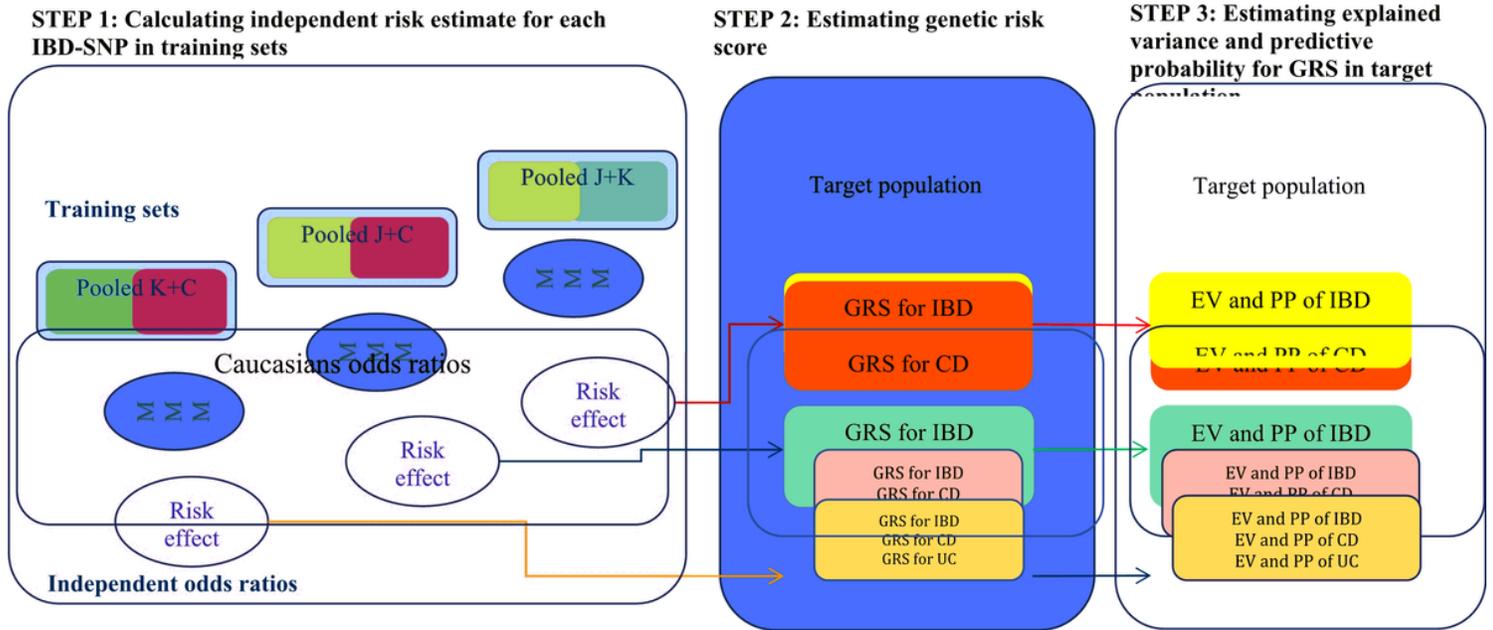


Figure 1

Study design EV: Explained variance, PP: posterior probabilities; J Japanese; K Korean; C Chinese; MMM software package For a training set of two East Asian populations, we calculated an independent risk estimate (OR) for IBD, CD and UC for each IBD-associated SNP. We combined the Korean and Chinese cohorts (ORKC), the Japanese and Chinese cohorts (ORJC) and finally the Japanese and Korean cohorts (ORJK). Using the MMM software package, we estimated independent odds ratios for the common IBD-associated SNPs in three East Asian populations, then we calculated the genetic risk scores, explained variance and posterior probabilities for our target populations. For Indians and Iranians, we included the risk allele and odds ratios as indicated for the Caucasian population in our trans-ethnic meta-GWAS

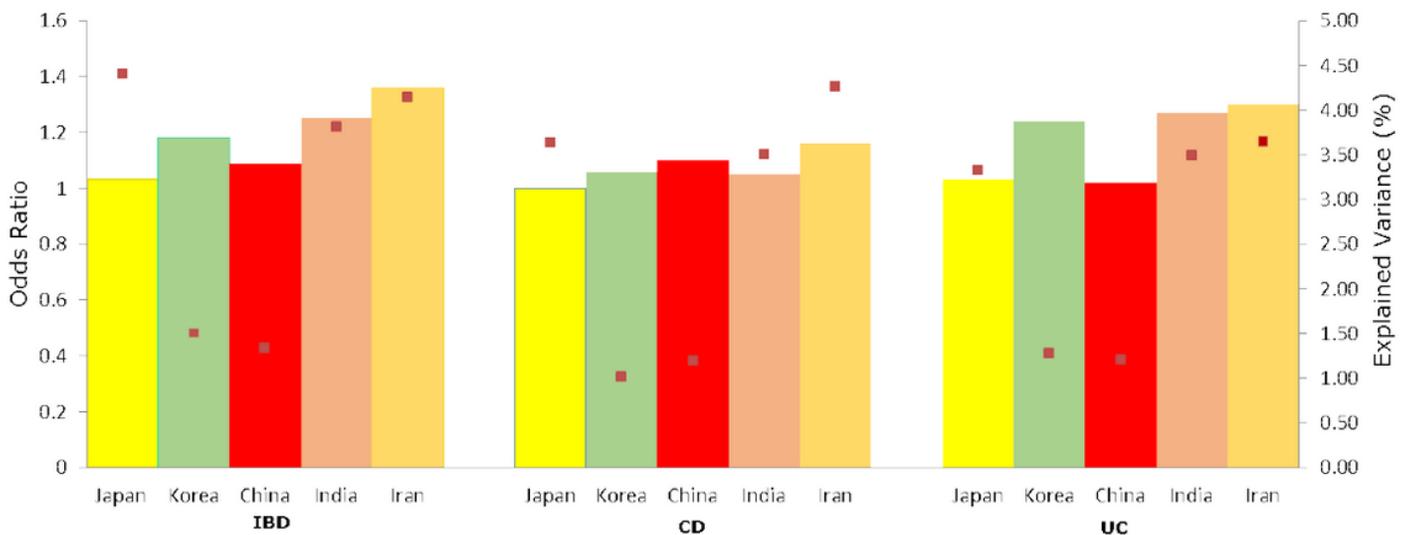


Figure 2

Odds ratios and explained variance in five Asian target populations, per phenotype and per ethnic cohort

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryGeneticpredictionofIBDS.Abedianetal.docx](#)