



Beyond susceptibility: do population-specific rare variants shape age at diagnosis in Korean inflammatory bowel disease?

Soo-kyung Park

Department of Gastroenterology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Korea

See Article on Page 847-857

Genome-wide association studies (GWAS) have identified hundreds of loci associated with inflammatory bowel disease (IBD), substantially advancing our understanding of host-microbe interactions, epithelial barrier function, and immune regulation [1,2]. However, most of the established IBD-associated variants are non-coding, confer modest effects individually, and have been discovered primarily in populations of European ancestry [2]. Therefore, their applicability to East Asian populations is limited. This is particularly important in patients with IBD, where susceptibility loci do not exert comparable effects across ancestries. NOD2, a major Crohn's disease (CD) determinant in European populations, has a more limited role in East Asian populations, whereas TNFSF15-related variants show stronger associations in Asian populations [3,4]. Population-specific genomic studies are thus not merely replications across different ethnic groups; they may reveal the biology that European ancestry studies underrepresent or miss entirely.

An equally important question is whether the genetic susceptibility determinants are the same as those governing when IBD becomes clinically apparent, and how aggressively it subsequently behaves. Evidence from CD and, more recently, cross-disease analyses suggests that susceptibility and prognosis have at least partly distinct genetic architectures [5,6]. In this issue of the *Korean Journal of Internal Medicine*, Cho et al. [7] extended this concept to Korean patients with IBD by using whole-exome sequencing (WES) to search for rare variants associated with age at diagnosis

and subsequent disease progression. Their study raises the clinically meaningful possibility that ancestry-informed rare variants may help to explain why some patients develop IBD earlier and potentially follow a more severe course.

One of the most notable strengths of this study is its analytical design. The authors first attempted conventional case-control analysis using Japanese and Han Chinese participants in the 1000 Genomes Project as external controls. A substantial distortion in the test statistics and systematic clustering attributable to the differences between exome capture and whole-genome sequencing was observed. Rather than interpreting potential artifactual signals, these results were discarded and a within-case Cox proportional hazards analysis of age at diagnosis among 341 Korean patients was employed [7]. This decision was methodologically sound and deserves further emphasis. Rare-variant association testing is extremely sensitive to technical artifacts because variants observed in only a handful of individuals can generate enormous effect estimates when platforms, capture kits, coverage, or calling pipelines differ [8]. By restricting the primary analysis to individuals that were affected and contrasting earlier with later diagnoses within that group, the authors reduced technical confounding factors and shifted the focus from susceptibility to phenotypic heterogeneity. This study reinforces a general principle in genomic epidemiology: the appropriate comparison group follows from the biological question, not from the convention.

The primary finding was the association between a rare GSG1 variant (rs146166808) and earlier age at diagnosis. The variant reached genome-wide significance ($p = 3.39 \times 10^{-8}$) and showed the strongest effect in the CD and com-

bined IBD analyses, with Kaplan–Meier curves indicating a substantially earlier diagnostic age among carriers [7]. In principle, rare coding variants can exert greater functional effects than common GWAS variants and may account for part of the missing heritability of complex traits [2]. The fact that most prioritized variants are rare and enriched in East Asian populations supports the value of population-specific sequencing.

Nevertheless, these estimates should be interpreted as signals rather than as measurements. The hazard ratios are extraordinarily large but rest on very few carriers, and the resulting confidence intervals [CIs] are correspondingly uninformative. Three of the five ulcerative colitis (UC) candidates—ZNF454, SALL3, and GLYATL1—share an identical hazard ratio of 212.95 (95% CI, 25.30–1,792.22), an arithmetic signature of single-carrier variants rather than three independent findings. In the progression analysis, the GSG1 association rests on two carriers, both of whom experienced events; the reported odds ratio of 12.52 carries a 95% CI of 0.59–264.79, compatible with a substantial effect and with none at all [7]. Asymptotic Cox and Firth-penalized tests were not well calibrated at minor allele counts in the single digits, and p-values in the 10^{-8} range from such data should not be equated with the same threshold in a common-variant GWAS [9]. Permutation or exact testing, together with the carrier counts reported alongside every effect estimate, allows readers to calibrate their confidence levels appropriately.

A related concern is the residual technical heterogeneity within case series. Cohort 3 was sequenced with a different capture kit on a different instrument, in an earlier era, from Cohorts 1 and 2, and the three cohorts differ markedly in subtype composition, smoking history, and mean age at diagnosis (25.6–34.0 yr) [7]. The cohort was included as a covariate; however, cohort membership is simultaneously a proxy for platform, calendar period, and phenotype; therefore, a rare variant that is differentially captured or denoted across kits will not be fully absorbed by that adjustment. The authors interpret the modest genomic inflation of their within-case analyses ($\lambda = 1.14$ – 1.22) as evidence of polygenicity; in a sample of 341 individuals tested largely at rare sites, small-sample miscalibration is an equally plausible explanation. Reporting per-cohort carrier distributions, read depth, genotype quality at prioritized sites, and orthogonal confirmation by Sanger sequencing, would be inexpensive and would substantially strengthen the claim that these are

biological rather than technical signals.

Population specificity framing has certain nuances. For most candidates, the case is clear: GSG1 rs146166808, for example, has an East Asian allele frequency roughly thirty-fold higher than the European frequency. However, the study's own supplementary allele-frequency data show that not every prioritized variant fits this description; IGLL1 rs139571703 has a global frequency $> 3\%$ and is more common in European and South Asian populations than in East Asian reference populations [7]. The general conclusion is well supported; however, the statement that all prioritized candidates are rare and predominantly East Asian requires qualification.

The biological interpretation of GSG1 is, however, preliminary. The predominant expression in intestinal neuroendocrine cells raises a plausible hypothesis regarding neuroendocrine signaling and intestinal inflammation; however, expression data cannot establish causality. Functional studies must determine whether rs146166808 alters GSG1 expression or protein function, and whether any such alteration affects epithelial integrity, neuroimmune communication, or inflammatory signaling. Replication, functional validation, and demonstrated predictive values beyond those of conventional clinical factors are prerequisites for clinical use.

This study also reported distinct CD and UC genetic architectures. Gene-level analyses identified 43 significant genes in CD, 27 in UC, and 14 in combined IBD, with ZNF286A the sole gene shared between subtypes; pathway analyses differed accordingly, with CD-associated genes enriched for axon guidance, monocyte chemotaxis, and NAD salvage, and UC-associated genes for calcium signaling, gliogenesis, and protein quality control [7]. These results should not be interpreted as evidence that neuronal pathways cause CD. However, they are consistent with the growing appreciation that enteric neural, epithelial, immune, and stromal compartments interact to maintain mucosal homeostasis [1,2], and that the axon-guidance signal offers a reasonable entry point for determining whether neuroimmune interactions contribute differently to the two subtypes. Enrichment is not a mechanism. Gene ontology and pathway databases reflect the accumulated structure of prior knowledge, with several pathways relying on a single gene (NAD salvage on NAPRT and RNA acetylation on NAT10), and pathway significance was declared at an uncorrected threshold. In a WES study of this size, gene-level signals are sensitive to the annotation and aggregation of rare variants.

Perhaps the most important conceptual limitation is phenotype. Age at diagnosis is clinically meaningful but not biological onset. Diagnostic delay is common in IBD, particularly in CD, and varies with disease location, presenting symptoms, healthcare utilization, referral patterns, available technology, and calendar period [10]. As the three cohorts were recruited across two decades and differed in mean diagnostic age, some of the observed variations may reflect changing diagnostic practices rather than genetic determinants of inflammation. Future studies should separate the age at symptom onset, age at the first objective evidence of intestinal inflammation, and age at formal diagnosis. A related caveat applies to the authors' screening of established susceptibility loci: the finding that known IBD variants do not explain variations in onset is limited by the fact that only 23 of more than a thousand screened associations were coded and therefore assayable by WES. This result describes exome coverage rather than the biology of onset.

The next priority is independent replication. Although 341 participants constituted a meaningful Korean WES cohort, this remains modest for rare-variant testing, and only 206 patients provided longitudinal data. Replication in independent Korean cohorts is essential, and coordinated studies in Japan, China, and elsewhere in East Asia would establish whether these signals are Korea-specific or reflect a broader East Asian architecture. Such efforts should use harmonized capture and joint variant naming to eliminate the technical heterogeneity identified in this study.

A broader implication is that IBD genetics should move beyond a binary susceptibility framework. A patient's genetic background may influence not only the development of IBD but also the age at which it declares itself, its phenotype and behavior, treatment responses, and long-term complications [5,6]. Such a multidimensional model is far more relevant to precision medicine than a catalogue of susceptibility loci. Routine genotyping of GSG1 or other rare candidate variants remains premature, and clinical utility requires reproducible associations, mechanistic plausibility, and demonstrable incremental performance over readily available clinical predictors.

Cho et al. took an important first step toward defining the role of population-specific rare variants in Korean patients with IBD. Their most durable contribution may be methodological and conceptual: a carefully constructed within-case analysis can reveal clinically relevant genetic signals that conventional susceptibility studies overlook, and East Asian

populations must be represented in genomic discovery rather than assumed to resemble European populations. As the incidence of IBD continues to increase across Asia, the challenge is to move from rare-variant discovery to functional validation and, ultimately, to determine whether ancestry-informed genomics can improve individualized prediction and management. The path from population-specific signals to precision medicine is lengthy; however, this study marks a credible beginning.

REFERENCES

1. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474:307-317.
2. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119-124.
3. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979-986.
4. Yang SK, Hong M, Zhao W, et al. Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* 2014;63:80-87.
5. Lee JC, Biasci D, Roberts R, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. *Nat Genet* 2017;49:262-268.
6. Yang Z, Pajuste FD, Zguro K, et al. Limited overlap between genetic effects on disease susceptibility and disease survival. *Nat Genet* 2025;57:2418-2426.
7. Cho M, Shin D, Yu J, Cheon JH, Kim JH. Population-specific rare variants influence age at diagnosis in Korean inflammatory bowel disease. *Korean J Intern Med* 2026;41:847-857.
8. Wu L, Schaid DJ, Sicotte H, Wieben ED, Li H, Petersen GM. Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet* 2015;52:10-16.
9. Ma C, Blackwell T, Boehnke M, Scott LJ; GoT2D investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet Epidemiol* 2013;37:539-550.
10. Vavricka SR, Spigaglia SM, Rogler G, et al. Systematic evaluation of risk factors for diagnostic delay in inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18:496-505.

Received : August 20, 2026

Accepted : August 22, 2026

Correspondence to

Soo-kyung Park, M.D.

Department of Gastroenterology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, 29 Saemunan-ro, Jongno-gu, Seoul 03181, Korea

Tel: +82-2-2001-8330, Fax: +82-2-2001-8360

E-mail: skparkmd@gmail.com

<https://orcid.org/0000-0001-8822-9632>

Conflicts of interest

The author discloses no conflicts.

Funding

None