Journal of Genetics, Vol. 97, No. 5, December 2018, pp. 1493–1507
https://doi.org/10.1007/s12041-018-1015-8

REVIEW ARTICLE

Genetics of ulcerative colitis: putting into perspective the incremental gains from Indian studies

GARIMA JUYAL[1]*, AJIT SOOD[2], VANDANA MIDHA[2] and B. K. THELMA[3]*

[1] *School of Biotechnology, Jawaharlal Nehru University, New Delhi 110 067, India*
[2] *Dayanand Medical College and Hospital, Ludhiana 141 001, India*
[3] *Department of Genetics, University of Delhi South Campus, New Delhi 110 021, India*
*For correspondence. E-mail: Garima Juyal, garimajuyal@gmail.com; B. K. Thelma, thelmabk@gmail.com.

Received 6 May 2018; revised 28 August 2018; accepted 19 September 2018; published online 14 November 2018

Abstract. Ulcerative colitis (UC), one of the two clinical subtypes of inflammatory bowel disease is perceived as a potential 'sleeping giant' in the Indian subcontinent. Clinical manifestation is overall believed to be the same across ethnic groups but overwhelming genetics from large European and fewer non-European studies have revealed shared as well as unique disease susceptibly signatures between them, pointing to population specific differences at genomic and environmental levels. A systematic recount of the four major eras in UC genetics spanning earliest linkage analysis, cherry picked candidate gene association studies, unbiased genomewide association studies, their logical extension in trans-ethnic setting (Immunochip study), lastly whole exome sequencing efforts for rare variant burden; and lessons learnt thereof in context of genetically distinct Indian population was attempted in this review. Genetic heterogeneity manifesting at allelic/locus level across these approaches has been the consistent finding through the range of pan India studies. On the other hand, these salient findings also highlight the limitations of even the best of these genetic leads for prognostic/clinical application. The imminent need, therefore, for the UC research community to adopt newer approaches/tools with improved study design to (i) gain better insight into genetic/mechanistic basis of disease; (ii) identify biomarkers of immediate translational value; and (iii) develop new/alternate therapeutic options is emphasized at the end.

Keywords. ulcerative colitis; inflammatory bowel disease; genetic heterogeneity; Indian population; genomewide association studies.

Introduction

Ulcerative colitis (UC) is a chronic, relapsing inflammatory disease of the gastrointestinal tract. Along with Crohn's disease (CD), UC is one of the two major forms of inflammatory bowel disease (IBD). However, due to high prevalence of UC among Asians in general and Indians in particular, this review is confined to genetics of UC, which is well documented to have a heritable component. UC affects individuals during their most productive years of life (20–40 years of age) and shows almost no gender bias (Kedia and Ahuja 2017). The inflammation in UC is mainly confined to colon and rectum, and its pattern is continuous and superficial. Based on the location and extent of inflammation involved, UC can be further subcategorized into proctitis (E1), left sided (E2) and pancolitis (E3) (Dubinsky 2017). These clinical presentations, together with differential drug response among patients and development of extraintestinal manifestations in some patients reiterate clinical heterogeneity among patients with UC.

UC has long been thought as a disease of developed nations. However, recent epidemiological studies suggest that it is on the rise in developing countries (Ng *et al.* 2013; Park *et al.* 2014; Kaplan 2015), indicating its emergence as a global disease. Multifactorial (genetic and nongenetic contributors) origin of UC has been well documented previously in disease aetiology (Loddo and Romano 2015) and in the preceding decade, UC (among a few others) has been at the frontline of success stories in complex trait research with several genetic loci uncovered (Ye and McGovern 2016). These findings primarily in populations of Caucasian ancestry have however failed to explain the total heritable component in UC. More recently, genetic analysis of UC in non-Europeans have identified several novel susceptibility genes/loci (Asano *et al.* 2009; Okada

et al. 2011; Yang *et al.* 2013; Juyal *et al.* 2015), indicating pronounced genetic heterogeneity in UC due to disparate genetic architecture and environmental attributes across populations. An indepth understanding of such differences in trans-ethnic settings therefore seems to be imminent to achieve the goals of early disease prediction and prevention, which are extremely important considering the emerging disease burden, particularly in the developing countries. This review which aims to highlight the salient genetic findings from India and other non-European populations is expected to facilitate such an effort.

UC can be labelled as a 'western disorder' emerging as a 'potential sleeping giant' in traditional non-IBD areas

Historically considered a disease of the developed societies (European ancestry populations), more recent reports, however, have shown that UC is an emerging disease in the Indian subcontinent (Sood *et al.* 2003; Singh *et al.* 2017) and other Asian countries. This shift may be due to a combination of factors including but not limited to increased awareness, availability of better facilities for diagnosis, rapid urbanization, better socioeconomic conditions, westernization (changes in dietary patterns and better hygienic conditions), etc. Notably, it has been observed that incidence and prevalence of UC vary greatly among Asian countries with the disease burden highest in India (Ng *et al.* 2016; Kedia and Ahuja 2017). Such differences in the incidence across time and geographical regions suggest role of environmental factors.

Evidence for a genetic aetiology

Aetiology of UC is not well understood but evidence gathered suggests that it is a complex/multifactorial disease involving one of the four components namely environmental changes, genetic variations, immune disequilibrium and dysbiosis of luminal microbiota that trigger inflammatory responses and their interactions. It is believed that variations in these factors account for induction of disease and influence the disease course and clinical diversity in different populations. Marked ethnic and geographic differences in incidence and prevalence of UC, concordance in twins, familial clustering (which was reflected in high sibling risk ratios), and association with genetic syndromes strengthen the role of genetics in UC (Ek *et al.* 2014). Nonetheless, these studies also suggested that significant nongenetic factors, primarily environmental or luminal, are required to 'trigger' UC in a genetically susceptible host. Motivated by these early observations that UC has a heritable component, significant disease gene mapping efforts have been made over the preceding decades. The quest for *bona fide* risk genes spanned over four distinct phases beginning with linkage mapping using multiple affected families, followed initially by candidate gene based association testing using unrelated affected individuals and healthy controls, and subsequently by genomewide and Immunochip association analysis; and finally, rare variant search by next-generation sequencing approach are briefly described in the following sections.

With the advent of genomewide scans, linkage mapping, which traces chromosomal regions that cosegregate with the disease in multiple member pedigrees, became a robust and extremely successful method in mapping highly penetrant loci affecting Mendelian traits (Altshuler *et al.* 2008). Subsequently, this method was increasingly being explored to map complex disease loci. From 1996, 10 genomewide linkage scans were performed in IBD which identified nine susceptibility loci (IBD1–IBD9) but with little replication across the scans (Mathew and Lewis 2004; Walters and Silverberg 2006). The principal insight learned from the lack of consistency of genomewide linkage results were that UC (unlike most single gene disorders) (i) exhibits clinical heterogeneity such as disease severity, location and extent; (ii) vary in their aetiological mechanisms, which might involve different biological pathways; and (iii) are driven by several low-penetrant risk genes. In other words, UC is a genetically heterogeneous disorder. Soon it was recognized that UC like other complex diseases results from the collective effects of polygenic variations, with each variant having a modest effect and present at a high frequency in human populations, commonly known as 'common disease/common variant' (CDCV) hypothesis (Risch and Merikangas 1996) and that linkage is not so well powered to detect them demanding new and more powerful approaches. Discovery of dense common genetic variations such as single-nucleotide polymorphisms (SNPs), indels, copy number variations (CNVs) throughout the genome following the human genome project (Sherry *et al.* 1999; Sachidanandam *et al.* 2001; McVean *et al.* 2004; International HapMap Consortium *et al.* 2007) offered a rather powerful and a timely tool to the gene hunters and facilitated candidate gene based association studies to start with followed by genomewide association studies (GWASs). In spite of their promise, hypothesis testing candidate–gene association approach yielded limited insights as most of the initial findings failed to replicate in subsequent studies. In case of UC, several candidate genes were tested predominantly in European populations; however, these failed to replicate in subsequent independent studies with failure rate being higher in non-European populations as compared to Europeans and quickly evidence of allelic and genetic heterogeneity in UC was appreciated. The reasons for discordant findings were mostly (i) sample size; (ii) small number of candidate genes and variants tested; (iii) population specific allele frequency profiles; and (iv) different linkage disequilibrium (LD) structures. These were partly addressed by additional SNP discovery projects including The SNP

Consortium (Thorisson and Stein 2003) and HapMap project (International HapMap Consortium *et al.* 2007) together with the advent of high-throughput microarray technologies enabling hypothesis free genomewide genotyping which upended the stasis in gene discovery. GWAS led to the discovery of an overwhelming genetic signature for UC (www.ebi.ac.uk/gwas/). These studies (i) highlighted known pathways but also identified unappreciated pathways raising novel hypotheses about disease pathogenesis; (ii) confirmed the participation of microbial component in the development of UC based on association of immune related genes; (iii) most importantly, demonstrated varying genetic contributors in different populations; and (iv) also revealed shared/unique biological mechanisms between UC and CD. GWASs era turned out to be only an interim respite for UC research community as the GWAS based discoveries and their meta-analysis showed only modest/weak effects on disease susceptibility and the bulk of the heritability could not be explained. Soon the complex trait researchers resorted to trans-ethnic GWAS. Apart from the other inherent advantages of such studies (Li and Keating 2014), identification of additional disease loci to narrow the missing heritability gap caught the attention. Notably, such a study in IBD conducted by International Inflammatory Bowel Disease Consortium (IIBDGC) utilized immune enriched, custom-designed genotyping array i.e. Immunochip and genotyped a much large sample set comprising 86,640 European individuals and 9846 individuals of East Asia, India or Iran descent. An additional 38 IBD risk associated loci were identified at genomewide significance in either the association analysis of individual ancestry groups ($P < 5\times10^{-8}$) or in the trans-ethnic meta-analysis that included all ancestries for UC ($n = 8$), CD ($n = 15$) or IBD ($n = 15$) (Liu *et al.* 2015). Simultaneously, UC genetics (like other complex traits) has witnessed a paradigm shift to common disease rare variant (CDRV) where structural and rare (minor allele frequency < 1%) variants mostly private are being identified using the next-generation sequencing approach.

From the above description, it is evident that genetic dissection of UC was performed with multiple approaches which evolved over the last two decades. However, most of these studies as mentioned before were performed primarily on populations of European ancestry. While these studies have expanded the genetic landscape of UC, they have been unable to explain the total disease variance despite large sample sizes. This highlights the importance of studying genetically divergent populations with unique genetic architecture and also exposed to unique environmental factors. This is elegantly exemplified by the series of genetic studies performed, for example in the ethnically distinct Indian population to identify shared as well as private genetic determinants underlying UC and the major insights gained thereof. Salient findings from these studies are summarized in the next section.

Current state of UC genetics: an Indian scenario

The early work on UC genetics in Indian population was primarily based on replicating the associations of famous candidate genes such as *IL-1Ra*, *TNF-alpha*, *GSTM1*/*T1*, and *NOD2* reported in Caucasian populations. While no association was observed for *IL-1Ra* and *TNF-alpha* with UC, *GSTM1*/*T1* showed significant association (table 1). Results of *NOD2* association in a north Indian cohort, on the other hand, were noteworthy. The minor alleles of all the three famous SNPs: 8, 12 and 13 were monomorphic in healthy individuals as well as UC patients in the north Indian study cohort. Even more interesting was the identification of a different variant, i.e. SNP5 which was associated modestly with UC in this cohort (Juyal *et al.* 2007). Of note, although SNPs 8, 12 and 13 are individually rare, they have been shown to be present on SNP5 haplotype in Caucasian populations (Juyal *et al.* 2007). Thus, this first Indian report provided evidence of allelic heterogeneity at *NOD2* in the north Indian UC cohort. Subsequently, several more novel variants have been reported in *NOD1* (Verma *et al.* 2009, 2012). The next candidate which received considerable attention was *MDR1*. The three most commonly investigated SNPs in this gene, namely 1236C>T (rs1128503), 2677G>T/A (rs2032582), and 3435C>T (rs1045642) were modestly associated with UC albeit differences in their minor allele frequencies (MAF) across north Indians and HapMap populations (Juyal *et al.* 2009). Similar to findings in *NOD2*, a significantly associated nonsynonymous SNP rs11209026 in yet another strong candidate gene namely *IL23R* in the Caucasian GWAS turned out to be monomorphic in Indian population. Such association studies with candidate genes carried out in several independent UC cohorts in India are summarized in table 1. Of note, SNPs 8, 12 and 13 in *NOD2* and *IL23R* SNP were also monomorphic in other non-European populations. For ease of comparison, these association findings are also included in table 1. Taken together, these unexpected results from heavy-hitter candidate genes provided initial clue for appreciable allelic and locus heterogeneity in UC pathogenesis and highlighted basal differences in genetic architecture across populations, which may partly explain nonreplication.

It was around the first documentation of *NOD2* association in the Indian population (2007) that the genetics of complex traits witnessed a technological revolution from hypothesis testing of 1–2 SNPs per gene based candidate gene association studies to the still popular GWAS approach. A few such studies performed on large sized UC cohorts almost all of them from Caucasian populations (Fisher *et al.* 2008; Franke *et al.* 2008; Silverberg *et al.* 2009; UK IBD Genetics Consortium *et al.* 2009; Asano *et al.* 2009) identified novel UC susceptibility genes but explained a very small fraction of the genetic contribution to disease susceptibility,

Table 1. Summary of candidate gene based association studies performed in Indian UC cohort and their status in Southeast Asians.

Gene (chr.)	Case	Control	Minor allele frequency*		Allelic *P* value	Genotypic *P* value	Reference	Association status in Southeast Asian UC cohorts (reference)
			Case	Control				
IL-1Ra (2)	82	141					Mittal *et al.* (2005)	
86-bp VNTR, intron 2								Associated (Nohara *et al.* 2003)
TNF-alpha (6)								
rs1800629	92	164						Associated (Sashio *et al.* 2002; Kim *et al.* 2003; Song *et al.* 2005; Cao *et al.* 2006)
GSTM1 (1) null polymorphism	85	164			<0.001		Mittal *et al.* (2007)	Associated (Ye *et al.* 2011)
GSTT1 (22) null polymorphism					<0.001			Associated (Qian *et al.* 2017)
NOD2 (16)								
rs2066844 (SNP8)	298	262	Monomorphic				Juyal *et al.* (2007)	
rs2066845 (SNP12)								
rs2066847 (SNP13)								
rs2066842 (SNP5)			0.17	0.12	0.008	0.005		
NOD1 (7)	95	102					Verma *et al.* (2009)	
4644T>A			0.26	0.09	0.002			
4773delG								
4785G>A								
5035T>C			0.16	0.32	0.002			
5098T>G			0.32	0.23	0.04			
MDR1 (7)								
rs3213619	270	274	Monomorphic				Juyal *et al.* (2009)	
rs1128503			0.37	0.43	0.05			Associated (Cao *et al.* 2015)
rs2032582								
rs1045642								Not associated (Osuga *et al.* 2006; Cao *et al.* 2015)
ADRA2A (10)	75	100					Sikander *et al.* (2010)	
rs1800544								Not investigated

Table 1 (*contd*)

Gene (chr.)	Case	Control	Minor allele frequency* Case	Minor allele frequency* Control	Allelic *P* value	Genotypic *P* value	Reference	Association status in Southeast Asian UC cohorts (reference)
NOD2 (16) rs2066844 (SNP8) rs2066845 (SNP12) rs2066847 (SNP13) rs2066842 (SNP5) rs2066843 rs1861759	411	442	Monomorphic				Mahurkar *et al.* (2011)	Monomorphic (Guo *et al.* 2004; Nakagome *et al.* 2010)
IL23R rs11209026			Monomorphic					Not associated (Peng *et al.* 2017); Monomorphic (Nakagome *et al.* 2010)
NOD1 (7) 11281T>C 11301T>C 11312T>G	95	162	0.18	0.28	0.021		Verma *et al.* (2012)	
11421A>G			0.24	0.09	0.003			
11438A>G			0.29	0.13	0.001			
11534C>T			0.35	0.18	0.001			
11585C>T			0.37	0.21	0.003			
IL4 (5) Repeat polymorphism, Intron 3	117	207					Ahirwar *et al.* (2012)	
IL10 rs1800896 rs3021097								Not associated (Kim *et al.* 2003)
TNF-alpha (6) rs1799964 rs1800630 rs1799724								Not associated (Cao *et al.* 2006)

Table 1 (*contd*)

Gene (chr.)	Case	Control	Minor allele frequency* Case	Minor allele frequency* Control	Allelic *P* value	Genotypic *P* value	Reference	Association status in Southeast Asian UC cohorts (reference)
XRCC-1 (19) rs25487	171	213				0.01	Bardia *et al.* (2012)	
APE-1 (19) rs3136820						0.04		
MIF (22) rs755622	139	176					Sivaram *et al.* (2012)	Associated (Fei *et al.* 2008)
CD14 (5) rs2569190			0.53	0.42	0.004	0.02		Associated (Kim *et al.* 2012)
TLR4 (9) rs4986790			0.12	0.06	0.01	0.03		Not associated (Okayama *et al.* 2002; Guo *et al.* 2005; Shen *et al.* 2010; Kim *et al.* 2012)
NOD2 (16) rs2066844 (SNP8) rs2066845 (SNP12) rs2066847 (SNP13) rs2067085	318	442	Monomorphic				Pugazhendhi *et al.* (2013)	Monomorphic (Guo *et al.* 2004; Nakagome *et al.* 2010)
rs2066842			0.12	0.09	0.03			
rs2066843			0.12	0.09	0.05			
rs1861759 rs2111235 rs5743266 rs2076753 rs5743291			Novel					
TLR4 (9) rs4986790	199	201	0.14	0.11	0.009		Meena *et al.* (2013)	Not associated (Cheng *et al.* 2015)
rs4986791			0.04	0.12	0.006			

Table 1 (*contd*)

Gene (chr.)	Case	Control	Minor allele frequency* Case	Minor allele frequency* Control	Allelic *P* value	Genotypic *P* value	Reference	Association status in Southeast Asian UC cohorts (reference)
XRCC3 (14)								
rs861539	171	213	0.25	0.19	0.041	0.03#	Bardia *et al.* (2014)	
RAD51								
rs1801320			0.29	0.21	0.012	0.02#		
hMSH2								
rs4987188			0.27	0.18	0.004	0.005#		
TNFSF15 (9)								
rs10114470	330	437	0.29	0.346	0.02	0.01	Baskaran *et al.* (2014a)	Not associated (Kakuta *et al.* 2006; Yang *et al.* 2011); rs3810936 associated in Japanese (Nakagome *et al.* 2010)
rs3810936			0.295	0.353	0.01	0.02		
rs6478108								
rs4263839								
rs6478109								Not associated (Kakuta *et al.* 2006)
rs7848647								Not associated (Kakuta *et al.* 2006; Yang *et al.* 2011)
rs7869487								
IRGM (5)								
rs1000113	400	448				0.0038	Baskaran *et al.* (2014b)	
rs4958847								Not associated (Lu *et al.* 2013; Moon *et al.* 2013)
rs9637876								
rs10059011								
rs11747270								
rs13361189								Not associated (Lu *et al.* 2013)
rs72553867								Not associated (Moon *et al.* 2013)
rs150227858								
rs150226250								
rs180802994			0.0691	0.0458	0.0394	0.016		

Table 1 (*contd*)

Gene (chr.)	Case	Control	Minor allele frequency* Case	Minor allele frequency* Control	Allelic *P* value	Genotypic *P* value	Reference	Association status in Southeast Asian UC cohorts (reference)
TLR1 (4)	328	350					Meena *et al.* (2015)	
rs5743611								Not associated (Kim *et al.* 2012)
TLR2 (4)								
rs5743708								Not associated (Shen *et al.* 2010; Kim *et al.* 2012; Lu *et al.* 2013)
TLR3 (4)								
rs113258886								
TLR5 (1)								
rs5744168			0.044	0.017	0.007			
rs2072493			0.09	0.11	0.021			
TLR6								
rs5743810								Not associated (Kim *et al.* 2012)
ATG16L1 (2)	249	393					Pugazhendhi *et al.* (2017)	
rs2241880								Not associated (Nakagome *et al.* 2010)
rs4663396								
rs3792106								
rs10210302						0.04		
rs3792109								
rs2241877			0.18	0.24	0.01	0.04		
rs6737398			0.35	0.41	0.05			
rs11682898								
rs4663402			0.08	0.04	0.001	0.003		
rs4663421			0.08	0.03	0.0002	0.0006		

#Significant association observed in different genetic models.
*Minor allele frequency shown only for associated SNPs.

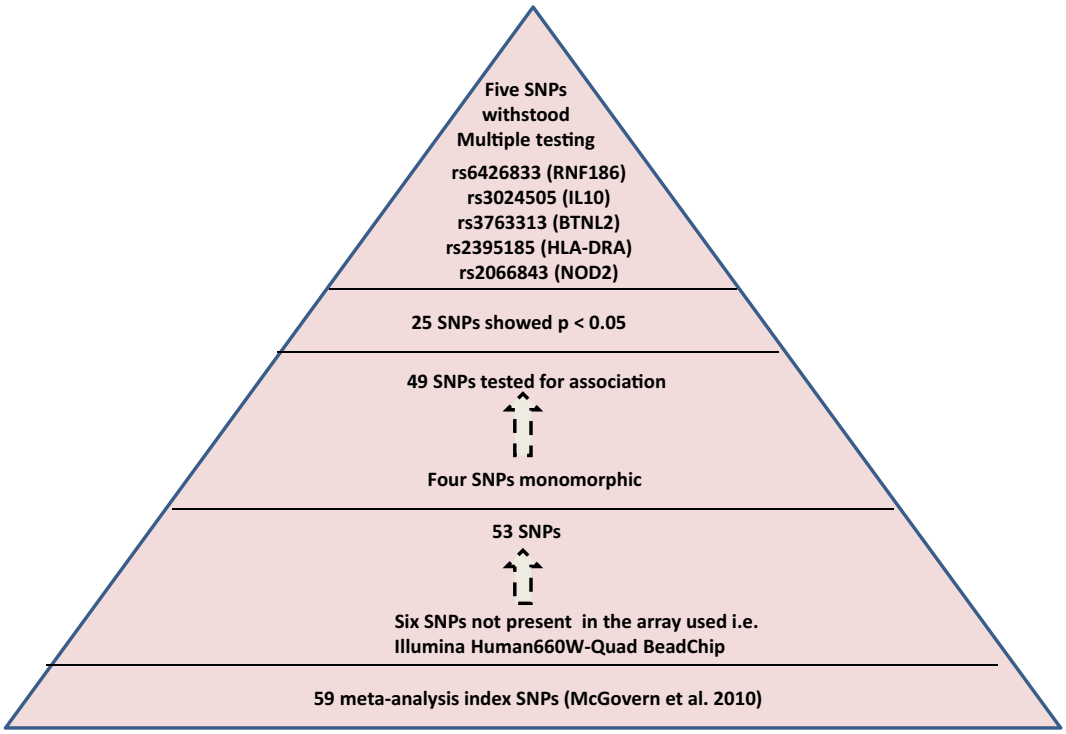

Figure 1. Limited replication of association of Caucasian UC susceptibility loci in a north Indian cohort (described in Juyal *et al.* 2011).

which resulted in meta-analysis of genomewide association scans to bolster the statistical power to identify additional loci (McGovern *et al.* 2010; Anderson *et al.* 2011). To get a sneak peek on allelic/locus heterogeneity in UC but now at a genomewide level, a total of 59 meta-analysis SNPs from 14 independent loci (attained $P < 10^{-5}$) (McGovern *et al.* 2010) were genotyped in 648 UC cases and 850 healthy controls (including the north Indian UC cohort previously analysed in candidate gene studies) which showed not so unexpected results (Juyal *et al.* 2011). Of note, six of the 59 index SNPs were not called in Indian samples and four were monomorphic. Of the remaining 49 SNPs, 25 showed association at $P < 0.05$ and of these, only five SNPs namely rs6426833 (*RNF186*, $P = 0.0004$), rs3024505 (*IL10*, $P = 0.001$), rs3763313 (*BTNL2*, $P = 0.00002$), rs2395185 (*HLA − DRA*, $P = 0.000002$) and rs2066843 (*NOD2*, $P = 0.0002$) withstood Bonferroni corrections (figure 1). The finding reaffirmed (i) only a partial concordance of Caucasian based meta-analysis results in the north Indian cohort; (ii) shared and different disease susceptibility across populations; (iii) disparity in the allele frequency of meta-analysis hits confirming differences in basal genetic architecture between populations; and (iv) limited utility of available commercial arrays for non-Caucasian population studies. This paved the way for novel gene hunt without *a priori* hypothesis and the first ever UC GWAS was conducted in the Indian population. A two tier study design comprising discovery (UC cases, 700; controls, 761) and validation phase (UC cases, 733; controls, 1148), and using an Illumina Human660W-Quad BeadChip facilitated identification of seven novel disease associated human leukocyte antigen (HLA)-independent genes/loci from chromosome 6. These included 3.8-1/*HCG26*, *BAT2*, *MSH5*, *HSPA1L*, *SLC44A4*, *CFB* and *NOTCH4*, which exceeded $P < 5 \times 10^{-8}$ in the combined analysis (table 2). Further, two novel HLA genes, namely*HLA-C* and *HLA-DQA2* were also identified in the discovery phase (Juyal *et al.* 2015). These GWAS findings yet again lent credence to genetic heterogeneity; pointed out shared and unique genetic aetiology across ethnic groups and highlighted the underrepresentation of appropriate markers in the array for non-European populations. Apart from the logical reasons i.e. gene–gene interactions and LD heterogeneity underlying these observations, varying environmental exposure (Mahid *et al.* 2006; IBD in EPIC Study Investigators *et al.* 2009; Kuenzig *et al.* 2016) across countries may also be contributors. For instance, epidemiological studies have suggested protective effect of tobacco smoke in UC and in those with established disease, smoking results in a less severe disease course (Parkes *et al.* 2014). It may be noteworthy that the north Indian GWAS cohort largely comprised nonsmokers (Juyal *et al.* 2015) as smoking causes social stigma in this part of the country. Such nongenetic factors may alter the effects of certain risk genes, resulting in apparent genetic heterogeneity, as evidenced in Indian studies. Thus, assessing genetic heterogeneity between the different populations in

combination with varying environmental exposures may well explain discordant findings across ethnic groups.

Overall, GWASs have made significant contributions to our understanding of UC biology but most of the genes identified there showed low/modest effect size leaving a substantial proportion of heritable component unexplained. CNVs, epigenetic effects, gene–gene interactions and gene–environment interactions have been considered to address the remaining unmapped heritability but actual extent of their contribution remains to be assessed. On the other hand, screening larger sample sizes to identify residual common and rare variant associations has been acknowledged to increase the narrow-sense heritability (i.e. proportion of trait variance which is due to additive genetic factors). This led to the creation of the International IBD Genetics Consortium (IIBDGC) (http://www.ibdgenetics.org/) which brought together several investigators and IBD GWAS datasets from across continents. This collaborative effort led to the identification of 71 new genetic loci for IBD in cohorts of European descent (Jostins *et al.* 2012). These encouraging findings together with evidence of shared IBD risk loci across diverse populations from previous studies and the notable success of trans-ethnic association studies in other complex diseases, including type-2 diabetes (Cook and Morris 2016) and rheumatoid arthritis (Okada *et al.* 2014) led to the first trans-ethnic association study of IBD. This effort was anticipated to identify additional IBD risk loci and compare the canvas of genetic risk of IBD across ancestrally diverse populations. Another feature of this study was the use of Immunochip, a custom designed genotyping array. This contained 196,524 variants across 186 known autoimmunity risk loci (Cortes and Brown 2010), which were previously described in European populations. Besides, the array also tagged rare variants which might have been previously overlooked. The chip was designed to assist in the replication, fine-mapping and discovery of loci associated with inflammatory and autoimmune diseases (Cortes and Brown 2010). Briefly, this study utilized genomewide or Immunochip data of 86,640 European individuals and Immunochip data from 9846 individuals of East Asian (Japan, Korea and Hong Kong), north Indian or Iranian descent and raised the number of known IBD risk loci to 200. This largest trans-ethnic study highlighted shared genetic risk across European and non-European cohorts for most IBD risk loci, although genetic heterogeneity remained with differences in minor allele frequency and or effect sizes at several important and known loci such as *TNFSF15*/*TNFSF8*, *IL23R*, etc. Such differences in effect size may reflect differences in gene–environment interactions between populations.

Despite this study having been insightful and the Immunochip had a dense coverage of immune related genes based on resequencing data (albeit from individuals of European ancestry), lack of association of the top most Indian GWAS hits, including *CFB*, *SLC44A4*, etc. in the European cohort included in the study encouraged further investigations. Keeping in view that the shared risk loci identified in this study were more between Europeans and East Asian cohorts (which had large sample sets), it was important to revisit the Immunochip datasets separately for the populations. In such an effort, genotype data already available for three divergent populations, namely Dutch (1729 cases and 1350 controls), Japanese (719 cases and 3263 controls) and north Indians (897 cases and 896 controls), were reanalysed with focus on two biologically relevant India GWAS genes, namely *CFB* and *SLC44A4*. Of the 28 SNPs in *CFB* present on the Immunochip, nearly 40% ($n = 13$) were monomorphic in north Indians and Japanese. This was not unexpected considering the Immunochip design, which is based on Caucasian SNP data. Notably, the reported north Indian UC GWAS index SNP rs4151657 within *CFB* (Juyal *et al.* 2015) showed strong allelic association in the Japanese ($P = 2.02 \times 10^{-8}$), but was nominally associated in the Dutch cohort ($P = 0.002$) (Gupta *et al.* 2016). In addition, another SNP rs537160 was suggestively significant ($P = 4.29 \times 10^{-5}$) in the Dutch cohort, which pointed to allelic heterogeneity yet again. This was reiterated by haplotype analysis which revealed a minimal three-marker haplotype, namely rs17201431 – rs2072634 – rs4151657 shared across north Indians and Japanese ($P < 10^{-8}$), but a different five-marker haplotype, namely rs4151651–rs4151652–rs17201431–rs512559–rs537160 significantly associated ($P = 2.07 \times 10^{-6}$) in the Dutch population after Bonferroni corrections. Similarly, GWAS index SNP rs2736428 within *SLC44A4*, on the other hand was also found to be significantly associated in Japanese ($P = 3.37 \times 10^{-9}$) but only nominally associated ($P = 0.002$) in the Dutch cohort. Other than this, 11 of the 22 SNPs within *SLC44A4* showed nominal association in all three ethnic groups, most of which were predicted to have regulatory effects. Thus, reanalysis of the Immunochip data helped detect comparable strength of association of India GWAS hits with Japanese UC cohort, missed in their original GWAS. It is likely that since Japanese GWAS was performed using Illumina HumanHap550v3 Genotyping BeadChip, these SNPs were not present in the array or they may have failed quality control. Alternatively, due to the smaller sample size in their GWAS, this association was not identified (Asano *et al.* 2009). It should be worth mentioning here that a recent study from the Han Chinese population provided further support for the role of *SLC44A4* in the pathogenesis of UC (Wu *et al.* 2017). Putting together these findings highlighted that genetic susceptibility underlying UC varies across ethnic groups. Further, although substantial proportions of disease associated genes may be shared across ethnic groups, population-wise detailed investigation of such risk genes warrant revisiting genomic data/generation of new data by resequencing approaches to identify critical variants relevant to the respective populations.

Table 2. Novel UC genes identified in genetically distinct north Indians (described in Juyal *et al.* 2015).

Gene	SNP	Risk allele	OR_{comb} (95% CI)*	Function	Supporting evidence for Indian GWAS hits in UC/IBD by other studies
CFB	rs4151657	G	1.54 (1.38–1.73)	A component of the alternative pathway of complement activation. The complement system is an important part of the innate immune system and the first-line host defence. The main effector functions of complement activation are opsonisation, chemotaxis, phagocytosis, and induction of inflammatory mediators	Ostvik *et al.* (2014) and Shaw (2018)
BAT2	rs2261033	G	1.34 (1.21–1.48)	Localized in the vicinity of the genes for TNF alpha and TNF beta. Involved in immune response (Choufi *et al.* 2006). Candidate gene for a related inflammatory disorder i.e. rheumatoid arthritis (Singal *et al.* 1999)	–
SLC44A4	rs2736428	A	1.43 (1.28–1.59)	Transporter that plays an important role in the absorption of the microbiota-generated thiamine pyrophosphate (TPP) in the large. Further support the hypothesis that the microbiota-generated TPP is absorbable and could contribute toward host thiamine homeostasis, especially toward cellular nutrition of colonocytes (Nabokina *et al.* 2014, 2016)	Wu *et al.* (2017)
MSH5	rs707939	A	1.51 (1.32–1.72)	Plays a role in DNA damage response and double-strand base repair, immunoglobulin diversity, and has been linked to neoplasia (including colorectal cancer), immune disease, and reproductive disorders	Kelsen *et al.* (2015)
HSPA1L	rs2075800	A	1.41 (1.27–1.56)	The HSP70 proteins play multiple roles in protein quality control of the cell, including refolding denatured proteins, preventing aggregation and intracellular protein transport. Have shown to modulate inflammatory response and exert anti-apoptotic functions, both of which are closely related to IBD pathogenesis. However, the distinct role of each HSP70 family member is not well understood, and their potential role in IBD has not been established	Takahashi *et al.* (2017)
HCG26	rs3749946	A	1.97 (1.66–2.33)	HLA complex group 26 (nonprotein coding); RNA gene	–
NOTCH4	rs549182	A	1.59 (1.37–1.85)	Part of NOTCH signalling family, which is involved in intestinal epithelium maintenance as well as in regeneration	Zheng *et al.* (2011)

*OR_{comb} refers to the values obtained on combined analysis of discovery and replication cohorts in Juyal *et al.* (2015).

Finally, Immunochip seems to be less informative in non-Europeans and therefore, efforts for population specific target arrays are required from the latter populations for fine mapping.

In summary, it is evident from the preceding sections that UC is an emerging disease in the Indian subcontinent (Sood *et al.* 2003; Singh *et al.* 2017) with likely differences in genetic predisposition warranting additional

genetic studies as well as characterizing the respective environmental background. This in other words, highlights the imminent need for systematic regional epidemiological surveys across the country particularly in view of dietary, geographical and socio-cultural differences. One of the nongenetic tractable components among these is the gut microbiome.

Future perspectives

Besides obtaining insights into the overall genetic underpinnings and biology of UC, the UC research community is also making efforts to identify genetic markers which have a potential to stratify patients into prognostic groups. Such stratification will enable unraveling of specific molecular pathways associated with disease course and therapeutic response, which may then serve as putative biomarkers and/or novel drug targets. It is true that patients have so far not been greatly benefitted from genetic findings. Nevertheless, over the preceding decades, the realm of UC genetics has made impressive progress and is continually and exponentially increasing, providing new and deeper insights into disease mechanism with implications for novel therapeutics, and also for drug repurposing. Genetic determinants identified (to some extent) have proved helpful in determining clinical subphenotypes, predicting disease course and development of new treatment modalities. However, such examples are more in CD as compared to UC, which can easily be explained based on a stronger genetic component in CD. Some of these clinically useful clinico-genetic correlates include: association of HLA with extensive disease and colectomy in UC (Haritunians *et al.* 2010); HLA and colonic CD (Silverberg *et al.* 2003); *NOD2* variants, ileal location and the need for CD-related surgery (Cleynen *et al.* 2013). Besides, an associated locus near *SMAD7* (Kennedy 2015) and *ITGAL* (James *et al.* 2011) have already been shown to have important applications in the development of new treatments. At the same time, as mentioned earlier, genetic findings have also provided evidence of potentially disparate genetic contributors to UC biology among ethnically diverse populations owing to genetic distinctness and different environmental disease triggers/modifiers etc. considering these and in view of limited genetic studies from India on IBD, we believe discovery efforts from this population must continue so as to identify common genetic variants of small effect and rare genetic variants of moderate/large effects and their cumulative contribution will undoubtedly be very insightful and may add to the heritability component. Given common variants are mostly shared across ethnically divergent populations, combining pan India IBD cohorts may prove more useful.

It must be acknowledged that genetic studies either on UC or any other complex traits have indeed shedlight into disease pathogenesis; these findings also indicate in parallel a few major limitations. Genetic markers alone will most likely never fully predict disease susceptibility, disease course and treatment outcome due to genetic heterogeneity; incomplete penetrance; modest to low frequency of the risk variants; unequivocal role of nongenetic factors in shaping the disease; and almost unaddressed clinical/phenotypic heterogeneity and therefore, search for other aetiological components is obviously necessary.

It may be relevant to mention here that nonreplication has just not been restricted to genetic findings. Laboratory markers available for IBD have also shown low specificity as their application remains limited beyond the populations in which they have been developed. For example, a recent study revealed presence of different serological antibodies between ethnically divergent IBD populations (Prideaux *et al.* 2012). Therefore, efforts towards biomarker discovery and novel drug target identification for Indian population are inevitable. Taken together that UC is both genetically and clinically heterogeneous, this enforces use of integrative, multidisciplinary OMICS approaches to obtain a range of bio-markers for effective/practical clinical utility. In addition, comprehensive clinical data would significantly enhance the momentum of biomarker discovery. Towards these efforts, analysis of the most tractable environmental component, i.e. gut microbiome in low-prevalence areas and which are now experiencing a rapid growth of the disease, would be an equally if not more important aspect. Therefore, newer approaches/tools with improved study design to (i) enhance the understanding of genetic/mechanistic basis of disease; (ii) identify biomarkers of immediate translational value; and (iii) develop newer alternate therapeutic options are envisaged to be the way forward.

Acknowledgements

Work on Indian UC patients covered in this review was supported in part by funds from the International Organization of Inflammatory Bowel Diseases (USA); Centre of Excellence in Genome Sciences and Predictive Medicine (#BT/01/COE/07/UDSC/2008 to BKT, AS and VM) from the Department of Biotechnology, Government of India; and Science and Engineering Research Board, Government of India (#SB/YS/LS-191/2014 to GJ). We acknowledge the International Inflammatory Bowel Disease Genetics Consortium for the Immunochip genotyping data.

References

Ahirwar D. K., Kesarwani P., Singh R., Ghoshal U. C. and Mittal R. D. 2012 Role of tumor necrosis factor-alpha (C-863A) polymorphism in pathogenesis of inflammatory bowel disease in northern India. *J. Gastrointest. Cancer* **43**, 196–204.

Altshuler D., Daly M. J. and Lander E. S. 2008 Genetic mapping in human disease. *Science* **322**, 881–888.

Anderson C. A., Boucher G., Lees C. W., Franke A., D'Amato M., Taylor K. D. *et al.* 2011 Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246–252.

Asano K., Matsushita T., Umeno J., Hosono N., Takahashi A., Kawaguchi T. *et al.* 2009 A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nat. Genet.* **41**, 1325–1329.

Bardia A., Tiwari S. K., Gunisetty S., Anjum F., Nallari P., Habeeb M. A. and Khan A. A. 2012 Functional polymorphisms in XRCC-1 and APE-1 contribute to increased apoptosis and risk of ulcerative colitis. *Inflamm. Res.* **61**, 359–65.

Bardia A., Tiwari S. K., Vishwakarma S. K., Habeeb M. A., Nallari P., Sultana S. A. *et al.* 2014 Haplotype analyses of DNA repair gene polymorphisms and their role in ulcerative colitis. *PLoS One* **9**, e108562.

Baskaran K., Pugazhendhi S. and Ramakrishna B. S. 2014a Protective association of tumor necrosis factor superfamily 15 (TNFSF15) polymorphic haplotype with ulcerative colitis and Crohn's disease in an Indian population. *PLoS One* **9**, e114665.

Baskaran K., Pugazhendhi S. and Ramakrishna B. S. 2014b Association of IRGM gene mutations with inflammatory bowel disease in the Indian population. *PLoS One* **9**, e106863.

Cao Q., Zhu Q., Wu M. L., Hu W. L., Gao M. and Si J. M. 2006 Genetic susceptibility to ulcerative colitis in the Chinese Han ethnic population: association with TNF polymorphisms. *Chin. Med. J. (Engl.)* **119**, 1198–1203.

Cao Y., Qu C., Chen Y., Li L. and Wang X. 2015 Association of ABCB1 polymorphisms and ulcerative colitis susceptibility. *Int. J. Clin. Exp. Pathol.* **8**, 943–947.

Cheng Y., Zhu Y., Huang X., Zhang W., Han Z. and Liu S. 2015 Association between TLR2 and TLR4 gene polymorphisms and the susceptibility to inflammatory bowel disease: a meta-analysis. *PLoS One* **10**, e0126803.

Choufi B., Chalabi N., Le Corre L., Delort L., Satih S., Bignon Y. *et al.* 2006 Gene expression in human acute cutaneous and hepatic graft versus host disease after allogeneic bone marrow transplantation. *Cancer Genom. Proteom.* **3**, 113–118.

Cleynen I., González J. R., Figueroa C., Franke A., McGovern D., Bortlík M. *et al.* 2013 Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut* **62**, 1556–1565.

Cook J. P. and Morris A. P. 2016 Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. *Eur. J. Hum. Genet.* **24**, 1175–1180.

Cortes A. and Brown M. A. 2010 Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101.

Dubinsky M. C. 2017 Reviewing treatments and outcomes in the evolving landscape of ulcerative colitis. *J. Postgrad. Med.* **129**, 538–553.

Ek W. E., D'Amato M. and Halfvarson J. 2014 The history of genetics in inflammatory bowel disease. *Ann. Gastroenterol.* **27**, 294–303.

Fisher S. A., Tremelling M., Anderson C. A., Gwilliam R., Bumpstead S., Prescott N. J. *et al.* 2008 Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat. Genet.* **40**, 710–712.

Fei B. Y., Lv H. X., Yang J. M. and Ye Z. Y. 2008 Association of MIF-173 gene polymorphism with inflammatory bowel disease in Chinese Han population. *Cytokine* **41**, 44–47.

Franke A., Balschun T., Karlsen T. H., Sventoraityte J., Nikolaus S., Mayr G. *et al.* 2008 Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat. Genet.* **40**, 1319–1323.

Guo Q. S., Xia B., Jiang Y., Qu Y. and Li J. 2004 NOD2 3020insC frameshift mutation is not associated with inflammatory bowel disease in Chinese patients of Han nationality. *World J. Gastroenterol.* **10**, 1069–1071.

Guo Q. S., Xia B., Jiang Y., Morré S. A., Cheng L., Li J. *et al.* 2005 Polymorphisms of CD14 gene and TLR4 gene are not associated with ulcerative colitis in Chinese patients. *J. Postgrad. Med.* **81**, 526–529.

Gupta A., Juyal G., Sood A., Midha V., Yamazaki K., Vich Vila A. *et al.* 2016 A cross-ethnic survey of CFB and SLC44A4, Indian ulcerative colitis GWAS hits, underscores their potential role in disease susceptibility. *Eur. J. Hum. Genet.* **25**, 111–122.

Haritunians T., Taylor K. D., Targan S. R., Dubinsky M., Ippoliti A., Kwon S. *et al.* 2010 Genetic predictors of medically refractory ulcerative colitis. *Inflamm. Bowel Dis.* **16**, 1830–1840.

IBD in EPIC Study Investigators, Tjonneland A., Overvad K., Bergmann M. M., Nagel G., Linseisen J. *et al.* 2009 Linoleic acid, a dietary n-6 polyunsaturated fatty acid, and the aetiology of ulcerative colitis: a nested case-control study within a European prospective cohort study. *Gut* **58**, 1606–1611.

International HapMap Consortium, Frazer K. A., Ballinger D. G., Cox D. R., Hinds D. A., Stuve L. L. *et al.* 2007 A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861.

James D. G., Seo D. H., Chen J., Vemulapalli C. and Stone C. D. 2011 Efalizumab, a human monoclonal anti-CD11a antibody, in the treatment of moderate to severe Crohn's Disease: an open-label pilot study. *Dig. Dis. Sci.* **56**, 1806–1810.

Jostins L., Ripke S., Weersma R. K., Duerr R. H., McGovern D. P., Hui K. Y. *et al.* 2012 Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124.

Juyal G., Amre D., Midha V., Sood A., Seidman E. and Thelma B. K. 2007 Evidence of allelic heterogeneity for associations between the NOD2/CARD15 gene and ulcerative colitis among North Indians. *Aliment. Pharmacol. Ther.* **26**, 1325–1332.

Juyal G., Midha V., Amre D., Sood A., Seidman E. and Thelma B. K. 2009 Associations between common variants in the MDR1 (ABCB1) gene and ulcerative colitis among North Indians. *Pharmacogenet. Genomics* **19**, 77–85.

Juyal G., Prasad P., Senapati S., Midha V., Sood A., Amre D. *et al.* 2011 An investigation of genome-wide studies reported susceptibility loci for ulcerative colitis shows limited replication in north Indians. *PLoS One* **6**, e16565.

Juyal G., Negi S., Sood A., Gupta A., Prasad P., Senapati S. *et al.* 2015 Genome-wide association scan in north Indians reveals three novel HLA-independent risk loci for ulcerative colitis. *Gut* **64**, 571–579.

Kakuta Y., Kinouchi Y., Negoro K., Takahashi S. and Shimosegawa T. 2006 Association study of TNFSF15 polymorphisms in Japanese patients with inflammatory bowel disease. *Gut* **55**, 1527–1528.

Kaplan G. G. 2015 The global burden of IBD: from 2015 to 2025. *Nat. Rev. Gastroenterol. Hepatol.* **12**, 720–727.

Kedia S. and Ahuja V. 2017 Epidemiology of inflammatory bowel disease in India: the great shift east. *Inflamm. Intest. Dis.* **2**, 102–115.

Kelsen J. R., Dawany N., Moran C. J., Petersen B. S., Sarmady M., Sasson A. *et al.* 2015 Exome sequencing analysis reveals variants in primary immunodeficiency genes in patients with very early onset inflammatory bowel disease. *Gastroenterology* **149**, 1415–1424.

Kennedy B. W. 2015 Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. *N. Engl. J. Med.* **372**, 2461.

Kim T. H., Kim B. G., Shin H. D., Kim J. W., Kim C. G., Kim J. S. *et al* . 2003 Tumor necrosis factor alpha and interleukin-10 gene polymorphisms in Korean patients with inflammatory bowel disease. *Korean J. Gastroenterol.* **42**, 377–386.

Kim E. J., Chung W. C., Lee K. M., Paik C. N., Jung S. H., Lee B. *et al.* 2012 Association between toll-like receptors/CD14 gene polymorphisms and inflammatory bowel disease in Korean population. *J. Korean Med. Sci*. **27**, 72–77.

Kuenzig M. E., Lee S. M., Eksteen B., Seow C. H., Barnabe C., Panaccione R. *et al.* 2016 Smoking influences the need for surgery in patients with the inflammatory bowel diseases: a systematic review and meta-analysis incorporating disease duration. *BMC Gastroenterol.* **16**, 143.

Li Y. R. and Keating B. J. 2014 Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med.* **6**, 91.

Liu J. Z., van Sommeren S., Huang H., Ng S. C., Alberts R., Takahashi A. *et al.* 2015 Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986.

Loddo I. and Romano C. 2015 Inflammatory bowel disease: genetics, epigenetics, and pathogenesis. *Front. Immunol.* **6**, 551.

Lu X. C., Tao Y., Wu C., Zhao P. L., Li K., Zheng J. Y. *et al.* 2013 Association between variants of the autophagy related gene–IRGM and susceptibility to Crohn's disease and ulcerative colitis: a meta-analysis. *PLoS One* **8**, e80602.

Mahid S. S., Minor K. S., Soto R. E., Hornung C. A. and Galandiuk S. 2006 Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin. Proc.* **81**, 1462–1471.

Mahurkar S., Banerjee R., Rani V. S., Thakur N., Rao G. V., Reddy D. N. *et al.* 2011 Common variants in NOD2 and IL23R are not associated with inflammatory bowel disease in Indians. *J. Gastroenterol. Hepatol.* **26**, 694–699.

Mathew C. G. and Lewis C. M. 2004 Genetics of inflammatory bowel disease: progress and prospects. *Hum. Mol. Genet.* **13**, 161–168.

McGovern D. P., Gardet A., Törkvist L., Goyette P., Essers J., Taylor K. D. *et al.* 2010. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.* **41**, 332–337.

McVean G. A. T., Myers S. R., Hunt S., Deloukas P., Bentley D. R. and Donnelly P. 2004 The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584.

Meena N. K., Ahuja V., Meena K. and Paul J. 2015 Association of TLR5 gene polymorphisms in ulcerative colitis patients of north India and their role in cytokine homeostasis. *PLoS One* **10**, e0120697.

Meena N. K., Verma R., Verma N., Ahuja V. and Paul J. 2013 TLR4 D299G polymorphism modulates cytokine expression in ulcerative colitis. *J. Clin. Gastroenterol.* **47**, 773–780.

Mittal R. D., Bid H. K. and Ghoshal U. C. 2005 IL-1 receptor antagonist (IL-1Ra) gene polymorphism in patients with inflammatory bowel disease in India. *Scand. J. Gastroenterol.* **40**, 827–831.

Mittal R. D., Manchanda P. K., Bid H. K. and Ghoshal U. C. 2007 Analysis of polymorphisms of tumor necrosis factor-alpha and polymorphic xenobiotic metabolizing enzymes in inflammatory bowel disease: study from northern India. *J. Gastroenterol. Hepatol.* **22**, 920–924.

Moon C. M., Shin D. J., Kim S. W., Son N. H., Park A., Park B. *et al.* 2013 Associations between genetic variants in the IRGM gene and inflammatory bowel diseases in the Korean population. *Inflamm. Bowel. Dis.* **19**, 106–114.

Nabokina S. M., Subramanian V. S. and Said H. M. 2016 The human colonic thiamine pyrophosphate transporter (hTPPT) is a glycoprotein and N-linked glycosylation is important for its function. *Biochim. Biophys. Acta* **1858**, 866–871.

Nabokina S. M., Inoue K., Subramanian V. S., Valle J. E., Yuasa H. and Said H. M. 2014 Molecular identification and functional characterization of the human colonic thiamine pyrophosphate transporter. *J. Biol. Chem.* **289**, 4405–4416.

Nakagome S., Takeyama Y., Mano S., Sakisaka S., Matsui T., Kawamura S. *et al.* 2010 Population-specific susceptibility to Crohn's disease and ulcerative colitis; dominant and recessive relative risks in the Japanese population. *Ann. Hum. Genet.* **74**, 126–136

Ng S. C., Tang W., Ching J. Y., Wong M., Chow C. M., Hui A. J. *et al.* 2013 Incidence and phenotype of inflammatory bowel disease based on results from the Asia-pacific Crohn's and colitis epidemiology study. *Gastroenterology* **145**, 158–165.

Ng W. K., Wong S. H. and Ng S. C. 2016 Changing epidemiological trends of inflammatory bowel disease in Asia. *Intest. Res.* **14**, 111.

Nohara H., Inoue N., Hibi T., Okita K. and Hinoda Y. 2003 Association between the interleukin-1 receptor antagonist polymorphism and ulcerative colitis with younger age at diagnosis. *Immunol. Lett.* **90**, 53–57.

Okada Y., Yamazaki K., Umeno J., Takahashi A., Kumasaka N., Ashikawa K. *et al.* 2011 HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. *Gastroenterology* **141**, 864–871.

Okada Y., Wu D., Trynka G., Raj T., Terao C., Ikari K. *et al.* 2014 Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381.

Okayama N., Fujimura K., Suehiro Y., Hamanaka Y., Fujiwara M., Matsubara T. *et al.* 2002 Simple genotype analysis of the Asp299Gly polymorphism of the Toll-like receptor-4 gene that is associated with lipopolysaccharide hyporesponsiveness. *J. Clin. Lab. Anal.* **16**, 56–58.

Ostvik A. E., Granlund Av, Gustafsson B. I., Torp S. H., Espevik T., Mollnes T. E. *et al.* 2014 Mucosal toll-like receptor 3-dependent synthesis of complement factor B and systemic complement activation in inflammatory bowel disease. *Inflamm. Bowel Dis.* **20**, 995–1003.

Osuga T., Sakaeda T., Nakamura T., Yamada T., Koyama T., Tamura T. *et al.* 2006 MDR1 C3435T polymorphism is predictive of later onset of ulcerative colitis in Japanese. *Biol. Pharm. Bull.* **29**, 324–329.

Park S. J., Kim W. H. and Cheon J. H. 2014 Clinical characteristics and treatment of inflammatory bowel disease: a comparison of Eastern and Western perspectives. *World J. Gastroenterol.* **20**, 11525–11537.

Parkes G. C., Whelan K. and Lindsay J. O. 2014 Smoking in inflammatory bowel disease: Impact on disease course and insights into the aetiology of its effect. *J. Crohns Colitis* **8**, 717–725.

Peng L. L., Wang Y., Zhu F. L., Xu W. D., Ji X. L. and Ni J. 2017 IL-23R mutation is associated with ulcerative colitis: A systemic review and meta-analysis. *Oncotarget* **8**, 4849–4863.

Prideaux L., De Cruz P., Ng S. C. and Kamm M. A. 2012 Serological antibodies in inflammatory bowel disease: a systematic review. *Inflamm. Bowel Dis.* **18**, 1340–1355.

Pugazhendhi S., Baskaran K., Santhanam S. and Ramakrishna B. S. 2017 Association of ATG16L1 gene haplotype with inflammatory bowel disease in Indians. *PLoS One* **12**, e0178291.

Pugazhendhi S., Santhanam S., Venkataraman J., Creveaux I. and Ramakrishna B. S. 2013 NOD2 gene mutations associate weakly with ulcerative colitis but not with Crohn's disease in Indian patients with inflammatory bowel disease. *Gene* **512**, 309–313.

Risch N. and Merikangas K. 1996 The future of genetic studies of complex human diseases. *Science* **273**, 1516–1517.

Sachidanandam R., Weissman D., Schmidt S. C., Kakol J. M., Stein L. D., Marth G. *et al.* 2001 A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933.

Sashio H., Tamura K., Ito R., Yamamoto Y., Bamba H., Kosaka T. *et al.* 2002 Polymorphisms of the TNF gene and the TNF receptor superfamily member 1B gene are associated with susceptibility to ulcerative colitis and Crohn's disease, respectively. *Immunogenetics* **53**, 1020–1027.

Shaw K. A., Cutler D. J., Okou D., Dodd A., Aronow B. J., Haberman Y. *et al.* 2018 Genetic variants and pathways implicated in a pediatric inflammatory bowel disease cohort. *Genes Immun.* https://doi.org/10.1038/s41435-018-0015-2.

Shen X. Y., Shi R. H., Wang Y., Zhang H. J., Zhou X. Q., Shen F. C. *et al.* 2010 Toll-like receptor gene polymorphisms and susceptibility to inflammatory bowel disease in Chinese Han and Caucasian populations. *Zhonghua Yi Xue Za Zhi* **90**, 1416–1420.

Sherry S. T., Ward M. and Sirotkin K. 1999 dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9**, 677–679.

Sikander A., Rana S. V., Sharma S. K., Sinha S. K., Arora S. K. and Prasad K. K. 2010 Association of alpha 2A adrenergic receptor gene (ADRAlpha2A) polymorphism with irritable bowel syndrome, microscopic and ulcerative colitis. *Clin. Chim. Acta* **411**, 59–63.

Silverberg M. S., Mirea L., Bull S. B., Murphy J. E., Steinhart A. H., Greenberg G. R. *et al.* 2003 A population- and family-based study of Canadian families reveals association of HLA DRB1*0103 with colonic involvement in inflammatory bowel disease. *Inflamm. Bowel Dis.* **9**, 1–9.

Silverberg M. S., Cho J. H., Rioux J. D., McGovern D. P., Wu J., Annese V. *et al.* 2009 Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat. Genet.* **41**, 216–220.

Singal D. P., Li J. and Lei K. 1999 Genetics of rheumatoid arthritis (RA): two separate regions in the major histocompatibility complex contribute to susceptibility to RA. *Immunol. Lett.* **69**, 301–306.

Singh P., Ananthakrishnan A. and Ahuja V. 2017 Pivot to Asia: inflammatory bowel disease burden. *Intest. Res.* **15**, 138–141.

Sivaram G., Tiwari S. K., Bardia A., Anjum F., Vishnupriya S. and Habeeb A. 2012 Macrophage migration inhibitory factor, Toll-like receptor 4, and CD14 polymorphisms with altered expression levels in patients with ulcerative colitis. *Hum. Immunol.* **73**, 201–205.

Song Y., Wu K. C., Zhang L., Hao Z. M., Li H. T., Zhang L. X. *et al.* 2005 Correlation between a gene polymorphism of tumor necrosis factor and inflammatory bowel disease. *Chin. J. Dig. Dis.* **6**, 170–174.

Sood A., Midha V., Sood N., Bhatia A. S. and Avasthi G. 2003 Incidence and prevalence of ulcerative colitis in Punjab, North India. *Gut* **52**, 1587–1590.

Takahashi S., Andreoletti G., Chen R., Munehira Y., Batra A., Afzal N. A. *et al.* 2017 De novo and rare mutations in the HSPA1L heat shock gene associated with inflammatory bowel disease. *Genome Med.* **9**, 8.

Thorisson G. A. and Stein L. D. 2003 The SNP Consortium website: past, present and future. *Nucleic Acids Res.* **31**, 124–127.

Qian J., Song Z., Lv Y., Huang X. and Mao B. 2017 Glutathione S-transferase T1 null genotype is associated with susceptibility to inflammatory bowel disease. *Cell Physiol. Biochem.* **41**, 2545–2552.

UK IBD Genetics Consortium, Barrett J. C., Lee J. C., Lees C. W., Prescott N. J., Anderson C. A. *et al.* 2009 Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat. Genet.* **41**, 1330–1334.

Verma R., Ahuja V. and Paul J. 2009 Frequency of single nucleotide polymorphisms in NOD1 gene of ulcerative colitis patients: a case-control study in the Indian population. *BMC Med. Genet.* **10**, 82.

Verma R., Ahuja V. and Paul J. 2012 Detection of single-nucleotide polymorphisms in the intron 9 region of the nucleotide oligomerization domain-1 gene in ulcerative colitis patients of North India. *J. Gastroenterol. Hepatol.* **27**, 96–103.

Walters T. D. and Silverberg M. S. 2006 Genetics of inflammatory bowel disease: current status and future directions. *Can. J. Gastroenterol.* **20**, 633–639.

Wu J., Cheng Y., Zhang R., Shen H., Ma L., Yang J. *et al.* 2017 Evaluating the association of common variants of the SLC44A4 gene with ulcerative colitis susceptibility in the Han Chinese population. *Genet. Test. Mol. Biomarkers* **21**, 1–5

Yang S. K., Jung Y., Hong M., Kim H., Ye B. D., Lee I. *et al.* 2011 No association between TNFSF15 and IL23R with ulcerative colitis in Koreans. *J. Hum. Genet*. **56**, 200–204.

Yang S. K., Hong M., Zhao W., Jung Y., Tayebi N., Ye B. D. *et al.* 2013 Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. *Inflamm. Bowel Dis.* **19**, 954–966.

Ye B. D. and McGovern D. P. B. 2016 Genetic variation in IBD: progress, clues to pathogenesis and possible clinical utility. *Expert Rev. Clin. Immunol.* **12**, 1091–1107.

Ye X., Jiang Y., Wang H., Chen L., Yuan S. and Xia B. 2011 Genetic polymorphisms of glutathione S-transferases are associated with ulcerative colitis in central China. *Cell Biochem. Biophys.* **60**, 323–328.

Zheng X., Tsuchiya K., Okamoto R., Iwasaki M., Kano Y., Sakamoto N. *et al.* 2011 Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis. *Inflamm. Bowel Dis.* **17**, 2251–2260.

Corresponding editor: H. A. Ranganath