Manuscript Doi: 10.1093/ecco-jcc/jjac124

NOD2 in Crohn's disease- unfinished business

James J Ashton PhD[1,2], Eleanor G Seaby[1], R Mark Beattie FRCPCH[2], and Sarah Ennis PhD[1]

1. Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton, UK
2. Department of Paediatric Gastroenterology, Southampton Children's Hospital, Southampton, UK

Correspondence to-

Professor Sarah Ennis

Human Genetics & Genomic Medicine

University of Southampton

Duthie Building (Mailpoint 808)

Southampton General Hospital

Southampton

SO16 6YD

UK

Tel: +44 (0)23 8079 8614

s.ennis@soton.ac.uk

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Acknowledgements

We would like to acknowledge support from Imogen S Stafford and Guo Cheng in the development of this article. Figures created with BioRender.com.

Conflicts of interest

The authors declare no conflicts of interest

Author Contributions

JJA and SE conceived the article. JJA gathered the data and wrote the article with help from all authors. All authors approved the final manuscript prior to submission

Funding

JJA is funded by an NIHR clinical lectureship (Southampton) and an ESPR post-doctoral award. There was no specific funding for this article.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Abstract

Studies of Crohn's disease consistently implicate *NOD2* as the most important gene in disease pathogenesis since first being identified in 2001. Since this point, genome-wide association, next-generation sequencing, and functional analyses have all confirmed a key role for *NOD2*, but despite this, *NOD2* also has significant unresolved complexity. More recent studies have reinvigorated an early hypothesis that *NOD2* may be a single-gene cause of disease, and the distinct ileal stricturing phenotype seen with *NOD2*-related disease presents an opportunity for personalised diagnosis, disease prediction and targeted therapy.

The genomics of *NOD2* has much that remains unknown, including the role of rare variation, phasing of variants across the haplotype block and the role of variation in the *NOD2*-regulatory regions. Here, we discuss the evidence and the unmet needs of *NOD2*-research, based on recently published evidence, and suggest methods that may meet these requirements.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Introduction

NOD2 is the most implicated gene in the aetiology of Crohn's disease (CD)[1]. The role of *NOD2* in development of CD has been acknowledged for over 20 years although the true extent of the involvement in disease pathogenesis is probably underestimated and remains largely unrecognised in the clinical setting. Heavily replicated genome wide association studies (GWAS) studies, contemporary next-generation sequencing data and novel functional annotation directly link *NOD2* dysfunction to intestinal fibrosis and suggest *NOD2's* contribution to Crohn's disease has significant and unresolved complexity[2–4]. A fuller understanding of the relationship between specific *NOD2* genetic variants, protein dysfunction and patient phenotypes has the potential to inform clinical management and have direct benefit to patients. In this article we discuss what is known about *NOD2*, including the evidence to suggest a significant potential for clinical translation, and discuss the limitations and gaps in current knowledge and data. Finally, we propose potential solutions to address this unmet need to translate the potential of *NOD2* as a clinical tool into clinical practice.

Clinical and molecular heterogeneity seen in CD has consistently pointed to multiple aetiologies, manifesting as different, but related, phenotypes with intestinal inflammation and a varied disease course. Differing complications and varying treatment response make better patient risk stratification a priority. The strong genetic component is well known, and was first suggested by twin studies, demonstrating a 50-60% heritability compared to other complex diseases[5]. Non-parametric linkage analysis by Hugot *et al* in 1996 implicated a locus on chromosome 16 [6]. The NOD2 gene was characterised in 2001 - now known to be expressed in granulocytes, dendritic cells and T-cells, with the highest expression in terminal ileal Paneth cells[7]. Further evidence detailing the protein function revealed the *NOD2* gene encodes an intracellular receptor for the bacterial peptidoglycan muramyl dipeptide (MDP), that upon bacterial sensing, forms an active oligomer that recruits adaptor proteins to illicit a downstream signalling cascade that ultimately results in inflammatory gene expression, or stimulates autophagy through association with *ATG16L1*[8].

Manuscript Doi: 10.1093/ecco-jcc/jjac124
The functional impact of damaging *NOD2* variation largely elicits its effect through loss of proinflammatory NOD-signalling and impaired autophagy (through inability to interact with *ATG16L1*)[9,10]. Collectively there is impaired clearance of bacteria, resulting in upregulation of alternative inflammatory pathways including IL-1b, IL-18 and activation of the NLRP3 inflammasome[9,10], figure 1. Knock-out *Nod2* (-/-) murine models do not develop spontaneous colitis in sterile conditions, but develop inflammation upon the introductions of bacteria, which suggests a key triggering role for the microbiome in *NOD2*-related CD[11]. Further data implicate variation in genes and protein complexes across the *NOD2*-signaling pathway in the development of CD[9]. The route through which loss-of-function variation in multiple genes within a pro-inflammatory pathway translates to chronic inflammation is complex and is evidence would suggest this is through activation of alternative inflammatory pathways in response to impaired bacterial clearance[9]. Epistasis and other gene-gene modifier effects are less studied due to limitations in modelling approaches, but it appears possible that oligogenic forms of Crohn's disease may exist, with the epicentre of genomic risk being at *NOD2*[9].

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Table 1- Glossary of key genetics/genomics terms

Term	**Definition**
Loci/Locus	A specific location within the genome where a gene, genetic marker or variant is located
Single nucleotide variant/polymorphism (SNV/P)	A single change in the DNA sequence (single base change) at a specific location in the genome. Used as a marker in genome-wide association studies to assess the association between genetic variation and a phenotype.
Linkage disequilibrium	The non-random association of alleles at different loci in the genome, often inherited as a block (a haplotype).
Coding vs non-coding region	Coding regions of the genome including exonic regions of genes which can be transcribed and translated to proteins. Non-coding DNA makes up around 99% of the human genome and comprises intronic, intergenic, regulatory, promotor and untranslated regions.
Variants- missense, frameshift, splicing, synonymous + nonsense	Variants (mutations) are changes within the genome compared to a reference genome. Humans harbour ~20 million variants in their genome (including coding and non-coding variants). Missense variants are changes that result in a single amino acid change in the translated protein. Frameshift variants change the reading frame of DNA, usually resulting in loss of function through introduction of a downstream and premature stop codon. Nonsense variants result in an immediate premature STOP codon and termination of translation.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

NOD2 association and functional studies

Genome wide association studies (GWAS) and subsequent meta-analyses have consistently reported *NOD2* as conferring the most significant genetic risk of Crohn's disease[12]. The number of loci identified as risk factors for CD has incremented steadily over the last fifteen years and includes additional genes in the *NOD2*-signalling pathway, such as *ATG16L1, CARD9* and *RIPK2*[13–15]. The proteins encoded by these genes impact innate immunity through close interaction with *NOD2*. *ATG16L1* is activated by muramyl dipeptide-stimulated *NOD2* as an alternative to typical downstream nuclear factor kappa B (NF-κB) transcription[9,10]. *RIPK2* has a key role in signal transduction from activated *NOD2* to downstream NF-κB activation, whilst *CARD9* acts synergistically with activated NOD2 to drive proinflammatory signalling[10]. The main pathway through which *NOD2* acts, alongside the additional immune pathways and the impact of variants in specific areas of the gene can be seen in figure 2.

When considering specific GWAS associations within the *NOD2* gene, the minor allele of the rs2066844 SNP encoding a missense change (R702W) is common to ~3.9% of north-west European population and confers an odds ratio of 2.0 ($p = 2.27 \times 10^{-217}$)[4,16]. Functional analysis of this variant through luceriferase reporter assays has confirmed reduced levels of NF-κB activation, greatly reduced response to lipopolysaccharide and peptidoglycan stimulation[8,17]. However, interpretation of the clinical relevance of a number of very strong association signals from across the *NOD2* locus has been hampered because many of the reported SNPs are located in non-coding regions of the gene (such as rs2357623 and rs72796367) or have been proven to be functionally benign (rs5743271 and rs184788345)[18]. It is logical to conclude that many of these SNPs are in linkage disequilibrium with actual causative variants. The specific causal variants may be rare and fall within regulatory or promotor regions, or lie on the same haplotype as damaging, but unidentified, variants. Fine-mapping analysis by Elding *et al* confirmed significant additional genetic heterogeneity in *NOD2*, although these data also fail to capture the full estimates of heritability related to linkage studies

Manuscript Doi: 10.1093/ecco-jcc/jjac124
implicated chromosome 16q, the genomic locus of *NOD2*, and points to the importance of regulatory and non-coding regions[19]. Rivas *et al* included detailed deep resequencing of GWAS loci in CD and which demonstrated haplotypes with multiple risk variants which may be indicative of additive effect of specific mutations[20]. Interestingly, this study evidenced further genetic heterogeneity and identified new haplotypes, not containing the established R702W, G908R, and L1007fs variants, that were independently associated with risk of Crohn's disease, including five rare variants achieving genome-wide significance (R311W, S431L, R703C, N852S, and M863V). Importantly, R311W, R703C and M863V, were not tested for impact on *NOD2* function as determined by cell line transfection of specific mutation, whereas S341L and N852S were shown to exhibit functional impact on *NOD2* localisation and MDP-induced NF-kB activation. Previous functional data from Chamaillard *et al* identified no functional deficit on NF-kB activation for R311W, R703C or M863V when assessed as single variants transfected into a cell line[21]. However, the identification of these variants as being significantly associated with Crohn's disease points to non-coding or non-assessed variants on the same haplotype being the actual cause of impaired *NOD2* function. Additionally, whilst functional assays based on induction of the *NOD2* pathway with MDP and assessment of NF-κB response are established and informative, these have often been used to assess the impact of single variants, rather than the confounding impact of multiple variants inherited in *cis* or compound heterozygosity where multiple variants are inherited in *trans* within an individual[1,22]. The widespread assessment of variation in individual patients is not possible as the number of permutations of different variant profiles is extremely high. Interestingly, Tanabe *et al* systematically introduced point mutations to affect 519/1040 amino acid residues within *NOD2* and reported on 806 missense/nonsense variants[22]. Many of these variants impacted on downstream signalling but despite this huge repository of functional evidence, the modelling of unique mutational profile of individual patients is not possible with current *in vitro* or *in silico* techniques. Despite this, there is potential for high-throughput modelling which may be able to include phased *NOD2* variant data in contemporary protein models, such as AlphaFold, may result in improved *in*

Manuscript Doi: 10.1093/ecco-jcc/jjac124

silico determination of functional impact[23]. A summary of the main studies reporting functional evidence derived from NOD2 variants observed in humans can be seen in table 2. It is important to note that the risk of CD conferred by *NOD2* is not replicated in Japanese or Chinese studies but has been associated through GWAS in African Americans populations[24–26]. Despite this, different *NOD2* variants, not detected in European studies performed on predominantly Caucasian populations, may confer risk of disease in the Asian population, or that environmental factors may protect against disease. It is also possible that CD is driven by alterative molecular pathways in the Japanese and Chinese populations.

The three 'common' coding risk variants

Studies most frequently report on the three relatively common coding variants; R702W, G908R and L1007fs, (with allele frequencies in gnomAD (all populations) of 0.02605, 0.01128 and 0.015 respectively)[16]. These variants confer odds ratios (OR) for CD of 2-4 as heterozygotes and 20-40 as homozygotes[27,28]. All three of these variants have confirmed functional impact on protein function and downstream inflammatory signalling, through reduced levels of NF-κB activation and absent or greatly reduced response to lipopolysaccharide and peptidoglycan stimulation[8,29,30]. However, there are multiple much rarer variants documented across the gene that are too infrequent to test in association studies of common variation and are too rare to routinely genotype in genotype-phenotype studies[31]. Determining the overall impact of multiple variants within the same gene has been hampered by the inability of either GWAS genotyping or short-read next generation sequencing (NGS) data to resolve which variants are inherited together on a single parental chromosome. Whilst such *phasing* of variants to either the maternally or paternally inherited chromosomes can be estimated for common variants, this statistical approach fails to assign rare or novel variants due to inability to assign these variants to an estimated haplotype[32].

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Next-generation sequencing

Exome sequencing data has added further evidence to the of role of variation in *NOD2* as a cause of CD. Independent studies have consistently reported on apparent autosomal recessive 'monogenic' *NOD2*-related disease, accounting for 7.3-10% of CD patients[3,33]. Exome sequencing across these two independent studies including adult and paediatric-onset patients, demonstrated a consistent pattern whereby carriage of an established *NOD2* 'risk' variant was found with a much rarer variant in trans on the second parental chromosome. Additionally, across the three independent cohorts for which data is reported (Toronto Sick Kids Hospital, Southampton Children's Hospital and RGC-GHS DiscovEHR adult IBD cohort), there is a strong signal connecting *trans NOD2* variants and a diagnosis of Crohn's disease specifically, as opposed to ulcerative colitis, 83%, 95% and 70.3% for respective cohorts. In data from Southampton Children's Hospital, all *potential* compound heterozygote *NOD2* variants across 20 paediatric patients were confirmed as being inherited in *trans* following segregation analysis, underpinning a recessive model (figure 3)[3]. Novel variants that often lack robust experimental evidence of function is repeatedly observed in these patients[3,33]. It is possible that there is a role for numerous other rare or novel *NOD2* variants acting in combination with more common *NOD2* variation in both paediatric and adult-onset CD. This appears likely to have been overlooked by studies focused on the three R702W, G908R and L1007fs variants identified by GWAS.

Within the gnomAD v2.1.1 database (https://gnomad.broadinstitute.org), consisting of over 125,000 whole exome sequences and 15,000 whole genome sequences, 219 variants of uncertain significance are observed in *NOD2*, compared to only 9 pathogenic or likely pathogenic variants[16]. Interestingly the well characterised L1007fs variant, highly implicated in Crohn's disease pathogenesis with multiple studies reporting impact on immune function, is only reported as a variant of 'conflicting interpretations of pathogenicity' within ClinVar and is observed as homozygote in 113 individuals[8,29,30,34]. This lack of characterisation of both coding and non-coding *NOD2* variants is hampering better understanding of *NOD2* in CD.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Although, initial studies implicating *NOD2* pointed towards a recessive model of inheritance, most studies from the last 20 years have not routinely assessed whether *NOD2* variants are inherited in *cis* or in *trans* [35]. This includes population-based analyses assessing the impact of multiple *NOD2* risk-alleles phase[36]. It appears at least possible that in healthy individuals, variants are more likely to be inherited in *cis*, and in those who develop CD the variants are in *trans*. This could be achieved by long-range sequencing of the entire *NOD2* haplotype block, allowing phasing of coding and non-coding variants, providing allele-specific data to improve risk stratification using *NOD2*. These studies would be complemented by experimental work assessing the functional impact of *NOD2* haplotypes in *cis* and *trans*, potentially through stimulation of peripheral blood mononuclear cells[37].

Specific very-early onset forms of *NOD2*-related Crohn's disease are being increasingly reported in the literature, although penetrance for *NOD2* variants appears far from 100% [38,39]. Girardelli *et al* identified a novel homozygous missense mutation, R426H, in a male patient with very-early onset Crohn's disease. This variant triggered an apparent gain-of-function phenotype, with hyperinflammatory response to MDP-stimulation[40]. NOD2 gain-of-function, hyperinflammatory, variants typically result in Blau syndrome, however these data point to a broader, complex phenotype.

Some caution must be taken in the interpretation of exome sequencing data, at least in part due to identification of many variants of unknown significance and difficulties of *in silico* modelling of the functional impact of multiple variants[40,41]. Next generation sequencing has also enabled identification of numerous monogenic forms of inflammatory bowel disease caused by genes directly interacting with *NOD2,* including loss of function variation in *RIPK2, XIAP* and *CARD9*[42]. *NOD2* has recently been included in a collated panel of monogenic IBD genes which are investigated as part of whole genome sequencing in very-early onset IBD (and selected other cases) in the United Kingdom[43]. However, the interpretation of this and other monogenic IBD genes, such as *XIAP, MVK* and *COL7A1*, which have penetrance below 100%, is complex [39].

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Mutation	Variant- DNA change and amino acid change	Functional evidence	Reference
Missense	c.113G>T p.R38M	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.315G> p.A105T	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	D113N	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.413G>A p.R138Q	Inability to interact with *RIPK2* and at least a 2-fold reduction in both basal activity and PGN-induced response	Parkhouse et al[17], Chamaillard et al[21]
Missense	c.469T>C p.W157R	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.484G>A p.V162I	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.566C>T p.T189M	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.703C>T p.R235C	Significantly reduced PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.743T>G p.L248R	Impaired membrane association and were unable to signal in response to MDP	Parkhouse et al[17]
Missense	c.866A>G p.N289S	>2-fold reduction in both basal activity and PGN-induced response	Chamaillard et al[21]
Missense	c.871G>A p.D291S	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.1042C>G p.L348V	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Nonsense	c.1065G>A p.W355X	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.1070A>C p.D357A	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.1087A>T p.I363F	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.1321G>A p.E441K	Borderline reduction in both basal activity and PGN-induced response	Chamaillard et al[21]
Missense	c.1387C>G p.P463A	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.1648C>G p.L550V	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense deletion	c.1671 delCCTGGG p.558delLG	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.1834G>A p.A612T	>2-fold reduction in both basal activity and PGN-induced response	Chamaillard et al[21]
Missense	c.1835C>T p.A612V	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.2104C>T p.R702W	Reduced levels of NF-κB activation, greatly reduced response to lipopolysaccharide and peptidoglycan stimulation	Bonen et al[8], Parkhouse et al[17]
Missense	c.2138G>A p.R713H	Major impairment of the PGN-response (<20% that of wildtype)	Chamaillard et al[21]
Missense	c.2368C>T p.P727C	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.2264C>T p.A755V	Reduced levels of NF-κB activity by 25-54%	Parkhouse et al[17]
Missense	c.2332G>A P.E778K	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.2368C>T p.R790W	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Missense	c.2475C>G p.N825K	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.2527G>A p.E843K	Major impairment of the PGN-response (<20% that of wildtype)	Chamaillard et al[21]
Missense	c.2546C>T p.A849V	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.2587A>G p.M863V	Significantly reduced basal NFκB production and PGN-induced NFκB activation compared to WT	Chamaillard et al[21]
Missense	c.2555A>G p.N852S	N825K has impaired membrane association and were unable to signal in response to Muramyl Dipeptide	Parkhouse et al[17]
Missense	c.2719T>C p.W907R	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]
Missense	c.2722G>C p.G908R	Reduced levels of NF-κB activation, greatly reduced response to lipopolysaccharide and peptidoglycan stimulation	Bonen et al[8]
Missense	p.A1007P	Reduced levels of NF-κB activity by >15%	Chamaillard et al[21]
Missense frameshift	c.G3016ins+C p.A1007fs	Reduced levels of NF-κB activation, no response to lipopolysaccharide and peptidoglycan stimulation	Bonen et al[8]
Nonsense	c.3055C>T p.R1019X	Reduced levels of NF-κB activity by >15% for MDP-induced NFκB production and activation of IL-8 promoter	Parkhouse et al[17]

Table 2- A selection of *NOD2* variants with published functional evidence demonstrating impact on protein function or downstream pathway activity from Bonen et al[8], Parkhouse et al[17] and Chamaillard et al[21]. Data aligned to reference genome GRCh38. Studies did not always perform comparable functional assays on all variants and there is possibility that multiple mechanisms of action are present for most variants, even if this has not functionally been proven in the reporting study.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

The stricturing NOD2 phenotype

Shortly after *NOD2*'s identification, deleterious variation within the gene was reported to have a strong phenotypic association with fibrostenotic disease behaviour, which is replicated in paediatric and adult-onset disease[44]. The phenotype appears to be consistent across adult and paediatric populations, although more data is available from paediatric studies specifically related to stricturing disease when accounting for rare *NOD2* variants. Studies also identified a strong association between deleterious *NOD2* variation and terminal ileal disease location, which was believed to account for the increased stricturing disease risk[45]. However more recent data from 'monogenic' *NOD2-related* patients confirm a strong stricturing phenotype, with an odds ratio of >10 compared to CD without *NOD2* variation, and this statistical association remained when accounting for terminal ileal disease location[3]. Furthermore, contemporary functional evidence points to a specific alteration of stromal regulation of fibroblasts through *STAT3* and gp130 ligands in patients harbouring *NOD2* variation, providing a mechanistic explanation for *NOD2* dysfunction to trigger fibrosis and leading to a stricturing phenotype[2]. Additionally, this strong genotype-potential phenotype connection reinforces the potential for *NOD2* to exist as a genomic tool for disease prediction, although the differing risk profiles of specific variants, or combinations of variants, has not yet been elucidated. There are additional data pointing towards significant variation in *NOD2* increasing the chance of complicated disease and the need for surgery, but the potential of this as a predictive tool has not yet been brought to the clinic[3,46]. Furthermore, a precise diagnosis of *NOD2-associated* disease may allow new, or known therapies, to be targeted at an early stage of disease, preventing significant complications and the need for subsequent surgery through the prevention of irreversible fibrosis, although the potential of this remains hypothetical with no studies providing evidence to date [47]. It appears increasingly feasible to utilise *NOD2* as a genomic biomarker in stratified randomised control trials of therapy. This may provide the ability to modify risk in patients at high risk of disease progression. Furthermore, there is the potential for post-hoc analyses of RCTs

Manuscript Doi: 10.1093/ecco-jcc/jjac124
that have already been conducted to identify patients most likely to respond to therapy based on underlying molecular cause of disease.

Conclusions and future research

Despite the strong evidence of *NOD2*'s role in CD there are many unanswered questions. Current independent estimates identify *NOD2* variation as potentially underpinning ~8% of CD, which would reflect 100,000s of individuals based on global CD prevalence[3,33]. *NOD2* appears to have the potential to translate into clinical practice, as a precise diagnosis, a prediction tool and as a potential therapeutic target. The potential strategies to resolve and implement this consist of 1) sequencing that fully capture important upstream non-coding regulatory variants, promoter variants and deep intronic variation; 2) phased data to elucidate recessive inheritance patterns in adults and children; 3) systematic functional interpretation of variants; and 4) thorough longitudinal phenotyping in the context of specific genotypes. Expanding this examination to *NOD2*-signalling pathway genes may result in improved resolution.

With contemporary sequencing techniques and access to vast numbers of patient samples, now may be the time to revisit some of these unresolved questions and implement clinical translation of *NOD2* sequencing.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Tables and Figures

Table 1- Glossary of key genetics/genomics terms

Table 2- *NOD2* variants identified within our inflammatory bowel disease cohort of 1086 patients which have published functional evidence demonstrating impact on protein function or downstream pathway activity. Data refer to reference genome GRCh38.

Figure 1- Comparison of normal intestinal immune function with *NOD2* variant induced immune dysregulation triggering impaired bacterial clearance and chronic increased alternative inflammatory pathways. *MDP- muramyl dipeptide*

Figure 2- The major NOD2 signalling pathway resulting in activation of NF-KB and additional proinflammatory signalling through the NLRP3 inflammasome and induction of autophagy through ATG16L1. Variants in NOD2 have been demonstrated to impact on 1) Deletions/stop gains prevent transcription or localisation, 2) Leucine-rich repeat variants such as G908R and L1007fs are located in the LRR and result in defective binding of MDP, likely preventing NOD2 oligomerisation and downstream signalling. Similarly, the R702W mutant is in the NOD domain and may also directly result in in defective oligomerisation, 3) Impair ATG16L1 induced autophagy, 4) CARD domain variants may reduce RIPK2 association with activated NOD2. Precise details of the impact of specific variants see table 2. The normal downstream sequalae of NOD2 activation can be seen in this figure which are variably impaired when patients harbour NOD2 variants functionally implicated in Crohn's disease, resulting in impaired immune response to bacteria, barrier breakdown, dysbiosis and chronic inflammation.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

Figure 3- Potential mechanism by which variants inherited on different chromosomes (in *trans*) will lead to two dysfunctional copies of the gene, whereas the same variants inherited on the same chromosome (in *cis*) would result in a preserved functional version of *NOD2* as there is a still a fully functional copy of the *NOD2* gene present. Typically, this would be elucidated through segregation analysis, where sequencing of the mother and father allows ascertainment of whether the variants are inherited on the same chromosome or separately on a maternal or paternal chromosome.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

References

1. Eckmann L., Karin M. NOD2 and Crohn's Disease: Loss or Gain of Function? *Immunity* 2005;**22**(6):661–7. Doi: 10.1016/J.IMMUNI.2005.06.004.

2. Nayar S., Morrison JK., Giri M., Gettler K., Chuang L., Walker LA., et al. A myeloid–stromal niche and gp130 rescue in NOD2-driven Crohn's disease. *Nature* 2021:1–9. Doi: 10.1038/s41586-021-03484-5.

3. Ashton JJ., Mossotto E., Stafford IS., Haggarty R., Coelho TAF., Batra A., et al. Genetic Sequencing of Pediatric Patients Identifies Mutations in Monogenic Inflammatory Bowel Disease Genes that Translate to Distinct Clinical Phenotypes. *Clinical and Translational Gastroenterology* 2020;**11**(2):e00129. Doi: 10.14309/ctg.0000000000000129.

4. Verstockt B., Smith KG., Lee JC. Genome-wide association studies in Crohn's disease: Past, present and future. *Clin Transl Immunology* 2018;**7**(1):e1001. Doi: 10.1002/cti2.1001.

5. Gordon H., Trier Moller F., Andersen V., Harbord M. Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies. *Inflamm Bowel Dis* 2015;**21**(6):1428–34. Doi: 10.1097/MIB.0000000000000393.

6. Hugot J-P., Laurent-Puig P., Gower-Rousseau C., Olson JM., Lee JC., Beaugerie L., et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;**379**(6568):821–3. Doi: 10.1038/379821a0.

7. Ogura Y., Bonen DK., Inohara N., Nicolae DL., Chen FF., Ramos R., et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**(6837):603–6. Doi: 10.1038/35079114.

8. Bonen DK., Ogura Y., Nicolae DL., Inohara N., Saab L., Tanabe T., et al. Crohn's disease-associated NOD2 variants share a signaling defect in response to lipopolysaccharide and peptidoglycan. *Gastroenterology* 2003;**124**(1):140–6. Doi: 10.1053/gast.2003.50019.

9. Ashton JJ., Boukas K., Stafford IS., Cheng G., Haggarty R., Coelho TAF., et al. Deleterious Genetic Variation Across the NOD Signaling Pathway Is Associated With Reduced NFKB Signaling Transcription and Upregulation of Alternative Inflammatory Transcripts in Pediatric Inflammatory Bowel Disease. *Inflammatory Bowel Diseases* 2022:1–11. Doi: 10.1093/IBD/IZAB318.

10. Caruso R., Warner N., Inohara N., Núñez G. NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity* 2014;**41**(6):898–908. Doi: 10.1016/j.immuni.2014.12.010.

11. Kobayashi KS., Chamaillard M., Ogura Y., Henegariu O., Inohara N., Nuñez G., et al. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 2005;**307**(5710):731–4. Doi: 10.1126/SCIENCE.1104911.

12. Jostins L., Ripke S., Weersma RK., Duerr RH., McGovern DP., Hui KY., et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**(7422):119–24. Doi: 10.1038/nature11582.

13. Franke A., McGovern DPB., Barrett JC., Wang K., Radford-Smith GL., Ahmad T., et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genetics 2010 42:12* 2010;**42**(12):1118–25. Doi: 10.1038/ng.717.

14. Barrett JC., Hansoul S., Nicolae DL., Cho JH., Duerr RH., Rioux JD., et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;**40**(8):955–62. Doi: 10.1038/NG.175.

15. Hampe J., Franke A., Rosenstiel P., Till A., Teuber M., Huse K., et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007;**39**(2):207–11. Doi: 10.1038/NG1954.

16. Karczewski KJ., Francioli LC., Tiao G., Cummings BB., Alföldi J., Wang Q., et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature 2020 581:7809* 2020;**581**(7809):434–43. Doi: 10.1038/s41586-020-2308-7.

17. Parkhouse R., Monie TP. Dysfunctional Crohn's Disease-Associated NOD2 Polymorphisms Cannot be Reliably Predicted on the Basis of RIPK2 Binding or Membrane Association. *Frontiers in Immunology* 2015;**6**. Doi: 10.3389/fimmu.2015.00521.

18. Caliskan M., Brown CD., Maranville JC. A catalog of GWAS fine-mapping efforts in autoimmune disease. *American Journal of Human Genetics* 2021;**108**(4):549. Doi: 10.1016/J.AJHG.2021.03.009.

19. Elding H., Lau W., Swallow DM., Maniatis N. Dissecting the genetics of complex inheritance: linkage disequilibrium mapping provides insight into Crohn disease. *Am J Hum Genet* 2011;**89**(6):798–805. Doi: 10.1016/J.AJHG.2011.11.006.

20. Rivas MA., Beaudoin M., Gardet A., Stevens C., Sharma Y., Zhang CK., et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genetics* 2011;**43**(11):1066–73. Doi: 10.1038/ng.952.

21. Chamaillard M., Philpott D., Girardin SE., Zouali H., Lesage S., Chareyre F., et al. Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. *Proc Natl Acad Sci U S A* 2003;**100**(6):3455–60. Doi: 10.1073/pnas.0530276100.

22. Tanabe T., Chamaillard M., Ogura Y., Zhu L., Qiu S., Masumoto J., et al. Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *EMBO Journal* 2004;**23**(7):1587–97. Doi: 10.1038/sj.emboj.7600175.

23. Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* 2021;**596**(7873):583–9. Doi: 10.1038/s41586-021-03819-2.

24. Inoue N., Tamura K., Kinouchi Y., Fukuda Y., Takahashi S., Ogura Y., et al. Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 2002;**123**(1):86–91. Doi: 10.1053/GAST.2002.34155.

25. Brant SR., Okou DT., Simpson CL., Cutler DJ., Haritunians T., Bradfield JP., et al. Genome-Wide Association Study Identifies African-Specific Susceptibility Loci in African Americans With Inflammatory Bowel Disease. *Gastroenterology* 2017;**152**(1):206-217.e2. Doi: 10.1053/j.gastro.2016.09.032.

26. Leong RWL., Armuzzi A., Ahmad T., Wong ML., Tse P., Jewell DP., et al. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Alimentary Pharmacology & Therapeutics* 2003;**17**(12):1465–70. Doi: 10.1046/J.1365-2036.2003.01607.X.

27. Helio T., Halme L., Lappalainen M., Fodstad H., Paavola-Sakki P., Turunen U., et al. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* 2003;**52**(4):558–62. Doi: 10.1136/gut.52.4.558.

28. Cuthbert AP., Fisher SA., Mirza MM., King K., Hampe J., Croucher PJP., et al. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 2002;**122**(4):867–74.

29. Salucci V., Rimoldi M., Penati C., Sampietro GM., van Duist MM., Matteoli G., et al. Monocyte-derived dendritic cells from Crohn patients show differential NOD2/CARD15-dependent immune responses to bacteria. *Inflamm Bowel Dis* 2008;**14**(6):812–8. Doi: 10.1002/IBD.20390.

30. Li J., Moran T., Swanson E., Julian C., Harris J., Bonen DK., et al. Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum Mol Genet* 2004;**13**(16):1715–25. Doi: 10.1093/HMG/DDH182.

31. Andreoletti G., Shakhnovich V., Christenson K., Coelho T., Haggarty R., Afzal NA., et al. Exome Analysis of Rare and Common Variants within the NOD Signaling Pathway. *Sci Rep* 2017;**7**:46454. Doi: 10.1038/srep46454.

32. Tewhey R., Bansal V., Torkamani A., Topol EJ., Schork NJ. The importance of phase information for human genomics. *Nature Reviews Genetics 2011 12:3* 2011;**12**(3):215–23. Doi: 10.1038/nrg2950.

Manuscript Doi: 10.1093/ecco-jcc/jjac124

33. Horowitz JE., Warner N., Staples J., Crowley E., Gosalia N., Murchie R., et al. Mutation spectrum of NOD2 reveals recessive inheritance as a main driver of Early Onset Crohn's Disease. *Sci Rep* 2021;**11**(1). Doi: 10.1038/S41598-021-84938-8.

34. Landrum MJ., Lee JM., Benson M., Brown G., Chao C., Chitipiralla S., et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* 2016;**44**(D1):D862–8. Doi: 10.1093/nar/gkv1222.

35. Economou M., A. TT., T. LK., V. TE., P. IJ. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 2004;**99**(12):2393–404. Doi: 10.1111/j.1572-0241.2004.40304.x.

36. Hugot JP., Zaccaria I., Cavanaugh J., Yang H., Vermeire S., Lappalainen M., et al. Prevalence of CARD15/NOD2 mutations in Caucasian healthy people. *Am J Gastroenterol* 2007;**102**(6):1259–67. Doi: 10.1111/J.1572-0241.2007.01149.X.

37. Coelho T., Mossotto E., Gao Y., Haggarty R., Ashton JJ., Batra A., et al. Immunological Profiling of Paediatric Inflammatory Bowel Disease Using Unsupervised Machine Learning. *Journal of Pediatric Gastroenterology & Nutrition* 2020;**70**(6):833–40. Doi: 10.1097/MPG.0000000000002719.

38. Uhlig HH., Muise AM. Clinical Genomics in Inflammatory Bowel Disease. *Trends in Genetics* 2017;**33**(9):629–41. Doi: 10.1016/j.tig.2017.06.008.

39. Uhlig HH., Schwerd T. From Genes to Mechanisms. *Inflammatory Bowel Diseases* 2016;**22**(1):202–12. Doi: 10.1097/MIB.0000000000000614.

40. Girardelli M., Loganes C., Pin A., Stacul E., Decleva E., Vozzi D., et al. Novel NOD2 Mutation in Early-Onset Inflammatory Bowel Phenotype. *Inflammatory Bowel Diseases* 2018;**24**(6):1204–12. Doi: 10.1093/ibd/izy061.

41. Kaczmarek-Ryś M., Hryhorowicz ST., Lis E., Banasiewicz T., Paszkowski J., Borejsza-Wysocki M., et al. Crohn's Disease Susceptibility and Onset Are Strongly Related to Three NOD2 Gene Haplotypes. *J Clin Med* 2021;**10**(17). Doi: 10.3390/JCM10173777.

42. Bolton C., Smillie CS., Pandey S., Elmentaite R., Wei G., Argmann C., et al. An Integrated Taxonomy for Monogenic Inflammatory Bowel Disease. *Gastroenterology* 2021. Doi: 10.1053/J.GASTRO.2021.11.014.

43. Infantile enterocolitis & monogenic inflammatory bowel disease (Version 1.38). Available at: https://panelapp.genomicsengland.co.uk/panels/176/. Accessed April 19, 2022.

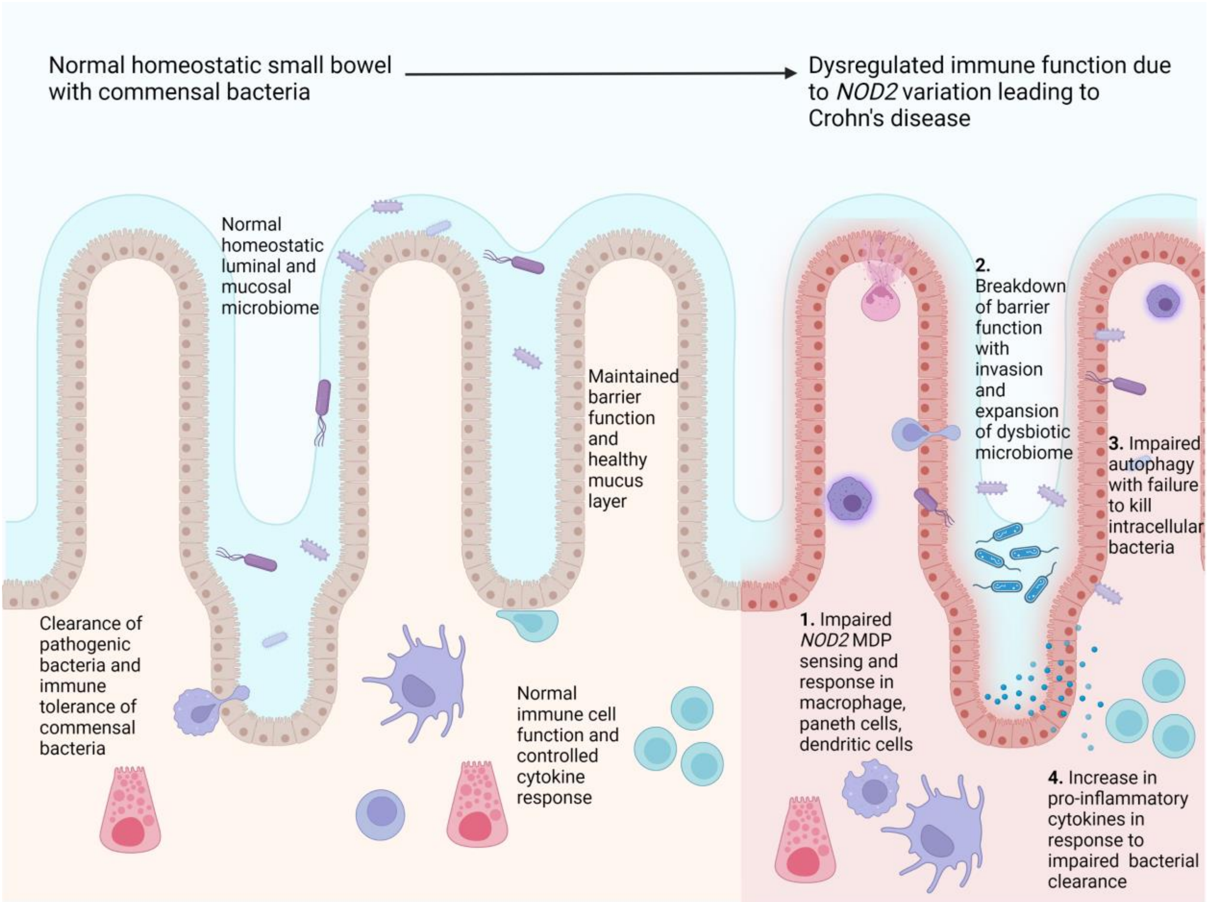
44. Abreu MT., Taylor KD., Lin Y-C., Hang T., Gaiennie J., Landers CJ., et al. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 2002;**123**(3):679–88.

45. Cleynen I., Boucher G., Jostins L., Schumm LP., Zeissig S., Ahmad T., et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet* 2016;**387**(10014):156–67. Doi: 10.1016/S0140-6736(15)00465-1.

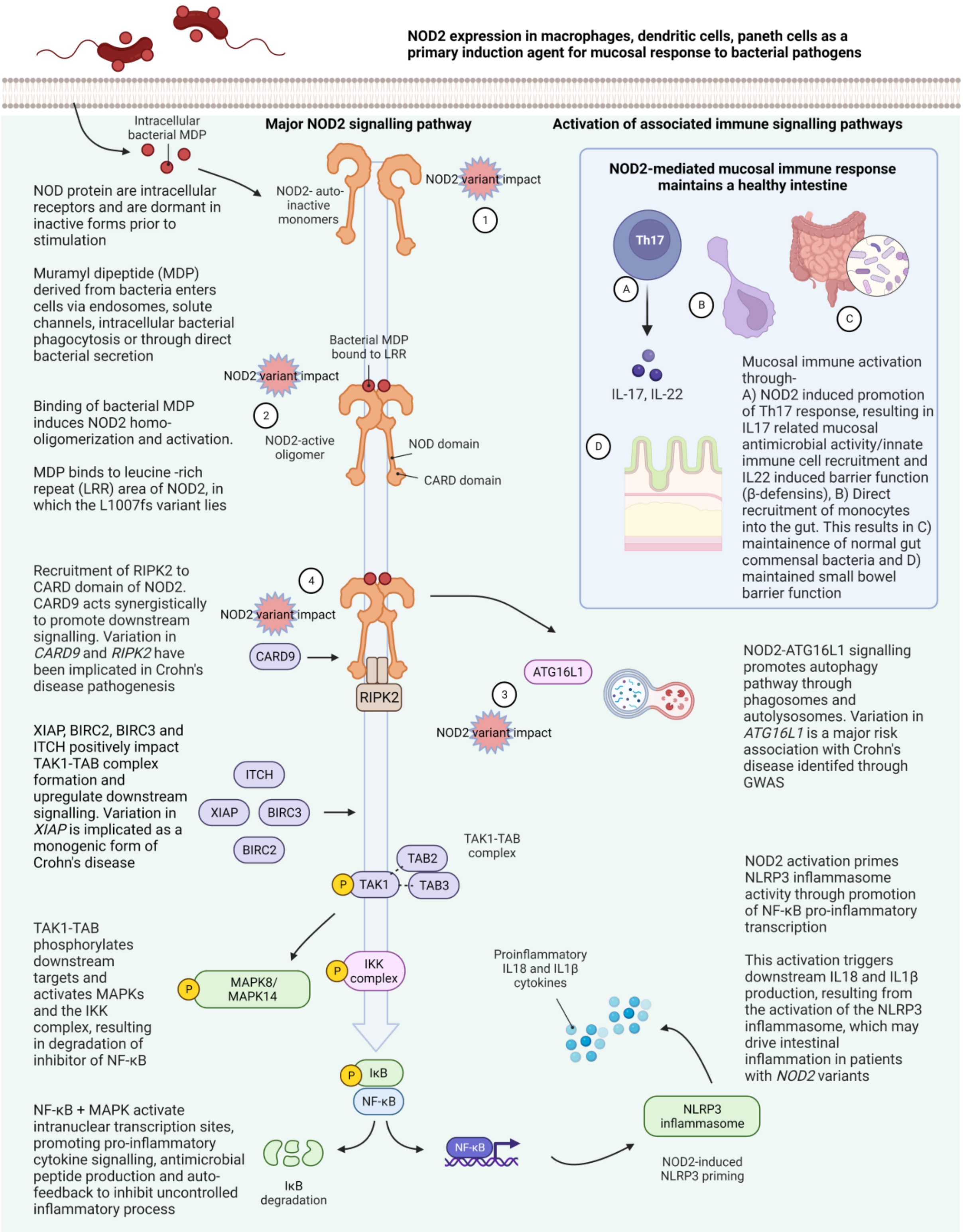
46. Adler J., Rangwalla SC., Dwamena BA., Higgins PD. The Prognostic Power of the NOD2 Genotype for Complicated Crohn's Disease: A Meta-Analysis. *American Journal of Gastroenterology* 2011;**106**(4):699–712. Doi: 10.1038/ajg.2011.19.

47. D'Haens G., Rieder F., Feagan BG., Higgins PDR., Panes J., Maaser C., et al. Challenges in the Pathophysiology, Diagnosis and Management of Intestinal Fibrosis in Inflammatory Bowel Disease. *Gastroenterology* 2019. Doi: 10.1053/j.gastro.2019.05.072.

Manuscript Doi: 10.1093/ecco-jcc/jjac124
Figure 1



Manuscript Doi: 10.1093/ecco-jcc/jjac124
Figure 2



Manuscript Doi: 10.1093/ecco-jcc/jjac124
Figure 3