HHS Public Access

Author manuscript

Nat Rev Gastroenterol Hepatol. Author manuscript; available in PMC 2022 June 22.

Published in final edited form as:

Nat Rev Gastroenterol Hepatol. 2022 June ; 19(6): 399–409. doi:10.1038/s41575-022-00593-y.

Multiomics to elucidate inflammatory bowel disease risk factors and pathways

Manasi Agrawal[1,2,✉],

Kristine H. Allin[2,3],

Francesca Petralia[4],

Jean-Frederic Colombel[1],

Tine Jess[2,3]

[1]The Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[2]Center for Molecular Prediction of Inflammatory Bowel Disease, Department of Clinical Medicine, Aalborg University, Copenhagen, Denmark.

[3]Department of Gastroenterology and Hepatology, Aalborg University Hospital, Aalborg, Denmark.

[4]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Abstract

Inflammatory bowel disease (IBD) is an immune-mediated disease of the intestinal tract, with complex pathophysiology involving genetic, environmental, microbiome, immunological and potentially other factors. Epidemiological data have provided important insights into risk factors associated with IBD, but are limited by confounding, biases and data quality, especially when pertaining to risk factors in early life. Multiomics platforms provide granular high-throughput data on numerous variables simultaneously and can be leveraged to characterize molecular pathways and risk factors for chronic diseases, such as IBD. Herein, we describe omics platforms that can advance our understanding of IBD risk factors and pathways, and available omics data on IBD and other relevant diseases. We highlight knowledge gaps and emphasize the importance of birth, at-risk and pre-diagnostic cohorts, and neonatal blood spots in omics analyses in IBD. Finally, we discuss network analysis, a powerful bioinformatics tool to assemble high-throughput data and derive clinical relevance.

Inflammatory bowel disease (IBD), including Crohn's disease and ulcerative colitis, is a progressive immune-mediated disease of the intestinal tract, for which there is currently no cure[1,2]. IBD pathophysiology is multifactorial, involving complex interactions of genetic, environmental, microbiome, immunological and potentially other risk factors[3].

✉ manasi.agrawal@mountsinai.org .

Author contributions

The authors contributed equally to all aspects of the article.

Although high-quality epidemiological data have provided important insights, they can be conflicting, and IBD risk factors and pathways remain inadequately understood[4]. Traditional epidemiological studies that typically utilize survey and administrative data carry limitations such as selection, confounding, misclassification and other biases, and oftentimes lack of detailed and quantitative data on relevant variables such as smoking, breastfeeding, infections and antibiotics, and those on the timing and duration of such exposures. Furthermore, the timing of exposure to risk factors is relevant; those that occur in the early life period when immune maturation is taking place might be important in IBD risk[5].

Some of these limitations can be bypassed through the use of omics platforms, which represent the machinery, infrastructure and software for high-throughput processing and analysis of omics data in biological samples to characterize the 'molecular epidemiology paradigm'[6]. Leveraging unique data sources, omics platforms can provide multidimensional insights into pathways from exposures to disease and are well adapted to unravelling the complex risk factors and interactions that mediate IBD (FIGS 1,2). Omics analyses represent a growing opportunity towards IBD prevention and prediction[7].

In this Perspective, we describe diverse omics platforms and the opportunities they represent towards understanding the molecular pathways of IBD in humans. When available, we discuss the literature on omics in IBD, focusing on birth, at-risk and pre-diagnostic human cohorts, highlighting knowledge gaps and sharing our viewpoints. When applicable, we summarize omics data from other chronic diseases as a framework for research in IBD.

Genomics

Genetic polymorphisms are the most consistent risk factors for IBD identified through techniques such as genome-wide association studies (GWAS) and whole-genome sequencing (reviewed in detail elsewhere[8]). These data indicate that although genetic polymorphisms contribute to IBD risk, they cannot explain it entirely. More than 250 loci have been implicated in IBD risk, but markedly fewer are mechanistically linked with IBD, such as variants in the *NOD2* (also known as *CARD15*), *IL23R* and *ATG16L1* genes that are involved in the innate immune response against bacterial antigens, host–microorganism interactions and autophagy[9,10]. There is marked heterogeneity in allele frequencies, effect sizes or a combination of both between individuals of European descent and those of East Asian ancestry[11,12]. Similarly, there are differences in genetic IBD risk scores between European and African American populations[13].

Emerging data suggest the relevance of genetic–non-genetic interactions in IBD risk. For example, clustering of individuals with IBD in Ashkenazi Jewish multiplex families by birth order and proximity in age suggested the role of a shared environment towards IBD risk[14]. Among first-degree relatives (FDRs) of individuals with IBD, antibiotic use before IBD diagnosis was associated with an increase in IBD risk, indicating genetic–non-genetic interaction[15]. Among FDRs of individuals with Crohn's disease in the Genetic, Environment and Microbiome (GEM) project, those carrying specific mutations in the *NOD2* gene had a higher relative abundance of the bacterial family Erysipelotrichaceae than FDRs without

the mutations, which might indicate mediation (causal effect) versus interaction (no causal effect) between host genetic and microbiome factors[16].

The exposome

The exposome is a broad term encompassing all exposures experienced throughout life, measured at both the population and individual levels, that can be consequences of exogenous or endogenous processes[17]. Traditionally, exposures have been assessed using survey and administrative data, which carry inherent limitations. Unbiased exposure measurement can be conducted in diverse ways. Exposure measurement can be direct (externally in the environment and internally in biological samples, the latter reflecting the 'internal dose'), and it is dependent on the pharmacokinetics of each compound measured[18]. Exposure measurement can be indirect, involving measurement of biological responses to the exposure, thereby reflecting the pharmacodynamic characteristics of the compound[18]. For example, cotinine, a metabolite of nicotine, is a biomarker for tobacco exposure[19]. Depending on the research question, different high-throughput platforms can be used to measure with high accuracy an expansive array of exposures, each one providing distinctive insights[18].

External data.

External exposures can be measured at the population level using remote sensing technologies such as distant measurement instruments, typically satellites. The geographic information system (GIS) is the infrastructure for recording and analysing spatial, geographic and temporal data, taking into account multiple overlapping layers of high-throughput data such as meteorology, air pollution, traffic, green space and biodiversity, variables that are otherwise challenging to quantify in granular detail[18,20]. GIS data are particularly useful for studying the effect of the environment on human health, an area of research that is increasingly being recognized as relevant[21,22].

Prenatal and early life exposure to air pollution, including polycyclic aromatic hydrocarbons (PAHs) and fine particulate matter ($PM_{2.5}$), have been associated with childhood obesity and cardiometabolic health using roadway traffic and residential near-roadway pollution exposure data[23–25]. The protective effects of green space, a measure of natural vegetation, on atopic sensitization and biomarkers of adverse health in children have been investigated[26,27], but such studies were limited in the context of IBD. In these studies, exposure to reactive oxygen species during childhood was associated with an increased risk of IBD ($P < 0.05$), but associations with other air pollutants were not statistically significant[28]. It has also been found that green space during early childhood is protective against later-onset IBD in a dose-dependent manner[29].

Sampling of air quality or assessing air pollution from traffic can be done using local sensors. For example, the New York City Community Air Survey measures data on pollutants at 100 locations throughout the city using stationary site monitors, which can be linked with zip code data. These data have been used to show that air pollution is associated with asthma-related urgent care visits and with increased childhood body weight[30,31].

Individual-level data can be measured using personal sensors, such as in the Columbia Center for Children's Environmental Health cohort study in the USA. Participants wore ambient air monitors with air sampling pumps over 2 days with a passive collection of $PM_{2.5}$ on a microfibre filter, and vapours and aerosols on a polyurethane foam cartridge. These measurements were assayed by gas chromatography–mass spectroscopy[19]. Using these data, prenatal PAHs and $PM_{2.5}$ exposures have been associated with decreased birthweight and birth length, and with childhood obesity in the offspring[23,32]. Silicone wristbands, which can passively absorb compounds the wearer is exposed to, and smartphone-based sensors that can record vital data, are other practical ways to gather omics data[18,33]. Although personal sensor platforms are being explored to predict IBD flares[34], they have not, to our knowledge, been used to understand the effect of exposures on IBD risk.

Internal, individual-level data.

Omics platforms can be used to measure diverse compounds with different physical and chemical properties, and at different stages of their metabolism. The measurement of external exposures, such as chemicals, depends largely on half-life and other pharmacokinetic properties. Metabolites of external exposure with long half-lives can serve as surrogates for these exposures; high-throughput analysis of low molecular mass metabolites (<2,000 Da) is referred to as metabolomics. Distinguishing metabolites of external exposures from those further downstream is key. Whereas the former represents exogenous exposures, the latter can indicate altered metabolic pathways, either before or after disease onset, depending on the timing of sample collection. For example, in a case–control study, plasma cotinine levels, reflecting tobacco exposure, were markedly higher in individuals who later developed IBD ($n = 96$; 70 with ulcerative colitis and 26 with Crohn's disease) than in age-matched and sex-matched controls ($n = 191$) with an odds ratio of 1.34 (95% CI 1.01–1.63)[35]. Metabolites of trichloroethylene, indicating exposure and biological response to the chemical, can be measured in plasma and be used to gain insights into the molecular pathways through which toxic effects due to trichloroethylene exposure occur[36].

Omics analyses can be untargeted in which all measurable molecules, such as environmental, dietary and endogenous products of metabolism in the sample, are analysed in a semiquantitative manner, or it can be targeted, implying quantitative analysis of specific molecules of interest[18,37]. For example, exposure to environmental tobacco smoke or organophosphate pesticide can be estimated through their metabolites plasma cotinine or chlorpyrifos, respectively. In a study that measured these metabolites, Perera et al. found a negative effect of environmental pollutants on birthweight, and birth length and head circumference[19]. Similarly, prenatal and childhood exposure to perfluoroalkyl and polyfluoroalkyl substances has been associated with adverse cardiometabolic risk during adolescence[38]. Heavy metals in biological samples have been associated with diseases later in life; for example, prenatal exposure to mercury negatively affects offspring height[39]. A nested case–control study examined the association between plasma perfluoroalkyl substances and the risk of IBD (73 participants with Crohn's disease and 143 controls, and 80 participants with ulcerative colitis and 159 controls) in a cohort of nurses, and showed inverse associations between plasma levels of three perfluoroalkyl substances and Crohn's disease ($P < 0.012$ for each association)[40]. Previous studies have shown conflicting data:

one study in 3,713 individuals in an occupational cohort, of whom 28 developed ulcerative colitis, found an association between perfluorooctanoic acid and ulcerative colitis, while another (n = 189) found no clear effect of perfluoroalkyl substances on IBD risk or intestinal inflammation[41,42].

A novel and innovative method to measure early life exposure to various compounds is the analysis of deciduous teeth. Growth of the teeth starts between 14 and 19 weeks of prenatal life and occurs in an incremental manner, with the neonatal line representing the transition from prenatal to postnatal life[43]. Various organic and inorganic compounds can be measured in teeth using mass spectrometry, and the timing of exposure is determined based on depth. A pilot study analysed heavy metals in deciduous teeth after natural shedding from 7 children with Crohn's disease, 5 children with ulcerative colitis and 16 healthy children[44]. Time-dependent differences in the levels of lead, copper, zinc and chromium were found in the teeth from children with IBD compared with the teeth from healthy children.

Metabolomics analyses have been conducted in samples from individuals with IBD. Distinct signatures of Crohn's disease were found in a pilot untargeted metabolomics analysis, in which the altered pathways were found to pertain to β-oxidation of fatty acids, amino acid metabolism and biotransformation of xenobiotics[45]. In Crohn's disease, alterations in amino acid and folate metabolism were found, and additionally sphingolipid metabolic pathways in ulcerative colitis correlated with serum metabolomics signatures such as carnosine ribose and choline, and faecal calprotectin, a biomarker of intestinal inflammation[46]. In a pilot analysis, phospholipid metabolites were found to be downregulated in Crohn's disease and ulcerative colitis compared with metabolite levels in individuals with no symptoms of IBD and normal levels of faecal calprotectin[47]. However, in studies conducted after IBD onset, it can be challenging to differentiate between causation and reverse causation.

Protein adductomics refers to the measurement of adducts that form between reactive small molecules and blood proteins such as serum albumin and haemoglobin. Protein adducts have longer half-lives than small molecules and can be measured over longer periods of time[48]. Exposure to chemicals such as PAHs, benzene and styrene has been studied using such analyses[49,50]. We suggest that these techniques could be applied to the exploration of the effects of environmental pollutants on IBD risk.

Lipidomics analysis, a subtype of metabolomics, refers to the large-scale untargeted or targeted measurement of lipid molecules in biological samples[51]. Altered lipid metabolism and specific signatures have been demonstrated in IBD, but after disease onset. Altered lipidomics signatures were reported in mucosal biopsy samples from 21 treatment-naive patients with newly diagnosed ulcerative colitis, 12 patients with ulcerative colitis in deep remission and 14 healthy individuals[52]. The investigators found disruption of mucosal lipid composition in patients with active ulcerative colitis compared with lipid composition in healthy individuals (specifically, variations in phosphatidylcholine, ceramide and sphingomyelin composition)[52]. Other studies determined lipidomics signatures in plasma from individuals with and without IBD, and found distinct differences in various lipid classes such as lysophosphatidylserine, phosphatidylserine and phosphatidylcholine[53,54].

Transcriptomics

The downstream effects of gene expression modulation by the exposome can be determined by measuring RNA levels, including mRNA, microRNA, small interfering RNA and non-coding RNAs, using next-generation sequencing[18]. Such transcriptomic analyses can be useful to study the pathway from exposure to outcome. For example, a study on the effect of $PM_{2.5}$ exposure on birthweight, using GIS data and placental transcriptome-wide gene co-expression RNA sequencing data, showed that expression of genes involved in amino acid transport and cellular respiration correlates with offspring birthweight and $PM_{2.5}$ exposure during a susceptibility window from 12 weeks before to 13 weeks into pregnancy[55].

Transcriptomic signatures have been studied in the context of IBD, albeit after disease onset. Through transcriptomic analysis of $CD8^+$ T cells in blood samples from 112 children with IBD and 19 children without IBD, distinct signatures were found in those with IBD, but transcription signatures were not correlated with disease outcome, estimated as a composite disease severity score[56].

Proteomics

Measurement of diverse proteins in biological samples provides information further downstream of gene expression and insights into disease pathogenesis. Although targeted measurements of specific proteins using enzyme-linked immunosorbent assays are a long-established technique, newer platforms that can measure a large number of proteins, such as multiplexed bead-based assays and high-resolution mass spectrometry, have expanded the field tremendously.

In a nested case–control study within the Proteomic Evaluation and Discovery in an IBD Cohort of Tri-service Subjects (PREDICTS) proteomic analyses of serological samples from individuals with Crohn's disease or ulcerative colitis were performed[57]. The study's participants were age-matched and sex-matched with healthy individuals (200 patients with Crohn's disease, 199 patients with ulcerative colitis and 200 healthy individuals) up to 5 years prior to IBD diagnosis. It was found that antibodies against microbial antigens, as well as specific inflammatory signatures pertaining to the complement cascade, innate immunity, and glycosaminoglycan and lysosome metabolism, were present up to 5 years prior to Crohn's disease diagnosis, but not prior to ulcerative colitis diagnosis[57]. These signatures, such as markers indicative of the complement cascade and innate immunity, were predictive of Crohn's disease, with higher accuracy closer to the diagnosis (areas under the receiver operating characteristic curve of 0.76 and 0.86, 5 years and 1 year before diagnosis, respectively)[57]. In a study that included a preclinical cohort, an inception cohort and twin pairs, proteomic signatures in individuals who later developed ulcerative colitis were found to be distinct from those in healthy individuals ($n = 72$ and $n = 140$, respectively)[58]. It was found that the proteins MMP10, CXCL9, CCL11 (also known as eotaxin), SLAMF1, CXCL11 and MCP1 (also known as CCL2) were upregulated prior to ulcerative colitis diagnosis. In the inception cohort, a multivariable model including these proteins was predictive of ulcerative colitis with an area under the curve of 0.92 (REF.[58]). Lastly, MMP10, CXCL9, CXCL11 and MCP1, but not CCL11 and SLAMF1, were elevated

in healthy twin siblings of individuals with ulcerative colitis[58]. In the MECONIUM study, offspring born to women with IBD ($n = 76$) were found to have significantly higher stool calprotectin levels than those born to women without IBD ($n = 213$) up to 3 years of age, after adjusting for sex, mode of delivery, feeding behaviour and exposure to antibiotics ($P = 0.015$ at 2 weeks to 3 months of age; $P = 0.00003$ at 12–36 months of age)[59].

Advances in proteomics platforms might bypass some of the limitations of traditional proteomics platforms. For example, nucleic acid programmable protein array (NAPPA), involving cDNA-based protein expression at the time of the assay, bypasses the need for advanced protein purification and considerations around proteins stability during storage, and increases the scope of protein analysis[60]. NAPPA has been applied to the identification of disease-specific signatures in many diseases, including IBD. A study of serum samples from 96 patients with Crohn's disease and 96 healthy individuals identified specific autoantibodies, including four novel IgA autoantibodies, which might prove to be useful in Crohn's disease diagnosis and management[61].

Another platform for analysis of immunological signatures is immunophenotyping, which involves determining the subtypes of immune cells using mass spectrometry. Kosoy et al. studied key immune cell signatures in patients with Crohn's disease or ulcerative colitis and healthy individuals as controls ($n = 728$, $n = 464$ and $n = 334$, respectively) using fluorescence-activated cell sorting analysis. The researchers found altered immune cell populations in patients with IBD, with lower levels of natural killer cells, B cells and $CD45RA^{-}CD8^{+}$ T cells, and further distinctions in the immunophenome based on IBD type and other disease-related variables such as disease duration, behaviour, location and IBD therapy[62].

Glycomics

Glycans are carbohydrate chains that attach to proteins or lipid molecules during post-translational modification, leading to structural, and thereby functional, modification[63]. Glycans serve as the interface between the cell surface and external factors, and the glycome, which is a measure of diverse glycan structures in an organism, is representative of function[64]. Glycans modulate immune function at multiple levels, such as T cell activation, humoral responses, host–microorganism interaction and malignant transformation, and are implicated in many chronic immune-mediated and gastrointestinal diseases[63,64]. Measurement of protein glycosylation in tissue or serum, both untargeted and targeted, using mass spectrometry platforms such as glycoblotting, can provide important insights into preclinical alterations and pathogenic pathways[65].

T cell receptor glycosylation.

A study on the glycosylation profiles of T cell receptors (TCRs) in the lamina propria of colonic biopsy samples from patients with ulcerative colitis on tissue histochemistry for glycans showed defective *N*-glycan branching on TCRs, mediated by lower expression of the *MGAT5* gene, which codes for a glycosyltransferase[66]. The investigators subsequently reported that mouse models null or heterozygous for *MGAT5* were susceptible to severe colitis, and treatment with *N*-acetylglucosamine led to a controlled T cell-mediated immune

response, improved colitis, as was shown by body weight loss and disease activity index, and decreased disease progression[67]. Similarly, supplementation with *N*-acetylglucosamine of ex vivo intestinal T cells from patients with ulcerative colitis improved branched TCRs *N*-glycosylation, with downregulation of T cell growth and T helper 1 (T_H1)/T_H17 response, and regulated T cell function[67]. Core fucosylation, which is the attachment of a fucose molecule to the innermost *N*-glycan in a glycoprotein, was found to be higher in mouse models of trinitrobenzene-induced colitis compared with mice without colitis, and in biopsy samples of inflamed mucosa from patients with IBD compared with their uninflamed tissue as well as compared with tissue from individuals without IBD[68]. Similarly, other studies have found differences in glycome signatures in serum samples from individuals with and without IBD using high-throughput omics platforms[69,70].

Oligosaccharides in breast milk.

Human breast milk contains nutrients and bioactive molecules that are essential for the infant's mucosal immune regulation, including host–microorganism interactions. One such group is human milk oligosaccharides (HMOs), glycans with highly complex and branched structures whose relevance for offspring gut maturation is being recognized[71]. Based on the expression levels of fucosyltransferase 2 and fucosyltransferase 3 in mothers, breast milk can be divided into four groups, each with a distinct HMO profile and composition[71,72]. These can be detected and quantified using liquid chromatography and mass spectrometry[73]. The relevance of HMO composition of breast milk for outcomes in infants is substantial. For example, necrotizing enterocolitis in preterm infants is linked to the colonization of the immature gut by pathogenic bacteria, and HMOs in breast milk are believed to mediate the lower risk of necrotizing enterocolitis in breastfed compared with formula-fed infants[74,75]. HMOs stimulate the proliferation of the genus *Bifidobacterium* and Bacteroidetes species, which dominate the gut microbiome of the healthy term infant[71]. Lastly, more mechanistic data show that HMOs isolated from human milk reduce attachment of enteropathogenic *Escherichia coli* (EPEC) to cultured human colonic epithelial cells as well as EPEC colonization in suckling mice[76].

We propose that HMO composition and effects on mucosal integrity and host–microorganism interactions can have implications for lowering the risk of IBD. Epidemiologically, breastfeeding has an important role in mucosal immune regulation in offspring, and although it is protective against IBD overall, this finding is not consistent across studies[5,77].

Epigenetics

Epigenetic changes imply alterations in gene expression through changes in DNA architecture such as DNA hypermethylation or hypomethylation, histone modification and chromatin remodelling, without changing the DNA sequence, which typically occur at CpG sites, which are DNA sequences where a guanine nucleotide follows a cytosine nucleotide in a linear sequence in the 5′ to 3′ direction. Epigenetic modifications can persist over cycles of cell division and replication, and can be inherited transgenerationally[78,79]. Through the application of epigenome-wide association studies (EWAS), investigators have linked

epigenetic signatures with exposures and diseases to study overall patterns of epigenetic alterations such as DNA methylation at CpG islands of gene promoter regions[79]. Next-generation sequencing provides more comprehensive and unbiased data, including those on allele-specific epigenetic alterations[80].

Diverse environmental insults such as pollution, smoking, heavy metals and inorganic chemicals have been linked with epigenetic alterations, especially when they occur in the early life period[21,81]. The association between prenatal maternal smoking and DNA methylation was demonstrated in a study that included individuals between the ages 16 and 48 years from five prospective birth cohorts. Analysis of the participants' serum samples showed that 69 differentially methylated CpGs in 36 genomic regions are associated with exposure to maternal smoke in the early life period[82]. Epigenetic modifications have been consistently linked with the programming of cellular function and biological ageing[21,79]. Epigenetic changes are also implicated in the loss of immune tolerance, dysregulated $CD8^+$ T cell function and autoimmune diseases such as rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis[78,83]. Increased methylation changes and methylome variation in $CD8^+$ T cells have been found with advancing age[84]. Many of the involved genes pertain to immune response and lineage differentiation, and might explain immunosenescence with advancing age[84].

Epigenetic alterations have been reported in the context of IBD. An EWAS analysis of DNA from whole-blood samples from 240 patients with newly diagnosed IBD (121 with Crohn's disease and 119 with ulcerative colitis) and 191 individuals as controls (74 symptomatic without IBD, and 117 healthy individuals) revealed 439 differentially methylated positions and 5 differentially methylated regions, and these findings were replicable in an independent cohort[85]. A study of DNA methylation signatures in intestinal mucosal samples from children with IBD (43 with Crohn's disease and 23 with ulcerative colitis) found distinct DNA methylation and transcriptomic signatures compared with the signatures in samples from 30 healthy individuals[86]. EWAS analysis was performed on peripheral blood mononuclear cells from 88 individuals with Crohn's disease and 61 with ulcerative colitis, and 39 healthy individuals. The investigators found 3,196 differentially methylated probes between patients with Crohn's disease and healthy individuals, and 1,481 differentially methylated probes between patients with ulcerative colitis and healthy individuals, with a 45% overlap between patients with Crohn's disease and patients with ulcerative colitis. Specifically, genes that were affected included *TRIM39–RPP21* (hypomethylated) and *TRAF6* (hypermethylated) in both patients with Crohn's disease and patients with ulcerative colitis. In addition, *TRAF6* mRNA expression was lower in patients with IBD than in healthy individuals, providing further downstream evidence of hypermethylation at this site and its effects[87]. In a study on the effect of prenatal exposure to smoke, 4 of 69 differentially methylated CpGs (two of which were *GFIP1* and *MIR548F3*, others unknown) were associated with an increased risk of IBD[82]. Finally, an EWAS analysis conducted on blood-derived $CD8^+$ T cells from 66 children with recently diagnosed Crohn's disease revealed that DNA methylation patterns of $CD8^+$ T cells were associated with chronological age, but not with Crohn's disease outcomes, such as treatment escalations and surgery[56].

Gut microbiome analysis

Data have consistently demonstrated differences in microbiome signatures using 16S ribosomal RNA (rRNA) gene sequencing and shotgun metagenomic sequencing, but with substantial heterogeneity. A systematic review published in 2020, based on data from 45 articles that compared intestinal microbiota in patients with IBD and individuals without IBD concluded that results were inconsistent. This discrepancy was mainly due to small sample sizes, lack of adjustment for potential confounders and lack of standardization of microbiome assessment methods, including sampling and storage, DNA extraction, sequencing and bioinformatic pipelines[88].

Similar to other omics data, it is difficult to determine whether the gut microbiota drives IBD or is altered due to IBD. To that end, studies in at-risk and preclinical cohorts are relevant. Using data from the MECONIUM cohort, researchers found that infants born to mothers with IBD had an aberrant faecal microbiota, with lower diversity, an enrichment in gammaproteobacteria and reduction in bifidobacteria compared with infants born to mothers without IBD[89]. Transfer of faecal microbiota from pregnant women with IBD and their offspring to germ-free mice resulted in a decrease in microbial diversity, and in class-switched memory B cells and regulatory T cells in the colonic lamina propria[89]. Differences in stool calprotectin levels in offspring of women with and without IBD in the MECONIUM cohort, as described previously, were associated with microbiome changes, with *Faecalibacterium*, *Bifidobacterium* and *Alistipes* being negatively correlated, and *Streptococcus* being positively correlated with stool calprotectin levels within 3 months of birth[59]. A whole-genome metagenomic shotgun sequencing performed in faecal samples from 99 twins (from 51 pairs), 495 healthy individuals and 99 unrelated patients with IBD demonstrated that IBD-like microbiome signatures found in twins with IBD were also present in healthy cotwins and unrelated patients with IBD[90]. Longitudinal follow-up will clarify the clinical relevance of these findings. A Mendelian randomization approach used human GWAS data and 16S faecal microbiota data in the MiBioGen consortium to examine causal links between the gut microbiota and ulcerative colitis[91]. The results suggested that lower abundance of the class Actinobacteria and its genus *Bifidobacterium* were causally related to ulcerative colitis[91]. Pre-diagnostic and post-diagnostic faecal biomarkers of ulcerative colitis were measured in faecal samples from 13 individuals in the GEM project who developed ulcerative colitis during follow-up[92]. The researchers found increased faecal proteolytic and elastase activity that was linked to an aberrant faecal microbiota prior to disease onset[92]. Elastase activity was negatively correlated with the relative abundance of *Adlercreutzia* and *Akkermansia*, and positively correlated with *Bacteroides vulgatus*, a taxon with proteolytic activity. Although the sample size was small, this study represents an important model for future longitudinal studies of IBD, including paired pre-diagnostic and post-diagnostic samples to elucidate the causal role of the microbiota in IBD.

In addition to longitudinal studies with pre-diagnostic samples, multiomic studies and high-resolution shotgun metagenomic sequencing studies have contributed to our understanding of the role of the microbiome in IBD. In the setting of the Integrative Human Microbiome Project, taxonomic, functional, and biochemical changes during disease flares were demonstrated, providing granular data on the molecular biology of IBD flares[93]. In another

study, diversity, growth and genes involved in virulence and antibiotic resistance were characterized at the level of the bacterial strain, pinpointing key bacterial species and representing potential new prevention and treatment targets[94]. In further support of a causal role of the gut microbiota in IBD, short-term use of faecal microbiota transplantation might have beneficial effects on IBD activity, including clinical remission and endoscopic remission[95].

Apart from bacteria, the intestinal microbiota also includes fungi and viruses (including phages infecting bacteria). Both fungi and viruses interact with the human immune system, and community differences in the fungome and virome have been reported in patients with IBD[96,97]. Study findings are heterogeneous, but the expansion of the fungal genus *Candida* and Caudovirales phages have consistently been associated with IBD[96]. Interestingly, in a study in 24 patients with ulcerative colitis receiving faecal microbiota transplantation and 15 patients with ulcerative colitis receiving placebo, patients with ulcerative colitis with a high abundance of *Candida* prior to faecal microbiota transplantation showed a greater clinical response following faecal microbiota transplantation than those with low *Candida* abundance[98]. Further, in recipients of faecal microbiota transplantation, reduction in *Candida* abundance following transplantation was correlated with reduced disease activity measured by the Mayo score. However, further studies are needed to improve our understanding of the mutual interactions among fungi in the microbiota, and their interactions with intestinal bacteria and with the human immune system.

Neonatal blood spots

Exposures during the early life period are likely to be especially relevant in IBD, as immune maturation occurs in this period[5]. Study of the early life period is associated with challenges such as the need for birth cohorts of sufficiently large sizes and long-term follow-up over decades to enable the disease to occur. This obstacle can be bypassed by the use of a unique resource known as neonatal blood spots (NBSs). These are drops of capillary blood obtained after birth via heel or finger prick, placed on standardized filter paper cards called Guthrie cards, allowed to dry and archived routinely in countries including Denmark and the USA to facilitate screening for congenital metabolic disorders and future research[99] (FIG. 3). NBSs require minimal resources for collection, processing and storage, and represent a massive, untapped resource for molecular epidemiological research pertaining to fetal and perinatal exposures. NBSs are being used in exposome and metabolome analyses[48].

Many countries have been archiving NBSs for over 30 years, providing an opportunity to study life-course diseases such as IBD. Since the implementation of the newborn screening programme in Denmark in 1982, NBSs have been collected routinely from all infants and stored in the Danish Neonatal Screening Biobank[100]. In Denmark, several impressive studies have been done using NBSs, especially in the area of toxoplasmosis, type 1 diabetes mellitus and mental health disorders[101]. In the USA as well, NBSs have been used in omics analyses. For example, an untargeted metabolomic analysis of NBSs was conducted to compare the abundance of small-molecule signatures in children who later developed acute lymphoblastic leukaemia (ALL)[102]. The investigators found that distinct

signatures pertaining to lipid and fatty acid pathways were associated with ALL, with further distinctions based on early versus late ALL diagnosis.

A case–control study in 384 patients with IBD and matched controls in Denmark determined the effect of vitamin D level at birth on paediatric IBD by measuring calcifediol, a form of vitamin D produced in the liver, in NBSs. They found no association between vitamin D level at birth and the risk of paediatric-onset IBD[103].

NBSs represent novel research opportunities, but their successful application in multiomics research pertaining to IBD is awaited. NBSs have been used in research for many years consistently in some countries and their value demonstrated, but their use might not be feasible in other countries—for example, those that lack adequate resources to maintain large-scale biobanks. Protocolized collection and storage, quality control and technical competence are important considerations. Ethical concerns around secondary use of NBSs, such as for research, are also important[101].

Network analysis

In analogy to pieces of a puzzle, once combined, high-dimensional multiomics data can provide information contributing to elucidation of the causal pathway in IBD, using advanced computational techniques. Over the past decade, network analysis has emerged as a powerful tool to integrate disparate omics datasets and to identify key drivers and biological pathways underlying disease initiation and progression[104,105]. These networks reveal the interplay of molecular variables and facilitate the identification of highly connected hubs, which can be targeted for therapeutic and preventive interventions, and are known as the IBD interactome[106].

To facilitate the estimation of such high-dimensional networks, it is key to borrow information from existing databases such as protein–protein interactions or perturbation experiments. These datasets provide complementary sources of information about regulatory relationships, such as network topology. For instance, protein–protein interaction databases, such as STRING, provide information about the network topology and the association across different proteins, whereas knockdown experiments provide information on causal relationships[107]. Various algorithms have been proposed to incorporate such prior information when inferring large-scale networks[108]. In the context of IBD, Peters et al. estimated a large regulatory network integrating DNA and RNA data to characterize the inflammatory component and pathways in IBD[109]. GWAS data were incorporated to identify the causal relationship of gene–gene associations via a Bayesian network model, as genes associated with causal variants are more likely to be causal themselves and drive the expression of other genes in the network. This work estimated causal relationships among loci previously linked to IBD based on GWAS, and contributed by providing an architecture of IBD at the molecular level for different stages of the disease. Such networks can cast light on the association between genes associated with IBD susceptibility loci and their downstream effects on other genes, revealing how genes work together to activate biological processes in IBD. In addition, given the role of environmental exposures towards IBD susceptibility, the Comparative Toxicogenomics Database might be an important resource in

co-expression networks. The use of this database could lead to the identification of genes or network structures induced by exposure to particular chemicals (for example, pesticides and food additives). Similarly, protein–protein interaction databases, knockout experiments from existing IBD studies and other databases such as the Human Glycome Project are relevant resources.

Conclusions

Rapidly expanding omics analytical platforms and tools represent a paradigm shift in molecular epidemiological research. Although there are limited omics data pertaining to IBD pathogenesis at this time, we believe that this is a growing field with tremendous opportunities. Diverse birth, at-risk and pre-diagnostic human cohorts have been established, and multiomics studies to characterize IBD pathways are underway, including those involving the use of NBSs (TABLE 1). Statistically well-powered studies, longitudinal follow-up of omics signatures in individuals before and after IBD onset, and replication of findings in independent cohorts are likely to be informative. Collection of granular clinical data and standardized sampling, storage and analysis pipelines remain essential to making reliable conclusions from individual studies and to enabling combined data analyses across multiple studies. Advances in network analysis could pave the path for advances in molecular epidemiology of IBD and other diseases, which might have important implications for the prediction of IBD onset and preventive strategies (BOX 1). Finally, these data could serve as a blueprint to elucidate pathways pertaining to other immune-mediated and chronic diseases and to improve public health overall (BOX 2).

Acknowledgements

The authors thank J. Gregory, Certified Medical Illustrator, Icahn School of Medicine at Mount Sinai, for the illustrations.

Competing interests

M.A. is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K23DK129762-01). J.-F.C. has received research grants from AbbVie, Janssen Pharmaceuticals and Takeda; has received payment for lectures from AbbVie, Amgen, Allergan, Ferring Pharmaceuticals, Shire and Takeda; has received consulting fees from AbbVie, Amgen, Arena Pharmaceuticals, Boehringer Ingelheim, Bristol Myers Squibb, Celgene Corporation, Eli Lilly, Ferring Pharmaceuticals, Galmed Research, Glaxo Smith Kline, Geneva, Iterative Scopes, Janssen Pharmaceuticals, Kaleido Biosciences, Landos, Otsuka, Pfizer, Prometheus, Sanofi, Takeda and TiGenix; and holds stock options in Intestinal Biotech Development. K.H.A., F.P. and T.J. declare no competing interests.

References

1. Torres J, Mehandru S, Colombel JF & Peyrin-Biroulet L Crohn’s disease. Lancet 389, 1741–1755 (2017). [PubMed: 27914655]
2. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L & Colombel JF Ulcerative colitis. Lancet 389, 1756–1770 (2017). [PubMed: 27914657]
3. Ananthakrishnan AN et al. Environmental triggers in IBD: a review of progress and evidence. Nat. Rev. Gastroenterol. Hepatol 15, 39–49 (2018). [PubMed: 29018271]
4. Piovani D et al. Environmental risk factors for inflammatory bowel diseases: an umbrella review of meta-analyses. Gastroenterology 157, 647–659.e4 (2019). [PubMed: 31014995]
5. Agrawal M et al. Early life exposures and the risk of inflammatory bowel disease: systematic review and meta-analyses. EClinicalMedicine 10.1016/j.eclinm.2021.100884 (2021).

6. Schulte PA et al. in Molecular Epidemiology: Principles and Practices, IARC Scientific Publication 163 Ch. 1 (eds Rothman N et al.) 1–7 (IARC, 2011).
7. Torres J et al. Results of the seventh scientific workshop of ECCO: precision medicine in IBD–prediction and prevention of inflammatory bowel disease. J. Crohns Colitis 15, 1443–1454 (2021). [PubMed: 33730755]
8. Graham DB & Xavier RJ Pathway paradigms revealed from the genetics of inflammatory bowel disease. Nature 578, 527–539 (2020). [PubMed: 32103191]
9. Rioux JD et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat. Genet 39, 596–604 (2007). [PubMed: 17435756]
10. Ogura Y et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411, 603–606 (2001). [PubMed: 11385577]
11. Liu JZ et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet 47, 979–986 (2015). [PubMed: 26192919]
12. Inoue N et al. Lack of common NOD2 variants in Japanese patients with Crohn's disease. Gastroenterology 123, 86–91 (2002). [PubMed: 12105836]
13. Somineni HK et al. Whole-genome sequencing of African Americans implicates differential genetic architecture in inflammatory bowel disease. Am. J. Hum. Genet 108, 431–445 (2021). [PubMed: 33600772]
14. Spencer EA, Helmus D, Telesco S, Colombel JF & Dubinsky MC Inflammatory bowel disease clusters within affected sibships in Ashkenazi Jewish multiplex families. Gastroenterology 159, 381–382 (2020). [PubMed: 32199878]
15. Gomes CF, Jensen CB, Allin K, Torres J & Burisch JM Sa509 risk factors associated with familial inflammatory bowel disease [abstract]. Gastroenterology 160, S–528 (2021).
16. Turpin W et al. Associations of NOD2 polymorphisms with Erysipelotrichaceae in stool of in healthy first degree relatives of Crohn's disease subjects. BMC Med. Genet 21, 204 (2020). [PubMed: 33059653]
17. Miller GW & Jones DP The nature of nurture: refining the definition of the exposome. Toxicol. Sci 137, 1–2 (2014). [PubMed: 24213143]
18. Niedzwiecki MM et al. The exposome: molecules to populations. Annu. Rev. Pharmacol. Toxicol 59, 107–127 (2019). [PubMed: 30095351]
19. Perera FP et al. Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. Env. Health Perspect 111, 201–205 (2003). [PubMed: 12573906]
20. Volk HE et al. Prenatal air pollution exposure and neurodevelopment: a review and blueprint for a harmonised approach within ECHO. Environ. Res 196, 110320 (2021). [PubMed: 33098817]
21. Peters A, Nawrot TS & Baccarelli AA Hallmarks of environmental insults. Cell 184, 1455–1468 (2021). [PubMed: 33657411]
22. Agrawal M, Hillenbrand CM, Allin KH & Torres J Early life greenspace and the risk of pediatric-onset inflammatory bowel disease: insights into the link between environmental and human health. Gastroenterology 161, 355–357 (2021). [PubMed: 33865842]
23. Rundle AG et al. Prenatal exposure to airborne polycyclic aromatic hydrocarbons and childhood growth trajectories from age 5–14 years. Environ. Res 177, 108595 (2019). [PubMed: 31352299]
24. McConnell R et al. A longitudinal cohort study of body mass index and childhood exposure to secondhand tobacco smoke and air pollution: the Southern California Children's Health Study. Env. Health Perspect 123, 360–366 (2015). [PubMed: 25389275]
25. Fleisch AF et al. Prenatal and early life exposure to traffic pollution and cardiometabolic health in childhood. Pediatr. Obes 12, 48–57 (2017). [PubMed: 26843357]
26. Ruokolainen L et al. Green areas around homes reduce atopic sensitisation in children. Allergy 70, 195–202 (2015). [PubMed: 25388016]
27. Ribeiro AI, Tavares C, Guttentag A & Barros H Association between neighbourhood green space and biological markers in school-aged children. Findings from the Generation XXI birth cohort. Environ. Int 132, 105070 (2019). [PubMed: 31387021]

28. Elten M et al. Ambient air pollution and the risk of pediatric-onset inflammatory bowel disease: a population-based cohort study. Environ. Int 138, 105676 (2020). [PubMed: 32217428]
29. Elten M et al. Residential greenspace in childhood reduces risk of pediatric inflammatory bowel disease: a population-based cohort study. Am. J. Gastroenterol 116, 347–353 (2020).
30. Lovinsky-Desir S et al. Air pollution, urgent asthma medical visits and the modifying effect of neighborhood asthma prevalence. Pediatr. Res 85, 36–42 (2019). [PubMed: 30337671]
31. Mao G et al. Individual and joint effects of early-life ambient exposure and maternal prepregnancy obesity on childhood overweight or obesity. Env. Health Perspect 125, 067005 (2017). [PubMed: 28669938]
32. Jedrychowski WA et al. Prenatal exposure to fine particles and polycyclic aromatic hydrocarbons and birth outcomes: a two-pollutant approach. Int. Arch. Occup. Env. Health 90, 255–264 (2017). [PubMed: 28168423]
33. Hirten RP et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J. Med. Internet Res 23, e26107 (2021). [PubMed: 33529156]
34. Hirten RP et al. Longitudinal autonomic nervous system measures correlate with stress and ulcerative colitis disease activity and predict flare. Inflamm. Bowel Dis 27, 1576–1584 (2020).
35. Widbom L et al. Elevated plasma cotinine is associated with an increased risk of developing IBD, especially among users of combusted tobacco. PLoS ONE 15, e0235536 (2020). [PubMed: 32614903]
36. Walker DI et al. High-resolution metabolomics of occupational exposure to trichloroethylene. Int. J. Epidemiol 45, 1517–1527 (2016). [PubMed: 27707868]
37. Patti GJ, Yanes O & Siuzdak G Innovation: Metabolomics: the apogee of the omics trilogy. Nat. Rev. Mol. Cell Biol 13, 263–269 (2012). [PubMed: 22436749]
38. Li N et al. Gestational and childhood exposure to per-and polyfluoroalkyl substances and cardiometabolic risk at age 12 years. Environ. Int 147, 106344 (2021). [PubMed: 33418195]
39. Jedrychowski WA et al. Depressed height gain of children associated with intrauterine exposure to polycyclic aromatic hydrocarbons (PAH) and heavy metals: the cohort prospective study. Environ. Res 136, 141–147 (2015). [PubMed: 25460630]
40. Lochhead P et al. Plasma concentrations of perfluoroalkyl substances and risk of inflammatory bowel diseases in women: a nested case control analysis in the Nurses' Health Study cohorts. Environ. Res 207, 112222 (2021). [PubMed: 34662575]
41. Steenland K, Zhao L & Winquist A A cohort incidence study of workers exposed to perfluorooctanoic acid (PFOA). Occup. Env. Med 72, 373–380 (2015). [PubMed: 25601914]
42. Xu Y et al. Inflammatory bowel disease and biomarkers of gut inflammation and permeability in a community with high exposure to perfluoroalkyl substances through drinking water. Environ. Res 181, 108923 (2020). [PubMed: 31759646]
43. Canturk N, Atsu SS, Aka PS & Dagalp R Neonatal line on fetus and infant teeth: an indicator of live birth and mode of delivery. Early Hum. Dev 90, 393–397 (2014). [PubMed: 24951074]
44. Nair N et al. Association between early-life exposures and inflammatory bowel diseases, based on analyses of deciduous teeth. Gastroenterology 159, 383–385 (2020). [PubMed: 32224130]
45. Lai Y et al. Serum metabolomics identifies altered bioenergetics, signaling cascades in parallel with exposome markers in Crohn's disease. Molecules 24, 449 (2019).
46. Kolho KL, Pessia A, Jaakkola T, de Vos WM & Velagapudi V Faecal and serum metabolomics in paediatric inflammatory bowel disease. J. Crohns Colitis 11, 321–334 (2017). [PubMed: 27609529]
47. Daniluk U et al. Untargeted metabolomics and inflammatory markers profiling in children with Crohn's disease and ulcerative colitis–a preliminary study. Inflamm. Bowel Dis 25, 1120–1128 (2019). [PubMed: 30772902]
48. Petrick LM, Uppal K & Funk WE Metabolomics and adductomics of newborn bloodspots to retrospectively assess the early-life exposome. Curr. Opin. Pediatr 32, 300–307 (2020). [PubMed: 31913157]
49. Funk WE, Waidyanatha S, Chaing SH & Rappaport SM Hemoglobin adducts of benzene oxide in neonatal and adult dried blood spots. Cancer Epidemiol. Biomark. Prev 17, 1896–1901 (2008).

50. Waidyanatha S, Zheng Y, Serdar B & Rappaport SM Albumin adducts of naphthalene metabolites as biomarkers of exposure to polycyclic aromatic hydrocarbons. Cancer Epidemiol. Biomark. Prev 13, 117–124 (2004).
51. Lydic TA & Goo YH Lipidomics unveils the complexity of the lipidome in metabolic diseases. Clin. Transl. Med 7, 4 (2018). [PubMed: 29374337]
52. Diab J et al. Lipidomics in ulcerative colitis reveal alteration in mucosal lipid composition associated with the disease state. Inflamm. Bowel Dis 25, 1780–1787 (2019). [PubMed: 31077307]
53. Iwatani S et al. Novel mass spectrometry-based comprehensive lipidomic analysis of plasma from patients with inflammatory bowel disease. J. Gastroenterol. Hepatol 35, 1355–1364 (2020). [PubMed: 32285970]
54. Murgia A et al. Italian cohort of patients affected by inflammatory bowel disease is characterised by variation in glycerophospholipid, free fatty acids and amino acid levels. Metabolomics 14, 140 (2018). [PubMed: 30830399]
55. Deyssenroth MA et al. Placental gene networks at the interface between maternal PM2.5 exposure early in gestation and reduced infant birthweight. Environ. Res 199, 111342 (2021). [PubMed: 34015297]
56. Gasparetto M et al. Transcription and DNA methylation patterns of blood-derived CD8(+) T cells are associated with age and inflammatory bowel disease but do not predict prognosis. Gastroenterology 160, 232–244.e7 (2021). [PubMed: 32814113]
57. Torres J et al. Serum biomarkers identify patients who will develop inflammatory bowel diseases up to 5 years before diagnosis. Gastroenterology 159, 96–104 (2020). [PubMed: 32165208]
58. Bergemalm D et al. Systemic inflammation in pre-clinical ulcerative colitis. Gastroenterology 161, 1526–1539.e9 (2021). [PubMed: 34298022]
59. Kim ES et al. Longitudinal changes in fecal calprotectin levels among pregnant women with and without inflammatory bowel disease and their babies. Gastroenterology 160, 1118–1130.e3 (2020). [PubMed: 33307026]
60. Ramachandran N et al. Self-assembling protein microarrays. Science 305, 86–90 (2004). [PubMed: 15232106]
61. Wang H et al. Identification of antibody against SNRPB, small nuclear ribonucleoprotein-associated proteins B and B', as an autoantibody marker in Crohn's disease using an immunoproteomics approach. J. Crohns Colitis 11, 848–856 (2017). [PubMed: 28204086]
62. Kosoy R et al. Deep analysis of the peripheral immune system in IBD reveals new insight in disease subtyping and response to monotherapy or combination therapy. Cell. Mol. Gastroenterol. Hepatol 12, 599–632 (2021). [PubMed: 33813036]
63. Verhelst X et al. Protein glycosylation as a diagnostic and prognostic marker of chronic inflammatory gastrointestinal and liver diseases. Gastroenterology 158, 95–110 (2020). [PubMed: 31626754]
64. Marth JD & Grewal PK Mammalian glycosylation in immunity. Nat. Rev. Immunol 8, 874–887 (2008). [PubMed: 18846099]
65. Miura Y et al. BlotGlycoABCTM, an integrated glycoblotting technique for rapid and large scale clinical glycomics. Mol. Cell Proteom 7, 370–377 (2008).
66. Dias AM et al. Dysregulation of T cell receptor N-glycosylation: a molecular mechanism involved in ulcerative colitis. Hum. Mol. Genet 23, 2416–2427 (2014). [PubMed: 24334766]
67. Dias AM et al. Metabolic control of T cell immune response through glycans in inflammatory bowel disease. Proc. Natl Acad. Sci. USA 115, e4651–e4660 (2018). [PubMed: 29720442]
68. Fujii H et al. Core fucosylation on T cells, required for activation of T-cell receptor signaling and induction of colitis in mice, is increased in patients with inflammatory bowel disease. Gastroenterology 150, 1620–1632 (2016). [PubMed: 26965517]
69. Trbojević Akmačić I et al. Inflammatory bowel disease associates with proinflammatory potential of the immunoglobulin G glycome. Inflamm. Bowel Dis 21, 1237–1247 (2015). [PubMed: 25895110]
70. Clerc F et al. Plasma N-glycan signatures are associated with features of inflammatory bowel diseases. Gastroenterology 155, 829–843 (2018). [PubMed: 29792883]

71. Bering SB Human milk oligosaccharides to prevent gut dysfunction and necrotizing enterocolitis in preterm neonates. Nutrients 10, 1461 (2018).
72. Han SM, Binia A, Godfrey KM, El-Heis S & Cutfield WS Do human milk oligosaccharides protect against infant atopic disorders and food allergy? Nutrients 12, 3212 (2020).
73. Austin S & Bénet T Quantitative determination of non-lactose milk oligosaccharides. Anal. Chim. Acta 1010, 86–96 (2018). [PubMed: 29447675]
74. Battersby C, Longford N, Mandalia S, Costeloe K & Modi N Incidence and enteral feed antecedents of severe neonatal necrotising enterocolitis across neonatal networks in England, 2012–13: a whole-population surveillance study. Lancet Gastroenterol. Hepatol 2, 43–51 (2017). [PubMed: 28404014]
75. Autran CA et al. Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants. Gut 67, 1064–1070 (2018). [PubMed: 28381523]
76. Manthey CF, Autran CA, Eckmann L & Bode L Human milk oligosaccharides protect against enteropathogenic Escherichia coli attachment in vitro and EPEC colonisation in suckling mice. J. Pediatr. Gastroenterol. Nutr 58, 165–168 (2014). [PubMed: 24048169]
77. Baron S et al. Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. Gut 54, 357–363 (2005). [PubMed: 15710983]
78. Meda F, Folci M, Baccarelli A & Selmi C The epigenetics of autoimmunity. Cell. Mol. Immunol 8, 226–236 (2011). [PubMed: 21278766]
79. Baccarelli A & Bollati V Epigenetics and environmental chemicals. Curr. Opin. Pediatr 21, 243–251 (2009). [PubMed: 19663042]
80. Meaburn E & Schulz R Next generation sequencing in epigenetics: insights and challenges. Semin. Cell Dev. Biol 23, 192–199 (2012). [PubMed: 22027613]
81. Czamara D et al. Integrated analysis of environmental and genetic influences on cord blood DNA methylation in newborns. Nat. Commun 10, 2548 (2019). [PubMed: 31186427]
82. Wiklund P et al. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. Clin. Epigenet 11, 97 (2019).
83. Deng Q et al. The emerging epigenetic role of CD8+ T cells in autoimmune diseases: a systematic review. Front. Immunol 10, 856 (2019). [PubMed: 31057561]
84. Tserel L et al. Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes. Sci. Rep 5, 13107 (2015). [PubMed: 26286994]
85. Ventham NT et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. Nat. Commun 7, 13507 (2016). [PubMed: 27886173]
86. Howell KJ et al. DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. Gastroenterology 154, 585–598 (2018). [PubMed: 29031501]
87. McDermott E et al. DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. J. Crohns Colitis 10, 77–86 (2016). [PubMed: 26419460]
88. Pittayanon R et al. Differences in gut microbiota in patients with vs without inflammatory bowel diseases: a systematic review. Gastroenterology 158, 930–946.e1 (2020). [PubMed: 31812509]
89. Torres J et al. Infants born to mothers with IBD present with altered gut microbiome that transfers abnormalities of the adaptive immune system to germ-free mice. Gut 69, 42–51 (2020). [PubMed: 31036757]
90. Brand EC et al. Healthy cotwins share gut microbiome signatures with their inflammatory bowel disease twins and unrelated patients. Gastroenterology 160, 1970–1985 (2021). [PubMed: 33476671]
91. Kurilshikov A et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat. Genet 53, 156–165 (2021). [PubMed: 33462485]
92. Galipeau HJ et al. Novel fecal biomarkers that precede clinical diagnosis of ulcerative colitis. Gastroenterology 160, 1532–1545 (2021). [PubMed: 33310084]

93. Lloyd-Price J et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019). [PubMed: 31142855]
94. Vich Vila A et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Sci. Transl. Med 10, eaap8914 (2018). [PubMed: 30567928]
95. Narula N et al. Systematic review and meta-analysis: fecal microbiota transplantation for treatment of active ulcerative colitis. Inflamm. Bowel Dis 23, 1702–1709 (2017). [PubMed: 28906291]
96. Iliev ID & Cadwell K Effects of intestinal fungi and viruses on immune responses and inflammatory bowel diseases. Gastroenterology 160, 1050–1066 (2021). [PubMed: 33347881]
97. Zuo T et al. Gut mucosal virome alterations in ulcerative colitis. Gut 68, 1169–1179 (2019). [PubMed: 30842211]
98. Leonardi I et al. Fungal trans-kingdom dynamics linked to responsiveness to fecal microbiota transplantation (FMT) therapy in ulcerative colitis. Cell Host Microbe 27, 823–829.e3 (2020). [PubMed: 32298656]
99. Barr DB et al. The use of dried blood spots for characterising children's exposure to organic environmental chemicals. Environ. Res 195, 110796 (2021). [PubMed: 33508256]
100. Nørgaard-Pedersen B & Simonsen H Biological specimen banks in neonatal screening. Acta Paediatr. Suppl 88, 106–109 (1999).
101. Nordfalk F & Ekstrøm CT Newborn dried blood spot samples in Denmark: the hidden figures of secondary use and research participation. Eur. J. Hum. Genet 27, 203–210 (2019). [PubMed: 30287898]
102. Petrick LM et al. Metabolomics of neonatal blood spots reveal distinct phenotypes of pediatric acute lymphoblastic leukemia and potential effects of early-life nutrition. Cancer Lett. 452, 71–78 (2019). [PubMed: 30904619]
103. Thorsen SU et al. Perinatal vitamin D levels are not associated with later risk of developing pediatric-onset inflammatory bowel disease: a Danish case-cohort study. Scand. J. Gastroenterol 51, 927–933 (2016). [PubMed: 26872831]
104. Petralia F et al. Integrated proteogenomic characterization across major histological types of pediatric brain cancer. Cell 183, 1962–1985.e31 (2020). [PubMed: 33242424]
105. Chen Y et al. Variations in DNA elucidate molecular networks that cause disease. Nature 452, 429–435 (2008). [PubMed: 18344982]
106. de Souza HSP, Fiocchi C & Iliopoulos D The IBD interactome: an integrated view of aetiology, pathogenesis and therapy. Nat. Rev. Gastroenterol. Hepatol 14, 739–749 (2017). [PubMed: 28831186]
107. Szklarczyk D et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452 (2015). [PubMed: 25352553]
108. Petralia F, Wang P, Yang J & Tu Z Integrative random forest for gene regulatory network inference. Bioinformatics 31, i197–i205 (2015). [PubMed: 26072483]
109. Peters LA et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. Nat. Genet 49, 1437–1449 (2017). [PubMed: 28892060]
110. Khalili H et al. Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. Gut 62, 1153–1159 (2013). [PubMed: 22619368]
111. Khalili H et al. Physical activity and risk of inflammatory bowel disease: prospective study from the Nurses' Health Study cohorts. BMJ 347, f6633 (2013). [PubMed: 24231178]
112. Lo CH et al. Ultra-processed foods and risk of Crohn's disease and ulcerative colitis: a prospective cohort study. Clin. Gastroenterol. Hepatol 10.1016/j.cgh.2021.08.031 (2021).
113. Turpin W et al. Analysis of genetic association of intestinal permeability in healthy first-degree relatives of patients with Crohn's disease. Inflamm. Bowel Dis 25, 1796–1804 (2019). [PubMed: 31251335]
114. Lee S-H et al. Anti-microbial antibody response is associated with future onset of Crohn's disease independent of biomarkers of altered gut barrier function, subclinical inflammation, and genetic risk. Gastroenterology 161, 1540–1551 (2019).

Box 1 |

The past, present and future of multiomics research in IBD

Established

- Clinical epidemiological studies
- Genome-wide association studies and whole-genome sequencing
- Early data on pre-diagnostic and microbiome signatures

Ongoing

- Developing omics platforms
- Assembly of birth, at-risk and pre-diagnostic cohorts globally
- Multiomics signatures of other chronic diseases as a framework for inflammatory bowel disease (IBD)

Future

- Network analyses to understand IBD pathogenesis
- Harmonized multiomics platforms across cohorts
- Multiomics signatures of IBD and validation in external cohorts

Goals

- Predict IBD
- Prevent IBD

Box 2 |

Key points and future directions

- While traditional epidemiological studies have provided focal insights into risk factors for inflammatory bowel disease (IBD), they carry inherent limitations, and IBD causal pathways are not well understood.
- Multiomics analyses, which leverage unbiased high-throughput data, are relevant to understanding IBD risk factors and pathways. Early life, at risk, and pre-diagnostic periods are of special interest in this regard.
- Genetic polymorphisms, such as those involved in the innate immune response to bacterial antigens, host–microorganism interactions and autophagy are the most consistent risk factors for IBD but explain risk only in part. Further study of genetic–non-genetic interactions is warranted.
- Environmental health-related exposures such as air pollutants and green space might modulate IBD risk. Geographic information system-based studies will help clarify the relevance of these exposures.
- Heavy metals might be linked to IBD risk, and large-scale targeted and untargeted exposome analyses are needed to explore this further. Similarly, data on metabolomics, adductomics and lipidomics signatures will be informative.
- Proteomics signatures pertaining to diverse inflammatory pathways are altered prior to Crohn’s disease and ulcerative colitis diagnosis, and might predict disease onset.
- Epigenetic alterations, which are linked to exposures such as maternal smoking during pregnancy, mediate immune dysregulation and the risk of immune-mediated diseases, including IBD.
- Microbiome data indicate clear differences in gut bacterial communities with IBD, but signatures prior to IBD onset will be helpful. In offspring of women with IBD, the gut microbiome is persistently dysbiotic and is associated with elevated faecal levels of calprotectin, a marker of intestinal mucosal inflammation.
- Network analyses to integrate omics data from complementary sources into cohesive models are critical to map the architecture of IBD pathways and infer causal relationships.
- Neonatal blood spots are an under-utilized resource for multiomics analysis of early life signatures in IBD.

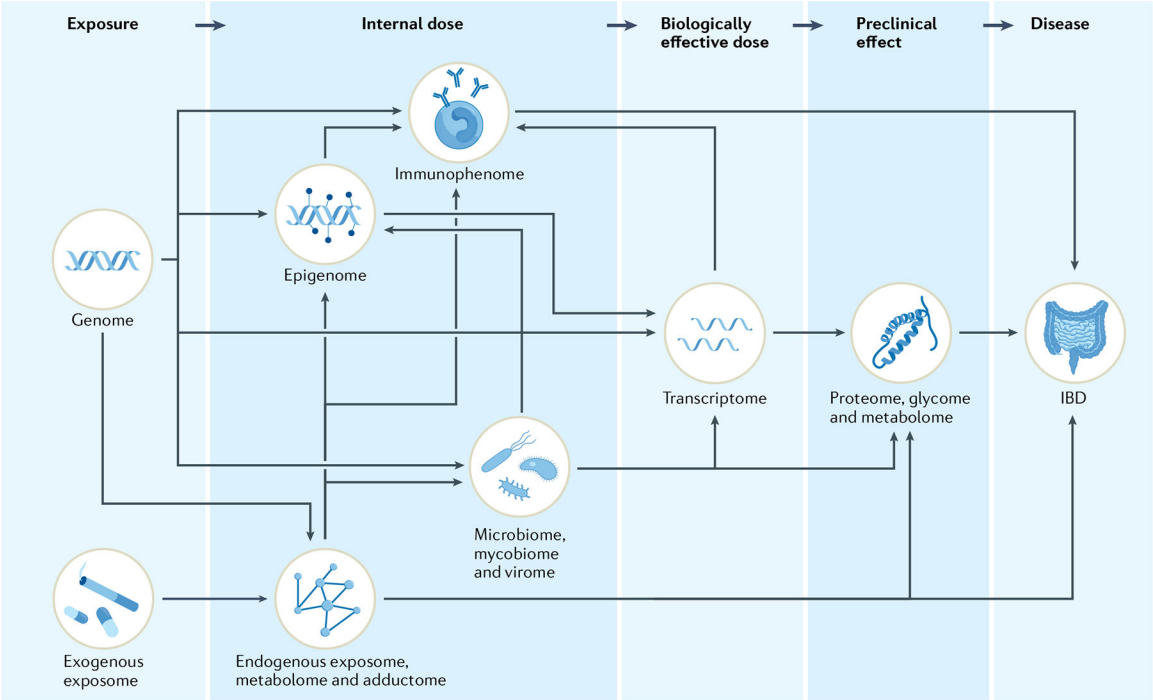

Fig. 1 |. Complex pathways lead to IBD.

Complex pathways starting with genetic and non-genetic risk factors and followed by downstream changes lead to inflammatory bowel disease (IBD). Altered signatures prior to disease onset can be ascertained using different omics platforms. These molecular mechanisms reflect the molecular epidemiology paradigm from exposure to disease and include internal dose, biologically effective dose and preclinical effect as intermediate steps[6]. Adapted from images courtesy of ©Mount Sinai Health System.

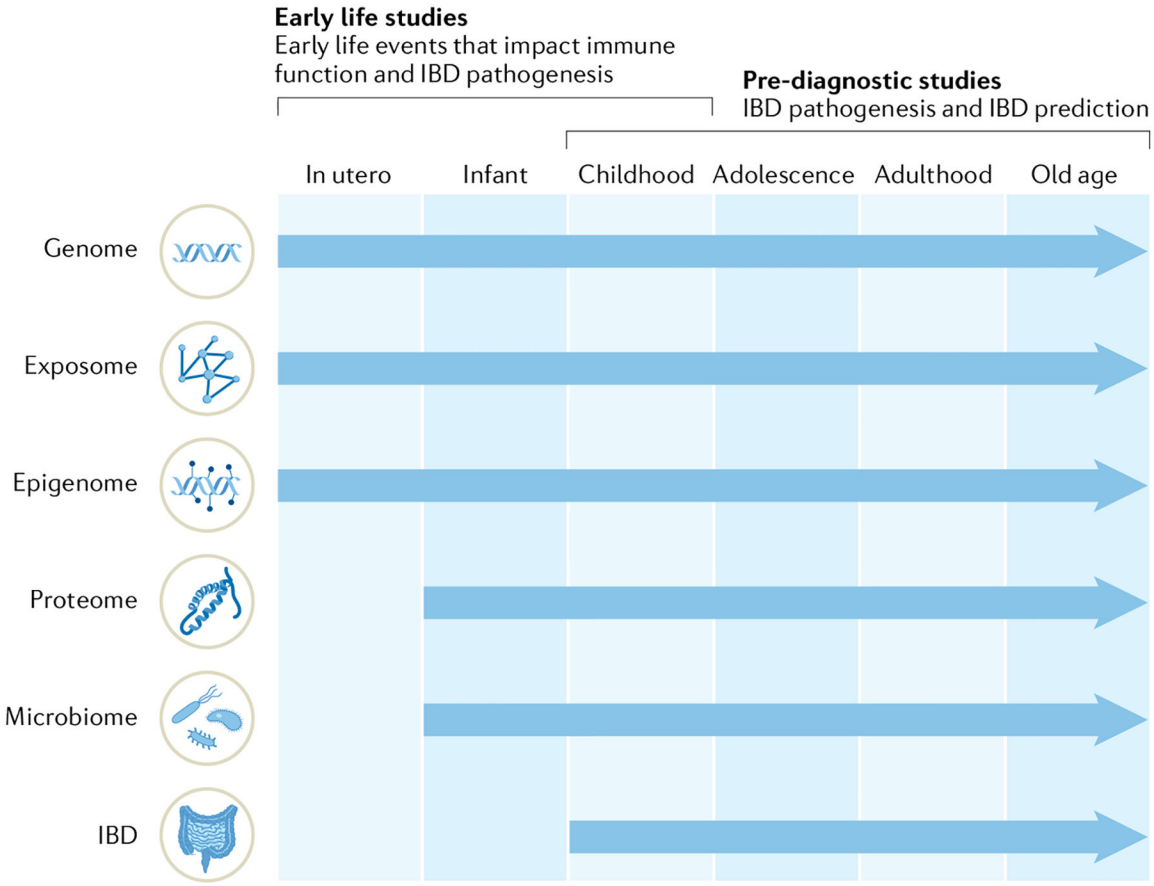

Fig. 2 |. Differences in the application of early life cohorts and pre-diagnostic cohorts in characterizing various omics signatures of IBD during different periods of life.

While early life cohorts can be used to understand early life events and omics signatures to understand changes in immune function and inflammatory bowel disease (IBD) pathogenesis, pre-diagnostic cohorts can help predict IBD risk and characterize IBD pathogenesis based on signatures closer to IBD diagnosis. Adapted from images courtesy of ©Mount Sinai Health System.

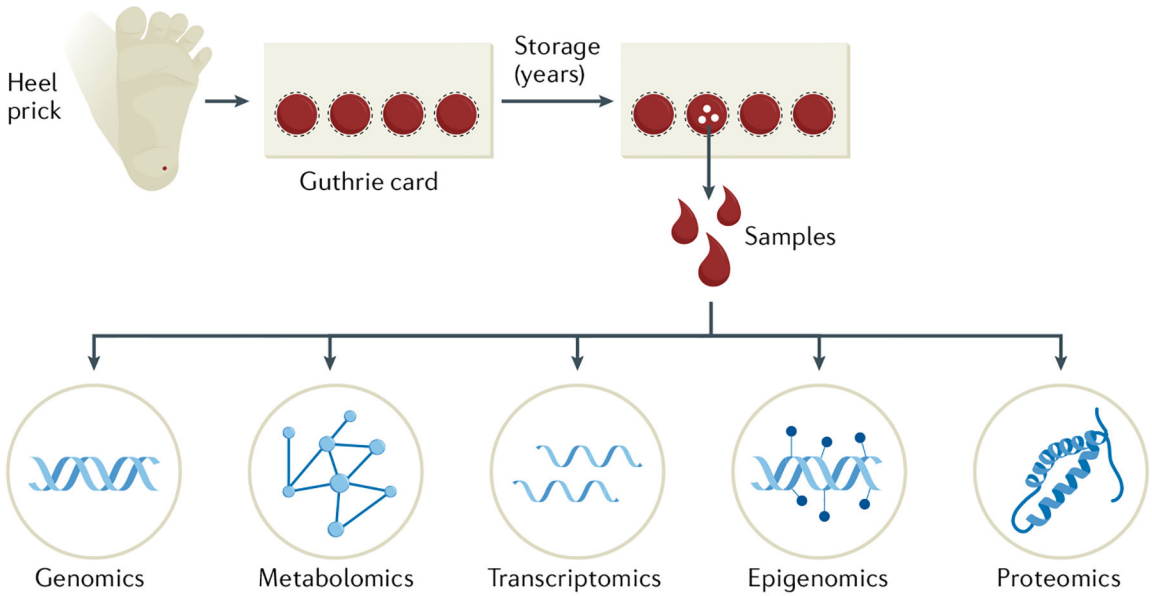

Fig. 3 |. Applications from neonatal screening.

Neonatal blood spots are collected after birth via heel or finger prick, stored on standardized filter paper cards, and can be archived for decades to facilitate screening for congenital metabolic disorders and for future research. Samples from neonatal blood spots can be used to characterize metabolomics, genetics and other omics signatures during the early life period that may be associated with diseases later in life. They represent a paradigm shift in molecular research pertaining to the early life period. Adapted from images courtesy of ©Mount Sinai Health System.

Table 1 |

A summary of prospective birth, at risk and pre-diagnostic cohorts to study IBD and related outcomes

Name of cohort, year	Number of enrolled participants	Cohort and/or study description	primary goal	Types and timing of biological samples collected	Types of omics analyses conducted/planned	Major published findings
Harvard Health Study cohorts, USA						
Nurses' Health Study, 1976	121,000	Pre-diagnostic; prospective cohorts of health-care providers	The non-genetic causes of IBD and other diseases	Blood (two separate collections), cheek cell swabs	Genomics, metabolomics, proteomics	Established the importance of various environmental risk factors in the development of Crohn's disease and ulcerative colitis[110–112]; reported on the association between plasma levels of perfluoroalkyl substances and IBD[40]
Nurses' Health Study 2, 1989	117,000					
Health Professionals Follow-Up Study, 1986	52,000					
Preclinical cohorts, Sweden						
Malmö Diet and Cancer cohort, 1991	28,000	Birth, FDRs of individuals with IBD, pre-diagnostic; prospective cohorts	The causes and multiomics signatures of IBD onset and course	Blood	Metabolomics, proteomics	Proteomic signatures in serum samples predictive of ulcerative colitis onset[58]
Swedish Newborn Dry Blood Spots cohort, 1975	Population-based			Dried blood spots		
Swedish IBD Twin cohort, 2003	>100 twin pairs			Blood, intestinal biopsies, saliva and stool		
Northern Sweden Health and Disease Study, 2011	135,000			Blood		
Other cohorts						
The Crohn's and Colitis Canada Genetics, Environmental, Microbial (GEM) Project, 2008	5,122	FDRs; prospective cohort of FDRs of individuals with Crohn's disease	The multiomics signatures of Crohn's disease	Blood, stool and urine	Genomics, metabolomics, proteomics, and microbiome analysis	Genetic risk profile of FDRs of individuals with Crohn's disease, interaction between genomic and microbiome signatures, faecal biomarkers preceding ulcerative colitis diagnosis, and antimicrobial antibody signatures predictive of Crohn's disease onset[16,92,113,114]
Mechanisms of Disease Transmission In Utero Through the Microbiome (MECONIUM) cohort, USA, 2014	133 mothers with IBD and their offspring; 299 mothers without IBD and their offspring	Early life, FDRs; prospective cohort of pregnant women with and without IBD and their offspring	The link between maternal and offspring multiomics signatures of IBD and transmission of the microbiota	Stool, saliva, vaginal swab, blood, placenta, cord blood, breast milk (mother), meconium, stool and buccal swab (offspring)	Metabolomics, proteomics, microbiome analysis	Microbiome signatures and elevated faecal calprotectin in offspring of women with versus without IBD[59,89]
Road to Prevention: A Mount Sinai Initiative, USA, 2016	667 individuals from multiplex	Birth, FDRs; cross-sectional study design of	The multiomics signatures of IBD onset	Blood, stool, deciduous teeth, hair, saliva	Genomics, exposome analysis, transcriptomics, proteomics,	Clustering of IBD in multiplex families[a](REF.[14])

Name of cohort, year	Number of enrolled participants	Cohort and/or study description	primary goal	Types and timing of biological samples collected	Types of omics analyses conducted/planned	Major published findings
	families[a]; 94 individuals from control families	multiplex fami1ies[a] with IBD with longitudinal follow-up and biospecimen collections from unaffected high-risk individuals			immunophenomics, microbiome analysis	
Twin cohort for the study of (pre) clinical IBD in the Netherlands (TWIN), 2017	124	Pre-diagnostic, FDRs; a prospective cohort of twin pairs >16 years of age discordant or concordant for IBD	The causes of IBD and the early multiomics signatures of IBD	Blood, urine, stool, oropharyngeal swabs, rectal biopsies	Proteomics, immunophenomics, microbiome and mycobiome analysis, IgA coating of bacteria	Microbiome signatures in twin pairs discordant for IBD[90]
Proteomic Evaluation and Discovery in an IBD Cohort ofTri-service Subjects (PREDICTS), USA, 2019	1,000 patients each with Crohn's disease or ulcerative colitis, and 500 healthy individuals as controls	Pre-diagnostic; a nested case-control study of subjects with incident IBD and healthy controls from a cohort of US military personnel	The multiomics signatures of IBD onset and course	Serum	Genomics, metabolomics, proteomics, glycomics, metagenomics, epigenetics	Proteomics signatures in serum samples predictive of Crohn's disease onset[57]
IBD Tooth Fairy Study, Portugal, 2019	65 patients with IBD, 102 healthy individuals	Birth, FDRs; collection and analysis of deciduous teeth of adults and children with and without IBD	The effect of early life exposures and the timing of exposures on subsequent IBD risk	Deciduous teeth	Exposome analysis, including metals and organic compounds	Differences in the levels of metals in teeth from children with and without IBD[44]
Mother-to-Infant Transfer of Bacteriome, Virome, Fungome and Metabolome in Health and Crohn's Disease (Mommy-CD), Hong Kong, 2019	1,000 healthy mothers and 620 infants; 15 pregnant mothers with IBD and 5 infants	Birth; prospective cohort of pregnant women with and without IBD and their offspring	The link between maternal and offspring multiomics signatures of IBD and transmission of the microbiome	Stool, saliva, vaginal swab, blood, placenta, cord blood, breast milk (mother); stool and saliva (father); meconium, stool and buccal swab (offspring)	Metabolomics, proteomics, microbiome analysis	NA
Center for Molecular Prediction of IBD (PREDICT), Denmark, 2021	10,000 IBD; 10,000 healthy individuals	Birth, pre-diagnostic; national population-based prospective cohort of individuals with and without IBD	The causes and multiomics signatures of IBD onset and course by combining omics data with nationwide longitudinal health register data on treatment, comorbidities and outcomes	Neonatal blood spots, pre-diagnostic and post-diagnostic serum samples, and on a subset stool	Genomics, geographic information system, metabolomics, proteomics, epigenetics, metagenomics, exposome	NA

FDR, first-degree relative; IBD, inflammatory bowel disease; NA, not applicable.

[a]Families with two or more members with IBD.