Genetic and dietary modulators of the inflammatory response in the gastro-intestinal tract of the BXD mouse genetic reference population

Xiaoxu Li[1], Jean-David Morel[1], Giorgia Benegiamo[1], Johanne Poisson[1], Alexis Bachmann[1], Alexis Rapin[1], Jonathan Sulc, Evan Williams[3], Alessia Perino[2], Kristina Schoonjans[2], Maroun Bou Sleiman[1,*], Johan Auwerx[1,*]

[1]Laboratory of Integrative Systems Physiology, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

[2]Laboratory of Metabolic Signaling, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

[3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg, Luxembourg.

*e-mail for correspondence: maroun.bousleiman@epfl.ch (MBS) or admin.auwerx@epfl.ch (JA)

Abstract

Inflammatory gut disorders, including inflammatory bowel disease (IBD), can be impacted by dietary, environmental and genetic factors. While the incidence of IBD is increasing worldwide, we still lack a complete understanding of the gene-by-environment interactions underlying inflammation and IBD. Here, we profiled the colon transcriptome of 52 BXD mouse strains fed with a chow or high-fat diet (HFD) and identified a subset of BXD strains that exhibit an IBD-like transcriptome signature on HFD, indicating that an interplay of genetics and diet can significantly affect intestinal inflammation. Using gene co-expression analyses, we identified modules that are enriched for IBD-dysregulated genes and found that these IBD-related modules share *cis*-regulatory elements that are responsive to the STAT2, SMAD3, and REL transcription factors. We used module quantitative trait locus (ModQTL) analyses to identify genetic loci associated with the expression of these modules. Through a prioritization scheme involving systems genetics in the mouse and integration with external human datasets, we identified *Muc4* and *Epha6* as the top candidates mediating differences in HFD-driven intestinal inflammation. This work provides insights into the contribution of genetics and diet to IBD risk and identifies two candidate genes, *MUC4* and *EPHA6*, that may mediate IBD susceptibility in humans.

Keywords: BXD genetic reference population (GRP), System genetics, Inflammatory bowel disease (IBD), High-fat diet (HFD), Chow diet (CD), Weighted gene correlation network analysis (WGCNA), Transcription factors (TFs), Module quantitative trait locus (ModQTL) mapping

Introduction

A long-term lipid-rich diet is associated with multiple metabolic disorders, such as obesity (Hasegawa *et al.*, 2020), cardiovascular disease (Lutsey, Steffen and Stevens, 2008; Maurya *et al.*, 2023), and systemic low-grade inflammation (Duan *et al.*, 2018; Christ, Lauterbach and Latz, 2019). The gastro-intestinal tract is the primary site of adaptation to dietary challenge, due to its roles in nutrient absorption, immunity and metabolism (Enriquez *et al.*, 2022). Dietary challenges or other environmental or genetic factors can lead to prolonged inflammation and eventually damage the gastro-intestinal tract (Huang *et al.*, 2017; Enriquez *et al.*, 2022). IBD encompasses several chronic inflammatory gut disorders, including ulcerative colitis (UC) and Crohn's disease (Chang, 2020; Adolph *et al.*, 2022). Patients with IBD, have a higher risk of developing colorectal cancer (CRC), one of the most lethal cancers (Kim and Chang, 2014; Shah and Itzkowitz, 2022). The incidence of IBD has increased worldwide (Alatab *et al.*, 2020; Freeman *et al.*, 2021) during the last decade, in part due to increased consumption of lipid-rich diets (Maconi *et al.*, 2010; Hou, Abraham and El-Serag, 2011). Furthermore, mouse studies show that HFD leads to more inflammation in the dextran sulfate sodium (DSS)-induced UC models compared to normal diet (Zhao *et al.*, 2020). However, the response to HFD is variable across individuals (Zeevi *et al.*, 2015) and the association between the lipid-rich diet and the risk of IBD in clinical studies is inconclusive (Kreuter *et al.*, 2019), possibly due to genetic factors underlying inter-individual variability in gut inflammation and dysbiosis (Baumgart and Sandborn, 2012). More than 200 risk genes associated with IBD were identified through human genome-wide association studies (GWAS) (Huang *et al.*, 2017), which have implicated epithelial function, microbe sensing and restriction, and adaptive immune response as drivers (Graham and Xavier, 2020; Kong *et al.*, 2023). However, there is still no effective treatment for IBD. Current therapies, such as anti-tumor necrosis factor alpha (TNF-α) antibodies (Rutgeerts *et al.*, 2005) and integrin α4β7 antibodies,

blocking leukocyte migration (Feagan *et al.*, 2013), can temporarily alleviate inflammation in a subset of patients (Rutgeerts *et al.*, 2005; Feagan *et al.*, 2013) but cause adverse effects (Harbord *et al.*, 2017) and fail to prevent relapses (Doherty *et al.*, 2018). Therefore, it is important to understand the gene-by-environment (GxE) interactions underpinning pre-clinical gut inflammation that eventually evolves into IBD, to aid in designing novel preventive and therapeutic strategies for intestinal inflammatory disorders.

Heterogeneity in clinical presentations as well as diversity in diet and lifestyle among human IBD patients render human genetic studies challenging (Molodecky *et al.*, 2011). Experiments in laboratory mice allow to control several environmental factors, such as temperature and diet, when exploring the genetic modulators of IBD and also enable the collection of several relevant tissues to help elucidate tissue-specific mechanisms (Nadeau and Auwerx, 2019; Li and Auwerx, 2020). In addition, to mirror the heterogeneity of human populations, genetically diverse populations, such as mouse genetic reference populations (GRPs), can be used in a systems genetics paradigm (Nadeau and Auwerx, 2019; Li and Auwerx, 2020). This not only allows the mapping of clinically relevant traits in controlled environments but also the characterization of intermediate molecular phenotypes from tissues that cannot easily be obtained in humans (Williams *et al.*, 2016; Li and Auwerx, 2020). For example, studies on the molecular basis of non-alcoholic fatty liver disease in the Collaborative Cross founder strains illustrated the importance of the genetic background in determining susceptibility to steatosis, hepatic inflammation and fibrosis (Benegiamo *et al.*, 2023). Moreover, the BXD GRP was used to identify genetic variants associated with metabolic phenotype variation, such as bile acid homeostasis (Li *et al.*, 2022), lipid metabolism in plasma (Jha, McDevitt, Halilbasic, *et al.*, 2018) and liver (Jha, McDevitt, Gupta, *et al.*, 2018) as well as mitochondrial dysregulation (Williams *et al.*, 2016), using GWAS or quantitative trait locus (QTL) mapping (Wu *et al.*,

2014; Williams *et al.*, 2016). Thus, large mouse GRPs are useful tools for identifying the tissue-specific mechanisms of complex diseases.

In order to decipher the genetic and environmental contributions to the development of intestinal inflammation, we measured the colon transcriptome of 52 BXD strains fed with CD or HFD (Williams *et al.*, 2016). HFD feeding from 8 to 29 weeks of age induced an IBD-like transcriptomic signature in colons of some, but not all, BXD strains, uncovering a subset of BXD strains that could be susceptible to HFD-induced IBD-like state. Gene co-expression analysis revealed two IBD-related modules in the colons of HFD-fed mice, one of which is likely under the control of a ModQTL. Through a systems genetics prioritization of genes under this ModQTL, we identified candidate IBD-related genes that we validated using GWAS in the UK Biobank (UKBB) for human IBD.

Results

HFD feeding leads to highly variable transcriptomic adaptations in the colon of BXD strains

For this study, we used an extensively characterized BXD mouse panel of 52 BXD strains fed with a chow diet (CD) or high-fat diet (HFD) from 8 to 29 weeks of age (Williams *et al.*, 2016; Jha, McDevitt, Gupta, *et al.*, 2018; Jha, McDevitt, Halilbasic, *et al.*, 2018), in which we mapped genetic determinants of metabolic traits in the liver (Williams *et al.*, 2016; Jha, McDevitt, Gupta, *et al.*, 2018) and plasma (Jha, McDevitt, Halilbasic, *et al.*, 2018). These mice underwent metabolic phenotyping, with many metabolic traits being altered by HFD (**Figure 1A**), and multiple organs were harvested and flash-frozen for future use (Williams *et al.*, 2016). Here, we focused on proximal colon samples from this population and performed microarray-based transcriptome analysis of this tissue (**Figure 1A**).

Principal component analysis (PCA) of all transcriptomes (**Figure 1—figure supplement 1A**) showed that the first principal component (PC1) separated mice by diet, indicative of a global diet effect in the population. Nevertheless, transcriptomes of several strains (such as BXD12, BXD84 and BXD81) on HFD had very similar PC1 values to their CD counterparts (**Figure 1—figure supplement 1B**), suggesting that they were resistant to dietary changes. Similarly, BXD strains did not cluster completely by diet based on hierarchical clustering analysis indicating that the genetic differences can override the impact of diet on the transcriptome in the colon (**Figure 1—figure supplement 1C**). To obtain a global, strain-independent, view of the HFD effect, we performed a differential expression analysis and identified 115 up- and 295 down-regulated differentially expressed genes (DEGs, absolute Log_2(Fold Change) > 0.5 and Benjamini-Hochberg (BH)-adjusted P value < 0.05, **Figure 1B**). Of note, *Cldn4*, one of claudins implicated in intestinal permeability (Ahmad *et al.*, 2017), was significantly down-regulated and serum amyloid A (*Saa1* and *Saa3*), which have been involved in the inflammatory response (Ye and Sun, 2015; Tannock *et al.*, 2018), were up-regulated upon HFD

(**Figure 1B**). Furthermore, gene set enrichment analysis (GSEA) showed an upregulation of inflammation, cell proliferation and translation, mitochondrial respiration, and stress response-related pathways upon HFD, while genes involved in the intermediate filament - that contribute to maintaining intestinal barriers (Misiorek *et al.*, 2016; Mun, Hur and Ku, 2022) - were down-regulated (**Figure 1C**). All in all, the transcriptome data are consistent with an HFD-induced downregulation of components of the intestinal barrier, enhanced permeability, induction of the unfolded protein response (UPR) and increased inflammation in BXD colons, much like HFD does in humans (Bischoff *et al.*, 2014). However, as in humans, not every strain exhibited the same response to dietary challenges. GSEA analyses applied individually to the diet effect in each strain showed a high degree of diversity in the inflammatory response (**Figure 1—figure supplement 1D**). For example, BXD44, 45 and 55, highlighted in red, were the 3 most susceptible strains to gut inflammation upon HFD, whereas BXD1, 67, and 85, colored in green, showed no significant enrichment in gut inflammation. This diversity in responses provided the basis for a systems genetics investigation of HFD-driven gut inflammation determinants in the BXD.

The transcriptomic response to HFD of a subset of BXD strains resembles DSS-induced ulcerative colitis (UC)

IBD is characterized by increasing inflammation in the gastro-intestinal tract (Adolph *et al.*, 2022). To investigate the disease relevance of the chronic inflammation seen in BXD colons upon HFD, we extracted the transcriptomic signatures from DSS-induced mouse UC models (Czarnewski *et al.*, 2019) and two IBD human studies (GSE16879 (Arijs *et al.*, 2009) and GSE83687 (Peters *et al.*, 2017), **Materials and Methods**) and used these signatures as custom gene sets in GSEA on the global HFD effect. DSS is widely used to induce UC in mouse models and disease severity increases over time (Czarnewski *et al.*, 2019). GSEA analyses showed that DSS-induced genes from days 4 (early inflammatory phase), 6 and 7 (acute

inflammatory phase) were significantly enriched in genes upregulated by HFD, especially the dysregulated genes in the later stage of DSS-induced UC (**Figure 1D, bottom panel**). Similarly, genes involved in human IBD (UC and Crohn's disease (CDs)) were also enriched in those same genes (**Figure 1D, top panel**). The same trend was observed for downregulated genes in mouse and human IBD, which were negatively enriched (**Figure 1D**), illustrating that HFD induced an IBD-like transcriptomic signature in BXD colons.

While the average response across all BXDs shared features of mouse and human IBD, we assessed the strain-specificity of this response by measuring each strain's response to IBD using GSEA (**Figure 2A**). Hierarchical clustering of the normalized enrichment scores (NES) in mouse IBD datasets classified the BXDs into three groups: susceptible strains highlighted in red (19 strains), intermediate strains represented in blue (11 strains), and resistant strains colored in green (17 strains) (**Figure 2A, top panel**). Of note, in line with colon histological lesions comparison of DSS-induced colitis mouse models in the literature (Mähler *et al.*, 1998), the C57BL/6J strain, one of the parental strains of the BXDs, was classified as one of the susceptible strains while the other parental strain DBA/2J belonged to the resistant group (**Figure 2A, top panel**), suggesting that genetic determinants inherited from the parental strains may determine the susceptibility of BXD strains to HFD-induced IBD-like inflammation in the colons.

To establish the functional relevance of this transcriptome-based classification on systemic inflammation, we compared plasma cytokine levels of these three groups under HFD (Williams *et al.*, 2016). Interestingly, the susceptible group have significantly lower levels of the anti-inflammatory cytokine - Interleukin (IL)-10 (**Figure 2B,** two-tailed t-test $p < 0.01$) and increased the proinflammatory cytokine - IL-15 (**Figure 2C,** two-tailed t-test $p < 0.0001$) compared to the resistant strains. *IL10* itself has been identified as an IBD-related candidate gene using GWAS in humans (Franke *et al.*, 2008) and IL-10-deficient mice are also well-

known mouse model for IBD research (Keubler *et al.*, 2015). IL-15 is another important cytokine involved in intestinal inflammation and is elevated in the human guts with IBD (Liu *et al.*, 2000). IL-15 knock-out mice are also reported to have less severe symptoms, such as weight loss and histological scores, following DSS administration (Yoshihara *et al.*, 2006). In summary, susceptibility to HFD-induced IBD-like inflammation in the colon, as assessed by changes in levels of genes associated with IBD, correlates with markers of the general inflammatory status of mice.

Identifying IBD-related gene modules in BXD colons

Since different BXD strains seem to exhibit different susceptibility to IBD, we set out to explore gene expression signatures underlying these differences. For that, we used Weighted Gene Co-expression Analysis (WGCNA) to construct CD- and HFD-specific gene co-expression networks to identify modules of co-expressed genes (**Figure 3A, Appendix 1 - Table 1**). Disease-associated modules were then defined as modules under HFD are significantly enriched in mouse DSS-induced UC signatures by an over-representation analysis (ORA, BH-adjusted P value < 0.05 and number of enriched genes > 5, **Figure 3A**). The HFD co-expression network consisted of 39 modules ranging in size from 34 to 1,853 genes and containing a total of 14,723 genes (**Appendix 1 - Table 1**). We visualized this network using Uniform Manifold Approximation and Projection (UMAP) (**Figure 3B**), reflecting that the majority modules were closely connected in the co-expression network.

Enrichment analyses indicated that modules HFD_M9 (484 genes), HFD_M16 (328 genes), and HFD_M28 (123 genes) were enriched with genes that are upregulated by DSS-induced colitis, while HFD_M15 (368 genes), HFD_M24 (159 genes), and HFD_M26 (135 genes) were significantly enriched with downregulated genes **(Figure 3C).** Of note, more than 20% of genes involved in HFD_M9 and HFD_M28 were part of the dysregulated genes of the acute

phase of mouse UC (day6 and day7) (**Figure 3C**). Interestingly, genes perturbed during IBD pathogenesis in humans were also enriched in HFD_M9 and HFD_M28 (**Figure 3C**).

While IBD-related genes were predominantly found in HFD modules, we also found that two modules, CD_M28 (185 genes) and CD_M32 (142 genes), in CD-fed mouse colons were associated with IBD (**Figure 3—figure supplement 1A)**. These two-modules significantly overlapped with the IBD-related HFD_M9 and HFD_M28 modules, respectively (BH-adjusted P value < 0.05) (**Figure 3—figure supplement 1B)**. Moreover, the molecular signatures underlying human UC and Crohn's disease were also clustered in these two modules (CD_M28 and CD_M32) under CD (**Figure 3—figure supplement 1C**). Collectively, the co-expression and enrichment analyses identify HFD_M9 and HFD_M28 as IBD-related modules on which we focus our subsequent investigation.

Identifying biological roles and transcriptional regulation of the IBD-related modules

To identify the biological function of the IBD-related modules, we performed enrichment analyses using the Hallmark database and the cell-type gene signatures (Kong *et al.*, 2023) (**Materials and Methods**). Genes in HFD_M9 were enriched in KRAS signaling and inflammation-related pathways, while HFD_M28 was enriched in IFN-α/γ responses (BH-adjusted P value < 0.05) (**Figure 4A**). Both modules were enriched in IFN-γ response genes (**Figure 4A**). IFN-γ is an essential cytokine for innate and adaptive intestinal immune responses (Brasseit *et al.*, 2018). It has been reported to play a key role in mouse (Ito *et al.*, 2006) and human (Tilg *et al.*, 2002) IBD pathogenesis, and was identified as a potential therapeutic target to alleviate inflammatory response in IBD (Li *et al.*, 2021). In addition, genes that are dysregulated in immune cells of Crohn's disease patients (Macrophages, B cell and immune cycling cells) were enriched in HFD_M9 (**Figure 4B**). In contrast, genes of HFD_M28 were not only enriched for genes that are dysregulated in immune cells, but also in intestinal epithelial cells of diseased individuals, such as Goblet and stem cells (**Figure 4B**). Overall,

HFD_M9 and HFD_M28 are both involved in inflammatory response, while genes involved in HFD_M28 also potentially influence intestinal epithelial barrier.

To identify transcriptional drivers of the two IBD-related modules, we performed a transcription factor (TF) enrichment analysis (**Materials and Methods**) and found that ZIC2, SMAD3, REL, FOSL1, and BATF are the top enriched transcription factors for the genes in HFD_M9 (**Figure 4C**), while the expression of genes in module HFD_M28 may be regulated by Interferon regulatory factors (IRFs, IRF1, IRF2, IRF7, and IRF9) and the signal transducer and activator of transcription families (STAT, STAT2) (**Figure 4D**). In fact, most of these TFs have been reported to be involved in gut inflammation. For example, *Smad3* mutant mice were more susceptible to intestinal inflammation (Yang *et al.*, 1999). Moreover, the IFN-STAT axis is essential to initiate the type-I IFN induction that is critical for human immune defense, such as IBD diseases (Stolzer *et al.*, 2021) and primary immunodeficiency diseases (Mogensen, 2019) as well as for disease tolerance (Mottis *et al.*, 2022). Collectively, we have identified TFs that likely control the expression of the two IBD-related modules to play an essential role in gut inflammation regulation.

Identifying ModQTLs for IBD-related modules and filtering of candidate genes

To analyze how the genotype impacts the IBD-like inflammatory response associated to HFD, we performed module QTL mapping analysis (ModQTL) for both IBD-related modules (HFD_M9 and HFD_28) (**Figure 5A**). We found a suggestive QTL for HFD_M28 (P value < 0.1), on chromosome 16 containing 552 protein-coding genes (**Figure 5A, Appendix 1 - Table 2**). The ModQTL analysis was also performed on the modules that are significantly enriched in IBD-downregulated genes (HFD_M15, HFD_M24, and HFD_M26), but no significant or suggestive QTLs were detected. Therefore, we focused on the QTL for IBD-induced genes in HFD_M28 and annotated its candidate genes based on three criteria (**Figure 5B**): (1) presence of high-impact genetic variants (such as missense and frameshift variants) in BXDs, (2)

association with inflammation based on literature mining (**Materials and methods**), (3) presence of *cis*-expression QTLs (eQTLs), that is, whether the expression of the gene is controlled by the QTL. The 27 genes satisfying at least two of the above criteria were considered as candidate genes driving the expression of module HFD_M28 (**Figure 5C**).

To further prioritize candidate genes regulating module HFD_M28, we applied GWAS to detect Crohn's disease- and UC-associated genetic variants using whole genome sequence (WGS) dataset in UKBB (**Figure 5C, Materials and Methods**). Interestingly, the genetic variants of two genes under the QTL peak, i.e, *ephrin type A receptor 6* (*EPHA6,* P value = 2.3E-06) (**Figure 5C**, **Figure 5—figure supplement 1A, Appendix 1 - Table 3**) and *Mucin 4* (*MUC4,* P value = 1.2E-06) (**Figure 5C**, **Figure 5—figure supplement 1B, Appendix 1 - Table 4)** were also associated with UC in humans. *EPHA6* belongs to Eph/Ephrin Signaling and this pathway has been associated with gut inflammation (Coulthard *et al.*, 2012) and proposed as a potential target to alleviate the inflammatory response in IBD (Grandi *et al.*, 2019), but the association between *EPHA6* and IBD is not explored yet. The Gene-Module Association Determination (G-MAD) (Li *et al.*, 2019) (https://systems-genetics.org/gmad) also revealed that expression of *Epha6* in mouse gastro-intestinal tract correlates with genes involved in inflammation-related pathways, such as IL-6 production and regulation of inflammatory response (**Figure 5D, Appendix 1 - Table 5**). *MUC4* is a transmembrane mucin (Gao *et al.*, 2021) and highly expressed in gastro-intestinal tract according to the human protein atlas (Uhlén *et al.*, 2015) (https://www.proteinatlas.org/humanproteome/tissue/intestine) (**Figure 5C**). The expression of *MUC4* in the human gastro-intestinal tract correlates with genes that are enriched for CRC and O-linked glycosylation based on G-MAD (Li *et al.*, 2019) (**Figure 5E, Appendix 1 - Table 5**). O-linked glycans are expressed by the intestinal epithelium to maintain barrier function, especially mucin type O-glycans, and gut disorders can be affected by dysfunction of O-linked glycosylation (Brazil and Parkos, 2022). Moreover,

276 *MUC4* is upregulated in enterocytes and Goblet cells in colons of Crohn's disease patients
277 (**Figure 5F)**. *MUC4* hence is a strong candidate because of its role in maintaining the intestinal
278 epithelium and controlling the gut inflammatory response (McGuckin *et al.*, 2011) and *EPHA6*
279 might be a novel candidate gene to impact gut inflammation. Based on the results of our QTL
280 mapping, human GWAS in UKBB, and existing literature, we hypothesize that *MUC4* and
281 *EPHA6* impact on colon integrity and inflammation and may be important players in gut
282 inflammation or IBD triggered by an unhealthy, lipid-rich diet.

283 However, it is unclear through what mechanisms the genetic variants in the candidate genes
284 affect IBD susceptibility. One possibility is that genetic variation leads to altered levels of
285 expression of the gene, ultimately affecting disease susceptibility. To test this possibility, we
286 examined the GTEx resource (GTEx Consortium, 2013) and found that *MUC4*, but not *EPHA6*,
287 has cis-eQTLs in the sigmoid and transverse colon. To establish likely causal links with IBD
288 incidence, we used these associations as instruments in a two-sample Mendelian randomization
289 (MR) (Hemani, Tilling and Smith, 2017; Hemani *et al.*, 2018) analysis. Using publicly
290 available GWAS summary statistics for IBD, Crohn's disease, and ulcerative colitis (Liu *et al.*,
291 2015; Elsworth *et al.*, 2020) as outcomes, we found suggestive evidence that increased
292 expression of *MUC4* in the sigmoid, but not transverse, colon may increase the risk of IBD
293 (nominal P value = 0.033, **Appendix 1 - Table 6**). No eQTLs were reported for *EPHA6* in the
294 colon, precluding us from investigating the potential consequences of changes in its expression
295 in these tissues.
296

Discussion

Dietary, environmental and genetic factors have all been reported to influence intestinal inflammation (Adolph *et al.*, 2022). Indeed, HFD can impair the intestinal epithelial barrier and trigger pre-clinical inflammation in the gastro-intestinal tract, eventually leading to inflammatory disorders of the gut (Enriquez *et al.*, 2022). In addition, genetic factors identified by GWAS can also predispose to IBD. For example, the interleukin-1 and -7 receptors (*IL-1R2* and *IL-7R*) were identified as candidate genes that regulate the immune response in IBD (Khor, Gardet and Xavier, 2011). However, the heterogeneity of diet and other environmental factors in human studies limits our ability to identify GxE interactions and pinpoint the genes and pathways involved in diet-induced gut inflammation. Studies in model organisms such as the mouse, where the environment can be carefully controlled, provide a valuable complement to human genetics studies that by nature are mainly observational (Nadeau and Auwerx, 2019; Li and Auwerx, 2020). Unfortunately, most mouse studies only evaluate mice from a single genetic background, limiting their generalizability and translatability to humans (Nadeau and Auwerx, 2019; Li and Auwerx, 2020). Conversely, GRPs such as the BXDs can mimic at least in part the heterogeneity of human populations and allow us to estimate the effect of GxE interactions on complex diseases (Jha, McDevitt, Gupta, *et al.*, 2018; Li and Auwerx, 2020).

Here, we utilized a panel of 52 BXD genetically diverse mouse strains fed with either HFD or CD to explore the genetic and dietary modulators of inflammation seen in the colon transcriptomes using systems genetics approaches. The colon transcriptomic response to HFD in this mouse population recapitulated several of the general features observed in DSS-induced UC mouse models and human IBD patients. In particular, we identified the upregulation of inflammation-related genes and the UPR as well as the downregulation of intercellular adhesion-related genesets as common signatures induced by HFD (Kreuter *et al.*, 2019). Moreover, our dataset not only was informative about the transcript changes of IBD at the

322 population level, but also unveiled extensive strain-specific effects that allowed us to classify
323 strains based on their propensity to develop IBD-like signatures. The fact that these
324 susceptibility groups also differed in anti- and pro-inflammatory plasma cytokine levels (IL-
325 10 and IL-15, respectively) suggests a relation between these tissue-specific transcriptional
326 signatures and systemic low-grade inflammation. Since gene interactions determine cellular
327 processes and the molecular functions of correlated genes are often similar (Nayak *et al.*, 2009),
328 we attempted to elucidate the mechanism underlying the diversity of IBD-like signatures and
329 chronic inflammation in BXD colons using gene co-expression analyses. This led us to identify
330 two IBD-related gene modules (HFD_M9 and HFD_M28).

331 As most differentially expressed genes are likely to be driven by and not be a cause of disease
332 (Porcu *et al.*, 2021), we attempted to understand whether the signatures in the colon are causes
333 or consequences of chronic inflammation. A first step was to characterize possible
334 transcriptional and genetic regulators of IBD-related modules. Enrichment analyses showed
335 that both IBD-associated modules largely consisted of immune response-related genes.
336 Specifically, genes involved in HFD_M9 and HFD_M28 are both differentially expressed in
337 immune cells in inflamed tissues of Crohn's disease patients (Kong *et al.*, 2023). Moreover,
338 the HFD_M28 module was enriched for TF motifs of STAT2 and IRF family, and HFD_M9
339 for SMAD3 and REL, which were illustrated to control the expression of these gut
340 inflammation-related genes, and influence the inflammatory response triggered by HFD in the
341 colon.

342 While we found IBD-related gene modules and the TFs driving their expression, the genetic
343 drivers of the diversity of gut inflammatory responses observed across the BXDs remained
344 elusive. To find candidate genes causing gut inflammation upon HFD, we then performed
345 Module QTL (ModQTL) analysis and allocated a suggestive ModQTL that may be controlling
346 one of IBD-related module (HFD_M28) under HFD. Importantly, through our prioritization

scheme for the genes under the ModQTL, we identify two plausible candidates, *Epha6* and *Muc4*, that have high-impact variants in the BXDs, are related to inflammation, and harbor variants in humans that are associated with IBD based on UKBB GWAS result. Mendelian randomization analysis suggests that higher expression of *MUC4* in the sigmoid colon may increase the risk of IBD. Furthermore, Muc4 knock-out mice have been shown to be more resistant to DSS-induced UC through upregulating the expression of *Muc2* (mucin secretion) and *Muc3* (transmembrane mucin) (Das *et al.*, 2016). A GWAS study also indicated that mutations in *EPHA6* increase risk for CRC (Guda *et al.*, 2015), but its potential association with IBD is a new finding. Therefore, these results point to important potential roles of *Muc4* and *Epha6* in gut chronic inflammation leading to inflammatory gut disorders.

Although studies in the BXD cohort are limited to variants present in the parental strains, C57BL/6J and DBA/2J, our analysis nevertheless shows how genetic diversity in this population allows us to detect the genetic modulators of chronic intestinal inflammation, that are more difficult to identify in widely used IBD mouse models on a single genetic background. In support of the generalizability of our data, the identified candidate genes in our mouse models were also associated to human UC, demonstrating that chronic inflammation induced upon HFD feeding may indeed be a prelude to human UC.

In conclusion, our systems genetics investigation of the colon in a controlled GRP, complemented with human GWAS studies, enabled the prioritization of modulators of IBD susceptibility that were generalizable to the human situation and may have clinical value.

Materials and methods

Population handling

Mice were studied as previously described (Williams *et al.*, 2016) and multiple organs were harvested for further analysis. Briefly, in groups of 3-5 animals from the same strain and diet, in isolator cages with individual air filtration (500 cm2, GM500, Tecniplast) and provided water ad libitum. Mice were fed CD ad libitum until 8 weeks of age. From 8 weeks to 29 weeks, half of the cohort was fed ad libitum HFD and the rest continued to be fed a CD (**Figure 1A**). CD composition: 18% kCal fat, 24% kCal protein and 58% kCal of carbohydrates (Teklad Global 18% Protein Rodent Diet 2018 chow diet, Envigo, Indianapolis, USA). HFD composition: 60.3% kCal fat, 18.4% kCal protein and 27.3% kCal of carbohydrates (Teklad Custom Diet TD.06414, Envigo, Indianapolis, USA). All mice were fasted overnight (from 6pm to 9am) prior to euthanasia. All procedures were approved by the veterinary office of canton Vaud under animal experimentation license number VD2257. In this work, proximal colons were extracted from the bio-banked samples and we did not use any new animals.

Transcriptome of the proximal colon in BXDs

A ~1 cm portion of the proximal half part of the colon was excised following euthanasia, washed in PBS and immediately stored in liquid nitrogen. Approximately 5 animals of the same strain fed the same diet were pooled at equal mass concentration for further RNA extraction. Total RNA was extracted using Direct-zol (Zymo Research) including the DNase digestion step. 100ng of total RNA was amplified using the Ambion® WT Expression Kit from Life Technologies (part number 4411974) and 5,500ng of cDNA was fragmented and labeled using the Affymetrix WT terminal labeling kit (part number 900671) all following manufacturers protocols. Labeled cDNA was hybridized on an Affymetrix Clariom S Assay microarray platform (GPL23038) in ~16 hours of incubation, then washed and stained using

an Affymetrix 450 Fluidics Station according to Affymetrix protocols. Finally, arrays were scanned on Affymetrix GSC3000 7G Scanner. Microarray data preprocessing was performed using apt-probeset-summarize from the Array Power Tool (APT) suite (v2.11.3) with the gc-sst-rma-sketch standard method and resulting expression values were log-transformed. Microarray probes targeting polymorphic regions in the BXD population were ignored in the process. For probesets targeting a same transcript, only the probeset with the highest value was considered.

Differential gene expression analysis

General differences in mRNA expression profiles between diets was assessed using Principal Component Analysis (PCA). Differential expression of individual transcripts between diets was assessed using the limma R Bioconductor package (version 3.48.3) (Ritchie *et al.*, 2015). Briefly, statistical significance was assessed using an empirical Bayes method (eBayes function) with an additive linear model accounting for diet and strain effect and adjusted P values were calculated by the Benjamini-Hochberg (BH) approach. Transcripts showing BH-adjusted P value below 0.05 and absolute Log_2 (Fold Change) above 0.5 were considered significantly associated with the effect of the diet.

Gene set enrichment analysis (GSEA) and Over-representation analysis (ORA)

Gene sets used in GSEA and ORA consisted of two parts: (1) the gene sets from the GO, KEGG, Hallmark, and Reactome databases were retrieved through the msigdbr R package (version 7.2.1) (Liberzon *et al.*, 2011). (2) the gene signatures of mouse and human IBD were used as custom gene sets (**Table 1**).

GSEA was performed using clusterProfiler R package (version 3.10.1) (Yu *et al.*, 2012) based on the log_2(Fold Change) ranking using parameters (*nPerm* = 100000, *minGSSize* = 30,

maxGSSize = 5000, *pvalueCutoff* = 1). The gene sets with absolute NES higher than 1 and BH-adjusted P value lower than 0.05 were identified as the significantly enriched gene sets.

ORA analysis was also performed using clusterProfiler R package (version 3.10.1) (Yu *et al.*, 2012) using parameters (minGSSize = 30, maxGSSize = 800). The gene sets with adjusted P value calculated by BH lower than 0.05 were identified as the significantly enriched gene sets.

Weighted gene correlation network analysis (WGCNA)

We used *WGCNA* R package (v1.51) (Langfelder and Horvath, 2008) to construct co-expression networks under CD and HFD, respectively. Firstly, the correlations between all pairs of gene across all BXDs fed with CD or HFD were calculated by Pearson correlation. Then, a best soft-thresholding power of 4 and 3 was chosen using pickSoftThreshold function with parameters (*networkType* = "signed hybrid", *blockSize* = 25000, *corFnc* = "bicor") for CD and HFD datasets in BXD colons separately. According to the calculated correlation coefficients, a network was constructed using parameters (*networkType* = "signed hybrid", *minModuleSize* = 30, *reassignThreshold* = 1e-6, *mergeCutHeight* = 0.15, *maxBlockSize* = 25000). The constructed co-expression gene modules were assigned color names and the module eigengenes were also identified for further analyses. To detect the preserved CD-modules in the co-expression modules under HFD, we defined gene modules under CD as custom genesets and performed ORA on each HFD-modules.

Transcription factor (TF) enrichment analysis

We first constructed a lognormal background distribution using the sequences of + 5kb region around the transcription starting site (TSS) of all genes and then downloaded the mouse HOCOMOCO-v10 (Kulakovskiy *et al.*, 2018) motifs from R package motifDB to perform TF enrichment analyses using R package *PWMenrich*. The significantly enriched motifs (P value < 0.001) were selected and then ranked based on the percentage of enriched promoters.

Module Quantitative Trait Locus (ModQTL) mapping in the BXDs

We first downloaded genotype information of each BXD mice from GeneNetwork (https://gn1.genenetwork.org/webqtl/main.py?FormID=sharinginfo&GN_AccessionId=600) and generated the kinship matrix of BXD mice using the leave-one-chromosome-out (LOCO) method. We then used the eigengenes of each module as phenotype input to perform Module QTL (ModQTL) with the R package *qtl2* (version 0.28) (Broman *et al.*, 2019) and the threshold of each QTL mapping analysis was obtained from a permutation test with 10,000 repeats. The peaks of QTL were calculated by *find_peaks* function with parameter: *prob=0.95.*

The same methods were also applied to gene expression QTL mapping (eQTL) and the significance threshold of each gene was obtained from a permutation test with 1,000 repeats. The significant peaks overlapped with the location of their corresponding gene were identified as *cis*-eQTL.

Literature mining

To explore the inflammation related genes, we first used candidate gene names and keywords ("IBD", "inflammatory bowel disease", "Ulcerative colitis", "Inflammation", "Inflammatory", "Crohn's disease") to search the title or abstract of associated literature using R package easyPubMed (version 2.13). Then, the genes involved in inflammation were confirmed by manual curation.

Genome-wide association study (GWAS) in UKBB

The phenotype data of inflamed Ulcerative colitis (Data-Field 131629, n = 6,459) and Crohn's disease (Data-Field 131627, n = 3,358) were firstly downloaded from UKBB (Bycroft *et al.*, 2018). 200,030 individuals with whole genome sequencing (WGS) (Halldorsson *et al.*, 2022) in UK Biobank were selected and then the population of European descent (including with 1,173 patients with Crohn's disease and 2,295 patients with UC) was extracted for further

GWAS analyses. Control individuals (n = 143,194) were included based on the following criteria: (1) Individuals without non- inflamed colitis (Data-Field 131631), Crohn's disease, and UC. (2) Individuals not taking any IBD-related medicine (**Appendix 1 - Table 7**).

WGS data provided by UK Biobank and used for GWAS were processed starting from pVCF files. We used REGENIE step1 to estimate population structure and then REGENIE step2 were applied to test associations between phenotypes and genetic variants and also included the following covariates in our model: the first 10 genetic principal components, age, sex, age:sex interaction, Body Mass Index (BMI), and smoking status. All data preparation and GWAS steps were run on DNAnexus.

Mendelian randomization (MR) analysis

eQTLs in sigmoid colon and transverse colon were selected in and their effect sizes obtained from the GTEx Portal on 2023-03-28 (v8, https://www.gtexportal.org/home/datasets, dbGaP Accession phs000424.v8.p2) (GTEx Consortium, 2013). No eQTLs were found for *EPHA6* but 147 and 87 eQTLs were found for *MUC4* in the sigmoid colon and transverse colon, respectively.

GWAS summary statistics for outcomes of interest, namely inflammatory bowel disease, ulcerative colitis, and Crohn's disease, were obtained from the IEU OpenGWAS project (Elsworth *et al.*, 2020), using the ieugwasr package (v0.1.5, https://mrcieu.github.io/ieugwasr/). As multiple GWAS exist for these diseases, we selected a representative one for each, prioritizing those with greater sample sizes and more cases, while still providing enough genetic variants for a large overlap with GTEx data (> 1 M SNPs). The selected GWAS were inflammatory bowel disease (ieu-a-31) (Liu *et al.*, 2015), ulcerative colitis (ieu-a-32) (Liu *et al.*, 2015), and non-cancer illness code, self-reported: Crohn's disease (ukb-b-8210) (Elsworth *et al.*, 2020).

For each outcome, *MUC4* eQTLs which were present in the outcome GWAS were pruned for independence (more than 10 kbp away or $r^2 < 0.01$) using Plink v1.90b6.21 (Purcell *et al.*, 2007) with the GTEx LD reference panel. In most cases this resulted in only a single eQTL being retained, with the exception of the sigmoid colon - Crohn's disease combination which resulted in two independent eQTLs.

Mendelian randomization (MR) was performed using the TwoSampleMR R package (version 0.5.6) (Hemani, Tilling and Smith, 2017; Hemani *et al.*, 2018). In the case with more than one eQTL, we used inverse-variance weighted MR, otherwise the Wald ratio. Because the magnitude of the normalized effect sizes provided by GTEx have no direct biological interpretation (https://gtexportal.org/home/faq#interpretEffectSize), the resulting causal effect estimates do not have an associated unit and cannot be translated into direct biological consequences. The direction (sign) of the effect remains interpretable.

Acknowledgments

We thank the Schoonjans' and Auwerx's lab members for technical assistance and discussions and Giacomo von Alvensleben for providing the GWAS analysis pipeline in human UKBB. The work in the JA laboratory was supported by grants from the Ecole Polytechnique Fédérale de Lausanne (EPFL), the European Research Council (ERC-AdG-787702), the Swiss National Science Foundation (SNSF 31003A_179435) and the Global Research Laboratory (GRL) National Research Foundation of Korea (NRF 2017K1A1A2013124). XL was supported by the China Scholarship Council (201906050019).

Author contributions

The study was conceived by XL, MBS and JA. EW, MBS and AB performed laboratory experiments. Data analyses were carried out by XL, AB, AR, JS and JP. XL and JA wrote the original manuscript. XL, MBS, JDM, GB, AP, KS and JA reviewed and edited the manuscript with contributions from all co-authors.

Competing interests

Authors declare no conflict of interest related to the work reported.

Data availability

The data that support the findings are available upon request to the corresponding authors (MBS and JA). The microarray data are available under the GEO numbers GSE225791. To review this dataset, please use this link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225791 and the review token is

522 exwtuucwpdqlzsr. Methods, materials and external resources are included in the **Materials**

523 **and Methods**.

524

References

Adolph, T.E. *et al.* (2022) 'The metabolic nature of inflammatory bowel diseases', *Nature Reviews Gastroenterology & Hepatology*, 19(12), pp. 753–767. Available at: https://doi.org/10.1038/s41575-022-00658-y.

Ahmad, R. *et al.* (2017) 'Obesity-induces Organ and Tissue Specific Tight Junction Restructuring and Barrier Deregulation by Claudin Switching', *Scientific Reports*, 7(1), p. 5125. Available at: https://doi.org/10.1038/s41598-017-04989-8.

Alatab, S. *et al.* (2020) 'The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017', *The Lancet Gastroenterology & Hepatology*, 5(1), pp. 17–30. Available at: https://doi.org/10.1016/S2468-1253(19)30333-4.

Arijs, I. *et al.* (2009) 'Mucosal Gene Expression of Antimicrobial Peptides in Inflammatory Bowel Disease Before and After First Infliximab Treatment', *PLOS ONE*, 4(11), p. e7984. Available at: https://doi.org/10.1371/journal.pone.0007984.

Baumgart, D.C. and Sandborn, W.J. (2012) 'Crohn's disease', *The Lancet*, 380(9853), pp. 1590–1605. Available at: https://doi.org/10.1016/S0140-6736(12)60026-9.

Benegiamo, G. *et al.* (2023) 'The genetic background shapes the susceptibility to mitochondrial dysfunction and NASH progression', *Journal of Experimental Medicine*, 220(4), p. e20221738. Available at: https://doi.org/10.1084/jem.20221738.

Bischoff, S.C. *et al.* (2014) 'Intestinal permeability – a new target for disease prevention and therapy', *BMC Gastroenterology*, 14(1), p. 189. Available at: https://doi.org/10.1186/s12876-014-0189-7.

Brasseit, J. *et al.* (2018) 'Divergent Roles of Interferon-γ and Innate Lymphoid Cells in Innate and Adaptive Immune Cell-Mediated Intestinal Inflammation', *Frontiers in Immunology*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2018.00023 (Accessed: 24 February 2023).

Brazil, J.C. and Parkos, C.A. (2022) 'Finding the sweet spot: glycosylation mediated regulation of intestinal inflammation', *Mucosal Immunology*, 15(2), pp. 211–222. Available at: https://doi.org/10.1038/s41385-021-00466-8.

554 Broman, K.W. *et al.* (2019) 'R/qtl2: Software for Mapping Quantitative Trait Loci with High-
555 Dimensional Data and Multiparent Populations', *Genetics*, 211(2), pp. 495–502. Available at:
556 https://doi.org/10.1534/genetics.118.301595.

557 Bycroft, C. *et al.* (2018) 'The UK Biobank resource with deep phenotyping and genomic data',
558 *Nature*, 562(7726), pp. 203–209. Available at: https://doi.org/10.1038/s41586-018-0579-z.

559 Chang, J.T. (2020) 'Pathophysiology of Inflammatory Bowel Diseases', *New England Journal*
560 *of Medicine*, 383(27), pp. 2652–2664. Available at: https://doi.org/10.1056/NEJMra2002697.

561 Christ, A., Lauterbach, M. and Latz, E. (2019) 'Western Diet and the Immune System: An
562 Inflammatory Connection', *Immunity*, 51(5), pp. 794–811. Available at:
563 https://doi.org/10.1016/j.immuni.2019.09.020.

564 Coulthard, M.G. *et al.* (2012) 'Eph/Ephrin Signaling in Injury and Inflammation', *The*
565 *American Journal of Pathology*, 181(5), pp. 1493–1503. Available at:
566 https://doi.org/10.1016/j.ajpath.2012.06.043.

567 Czarnewski, P. *et al.* (2019) 'Conserved transcriptomic profile between mouse and human
568 colitis allows unsupervised patient stratification', *Nature Communications*, 10(1), p. 2892.
569 Available at: https://doi.org/10.1038/s41467-019-10769-x.

570 Das, S. *et al.* (2016) 'Mice deficient in Muc4 are resistant to experimental colitis and colitis-
571 associated colorectal cancer', *Oncogene*, 35(20), pp. 2645–2654. Available at:
572 https://doi.org/10.1038/onc.2015.327.

573 Doherty, G. *et al.* (2018) 'European Crohn's and Colitis Organisation Topical Review on
574 Treatment Withdrawal ["Exit Strategies"] in Inflammatory Bowel Disease', *Journal of*
575 *Crohn's and Colitis*, 12(1), pp. 17–31. Available at: https://doi.org/10.1093/ecco-jcc/jjx101.

576 Duan, Y. *et al.* (2018) 'Inflammatory Links Between High Fat Diets and Diseases', *Frontiers*
577 *in Immunology*, 9. Available at:
578 https://www.frontiersin.org/articles/10.3389/fimmu.2018.02649 (Accessed: 19 January 2023).

579 Elsworth, B. *et al.* (2020) 'The MRC IEU OpenGWAS data infrastructure'. bioRxiv, p.
580 2020.08.10.244293. Available at: https://doi.org/10.1101/2020.08.10.244293.

581 Enriquez, J.R. *et al.* (2022) 'A dietary change to a high-fat diet initiates a rapid adaptation of
582 the intestine', *Cell Reports*, 41(7). Available at: https://doi.org/10.1016/j.celrep.2022.111641.

583 Feagan, B.G. *et al.* (2013) 'Vedolizumab as Induction and Maintenance Therapy for Ulcerative
584 Colitis', *New England Journal of Medicine*, 369(8), pp. 699–710. Available at:
585 https://doi.org/10.1056/NEJMoa1215734.

586 Franke, A. *et al.* (2008) 'Sequence variants in IL10, ARPC2 and multiple other loci contribute
587 to ulcerative colitis susceptibility', *Nature Genetics*, 40(11), pp. 1319–1323. Available at:
588 https://doi.org/10.1038/ng.221.

589 Freeman, K. *et al.* (2021) 'The incidence and prevalence of inflammatory bowel disease in UK
590 primary care: a retrospective cohort study of the IQVIA Medical Research Database', *BMC*
591 *Gastroenterology*, 21(1), p. 139. Available at: https://doi.org/10.1186/s12876-021-01716-6.

592 Gao, X.-P. *et al.* (2021) 'Integrative Analysis of MUC4 to Prognosis and Immune Infiltration
593 in Pan-Cancer: Friend or Foe?', *Frontiers in Cell and Developmental Biology*, 9. Available at:
594 https://www.frontiersin.org/articles/10.3389/fcell.2021.695544 (Accessed: 10 February 2023).

595 Graham, D.B. and Xavier, R.J. (2020) 'Pathway paradigms revealed from the genetics of
596 inflammatory bowel disease', *Nature*, 578(7796), pp. 527–539. Available at:
597 https://doi.org/10.1038/s41586-020-2025-2.

598 Grandi, A. *et al.* (2019) 'Targeting the Eph/Ephrin System as Anti-Inflammatory Strategy in
599 IBD', *Frontiers in Pharmacology*, 10, p. 691. Available at:
600 https://doi.org/10.3389/fphar.2019.00691.

601 GTEx Consortium (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nature Genetics*,
602 45(6), pp. 580–585. Available at: https://doi.org/10.1038/ng.2653.

603 Guda, K. *et al.* (2015) 'Novel recurrently mutated genes in African American colon cancers',
604 *Proceedings of the National Academy of Sciences*, 112(4), pp. 1149–1154. Available at:
605 https://doi.org/10.1073/pnas.1417064112.

606 Halldorsson, B.V. *et al.* (2022) 'The sequences of 150,119 genomes in the UK Biobank',
607 *Nature*, 607(7920), pp. 732–740. Available at: https://doi.org/10.1038/s41586-022-04965-x.

608 Harbord, M. *et al.* (2017) 'Third European Evidence-based Consensus on Diagnosis and
609 Management of Ulcerative Colitis. Part 2: Current Management', *Journal of Crohn's and*
610 *Colitis*, 11(7), pp. 769–784. Available at: https://doi.org/10.1093/ecco-jcc/jjx009.

611 Hasegawa, Y. *et al.* (2020) 'Long-term effects of western diet consumption in male and female
612 mice', *Scientific Reports*, 10(1), p. 14686. Available at: https://doi.org/10.1038/s41598-020-
613 71592-9.

614 Hemani, G. *et al.* (2018) 'The MR-Base platform supports systematic causal inference across
615 the human phenome', *eLife*. Edited by R. Loos, 7, p. e34408. Available at:
616 https://doi.org/10.7554/eLife.34408.

617 Hemani, G., Tilling, K. and Smith, G.D. (2017) 'Orienting the causal relationship between
618 imprecisely measured traits using GWAS summary data', *PLOS Genetics*, 13(11), p. e1007081.
619 Available at: https://doi.org/10.1371/journal.pgen.1007081.

620 Hou, J.K., Abraham, B. and El-Serag, H. (2011) 'Dietary Intake and Risk of Developing
621 Inflammatory Bowel Disease: A Systematic Review of the Literature', *Official journal of the*
622 *American College of Gastroenterology | ACG*, 106(4), p. 563. Available at:
623 https://doi.org/10.1038/ajg.2011.44.

624 Huang, H. *et al.* (2017) 'Fine-mapping inflammatory bowel disease loci to single-variant
625 resolution', *Nature*, 547(7662), pp. 173–178. Available at:
626 https://doi.org/10.1038/nature22969.

627 Ito, R. *et al.* (2006) 'Interferon-gamma is causatively involved in experimental inflammatory
628 bowel disease in mice', *Clinical and Experimental Immunology*, 146(2), pp. 330–338.
629 Available at: https://doi.org/10.1111/j.1365-2249.2006.03214.x.

630 Jha, P., McDevitt, M.T., Halilbasic, E., *et al.* (2018) 'Genetic Regulation of Plasma Lipid
631 Species and Their Association with Metabolic Phenotypes', *Cell Systems*, 6(6), pp. 709-721.e6.
632 Available at: https://doi.org/10.1016/j.cels.2018.05.009.

633 Jha, P., McDevitt, M.T., Gupta, R., *et al.* (2018) 'Systems Analyses Reveal Physiological Roles
634 and Genetic Regulators of Liver Lipid Species', *Cell Systems*, 6(6), pp. 722-733.e6. Available
635 at: https://doi.org/10.1016/j.cels.2018.05.016.

636 Keubler, L.M. *et al.* (2015) 'A Multihit Model: Colitis Lessons from the Interleukin-10–
637 deficient Mouse', *Inflammatory Bowel Diseases*, 21(8), pp. 1967–1975. Available at:
638 https://doi.org/10.1097/MIB.0000000000000468.

639 Khor, B., Gardet, A. and Xavier, R.J. (2011) 'Genetics and pathogenesis of inflammatory
640 bowel disease', *Nature*, 474(7351), pp. 307–317. Available at:
641 https://doi.org/10.1038/nature10209.

642 Kim, E.R. and Chang, D.K. (2014) 'Colorectal cancer in inflammatory bowel disease: The risk,
643 pathogenesis, prevention and diagnosis', *World Journal of Gastroenterology*, 20(29), pp.
644 9872–9881. Available at: https://doi.org/10.3748/wjg.v20.i29.9872.

645 Kong, L. *et al.* (2023) 'The landscape of immune dysregulation in Crohn's disease revealed
646 through single-cell transcriptomic profiling in the ileum and colon', *Immunity*, 0(0). Available
647 at: https://doi.org/10.1016/j.immuni.2023.01.002.

648 Kreuter, R. *et al.* (2019) 'The role of obesity in inflammatory bowel disease', *Biochimica et*
649 *Biophysica Acta (BBA) - Molecular Basis of Disease*, 1865(1), pp. 63–72. Available at:
650 https://doi.org/10.1016/j.bbadis.2018.10.020.

651 Kulakovskiy, I.V. *et al.* (2018) 'HOCOMOCO: towards a complete collection of transcription
652 factor binding models for human and mouse via large-scale ChIP-Seq analysis', *Nucleic Acids*
653 *Research*, 46(Database issue), pp. D252–D259. Available at:
654 https://doi.org/10.1093/nar/gkx1106.

655 Langfelder, P. and Horvath, S. (2008) 'WGCNA: an R package for weighted correlation
656 network analysis', *BMC Bioinformatics*, 9(1), p. 559. Available at:
657 https://doi.org/10.1186/1471-2105-9-559.

658 Li, H. *et al.* (2019) 'Identifying gene function and module connections by the integration of
659 multispecies expression compendia', *Genome Research* [Preprint]. Available at:
660 https://doi.org/10.1101/gr.251983.119.

661 Li, H. *et al.* (2022) 'Integrative systems analysis identifies genetic and dietary modulators of
662 bile acid homeostasis', *Cell Metabolism*, 34(10), pp. 1594-1610.e4. Available at:
663 https://doi.org/10.1016/j.cmet.2022.08.015.

664 Li, H. and Auwerx, J. (2020) 'Mouse Systems Genetics as a Prelude to Precision Medicine',
665 *Trends in genetics: TIG*, 36(4), pp. 259–272. Available at:
666 https://doi.org/10.1016/j.tig.2020.01.004.

667 Li, J.-Y. *et al.* (2021) 'IRF/Type I IFN signaling serves as a valuable therapeutic target in the
668 pathogenesis of inflammatory bowel disease', *International Immunopharmacology*, 92, p.
669 107350. Available at: https://doi.org/10.1016/j.intimp.2020.107350.

670 Liberzon, A. *et al.* (2011) 'Molecular signatures database (MSigDB) 3.0', *Bioinformatics*,
671 27(12), pp. 1739–1740. Available at: https://doi.org/10.1093/bioinformatics/btr260.

672 Liu, J.Z. *et al.* (2015) 'Association analyses identify 38 susceptibility loci for inflammatory
673 bowel disease and highlight shared genetic risk across populations', *Nature Genetics*, 47(9),
674 pp. 979–986. Available at: https://doi.org/10.1038/ng.3359.

675 Liu, Z. *et al.* (2000) 'IL-15 is highly expressed in inflammatory bowel disease and regulates
676 local T cell-dependent cytokine production', *Journal of Immunology (Baltimore, Md.: 1950)*,
677 164(7), pp. 3608–3615. Available at: https://doi.org/10.4049/jimmunol.164.7.3608.

678 Lutsey, P.L., Steffen, L.M. and Stevens, J. (2008) 'Dietary Intake and the Development of the
679 Metabolic Syndrome', *Circulation*, 117(6), pp. 754–761. Available at:
680 https://doi.org/10.1161/CIRCULATIONAHA.107.716159.

681 Maconi, G. *et al.* (2010) 'Pre-illness changes in dietary habits and diet as a risk factor for
682 inflammatory bowel disease: A case-control study', *World Journal of Gastroenterology*,
683 16(34), pp. 4297–4304. Available at: https://doi.org/10.3748/wjg.v16.i34.4297.

684 Mähler, M. *et al.* (1998) 'Differential susceptibility of inbred mouse strains to dextran sulfate
685 sodium-induced colitis', *American Journal of Physiology-Gastrointestinal and Liver
686 Physiology*, 274(3), pp. G544–G551. Available at:
687 https://doi.org/10.1152/ajpgi.1998.274.3.G544.

688 Maurya, S.K. *et al.* (2023) 'Western Diet Causes Heart Failure With Reduced Ejection Fraction
689 and Metabolic Shifts After Diastolic Dysfunction and Novel Cardiac Lipid Derangements',
690 *JACC: Basic to Translational Science*, 0(0). Available at:
691 https://doi.org/10.1016/j.jacbts.2022.10.009.

McGuckin, M.A. *et al.* (2011) 'Mucin dynamics and enteric pathogens', *Nature Reviews Microbiology*, 9(4), pp. 265–278. Available at: https://doi.org/10.1038/nrmicro2538.

Misiorek, J.O. *et al.* (2016) 'Keratin 8-deletion induced colitis predisposes to murine colorectal cancer enforced by the inflammasome and IL-22 pathway', *Carcinogenesis*, 37(8), pp. 777–786. Available at: https://doi.org/10.1093/carcin/bgw063.

Mogensen, T.H. (2019) 'IRF and STAT Transcription Factors - From Basic Biology to Roles in Infection, Protective Immunity, and Primary Immunodeficiencies', *Frontiers in Immunology*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2018.03047 (Accessed: 19 January 2023).

Molodecky, N.A. *et al.* (2011) 'Challenges associated with identifying the environmental determinants of the inflammatory bowel diseases', *Inflammatory Bowel Diseases*, 17(8), pp. 1792–1799. Available at: https://doi.org/10.1002/ibd.21511.

Mottis, A. *et al.* (2022) 'Tetracycline-induced mitohormesis mediates disease tolerance against influenza', *The Journal of Clinical Investigation*, 132(17). Available at: https://doi.org/10.1172/JCI151540.

Mun, J., Hur, W. and Ku, N.-O. (2022) 'Roles of Keratins in Intestine', *International Journal of Molecular Sciences*, 23(14), p. 8051. Available at: https://doi.org/10.3390/ijms23148051.

Nadeau, J.H. and Auwerx, J. (2019) 'The virtuous cycle of human genetics and mouse models in drug discovery', *Nature Reviews. Drug Discovery*, 18(4), pp. 255–272. Available at: https://doi.org/10.1038/s41573-018-0009-9.

Nayak, R.R. *et al.* (2009) 'Coexpression network based on natural variation in human gene expression reveals gene interactions and functions', *Genome Research*, 19(11), pp. 1953–1962. Available at: https://doi.org/10.1101/gr.097600.109.

Peters, L.A. *et al.* (2017) 'A functional genomics predictive network model identifies regulators of inflammatory bowel disease', *Nature Genetics*, 49(10), pp. 1437–1449. Available at: https://doi.org/10.1038/ng.3947.

718 Porcu, E. *et al.* (2021) 'Differentially expressed genes reflect disease-induced rather than
719 disease-causing changes in the transcriptome', *Nature Communications*, 12(1), p. 5647.
720 Available at: https://doi.org/10.1038/s41467-021-25805-y.

721 Purcell, S. *et al.* (2007) 'PLINK: A Tool Set for Whole-Genome Association and Population-
722 Based Linkage Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575.
723 Available at: https://doi.org/10.1086/519795.

724 Ritchie, M.E. *et al.* (2015) 'limma powers differential expression analyses for RNA-
725 sequencing and microarray studies', *Nucleic Acids Research*, 43(7), p. e47. Available at:
726 https://doi.org/10.1093/nar/gkv007.

727 Rutgeerts, P. *et al.* (2005) 'Infliximab for Induction and Maintenance Therapy for Ulcerative
728 Colitis', *New England Journal of Medicine*, 353(23), pp. 2462–2476. Available at:
729 https://doi.org/10.1056/NEJMoa050516.

730 Shah, S.C. and Itzkowitz, S.H. (2022) 'Colorectal Cancer in Inflammatory Bowel Disease:
731 Mechanisms and Management', *Gastroenterology*, 162(3), pp. 715-730.e3. Available at:
732 https://doi.org/10.1053/j.gastro.2021.10.035.

733 Stolzer, I. *et al.* (2021) 'An IFN-STAT Axis Augments Tissue Damage and Inflammation in a
734 Mouse Model of Crohn's Disease', *Frontiers in Medicine*, 8. Available at:
735 https://www.frontiersin.org/articles/10.3389/fmed.2021.644244 (Accessed: 12 January 2023).

736 Tannock, L.R. *et al.* (2018) 'Serum amyloid A3 is a high density lipoprotein-associated acute-
737 phase protein', *Journal of Lipid Research*, 59(2), pp. 339–347. Available at:
738 https://doi.org/10.1194/jlr.M080887.

739 Tilg, H. *et al.* (2002) 'Treatment of Crohn's disease with recombinant human interleukin 10
740 induces the proinflammatory cytokine interferon γ', *Gut*, 50(2), pp. 191–195. Available at:
741 https://doi.org/10.1136/gut.50.2.191.

742 Uhlén, M. *et al.* (2015) 'Tissue-based map of the human proteome', *Science*, 347(6220), p.
743 1260419. Available at: https://doi.org/10.1126/science.1260419.

744 Williams, E.G. *et al.* (2016) 'Systems proteomics of liver mitochondria function', *Science*,
745 352(6291). Available at: https://doi.org/10.1126/science.aad0189.

746 Wu, Y. *et al.* (2014) 'Multilayered genetic and omics dissection of mitochondrial activity in a
747 mouse reference population', *Cell*, 158(6), pp. 1415–1430. Available at:
748 https://doi.org/10.1016/j.cell.2014.07.039.

749 Yang, X. *et al.* (1999) 'Targeted disruption of SMAD3 results in impaired mucosal immunity
750 and diminished T cell responsiveness to TGF-β', *The EMBO Journal*, 18(5), pp. 1280–1291.
751 Available at: https://doi.org/10.1093/emboj/18.5.1280.

752 Ye, R.D. and Sun, L. (2015) 'Emerging functions of serum amyloid A in inflammation',
753 *Journal of Leukocyte Biology*, 98(6), pp. 923–929. Available at:
754 https://doi.org/10.1189/jlb.3VMR0315-080R.

755 Yoshihara, K. *et al.* (2006) 'Role of interleukin 15 in colitis induced by dextran sulphate
756 sodium in mice', *Gut*, 55(3), pp. 334–341. Available at:
757 https://doi.org/10.1136/gut.2005.076000.

758 Yu, G. *et al.* (2012) 'clusterProfiler: an R Package for Comparing Biological Themes Among
759 Gene Clusters', *OMICS: A Journal of Integrative Biology*, 16(5), pp. 284–287. Available at:
760 https://doi.org/10.1089/omi.2011.0118.

761 Zeevi, D. *et al.* (2015) 'Personalized Nutrition by Prediction of Glycemic Responses', *Cell*,
762 163(5), pp. 1079–1094. Available at: https://doi.org/10.1016/j.cell.2015.11.001.

763 Zhao, D. *et al.* (2020) 'High-Fat Diet Promotes DSS-Induced Ulcerative Colitis by
764 Downregulated FXR Expression through the TGFB Pathway', *BioMed Research International*,
765 2020, p. e3516128. Available at: https://doi.org/10.1155/2020/3516128.

766
767

Figure Legends

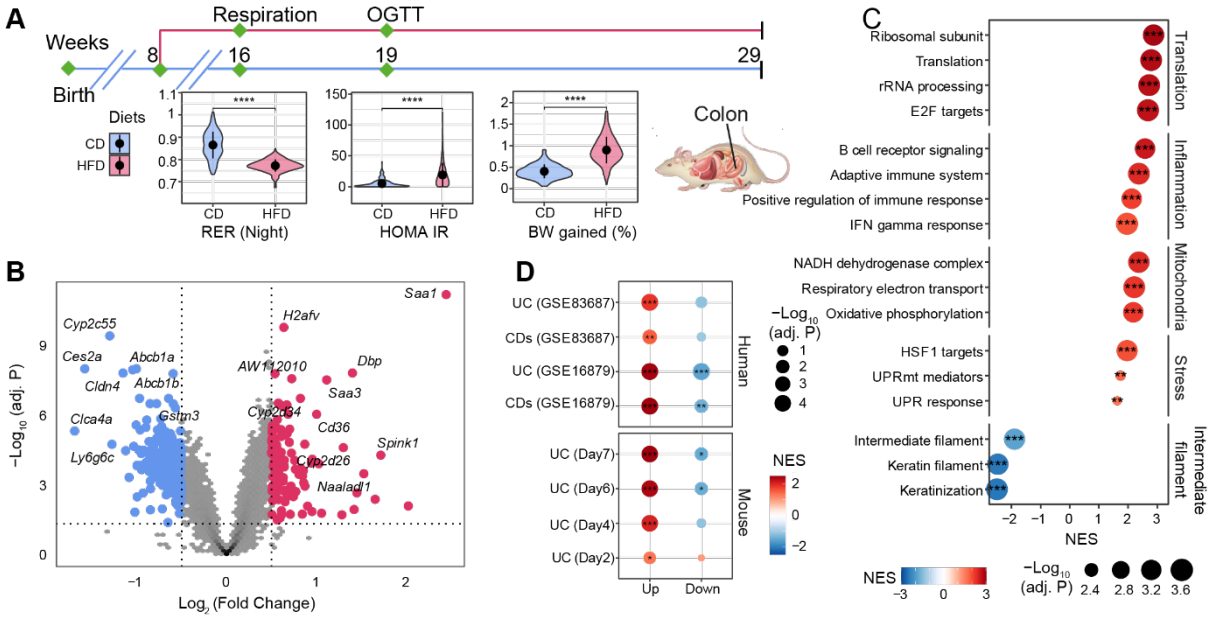


Figure 1. The effect of long-term HFD on gene expression in BXD colons. (**A**) Graphical representation of a pipeline of a previously described BXD mouse study (Williams *et al.*, 2016). Mice were fed HFD or CD starting from 8 weeks of age and metabolic phenotyping was performed as indicated. Mice were sacrificed at 29 weeks of age and multiple organs were collected and frozen for further analyses. BXD colon transcriptomes were analyzed in this study. CD: chow diet indicated in blue, HFD: high-fat diet indicated in red. P values were calculated by two-tailed Student's t-test and indicated as follows: * P <0.05; ** P <0.01; *** P <0.001; **** P <0.0001. (**B**) Volcano plot showing the HFD effect on BXD colon transcriptomes compared to CD and the up- and down-regulated differentially expressed genes (DEGs, absolute Log_2(Fold Change) > 0.5 and BH-adjusted P value (adj. P) < 0.05) were highlighted in red and blue, respectively. (**C**) Gene set enrichment analysis showing the effect of HFD on gene expression in BXD colons. Gene sets were grouped into five categories: Translation, Inflammation, Mitochondria, Stress, and Intermediate filament. Normalized enrichment scores (NES) were represented by color and -Log_{10} (BH-adj. P) were represented by dot size and indicated as follows: * Adjusted P value <0.05; ** Adjusted P value <0.01; *** Adjusted P value <0.001. (**D**) Enrichment analysis of molecular signatures of mouse and human IBD on the transcriptome of BXD colons. UC: Ulcerative colitis, CDs: Crohn's disease

The following figure supplement is available for figure 1:

Figure supplement 1. Transcriptome profile in BXD colons.

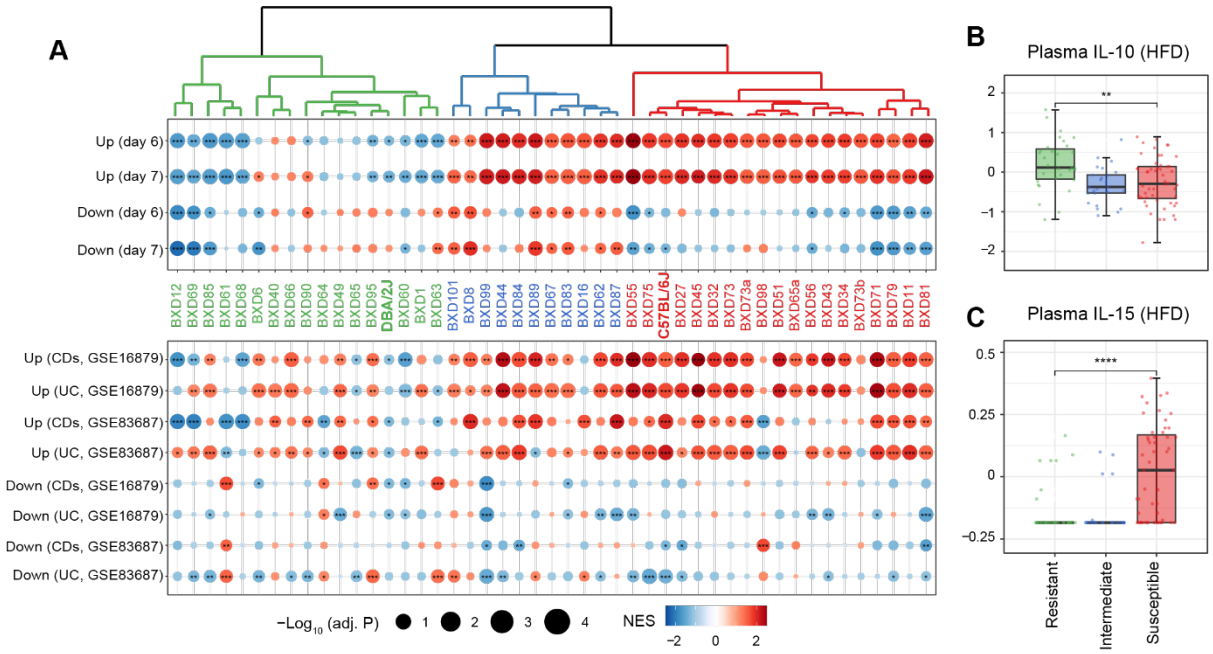
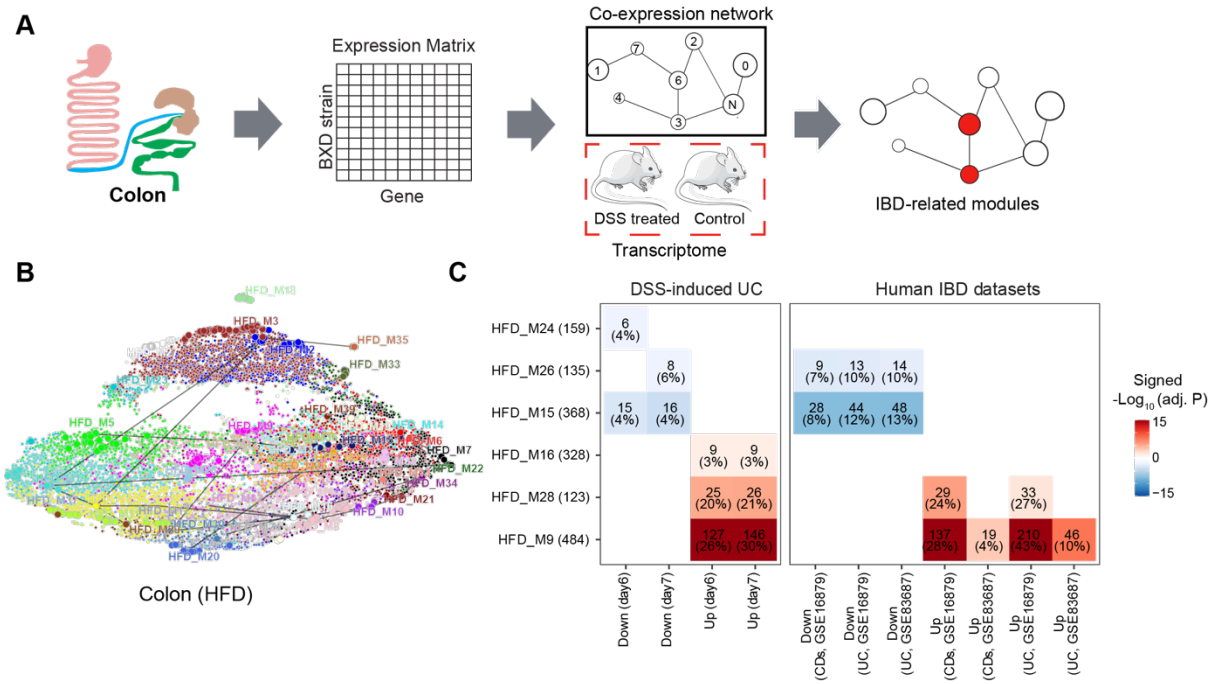


Figure 2. Identifying susceptible strains to HFD-induced IBD-like inflammation and their effect on plasma cytokines. (**A**) Enrichment analysis of the molecular signatures in human (**bottom panel**) and mouse IBD (**top panel**) models on the gene expression of individual BXD colons. Dendrogram showing the propensity to develop a mouse IBD-like signature among BXD strains upon HFD. BXD strains were divided into three clusters: susceptible (in red), intermediate (in blue), and resistant strains (in green). Normalized enrichment scores (NES) were represented by color and -Log_{10}(BH-adjusted P values) were represented by dot size and indicated as follows: * Adjusted P value <0.05; ** Adjusted P value <0.01; *** Adjusted P value <0.001. UC: Ulcerative colitis, CDs: Crohn's disease (**B, C**) Boxplots showing the effect of susceptible strains on plasma IL-10 (**B**) and IL-15 (**C**) level compared to resistant strains. P values were calculated by two-tailed Student's t-test and indicated as follows: * P <0.05; ** P <0.01; *** P <0.001; **** P <0.0001.



804

805 **Figure 3. Identifying IBD-related gene modules.** (**A**) Pipeline for exploring the IBD-
806 associated gene modules. A co-expression gene network was constructed based on the
807 transcriptome of BXD colons under HFD. IBD-associated modules were then defined as gene
808 modules under HFD that are significantly clustered in mouse DSS-induced UC signatures. (**B**)
809 UMAP representation of the co-expression gene network under HFD. 39 co-expression
810 modules are represented in the corresponding color and the correlated modules (Spearman
811 correlation coefficient between the eigengene of modules > 0.7) were linked by a grey line. (**C**)
812 Heatmap showing the enrichment of co-expression modules in mouse and human IBD gene
813 signatures and the number and percentage of enriched genes were labeled. The number of
814 enriched genes divided by the number of genes involved in the respective module was defined
815 as the percentage of enriched genes. Signed $-Log_{10}$(adj. P) indicated BH-adjusted P value and
816 the enriched gene set. Enriched modules of up- and down-regulated genes upon disease were
817 highlighted in red and blue, respectively. UC: Ulcerative colitis, CDs: Crohn's disease.

818 The following figure supplement is available for figure 3:

819 Figure supplement 1. Exploring IBD-associated co-expression modules under CD.

820

821

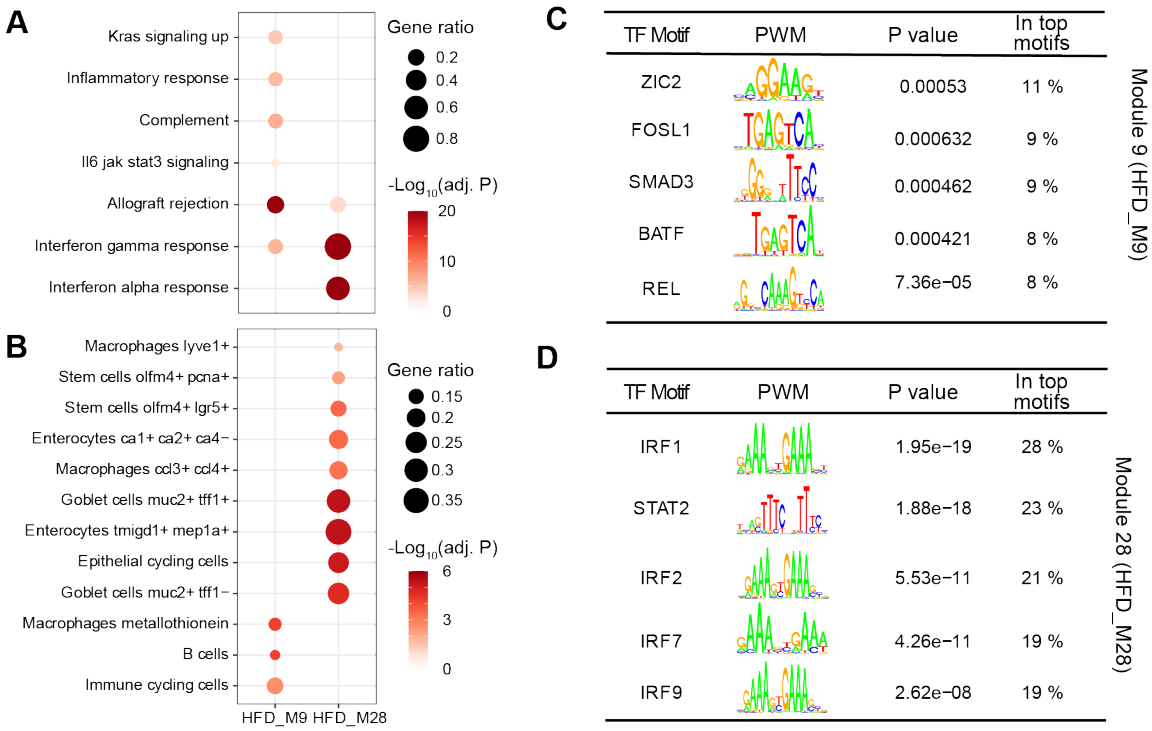
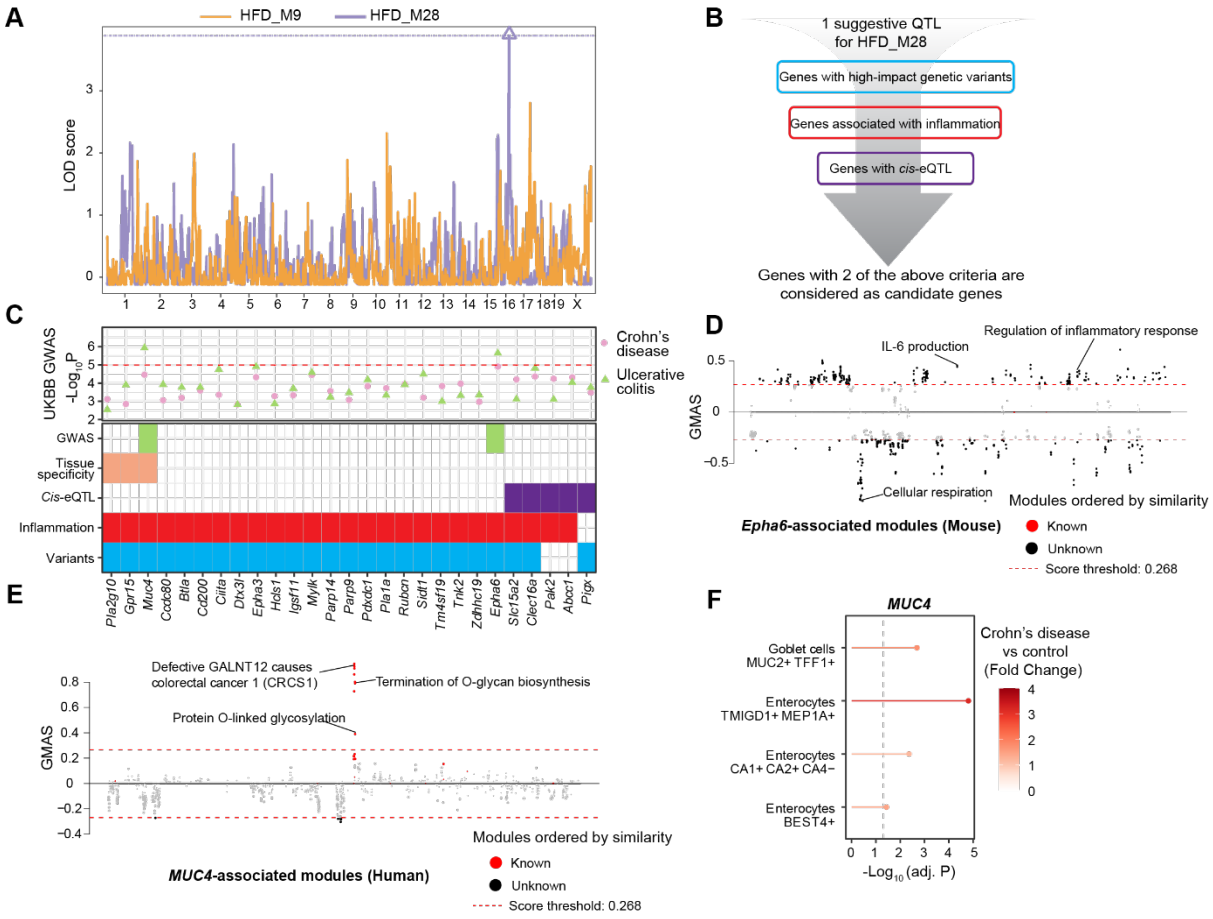


Figure 4. Biological interrogation of identified IBD-related modules. (**A**, **B**) Dot plots showing the enrichment of IBD-related modules in hallmark genesets (**A**) and cell-type gene signatures of inflamed colon in Crohn's disease patients (**B**). Gene ratios higher than 0.1 are shown and represented by dot size. Dots are colored by -log$_{10}$(BH-adjusted P values). (**C**, **D**) The enriched motifs for promoters of the genes involved in module HFD_M9 (**C**) and HFD_M28 (**D**). The significantly enriched motifs (P value < 0.001) were ranked based on the percentage of enriched promoters (In top motifs) and then the top five TFs were selected. TF: Transcription factor. PWM: Positional weight matrix.



832

833 **Figure 5. ModQTL mapping for two IBD-related modules and the prioritization of**
834 **candidate genes.** (**A**) Manhattan plot showing the ModQTL mapping result for disease-related
835 modules HFD_M9 and HFD_M28. ModQTL maps of HFD_M9 and HFD_M28 are indicated
836 in orange and purple, respectively. The threshold calculated by permutation test ($P < 0.1$) for
837 HFD_M28 is represented by a purple dashed line. (**B**) The filtering criteria for selecting
838 candidate genes under the ModQTL peak for HFD_M28. Genes with 2 of the described criteria
839 are considered as candidate genes. (**C**) The most significant associations between 27 candidate
840 genes under the ModQTL peak and Crohn's disease or UC identified through GWAS according
841 to whole genome sequence in the human UKBB are shown in the scatter plot (**top panel**).
842 Crohn's disease and UC are indicated by pink circle and green triangle, respectively. The
843 threshold ($-Log_{10}P$ value = 5) is represented by a red dashed line. Heatmap showing the
844 identified 27 candidate genes of module HFD_M28 (**bottom panel**). Variants colored in blue
845 indicate genes with high-impact genetic variants in BXD mice (including missense, frameshift,
846 initiator codon, splice donor, splice acceptor, in-frame deletion, in-frame insertion, stop lost,
847 stop gained). Inflammation is indicated in red and represents genes associated with
848 inflammation based on literature mining. *Cis*-eQTL colored in purple indicates genes with *Cis*-

eQTLs. Tissue specificity colored in orange means genes that are highly expressed in human intestine (data were downloaded from human protein atlas, https://www.proteinatlas.org/humanproteome/tissue/intestine (Uhlén *et al.*, 2015)). GWAS result of UC in UKBB colored in green indicates that genes are significantly associated with human UC. (**D**, **E**) Manhattan plots showing the associated gene expression modules of *Epha6* in mouse gastro-intestinal tract (**D**) and that of *MUC4* in human gastro-intestinal tract (**E**) (data from https://systems-genetics.org/gmad (Li *et al.*, 2019)). The threshold is represented by the red dashed line (absolute Gene-Module Association Score (GMAS) >= 0.268). Terms above the threshold are identified as the significant associated terms. GO terms or gene modules are ranked by similarity. Known associated terms are shown as red dots and new significant associated terms are colored in black. (**F**) Dot plot showing that the expression of *MUC4* was higher in four cell types of human inflamed colon with Crohn's disease.

The following figure supplement is available for figure 5:

Figure supplement 1. Genome-wide association studies (GWAS) for Ulcerative colitis (UC) and Crohn's disease (CDs) in humans.

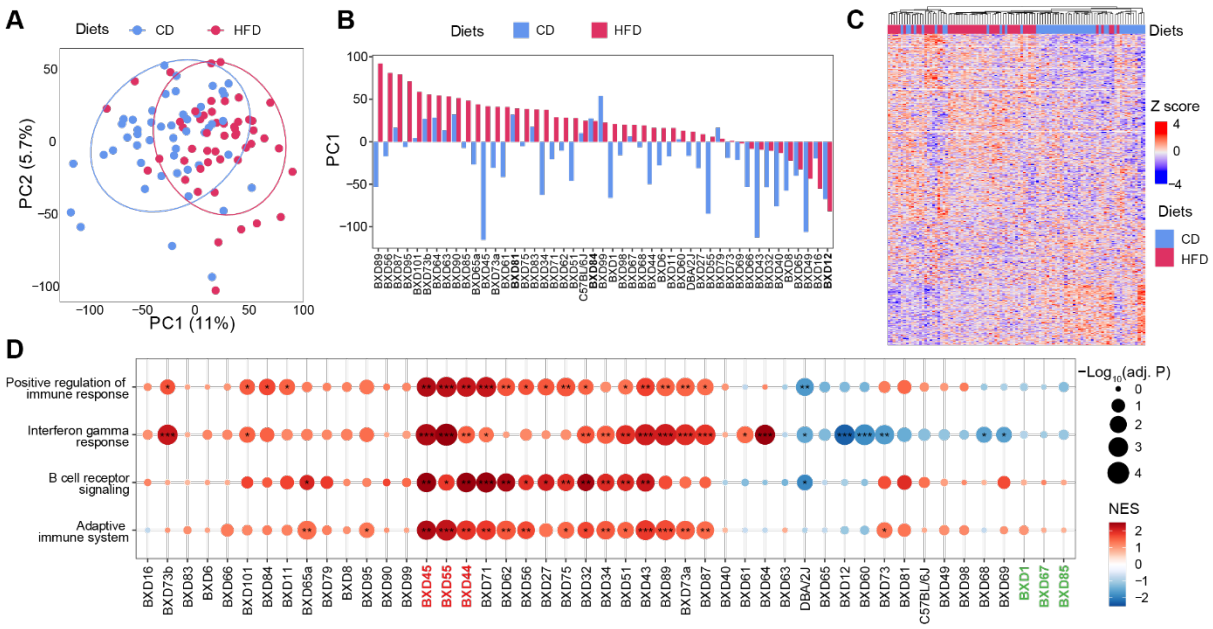


Figure 1—figure supplement 1. Transcriptome profile in BXD colons. (**A**) Principal-component analysis (PCA) of the microarray profiles of BXD colons under high-fat (HFD, indicated in red) or chow diet (CD, represented in blue). (**B**) Bar plot showing the primary principle (PC1) calculated by PCA of the colon transcriptomes in each BXD strain fed with CD (blue) or HFD (red). BXD strains highlighted in bold mean they are more resistant to dietary challenges. (**C**) Heatmap showing unsupervised hierarchical clustering of colon transcriptome in both the CD and HFD fed BXDs. (**D**) Enrichment analysis of inflammation-related genesets showing the effect of HFD on gene expression in individual BXD colon. Normalized enrichment scores (NES) were represented by color and -Log_{10}(BH-adjusted P values) were represented by dot size and indicated as follows: * Adjusted P value < 0.05; ** Adjusted P value < 0.01; *** Adjusted P value < 0.001. BXD strains highlighted in red represent the 3 most susceptible strains to gut inflammation upon HFD. BXD strains colored in green show no significant enrichment in gut inflammation.

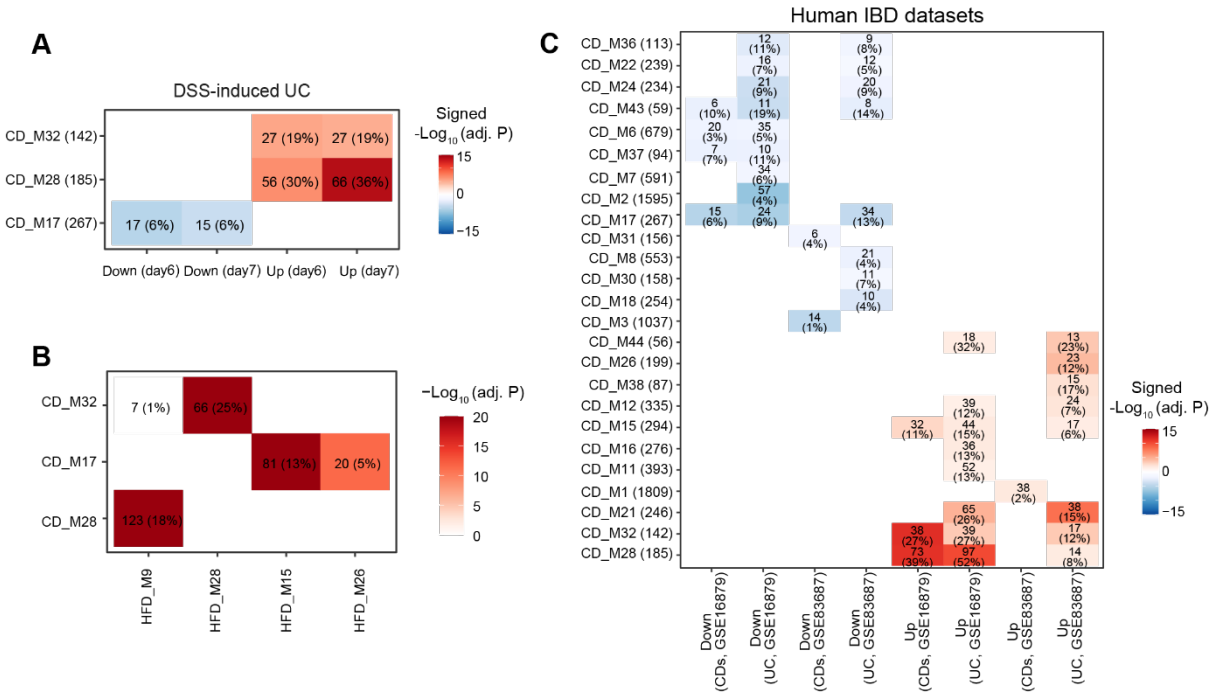


Figure 3—figure supplement 1. Exploring IBD-associated co-expression modules under CD. (**A**) Heatmap showing modules enriched in mouse IBD gene signatures and the number and percentage of enriched genes were labeled. The number of enriched genes divided by the number of genes involved in the respective module was defined as the percentage of enriched genes. Signed $-Log_{10}$(adj. P) indicated BH-adjusted P value and the enriched gene set. Enriched modules of up- and down-regulated genes upon disease were highlighted in red and blue, respectively. (**B**) Heatmap showing the similarity of co-expression modules identified under CD or HFD. The number and percentage of overlapped genes were labeled. The number of overlapped genes divided by the number of genes involved in the respective module in HFD was defined as the percentage of overlapped genes. Adjusted P values calculated by BH were indicated by color. (**C**) Heatmap showing the modules under CD enriched in human IBD gene signatures and the number and percentage of enriched genes were labeled. UC: Ulcerative colitis, CDs: Crohn's disease.

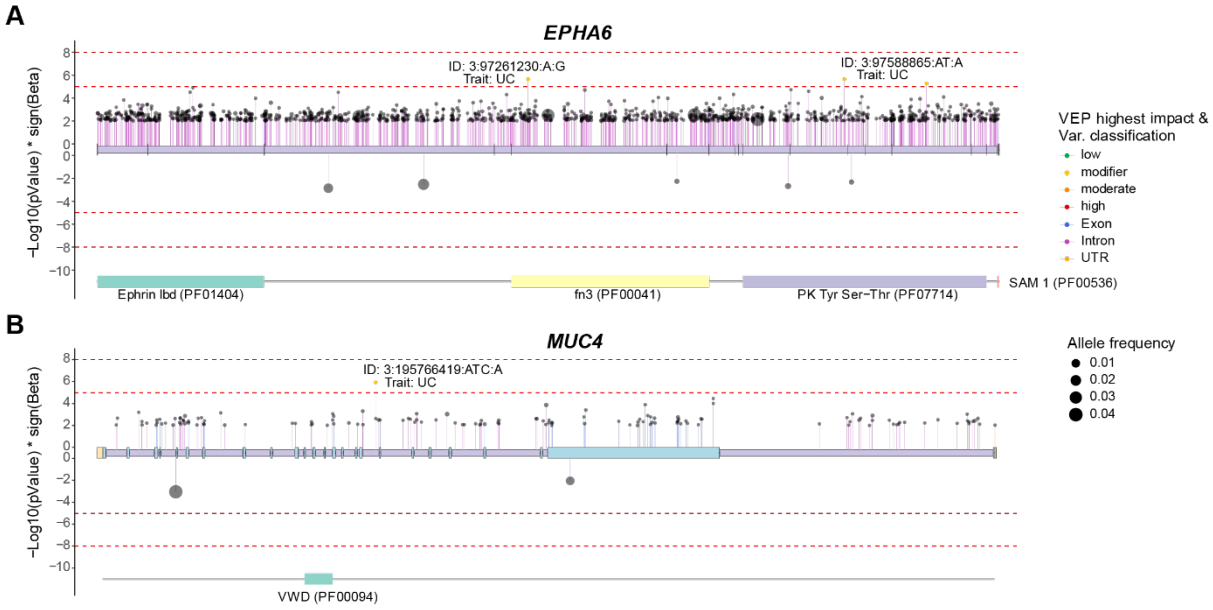


Figure 5—figure supplement 1. Genome-wide association studies (GWAS) for Ulcerative colitis (UC) and Crohn's disease (CDs) in humans. (**A**, **B**) Lollipop plots showing UC- or CDs-associated genetic variants of *EPHA6* (**A**) and *MUC4* (**B**) based on UKBB whole genome sequence data. Variants effect were predicted and their classification are represented by color. VEP: Variant Effect Prediction.

903 **Tables**

904 Table 1: Gene signatures of mouse and human IBDs

Disease	Source	DEGs	Thresholds	Species	Tissues
DSS-induced UC	GSE131032 (Czarnewski *et al.*, 2019)	Author-provided DEGs	absolute log2FC >1 & BH-adjusted P value < 0.05	Mouse	Colon
Crohn's disease & UC	GSE16879 (Arijs *et al.*, 2009; Li *et al.*, 2019)	Author-provided DEGs	absolute FC >1 & BH-adjusted P value <0.01	Human	Colon
Crohn's disease & UC	GSE83687 (Peters *et al.*, 2017)	DEGs computed by limma package (Ritchie et al., 2015)	absolute log2FC >1 & BH-adjusted P value <0.001	Human	Colon
Crohn's disease	SCP1884 (Kong *et al.*, 2023)	Author-provided DEGs	DE coefficient >1 & BH-adjusted P value <0.05	Human	Colon

905

906

907 **Table Supplements and Legends**

908 **Appendix 1 - Table 1. Co-expression networks under CD or HFD. Related to Figure 3**

909 **Appendix 1 - Table 2. Genes under QTL peak of module HFD_M28. Related to Figure 5**

910 **Appendix 1 - Table 3. Associations between genetic variants of EPHA6 and IBD. Related**

911 **to Figure 5 and its figure supplement 1**

912 **Appendix 1 - Table 4. Associations between genetic variants of MUC4 and IBD. Related**

913 **to Figure 5 and its figure supplement 1**

914 **Appendix 1 - Table 5. G-MAD result for *Epha6* in mice and *MUC4* in humans. Related to**

915 **Figure 5**

916 **Appendix 1 - Table 6. MR result for MUC4.**

917 **Appendix 1 - Table 7. Medicine for human IBD.**

918

919

920