OPEN ACCESS

Original research

Gut virome-colonising *Orthohepadnavirus* genus is associated with ulcerative colitis pathogenesis and induces intestinal inflammation *in vivo*

Luca Massimino,[1,2] Orazio Palmieri,[3] Amanda Facoetti,[2,4] Davide Fuggetta,[2,4] Salvatore Spanò,[1,2] Luigi Antonio Lamparelli ,[5,6] Silvia D'Alessio,[7] Stefania Cagliani,[2,4] Federica Furfaro,[1] Ferdinando D'Amico,[1] Alessandra Zilli,[1] Gionata Fiorino,[1,4] Tommaso Lorenzo Parigi ,[1,4] Daniele Noviello ,[8] Anna Latiano,[3] Fabrizio Bossa ,[3] Tiziana Latiano,[3] Alessandra Pirola,[9] Luca Mologni,[10] Rocco Giovanni Piazza,[10,11] Danilo Abbati,[2] Francesco Perri,[3] Chiara Bonini,[2,4] Laurent Peyrin-Biroulet ,[12,13] Alberto Malesci,[1,4] Vipul Jairath ,[14] Silvio Danese ,[1,2,4] Federica Ungaro [1,2,4,6]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2022-328375).

For numbered affiliations see end of article.

Correspondence to
Dr Federica Ungaro, Department of Gastroenterology and Digestive Endoscopy, Division of Immunology, Transplantation and Infectious Disease, IRCCS Ospedale San Raffaele, Milano, 20132, Italy;
ungaro.federica@hsr.it

LM, OP and AF are joint first authors.
SD and FU are joint senior authors.

Received 29 July 2022
Accepted 6 February 2023

To cite: Massimino L, Palmieri O, Facoetti A, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2022-328375

ABSTRACT

Objectives Ulcerative colitis (UC) is a chronic inflammatory disorder of unknown aetiology. Gut virome dysbiosis is fundamental in UC progression, although its role in the early phases of the disease is far from fully understood. Therefore, we sought to investigate the role of a virome-associated protein encoded by the *Orthohepadnavirus* genus, the hepatitis B virus X protein (HBx), in UC aetiopathogenesis.

Design HBx positivity of UC patient-derived blood and gut mucosa was assessed by RT-PCR and Sanger sequencing and correlated with clinical characteristics by multivariate analysis. Transcriptomics was performed on HBx-overexpressing endoscopic biopsies from healthy donors.

C57BL/6 mice underwent intramucosal injections of liposome-conjugated HBx-encoding plasmids or the control, with or without antibiotic treatment. Multidimensional flow cytometry analysis was performed on colonic samples from HBx-treated and control animals. Transepithelial electrical resistance measurement, proliferation assay, chromatin immunoprecipitation assay with sequencing and RNA-sequencing were performed on *in vitro* models of the gut barrier. HBx-silencing experiments were performed *in vitro* and *in vivo*.

Results HBx was detected in about 45% of patients with UC and found to induce colonic inflammation in mice, while its silencing reverted the colitis phenotype *in vivo*. HBx acted as a transcriptional regulator in epithelial cells, provoking barrier leakage and altering both innate and adaptive mucosal immunity *ex vivo* and *in vivo*.

Conclusion This study described HBx as a contributor to the UC pathogenesis and provides a new perspective on the virome as a target for tailored treatments.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Gut virome dysbiosis is fundamental to ulcerative colitis (UC) progression, although it remains unclear how it may be involved in the early phases of chronic inflammation.

WHAT THIS STUDY ADDS

⇒ The hepatitis B virus X protein-harbouring *Orthohepadnavirus* genus, detected in a subcohort of patients with UC, affects intestinal mucosal barrier integrity by impacting the immune cell transcriptional state and induces intestinal inflammation *in vivo*.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study proposes the virome as a plausible target for tailored treatments in UC, possibly leading to an entirely new approach to therapeutic intervention.

INTRODUCTION

Ulcerative colitis (UC), one of the major forms of inflammatory bowel disease (IBD), is a chronic inflammatory condition of the colon characterised by a continuous pattern of epithelial ulceration causing bleeding and abdominal pain in patients, for which definite treatment is still lacking.[1] Despite UC's aetiology remains unknown, intestinal dysbiosis has been recognised as pivotal for its progression.[1] Among several intestinal commensals, the contribution of viruses to the pathogenesis of IBD has gained interest during the last years.[2 3] On one hand, bacteriophages can shape the bacterial composition of the intestinal communities,[4] on the other eukaryotic-targeting viruses have been proven to interact with the host and contribute to intestinal inflammation.[5] In this regard, Adiliaghdam *et al* have recently shown an outstanding proof-of-concept depicting the virome, isolated from patients with long-lasting IBD who underwent surgery, as an autonomous contributor to the disease phenotype

at the level of the innate immunity *in vitro* and in experimental colitis models.[6] However, the paramount question that remains unanswered is whether and which specific gut virome-associated factors contribute to IBD aetiopathogenesis interacting with the host's mucosal immunity.[7]

Previous studies have highlighted the hepatitis B viral protein X (HBx), belonging to the eukaryotic-targeting *Orthohepadnavirus* genus, to be upregulated specifically in the gut virome of early diagnosed, treatment-naïve paediatric patients with UC by comparison with healthy and Crohn's disease (CD) mucosae,[8] shedding light on the possible involvement of this viral entity in UC aetiopathogenesis.

Based on these premises, we here aim to understand the role(s) and function(s) of HBx in UC pathogenesis. We show that HBx impairs the gut mucosal immune defence, particularly at the level of the epithelial barrier, inducing *per se* colitis-like symptoms in mice. We also observe that HBx shapes the transcriptomic state of intestinal cells, acting as a DNA-binding protein and transcriptional regulator of genes known to be involved in intestinal inflammation and epithelial cell biology.

Our results identify HBx as a novel key orchestrator of UC pathogenesis, proving the concept that specific virome-derived factors may be directly involved in initiating and sustaining chronic intestinal inflammation and therefore become reasonable targets of innovative therapies implementing the current protocols with tailored disease management.

METHODS

Please refer to the online supplemental material 1 for comprehensive details.

RESULTS

The *Orthohepadnavirus* genus-belonging HBx transcript is detected in a cohort of patients with UC

Hepadnaviridae is a family of small enveloped double-stranded DNA viruses with mainly hepatotropic features that provoke either transient or persistent infections, such as hepatitis.[9] They include five genera, among which *Orthohepadnavirus* is the sole genus having mammals as a natural host.[9] While all genera share the expression of three major sets of proteins (precore/core, polymerase and preS/S), only *Orthohepadnavirus* carries an additional open reading frame, known as HBx.[9] HBx, largely studied in the context of persistent hepatitis B virus (HBV) infection in the liver,[10] has been described as interfering with the cellular transcriptional machinery and impacting different biological processes, which range from DNA damage repair to immune system activation.[11 12]

By metatranscriptomics, we have recently shown that HBx was specifically enriched in the gut virome of a cohort of paediatric patients with UC and absent in patients with CD and healthy subjects, pinpointing this viral entity as a possible direct contributor to intestinal inflammation onset.[8] Of note, during this study, HBx transcript was detected in both immune and non-immune cell populations isolated from HBx-positive UC patient-derived colonic mucosae (online supplemental figures 1A–C).

We then confirmed the HBx positivity in colonic biopsies in two independent cohorts of adult patients with UC, where we detected the transcript in 41%–57% of cases (figure 1A,B, and online supplemental figure 1). No differences were found between HBx-positive and HBx-negative patients with UC in terms of clinical characteristics and disease manifestation/progression. Indeed, by multivariate analysis, no direct associations were found between HBx transcript positivity and standard clinical management data including, but not limited to, age, gender, treatment protocols and faecal calprotectin levels as a proxy for disease severity[13] (figure 1C and online supplemental figure 2).

Since HBx is a protein encoded also by human HBV, its presence in the gut could be due to ongoing or former HBV infections of these patients. To exclude this option, anti-HBV serum positivity (anti-HBc) was correlated with HBx positivity, but no statistically significant correlation was found (figure 1C). We thus wondered whether the HBx encoding sequence might be part of the genetic background. Surprisingly, only 4% of patients with UC (cohort 1, figure 1B) showed the HBx encoding sequence in their genomic DNA. This finding might be explained by the fact that the *Hepadnaviridae* viruses only occasionally and accidentally integrate into the host's genome,[14] thus not being part of the genetic background of these individuals. To further support this hypothesis, we screened DNA samples from 52 UC trio probands (figure 1D) and the same percentage of HBx positivity (4%) was confirmed in the peripheral blood, composed of circulating immune and non-immune cells, which can derive also from the intestine.[15] Moreover, no vertical transmission was observed (figure 1E).

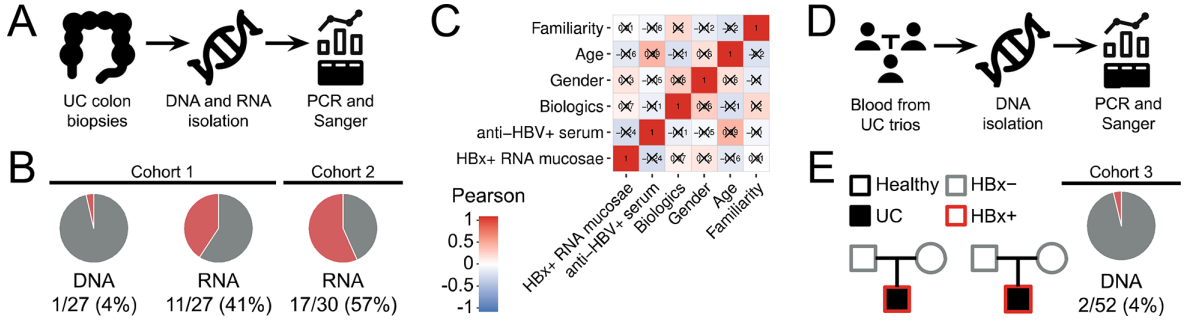


Figure 1 Hepatitis B virus X protein (HBx) transcript positivity characterises a cohort of patients with UC. (A) Schematic representation of the experimental workflow for HBx transcript detection in patients with UC. (B) Pie charts summarising HBx positivity in genomic DNA or RNA of UC cohort 1 (Humanitas Research and Clinical Institute, Milan, Italy), and RNA in UC cohort 2 (Fondazione Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy). (C) Correlation heatmap showing multivariate analysis results expressed as Pearson's coefficients (from 1 to −1). X marks non-significant at p<0.05 (adjustment: Holm). (D) Schematic representation of the experimental workflow for assessing HBx positivity in genomic DNA of patients with UC. (E) On the left, the inheritance tree shows the absence of vertical transmission of HBx genomic DNA in UC cohort 3-derived blood samples (Casa Sollievo della Sofferenza). On the right, is a pie chart summarising the results. Icons from Streamline (https://app.streamlinehq.com).

Collectively, these results highlighted the intestinal HBx transcript positivity to define a cohort of patients with UC and to be not related to anti-HBV serological reactivity nor part of the host's genetics.

HBx impacts the physiological functions of the intestinal mucosa both *ex vivo* and *in vivo*

Although HBx was detected in specific cell compartments (online supplemental figure 1), its impact on the gut mucosa was not elucidated yet. Therefore, to understand how a gut virome-derived factor may contribute to UC pathogenesis, we performed a transcriptomic analysis of healthy donor-derived mucosal biopsies transduced with lentiviruses carrying either the HBx-IRES-GFP encoding sequence or the GFP as control (online supplemental figure 3A). HBx overexpression was found to impact the gene expression profile of intestinal tissues by upregulating 1361 and downregulating 204 genes (online supplemental figure 3B), resulting in the enrichment of biological processes related to antimicrobial response (including the toll-like receptor-mediated signalling), pro-inflammatory pathways (tumour necrosis factor (TNF)-α, interleukins and chemokines), epithelial cell regeneration (Wnt pathway) and the downregulation of interferons (IFN)-related and response to virus-related pathways (online supplemental figure 3C), indicating HBx to interfere with the correct physiology of the tissue and to contribute to UC-specific cellular and molecular alterations.[1]

These results prompted us to further investigate the HBx functions directly *in vivo* to support the pathogenic role of HBx in intestinal inflammation. To this end, C57BL/6 wild-type mice were intrarectally injected with liposomes conjugated with either HBx-IRES-GFP-encoding (hereafter referred to as HBx) or GFP-encoding plasmid as control (hereafter referred to as GFP) (figure 2A). After 15 days, HBx-treated mice started developing clinical symptoms of colitis, as indicated by the increased Disease Activity Index (DAI) (figure 2B), a higher endoscopic score (figure 2C) and reduced colon length (figure 2D), by comparison with the GFP-treated animals. The endoscopic score combines various parameters, including colon translucency, mucosal granularity (representing oedema and small erosions) and changes in the vascular pattern. While in GFP-treated mice the mucosa was translucent, with low granularity and visible blood vessels, in HBx-induced animals we observed a significant increase in colon thickness and granularity, with clear signs of mucosal ulcerations and merely visible blood vessels (online supplemental figures 4A–D).

To evaluate whether HBx-induced colitis is influenced by microbiota, mice were treated for 2 weeks with a broad-spectrum antibiotic cocktail and then administered with either GFP-carrying or HBx-carrying liposomes for 15 days to evaluate the early effects of the viral factor in these conditions. Groups without antibiotics were used as control (online supplemental figure 5A). Interestingly, antibiotic treatment did not impact the overall DAI (online supplemental figure 5B), indicating that the HBx-induced effects did not depend on the microbiota composition.

Similar results were observed at the tissue level. In fact, both the colon lengths and endoscopic scores of HBx-treated mice did not change with statistical significance on antibiotic administration (online supplemental figures 5C–E), indicating that microbiota did not influence the HBx-induced colitis-like tissue hallmarks in mice.

Notably, the increased diarrhoea and faecal blood observed in HBx-treated mice (online supplemental figures 4C,D), along with the mucosal ulceration, led us to propose HBx as a factor that impairs barrier function in the intestinal mucosa, thus promoting gut inflammation. The barrier alteration then induces the translocation of commensal bacteria and microbial products from the gut lumen into the bowel wall, leading to acute mucosal inflammation that, if not resolved by an adequately mounting immune response, leads to chronic intestinal inflammation. This is in line with the leading theory explaining UC pathogenesis.[16]

To evaluate the immunological changes in our model, we performed a FACS analysis of murine colonic tissues. This revealed that the number of dendritic cells (DCs), $CD8^+$ T cells and neutrophils was significantly reduced in HBx-treated versus GFP-treated mice, while macrophages (MΦ) and endothelial cells were only slightly affected (figure 2E,F and online supplemental figure 6). Since the first defence against pathogens (innate immunity and T cells) was impaired and the low number of neutrophils may hamper the correct resolution of inflammation,[17] such an immune state of the colonic mucosa on HBx stimulation may explain why mice developed colitis-like symptoms.[17] This immune landscape is in line with UC-specific pathological traits.[17 18] In contrast with the higher levels of ulcerations (a typical sign of epithelial cell loss and crypt damage[19 20]), intestinal epithelial cell number was significantly increased by HBx compared with controls (figure 2E,F, and online supplemental figure 6).

Since the epithelial barrier alteration is crucial in initiating UC pathogenesis,[16] we sought to better elucidate this point by exploiting the Caco-2 cell line as an *in vitro* model of the intestinal epithelial barrier.[21] HBx-overexpressing Caco-2 cells displayed an increased proliferation rate by comparison with the GFP-transduced cells (figure 3A) and a reduced transepithelial electrical resistance (TEER) (figure 3B), indicating the barrier leakage. These data were also confirmed *in vitro* in epithelial organoid cultures (figure 3C,D) transduced with either GFP or HBx-IRES-GFP lentiviruses (figure 3E), which showed upregulation of stemness markers (*LRG5* and *OLFM4*) and downregulation of epithelial barrier markers (*MUC1*, *MUC13* and *TJP1*) in HBx-transduced cells compared with GFP (figure 3F,G), suggesting the viral protein to increase proliferation capabilities and to affect barrier function in the intestinal epithelium.

Collectively, these results demonstrate that HBx drives colitis *in vivo* by acting at the level of the epithelial barrier and, in parallel, affects the gut mucosal immune cells, likely not mounting an adequate defence against the invading microorganisms and thus fostering a persistent inflammation unable to self-resolve.

HBx acts as a transcriptional regulator in the intestinal epithelium

The findings collected so far prompted us to study more in detail the effect of HBx on the gut epithelium. Since HBx was previously found to bind the host's DNA and regulate transcription during HBV infection,[10 11 22 23] we performed a chromatin immunoprecipitation assay with sequencing (ChIP-Seq) analysis on V5-tagged HBx-transduced Caco-2 cells to understand the early effects of HBx (figure 4A). For comprehensive chromatin profiling, we took advantage of the open-source data available at ENCODE (https://www.encodeproject.org/)[24] and analysed several other epigenetic markers together with HBx-V5 occupancy. HBx ChIP-Seq profile closely resembled that of a typical transcription factor with discrete narrow peaks (online supplemental figures 7A,B) and was enriched for specific DNA motifs (figure 4B), demonstrating that HBx retained its chromatin-binding behaviour even in non-hepatic cells. To better

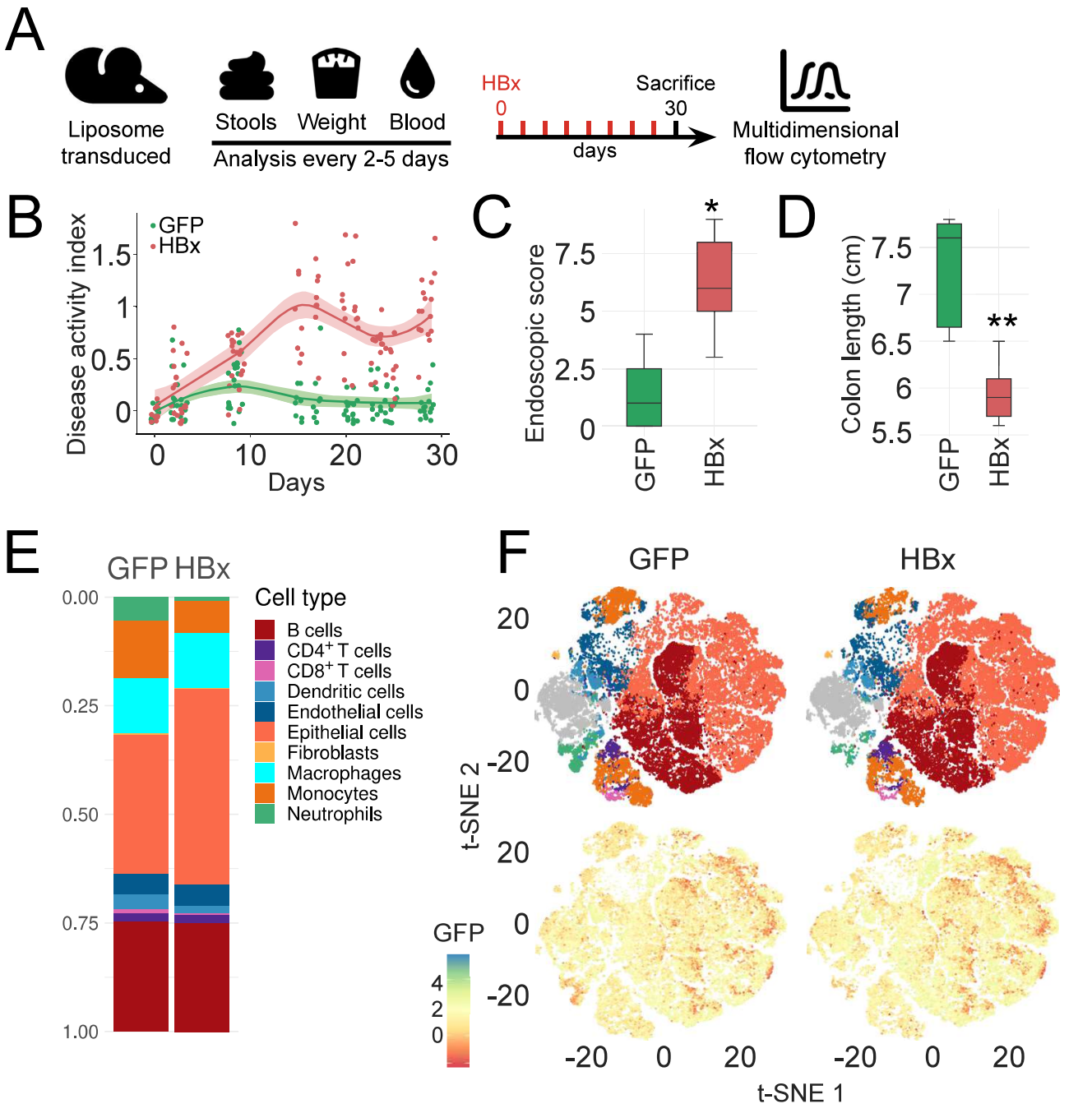


Figure 2 Hepatitis B virus X protein (HBx) induces colitis symptoms in mice and shapes the colonic mucosal immunity. (A) Schematic experimental workflow of *in vivo* HBx-induced colitis (n=6/group, two independent experiments). (B) Disease Activity Index (DAI) of mice treated with either GFP-carrying or HBx-carrying liposomes. Differences between groups are statistically significant. Statistical analysis was performed with two-way analysis of variance, with Bonferroni's postcorrection. Results with a p value <0.05 were considered significant. * P<.05; ** P<.01. (C,D) Box plots showing endoscopic score (C) and colon length (D) of HBx-induced colitic mice. (E, F) Relative cell population abundance (E) and t-distributed stochastic neighbour embedding (t-SNE) multidimensional scaling (F) in HBx-induced colitis mice versus the GFP control. Immunophenotyping was performed on n=3 mice/group. Icons from Streamline (https://app.streamlinehq.com).

characterise these regions, peaks were mapped to the nearest genes and annotated accordingly. Interestingly, the great majority of the peaks were found within intergenic regions (figure 4C), typically not accessible to transcription factors (low DNAse I hypersensitivity), and far from canonical promoter markers (ie, H3K4me3 and H3K27me3), as assessed by Pearson's correlation and Hidden-Markov chromatin modelling (figure 4D,E). These data depicted HBx as a possible pioneering factor,[25] binding putative enhancers in a heterochromatin state and therefore not commonly used by Caco-2 cells. To understand the

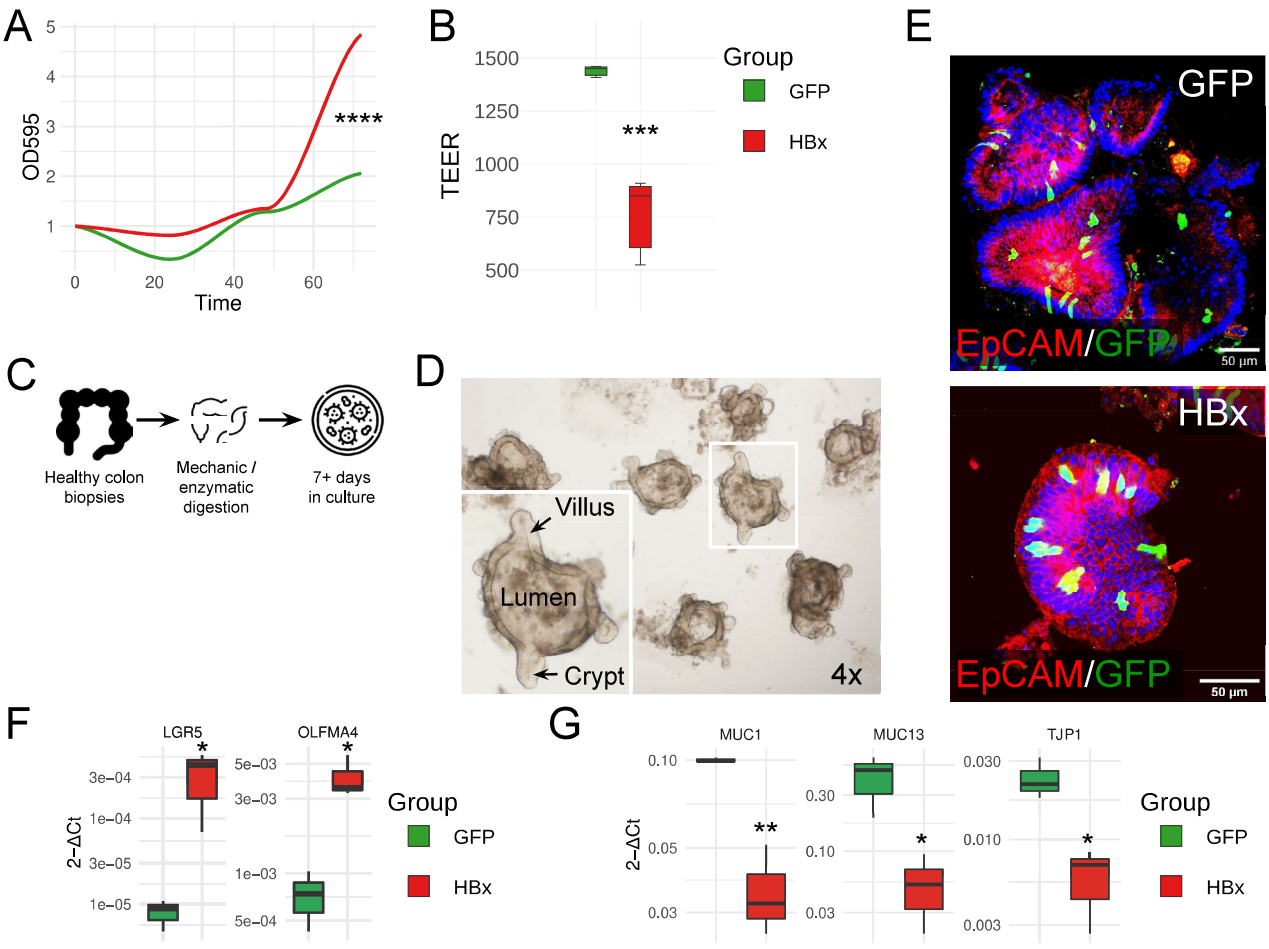


Figure 3 Hepatitis B virus X protein (HBx) alters epithelial barrier functions. (A) Graph showing cellular growth rate between HBx-transduced and GFP-transduced epithelial cell line. Experiments were performed in triplicates in three independent experiments. Statistical analysis was performed with two-way analysis of variance with Bonferroni's postcorrection. (B) Box plot showing transepithelial electrical resistance (TEER, expressed as Ω/cm²) measurements on HBx-transduced and GFP-transduced epithelial cell lines. Statistical analysis was performed with Student's t-test (C) Schematic representation of organoid isolation. (D) After 7 days in culture, organoids become mature and structured, with crypt and villus domains. (E) Immunofluorescence images showing GFP-transduced and HBx-transduced intestinal organoid structures and the positivity for the epithelial marker EpCAM (red) and the GFP (green). (F, G) Box plots showing real-time PCR results for the stem cell (F) and epithelial barrier markers (G) in organoids transduced with either the GFP-carrying or HBx-carrying lentiviruses, expressed as 2-ΔCT (GAPDH was used as the housekeeping genes). Statistical analysis was performed with Student's t-test. *P<0.05; **p<0.01; ***p<0.005; ****p<0.001. Icons from Streamline (https://app.streamlinehq.com).

consequences of such chromatin binding, we performed RNA-sequencing (RNA-Seq) and transcriptome profiling of the same cells. We observed a modest but specific dysregulation of gene expression, with 468 upregulated and only 11 downregulated genes on HBx-V5 overexpression (figure 4F), with the functional enrichment of Gene Ontology categories related to cell proliferation and tissue regeneration, pro-inflammatory signals, epithelial-to-mesenchymal transition and transcriptional regulation (figure 4G), supporting the hypothesis that HBx shapes the overall transcriptional state of epithelial cells towards a pro-inflammatory phenotype.[26] Furthermore, when the differentially expressed genes were intersected with HBx ChIP targets, 20 genes were found to be modulated while being directly bound by HBx, mainly within the intergenic regions (figure 4H–J). These genes were found to be involved in the regulation of epithelial barrier functions (FERM domain containing 4B (FRMD4B); fibulin 7 (FBLN7); mucin 17 cell surface-associated (MUC17)), intracellular signalling (platelet-derived growth factor receptor alpha (PDGFRA); sonic Hedgehog signalling molecule (SHH); phosphodiesterase 4D (PDE4D); transmembrane protein 242 (TMEM242)), cell division (establishment of sister chromatid cohesion N-acetyltransferase 1 (ESCO1); leucine-rich repeat and coiled-coil centrosomal protein 1 (LRRCC1); thyroid hormone receptor interactor 13 (TRIP13); tubulin beta 2B class IIb (TUBB2B), zinc finger CCHC-type containing 7 (ZCCHC7)) and transcription (cAMP-responsive element-binding protein 5 (CREB5); LUC7 like 2 (LUC7L2), pre-mRNA splicing factor) (online supplemental table 3). Their altered expression may explain the increased proliferation rate and epithelial barrier leakage shown by the HBx-overexpressing epithelial Caco-2 cell line. Interestingly, many of these genes were already found to be associated with UC pathogenesis.[27–33]

Conclusively, HBx was shown to impair the functions of epithelial cells by directly binding to intergenic DNA regions,

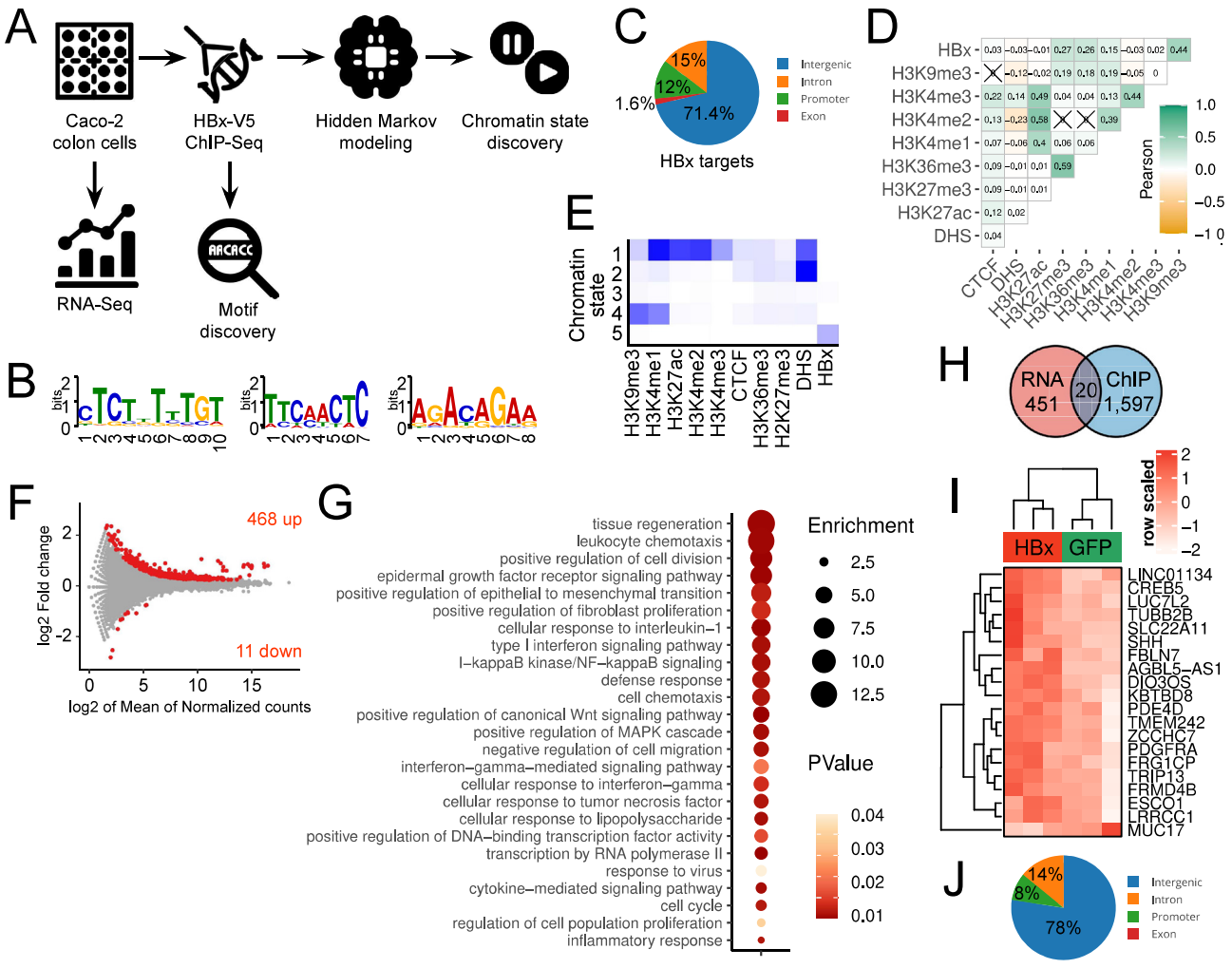


Figure 4 Hepatitis B virus X protein (HBx) regulates the expression of ulcerative colitis (UC)-related genes by binding to enhancer regions. (A) Schematic experimental workflow of the chromatin immunoprecipitation assay with sequencing (ChIP-Seq) and RNA-sequencing (RNA-Seq) analyses of HBx-V5-transduced and GFP-transduced cells. (B) Top three representative DNA motifs enriched in HBx peaks. (C) Pie chart showing gene body-centric annotation of all HBx peaks. (D) Correlation heatmap showing ChIP-Seq coverage multivariate analysis results expressed as Pearson's coefficients (from 1 to −1). (E) Heatmap showing the relative abundance of the different ChIP-Seq signals within the chromatins states found by hidden Markov modelling. (F, G) MA-plot (F) and Gene Ontology plot (G) showing differentially expressed genes and the resulting dysregulated biological processes, respectively, in HBx-expressing versus GFP-expressing Caco-2 cells. (H) Venn diagram showing the intersection between the differentially expressed genes and HBx putative targets. (I, J) Gene expression heatmap (I) showing the 20 genes found to be differentially expressed while directly bound by HBx within their respective enhancers, summarised in the pie chart (J) showing gene body-centric annotation. Icons from Streamline (https://app.streamlinehq.com).

affecting gene expression in epithelial cells that concomitantly acquired a stem-cell-like phenotype (by SHH overexpression) and reduced mucin expression (MUC17), together with the activation of other genes already reported to be involved in UC pathogenesis.

In vitro and *in vivo* silencing of HBx reverts colitis-like symptoms and restores the gut barrier functions

To pinpoint the relevance of HBx as a possible therapeutic target, we assessed the efficacy of HBx inhibition *in vitro*. To this aim, HBx-overexpressing cells were treated with three different HBx-targeting small interfering RNAs (siRNAs) (online supplemental figure 8A). HBx-overexpressing Caco-2 reduced their proliferation rate and recovered the TEER values by comparison with the scramble-treated cells (online supplemental figures 8B,C). Also, tight junction-related marker expression was recovered on HBx-targeting siRNA administration as compared with the control cells (online supplemental figure 8D).

We then tested siRNA HBx-targeting efficacy *in vivo*. HBx-induced colitic mice were treated with the viral factor-targeting siRNAs after colitis induction (figure 5A). Interestingly, the treatment with HBx-targeting siRNAs caused a prompt reversion of the inflammatory symptoms in terms of reduced DAI, bleeding, stool consistency, endoscopic scores and recovery of the colon length (figure 5B–D and online supplemental figures 9A–D), in contrast with animals receiving scrambled siRNAs. These results were paralleled by the recovery of the expression of the epithelial barrier integrity indicators Muc13 and *Tjp1* (online

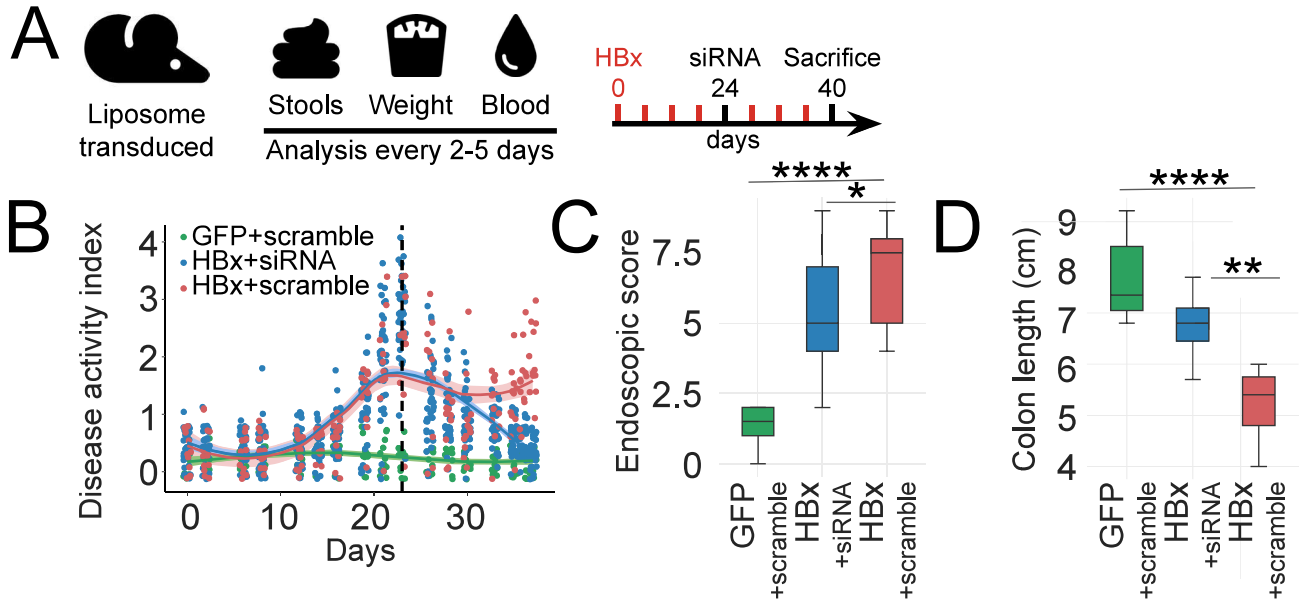


Figure 5 Hepatitis B virus X protein (HBx)-induced phenotype is reversible *in vivo*. (A) Schematic experimental workflow of *in vivo* HBx-induced colitis and small interfering RNA (siRNA)-mediated rescue treatment (n=6 mice/group, 2 independent experiments). (B) Disease Activity Index (DAI) of mice treated with either GFP-carrying or HBx-carrying liposomes and HBx-targeting siRNAs or the scramble. The black dotted line indicates the commencement of siRNA administration. Differences between groups are statistically significant. Statistical analysis was performed with two-way analysis of variance, with Bonferroni's postcorrection. Results with a p value <0.05 were considered significant. (C, D) Box plots showing endoscopic score (C) and colon length (D) of GFP-carrying or HBx-carrying liposomes and HBx-targeting siRNAs or the scramble. Statistical analysis was performed with Student's t-test. *P<0.05; **p<0.01; ****p<0.001. Icons from Streamline (https://app.streamlinehq.com).

supplemental figure 10A), the modulation of Notch signalling (online supplemental figure 10B) and the decreased expression of the pro-inflammatory *Tnf* and *Il1b* (online supplemental figure 10C), demonstrating that HBx inhibition reverted the pro-inflammatory phenotype, reestablishing the barrier integrity and the inflammation-related defence response.[34]

DISCUSSION

The two main types of IBD, UC and CD, differ in localisation, pattern of inflammation, type of immune cell infiltrate and complications developed.[35] We recently proposed the two diseases to be different also in terms of virome composition, pointing out HBx to be specific for patients with UC (treatment-naïve and early diagnosed) and absent in the gut mucosa of patients with CD and healthy patients.[8] Given these premises, it is reasonable to speculate that HBx could be associated with UC pathogenesis and not involved in CD.

The human microbiota is a densely populated community of different entities important for maintaining tissue homeostasis and stimulating the host's immunity.[36] While many studies have already investigated and assessed the roles and functions of bacteria in the gut, few notions are available for the viral commensals of the intestine, although they have been previously proposed to influence the overall intestinal microbiota composition.[4 5 37] Importantly, eukaryotic viruses colonising the mucosal surfaces may infect host cells without symptoms, and, even if asymptomatic, a virus-carrying host may harbour a persistent immune response with a 'continuum' of inflammatory mediators that might increase host susceptibility to disease.[7] Moreover, viruses are sensed by the host's specific molecules, thus evoking a specific transcriptional response in both infected and bystander cells with the consequent release of inflammatory signals, ultimately influencing systemic immunity.[7] For these reasons, eukaryotic-targeting viruses have recently attracted considerable interest in intestinal diseases, including UC.[5 38] Importantly, viruses isolated from patients with long-lasting inflammation have been demonstrated to autonomously induce intestinal inflammation in experimental models of colitis.[6]

As a step further, our work identifies for the first time a specific *Orthohepadnavirus*-derived factor, namely HBx, as being associated with UC early pathogenesis[8] and being a direct inducer of colitis-like symptoms in mice by provoking epithelial ulceration and mucosal barrier leakage, leading to the impairment of gut defence. This is an additional piece to the puzzle of IBD pathogenesis that putatively identifies one of the early causes of intestinal inflammation.

Of note, compelling evidence has recently pointed out viruses to be strongly associated with the onset of autoimmune diseases, such as the Epstein-Barr virus in the pathogenesis of multiple sclerosis.[39] A similar scenario may be that of the *Orthohepadnaviruses* found within the intestine that latently stimulate the immune system through the expression of HBx, which may predispose an individual to develop UC by directly regulating the expression of specific genes known to participate in UC pathogenesis.[27–33]

Furthermore, since HBx is a protein encoded by a hepatotropic viral family, its presence in the colon was largely unexpected, although non-human HBV lymphotropism and its ability to use lymphoid cells as extrahepatic reservoirs have been recently described also in lymphoid tissues, including spleen and

lymph nodes.[40–44] These findings support our and other studies reporting the virome to be made up of a large plethora of entities not necessarily colonising their preferential tissues but residing on the mucosal surfaces while expressing antigenic molecules and stimulating tissue immunity without activating the canonical infection cycle.[7 8]

Most likely, HBx might exert its detrimental role in the intestinal mucosa by affecting the host's defence against pathogens, altering mucosal homeostasis and ultimately causing the perpetuation of intestinal inflammation in patients with UC.

This possible scenario is corroborated by *in vitro* and *in vivo* experiments. Indeed, epithelial cells *in vitro* were affected in their transcriptional state and lost their capability to maintain the barrier on HBx administration. Likewise, mice showed colitis-like symptoms and barrier leakage with an altered immune milieu where DCs, $CD8^+$ T cells and neutrophils were reduced in number. Notably, DCs are a unique cell subset specialised in the production of type I IFNs. They promote antiviral immune responses and are implicated in the pathogenesis of autoimmune diseases characterised by a type I IFN signature.[45] Neutrophils are regarded as the first line of defence in the innate arm of the immune system. They capture and destroy invading microorganisms through phagocytic mechanisms and the formation of neutrophil extracellular traps after detecting pathogens.[46] $CD8^+$ T cells include cytotoxic T cells, important for killing virally infected cells, and $CD8^+$ suppressor T cells, which restrain certain types of the immune response.[47] Therefore, it is reasonable that, in UC pathogenesis, the HBx-induced impairment of these populations in the gut facilitates a persistent inflammation in the mucosa after the translocation of invading microorganisms across the ulcerated epithelium.

Notably, our data showed that the microbiota did not influence the HBx-mediated colitis in mice, at least during the induction phase, indicating that HBx does not cooperate with other commensals to exert its action. This is in line with the concept that antibiotic treatment is not effective in inducing remission in patients with UC.[48]

HBx was extensively described in the context of HBV-associated hepatocellular carcinoma for its ability to induce epigenetic modifications in the host's cells.[49] Whereas, to the best of our knowledge, no evidence is available in the setting of UC-associated colorectal carcinogenesis except for some association studies.[50]

Given HBx's prominent role in liver carcinogenesis, a similar effect could be speculated/hypothesised in colitis-associated colorectal cancer and investigated in future studies.

Interestingly, in colons derived from HBx-positive patients with UC, HBx transcript is detected, despite the absence of HBx DNA integration in the majority of the cases tested. This might be consistent with the notion that the Hepadnaviridae life cycle consists of its entry into the host cells, the release of the genome as a relaxed-circular DNA into the host cell's nucleus and conversion to covalently closed circular DNA, which ultimately serves as a template for the transcription of the viral RNAs, including HBx.[51] Occasionally, a part of the incoming Hepadnaviridae DNA is integrated into the host's genome, and, while remaining replication-incompetent, it can act as a template for the production of proteins, possibly related to viral-specific immune tolerance or the development of viral infection-related pathogenesis.[51] Additionally, in our study, the rare HBx DNA integration (4% of the cases tested) was demonstrated to be not inherited. In this regard, we do not exclude that HBx genomic DNA detected in the peripheral blood may derive from immune and non-immune cells with an intestinal origin.[15] Further studies will be required to better elucidate the source of this genomic integration.

HBx positivity was found to not correlate with anti-HBV reactivity. These results suggest, on one hand, the lack of correlation between active HBV infection and the symptomatology observed in our patients, on the other hand, they may suggest that HBx comes from non-human *Orthohepadnaviruses* other than human HBV, raising concerns about how this viral factor was acquired. We speculate that HBx positivity may occur after environmental exposure, for example, during an event of zoonotic spillover. Curiously, when the *Hepadnaviridae* family was searched within the Serratus viral discovery database, the majority of the hits were found in *Trichobilharzia* samples, a notorious human zoonotic agent colonising water during its life cycle (online supplemental results and online supplemental figures 11 and 12), suggesting that exposure to contaminated waters might be proposed as a risk for acquiring the viral factor. This scenario would be in line with previous association studies suggesting environmental factors among the triggers of chronic inflammatory conditions, including UC.[52]

Despite the lack of longitudinal studies demonstrating the association of zoonotic spillover with the increased incidence of UC in the population, we do not exclude that exposure to Hepadnaviridae could happen after contact with specific environmental agents carrying HBx, which may chronically stimulate the immune and non-immune gut compartments. Of note, zoonotic spillover was already shown for other disease outbreaks ('waterborne zoonoses: identification, causes, and control', https://www.who.int/publications/i/item/9241562730). Nevertheless, we are fully aware that this hypothesis needs to be tested in future studies aiming at the discovery of the real source of HBx positivity in the population.

Another worthwhile aspect of our study is the absence of clinical differences between HBx-positive and HBx-negative patients with UC in the cohorts analysed, while HBx was found to induce marked colitis-like symptoms in mice. This might be explained considering two main pieces of evidence: (i) these patients were not paediatric (as in our first study) and (ii) they were suffering from long-lasting inflammation while receiving anti-inflammatory treatments. All these factors might have impacted differential HBx-related clinical manifestations because of other insults intervening during the course of the disease. By contrast, mouse models were performed on healthy wild-type animals, where the first hit was the viral protein only, thus mimicking the early mechanisms leading to intestinal inflammation independently of other actors.

The currently available treatments for UC are all designed to block single factors involved in the mucosal inflammatory process, such as monoclonal antibodies against cytokines (anti-TNF) or small molecules directed at signalling molecules (JAKs). All these approaches are partially effective as these agents lose effectiveness and patients have disease relapse. The main reason for these failures is that UC is a complex disease where numerous biological functions (eg, protein translation, pro-inflammatory molecule secretion, cell proliferation, apoptosis, etc) exert integrated and complementary roles that foster disease persistence. As a result, blocking one single pathway (ie, JAK/STAT) or one single cytokine (ie, TNF) may not be enough to permanently inhibit an overwhelmingly complex inflammatory process.

Our study sheds light on the possibility that a viral genus may induce colitis in humans and the chance that targeting virome-derived factors in specific cohorts of patients with UC may offer a whole new therapeutic potential. First evidence has been

provided by the siRNA-mediated silencing of HBx, effective in recovering the pro-inflammatory phenotype *in vivo*.

Although the gaps between the *in vitro*, *ex vivo* and in vivo experimental results shown in the current study do exist and the path up to the clinical applications is long, the reality is that the treatment of UC is far from satisfactory and has reached a therapeutic 'ceiling'. The paradigm shift in UC pathogenesis that we propose here might result in alternative therapeutic approaches, at least for HBx-positive patients with UC, ultimately ameliorating patients' quality of life.

Author affiliations
[1]Department of Gastroenterology and Digestive Endoscopy, IRCCS Ospedale San Raffaele, Milano, Italy
[2]Division of Immunology, Transplantation and Infectious Disease, IRCCS Ospedale San Raffaele, Milano, Italy
[3]Division of Gastroenterology and Endoscopy, Fondazione IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Foggia, Italy
[4]Faculty of Medicine, Università Vita Salute San Raffaele, Milano, Italy
[5]IBD Center, IRCCS Humanitas Research Hospital, Rozzano, Italy
[6]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[7]PhoenixLAB, Lodi, Italy
[8]Department of Pathophysiology and Transplantation, University of Milan, Milano, Italy
[9]GalSeq s.r.l, Milan, Italy
[10]Department of Medicine and Surgery, University of Milan-Bicocca, Monza, Italy
[11]Hematology and Clinical Research Unit, San Gerardo Hospital, Monza, Italy
[12]Inserm NGERE, University of Lorraine, Nancy, France
[13]Department of Hepato-Gastroenterology, University Hospital Centre Nancy, Nancy, France
[14]Department of Medicine, Division of Gastroenterology, Western University, London, Ontario, Canada

Twitter Luca Massimino @lucamassimino

Acknowledgements We thank Giuseppe Biscaglia MD, Grazia Anna Niro MD, Maria Guerra BSc and Lucia Zaffarano BSc for their support and assistance throughout the study.

Contributors LM and FU: conceptualisation and writing the original draft; FU, OP, AF, DF, SS, SD'A, DN, AL, FB, TL, AP: data search and acquisition; LM, LAL, DA: bioinformatics and statistical analysis, FF, FD'A, AZ: human sample collection; AF, SC, GF, LMO, RGP, FP, CB, LP-B, AM, TLP, VJ, SD, FU: review and editing; SD, FU: supervision; FU: resources and funding acquisition. SD and FU are jointly supervised this work; FU: guarantor of the study.

Funding This study received support from the 'Fondazione Cariplo per la Ricerca Biomedica' under grant agreement #2018-0112 to FU, 'Fondazione AMICI ONLUS' for Research Prize 2020 to FU and the European Crohn's and Colitis Organization (ECCO) under the grant agreement 2017 to FU; Italian Ministry of Health (Ricerca Corrente 2022-2024) to OP and AL and by the Italian Association for Cancer Research (Fondazione AIRC) Investigator Grant 2018 ID 22082 to RP.

Competing interests SD has served as a speaker, consultant and advisory board member for Schering Plough, Abbott (AbbVie) Laboratories, Merck and Co, UCB Pharma, Ferring, Cellerix, Millenium Takeda, Nycomed, Pharmacosmos, Actelion, Alfa Wasserman, Genentech, Grunenthal, Pfizer, AstraZeneca, Novo Nordisk, Vifor and Johnson and Johnson. LP-B has served as consultant for Merck, AbbVie, Janssen, Genentech, Ferring, Tillots, Vifor, Pharmacosmos, Celltrion, Takeda, Biogaran, Boerhinger-Ingelheim, Lilly, Pfizer, Jndex Pharmaceuticals, Amgen, Sandoz, Celgene, Biogen, Samsung Bioepis, Alma, Sterna, Nestlé, Enterome, Mylan, HAC-Pharma, Tigenix, and has served as speaker for Merck, AbbVie, Janssen, Genentech, Ferring, Tillots, Vifor, Pharmacosmos, Celltrion, Takeda, Boerhinger-Ingelheim, Pfizer, Amgen, Biogen, Samsung Bioepis. VJ has received has received consulting/advisory board fees from AbbVie, Alimentiv Inc (formerly Robarts Clinical Trials), Arena Pharmaceuticals, Asahi Kasei Pharma, Asieris, Bristol Myers Squibb, Celltrion, Eli Lilly, Ferring, Flagship Pioneering, Fresenius Kabi, Galapagos, GlaxoSmithKline, Genentech, Gilead, Janssen, Merck, Mylan, Pandion, Pendopharm, Pfizer, Protagonist, Reistone Biopharma, Roche, Sandoz, Second Genome,Takeda, Teva, Topivert, Vividion; speaker's fees from, AbbVie, Ferring, Galapagos, Janssen Pfizer Shire, Takeda, Fresenius Kabi. The other authors declare no conflicts of interest.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was approved by the independent ethic committees (protocol number CE Humanitas ID 975, protocol number CE Casa Sollievo Della Sofferenza: 12701/2008). Patients in this study have been sufficiently anonymised. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Raw and processed data were deposited into the NBCI GEO repository with accession number GSE204665. Sanger sequencing results are available at OP978007-OP978010 in NCBI GenBank Caco-2 raw sequencing data for ChIP-Seq analysis were downloaded from ENCODE and they were mentioned as a reference in the text. Results, analytic methods and study materials will be made available on request to the authors.

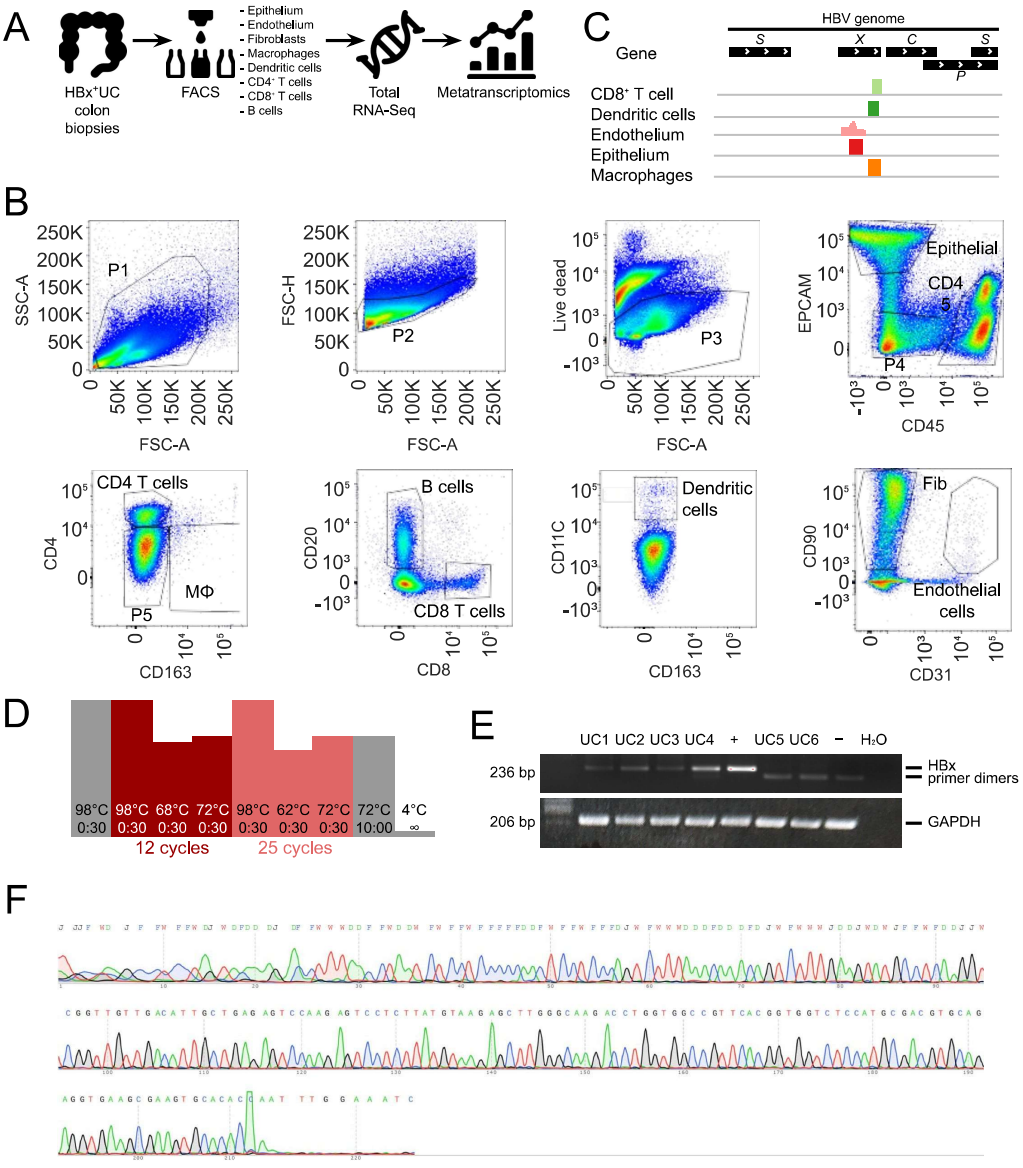
ORCID iDs
Luigi Antonio Lamparelli http://orcid.org/0000-0002-9571-2582
Tommaso Lorenzo Parigi http://orcid.org/0000-0002-3398-0231
Daniele Noviello http://orcid.org/0000-0002-3482-0109
Fabrizio Bossa http://orcid.org/0000-0002-4518-9949
Laurent Peyrin-Biroulet http://orcid.org/0000-0003-2536-6618
Vipul Jairath http://orcid.org/0000-0001-8603-898X
Silvio Danese http://orcid.org/0000-0001-9867-8861
Federica Ungaro http://orcid.org/0000-0001-5395-7795

REFERENCES

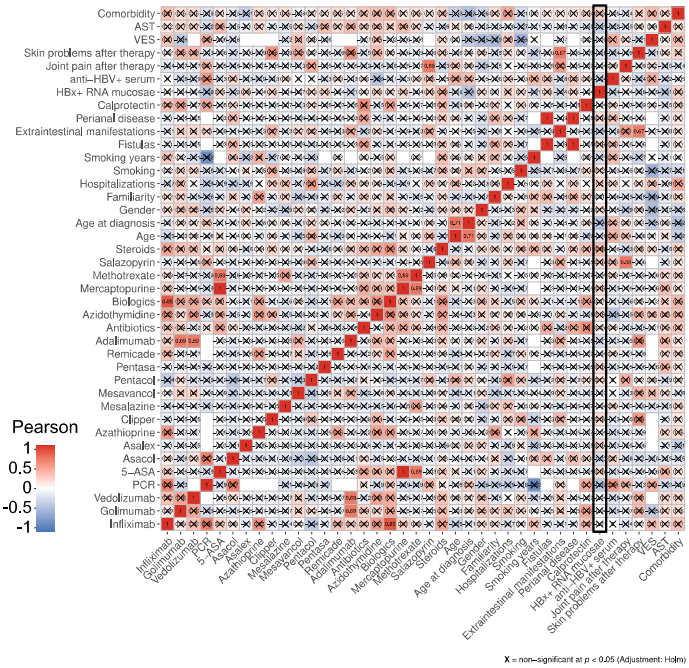
1 Massimino L, Lamparelli LA, Houshyar Y, *et al*. The inflammatory bowel disease transcriptome and metatranscriptome meta-analysis (IBD tamma) framework. *Nat Comput Sci* 2021;1:511–5.
2 Matijašić M, Meštrović T, Paljetak HČ, *et al*. Gut microbiota beyond bacteria-mycobiome, virome, archaeome, and eukaryotic parasites in IBD. *Int J Mol Sci* 2020;21:2668.
3 Fernandes MA, Verstraete SG, Phan TG, *et al*. Enteric virome and bacterial microbiota in children with ulcerative colitis and Crohn disease. *J Pediatr Gastroenterol Nutr* 2019;68:30–6.
4 Norman JM, Handley SA, Baldridge MT, *et al*. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015;160:447–60.
5 Ungaro F, Massimino L, D'Alessio S, *et al*. The gut virome in inflammatory bowel disease pathogenesis: from metagenomics to novel therapeutic approaches. *United European Gastroenterol J* 2019;7:999–1007.
6 Adiliaghdam F, Amatullah H, Digumarthi S, *et al*. Human enteric viruses autonomously shape inflammatory bowel disease phenotype through divergent innate immunomodulation. *Sci Immunol* 2022;7:eabn6660.
7 Virgin HW. The virome in mammalian physiology and disease. *Cell* 2014;157:142–50.
8 Ungaro F, Massimino L, Furfaro F, *et al*. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* 2019;10:149–58.
9 Magnius L, Mason WS, Taylor J, *et al*. ICTV virus taxonomy profile: hepadnaviridae. *J Gen Virol* 2020;101:571–2.
10 Slagle BL, Bouchard MJ. Role of HBx in hepatitis B virus persistence and its therapeutic implications. *Curr Opin Virol* 2018;30:32–8.
11 Bouchard MJ, Schneider RJ. The enigmatic X gene of hepatitis B virus. *J Virol* 2004;78:12725–34.
12 Benhenda S, Cougot D, Buendia M-A, *et al*. Chapter 4 hepatitis B virus X protein. In: *Elsevier*. 2009: 75–109.
13 Kawashima K, Ishihara S, Yuki T, *et al*. Fecal calprotectin level correlated with both endoscopic severity and disease extent in ulcerative colitis. *BMC Gastroenterol* 2016;16:47.
14 Tu T, Zhang H, Urban S. Hepatitis B virus DNA integration: in vitro models for investigating viral pathogenesis and persistence. *Viruses* 2021;13.
15 Pantel K, Denève E, Nocca D, *et al*. Circulating epithelial cells in patients with benign colon diseases. *Clin Chem* 2012;58:936–40.
16 Neurath MF. Cytokines in inflammatory bowel disease. *Nat Rev Immunol* 2014;14:329–42.

17 Ungaro F, Tacconi C, Massimino L, *et al*. MFSD2A promotes endothelial generation of inflammation-resolving lipid mediators and reduces colitis in mice. *Gastroenterology* 2017;153:1363–77.
18 Corridoni D, Antanaviciute A, Gupta T, *et al*. Single-cell atlas of colonic CD8+ T cells in ulcerative colitis. *Nat Med* 2020;26:1480–90.
19 Schulzke JD, Ploeger S, Amasheh M, *et al*. Epithelial tight junctions in intestinal inflammation. *Ann N Y Acad Sci* 2009;1165:294–300.
20 Gitter AH, Bendfeldt K, Schulzke JD, *et al*. Leaks in the epithelial barrier caused by spontaneous and TNF-alpha-induced single-cell apoptosis. *FASEB J* 2000;14:1749–53.
21 Verhoeckx K, Cotter P, López-Expósito I, *et al*. *The impact of food bioactives on health*. Cham: Springer, 2015.
22 Guerrieri F, Belloni L, D'Andrea D, *et al*. Genome-wide identification of direct HBx genomic targets. *BMC Genomics* 2017;18:184.
23 Belloni L, Pollicino T, De Nicola F, *et al*. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc Natl Acad Sci U S A* 2009;106:19975–9.
24 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
25 Colasante G, Rubio A, Massimino L, *et al*. Direct neuronal reprogramming reveals unknown functions for known transcription factors. *Front Neurosci* 2019;13:283.
26 Okamoto R, Watanabe M. Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease. *J Gastroenterol* 2016;51:11–21.
27 Parmar AS, Lappalainen M, Paavola-Sakki P, *et al*. Association of celiac disease genes with inflammatory bowel disease in Finnish and Swedish patients. *Genes Immun* 2012;13:474–80.
28 Kumagai S, Ohtani H, Nagai T, *et al*. Platelet-derived growth factor and its receptors are expressed in areas of both active inflammation and active fibrosis in inflammatory bowel disease. *Tohoku J Exp Med* 2001;195:21–33.
29 Spadaccini M, D'Alessio S, Peyrin-Biroulet L, *et al*. PDE4 inhibition and inflammatory bowel disease: a novel therapeutic avenue. *Int J Mol Sci* 2017;18.
30 Yang L, Bian Y, Li Z, *et al*. Identification of potential biomarkers and pathways in ulcerative colitis with combined public mrna and mirna expression microarray data analysis. *J Gastrointest Oncol* 2019;10:847–58.
31 Drobin K, Assadi G, Hong M-G, *et al*. Targeted analysis of serum proteins encoded at known inflammatory bowel disease risk loci. *Inflamm Bowel Dis* 2019;25:306–16.
32 Senapati S, Ho SB, Sharma P, *et al*. Expression of intestinal MUC17 membrane-bound mucin in inflammatory and neoplastic diseases of the colon. *J Clin Pathol* 2010;63:702–7.
33 Nielsen CM, Williams J, van denGR, *et al*. Hh pathway expression in human gut tissues and in inflammatory gut diseases. *Lab Invest* 2004;84:1631–42.
34 Lu Y, Li X, Liu S, *et al*. Toll-like receptors and inflammatory bowel disease. *Front Immunol* 2018;9:72.
35 D'Alessio S, Ungaro F, Noviello D, *et al*. Revisiting fibrosis in inflammatory bowel disease: the gut thickens. *Nat Rev Gastroenterol Hepatol* 2022;19:169–84.
36 Houshyar Y, Massimino L, Lamparelli LA, *et al*. Going beyond bacteria: uncovering the role of archaeome and mycobiome in inflammatory bowel disease. *Front Physiol* 2021;12:783295.
37 Mukhopadhya I, Segal JP, Carding SR, *et al*. The gut virome: the "missing link" between gut bacteria and host immunity? *Therap Adv Gastroenterol* 2019;12:1756284819836620.
38 Zuo T, Lu X-J, Zhang Y, *et al*. Gut mucosal virome alterations in ulcerative colitis. *Gut* 2019;68:1169–79.
39 Bjornevik K, Cortese M, Healy BC, *et al*. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 2022;375:296–301.
40 Lau KCK, Burak KW, Coffin CS. Impact of hepatitis B virus genetic variation, integration, and lymphotropism in antiviral treatment and oncogenesis. *Microorganisms* 2020;8:1470.
41 Michalak TI, Mulrooney PM, Coffin CS. Low doses of hepadnavirus induce infection of the lymphatic system that does not engage the liver. *J Virol* 2004;78:1730–8.
42 Lew YY, Michalak TI. In vitro and in vivo infectivity and pathogenicity of the lymphoid cell-derived woodchuck hepatitis virus. *J Virol* 2001;75:1770–82.
43 Mulrooney-Cousins PM, Michalak TI. Repeated passage of wild-type woodchuck hepatitis virus in lymphoid cells does not generate cell type-specific variants or alter virus infectivity. *J Virol* 2008;82:7540–50.
44 Korba BE, Cote PJ, Gerin JL. Mitogen-induced replication of woodchuck hepatitis virus in cultured peripheral blood lymphocytes. *Science* 1988;241:1213–6.
45 Bencze D, Fekete T, Pázmándi K. Type I interferon production of plasmacytoid dendritic cells under control. *Int J Mol Sci* 2021;22:4190.
46 Rosales C. Neutrophil: a cell with many roles in inflammation or several cell types? *Front Physiol* 2018;9:113.
47 Deng Q, Luo Y, Chang C, *et al*. The emerging epigenetic role of CD8+T cells in autoimmune diseases: a systematic review. *Front Immunol* 2019;10:856.
48 Gordon M, Sinopoulou V, Grafton-Clarke C, *et al*. Antibiotics for the induction and maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev* 2022;5:CD013743.
49 Chaturvedi VK, Singh A, Dubey SK, *et al*. Molecular mechanistic insight of hepatitis B virus mediated hepatocellular carcinoma. *Microb Pathog* 2019;128:184–94.
50 Massimino L, Lovisa S, Antonio Lamparelli L, *et al*. Gut eukaryotic virome in colorectal carcinogenesis: is that a trigger? *Comput Struct Biotechnol J* 2021;19:16–28.
51 Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. *Antiviral Res* 2020;182:104925.
52 Abegunde AT, Muhammad BH, Bhatti O, *et al*. Environmental risk factors for inflammatory bowel diseases: evidence based literature review. *World J Gastroenterol* 2016;22:6296–317.

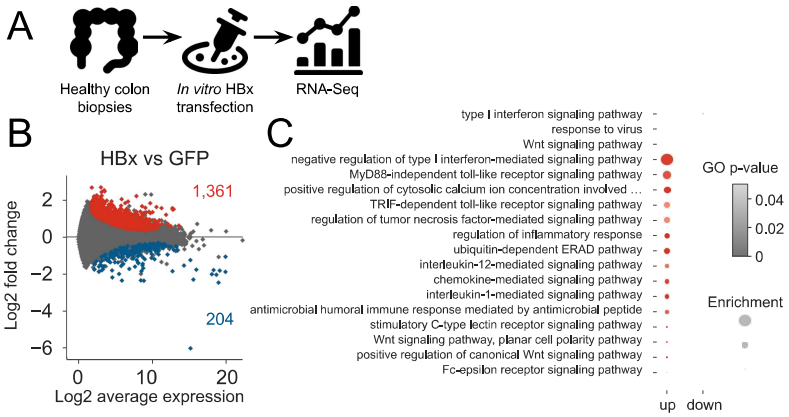
Supplementary figure 1



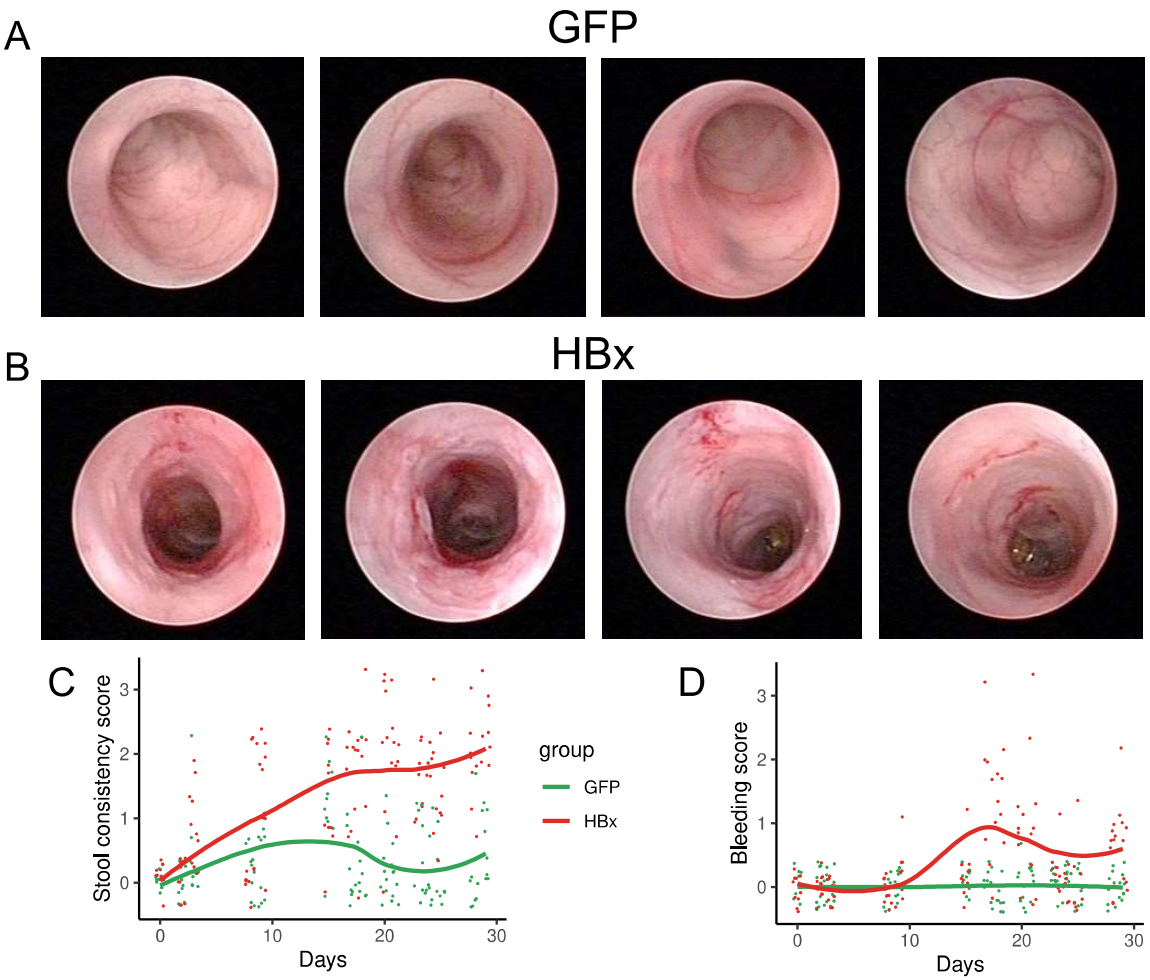
Supplementary figure 2



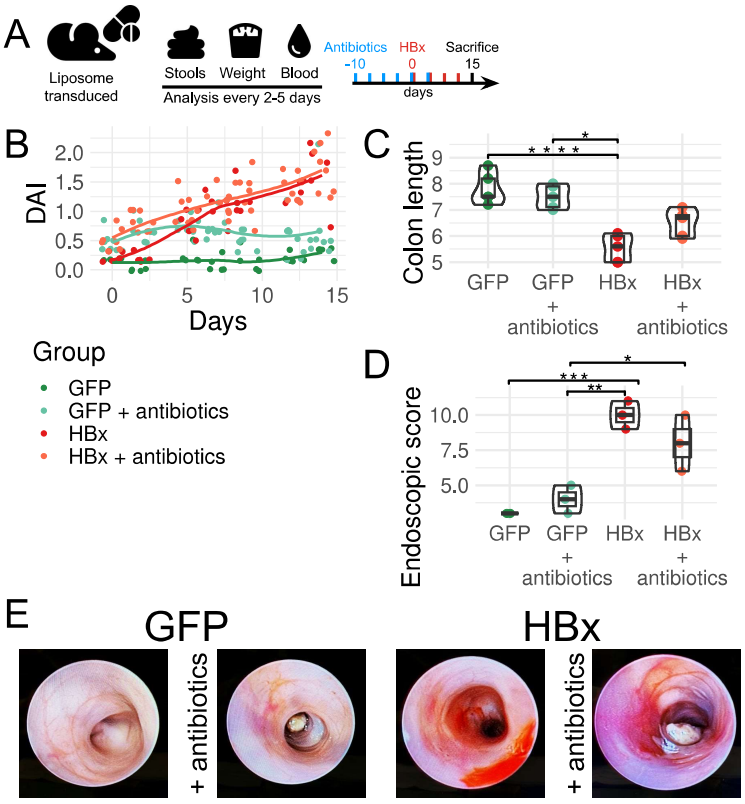
Supplementary figure 3



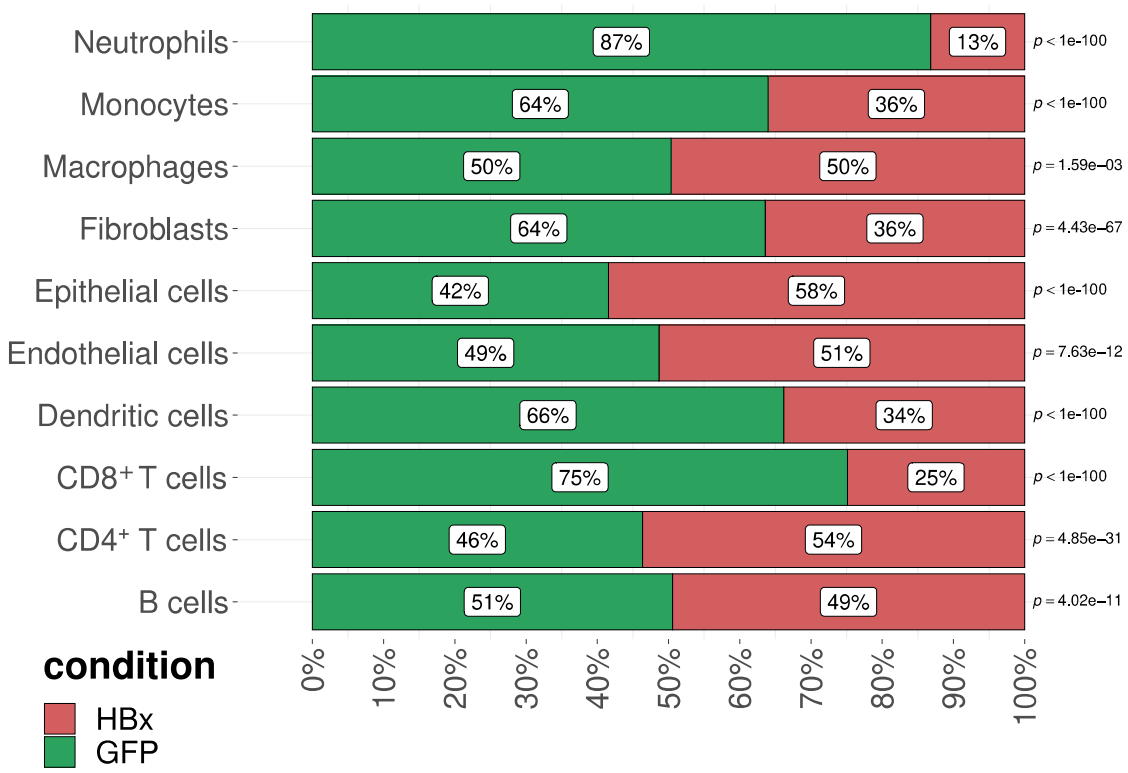
Supplementary figure 4



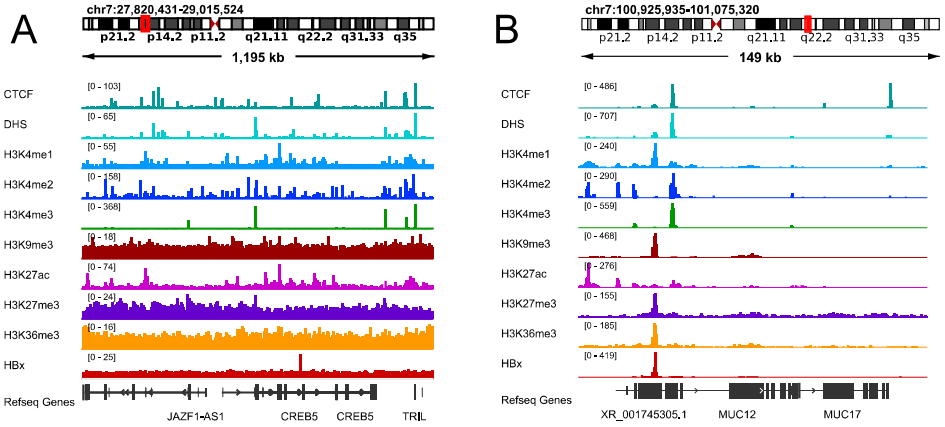
Supplementary figure 5



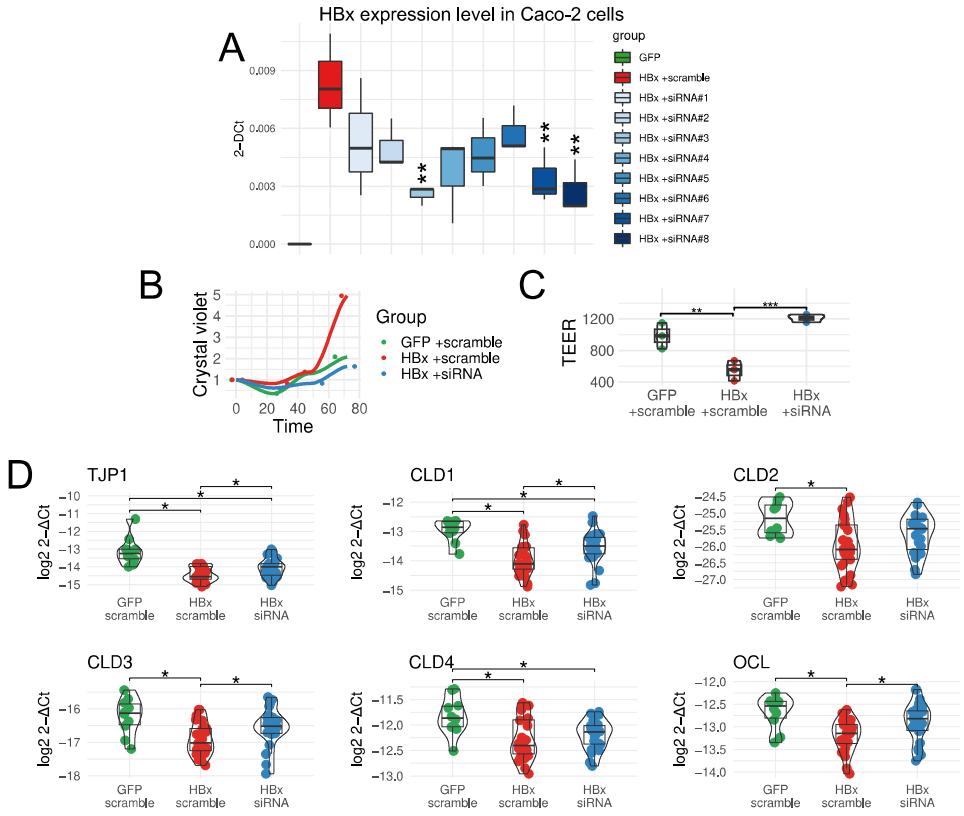
Supplementary figure 6



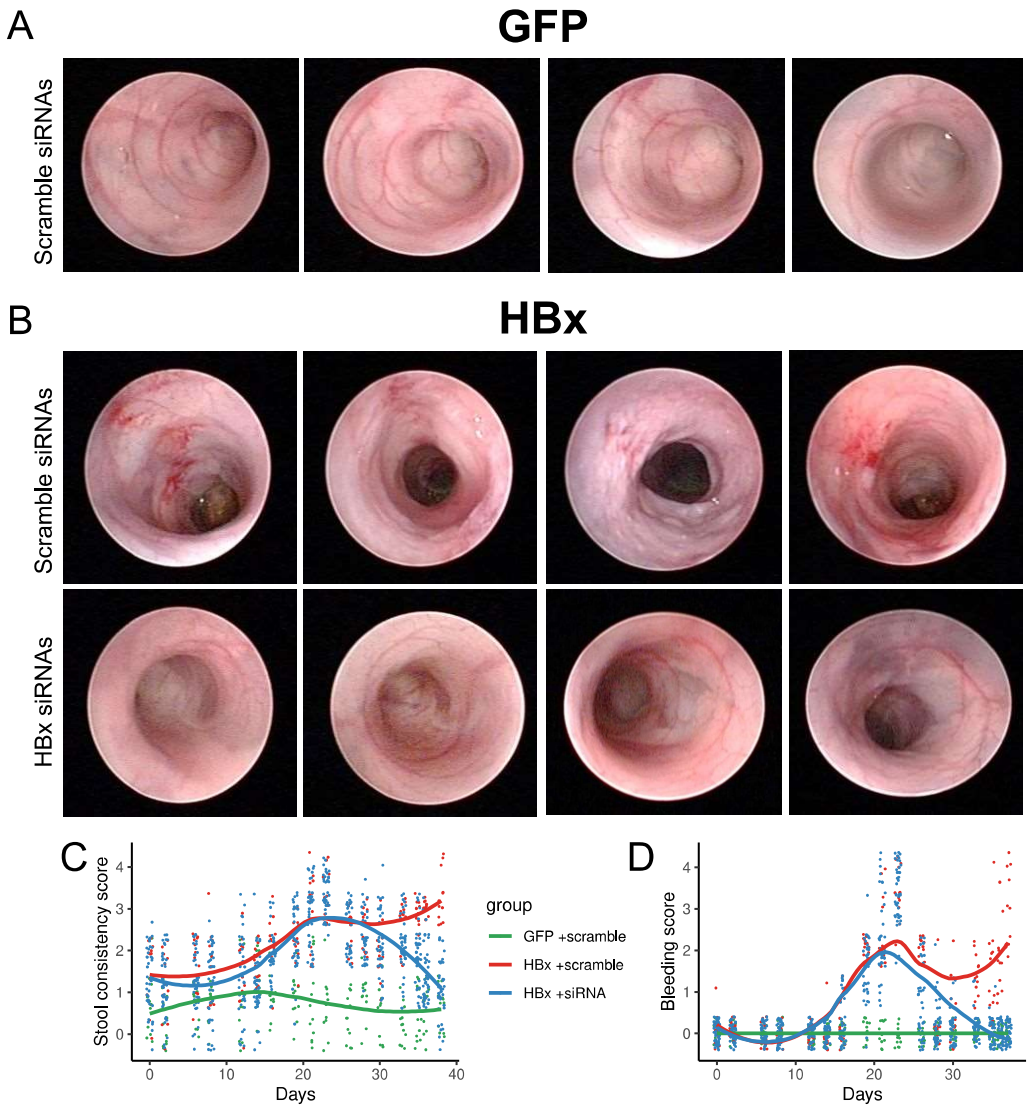
Supplementary figure 7



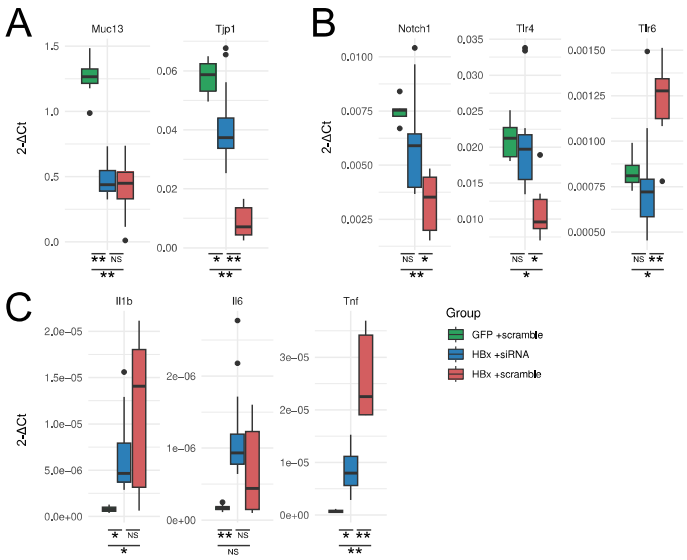
Supplementary figure 8



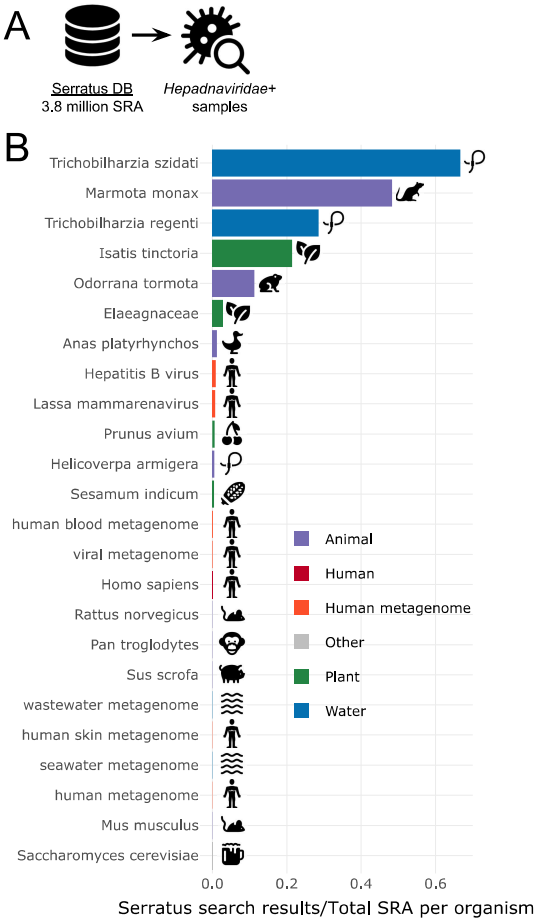
Supplementary figure 9



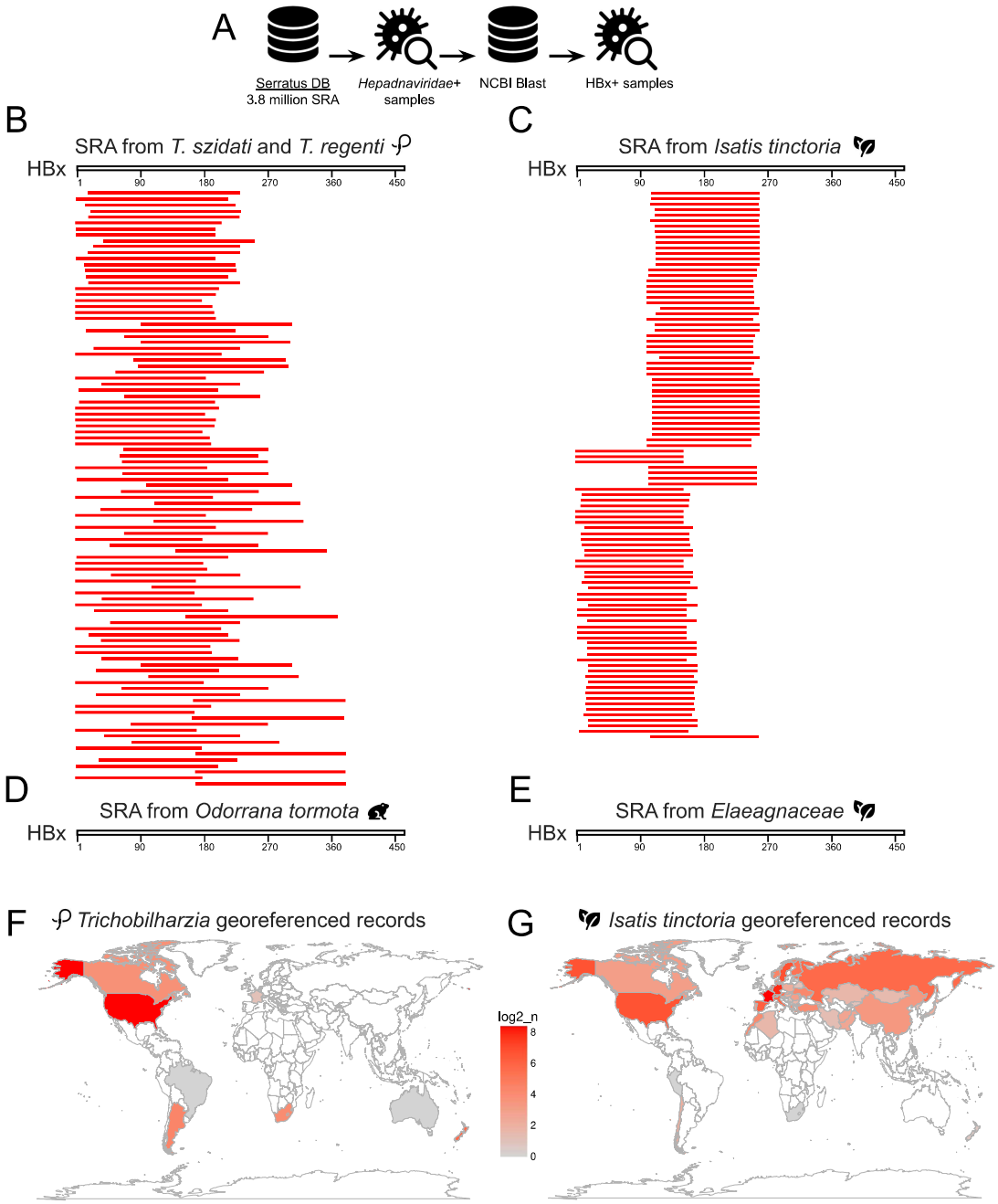
Supplementary figure 10



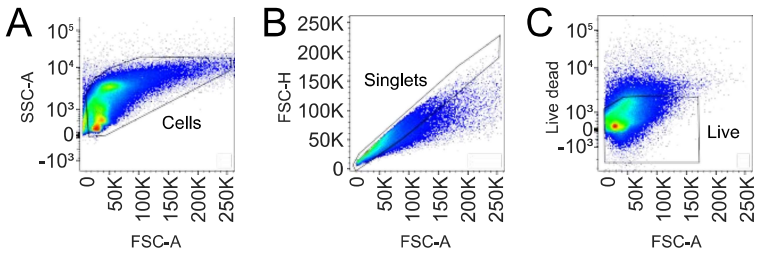
Supplementary figure 11



Supplementary figure 12



Supplementary figure 13



Supplementary results

HBx may be acquired by zoonotic spillover

To explain why HBx may be present in UC patients, we supposed that HBx positivity may be acquired after environmental exposure to this viral entity, for example, during an event of zoonotic spillover.

To corroborate this hypothesis reads belonging to the *Hepadnaviridae* family were searched with Serratus[1] within 3.8 million NGS runs from the NCBI SRA data repository (https://www.ncbi.nlm.nih.gov/sra), normalized on the total runs per species (**Supplementary Figure 11A**). Interestingly, half of the runs from *Trichobilharzia*, a notorious human zoonotic agent colonizing water during its zoonotic life cycle [2,3], were found positive, followed by those from *Isatis tinctoria*, a flowering plant found worldwide, and *Odorrana tormota* (Chinese torrent frog) (**Supplementary Figure 11B**). A high number of runs belonging to *Marmota monax* (eastern woodchuck) was highly expected, as this animal is naturally infected with Woodchuck Hepatitis Virus (WHV)[4] (Supplementary Figure 11B). All these organisms, including humans, may accidentally get in contact with *Trichobilharzia*-contaminated water during their lifespan. As mentioned in the main text, it is noteworthy that only mammalian-infecting *Orthohepadnavirus* carries the HBx ORF, thus raising concerns regarding the possibility that *Trichobilharzia*, or *the Isatis tinctoria,* might be at the basis of the HBx zoonotic transmission. Therefore, we searched for the HBx sequence within the runs (**Supplementary Figure 12A**), and we interestingly found that both *Trichobilharzia* and the plant *Isatis tinctoria*, but not *Odorrana tormota*, nor *Elaeagnaceae*, were carrying the HBx encoding sequence (**Supplementary Figures 12B-12E**).

To evaluate the global distribution of Trichobilharzia and Isatis tinctoria, the Global Biodiversity Information Facility database (https://www.gbif.org/) was interrogated to map 404 georeferenced records reported from 1949 to 2022, and we found Thrichobilharzia spreading in USA, New Zealand, Argentina, South Africa, Canada, North America, East Boreal, France, Asia, Australia, Brazil, Iceland, United Kingdom (Supplementary Figure

12F and Supplementary Table 1), while Isatis tinctoria was highly widespread in Europe, including Italy, and USA (Supplementary Figure 12G and Supplementary Table 2).

It has been recently reported that the highest prevalence/incidence of UC occurs in the USA and European countries[5,6]. While being aware this does not still represent a causal link between the environmental exposure to these specific factors and disease pathogenesis, we observed that the UC global burden overlaps the geographical distribution of the Trichobilharzia and Isatis T. spreading, further supporting a possible association between these events.

Nevertheless, further association studies, and/or systematic reviews, and meta-analyses are required to demonstrate this hypothesis and exclude any *epiphenomenon*.

Supplementary methods

Biopsy collection and patient inclusion criteria

Endoscopic biopsies collected from patients with UC undergoing endoscopy were used. The degree of disease activity active state was diagnosed by physicians based on clinical, endoscopic, and histological criteria. Clinical, endoscopic, and histological severity were assessed using the Mayo clinical score, Mayo endoscopic score[9], and the Riley histological scoring system[10], respectively. Active mucosal biopsies were defined by endoscopic Mayo score ≥2 and Riley Index ≥3.

Healthy endoscopic biopsies were negative for anti-HBV antibodies obtained from patients undergoing routine check-ups for non-IBD-related diseases.

All patients were tested according to clinical practice guidelines for anti-HBcAg total (IgM+IgG) HBs levels by chemiluminescent immunoassay (Abbott, Germany) and liver function enzymes. This was a multicenter observational retrospective study enrolling patients from 2008 to 2018 in two clinical centers (Humanitas Research and Clinical Institute, Milan, Italy, and Fondazione IRCCS Casa Sollievo Della Sofferenza, Foggia, Italy) and approved by the independent ethic committees (Protocol number CE Humanitas ex D.M.8/2/2013-183/14, protocol number CE Casa Sollievo Della Sofferenza: 12701/2008). All subjects provided written informed consent.

In vivo experiments

Female C57BL/6 mice were maintained in a specific-pathogen-free (SPF) animal facility and used at 8 weeks of age. All experiments were performed following the guidelines established in the Principle of Laboratory Animal Care (directive 86/609/EEC) and were approved by the Italian Ministry of Health. The procedures involving mice conformed to institutional guidelines in agreement with national and international law and were approved by the ethics committee of the Humanitas Research Hospital. On the basis of our experience with animal models and according to animal-welfare policy (directive 86/609/EEC), which strongly suggests the use of a limited number of animals, we estimated that two experiments with n=6 mice per group would allow us to reach statistical significance. Mice were randomized, and experimental mice allocation and evaluation of clinical disease parameters by a dedicated lab technician were blinded.

HBV screening of patients

Anti-HBc antibodies are the best marker for prior exposure to HBV as they become detectable before anti-HBs and remain positive indefinitely after[11]. Moreover, and practical for our study purposes, anti-HBc antibodies are not influenced by prior vaccination (unlike anti-HBs).

Patients positive for anti-HBc antibodies (IgM+IgG) have encountered the virus at a certain point in life and either cleared the infection (roughly in 95% of cases) or have an ongoing chronic infection (5%)[12]. To rule out the latter, anti-HBs testing was performed in all anti-HBc-positive patients, and no chronic infection was found. Moreover, no patient in the whole cohort, regardless of HBV serology, had deranged liver function tests (AST and ALT), further excluding infections, especially acute ones. In addition, the clinical records of all patients were reviewed and no mention was found of any liver abnormalities, HBV infection, other viral hepatitis, or HIV.

Reverse transcription-polymerase chain reaction (RT-PCR) and Sanger sequencing

Total RNA was extracted from the fresh frozen biopsy samples using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA; cat. #217004) and gentleMACS dissociator (Miltenyi Biotec, Bergisch-Gladbach, Germany; cat. #130-096-335) according to manufacturers' instructions. RNA quality was evaluated on the Bioanalyzer 2100 microcapillary electrophoresis system (Agilent Technologies, Palo Alto, CA, USA). Genomic DNA was extracted from endoscopic biopsies or blood by using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA, cat. #69504) according to the manufacturer's instructions.

A mixture containing 0.1µg of total RNA from each sample was reverse transcribed for 10 min at 25°C, and 2 h at 37°C using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA; cat. #4368814).

For GAPDH, PCR was performed in a final volume of 25 µL containing 2.5 µL 10× PCR Buffer, 2.5mM dNTPs, 25mM MgCl2, 15pM of forward (5′-GAAGGTGAAGGTCGGAGTC-3') and reverse (5′-GAAGATGGTGATGGGATTTC-3') primers, 0.75U AmpliTaq Gold polymerase (Thermo Fisher Scientific, Waltham, MA, USA; cat. #N808-0245) and 1µL cDNA. Cycling PCR conditions consisted of an initial 10 min denaturation step at 94°C, followed by 35 cycles of 94°C for 1 min, 58°C for 1 min, and 72°C for 1 min, with a final extension at 72°C for 10 min.

For HBx detection, PCR reactions (25µL) were performed in 1X Phusion GC Buffer, 3.5mM $MgCl_2$, 0.1U Phusion High-Fidelity (Thermo Fisher Scientific, Waltham, Massachusetts, USA; cat. #F530L), 0.2mM for forward (5′-GTGTGCACTTCGCTTCACCT -3') and reverse (5′-GTGCTGGTGAACAGACCAAT-3') primers, 0.5 mM dNTPs, DMSO 4% and 1 µL cDNA (or 25 ng of genomic DNA, when trios were analyzed). After initial denaturation for 30 sec at 98°C, the reaction was subjected to amplification for 12 cycles at 98° C for 30 sec, at 68°C (ending at an annealing temperature of 62°C) for 30 sec and at 72°C for 30 sec. It follows 25 cycles: 98°C for 30 sec, 62°C for 30 sec, and 72°C for 30 sec. The final extension was 72°C for 10 min (**Supplementary Figure 1D**). PCR products were visualized by GelRed Nucleic

Acid Stain (Biotium, Fremont, CA, USA, cat. #41003) on 3% agarose gels in presence of a pGEM DNA marker ladder (Promega, Madison, WI, USA, cat. #G1741).

A sequencing reaction was performed on positive samples, including the negative for further control to verify the correspondence between the amplicons and the selected HBx sequence. Briefly, PCR products were purified with a vacuum system (Millipore, Bedford, MA, USA, cat. #MSNU03050) according to the manufacturer's recommendation. The amplicons were sequenced from both ends using an aliquot (3.2pM) of the PCR reaction primer in presence of BigDye Terminator Cycle Sequencing Kit v. 1.1 (Thermo Fisher Scientific, Waltham, MA, USA, cat. #4337450) according to the manufacturer's recommendation. After purification by using centrisep columns (Princeton Separations, Adelphia, NJ, USA, cat. #CS-901), sequencing reactions were loaded onto 3500 DX Genetic Analyzer capillaries (Applied Biosystems, Waltham, MA, USA**)** and analyzed using the Sequencing Analysis software v5.4.

Sanger sequencing results are available at OP978007-OP978010 in NCBI GenBank.

Quantitative real-time PCR

Total RNA was extracted from Caco-2 cells and mouse colon tissues using the PureZOL RNA isolation reagent (Bio-Rad, Hercules, CA, USA, cat. #7326890) according to the manufacturer's instructions. RNA retrotranscription was performed with High Capacity cDNA Reverse Transcription Kits (Applied Biosystems, Waltham, MA, USA, cat. #11732641001). The quantitative real-time PCR reaction was performed with SYBR® Green Real-Time PCR Master Mix (Thermo Fisher Scientific, Waltham, Massachusetts, USA, cat. #4385610) according to the manufacturer's instructions. The primer pairs used are summarized in **Supplementary Table 4** and **Table 5**. The reactions were performed and analyzed on the ViiA7 Real-Time PCR System (Applied Biosystems, Waltham, Massachusetts, USA). GAPDH expression was used as the housekeeping gene. Data were calculated using the 2^-ΔCt method.

Serratus and NCBI Blast

The Serratus open-science viral discovery platform was queried for *Hepadnavirus* in March 2022. The total number of SRA reads per species was obtained from NCBI SRA and used to normalize Serratus hits. SRA reads belonging to the top 10 enriched species were then subjected to NCBI Blast[7] to search for the *HBx* coding sequence. Plots were made with ggplot2[8].

Biodiversity data analysis

Georeferenced records for Trichobilharzia and Isatis tinctoria were downloaded from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/), reshaped with tidyverse[9], and plotted with ggplot2[8].

Molecular cloning

To generate the EF1α-HBx-IRES-GFP lentiviral vector, the HBx encoding sequence was cloned downstream of the EF1α promoter and upstream of the IRES sequence in the pHIV-IRES luciferase vector (Addgene, Watertown, MA, USA, cat. plasmid #21375), previously modified by substituting the luciferase sequence with the GFP. The V5-tagged HBx-carrying lentiviral vector was obtained by inserting upstream the V5 tag in the HBx encoding sequence without the stop codon in the pCAG-V5 vector, kindly given by Dr. Alessandro Sessa of the San Raffaele Research Institute. Then, the HBx-V5 sequence was inserted downstream of the EF1α promoter-carrying lentiviral vector. The EF1α-GFP vector was used as a control, as previously described[10].

Lentiviral transduction of endoscopic biopsies and cells

Lentiviral particles harboring V5-tagged HBx, HBx-IRES-GFP, or GFP were produced by transient transfection of 293T cells according to standard protocols[11] and as previously described[10]. This cell line is competent for replicating vectors carrying the SV40 region of replication. Mucosal biopsies collected from healthy donors were washed twice in RPMI medium (Lonza, Basel, Switzerland, cat. #12-167Q) supplemented with Fetal Bovine Serum (Euroclone, Pero, Italy, cat. #ECS0186L), UltraGlutamine™ I (Lonza, Basel,

Switzerland, cat. #BE17-605E/U1, final concentration: 2mM), HEPES Buffer (Euroclone, Pero, Italy, cat. #ECM0180D, final concentration: 10mM), and with an antibiotics mixture of Penicillin-Streptomycin-Amphotericin B (Lonza, Basel, Switzerland, cat. #17-745E, final concentration: 100mU/mL; 100mg/mL; 0.25mg/mL), Gentamicin Sulfate (Lonza, Basel, Switzerland, cat. #17-518Z, final concentration: 0.5mg/mL), and Puromycin (Invivogen, San Diego, CA, USA cat. #ant-pr-1, final concentration: 0.1mg/mL). Each biopsy was washed, plated, and transduced with lentiviral particles carrying either the HBx-IRES-GFP or the GFP encoding sequences as control. Transduced biopsies were maintained at 37°C for 48 hours and then washed twice in phosphate-buffered saline (PBS) 1x. Biopsies consequently underwent digestion using collagenase type 2 (Worthington Biochemical Corporation, Lakewood, NJ, USA, cat. #LS004176, final concentration: 7.5mg/mL), followed by gentle compression through 70µm and 100µm cell strainer (Falcon, Corning, NY, USA, cat. #352350, 352360) to extrude cell clumps. The obtained cellular pool was collected after centrifuge and frozen for the RNA-Seq analysis.

For Caco-2, 1×10^6 cells were transduced with lentiviral particles carrying either the V5-tagged HBx-encoding sequence or the HBx-IRES-GFP or the empty vector as control. For siRNA treatment, Caco-2 cells were transfected with 0.2nmol of HBx-targeting siRNAs or the relative scramble controls, by using Lipofectamine 2000® (Life Technology, Carlsbad, CA) according to the manufacturer's instructions. HBx-targeting siRNA sequences are listed in Supplementary Table 6. For the experiments, siRNA#3, #7, and #8 were used.

Prior to ChIP analysis, cells were then maintained confluent in culture for 72 hours and then processed.

For epithelial organoid transduction, cells were collected in a 15mL tube and treated with a medium containing lentiviral particles carrying either the HBx-IRES-GFP or GFP-encoding sequence for 4 hours at 37°C. After incubation, organoids were washed and centrifuged twice and moved to a 5mL round bottom FACS tube for the following analysis.

Two independent experiments with endoscopic biopsies, Caco-2 cells, and epithelial organoids were performed at least in triplicate.

Epithelial organoid derivation

Healthy donor-derived endoscopic biopsies were collected in PBS1X without calcium and magnesium on ice and washed 3 times in 10mL of Complete Chelating Solution (1X CCS), made of Incomplete Chelating Solution (ICS 5X, 500mL MilliQ H2O, 2.7g KH2PO4, 0.3g KCl, 2.49g Na2HPO4-2H2O, 14g NaCl, 37.5g Sucrose, 25g D-Sorbitol) diluted 1:5 in MilliQ H2O and supplemented with DTT 0.5µM (PanReac Applichem, Chicago, IL, USA, Cat.#A9448). Biopsies were then incubated at 4°C for 30 minutes in a solution of CCS containing EDTA 0.5M. After digestion, crypts from biopsies were obtained by gentle mechanical dissociation. Crypts were subsequently transferred into a 15mL tube and centrifuged at 800RPM for 5 minutes. The obtained crypt-enriched fraction was resuspended in Matrigel (Corning, New York, NY, USA, cat. #356231)-advanced DMEM/F12 medium (Lonza, Basel, Switzerland, cat. #BE04687F/1) to a ratio of 1:1 and then seeded in a 24-well plate and incubated for 20 minutes at 37°C/5%CO_2 to allow the Matrigel dome to solidify. Afterward, 700uL of WENR culturing medium (**Supplementary Table 7**) was supplemented with Y27632 (Tocris, Bristol, UK, cat. #1254) and CHIR99201 (Tocris, Bristol, UK, cat. #4423) and added to the cell culture. CHIR 99021 and Y27632 were added only on the first and the second day, then only WENR was used.

RNA extraction, library preparation, and transcriptomics

RNA was extracted from human and mouse colon tissues using the PureZOL RNA isolation reagent (Bio-Rad, Hercules, CA, USA, cat. #7326890) according to the manufacturer's instructions. Library preparation and RNA-Seq were performed at the Galseq Srl NGS facility in Milan. For the RNA-Seq, libraries were generated with the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® and NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Ipswich, MA, USA) and sequenced on Illumina Hiseq instrument using a 2x150 paired-end strategy, with a depth of 25 million clusters/sample.

Transcriptomics analyses were performed as previously described[12]. Briefly, FASTQ reads were quality-checked and adaptor-trimmed with Trimmomatic[13] and then mapped to the human hg38 reference genomes with STAR[14]. Differential gene expression and

microbiota abundance were performed with DESeq2[15]. Human functional enrichment analysis was performed with geneSCF[16]. Plots were made with ggplot2[8] or IGV[17]. Raw and processed data were deposited into the NBCI GEO repository with accession number GSE204665.

HBx-induced colitis experiments

8-week-old female C57BL/6 mice were intrarectally injected with 50µl of liposomes conjugated with 1µg/animal of HBx-IRES-GFP- or GFP-encoding plasmids for 20 days. Of note, after about 20 days, the concentration of both plasmids was doubled since mice likely developed resistance to the HBx insult. This is well in line with the concept that HBx, similarly to other viral-derived factors, can coexist with the host because of a sort of immunization process[18]. For HBx silencing *in vivo,* liposome conjugated HBx-targeting or scramble siRNAs were administered every two days by intramucosal injections starting from day 24. The efficacy of the siRNA-mediated HBx silencing was performed in HBx-overexpressing Caco-2 cells undergoing liposome-conjugated siRNA transduction as previously reported[19]. The siRNA sequences are listed in **Supplementary Table 6**. For the experiments, siRNA#3, #7, and #8 were used.

Liposome-conjugated plasmids were made in Opti-MEM™IReduced Serum Medium (Thermo Fisher Scientific, Waltham, MA, USA, cat. #31985062) using Lipofectamine™ 2000 Transfection Reagent (Thermo Fisher Scientific, Waltham, MA, USA, cat. #11668019) according to the manufacturer's instructions. Injections were performed every other day, for a total of three injections a week. Before each injection, the body weight, presence of gross blood in the feces, and stool consistency were recorded. The DAI was determined by scoring changes in weight loss (0=none, 1=1%–5%, 2=5%–10%, 3=10%–20%, 4=20%), stool consistency (0=normal, 2=loose, 4=diarrhea), and rectal bleeding (0=normal, 2=occult bleeding, 4=gross bleeding). For monitoring colitis, a high-resolution mouse video endoscopic system was developed and used as previously described[20]. At the end of the experiment, mice were sacrificed, colon length was measured, and organs were collected for further analysis.

To deplete the gut microbiota, 8 weeks female C57BL/6 mice were treated with a cocktail of broad-spectrum antibiotics (ampicillin 1g/L, neomycin sulfate 1g/L, vancomycin 0.5g/L, and metronidazole 0.2g/L, Sigma-Aldrich Co. Ltd, MO, USA) dissolved in drinking water for 15 consecutive days, as previously described[21]. The drinking solution was provided ad libitum and renewed every 2 days. On day 10 of antibiotic treatment mice started to receive intrarectal injections of HBx-IRES-GFP- or GFP-encoding plasmids, every 2 days for 15 days, as described in the Supplementary information ("HBx-induced colitis experiments").

Fluorescence-activated cell sorting (FACS)

Freshly collected UC patient-derived biopsies were incubated in collagenase type 2 (Worthington Biochemical Corporation, Lakewood, NJ, USA, cat. #LS004177 final concentration: 7.5mg/mL) then smashed in a 100 µm cell strainer (Falcon, Corning, NY, USA, cat. #352360) and 70µm cell strainer (Falcon, Corning, NY, USA, cat. #352350) using complete RPMI 1640 (Lonza, Basel, Switzerland, cat. #12-167Q). Cell suspensions were then washed in Sorting buffer (PBS/HBSS w/o Calcium and Magnesium supplemented with 10% fetal calf serum) and incubated with the Aqua stain Qdot 525 (Thermo Fisher, Waltham, MA, USA, L34957; 0,5:100) at room temperature in the dark for 20 minutes. Cells were sorted to isolate and collect the following mucosal cell populations based on the expression of specific cell markers: Endothelial cells (CD31 FITC; BD Bioscience, Franklin Lakes, NJ, USA, cat. #555545; 20:100), Fibroblasts (CD90 APC; BD Bioscience, Franklin Lakes, NJ, USA, cat. #561971; 5:100), Epithelial cells (EpCAM PE; BD Bioscience, Franklin Lakes, NJ, USA, cat. #347198; 20:100), leukocytes (CD45 PE CF594; BD Bioscience, Franklin Lakes, NJ, USA, cat. #3346961; 2,5:100), $CD4^+$ T cells (CD4 PE-Cy7; BD Bioscience, Franklin Lakes, NJ, USA, cat. #557852; 2,5:100) CD8+ T cells (CD8 BV780; Biolegend, San Diego, CA, USA, cat. #56382; 0,63:100), B cells (CD20 APC-Cy7; BD Bioscience, Franklin Lakes, NJ, USA, cat. #560734; 2,5:100), Macrophages (CD163 BV650; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563888; 5:100), Myeloid cells (CD11b BV605; BD Bioscience, Franklin Lakes, NJ, USA, cat. #562721; 5:100), Dendritic cells (CD11c BV711; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563130; 5:100), Neutrophils (CD66b; BD Bioscience, Franklin

Lakes, NJ, USA, cat. #562940; 5:100). Soon after the staining, cells were sorted using BD FacsAriaIII and DIVA software (BD, Franklin Lakes, NJ, USA) and the different cell populations underwent library preparation and RNA-Seq at the Galseq Srl NGS facility in Milan.

For murine samples, freshly collected colonic mucosa were incubated in collagenase type 2 (Worthington Biochemical Corporation, Lakewood, NJ, USA, cat. #LS004177 final concentration: 7.5mg/mL) and then smashed in a 100µm cell strainer (Falcon, Corning, NY, USA, cat. #352360) and 70µm cell strainer (Falcon, Corning, NY, USA, cat. #352350) using complete RPMI 1640 (Lonza, Basel, Switzerland, cat. #12-167Q).

The cell suspension was then washed in complete RPMI 1640, and then in FACS Buffer (PBS/HBSS w/o Calcium and Magnesium supplemented with 5% fetal calf serum). Cells were then counted. 1×10^6 cells/colonic cell preparation were incubated with the Aqua stain Qdot 525 (Thermo Fisher, Waltham, MA, USA, L34957; 0,5:100) in the dark for 20 minutes at room temperature. To stain the subpopulations and determine the immunophenotype, the following antibodies were used: Epithelial cells (EpCam PE; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563477; 1:100), Endothelial cells (CD31 BV421; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563356; 3:100), Fibroblasts (CD90 APC; BD Bioscience, Franklin Lakes, NJ, USA, cat. #553007; 0,5:100), Leukocytes (CD45 PE CF594; BD Bioscience, Franklin Lakes, NJ, USA, cat. #562420; 0,625:100), $CD4^+$ T cell (CD4 PE-Cy7; BD Bioscience, Franklin Lakes, NJ, USA, cat. #552775; 1,25:100), $CD8^+$ T cell (CD8 BV780; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563332; 1,25:100), B Cells (CD20 APC-Cy7; Biolegend, San Diego, CA, USAcat. #150418; 2:100), Monocytes (Ly6C PerCP Cy5.5;BD Bioscience, Franklin Lakes, NJ, USA, cat. #560525; 1:100), Macrophages (F4/80 BV650; BD Bioscience, Franklin Lakes, NJ, USA, cat. # 743282; 1,25:100), Myeloid cells (CD11b BV605; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563015; 0,625:100), Dendritic cells (CD11c BV711; BD Bioscience, Franklin Lakes, NJ, USA, cat. #563048; 1:100), Neutrophils (Ly6G Alexa Fluor 700; BD Bioscience, Franklin Lakes, NJ, USA, cat. #561236; 1:100). Cells were incubated in the dark for 30 minutes at 4˚C, followed by fixation and permeabilization with Fixation/Permeabilization solution (Perm/Wash BD, BD Bioscience, Franklin Lakes, NJ, USA cat. #557885) for 20 minutes at 4˚C in the dark. Then anti-GFP Alexa Fluor 488

(Biolegend, San Diego, CA, USA, cat. #3380008; 1:100) was used to amplify the signals from GFP cells. Samples were acquired with BD LSRII Fortessa equipped with Diva Software (BD, Franklin Lakes, NJ, USA). Compensated and biexponentially transformed FCS of acquired samples, singlets, and living cell files (**Supplementary Figures 13A-13C**) were exported for subsequent multidimensional scaling, clustering, and cell-type identification performed through cytoChain[18].

Immunofluorescence analysis

For immunofluorescence analysis of epithelial organoids, the cells were fixed with a solution of PBS with 4% PFA for 30 minutes on ice and then incubated in blocking buffer (PBS1x with 5% with 10% Normal Goat Serum (Thermo Fisher Scientific, Waltham, MA, USA, Cat. #31872) and 0,5% Triton X-100 (Sigma-Aldrich, Darmstadt, Germany, cat. #X100-100mL) 2 hours at room temperature. After washing, cells were incubated with mouse anti-Human EpCam (Santa Cruz Biotechnology, Santa Cruz, CA, USA, cat. #SC-66020; 1:100) and rabbit anti-GFP (Thermo Fisher Scientific, Waltham, MA, USA, cat. #A-21311; 1:500) primary antibodies diluted in blocking buffer overnight at 4°C. Thus, samples were washed and then incubated with secondary antibodies Alexa Fluor 594 dye Goat anti-mouse and Alexa Fluor 488 dye goat anti-rabbit (Thermo Fisher Scientific, Waltham, MA, USA, Cat. #A11005 and #A11034; 1:1.000 both) for 2 hours at room temperature at dark. DAPI (Invitrogen, Waltham, MA, USA, 1:25.000) was added at 5ug/ml concentration in PBS 1x for 15 minutes. Organoids were mounted with Vectashield (Vector Laboratories, Newark, NJ, USA, Cat.#H-1000) on glass and analyzed with a confocal microscope.

Proliferation and permeability assays

For proliferation assay, HBx-IRES-GFP (HBx) transduced and the GFP control Caco-2 cell lines were seeded in 96-well cell culture plates (1×10^3 cells/well) in 10% FBS DMEM medium supplemented with antibiotics. On days 0, 1, 2, and 3 cells were stained with 0.2% crystal violet (Sigma-Aldrich, Darmstadt, Germany, cat. #C0775) dissolved in ethanol. The uptake of dye by cells on plates was eluted with 33% acetic acid in water. Plates were gently shaken for 20 min and the absorbance at 595nm was measured by a

Versamax microplate reader (Molecular Devices). The optical density of each sample was normalized to day 0 of each sample.

For cell permeability assay, HBx and GFP transduced cells were seeded on 0.4μm pore-size Transwell Permeable Supports (Corning Costar, Cambridge, MA, USA, cat. #3401), cultured in a complete culture medium, and allowed to form a monolayer, as previously described[10]. After the establishment of a stable monolayer, cells were assayed for permeability using a Millicall-ERS volt-ohm meter (World Precision Instruments, New Haven, CT) to measure the transepithelial electrical resistance (TEER). The percentage of change in the transendothelial electrical resistance was calculated by comparison with the control cells (GFP).

ChIP, ChIP-Seq analysis

Caco-2 raw sequencing data regarding informative chromatin modifications, CTCF binding profile, and DNA hypersensitivity were downloaded from ENCODE[22]. Chromatin immunoprecipitation, library preparation, and ChIP-Seq were performed at the Galseq Srl NGS facility in Milan. Libraries were generated with the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) and sequenced on Illumina Hiseq instrument using a 2x150 paired-end strategy, with a depth of 20 million clusters/sample. Computational analysis was performed as previously described[23]. Briefly, FASTQ reads were quality-checked and adaptor-trimmed with Trimmomatic[13] and then mapped to the human hg38 reference genomes with Bowtie2[24]. ChIP peak calling and gene-centric annotation were performed with MACS3[25] and ChIPseeker[26], respectively. Motif discovery was performed with MEME[27]. Chromatin state correlation and Hidden Markov modeling were performed with bamCoverage[28] and ChromHMM[29], respectively. Plots were made with the ggstatplot[30] R package or IGV[17]. Raw and processed data were deposited into the NBCI GEO repository with accession number GSE204665.

Statistical analysis

The sample size was calculated based on the effect size derived from preliminary data. The Shapiro-Wilk normality test assesses homoskedasticity. Dispersions were displayed with ggplot2[8] as box plots (median +/- 1.58*IQR/$\sqrt{n}$), or trend lines for longitudinal analyses. Differences between the groups were calculated in GraphPad Prism and R, using a two-tailed t-test, one-way ANOVA, or two-way ANOVA, with Tukey's or Bonferroni's post-hoc correction, where appropriate. Multivariate analysis through Pearson's R^2 correlation coefficients and plotting were performed with the ggstatplot R package[30]. Tests were considered statistically significant when the P-value ≤ 0.05.

Supplementary Tables

Supplementary Table 1. GBIF georeferenced records for Trichobilharzia.

Region	n
USA	202
New Zealand	38
Argentina	16
South Africa	13
Canada	11
North America	6
East Boreal	3
France	2
Asia	1
Australia	1

Brazil	1
Iceland	1
United Kingdom	1

Supplementary Table 2. GBIF georeferenced records for Isatis tinctoria.

Region	**n**
France	9101
Germany	5455
Netherlands	4149
Sweden	1558
USA	1437
Switzerland	1096
Spain	831
Italy	548
Belgium	505
Russia	501
Finland	308
Estonia	306
Austria	250
United Kingdom	147
Norway	130
Denmark	107
Ukraine	103
Morocco	93
Turkey	84
Greece	77
Åland	73
Czechia	63
Luxembourg	52

China	40
Canada	28
Romania	25
Croatia	21
Afghanistan	19
Pakistan	17
Portugal	17
Slovakia	16
Poland	12
Armenia	10
Chile	10
Japan	10
Hungary	9
Bulgaria	7
Georgia	7
Algeria	6
Kazakhstan	6
Mongolia	5
Iran	4
North Korea	4
Belarus	3
Kyrgyzstan	3
Moldova	3
Montenegro	3
Serbia	3
New Zealand	2
Slovenia	2
Iceland	1
Ireland	1

Latvia	1
Lebanon	1
North Macedonia	1
Peru	1
South Africa	1

Supplementary Table 3. Functional annotation of the HBx target genes.

ID	Gene Name	Species	GOTERM_BP_DIRECT
FRMD4B	FERM domain containing 4B(FRMD4B)	Homo sapiens	GO:0090162~establishment of epithelial cell polarity,
LUC7L2	LUC7 like 2, pre-mRNA splicing factor(LUC7L2)	Homo sapiens	GO:0006376~mRNA splice site selection,
CREB5	cAMP responsive element binding protein 5(CREB5)	Homo sapiens	GO:0006357~regulation of transcription from RNA polymerase II promoter,GO:0045893~positive regulation of transcription, DNA-templated,
ESCO1	establishment of sister chromatid cohesion N-acetyltransferase 1(ESCO1)	Homo sapiens	GO:0006275~regulation of DNA replication,GO:0007062~sister chromatid cohesion,GO:0018394~peptidyl-lysine acetylation,GO:0034421~post-translational protein acetylation,
FBLN7	fibulin 7(FBLN7)	Homo sapiens	GO:0007155~cell adhesion,
KBTBD8	kelch repeat and BTB domain containing	Homo sapiens	GO:0006417~regulation of translation,GO:0006513~protein monoubiquitination,GO:0014029~neural crest

	8(KBTBD8)		formation,GO:0014032~neural crest cell development,
LRRCC1	leucine rich repeat and coiled-coil centrosomal protein 1(LRRCC1)	Homo sapiens	GO:0007049~cell cycle,GO:0051301~cell division,
MUC17	mucin 17, cell surface associated(MUC 17)	Homo sapiens	GO:0019725~cellular homeostasis,
PDE4D	phosphodiesterase 4D(PDE4D)	Homo sapiens	GO:0002027~regulation of heart rate,GO:0006198~cAMP catabolic process,GO:0007165~signal transduction,GO:0010469~regulation of receptor activity,GO:0010880~regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum,GO:0019933~cAMP-mediated signaling,GO:0032729~positive regulation of interferon-gamma production,GO:0032743~positive regulation of interleukin-2 production,GO:0032754~positive regulation of interleukin-5 production,GO:0033137~negative regulation of peptidyl-serine phosphorylation,GO:0043951~negative regulation of cAMP-mediated signaling,GO:0045822~negative regulation of heart contraction,GO:0050852~T cell receptor signaling pathway,GO:0060314~regulation of ryanodine-sensitive calcium-release channel activity,GO:0061028~establishment of endothelial barrier,GO:0071320~cellular

			response to cAMP,GO:0071872~cellular response to epinephrine stimulus,GO:0071875~adrenergic receptor signaling pathway,GO:0086004~regulation of cardiac muscle cell contraction,GO:0086024~adrenergic receptor signaling pathway involved in positive regulation of heart rate,GO:1901844~regulation of cell communication by electrical coupling involved in cardiac conduction,GO:1901898~negative regulation of relaxation of cardiac muscle,
PDGFRA	platelet derived growth factor receptor alpha(PDGFRA)	Homo sapiens	GO:0001553~luteinization,GO:0001701~in utero embryonic development,GO:0001775~cell activation,GO:0002244~hematopoietic progenitor cell differentiation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0007204~positive regulation of cytosolic calcium ion concentration,GO:0007275~multicellular organism development,GO:0008210~estrogen metabolic process,GO:0008284~positive regulation of cell proliferation,GO:0010544~negative regulation of platelet activation,GO:0010863~positive regulation of phospholipase C activity,GO:0014068~positive regulation of phosphatidylinositol 3-kinase signaling,GO:0018108~peptidyl-tyrosine phosphorylation,GO:0023019~signal transduction involved in regulation of gene expression,GO:0030198~extracellular matrix organization,GO:0030324~lung development,GO:0030325~adrenal gland development,GO:0030335~positive regulation of cell migration,GO:0030539~male genitalia development,GO:0033327~Leydig cell

			differentiation,GO:0033674~positive regulation of kinase activity,GO:0034614~cellular response to reactive oxygen species,GO:0035790~platelet-derived growth factor receptor-alpha signaling pathway,GO:0038091~positive regulation of cell proliferation by VEGF-activated platelet derived growth factor receptor signaling pathway,GO:0042060~wound healing,GO:0042475~odontogenesis of dentin-containing tooth,GO:0043552~positive regulation of phosphatidylinositol 3-kinase activity,GO:0046777~protein autophosphorylation,GO:0048008~platelet-derived growth factor receptor signaling pathway,GO:0048015~phosphatidylinositol-mediated signaling,GO:0048146~positive regulation of fibroblast proliferation,GO:0048557~embryonic digestive tract morphogenesis,GO:0048701~embryonic cranial skeleton morphogenesis,GO:0048704~embryonic skeletal system morphogenesis,GO:0050872~white fat cell differentiation,GO:0050920~regulation of chemotaxis,GO:0055003~cardiac myofibril assembly,GO:0060021~palate development,GO:0060325~face morphogenesis,GO:0060326~cell chemotaxis,GO:0061298~retina vasculature development in camera-type eye,GO:0070374~positive regulation of ERK1 and ERK2 cascade,GO:0070527~platelet aggregation,GO:0071230~cellular response to amino acid stimulus,GO:0072277~metanephric glomerular capillary formation,GO:2000249~regulation of actin cytoskeleton reorganization,GO:2000739~regulation of

			mesenchymal stem cell differentiation,
SLC22A11	solute carrier family 22 member 11(SLC22A11)	Homo sapiens	GO:0015698~inorganic anion transport,GO:0015711~organic anion transport,GO:0046415~urate metabolic process,GO:0055085~transmembrane transport,

SHH	sonic hedgehog signaling molecule(SHH)	Homo sapiens	GO:0000122~negative regulation of transcription from RNA polymerase II promoter,GO:0001569~branching involved in blood vessel morphogenesis,GO:0001570~vasculogenesis,GO:0001656~metanephros development,GO:0001658~branching involved in ureteric bud morphogenesis,GO:0001708~cell fate specification,GO:0001755~neural crest cell migration,GO:0001947~heart looping,GO:0002052~positive regulation of neuroblast proliferation,GO:0002076~osteoblast development,GO:0002320~lymphoid progenitor cell differentiation,GO:0003140~determination of left/right asymmetry in lateral mesoderm,GO:0006897~endocytosis,GO:0007224~smoothened signaling pathway,GO:0007228~positive regulation of hh target transcription factor activity,GO:0007267~cell-cell signaling,GO:0007389~pattern specification process,GO:0007398~ectoderm development,GO:0007405~neuroblast proliferation,GO:0007411~axon guidance,GO:0007417~central nervous system development,GO:0007418~ventral midline development,GO:0007442~hindgut morphogenesis,GO:0007507~heart development,GO:0007596~blood coagulation,GO:0008209~androgen metabolic process,GO:0008284~positive regulation of cell proliferation,GO:0009880~embryonic pattern specification,GO:0009949~polarity specification of anterior/posterior axis,GO:0009953~dorsal/ventral pattern formation,GO:0010468~regulation of gene expression,GO:0010628~positive regulation of

			gene expression,GO:0010629~negative regulation of gene expression,GO:0014003~oligodendrocyte development,GO:0014706~striated muscle tissue development,GO:0014858~positive regulation of skeletal muscle cell proliferation,GO:0014902~myotube differentiation,GO:0016539~intein-mediated protein splicing,GO:0016540~protein autoprocessing,GO:0021513~spinal cord dorsal/ventral patterning,GO:0021522~spinal cord motor neuron differentiation,GO:0021794~thalamus development,GO:0021904~dorsal/ventral neural tube patterning,GO:0021930~cerebellar granule cell precursor proliferation,GO:0021938~smoothened signaling pathway involved in regulation of cerebellar granule cell precursor cell proliferation,GO:0021978~telencephalon regionalization,GO:0030010~establishment of cell polarity,GO:0030162~regulation of proteolysis,GO:0030177~positive regulation of Wnt signaling pathway,GO:0030324~lung development,GO:0030326~embryonic limb morphogenesis,GO:0030336~negative regulation of cell migration,GO:0030539~male genitalia development,GO:0030850~prostate gland development,GO:0030878~thyroid gland development,GO:0030900~forebrain development,GO:0030901~midbrain development,GO:0030902~hindbrain development,GO:0031016~pancreas development,GO:0031069~hair follicle morphogenesis,GO:0032435~negative regulation of proteasomal ubiquitin-dependent protein catabolic process,GO:0033077~T cell

			differentiation in thymus,GO:0033089~positive regulation of T cell differentiation in thymus,GO:0033092~positive regulation of immature T cell proliferation in thymus,GO:0034244~negative regulation of transcription elongation from RNA polymerase II promoter,GO:0034504~protein localization to nucleus,GO:0035115~embryonic forelimb morphogenesis,GO:0035116~embryonic hindlimb morphogenesis,GO:0042127~regulation of cell proliferation,GO:0042130~negative regulation of T cell proliferation,GO:0042307~positive regulation of protein import into nucleus,GO:0042475~odontogenesis of dentin-containing tooth,GO:0042481~regulation of odontogenesis,GO:0042733~embryonic digit morphogenesis,GO:0043010~camera-type eye development,GO:0043066~negative regulation of apoptotic process,GO:0043369~CD4-positive or CD8-positive, alpha-beta T cell lineage commitment,GO:0045059~positive thymic T cell selection,GO:0045060~negative thymic T cell selection,GO:0045109~intermediate filament organization,GO:0045445~myoblast differentiation,GO:0045596~negative regulation of cell differentiation,GO:0045880~positive regulation of smoothened signaling pathway,GO:0045893~positive regulation of transcription, DNA-templated,GO:0045944~positive regulation of transcription from RNA polymerase II promoter,GO:0046638~positive regulation of alpha-beta T cell differentiation,GO:0046639~negative regulation of alpha-beta T cell differentiation,GO:0048468~cell

			development,GO:0048538~thymus development,GO:0048557~embryonic digestive tract morphogenesis,GO:0048617~embryonic foregut morphogenesis,GO:0048643~positive regulation of skeletal muscle tissue development,GO:0048645~animal organ formation,GO:0048663~neuron fate commitment,GO:0048706~embryonic skeletal system development,GO:0048709~oligodendrocyte differentiation,GO:0048714~positive regulation of oligodendrocyte differentiation,GO:0048745~smooth muscle tissue development,GO:0048754~branching morphogenesis of an epithelial tube,GO:0048839~inner ear development,GO:0048859~formation of anatomical boundary,GO:0048864~stem cell development,GO:0051155~positive regulation of striated muscle cell differentiation,GO:0051781~positive regulation of cell division,GO:0060020~Bergmann glial cell differentiation,GO:0060021~palate development,GO:0060070~canonical Wnt signaling pathway,GO:0060174~limb bud formation,GO:0060428~lung epithelium development,GO:0060439~trachea morphogenesis,GO:0060445~branching involved in salivary gland morphogenesis,GO:0060447~bud outgrowth involved in lung branching,GO:0060458~right lung development,GO:0060459~left lung development,GO:0060463~lung lobe morphogenesis,GO:0060484~lung-associated mesenchyme development,GO:0060516~primary prostatic bud elongation,GO:0060523~prostate epithelial cord elongation,GO:0060662~salivary

			gland cavitation,GO:0060664~epithelial cell proliferation involved in salivary gland morphogenesis,GO:0060685~regulation of prostatic bud formation,GO:0060738~epithelial-mesenchymal signaling involved in prostate gland development,GO:0060769~positive regulation of epithelial cell proliferation involved in prostate gland development,GO:0060782~regulation of mesenchymal cell proliferation involved in prostate gland development,GO:0060783~mesenchymal smoothened signaling pathway involved in prostate gland development,GO:0060840~artery development,GO:0060916~mesenchymal cell proliferation involved in lung development,GO:0061053~somite development,GO:0061189~positive regulation of sclerotome development,GO:0071285~cellular response to lithium ion,GO:0071542~dopaminergic neuron differentiation,GO:0072136~metanephric mesenchymal cell proliferation involved in metanephros development,GO:0072205~metanephric collecting duct development,GO:0090090~negative regulation of canonical Wnt signaling pathway,GO:0090370~negative regulation of cholesterol efflux,GO:0097190~apoptotic signaling pathway,GO:1900175~regulation of nodal signaling pathway involved in determination of lateral mesoderm left/right asymmetry,GO:1900180~regulation of protein localization to nucleus,GO:1904339~negative regulation of dopaminergic neuron differentiation,GO:1905327~tracheoesophageal septum formation,GO:2000062~negative

			regulation of ureter smooth muscle cell differentiation,GO:2000063~positive regulation of ureter smooth muscle cell differentiation,GO:2000357~negative regulation of kidney smooth muscle cell differentiation,GO:2000358~positive regulation of kidney smooth muscle cell differentiation,GO:2000729~positive regulation of mesenchymal cell proliferation involved in ureter development,GO:2001054~negative regulation of mesenchymal cell apoptotic process,

TRIP13	thyroid hormone receptor interactor 13(TRIP13)	Homo sapiens	GO:0001556~oocyte maturation,GO:0006302~double-strand break repair,GO:0006355~regulation of transcription, DNA-templated,GO:0006366~transcription from RNA polymerase II promoter,GO:0007094~mitotic spindle assembly checkpoint,GO:0007130~synaptonemal complex assembly,GO:0007131~reciprocal meiotic recombination,GO:0007141~male meiosis I,GO:0007144~female meiosis I,GO:0007283~spermatogenesis,GO:0007286~spermatid development,GO:0048477~oogenesis,GO:0051598~meiotic recombination checkpoint,
TMEM242	transmembrane protein 242(TMEM242)	Homo sapiens	GO:0033615~mitochondrial proton-transporting ATP synthase complex assembly,
TUBB2B	tubulin beta 2B class IIb(TUBB2B)	Homo sapiens	GO:0000226~microtubule cytoskeleton organization,GO:0000278~mitotic cell cycle,GO:0001764~neuron migration,GO:0007017~microtubule-based process,GO:0050804~modulation of synaptic transmission,GO:1902669~positive regulation of axon guidance,GO:1990403~embryonic brain development,
ZCCHC7	zinc finger CCHC-type containing 7(ZCCHC7)	Homo sapiens	

Supplementary Table 4. List of primer pairs for gene expression analysis of murine samples.

	Gene name	Forward sequence	Reverse sequence
Primers	***Il1b***	AGTTGACGGACCCCAAAA	AGCTGGATGCTCTCATCAGG
	Il6	TAGTCCTCCCTACCCCAATTTC	TTGGTCCTTAGCCACTCCTTC
	Tnfa	GCCTGTAGCCCACGTCGTAG	TCCTCCACTTGGTGGTTTG
	Notch1	GATGGCCTCAATGGGTACAAG	TCGTTGTTGTTGATGTCACAGT
	Tlr6	AGCCAAGACAGAAAACCCATC	GGGGTCATGCTTCCGACTAT
	Tlr4	ATGGCATGGCTTACACCACC	GAGGCCAATTTTGTCTCCACA
	Muc1	TACCCTACCTACCACACTCACG	CTGCTACTGCCATTACCTGC
	Muc13	AGGGAGGGAGAAGGTAGCAG	GCTTTTGTGGCTTTCCCAGG
	Tjp1	GGAGCAGGCTTTGGAGGAG	TGGGACAAAAGTCCGGGAAG
	Gapdh	CGTGTTCCTACCCCCAATGT	TGTCATCATACTTGGCAGGTT

Supplementary Table 5. List of primer pairs for gene expression analysis of human samples.

	Gene name	Forward sequence	Reverse sequence
C	***CLDN1***	CAGAGCACCGGGCAGATCCA	GGCACGGGTTGCTTGCAATG
	CLDN2	CTGAGCTCACAGGCCATTCA	CCACAAGCAGCCTCAAGAAG
	CLDN3	CACTGCCACAGGACCTTCAG	AGGATGGCCACCACGATGA

	CLDN4	GCAACATTGTCATCACCTCGCAG	ACACCTTGCACTGCATCTGG
	OCLM	CACCTCCTTACAGGCCTGAT	GAGTAGGCTGGCTGAGAGA
	TJP1	ACCAGTAAGTCGTCCTGATCC	TCGGCCAAATCTTCTCACTCC
	GAPDH	CATGAGAAGTATGACAACAGC	AGTCCTTCCACGATACCAAAG

Supplementary Table 6. List of HBx-targeting siRNA sequences.

	HBx gene target sequence
HBx.siRNA1	UUGACAUUGCUGAGAGUCCAA
HBx.siRNA2	UAUGCCUCAAGGUCGGUCGUU
HBx.siRNA3	UUUAAACAAACAGUCUUUGAA
HBx.siRNA4	UCGGUCGUUGACAUUGCUGAG
HBx.siRNA5	UUCACGGUGGUCUCCAUGCGA
HBx.siRNA6	AGUCUUUAAACAAACAGUCUU
HBx.siRNA7	UCUUUAAACAAACAGUCUUUG
HBx.siRNA8	UUAAACAAACAGUCUUUGAAG

Supplementary Table 7. WENR culturing medium composition.

Compound	**Stock Mol/L**	**Final Mol/L**	**Final Volume (mL) in 100 ml**	Company; cat. #

WNT3a	2	1	50	ATCC; cat. #CRL-2647
R- Spondin	5	1	20	Kindly provided by Leland Standford Junior University; Prof. Brian Huang
Nogging	10	1	10	Kindly provided byHubrecht Insitute; Prof. J. den Hertog
B27	50	1	2	LifeTechnologies; 17504044
N2	100	1	1	LifeTechnologies; 17502001
NAC	0.5	0.001	0.2	Sigma-Aldrich; A7250
Nicotinammide	1	0.01	1	Sigma-Aldrich; 72340-100
A83-01	0.005	0.0000005	0.01	Tocris; 2939
SB431514	0.01	0.000001	0.01	Tocris; 1614
SB202190	0.03	0.000003	0.01	Tocris; 1264
EGF	0.1	0.00005	0.05	Peprotech AF-100-15
Prostaglandin E2	0.001	0.0000001	0.001	Tocris; 2296
Gastrin	0.0001	0.00000001	0.01	Tocris; 3006
Advanced DMEM/F12			16	Lonza; #BE04687F/1

References

1 Edgar RC, Taylor J, Lin V, *et al.* Petabase-scale sequence alignment catalyses

viral discovery. *Nature* 2022;602:142–7. doi:10.1038/s41586-021-04332-2

2 Soldánová M, Selbach C, Sures B. The Early Worm Catches the Bird? Productivity and Patterns of Trichobilharzia szidati Cercarial Emission from Lymnaea stagnalis. *PLoS ONE* 2016;11:e0149678. doi:10.1371/journal.pone.0149678

3 Gulyás K, Soldánová M, Orosová M, *et al.* Confirmation of the presence of zoonotic Trichobilharzia franki following a human cercarial dermatitis outbreak in recreational water in Slovakia. *Parasitol Res* 2020;119:2531–7. doi:10.1007/s00436-020-06751-y

4 Tennant BC, Gerin JL. The woodchuck model of hepatitis B virus infection. *ILAR J* 2001;42:89–102. doi:10.1093/ilar.42.2.89

5 Ng SC, Shi HY, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017;390:2769–78. doi:10.1016/S0140-6736(17)32448-0

6 Kobayashi T, Siegmund B, Le Berre C, *et al.* Ulcerative colitis. *Nat Rev Dis Primers* 2020;6:74. doi:10.1038/s41572-020-0205-x

7 Altschul SF, Gish W, Miller W, *et al.* Basic local alignment search tool. *J Mol Biol* 1990;215:403–10. doi:10.1016/S0022-2836(05)80360-2

8 Ginestet C. ggplot2: Elegant Graphics for Data Analysis. *J Royal Statistical Soc A* 2011;174:245–6. doi:10.1111/j.1467-985X.2010.00676_9.x

9 Wickham H, Averick M, Bryan J, *et al.* Welcome to the tidyverse. *JOSS* 2019;4:1686. doi:10.21105/joss.01686

10 Ungaro F, Tacconi C, Massimino L, *et al.* MFSD2A Promotes Endothelial Generation of Inflammation-Resolving Lipid Mediators and Reduces Colitis in Mice. *Gastroenterology* 2017;153:1363-1377.e6. doi:10.1053/j.gastro.2017.07.048

11 Dull T, Zufferey R, Kelly M, *et al.* A third-generation lentivirus vector with a conditional packaging system. *J Virol* 1998;72:8463–71. doi:10.1128/JVI.72.11.8463-8471.1998

12 Ungaro F, Massimino L, Furfaro F, *et al.* Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* 2019;10:149–58. doi:10.1080/19490976.2018.1511664

13 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–20. doi:10.1093/bioinformatics/btu170

14 Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. doi:10.1093/bioinformatics/bts635

15 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. doi:10.1186/s13059-014-0550-8

16 Subhash S, Kanduri C. GeneSCF: a real-time based functional enrichment tool with support for multiple organisms. *BMC Bioinformatics* 2016;17:365. doi:10.1186/s12859-016-1250-z

17 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics* 2013;14:178–92. doi:10.1093/bib/bbs017

18 Virgin HW. The virome in mammalian physiology and disease. *Cell* 2014;157:142–50. doi:10.1016/j.cell.2014.02.032

19 Ungaro F, Colombo P, Massimino L, *et al.* Lymphatic endothelium contributes to colorectal cancer growth via the soluble matrisome component GDF11. *Int J Cancer* 2019;145:1913–20. doi:10.1002/ijc.32286

20 D'Alessio S, Correale C, Tacconi C, *et al.* VEGF-C-dependent stimulation of lymphatic function ameliorates experimental inflammatory bowel disease. *J Clin Invest* 2014;124:3863–78. doi:10.1172/JCI72189

21 Tan J, Gong J, Liu F, *et al.* Evaluation of an antibiotic cocktail for fecal microbiota transplantation in mouse. *Front Nutr* 2022;9:918098. doi:10.3389/fnut.2022.918098

[dataset] 22 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74. doi:10.1038/nature11247

23 Sessa A, Fagnocchi L, Mastrototaro G, *et al.* SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron* 2019;104:271-289.e13. doi:10.1016/j.neuron.2019.07.013

24 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9. doi:10.1038/nmeth.1923

25 Gaspar JM. Improved peak-calling with MACS2. *BioRxiv* Published Online First: 17 December 2018. doi:10.1101/496521

26 Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31:2382–3. doi:10.1093/bioinformatics/btv145

27 Bailey TL, Boden M, Buske FA, *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009;37:W202-8. doi:10.1093/nar/gkp335

28 Ramírez F, Dündar F, Diehl S, *et al.* deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;42:W187-91.

doi:10.1093/nar/gku365

29 Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;9:215–6. doi:10.1038/nmeth.1906

30 Patil I. Visualizations with statistical details: The “ggstatsplot” approach. *JOSS* 2021;6:3167. doi:10.21105/joss.03167

Supplementary figure legends

Supplementary Figure 1. HBx was detected in specific mucosal cell populations from UC patients. (**A**) Schematic experimental workflow for the sorting of the indicated cell populations from UC patient-derived colonic mucosae (n=24 total samples). (**B**) Gating strategy to isolate live, epithelial cells, $CD4^+$ T cells and macrophages, B and $CD8^+$ T cells, dendritic cells, fibroblast, and endothelial cells, as indicated in each plot. (**C**) HBV genome (Genes S, P, X, and C) mapping of FASTQ reads from RNA-Seq detects HBx transcript only in $CD8^+$ T cells, dendritic cells, endothelium, epithelium, and macrophages. (**D**) Schematic representation of the Touchdown PCR (starting at 68°C and ending at an annealing temperature of 62°C) amplification cycles. (**E**) Gel electrophoresis image showing HBx amplicon in a representative cohort of UC patients (UC 1, 2, 3, 4, 5, 6 patients at their first screening). "+" sample symbol indicates the HBx+ UC patient tested in the previously published work(12) GAPDH was used as an internal control. (**F**) Representative read from Sanger sequencing showing perfect amplicon alignment to the HBx reference sequence.

Supplementary Figure 2. HBx correlates with no UC clinical parameters.
Correlation heatmap showing multivariate analysis results expressed as Pearson coefficients (from 1 to -1). Xs mark non-significant at $P<0.05$ (Adjustment: Holm). The analysis was performed on UC patients recruited at Fondazione Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy.

Supplementary Figure 3. HBx overexpression in colonic mucosa activates proinflammatory biological processes.
(**A**) Schematic experimental workflow of healthy colonic biopsies (n=3 healthy donors) transduction with HBx followed by RNA-Seq. (**B**) MA plot showing differentially expressed genes between HBx-overexpressing colonic biopsies and the GFP-transduced controls. Red and blue indicate statistically significant up- and down-regulated genes, respectively. (**C**) GO plot showing dysregulated biological processes in HBx-transduced biopsies by comparison with GFP. Icons from Streamline (https://app.streamlinehq.com).

Supplementary Figure 4. HBx induces mucosal barrier damage *in vivo*.
(**A, B**) Representative images of endoscopy performed in GFP- (A) and HBx-treated (B) mice. (**C, D**) Graph showing statistically significant differences in stool consistency (C) and bleeding scores (D) in GFP- and HBx-treated mice. Statistical analysis was performed with Two-way ANOVA, with Bonferroni's post-correction.

Supplementary Figure 5. HBx induces colitis with no contributions by the microbiota. (**A**) Schematic experimental workflow of *in vivo* HBx-induced colitis (n=5-8/group). (**B**) Disease Activity Index (DAI) of mice treated with either GFP- or HBx-carrying liposomes with or without antibiotic treatment. (**C, D**) Box plots showing colon length (C) and endoscopic scores (D) of HBx-induced colitic mice with or without antibiotic treatment. (**E**) Representative images of endoscopy performed in GFP- and HBx-treated mice with or without antibiotic treatment. Statistical analysis was performed with Two-way ANOVA, with Bonferroni's post-correction. *P<.05; **P<.01; ***P<.005; ****P<.001

Supplementary Figure 6. HBx shapes colonic mucosal immunity.
Relative cell population abundance in HBx-induced colitis mice versus the GFP control quantified by FACS analysis.

Supplementary Figure 7. HBx regulates the expression of UC-related genes by binding enhancer regions.
(**A, B**) Integrative Genome Viewer (IGV) screenshots for two representative loci showing HBx peak localization, together with all other ChIP-Seq signals.

Supplementary Figure 8. HBx silencing re-establishes the barrier functions and reduces proinflammatory phenotype *in vitro*.
(**A**) Box plots showing the HBx expression level in HBx-transduced Caco-2 upon the indicated siRNA administration. (**B**) Graph showing cellular growth rate between HBx- and GFP-transduced epithelial cell line upon either the HBx-targeting siRNA administration or the scramble control. Experiments were performed in triplicates in three independent experiments. (**C**) Box plots showing Transepithelial Electrical Resistance (TEER, expressed as Ω/cm²) measurements on HBx- and GFP-transduced epithelial cell lines upon either the HBx-targeting siRNA administration or the scramble control. (**D**) Box plots showing Real-Time PCR results for the epithelial barrier markers in Caco-2 cells transduced with either the GFP- or HBx-carrying lentiviruses upon HBx-targeting siRNA administration or the scramble control, expressed as 2-ΔCT (GAPDH was used as the housekeeping genes). Statistical analysis was performed with One-way or Two-way Anova with Bonferroni's post-correction. *P<.05; **P<.01; ***P<.005; ****P<.001.

Supplementary Figure 9. HBx silencing restores the gut mucosa *in vivo*.
(**A, B**) Representative images of endoscopy performed in GFP (A) and HBx-induced (B) colitic mice treated with scramble or HBx-targeting siRNAs. (**C, D**) Graph showing stool consistency (C) and bleeding scores (D) in GFP and HBx-induced colitic mice treated with scramble or HBx-targeting siRNAs. Statistical analysis was performed with Two-way ANOVA, with Bonferroni's post-correction. Results with a P-value<0.05 were considered significant.

Supplementary Figure 10. HBx silencing re-establishes the barrier functions and reduces proinflammatory phenotype *in vivo*.

(**A-C**) Box plots showing relative expression of genes related to the epithelial barrier (A), anti-microbial response (B), and pro-inflammatory response (C) in the colonic mucosa of GFP and HBx liposome-transduced mice upon HBx-targeting siRNA administration ot scramble control. Expression levels are expressed as 2-ΔCt; GAPDH was used as the housekeeping gene. N=3/group, 2 independent experiments. Statistical analysis was performed with Two-way ANOVA, with Tukey's post-correction. *P indicates the comparison HBx versus GFP; $P indicates the comparison HBx versus siRNA HBx; *P<.05; **P<.005.

Supplementary Figure 11. The *Hepadnaviridae* family is detected in different organisms.

(**A**) Schematic representation of the computational query for *Hepadnaviridae* reads within the Serratus database (DB). (**B**) Bar graph showing the relative number of positive reads per species, color-coded for macro categories. Icons from Streamline (https://app.streamlinehq.com).

Supplementary Figure 12. HBx BLAST alignments and biodiversity data of *Trichobilharzia* and *Isatis tinctoria*.
(**A**) Schematic representation of the computational workflow. The reads found to be positive for Hepadnaviridae in the Serratus DB were mapped to the HBx reference sequence with NCBI BLAST. (**B-E**) *Trichobilharzia* (B) and *Isatis Tinctoria* (C) read alignment to the HBx reference sequence. No alignment was found when the mapping was attempted with the reads belonging to *Odorrona tormota* (D), or *Elaeagnaceae* (E). (**F-G**) Map of georeferenced record observations of Trichobilharzia (F) and Isatis Tinctoria (G) from the Global Biodiversity Information Facility (GBIF), representative of their global distribution. Icons from Streamline (https://app.streamlinehq.com).

Supplementary Figure 13. FACS analysis of HBx-treated mice.
(**A-C**) Gating strategy to isolate live cells prior to downstream immune profiling with cytoChain.