



Review

# Gut eukaryotic virome in colorectal carcinogenesis: Is that a trigger?

Luca Massimino<sup>a,1</sup>, Sara Lovisa<sup>b,1</sup>, Luigi Antonio Lamparelli<sup>a</sup>, Silvio Danese<sup>a,b</sup>,  
Federica Ungaro<sup>a,b,\*</sup>



<sup>a</sup> Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

<sup>b</sup> IBD Center, Laboratory of Gastrointestinal Immunopathology, Humanitas Clinical and Research Center, Rozzano, Milan, Italy

ARTICLE INFO

Article history:

Received 30 September 2020  
Received in revised form 24 November 2020  
Accepted 28 November 2020  
Available online 07 December 2020

Keywords:

Colorectal cancer  
Inflammation  
Genetic alterations  
Immunity  
Microbiota  
Virome  
Bioinformatics  
Computational biology

ABSTRACT

The human gut microbiota is composed of bacteria and viruses that might be associated with colorectal cancer (CRC) onset and progression. Indeed, although viral infections have been reported to be the primary trigger in many diseases, the role of eukaryotic viruses populating the gut mucosa during early colorectal carcinogenesis is underinvestigated. Human eukaryotic viruses in the gut were found to induce alterations of the immune homeostasis so that some viral-dependent mechanisms likely able to induce DNA alterations in the bowel wall have been proposed, although no demonstration is available yet. However, thanks to the latest advancements in computational biology and the implementation of the bioinformatic pipelines, the option of establishing a direct causative link between intestinal virome and CRC will be possible soon, hopefully paving the way to innovative therapeutic strategies blocking or reverting the CRC pathogenesis.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

2. The human gut virome .....	17
3. CRC pathogenesis and its hallmarks: A summary in brief. ....	18
4. The eukaryotic enteric virus effects on intestinal homeostasis: Friend or foe? .....	19
5. Eukaryotic viruses and their contribution to carcinogenesis: Is that possible in the intestine? .....	20
6. The challenge of the virome discovery: What was done and what is next .....	22
6.1. Computational approaches for virome studies .....	23
7. Summary and outlook .....	24
8. Literature search strategy .....	26
Author contributions. ....	26
Declaration of Competing Interest .....	26
Acknowledgments .....	26
References .....	26

## 1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy is classified as fourth cancer for mortality rate, with

several factors influencing its aetiology, ranging from genetic predisposition to daily habits, dietary intake of red meat, particularly processed meat, and prolonged exposure to some form of environmental stressors [1] and carcinogens in the workplace [2]. CRC is characterized by inflammation [3], prominent angiogenesis [4,5], and lymphangiogenesis [6–8], the former primarily designated for sustaining cellular overgrowth, the latter causative of the high risk of CRC-metastasis mainly up to the liver and,

\* Corresponding author at: IBD Center, Laboratory of Gastrointestinal Immunopathology, Humanitas Clinical and Research Center, Rozzano, Milan, Italy.

E-mail address: [federica.ungaro@humanitasresearch.it](mailto:federica.ungaro@humanitasresearch.it) (F. Ungaro).

<sup>1</sup> These authors share co-first authorship.

to a lesser extent, to the lungs, bones, and brain [3]. The etiopathogenesis of CRC arises from a neoplastic transformation of epithelial cells due to irreparable DNA damage in the bowel wall, where cells start to aberrantly proliferate and accumulate mutational events [9,10].

One of the features associated with CRC pathogenesis is intestinal dysbiosis [11]. The human microbiota, comprising bacteria, viruses, fungi, protists, and archaea, represents an integral part of our body and is key for tissue homeostasis, while it influences the host's immunity and metabolism. Numerous microorganisms live as symbiotic commensals of our body and carry approximately 150 times more genes than are found in the entire human genome [12]. They represent a multifaceted community (because of the coexistence of different species within the same tissue) colonizing several tissues and organs such as skin, mammary glands, placenta, seminal fluid, uterus, ovarian follicles, lung, saliva, oral mucosa, conjunctiva, biliary tract, and, to a greater extent, the gastrointestinal tract, where the colorectum is the area most densely populated by the microbial species [12].

In physiological conditions, the human microbial species coexist in a constant equilibrium, where one species may prevail over the other and participates in the regulation of the relative abundance of the other commensals. Nevertheless, changes in microbiota composition influence the host's physiology and, by impacting the immune, endocrine, and nervous system, these variations are associated with a wide array of illnesses, ranging from major depressive disorder [13,14] to inflammatory bowel disease (IBD) [15], to colorectal cancer [11].

Although the most abundant and studied gut colonizers, in both healthy and pathological conditions [15,16], are the bacteria [17], a growing body of evidence highlighted dysbiosis of gut virome, composed of both bacteriophages and eukaryotic-targeting viruses, to be responsible for shaping the overall intestinal microbiota composition [15]. As an example, bacteriophages populating the gut microbiota act as predators of bacterial species, maintaining the bacterial species diversity of the intestinal tract through predator–prey relationships [18]. This results in a direct or indirect effect on the host's health and physiology [19,20]. It is noteworthy that phage infection can lead to virulent or temperate replication cycles. Although the role of these lifestyles within the gut is still being investigated, it is thought that there may be a predominance of temperate phage, which can incorporate into the bacterial chromosome as latent prophages.

More neglected actors in the virome are the eukaryotic-targeting viruses, which only recently have been attracting interest in gastrointestinal diseases, such as IBD [15,18,21]. Some studies highlighted eukaryotic-targeting viral infections as responsible for at least a sixth of the global cancer burden [22,23]. Despite the enormous amount of evidence associating eukaryotic viral infections with the higher risk of colorectal carcinogenesis [24], the investigation of the specific role of the eukaryotic gut virome in the CRC pathogenesis remains to be fully elucidated.

The understanding of the role of the human gut virome in health and illness is in its infancy because of the only recently developed computational approaches, which now allow uncovering both DNA and RNA viruses at the same time [21] highlighting virome dysbiosis as associated with diseases such as periodontitis [25], IBD [21,26], cystic fibrosis [27,28], and cancer [24,29].

This review will attempt to provide insights on the possible mechanisms of eukaryotic virus-induced colorectal carcinogenesis and to suggest the most appropriate computational workflows for virome studies in the context of CRC. This will hopefully help further studies to prompt the virome discovery field, ultimately serving as the starting point for the development of innovative therapies blocking early phases of CRC, rather than contrasting symptoms.

## 2. The human gut virome

On the earth, we coexist together in mutualism, other times, we can compete for nutrients. This is true at every complexity level of living beings: mammals and bacteria, birds and protists, algae and fungi, all these organisms can populate the same area at the same time, either without competing for food or other sources of energy or fighting to obtain a more comfortable place to live within. In this scenario, not properly living organisms, but better defined as infectious agents, are the viruses, which need invading cells (either prokaryotic or eukaryotic) to replicate themselves and take advantage of their whole molecular apparatus to propagate and multiply their genetic information, in millions at a time. Every virus may differ for its impact on human health, ranging from mild illness (rhinovirus of the common cold) to the devastating pandemic in 2020 due to the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2).

Besides the different places of the earth where we can find different viral species, the human body is a place where a variety of viruses can exist, either latently stimulating the immune system and promoting its development without clinical manifestation [30], or being contrasted, sometimes with serious complications for human health [31].

Following the Baltimore Virus classification, viruses can belong to 7 different groups, based on the combination of their nucleic acid (DNA or RNA), strandedness (single-stranded or double-stranded), sense, and method of replication. RNA- and DNA-viruses have long been associated with human physiology, colonize different niches of the body, and form the human virome, which includes eukaryotes-infecting viruses (the eukaryotic virome), bacteriophages or phages (the bacterial virome), and viruses that infect archaea (the archaeal virome) [32]. These various types of viruses differ in multiple ways, but they all share the capability to exist with both “lytic” life cycles (during which the host is killed) and “latent” life cycles, when the viruses stably reside within a living cell either integrated into the host chromosome (e.g. prophages) or as an episome (e.g. herpesviruses) until they reactivate and become infectious. This variety of survival opportunities allows viruses to elude the body's defense responses and to evolve together with its host in mutualistic and/or symbiotic relationships [18,30].

Similar to the other components of the microbiota, the virome is large and diverse, stabilized in adulthood, and highly personalized, inhabiting different parts of the body, mainly the mucosal surfaces [32,33]. However, the intestine can be considered as the main residence of both prokaryotic and eukaryotic human virome [34]. Prokaryotic-infecting viruses in the gut are more highly abundant than bacteria, thus indicating that their existence within the microbiota can modify the relative proportions of the intestinal bacterial strains by killing their hosts during the lytic release of viral particles [15]. Additionally, they integrate into bacterial genomes, finally influencing the physiology of the animal host [35]. Likewise, the eukaryotic virome may influence diseases [15,21,32,36,37] and has been proposed as a sort of “influencer” of intestinal immunity, because of the capability of viral protein to latently stimulate the immune response [30,32,38] by directly interacting with the molecular apparatus of the host infected cells [39].

The main colonizers of the gut virome are mostly DNA bacteriophages [35,40,41]. Interestingly, cross-assembly phages (CrAssphage) have been found as the prevalent constituent of the human gut viral population, possibly representing a new viral family of bacteriophages recently found to be present in thousands of human-feces-associated environments around the world, confirming it as a strong marker for fecal contamination [42].

Thanks to metatranscriptomics studies uncovering the presence of both RNA- and DNA-based entities allowing also the characterization of the eukaryotic virome, the most abundant eukaryotic viral order found in the gut mucosa is the *Herpesvirales*, followed by the *Picornavirales* and *Tymovirales*, independently of the disease and intestinal location [21]. At the family level, the gut eukaryotic virome composition is made of viral entities belonging to the *Adenoviridae*, *Anelloviridae*, *Astroviridae*, *Parvoviridae*, *Picornaviridae*, and *Picobirnaviridae* families, that may be either symptomatic or latent for long in healthy people [15]. It is noteworthy that the gut virome is made not only of enterotropic viruses, such as rotavirus or norovirus [43] but also of allotropic viral entities, such as *Pneumoviridae* [44], *Herpesviridae* [45], *Hepeviridae*, and *Hepadnaviridae*, together with insect and plant-infecting viruses [21].

These pieces of evidence lead to the conclusion that the prolonged co-existence of these viral entities with the host has some implications. For example, during mammalian evolution endogenous retroviruses (ERVs) have been integrating into the genome, consistently impacting on mammalian physiology, as in the case of syncytin proteins, important for the placental development and derived from the ERV *env* genes, or the ERV-derived interferon-inducible enhancer elements [46].

Also, even if rarely, because of the high level of mutations during the genomic replication, ERVs may encode for some viral products which can activate the host's immunity, affect gene expression and promote carcinogenesis [43].

In line with this, to look at specific eukaryotic viral infections as triggers for CRC onset is reasonable, taking also into consideration that other oncoviruses can induce carcinogenesis [23,47], as we will discuss later in this review.

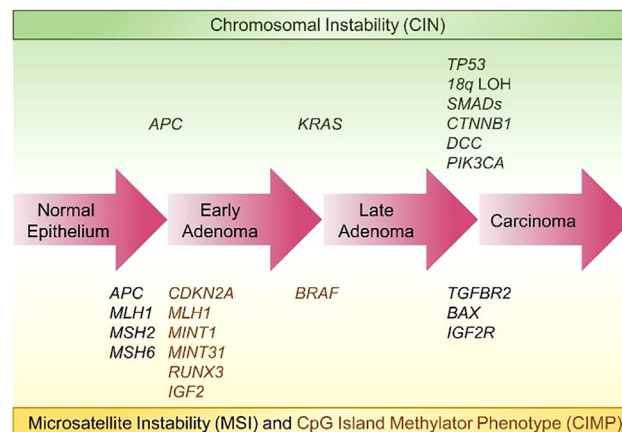
### 3. CRC pathogenesis and its hallmarks: A summary in brief

CRC arises through the adenoma-carcinoma progression, whereby the normal colorectal epithelium undergoes a series of transformation events inducing the initial formation of adenomas which then progresses into carcinomas, subsequently leading to invasive and metastatic tumors [48].

Besides the well-characterized genomic instability, schematized in Fig. 1, and the tumor-associated inflammation, very recently reviewed elsewhere [3], perturbation of the intestinal microbiome-associated homeostasis is classified as a hallmark of CRC.

Of note, an increasing number of studies strongly indicate a fundamental contribution of the gut microbiota to the initiation and progression of CRC by modulating the host's immunity and inducing inflammation-mediated tumor-promoting mechanisms (Fig. 2).

In a spontaneous transgenic mouse model of serrated polyps, the development of the cecal polyps was associated with local epithelial barrier leakage, bacterial invasion, and marked accumulation of neutrophils. Anti-Ly6G treatment, antibiotic treatment, or embryo rederivation were able to markedly inhibit the formation of these polyps, therefore indicating microbiome and the microbiome-induced inflammation as central drivers or intestinal carcinogenesis [49]. Similarly, a spontaneous model of CRC by *Zeb2* overexpression in the intestinal epithelium confirmed the microbiota-dependent initiation of carcinogenesis as cancer development was completely prevented by treatment with broad-spectrum antibiotics or germ-free rederivation. This microbiota-induced CRC was at least in part mediated by the local infiltration of myeloid immune cells and the release of inflammatory cytokines as demonstrated by the GR-1-depleting antibody-mediated protection against CRC [50]. This close interplay among microbiota, inflammation, and CRC is also evident during colitis-associated



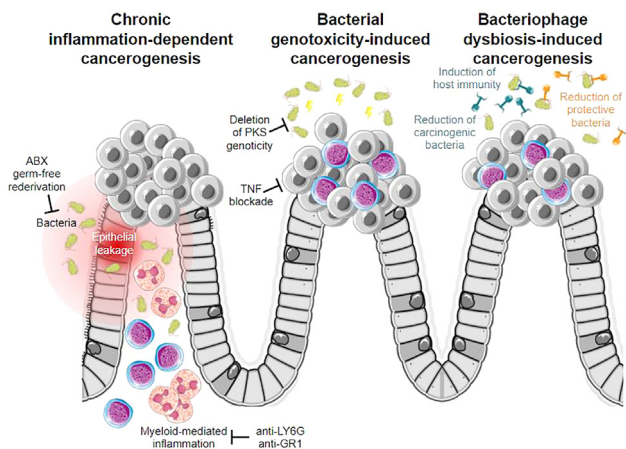
**Fig. 1.** Genomic alterations in CRC pathogenesis. The prominent molecular hallmark of CRC pathogenesis is represented by the genomic instability resulting from three distinct pathways: chromosomal instability (CIN), microsatellite instability (MSI), and CpG island methylator phenotype (CIMP). CIN is characterized by an imbalance in the chromosome number (aneuploidy) and loss of parental alleles (loss of heterozygosity, LOH) due to defects in the chromosomal segregation, mitotic checkpoints, and DNA damage repair machinery, eventually leading to gene copy number variation and mutational burden. Inactivating mutations, allelic deletions, or promoter hypermethylation of the adenomatous polyposis (*APC*) gene, causing the hyperactivation of the WNT signaling pathway, represent the precursory genetic event in CRC. Subsequent activating mutation in the *KRAS* gene increases Ras signaling and downstream effector pathways, thus promoting cell proliferation and survival. Mutation and LOH of the p53 tumor suppressor gene (*TP53*) represent the third major genetic event enhancing inflammation and marking the transition from adenoma to carcinoma. Mutation or allelic loss in other genes, such as *CTNNB1*, *SMAD2*, *SMAD4*, *DCC*, and *PIK3CA*, and aberrant overexpression of *COX-2*, represent genetic events closely associated with CRC progression. MSI is the result of defects in the DNA mismatch repair system due to mutational or epigenetic inactivation of *MLH1*, *MSH2*, and *MSH6* genes. The consequent insertion or deletion errors occurring mainly in the microsatellite regions, sequences of tandem DNA repeats, results in frame-shift mutations in the coding region of tumor-associated genes, such as *TGFBR2* and *BAX*. The CIMP subtype is defined by transcriptional inactivation by DNA hypermethylation at promoter CpG island of genes involved in cell cycle or tumor-suppressive activities (*CDKN2A*, *MLH1*, *MINT1*, *RUNX3*, *IGF2*, among others). CIMP CRC is characterized by MSI and is highly associated with *BRAF* mutations.

colon cancer (CAC) development. Indeed, in an experimental model of CAC, the associated intestinal dysbiosis was found to enhance the inflammation-induced colorectal tumorigenesis [51].

Multiple pieces of evidence indicate that the inflammatory state impacts microbiome homeostasis by promoting the expansion of microorganisms with genotoxic capabilities. As an example, in the *Il10* knock-out spontaneous model of colitis, the deletion of the *pks* genotoxic island decreases tumor growth and invasion with no effects on inflammation [52]. Similar results were confirmed in human studies, where the direct genotoxic activity of the intestinal bacterial microbiome in human CRC was indeed recently observed [53]. Additionally, targeting intestinal inflammation with TNF blockade was able to revert the carcinogenic status of the microbiota to non-carcinogenic status and to protect from CRC development [54].

As another consistent component of the gut microbiota, bacteriophages in the gut have been long proposed to be indirectly associated with CRC, and some studies are currently attempting to depict their involvement in gut carcinogenesis [37]. A plethora of bacteriophages possibly involved at different stages of CRC has been compiled [24], and some of them displayed a dual role during CRC, since have emerged as both causatives of the depletion of favorable intestinal bacterial commensals, thus resulting detrimental [24,36], and as inhibitors of the colonization by carcinogenic bacteria, therefore exerting a protective role from CRC development [55].



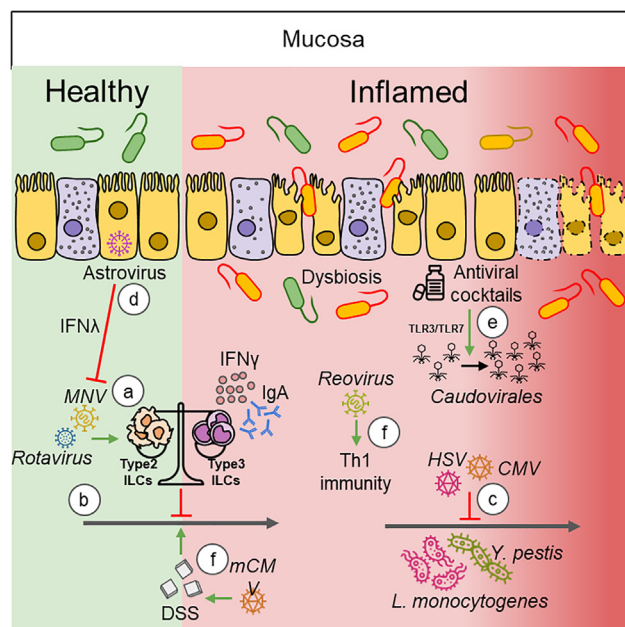


**Fig. 2.** Mechanisms of microbiome-induced intestinal cancerogenesis. Bacterial-induced cancerogenesis is mainly exerted by two mechanisms: induction of intestinal inflammation, characterized by the infiltration of myeloid immune cells and the release of inflammatory cytokines, or by direct genotoxic activity. Ablation of the bacterial component by antibiotics (ABX) treatment or germ-free rederivation, reduction of inflammation by anti-Ly6G, anti-GR1 or anti-TNF treatments, and inhibition of bacterial genotoxic capabilities by deletion of the *pkS* island were all demonstrated to efficiently protect from CRC development. Bacteriophages were shown to be able to shape the bacterial community in opposite ways: they can promote CRC by inducing dysbiosis with reduction of favorable commensal bacteria or impair CRC development by reducing the colonization by carcinogenic bacteria and inducing the host immunity. The cartoon has been produced with google slides (<https://www.google.com/slides/about/>) and its freely available addons (<https://www.google.com/slides/about/>) and with SMART (<https://www.smart.servier.com>).

On one hand, these studies shed light on the importance of the tight connection between gut homeostasis and microbiota during colorectal carcinogenesis, on the other pointed out bacteria and bacteriophage to indirectly impact the host's cells, raising questions about the possible roles harbored by other colonizers of the intestinal mucosa, such as eukaryotic viruses, in triggering CRC. We will discuss this point shortly.

**4. The eukaryotic enteric virus effects on intestinal homeostasis: Friend or foe?**

The virome-host coevolution causes the ability of viral entities to exert both beneficial and detrimental effects. Several animal studies explained how particular enteric viruses can benefit the health of an organism. Murine norovirus (MNV) was found to revert bacterial dysbiosis-induced architectural and immune defects in the gastrointestinal tract by restoring small intestinal crypt-villus structures and Paneth cell functions, inducing lymphocyte upregulation with the production of IFN $\gamma$  and IgA release thus balancing the type 2 and 3 innate lymphoid cells (ILCs) ratio *in vivo* in a type 1 IFN signaling-dependent mechanism [56] (Fig. 3a). This happened not only in homeostatic conditions but also during *Citrobacter rodentium* infection or dextran sodium sulfate (DSS)-induced colitis (Fig. 3b), thus pointing out the possible protective role that some enteric viruses can exert. This is not the only evidence reporting a eukaryotic virus to be beneficial. Indeed, previous literature demonstrated that mice latently infected with either murine gammaherpesvirus 68 or murine cytomegalovirus (mCMV), genetically highly comparable to the human pathogens Epstein-Barr virus and human cytomegalovirus respectively, are resistant to infection by the bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis* [57] (Fig. 3c). This study added another important key point, highlighting that latent infection induced the continuous and persistent immune stimulation through the antiviral IFN $\gamma$  and systemic macrophage activation at basal levels.



**Fig. 3.** The impact of enteric viruses on intestinal mucosa homeostasis. (a) Murine norovirus (MNV) protects mucosa by maintaining bacterial homeostasis and epithelial barrier integrity, as inducing lymphocyte upregulation with the production of IFN $\gamma$  and IgA release and balancing type 2/3 innate lymphoid cells (ILCs). (b) MNV protects against DSS-induced colitis. (c) Murine gammaherpesvirus 68 (HSV) or murine cytomegalovirus (mCMV) counteract infections by the bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis*. (d) Murine astrovirus protects against MNV and rotavirus infections in IFN- $\lambda$  signaling dependent. (e) Antiviral drug cocktail increased the susceptibility to DSS via Toll-like receptor (TLR)3 and TLR7-triggered signaling, with the final increased level of Caudovirales. (f) CMV infection leads to a more severe DSS-induced colitis. (g) Reovirus infection promotes TH1 immunity. The cartoon has been produced with google slides (<https://www.google.com/slides/about/>) and its freely available addons.

This effect probably shapes the immune environment, which ultimately protects against subsequent infections, endowing such a latency with the capability to build a symbiotic relationship with the host's immunity [57]. Regarding the effects of the virome on modulating enteric infections, a recent study reported that a specific strain of the murine astrovirus in the gut can complement primary immunodeficiency, thus protecting against murine norovirus and rotavirus infections in IFN- $\lambda$  signaling dependent manner in gut epithelial cells [58] (Fig. 3d).

Besides this body of evidence, depicting individual viral infection as a modulator of intestinal immunity, other studies have made more consistent the concept that the whole gut virome intervenes in the maintenance of gut homeostasis. This was the case of the study by Yang and colleagues, demonstrating that the depletion of the gut viral entities through antiviral drug cocktail administration increased the susceptibility to DSS in mice via Toll-like receptor (TLR)3 and TLR7-triggered signalling [59], with the final increased level of bacteriophage Caudovirales (Fig. 3e), similar to the observations in human IBD patients [18]. However, the boundary between beneficial and detrimental effects exerted by the virome is subtle. As mentioned above, while the depletion of the viral community may be detrimental against colitis [59], specific latent mCMV infection leads to a more severe DSS-induced colitis [60] (Fig. 3f). Another example is represented by the reovirus infection, known as an avirulent pathogen that elicits protective immunity that nevertheless disrupts the intestinal immune homeostasis by suppressing peripheral regulatory T cell (pTreg) conversion and promoting interferon regulatory factor 1-driven TH1 immunity to dietary antigen, eventually triggering the development of celiac disease [58] (Fig. 3g).

Altogether these data on one hand support and sustain the virome as an essential contributor to intestinal homeostasis, on the other raise concerns about the possibility that the eukaryotic resident virome could be responsible for disease pathogenesis alone or in combination with environmental and genetic factors. This can be due to the latent stimulation of gastrointestinal immunity, as already discussed above. Indeed, continuous stimulation of the gastrointestinal tract may explode in chronic intestinal inflammation, such as IBD with an exacerbated immune response and uncontrolled inflammatory milieu [15]. This theory has been sustained by our and others' studies [15,21,32], even if no demonstration is available so far to support the causative role of virome in the pathogenesis of chronic intestinal inflammation.

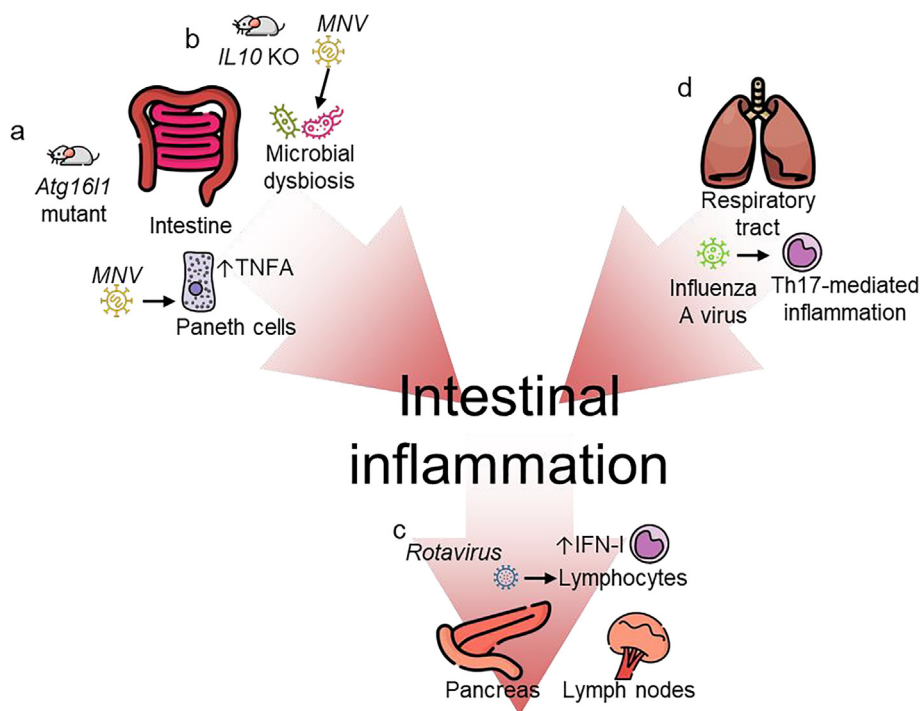
However, some indications have come from the models of DSS-induced experimental colitis in mice with mutations in the *Atg16l1* gene, whose alterations have been associated with Crohn's disease (CD) pathogenesis. In these mice, MNV infection accelerated colitis development through the virally-induced TNF $\alpha$ -dependent Paneth cell necroptosis (Fig. 4a) [61,62], similarly to what happened in *Il10* KO mice, where mucosal inflammation was induced by MNV and driven by microbiota [63] (Fig. 4b). These results led to the conclusion that enteric virus infections may trigger disease in a genetically predisposed individual. Also, enteric viruses can influence distant organs, as in the case of rotavirus infection, that can accelerate autoimmune diabetes onset in non-obese diabetic (NOD) mice, through the activation of lymphocytes in the pancreatic lymph nodes via IFN-I signalling [64,65] (Fig. 4c). Conversely, extraintestinal viruses may influence the gut microenvironment, as in the case of the *Influenza A virus*, that primarily affects the respiratory tract, but damages the intestinal tissue via microbiota-mediated Th17 inflammation [66] (Fig. 4d). These are few examples of a large amount of evidence describing possible immune activation after enteric virus infection leading to disease development, in organs that might also be far from the primary site of the viral insult [43].

### 5. Eukaryotic viruses and their contribution to carcinogenesis: Is that possible in the intestine?

The association between viral infections and neoplastic lesions has been largely studied and well-established [67,68] so that human cancers associated with oncoviral infections represent 15–20% of the worldwide cancer incidence [69]. Their oncogenic potential is now known to be mediated by various mechanisms, such as mutagenic integration into the host's genome and expression of oncogenic viral proteins [70], as well as their impact on cellular and inflammatory pathways [24]. The general mechanisms responsible for inducing tumorigenesis in the infected tissues are two: while the oncogenic RNA viruses cause cancer through mechanisms like the chronic inflammation and production of reactive oxygen species (ROS), the DNA viruses interfere with proteins regulating cell division checkpoints, apoptosis, telomere length regulation, and host cell's DNA repair mechanisms [71].

The greatest challenge in building a comprehensive list of oncoviruses is attributed to the lack of a unique and reproducible gold standard approach to identify them [15]. However, due to the rapid signs of progress in computational biology, the latest advances allowed steps forward in this field, and more precise information concerning the existence of viruses within the different niches of the human organism has been gained [69], as we will discuss later in this review.

Another issue concerns the difficulty of demonstrating a direct causality between viruses and carcinogenesis because of the latency period often occurring after a viral infection. Usually, human oncoviruses provoke latent or pseudo-latent infections, in which case they do not replicate to form viral particles and their lytic replication is reduced or even absent [72]. This aspect complicates the identification of a direct virus-tumor causality since oncoviruses remain hidden for a long time before establishing as carcinogenic actors and, in some cases, two decades are necessary before the cancer onset [73].



**Fig. 4.** Enteric and enterotropic viruses may influence the intestine. (a) MNV infection in *Atg16l1* mutation-carrying mice accelerates colitis development through TNF $\alpha$ -dependent Paneth cell necroptosis. (b) In *Il10* KO mice, MNV induces microbiota-driven intestinal inflammation. (c) Rotavirus infection activates lymphocytes in the pancreatic lymph nodes via IFN-I signalling. (d) *Influenza A virus* damages the intestinal tissue via Th17-mediated inflammation. The cartoon has been produced with google slides (<https://www.google.com/slides/about/>) and its freely available addons.

Biologically, this viral latency serves to escape immune surveillance by turning off unnecessary viral proteins that might be recognized and attacked by the immune system and thus counteracted. Upon latency, viruses exist as a naked nucleic acid, often as a plasmid or episome, which relies on the host cell's machinery to replicate whenever the cell divides. Nevertheless, the majority of virus-induced human cancers carry multiple viral integration events in their genomes [69], which induce mutagenesis, or the production of exogenous viral proteins. It is noteworthy that there are two types of viral persistence in the body: at the cellular level, where viruses infect a specific cell lineage during the lifetime, and at the organism level, where the viruses persist over the body in its entirety [74]. This could be important when we delineate some viruses as promoters and stimulators of the immune response in other organs, although they prefer and affect a particular cell type.

So far, the list of oncoviruses includes several entities, including the Epstein-Barr virus (EBV, also known as human herpesvirus 4; HHV4). The DNA-based EBV infects over 90% of the world's population during childhood and has been associated with a plethora of human malignancies in different organs [70,75,76]. It primarily infects B lymphocytes [77] and strongly impacts cellular and molecular machinery through EBV-derived oncoproteins that ultimately favor the tumor cell immune-escape and proliferation [78–81].

Human T-Lymphotropic virus type I (HTLV-1), annotated as an oncogenic RNA-based retrovirus, induces non-Hodgkin's peripheral T-cell malignancy called adult T-cell leukemia/lymphoma (ATL) in 5% of infected individuals [82], by randomly integrating into the host genome into fragile chromosomal regions [82,83]. Of note, besides ATL, HTLV-1 is associated with other pathological conditions, such as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP), HTLV-1-associated uveitis (HU), and various dermatological conditions [83], most likely related to high viral load and/or overstimulation of dendritic cells [84,85], thus indicating that a virus can interfere with the biology of many niches and not only with its cellular host.

Other examples of oncoviruses are the Hepatitis B Virus (HBV), and the Human Papilloma Virus (HPV), the former annotated as the main cause of the hepatocarcinoma (HCC) because of the random viral genome insertional events into the host's genome [86], the latter causative of the squamous neoplasia of the anogenital and head-and-neck regions [87] via the induction of genomic instability into the host cell [88,89].

Human Herpesvirus-8 (HHV-8), also known as Kaposi's sarcoma-associated herpesvirus (KSHV) [90], spends its latency period in the B-lymphocytic cells and vascular endothelium, intermittently switching to lytic replication, especially upon immunodeficiency and stressing conditions [91]. It encodes for transforming proteins and anti-apoptotic factors, promoting also angiogenesis, the release of host and viral cytokines, and cell proliferation [90].

Merkel cell polyomavirus (MCV or MCPyV) [75] is the major causative factor of Merkel cell carcinomas (MCC) first identified in 2008 as integrated into MCC-derived cells' genome at several different chromosomal sites with an identical clonal viral DNA integration [92], occurring early, if not at the beginning, in the MCC pathogenesis and without the involvement of cellular tumor suppressor genes or oncogenes [93].

These examples of oncoviruses can become part of the virome [94], ultimately impacting the tissue homeostasis [94]. Consequently, it is reasonable to look at specific entities colonizing the gut eukaryotic virome as possible triggers of intestinal cancer-causing mutational events and cancer-associated inflammation. Such a possibility is further supported by data reporting the disease-associated viruses, such as herpesviruses, polyomaviruses, papillomaviruses, polyomaviruses, hepatitis B virus, hepatitis C

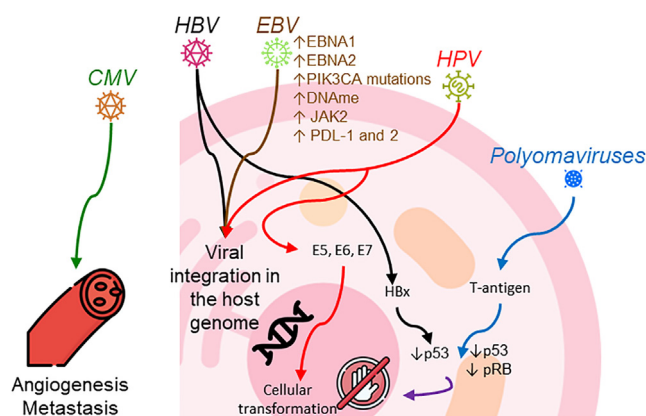
virus, and human immunodeficiency virus (HIV) to colonize the gut viromes of some individuals [21,95].

In line with this, changes in gut virome composition have been already associated with CRC, and specific viruses populating the virome have been uncovered as associated with the colorectal carcinogenic process [24].

A meta-analysis performed on 16 different studies, involving a total of 1436 patients, showed that the overall prevalence of HPV was about 31.9%, with the lowest in Europe (up to 5%). HPV 18 strain was frequently associated with CRC in Asia and Europe, whereas the 16 was the most prevalent in South America. Also, HPV positivity was found to increase the CRC risk [96] and persistent HPV infection is linked to anal and rectal cancer [97]. As already discussed above, the HPV-induced oncogenic mechanism might be due to the viral DNA integration into the host's genome with the subsequent transformation of the affected cells. However, the exact mechanism is yet to be investigated, even if putative actors encompass the HPV genome-encoded oncoviral proteins E5, E6, and E7 that may intervene in the neoplastic lesion onset in colorectal tissues [24] (Fig. 5, in red).

Human polyomavirus, including MCPyV, has also been associated with CRC [98]. Indeed, viral DNA sequences were displayed to be present in about 90% of the analyzed CRC samples, in both tumor areas and healthy margins. Furthermore, the associated lymphocytes were carrying polyomavirus infections [99]. Physiologically, polyomaviruses encode the T-antigen, an oncogenic protein mainly interacting with and inactivating p53 and pRB, besides interfering with other signaling pathways [100] and probably inducing CRC (Fig. 5, in blue).

Herpesviruses have been frequently detected in colon and rectum cancer. 21% of rectal cancers and 18% of colon cancers examined in a study contained the herpesviruses EBV, CMV, or HHV6 [101]. In the case of CMV, few studies demonstrated the disease-free survival rate to be lower in patients carrying CMV positive



**Fig. 5.** Possible mechanisms of viral-induced colorectal carcinogenesis. HPV genome can integrate within the host's, encoding the oncoviral proteins E5, E6, and E7 that may intervene in the neoplastic lesion onset in colorectal tissues (in red). Polyomaviruses encode the T-antigen, interacting with and inactivating p53 and pRB, finally contributing to colorectal neoplastic transformation (in blue). CMV infections may induce angiogenesis, resistance to apoptosis, cellular invasion, and metastasis (in green). EBV integrates within the host's genome, encoding nuclear antigens (EBNA1 and EBNA2), which subsequently results in PIK3Ca mutations, DNA hypermethylation, amplification of JAK2, and over-expression of PD-L1 and PD-L2 (in brown). HBV may induce carcinogenesis in the colon by integrating into the host's genome and thus encoding the Hepatitis B protein X (HBx). HBx interacts with p53, finally inducing cellular proliferation (in black). The violet arrow indicates the p53 and pRB let the cellular checkpoints to fail (hand-stop icon), ultimately causing cellular transformation. The cartoon has been produced with google slides (<https://www.google.com/slides/about/>) and its freely available addons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



tumors compared to those with the CMV negative [102]. It is plausible that such an association may derive from the virus's ability to induce angiogenesis, resistance to apoptosis, cellular invasion, and metastasis [103] (Fig. 5, in green).

EBV was largely studied in the context of CRC pathogenesis, albeit discrepant results have come out, mainly due to the different detection approaches utilized. Overall, EBV displayed an association with CRC [104]. A study from Salyakina and colleagues reported a common co-infection of EBV with other viruses in 20% of the CRC samples (CMV and HHV-6B) [105]. One of the proposed mechanisms inducing colorectal carcinogenesis relies on the integration of the viral DNA within the host's genome, leading to alterations of the DNA methylation pattern [106], which is known to be one of the molecular features of CRC pathogenesis [107]. Also, another possible mechanism is the production of EBV nuclear antigens (EBNA1 and EBNA2), which subsequently results in PIK3Ca mutations, DNA hypermethylation, amplification of JAK2, and over-expression of PD-L1 and PD-L2 [24] (Fig. 5, in brown).

Recently, the association between HBV infection and CRC was also strongly pointed out. Indeed, HBV infection was significantly associated with the risk of CRC in a cohort of Taiwanese subjects [108]. This study confirmed what was previously reported by Song and colleagues, who found HBV seropositive patients to display a higher CRC risk by comparison with the HBV negative subjects [109]. One possible mechanism for HBV to induce carcinogenesis in the colon might mimic that of HCC: it integrates into the host's genome and encodes for the Hepatitis B protein X (HBx), where the inactivation of p53 protein allows the cell to remain in the S-phase [24] (Fig. 5, in black). Possible mechanisms of oncovirus-induced colorectal carcinogenesis are summarized in Table 1.

Although these studies pinpointed the association of these eukaryotic viruses with CRC risk and pathogenesis, so far evidence for the causality existing between viral infections and gut carcinogenesis is yet to be uncovered. However, with the advent of innovative sequencing technologies and bioinformatic approaches, the possibility to uncover the mechanisms underlying the possible virus-induced colorectal carcinogenesis is now more real and feasible in the short-term future, although with some limitations.

## 6. The challenge of the virome discovery: What was done and what is next

The difficulties experienced in the study of the virome in any organism's niche appear evident when we realize that very few studies in the literature have been dedicated to the investigation of viruses if compared to those studying the bacterial population

[110]. In the past, some of the great limitations in studying the virome was the low sensitivity of sequencing technologies and few computational approaches not being able to thoroughly analyze the whole microbiome composition. This has been representing for long the first great challenge in studying the virome in different diseases, including the CRC.

Nevertheless, once recognized the important role of viruses in the disease pathogenesis [110], much more attention has been devoted to this field, and the rapid evolution of the computational and sequencing approaches in the last decades made the virome discovery significantly easier, even if some hurdles need to be still overcome.

A second consistent challenge in virome studies was to establish its causal role in disease pathogenesis. Indeed, many of the studies addressing virome profiling are just descriptive, and only associations between changes in virome composition and the pathogenic events in different disease contexts have been highlighted. In parallel, what is also missing is a consensus in identifying viral entities, differently from what happens for the bacteria. Indeed, whereas bacterial or fungal microbiome can be annotated based on the 16S rRNA and internal transcribed spacer (ITS) *loci*, respectively, the same is not possible for the viruses, since sequences conserved among the different types of viruses are usually lacking, impeding to analyze the virome systematically.

These hurdles make virome studies not only harder than the bacterial or fungal, but also more expensive in terms of experimental sequencing depth, computational infrastructure requirements, and computing costs.

Another challenge is represented by the lack of complete viral genome annotations, so that, when viral sequences are aligned to the reference genome, many of the reads remain unclassified (the so-called viral "dark matter") [111].

The viral dark matter of the virome, classified as viral sequences that do not align to any reference genome, may include the novel, highly divergent viruses that are unrecognizable and thus ignored, albeit they represent from 40% to 90% of the total sequences generated [29,112]. Therefore, this represents a great limitation in the whole virome profiling, leading to an incomplete and potentially biased analysis. To solve this issue, experimental approaches were designed to increase the viral sequences detected, such as the optimization of viral propagation methods and the viral tagging with fluorescent dyes, both resulted in the most useful approaches to correct viral detection, including the definition of the viral dark matter [113,114]. Being the viral dark matter a major complication in virome studies, this topic was already addressed in many studies and extensively reviewed elsewhere [110,111].

**Table 1**  
Possible molecular and cellular mechanisms inducing CRC.

Oncovirus	Evidence in CRC	Possible CRC-inducing mechanisms
HPV	<ul style="list-style-type: none"> <li>HPV positivity increases the CRC risk [115]</li> <li>Persistent HPV infection is linked to anal and rectal cancer [116]</li> </ul>	<ul style="list-style-type: none"> <li>HPV DNA integration into the host's genome and cell transformation;</li> <li>HPV-encoded E5, E6, and E7 proteins may intervene in the neoplastic lesion onset [36]</li> </ul>
MCPyV	<ul style="list-style-type: none"> <li>MCPyV DNA found in 90% of CRC samples analyzed, especially in lymphocytes [118].</li> </ul>	<ul style="list-style-type: none"> <li>MCPyV-induced CRC might rely on the T-antigen-mediated inactivation of p53 and pRB, as well as other signaling pathways [119].</li> </ul>
CMV	<ul style="list-style-type: none"> <li>CMV positivity tumors predicts low disease-free survival rate in CRC patients [121].</li> </ul>	<ul style="list-style-type: none"> <li>It is likely to induce angiogenesis, resistance to apoptosis, cellular invasion, and metastasis [122]</li> </ul>
EBV	<ul style="list-style-type: none"> <li>EBV positivity is associated with CRC pathogenesis [123]</li> <li>In 20% of CRC cases a coinfection by both EBV and CMV (or HHV-6B) was observed [124]</li> </ul>	<ul style="list-style-type: none"> <li>EBV DNA integration into the host's genome [125]</li> <li>Alteration of DNA methylation pattern [125]</li> <li>EBV nuclear antigens EBNA1 and 2 induce PIK3Ca mutations, DNA hypermethylation, JAK2 activation along with PD-L1 and PD-L2 overexpression [36].</li> </ul>
HBV	<ul style="list-style-type: none"> <li>HBV seropositivity increase CRC risk [109]</li> </ul>	<ul style="list-style-type: none"> <li>HBV-induced CRC might rely on the HBx-mediated inactivation of p53, sticking cells into the S-phase [36]</li> </ul>

Another constant bias found in many virome studies is the analysis of DNA extracted from either mucosal tissues or the stools (this is the case of the intestine). This led to the exclusion of the RNA viruses from the analysis, being thus ignored for long, though they represent the majority of microbial entities in the intestinal tract [110]. Also, many and maybe the majority of the works included the analysis of the stools, because of their easy-to-access nature. However, this put a magnification lens on the bacteriophages (found in the stools and parasitizing bacteria), inevitably disregarding eukaryotic-infecting viruses, which were only identified by chance upon stool analysis and therefore underestimated. Hence, to achieve a more comprehensive description of the virome, it is crucial that RNA viruses are preserved in the specimens and that both the RNA and DNA fractions are incorporated into the sequencing analysis.

Our basic knowledge concerning the viral component of the microbiota has been relying for a long on the use of culture (plaque)-based techniques for isolating phages from biological samples using a cultured bacterial host. This has been largely applied as an essential tool to understand phage-host interactions. The main advantage is that the bacterial host is known, and thus the system can be easily manipulated *ad hoc* for testing hypothesis in real-time. However, only a few bacteria can be cultured [117] and the *in vivo* translation of the *in vitro* findings is not that straightforward, because the laboratory conditions used do not reflect the complexity of the microbial population *in vivo* [118]. Although many novel eukaryotic viruses or phages are identified every year and added to the list of known viruses [119], culture systems for only a very limited number have been described to date [116,120,121]. Thus, besides the difficulties in identifying viruses in human samples (mainly in the past), there is a current urgency in developing culture models to study mechanisms associated with specific viruses. As far as it is demanding, more efforts are needed to address that point. As an example, the use of organoid systems to propagate human noroviruses and study their mechanism of replication seemed to be an appropriate tool in some works [122]. However, the route to full comprehension is still long and needs both protocol optimization and standardization.

The most and still used method to detect a viral sequence within a sample is the Polymerase Chain Reaction (PCR), which detects and quantifies both DNA and RNA viruses by simply designing primers targeting a known viral genomic region, followed by amplicon confirmation by Sanger sequencing [21]. The big hurdle is the discrimination of viral/phage abundance and diversity because of the limited available sequence data which primers can be designed on [118].

As a matter of fact, metagenomics represents an unbiased approach to uncover the viral genome within a sample. There are two main approaches for metagenomics studies: amplicon sequencing and whole-genome shotgun sequencing [123]. With amplicon sequencing, only a selected pool of genes is sequenced. Generally, the 16S rRNA genes are used for bacterial taxonomy-based classification of the species in a sample. However, this kind of approach is not suitable for virome profiling since viruses do not have any ribosomal genes. Besides, the viral composition cannot be defined by just aligning sequences to known genes, since most of the viruses are not annotated yet. Therefore, shotgun sequencing is preferentially used for the virome studies: here, the whole DNA is sequenced, non-interesting reads (like in our case human, bacterial and fungal) are removed and the remaining DNA is *de novo* assembled into contigs [123]. These contigs can represent hypothetically and optimistically complete viral genomes which can be studied from a functional (gene prediction) and a taxonomy point of view. However, this last method requires huge efforts for the starting steps necessary to recognize and differentiate the viral species contained in the sample, as well as heavy computational

costs for the *de novo* assembly step. Also, only DNA viruses can be studied, and thus RNA viruses are excluded [15]. This issue is overcome thanks to metatranscriptomics: by analyzing the entire transcriptome of a specimen and by annotating specific viral reads to the respective reference genomes, the discovery of both RNA and DNA viruses became possible [21].

### 6.1. Computational approaches for virome studies

Together with the latest advance in the Next-Generation Sequencing (NGS) to develop the most fruitful approach to uncover viral entities [15], great efforts were made to build systematic bioinformatics pipelines able to elucidate the composition of the virome ever since.

Currently, there is no gold standard pipeline to unveil the virome composition of mucosal tissues. Nevertheless, over the years, many computational approaches have been developed and tested. The majority are based on the exact alignment of the sequencing reads to the annotated viral genome references. Some approaches, less represented but not less important, are based on the k-mer classification. k-mer based methods classify the property of k-mers (substrings of length k in DNA sequence data [124]) in genome sequences in terms of frequencies or positions.

The workflow (Fig. 6) shows the typical researcher decision-making process while setting up a metagenomic/metatranscriptomics experiment. The real bottleneck is having or not an expert bioinformatician and the hardware necessary to perform such extensive and time-consuming types of analysis. Having this, a researcher would only need to be careful of the coverage required to properly represent the virome of interest. Indeed, some software requires the *de novo* assembly of either DNA or RNA viral genomes before read alignment and therefore needs high coverage, nowadays achieved only with specific enrichment procedures. Nevertheless, an increasing number of programs able to work with small sequences instead of “big” genomes is providing researchers with the possibility to investigate even under-represented and sometimes neglected viral niches.

Table 2 summarises currently available bioinformatics pipelines with their strengths and limitations. Having or not a web-app is pivotal for those researchers that do not have a computational biologist in their team. These pipelines are commonly not customizable, cannot be performed in bulk, and sensitive data need to be uploaded in public servers, but they nevertheless represent a valuable resource. When performing this kind of analysis the results are greatly improved by *de novo* genome assembly before mapping.

All these tools are publicly available and display benefits as well as limitations. Therefore, they should be used after careful consideration of sample type and scientific question.

Although the numerous steps forward for implementing and improving the virome analysis, many issues still need to be overcome, such as the sensitivity of the sequencing techniques and false-positive detections, due to not-standardized technical procedures for viral sequence extraction, limitations related to viral genomes annotations in the databases, often incorrect or poorly curated, and low sequencing coverage. Indeed, because of the underrepresentation of the viral genomes compared to the host's, the sequencing analysis needs to be deeper than that observed in common NGS experiments (i.e., about 50 million reads per sample for a total RNA-Seq) to be able to read thousands instead of dozens of viral genome mapping reads. However, such an approach becomes extremely expensive, rendering this type of analysis more challenging. Furthermore, particular viruses can fail to be detected because of their very high homology to other viruses, or sample cross-contamination [15,24].



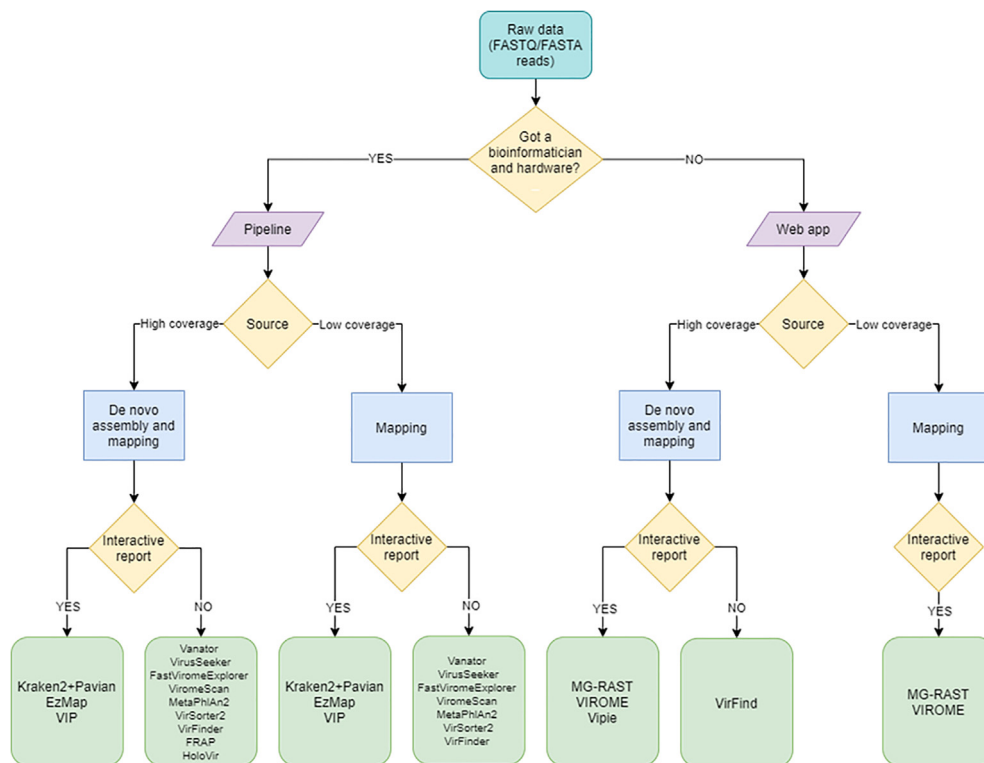


Fig. 6. The workflow of decision-making processes to set up a metagenomic/metatranscriptomic study.

## 7. Summary and outlook

CRC is a process that takes place in the distal colon, starting from a neoplastic lesion occurring in the bowel wall. From there, cells aberrantly proliferate, accumulating DNA alterations as well as promoting pro-inflammatory signaling and immune response [3,9].

In such a complex scenario, intestinal dysbiosis has been long associated with CRC, even if attention was much more devoted to bacteria than to other components of the intestinal microbiota. Among these, viruses, and mainly bacteriophages [37], have just begun attracting interest as modulators of the bacterial composition of the gut microbiota, also in CRC [37,125]. Hence, great attention was given to the bacteriophages colonizing CRC patients' intestines, disregarding eukaryotic-targeting viruses that infect the host's gut mucosa and might lead to CRC initiation and progression. This is indeed an open question: could eukaryotic viruses trigger CRC by inducing mutational events that transform the epithelial cells in the gut, finally leading to tumorigenesis? Although this might represent a real breakthrough in uncovering the mechanisms of CRC, ultimately paving the way to possible innovative treatments blocking the cause rather than counteracting the CRC-associated characteristics (i.e. tumor-associated inflammation), some limitations in this field need to be overcome first. Indeed, so far some of the major limitations in studying the virome have been the low sensitivity of sequencing technologies and the limited availability of computational approaches that were not able to thoroughly analyze the whole microbiome composition. However, new cutting-edge technologies in nucleic acid sequencing, as well as the innovative bioinformatics and statistics tools (Table 2 and Fig. 6) developed for analyzing a large amount of computational data will now provide us with the possibility to perform a more sensitive and accurate virome analysis, eventually identifying possible viral entities responsible for CRC onset.

Hence, putting together recent technological advances, as well as the development of adequate bioinformatics pipelines for virome discovery, a comprehensive picture of the entire virome could be fully delineated, providing details on taxa, orders, and families that might contribute to virus-induced disease aetiogenesis, including CRC. To render the computational approach much more successful, patient selection criteria need to be adequately formulated when human samples are analyzed. Indeed, to unveil the causative role of a specific viral entity during CRC pathogenesis, it would be of extreme importance to analyze samples from subjects with early disease, to increase the probability of identification of those viruses directly correlated with the aetiogenesis and not reactivated because of cancer-associated inflammation, immune suppressant treatments or subsequent viral infections. Importantly, genomic studies in combination with metagenomics/metatranscriptomics will be needed to associate highly frequent CRC-associated alterations with higher levels of specific eukaryotic viruses. Nevertheless, *in vitro* and *in vivo* validation of *in silico* data will result mandatory for establishing and demonstrating the causal role harbored by specific viral entities in inducing colorectal carcinogenesis.

Therefore, through high accuracy in both bioinformatics analysis and patient selection, as well as the experimental validation of computational results, it would be possible to reduce the risk of inappropriate and elusive studies, providing much more reliable results, also endowed with clinical implications. Indeed, whether a virus is found to be correlated with a specific CRC-associated phenotype, the use of antiviral drugs, inhibitors, or vaccines could block or maybe revert the carcinogenic process, hopefully reducing CRC relapse or ameliorating the pathogenic course. Moreover, viral entities could be used as markers of CRC occurrence or relapse. All these possibilities will be hopefully tested soon and this field will possibly gain all the attention it deserves.

**Table 2**  
Summary of the current available bioinformatics pipelines with their strengths (green) and limitations (red).

Software	Link	Web app	De novo assembly	Viral classes analyzed (based on baltimore classification)	Interactive report	Aligner	Alignment
MG-RAST	<a href="https://www.mg-rast.org/">https://www.mg-rast.org/</a>	Available	Not needed	AI	Available	Bowtie	reference genome
VIROME	<a href="http://virome.dbi.udel.edu/">http://virome.dbi.udel.edu/</a>	Available	Not needed	AI	Available	blast	reference genome
VirFind	<a href="http://virfind.org/">http://virfind.org/</a>	Available	Needed	I, II, VI (DNA viruses)	Not available	blast	de novo and reference genome
Vipie	<a href="http://vipie.rd.tuni.it/vipie/index.html">http://vipie.rd.tuni.it/vipie/index.html</a>	Available	Needed	I, II, VI (DNA viruses)	Available	blast	de novo genome
MetaVir 2	<a href="http://metavir-meb.univ-spclermont.fr/">http://metavir-meb.univ-spclermont.fr/</a>	Discontinued					
Kraken2 Pavian	<a href="https://ccb.jhu.edu/software/kraken2/">https://ccb.jhu.edu/software/kraken2/</a> <a href="https://github.com/fbreitwieser/pavian">https://github.com/fbreitwieser/pavian</a>	Not available	Not needed	AI	Available	internal classifier	reference genome
EzMap	<a href="https://github.com/dekonin9-lab/ezmap">https://github.com/dekonin9-lab/ezmap</a>	Not available	Not needed	AI	Available	bowtie2 and blast	reference genome
VIP	<a href="https://github.com/keyfabvdc/VIP/">https://github.com/keyfabvdc/VIP/</a>	Not available	Needed	AI	Available	internal classifier	de novo and reference genome
Vanator	<a href="https://sourceforge.net/projects/vanator-cvr/">https://sourceforge.net/projects/vanator-cvr/</a>	Not available	Needed	AI	Not available	BLAST	de novo and reference genome
VirusSeeker	<a href="http://pathology.wustl.edu/virusseeker/index.htm">http://pathology.wustl.edu/virusseeker/index.htm</a>	Not available	Not needed	AI	Not available	blast	reference genome
FastViromeExplorer	<a href="https://fastviromeexplorer.readthedocs.io/en/latest/">https://fastviromeexplorer.readthedocs.io/en/latest/</a>	Not available	Not needed	AI	Not available	kallisto	reference genome
ViromeScan	<a href="http://sourceforge.net/projects/viromescan/">http://sourceforge.net/projects/viromescan/</a>	Not available	Not needed	AI	Not available	bowtie2	reference genome
MetaPhiAn2	<a href="https://huttenhower.sph.harvard.edu/metaphlan">https://huttenhower.sph.harvard.edu/metaphlan</a>	Not available	Not needed	AI	Not available	internal classifier	reference genome
VirSorter2	<a href="https://github.com/jiarong/VirSorter2">https://github.com/jiarong/VirSorter2</a>	Not available	Not needed	AI	Not available	internal classifier and blast	reference genome
VirFinder	<a href="https://github.com/jessieren/VirFinder">https://github.com/jessieren/VirFinder</a>	Not available	Not needed	AI	Not available	internal classifier	reference genome
FRAP	<a href="https://github.com/yinacobian/frap">https://github.com/yinacobian/frap</a>	Not available	Needed	I, II, VI (DNA viruses)	Not available	internal classifier	de novo and reference genome
HoloVir	<a href="https://github.com/plaffy/HoloVir">https://github.com/plaffy/HoloVir</a>	Not available	Needed	I, II, VI (DNA viruses)	Not available	internal classifier	de novo and reference genome

## 8. Literature search strategy

A large and comprehensive literature through Medline (PubMed) and Google has been conducted to identify all relevant citations published within the last thirty years by using the following terms either alone or in combination: “colon”, “carcinoma”, “colon carcinoma”, “colorectal cancer”, “microbiota”, “intestinal homeostasis and colorectal cancer”, “virome”, “gut virome and colorectal cancer”, “eukaryotic viruses and colorectal carcinogenesis”, “viral-induced mutagenesis”, “virus-induced carcinogenesis”. Highly regarded relevant articles were not excluded *a priori*. Only studies exploring cellular, molecular, and clinical characteristics of colorectal cancer, gut virome, and microbiota have been selected. We also searched the reference lists of key review articles for additional papers we considered to be relevant to this manuscript.

## Author contributions

LM, SL, LAL, FU: conceptualization, writing, review, and editing; SD: supervision, review, and editing; FU funding acquisition.

## Declaration of Competing Interest

S.D. 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, Astra Zeneca, Novo Nordisk, Vifor, and Johnson and Johnson. The other authors declare no conflicts of interest.

## Acknowledgments

The authors would like to acknowledge the “Fondazione AIRC per la Ricerca sul Cancro” under the “My First AIRC Grant” agreement #23647 to FU; “Fondazione Cariplo per la ricerca Biomedica” under the grant agreement # 2018-0112 to FU; and Fondazione Umberto Veronesi for the 2020 Post-Doctoral Fellowship to SL.

## References

- Day WW. Environmental and occupational determinants of cancer : Interventions for primary prevention World Water Day : Water for cities Global disease burden involving chemicals WHO Framework for Risk Assessment of Combined Exposures to Multiple Chemicals Healthy 2011.
- Takala J, Hämäläinen P, Saarela KL, Yun LY, Manickam K, Jin TW, et al. Global estimates of the burden of injury and illness at work in 2012. *J Occup Environ Hyg* 2014;11:326–37. <https://doi.org/10.1080/15459624.2013.863131>.
- Ungaro F, D'Alessio S, Danese S. The role of pro-resolving lipid mediators in colorectal cancer-associated inflammation: implications for therapeutic strategies. *Cancers (Basel)* 2020;12:2060. <https://doi.org/10.3390/cancers12082060>.
- Mathonnet M. Hallmarks in colorectal cancer: angiogenesis and cancer stem-like cells. *World J Gastroenterol* 2014;20:4189. <https://doi.org/10.3748/wjg.v20.i15.4189>.
- Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74. <https://doi.org/10.1016/j.cell.2011.02.013>.
- Tacconi C, Correale C, Gandelli A, Spinelli A, Dejana E, D'Alessio S, et al. Vascular endothelial growth factor C disrupts the endothelial lymphatic barrier to promote colorectal cancer invasion. *Gastroenterology* 2015;148(1438–1451):. <https://doi.org/10.1053/j.gastro.2015.03.005e8>.
- Tacconi C, Ungaro F, Correale C, Arena V, Massimino L, Detmar M, et al. Activation of the VEGFC/VEGFR3 pathway induces tumor immune escape in colorectal cancer. *Cancer Res* 2019;79:4196–210. <https://doi.org/10.1158/0008-5472.CAN-18-3657>.
- Ungaro F, Colombo P, Massimino L, Ugolini GS, Correale C, Rasponi M, et al. Lymphatic endothelium contributes to colorectal cancer growth via the soluble matrix component GDF11. *Int J Cancer* 2019;ijc.32286. <https://doi.org/10.1002/ijc.32286>.
- Lei H, Tao K. Somatic mutations in colorectal cancer are associated with the epigenetic modifications. *J Cell Mol Med* 2020;jcmm.15799. <https://doi.org/10.1111/jcmm.15799>.
- Chang P-Y, Chen J-S, Chang S-C, Wang M-C, Chang N-C, Wen Y-H, et al. Acquired somatic TP53 or PIK3CA mutations are potential predictors of when polyps evolve into colorectal cancer. *Oncotarget* 2017;8:72352–62. <https://doi.org/10.18632/oncotarget.20376>.
- Perillo F, Amoroso C, Strati F, Giuffrè MR, Díaz-Basabe A, Lattanzi G, et al. Gut microbiota manipulation as a tool for colorectal cancer management: recent advances in its use for therapeutic purposes. *Int J Mol Sci* 2020;21:5389. <https://doi.org/10.3390/ijms21155389>.
- Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. *Nat Med* 2018;24:392–400. <https://doi.org/10.1038/nm.4517>.
- Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, et al. Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav Immun* 2015;48:186–94. <https://doi.org/10.1016/j.bbi.2015.03.016>.
- Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Mol Psychiatry* 2016;21:786–96. <https://doi.org/10.1038/mp.2016.44>.
- Ungaro F, Massimino L, D'Alessio S, Danese S. The gut virome in inflammatory bowel disease pathogenesis: From metagenomics to novel therapeutic approaches. *United Eur Gastroenterol J* 2019;7:999–1007. <https://doi.org/10.1177/2050640619876787>.
- Chen Y, Yang Y, Gu J. Clinical implications of the associations between intestinal microbiome and colorectal cancer progression. *Cancer Manag Res* 2020;12:4117–28. <https://doi.org/10.2147/CMAR.S240108>.
- Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature* 2012;489:220–30. <https://doi.org/10.1038/nature11550>.
- Norman JM, Handley SA, Baldrige MT, Droit L, Liu CY, Keller BC, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015;160:447–60. <https://doi.org/10.1016/j.cell.2015.01.002>.
- Caputo M, Zoch-Lesniak B, Karch A, Vital M, Meyer F, Klawonn F, et al. Bacterial community structure and effects of picornavirus infection on the anterior nares microbiome in early childhood. *BMC Microbiol* 2019;19:1. <https://doi.org/10.1186/s12866-018-1372-8>.
- Yinda CK, Vanhulle E, Conceição-Neto N, Beller L, Deboutte W, Shi C, et al. Gut virome analysis of cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *MSphere* 2019;4. <https://doi.org/10.1128/mSphere.00585-18>.
- Ungaro F, Massimino L, Furfaro F, Rimoldi V, Peyrin-Biroulet L, D'Alessio S, et al. Metagenomic analysis of intestinal mucosa revealed a specific eukaryotic gut virome signature in early-diagnosed inflammatory bowel disease. *Gut Microbes* 2018;1–10. <https://doi.org/10.1080/19490976.2018.1511664>.
- Rositch AF. Global burden of cancer attributable to infections: the critical role of implementation science. *Lancet Glob Heal* 2020;8:e153–4. [https://doi.org/10.1016/S2214-109X\(20\)30001-2](https://doi.org/10.1016/S2214-109X(20)30001-2).
- de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Heal* 2020;8:e180–90. [https://doi.org/10.1016/S2214-109X\(19\)30488-7](https://doi.org/10.1016/S2214-109X(19)30488-7).
- Emler C, Ruffin M, Lamendella R. Enteric Virome and Carcinogenesis in the Gut. *Dig Dis Sci* 2020;65:852–64. <https://doi.org/10.1007/s10620-020-06126-4>.
- Pride DT, Abeles SR, Robles-Sikisaka R, Naidu M, Boehm TK, Ly M, et al. Altered oral viral ecology in association with periodontal disease. *MBio* 2014;5:e01133–1214. <https://doi.org/10.1128/mbio.01133-14>.
- Fernandes MA, Verstraete SG, Phan TG, Deng X, Stekol E, LaMere B, et al. Enteric virome and bacterial microbiota in children with ulcerative colitis and crohn disease. *J Pediatr Gastroenterol Nutr* 2019;68:30–6. <https://doi.org/10.1097/MPG.0000000000002140>.
- Nosrat B, Silva J, Rohwer F, Conrad D, Schmieder R, Tammadoni S, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* 2009;4:.. <https://doi.org/10.1371/journal.pone.0007370e7370>.
- Moran Losada P, Chouvarine P, Dorda M, Hedtfeld S, Mielke S, Schulz A, et al. The cystic fibrosis lower airways microbial metagenome. *ERJ Open Res* 2016;2:00096–2015. <https://doi.org/10.1183/23120541.00096-2015>.
- Santiago-Rodríguez TM, Hollister EB. Human virome and disease: high-throughput sequencing for virus discovery, identification of phage-bacteria dysbiosis and development of therapeutic approaches with emphasis on the human gut. *Viruses* 2019;11:656. <https://doi.org/10.3390/v11070656>.
- Pfeiffer JK, Virgin HW. Transkingdom control of viral infection and immunity in the mammalian intestine. *Science (80-)* 2016;351:aad5872–aad5872. <https://doi.org/10.1126/science.aad5872>.
- Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili S-M, Bahreini E. A comprehensive review of COVID-19 characteristics. *Biol Proced Online* 2020;22:19. <https://doi.org/10.1186/s12575-020-00128-2>.
- Virgin HW. The virome in mammalian physiology and disease. *Cell* 2014;157:142–50. <https://doi.org/10.1016/j.cell.2014.02.032>.
- Matijašić M, Meštrović T, Paljetak HC, Perić M, Barešić A, Verbanac D. Gut microbiota beyond bacteria-mycobiome, virome, archaeome, and eukaryotic parasites in IBD. *Int J Mol Sci* 2020;21:.. <https://doi.org/10.3390/ijms21082668>.



- [34] Seo S-U, Kweon M-N. Virome–host interactions in intestinal health and disease. *Curr Opin Virol* 2019;37:63–71. <https://doi.org/10.1016/j.coviro.2019.06.003>.
- [35] Manrique P, Dills M, Young M. The human gut phage community and its implications for health and disease. *Viruses* 2017;9:141. <https://doi.org/10.3390/v9060141>.
- [36] Handley SA, Devkota S. Going viral: a novel role for bacteriophage in colorectal cancer. *MBio* 2019;10. <https://doi.org/10.1128/mBio.02626-18>.
- [37] Szeto C-H, Lam T-Y-T, Li X, Sung J-J-Y, Chan FKL, Chan MTV, et al. Alterations in enteric virome are associated with colorectal cancer and survival outcomes. *Gastroenterology* 2018;155(529–541):. <https://doi.org/10.1053/j.gastro.2018.04.018e5>.
- [38] Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell* 2009;138:30–50. <https://doi.org/10.1016/j.cell.2009.06.036>.
- [39] Metzger R, Krug A, Eisenächer K. Enteric virome sensing—its role in intestinal homeostasis and immunity. *Viruses* 2018;10:146. <https://doi.org/10.3390/v10040146>.
- [40] Hoyles L, McCartney AL, Neve H, Gibson GR, Sanderson JD, Heller KJ, et al. Characterization of virus-like particles associated with the human faecal and caecal microbiota. *Res Microbiol* 2014;165:803–12. <https://doi.org/10.1016/j.resmic.2014.10.006>.
- [41] Castro-Mejía JL, Muhammed MK, Kot W, Neve H, Franz CMAP, Hansen LH, et al. Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* 2015;3:64. <https://doi.org/10.1186/s12915-015-0131-4>.
- [42] Garmeaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol* 2019;17:84. <https://doi.org/10.1186/s12915-019-0704-y>.
- [43] Neil JA, Cadwell K. The intestinal virome and immunity. *J Immunol* 2018;201:1615–24. <https://doi.org/10.4049/jimmunol.1800631>.
- [44] Zuo T, Lu X-J, Zhang Y, Cheung CP, Lam S, Zhang F, et al. Gut mucosal virome alterations in ulcerative colitis. *Gut* 2019;gutjnl-2018-318131. <https://doi.org/10.1136/gutjnl-2018-318131>.
- [45] Wang W, Jovel J, Halloran B, Wine E, Patterson J, Ford G, et al. Metagenomic analysis of microbiome in colon tissue from subjects with inflammatory bowel diseases reveals interplay of viruses and bacteria. *Inflamm Bowel Dis* 2015;21:1419–27. <https://doi.org/10.1097/MIB.0000000000000344>.
- [46] Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* (80-) 2016;351:1083–7. <https://doi.org/10.1126/science.aad5497>.
- [47] Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol* 2019. <https://doi.org/10.1038/s41575-019-0189-8>.
- [48] Markowitz SD, Bertagnoli MM. Molecular basis of colorectal cancer. *N Engl J Med* 2009;361:2449–60. <https://doi.org/10.1056/NEJMra0804588>.
- [49] Bongers G, Pacer ME, Geraldino TH, Chen L, He Z, Hashimoto D, et al. Interplay of host microbiota, genetic perturbations, and inflammation promotes local development of intestinal neoplasms in mice. *J Exp Med* 2014;211:457–72. <https://doi.org/10.1084/jem.20131587>.
- [50] Slowicka K, Petta I, Blancke G, Hoste E, Dumas E, Sze M, et al. Zeb2 drives invasive and microbiota-dependent colon carcinoma. *Nat Cancer* 2020;1:620–34. <https://doi.org/10.1038/s43018-020-0070-2>.
- [51] Hu B, Elinav E, Huber S, Strowig T, Hao L, Hafemann A, et al. Microbiota-induced activation of epithelial IL-6 signaling links inflammasome-driven inflammation with transmissible cancer. *Proc Natl Acad Sci* 2013;110:9862–7. <https://doi.org/10.1073/pnas.1307575110>.
- [52] Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan T-J, et al. Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science* (80-) 2012;338:120–3. <https://doi.org/10.1126/science.1224820>.
- [53] Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli*. *Nature* 2020;580:269–73. <https://doi.org/10.1038/s41586-020-2080-8>.
- [54] Yang Y, Gharaibeh RZ, Newsome RC, Jobin C. Amending microbiota by targeting intestinal inflammation with TNF blockade attenuates development of colorectal cancer. *Nat Cancer* 2020;1:723–34. <https://doi.org/10.1038/s43018-020-0078-7>.
- [55] Gogokhia L, Buhrke K, Bell R, Hoffman B, Brown DG, Hanke-Gogokhia C, et al. Expansion of bacteriophages is linked to aggravated intestinal inflammation and colitis. *Cell Host Microbe* 2019;25(285–299):. <https://doi.org/10.1016/j.chom.2019.01.008e8>.
- [56] Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* 2014. <https://doi.org/10.1038/nature13960>.
- [57] Barton ES, White DW, Cathelyn JS, Brett-McClellan KA, Engle M, Diamond MS, et al. Herpesvirus latency confers symbiotic protection from bacterial infection. *Nature* 2007;447:326–9. <https://doi.org/10.1038/nature05762>.
- [58] Ingle H, Lee S, Ai T, Orvedahl A, Rodgers R, Zhao G, et al. Viral complementation of immunodeficiency confers protection against enteric pathogens via interferon- $\lambda$ . *Nat Microbiol* 2019;4:1120–8. <https://doi.org/10.1038/s41564-019-0416-7>.
- [59] Yang J-Y, Kim M-S, Kim E, Cheon JH, Lee Y-S, Kim Y, et al. Enteric viruses ameliorate Gut inflammation via toll-like receptor 3 and toll-like receptor 7-mediated interferon- $\beta$  production. *Immunity* 2016;44:889–900. <https://doi.org/10.1016/j.immuni.2016.03.009>.
- [60] Onyeagocha C, Hossain MS, Kumar A, Jones RM, Roback J, Gewirtz AT. Latent cytomegalovirus infection exacerbates experimental colitis. *Am J Pathol* 2009;175:2034–42. <https://doi.org/10.2353/ajpath.2009.090471>.
- [61] Cadwell K, Patel KK, Maloney NS, Liu T-C, Ng ACY, Storer CE, et al. Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell* 2010;141:1135–45. <https://doi.org/10.1016/j.cell.2010.05.009>.
- [62] Matsuzawa-Ishimoto Y, Shono Y, Gomez LE, Hubbard-Lucey VM, Cammer M, Neil J, et al. Autophagy protein ATG16L1 prevents necroptosis in the intestinal epithelium. *J Exp Med* 2017;214:3687–705. <https://doi.org/10.1084/jem.20170558>.
- [63] Basic M, Keubler LM, Buettner M, Achard M, Breves G, Schröder B, et al. Norovirus triggered microbiota-driven mucosal inflammation in interleukin 10-deficient mice. *Inflamm Bowel Dis* 2014;20:431–43. <https://doi.org/10.1097/01.MIB.0000441346.86827.ed>.
- [64] Pane JA, Webster NL, Coulson BS. Rotavirus activates lymphocytes from non-obese diabetic mice by triggering toll-like receptor 7 signaling and interferon production in plasmacytoid dendritic cells. *PLoS Pathog* 2014;10:.. <https://doi.org/10.1371/journal.ppat.1003998>e1003998.
- [65] Pane JA, Fleming FE, Graham KL, Thomas HE, Kay TWH, Coulson BS. Rotavirus acceleration of type 1 diabetes in non-obese diabetic mice depends on type 1 interferon signalling. *Sci Rep* 2016;6:29697. <https://doi.org/10.1038/srep29697>.
- [66] Wang J, Li F, Wei H, Lian Z-X, Sun R, Tian Z. Respiratory influenza virus infection induces intestinal immune injury via microbiota-mediated Th17 cell-dependent inflammation. *J Exp Med* 2014;211:2397–410. <https://doi.org/10.1084/jem.20140625>.
- [67] Martin D, Gutkind JS. Human tumor-associated viruses and new insights into the molecular mechanisms of cancer. *Oncogene* 2008;27:531–42. <https://doi.org/10.1038/onc.2009.351>.
- [68] Vyshenska D, Lam KC, Shulzhenko N, Morgun A. Interplay between viruses and bacterial microbiota in cancer development. *Semin Immunol* 2017. <https://doi.org/10.1016/j.smim.2017.05.003>.
- [69] Chen X, Kost J, Sulovari A, Wong N, Liang WS, Cao J, et al. A virome-wide clonal integration analysis platform for discovering cancer viral etiology. *Genome Res* 2019;29:819–30. <https://doi.org/10.1101/gr.242529.118>.
- [70] Khoury JD, Tannir NM, Williams MD, Chen Y, Yao H, Zhang J, et al. Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J Virol* 2013;87:8916–26. <https://doi.org/10.1128/JVI.00340-13>.
- [71] McLaughlin-Drubin ME, Munger K. Viruses associated with human cancer. *Biochim Biophys Acta – Mol Basis Dis* 2008;1782:127–50. <https://doi.org/10.1016/j.bbadis.2007.12.005>.
- [72] zur Hausen H. The search for infectious causes of human cancers: where and why. *Virology* 2009;392:1–10. <https://doi.org/10.1016/j.virus.2009.06.001>.
- [73] Schiller JT, Lowy DR. Virus Infection and Human Cancer: An Overview, 2014, p. 1–10. [https://doi.org/10.1007/978-3-642-38965-8\\_1](https://doi.org/10.1007/978-3-642-38965-8_1).
- [74] Pandey NV. DNA viruses and cancer: insights from evolutionary biology. *VirusDisease* 2020;31:1–9. <https://doi.org/10.1007/s13337-019-00563-0>.
- [75] Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat Rev Cancer* 2010;10:878–89. <https://doi.org/10.1038/nrc2961>.
- [76] Cavallo S. Immune-mediated genesis of multiple sclerosis. *J Transl Autoimmun* 2020;3:.. <https://doi.org/10.1016/j.jtauto.2020.100039>100039.
- [77] Hatto OL, Harris-Arnold A, Schaffert S, Krams SM, Martinez OM. The interplay between Epstein-Barr virus and B lymphocytes: implications for infection, immunity, and disease. *Immunol Res* 2014;58:268–76. <https://doi.org/10.1007/s12026-014-8496-1>.
- [78] Satoh T, Wada R, Yajima N, Imaizumi T, Yagihashi S. Tumor microenvironment and RIG-I signaling molecules in Epstein-Barr virus-positive and -negative classical Hodgkin lymphoma of the elderly. *J Clin Exp Hematop* 2014;54:75–84. <https://doi.org/10.3960/jslr.54.75>.
- [79] Wu R, Sattarzadeh A, Rutgers B, Diepstra A, van den Berg A, Visser L. The microenvironment of classical Hodgkin lymphoma: heterogeneity by Epstein-Barr virus presence and location within the tumor. *Blood Cancer J* 2016;6:e417–e417. <https://doi.org/10.1038/bcj.2016.26>.
- [80] Zhao J, Liang Q, Cheung K-F, Kang W, Lung RWM, Tong JHM, et al. Genome-wide identification of Epstein-Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer* 2013;119:304–12. <https://doi.org/10.1002/ncr.27724>.
- [81] Marquitz AR, Mathur A, Shair KHY, Raab-Traub N. Infection of Epstein-Barr virus in a gastric carcinoma cell line induces anchorage independence and global changes in gene expression. *Proc Natl Acad Sci* 2012;109:9593–8. <https://doi.org/10.1073/pnas.1202910109>.
- [82] Panfil AR, Martinez MP, Ratner L, Green PL. Human T-cell leukemia virus-associated malignancy. *Curr Opin Virol* 2016;20:40–6. <https://doi.org/10.1016/j.coviro.2016.08.009>.
- [83] Giam C-Z, Semmes O. HTLV-1 infection and adult T-cell leukemia/lymphoma—A tale of two proteins: tax and HBZ. *Viruses* 2016;8:161. <https://doi.org/10.3390/v8060161>.
- [84] Olindo S, Lézin A, Cabre P, Merle H, Saint-Vil M, Edimonana Kaptue M, et al. HTLV-1 proviral load in peripheral blood mononuclear cells quantified in 100 HAM/TSP patients: a marker of disease progression. *J Neurol Sci* 2005;237:53–9. <https://doi.org/10.1016/j.jns.2005.05.010>.
- [85] Tattermusch S, Skinner JA, Chaussabel D, Banchemau J, Berry MP, McNab FW, et al. Systems biology approaches reveal a specific interferon-inducible

- signature in HTLV-1 associated myelopathy. *PLoS Pathog* 2012;8:. <https://doi.org/10.1371/journal.ppat.1002480>.
- [86] Hilleman MR. Critical overview and outlook: pathogenesis, prevention, and treatment of hepatitis and hepatocarcinoma caused by hepatitis B virus. *Vaccine* 2003;21:4626–49. [https://doi.org/10.1016/S0264-410X\(03\)00529-2](https://doi.org/10.1016/S0264-410X(03)00529-2).
- [87] Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer* 2010;10:550–60. <https://doi.org/10.1038/nrc2886>.
- [88] Wentzensen N, Vinokurova S, von Doerberitz MK. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res* 2004;64:3878–84. <https://doi.org/10.1158/0008-5472.CAN-04-0009>.
- [89] Ziegert C, Wentzensen N, Vinokurova S, Kisselov F, Eienkel J, Hoeckel M, et al. A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. *Oncogene* 2003;22:3977–84. <https://doi.org/10.1038/sj.onc.1206629>.
- [90] Dow DE, Cunningham CK, Buchanan AM. A review of human herpesvirus 8, the Kaposi's sarcoma-associated herpesvirus, in the pediatric population. *J Pediatric Infect Dis Soc* 2014;3:66–76. <https://doi.org/10.1093/pids/pit051>.
- [91] Rewane A, Tadi P. Herpes Virus Type 8 (HHV 8). 2020.
- [92] Feng H, Shuda M, Chang Y, Moore PS. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* (80-) 2008;319:1096–100. <https://doi.org/10.1126/science.1152586>.
- [93] Starrett GJ, Marcellus C, Cantalupo PG, Katz JP, Cheng J, Akagi K, et al. Merkel cell polyomavirus exhibits dominant control of the tumor genome and transcriptome in virus-associated merkel cell carcinoma. *MBio* 2017;8. <https://doi.org/10.1128/mBio.02079-16>.
- [94] Stern J, Miller G, Li X, Saxena D. Virome and bacteriome: two sides of the same coin. *Curr Opin Virol* 2019;37:37–43. <https://doi.org/10.1016/j.coviro.2019.05.007>.
- [95] Mukhopadhyay I, Segal JP, Carding SR, Hart AL, Hold GL. The gut virome: the 'missing link' between gut bacteria and host immunity? *Therap Adv Gastroenterol* 2019;12:175628481983662. <https://doi.org/10.1177/1756284819836620>.
- [96] Damin DC, Ziegelmann PK, Damin AP. Human papillomavirus infection and colorectal cancer risk: a meta-analysis. *Color Dis* 2013;15:e420–8. <https://doi.org/10.1111/codi.12257>.
- [97] Lui RN, Tsoi KKF, Ho JMW, Lo CM, Chan FCH, Kyaw MH, et al. Global increasing incidence of young-onset colorectal cancer across 5 continents: a jointpoint regression analysis of 1,922,167 cases. *Cancer Epidemiol Biomarkers Prev* 2019;28:1275–82. <https://doi.org/10.1158/1055-9965.EPI-18-1111>.
- [98] Prado J, Monezi T, Amorim A, Lino V, Paladino A, Boccardo E. Human polyomaviruses and cancer: an overview. *Clinics* 2018;73. <https://doi.org/10.6061/clinics/2018/e558s>.
- [99] Casini B, Borgese L, Del Nonno F, Galati G, Izzo L, Caputo M, et al. Presence and incidence of DNA sequences of human polyomaviruses BKV and JCV in colorectal tumor tissues. *Anticancer Res* n.d.;25:1079–85.
- [100] Rollison DE. JC virus infection. *J Clin Gastroenterol* 2010;1. <https://doi.org/10.1097/MCG.0b013e3181e0084b>.
- [101] Cantalupo PG, Katz JP, Pipas JM. Viral sequences in human cancer. *Virology* 2018;513:208–16. <https://doi.org/10.1016/j.virol.2017.10.017>.
- [102] Chen H-P, Jiang J-K, Lai P-Y, Chen C-Y, Chou T-Y, Chen Y-C, et al. Tumoral presence of human cytomegalovirus is associated with shorter disease-free survival in elderly patients with colorectal cancer and higher levels of intratumoral interleukin-17. *Clin Microbiol Infect* 2014;20:664–71. <https://doi.org/10.1111/1469-0691.12412>.
- [103] Cinatl J, Vogel J-U, Kotchetkov R, Wilhelm Doerr H. Oncomodulatory signals by regulatory proteins encoded by human cytomegalovirus: a novel role for viral infection in tumor progression. *FEMS Microbiol Rev* 2004;28:59–77. <https://doi.org/10.1016/j.femsre.2003.07.005>.
- [104] Bedri S, Sultan AA, Alkhalaf M, Al Moustafa A-E, Vranic S. Epstein-Barr virus (EBV) status in colorectal cancer: a mini review. *Hum Vaccin Immunother* 2019;15:603–10. <https://doi.org/10.1080/21645515.2018.1543525>.
- [105] Salyakina D, Tsinoremas NF. Viral expression associated with gastrointestinal adenocarcinomas in TCGA high-throughput sequencing data. *Hum Genomics* 2013;7:23. <https://doi.org/10.1186/1479-7364-7-23>.
- [106] Karpinski P, Myszk A, Ramsey D, Kielan W, Sasiadek MM. Detection of viral DNA sequences in sporadic colorectal cancers in relation to CpG island methylation and methylator phenotype. *Tumor Biol* 2011;32:653–9. <https://doi.org/10.1007/s13277-011-0165-6>.
- [107] Tse JWT, Jenkins LJ, Chionh F, Mariadason JM. Aberrant DNA methylation in colorectal cancer: what should we target?. *Trends in Cancer* 2017;3:698–712. <https://doi.org/10.1016/j.trecan.2017.08.003>.
- [108] Su F-H, Le TN, Muo C-H, Te SA, Sung F-C, Yeh C-C. Chronic hepatitis B virus infection associated with increased colorectal cancer risk in taiwanese population. *Viruses* 2020;12:97. <https://doi.org/10.3390/v12010097>.
- [109] Song C, Lv J, Liu Y, Chen JG, Ge Z, Zhu J, et al. Associations between hepatitis B virus infection and risk of all cancer types. *JAMA Netw Open* 2019;2:. <https://doi.org/10.1001/jamanetworkopen.2019.5718>.
- [110] Wang D. 5 challenges in understanding the role of the virome in health and disease. *PLOS Pathog* 2020;16:. <https://doi.org/10.1371/journal.ppat.1008318>.
- [111] Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. *Virus Res* 2017;239:136–42. <https://doi.org/10.1016/j.virusres.2017.02.002>.
- [112] Guerin E, Hill C. Shining light on human gut bacteriophages. *Front Cell Infect Microbiol* 2020;10:. <https://doi.org/10.3389/fcimb.2020.00481>.
- [113] Wu Q, Wang Y, Cao M, Pantaleo V, Burgyan J, Li W-X, et al. Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. *Proc Natl Acad Sci* 2012;109:3938–43. <https://doi.org/10.1073/pnas.1117815109>.
- [114] Zhang Z, Qi S, Tang N, Zhang X, Chen S, Zhu P, et al. Discovery of replicating circular RNAs by RNA-Seq and computational algorithms. *PLoS Pathog* 2014;10:. <https://doi.org/10.1371/journal.ppat.1004553>.
- [115] Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 2015;3:. <https://doi.org/10.7717/peerj.985>.
- [116] Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *Elife* 2015;4:. <https://doi.org/10.7554/elife.08490>.
- [117] Rappé MS, Giovannoni SJ. The uncultured microbial majority. *Annu Rev Microbiol* 2003;57:369–94. <https://doi.org/10.1146/annurev.micro.57.030502.090759>.
- [118] Ogilvie LA, Jones BV. The human gut virome: a multifaceted majority. *Front Microbiol* 2015;6. <https://doi.org/10.3389/fmicb.2015.00918>.
- [119] Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. *Philos Trans R Soc B Biol Sci* 2012;367:2864–71. <https://doi.org/10.1098/rstb.2011.0354>.
- [120] Chiu CY, Greninger AL, Chen EC, Haggerty TD, Parsonnet J, Delwart E, et al. Cultivation and serological characterization of a human Theiler's-like cardiomyocytovirus associated with diarrheal disease. *J Virol* 2010;84:4407–14. <https://doi.org/10.1128/JVI.02536-09>.
- [121] Janowski AB, Bauer IK, Holtz LR, Wang D. Propagation of astrovirus VA1, a neurotropic human astrovirus, in cell culture. *J Virol* 2017;91. <https://doi.org/10.1128/JVI.00740-17>.
- [122] Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, et al. Replication of human noroviruses in stem cell-derived human enteroids. *Science* 2016;353:1387–93. <https://doi.org/10.1126/science.aaf5211>.
- [123] Liu Y-X, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 2020. <https://doi.org/10.1007/s13238-020-00724-8>.
- [124] Melsted P, Pritchard JK. Efficient counting of k-mers in DNA sequences using a bloom filter. *BMC Bioinf* 2011;12:333. <https://doi.org/10.1186/1471-2105-12-333>.
- [125] Hannigan GD, Duhaime MB, Ruffin MT, Koumpouras CC, Schloss PD. Diagnostic potential and interactive dynamics of the colorectal cancer virome. *MBio* 2018;9. <https://doi.org/10.1128/mBio.02248-18>.