65760_Auto_Edited - 副本.docx

WORD COUNT 5546
TIME SUBMITTED 20-SEP-2021 10:57PM
PAPER ID 76841521

Name of Journal: *World Journal of Clinical Oncology*

Manuscript NO: 65760

Manuscript Type: REVIEW

Systems biology and OMIC data integration to understand gastrointestinal cancers

Bispo IMC *et al*. An overview in gastrointestinal cancers OMIC data

Iasmin Moreira Costa Bispo, Henry Paul Granger, Palloma Porto Almeida, Patricia Belini Nishiyama, Leandro Martins de Freitas

Abstract

Gastrointestinal cancers are a set of diverse diseases affecting many parts/organs. The five most frequent gastrointestinal cancer types are esophageal, gastric, liver, pancreas, and colorectal; together, they affect 5 million new cases and cause the death of 3.5 million people annually. We bring information about molecular changes crucial to tumorigenesis and can give information about behavior and prognosis. During the formation of cancer cells, the genomic changes are microsatellite instability with multiple chromosomal arrangements in gastric and colorectal cancer. The genomically stable subtype is observed in gastric and pancreatic cancer. Besides these genomic subtypes, the colorectal cancer also has epigenetic modification (hypermethylation) associated with poor prognosis. The pathway information highlights the functions shared by gastrointestinal cancers such as apoptosis, focal adhesion, PAK pathway, PI3K-Akt signaling pathway, TGF-β pathway, and Toll-like receptor signaling pathway. These pathways show survival, cell proliferation, and cell motility. Besides, the immune response and inflammation is also an essential element in the shared functions. We also retrieved information on protein-protein interaction from the STRING database, and we found the proteins AKT1, CTNNB1, EP300, TP53, and TP53BP1 as central nodes in the network. The protein expression is reported to be associated with overall survival in some gastrointestinal cancers. The TP53BP1 low expression in colorectal cancer, EP300 high expression in esophageal cancer, high expression of AKT1/TP53, or CTNNB1 Low expression in gastric cancer is associated with a poor prognosis. The Kaplan Meier plotter database also confirms the association between the five central genes' expression and gastric cancer survival. In conclusion, gastrointestinal cancers are very diverse at the molecular level. However, the shared mutations and protein pathways might be used to understand better and reveal diagnostic/prognostic or drug targets.

Key Words: Gastrointestinal cancers; Genome; Cellular pathways; Protein-protein interaction; Prognosis; OMIC data

Bispo IMC, Granger HP, Almeida PP, Nishiyama PB, Freitas LM. Systems biology and OMIC data integration to understand gastrointestinal cancers. *World J Clin Oncol* 2021; In press

Core Tip: We highlight the genomic mutations and cellular pathways shared by gastrointestinal cancers. These are responsible for the cell's behavior that allows unlimited cell replication and invasion of other tissues. Using the STRING database, we found AKT1, CTNNB1, EP300, TP53, and TP53BP1 as central nodes in the gastrointestinal cancers protein network. The central nodes' expression is associated with poor survival in some gastrointestinal cancers. The Kaplan Meier plotter database also confirms the association between these five central genes expression and gastric cancer survival. Together, this information point to central and shared aspects of the most frequent gastrointestinal cancers.

INTRODUCTION

In 2020, the number of cancer cases in the digestive system was 5 million and 3.5 million deaths worldwide[1,2]; the system with the highest number of cases and also among the highest percentage of deaths[3] (Table 1). The cancer types in this system can be classified as the organ origin and cell type. The most frequent are esophageal, gastric, liver, pancreas, and colorectal[2,3]. The cancers in the gastric, liver, and colorectal are among the most common causes of cancer deaths annually[2]. Gastrointestinal cancers also have specific molecular changes in genetic/genome, epigenetics, gene expression, and cellular pathways contributing to tumor behavior. This information might be helpful in diagnosis, prognosis, and new drug development.

Esophageal cancer

Esophageal cancer (EC) has two subtypes: esophageal squamous cell cancer (ESCC) and esophageal adenocarcinoma (EAC)[4]. The incidence of ESCC increases globally and predominantly in Eastern Asia and Eastern/Southern[4–7]. However, the ESCC decreases

while EAC increases in the USA and a few European countries[5]. The ESSC and EAC incidence differences are geographically observed in sex and ethnic patterns[4,5].

There is also a well-established genetic factor associated with sex, and although it is still not well understood, it is known that there is a prevalence of 2.5:1-4.4:1 of men to women[4,6]. Studies indicate a protective effect of female sex hormones, including providing a lower risk of cancer for women who have previously breastfed. Nevertheless, environmental factors also influence this prevalence as, for example, men tend to abuse alcohol and tobacco, which are primary risk factors for the manifestation of EC[4,8].

The risk factors to ESCC are smoking, low vegetables/fruit diet, and alcohol consumption[9], while to EAC, the risk factors are obesity, gastroesophageal reflux disease[9,10]. When alcohol and tobacco are used together, there is an increased risk. This combination is believed to be responsible for 70%-90% of cases, mainly because they cause chronic irritation and inflammation of the esophageal mucosa. In the case of obesity, the greater the abdominal circumference, the greater the intra-abdominal pressure increases the probability of developing gastroesophageal reflux[4,6,11–14].

Early diagnosis is fundamental to improve prognosis. However, dysplasia usually is asymptomatic[4,11,12,15]. Besides, it manifests itself at an average age of 67 years, when there is a high incidence of metastasis, mainly in lymph nodes, liver, lungs, and bones[11,12]. These features make the EC an aggressive malignancy with a 15%-23% five-year survival rate[9,10].

Gastric cancer

Gastric cancer (GC) is the fourth in incidence and mortality word wild[1,2]. The primary risk factors for gastric cancer are genetics, diet (high amount of salt and low consumption of fruits and vegetables), *Helicobacter pylori* or *Epstein–Barr* virus infection, smoking, alcohol intake, and sedentary life[16–19]. The principal risk factor for gastric cancer is *H. pylori* infection, accounting for 80% of the cases. Although the decreasing incidence of *H. pylori* infection, gastric cancer deaths are still high. While the primary

risk factor is the *H. pylori* infection, many genes were already associated with GC[16,18,20], and some genetic variations that can interact with *H. pylori* increase the gastric cancer risk[21,22]. The incidence of gastric cancer is higher in males (1.32-2.2) and Eastern/Central Asia and Latin America[16,18].

Obesity can induce inflammation of the stomach lining through TNF, IL-6, and CCL2. In contrast, a diet rich in fruits and vegetables has proven to be an ally in cancer prevention because it contains numerous antioxidants that prevent metabolic damage, especially vitamin C[18,23].

A relevant factor in the decline of gastric cancer has been the successful prevention and treatment of infections by *H. pylori*[18]. According to the International Agency for Research on Cancer, this is a carcinogen from group 1, meaning there is sufficient evidence of humans' carcinogenicity[24,25]. *H. pylori* infection affects more than half of the world population, and its eradication may considerably decrease the chances of stomach cancer. However, it would increase the chances of esophageal adenocarcinoma. Though, it is not yet known how this esophageal protection mechanism occurs[18,25,26].

Hepatocellular carcinoma

There are about 1 million new cases of liver cancer each year, with hepatocellular carcinoma (HCC) responsible for the majority of cases (90%) being the second most common cause of cancer death worldwide[27,28].

The HCC presents a poor prognosis due to late diagnosis. Due to the heterogeneity of the tumor, there is difficulty in drawing a treatment line for HCC, and multiple different tumors may occur in a single patient[28,29]. This heterogeneity can be caused by environmental factors and/or genomic and biological changes caused by the tumor lesion[28].

Cirrhosis and non-alcoholic fatty liver disease are risk factors associated with alcohol abuse and obesity that can lead to the onset of HCC. Other genetic factors such

as diabetes, exposure to carcinogens such as aflatoxins, and biological factors, especially hepatitis virus infection, can be highlighted[29].

The HCC development is a multistep process. It starts as a chronic liver disease that leads to inflammation, fibrosis, or aberrant hepatocyte regeneration. This set of conditions can progress to cirrhosis and later malignancy. The causes of this inflammation can be HBV/HCV infection, fatty liver disease, excessive alcohol intake, and aflatoxin consumption[27,30]. The outcome of this inflammation can be influenced by epigenetics and the immunological response in the tumor microenvironment to create a preneoplastic lesion until producing cells with highly proliferative, invasive, and survival skills[27].

The geographic regions most affected by HCC are Southeast Asia and sub-Saharan Africa, where there is a picture of endemic infection by the hepatitis virus and high exposure to aflatoxin, which are responsible for 70%-90% of cases in these places[29]. Currently, there is no line of therapy based on biomarkers suitable for HCC although there are already some candidate genes[31].

Pancreatic cancer

Pancreatic cancer, characterized by pancreatic ductal adenocarcinoma (PDAC), is the seventh leading cause of cancer death worldwide[32]. Its incidence is higher in Europe, followed by North America and Oceania, mainly in people over 70 years old. Incidence and mortality increase with increasing age and are more common in men than women[33].

It is highly fatal because it presents aggressive growth and a lack of symptoms in the disease's initial stage. As the tumor progresses, a picture of nonspecific symptoms begins, including jaundice, weight loss, abdominal pain, and fatigue[33]. About 80% of the diagnosis is made in advanced clinical stages, leading to a 5-year prognosis of survival after surgery[34]. Surgical resection is the single strategy capable of curing pancreatic cancer. Besides, using chemotherapy concomitantly improves survival rates[35].

The main risk factors for the onset of pancreatic adenocarcinoma are smoking, alcohol, obesity, *H. pylori*, and type 2 diabetes[35]. Other factors such as fat infiltration into the pancreas have been associated with the development of intraepithelial neoplasms. Pancreatic cancer can also arise from genetic factors that can cause familial syndromes, such as Peutz-Jegheirs syndrome[32]. History of pancreatic cancer in first-degree relatives leads to a 2 to 3-fold increase in the risk of incidence due to inherited genetic predispositions[36].

Colorectal cancer

Colorectal cancer (CRC) is the second in the number of cases (1.3 million) and the third in the number of deaths (540 thousand) annually[2]. The CRC is responsible for about 10% of cancer-related death worldwide, and in the last 45 years, there was an increase in this mortality rate[37]. Its incidence is higher in developed countries like Australia and New Zealand, followed by countries in Europe, East Asia, and North America, and the frequency increases as individuals age, usually appearing in people over 50 years old[38].

The tumor can originate in both the colon and the rectum, but they usually fuse because they have similar clinical and biological characteristics, with adenocarcinoma as the main cell type of the tumor[38]. Many factors are associated with this increase in the diagnosis/mortality rate, such as an increase in life expectancy, poor dietary habits, and risk factors: smoking, red meat consumption, sedentary lifestyle, obesity, alcohol intake, and genetics[37,39–41]. These factors change the genetic/molecular in colon epithelial cells deactivating suppressor tumor genes and activating oncogenes to create aggressive and malignant behavior[41].

In the early stages, the disease has no clinical manifestation, and the patient may be asymptomatic for years but as the disease progresses and progresses to a more severe condition, symptoms such as changes in intestinal motility, hidden or evident colorectal bleeding, cramps, loss of weight, weakness, and fatigue are manifesting[38].

GENOME DATA IN GASTROINTESTINAL CANCERS

Esophageal cancer

There are several generalized genomic changes when esophageal carcinoma cells are analyzed. The most evident is a somatic mutation in TP53 that appears in about 83% of cells. The p53 protein is a tumor suppressor and one of the most important transcription factors for regulating proliferation, apoptosis, autophagy, and cell cycle. However, this gene has a high percentage of mutation in cancer cases, reaching 75% in tumor cells[12,42].

There are also changes in genes that control cell cycle and differentiation, including CDKN2A, NFE2L2, CHEK1/2, NOTCH1/3. Others may appear overexpressed, such as CCND1, CDK4/6[12,43–46]. The BTG3 protein could regulate the cell cycle's progression; its low expression is related to esophageal adenocarcinoma's appearance, and its level of expression is directly correlated with lymph node metastasis[12,47].

The presence of mutations in the growth factors in cancer cells is well documented in the literature. Overexpression of EGRF in carcinoma cells is associated with lymph node metastasis, and its level of expression also influences the patient's clinical stage. Another growth factor correlated with esophageal carcinoma is VEGF-C encoded by the FLT1 gene, and its levels in the tissues are correlated with tumor stages and the state of metastasis[12,42,48].

Using next-generation sequencing, frequent mutations in carcinoma cells have been observed in the KMT2D, SETD2, NOTCH1, RB1, CDKN2A, BAP1, FOXO3, and MSH6 genes compared to adenocarcinoma. It was also observed that some CNVs in FGF3, FGF4, FGF19, and CCND1 were more expressed in carcinoma when compared to adenocarcinoma[49].

Gastric cancer genome

Besides the infectious pathogens causes, the genetic data help to classify the GC in three more subtypes: microsatellite instability (21.7%), genomically stable (19.6%), and chromosomally unstable tumors (49.1%)[50].

Although the infectious pathogens are an environmental cause of GC, the GC caused by the infectious pathogens is associated with genetic modifications like PIK3CA mutations or gene amplification of JAK, PD-L1/2, or ERBB2. The infectious pathogen can also induce epigenetic modifications in this type of CG as DNA methylation in the PTEN gene promoter[51] and tumor-suppressor gene adenomatous polyposis coli (APC)[52]. The microsatellite instability is more associated with a high number of truncating or missense mutations. The genes with the highest number of mutations in microsatellite instability GC are EGFR, ERBB3, KRAS/NRAS, and PIK3CA[53].

Genomically stable tumors present a high number of mutations, especially genes well associate with cancer. The gene RHOA works as signal transduction inducing cell proliferation, actin cytoskeleton structure, cell movement associated with metastasis[54,55]. The genes CLDN18 and ARHGAP are frequently translocated in genomically stable GC tumors. The gene CDH1 encodes a cell-cell adhesion protein, and it is also currently mutated in this type of cancer[56]. Furthermore, CDH1 has a role in cell proliferation, invasive behavior, and migration[57–59]. In the CDH1 gene, autosomal dominant mutations increase stomach cancer risk, especially when one of its copies is lost, generating a scenario of diffuse hereditary gastric cancer[18].

The chromosomal alterations involve gene amplification of EGFR, ERBB2/3, KRAS/NRAS, and RASA1; gene deletion of PTEN. These genetic modifications probably would result in gene activation or deactivation, which together would result in tumor cell phenotypes. The genes EGFR, ERBB2/3, JAK2, FGFR2, MET, KRAS/NRAS, and PIK3CA, are predicted to be active, while RASA1, PTEN, PIK3R1 would be inactive.

Hepatocellular Carcinoma

There are numerous genetic changes in HCC cells, including mutations, changes in the number of copies, and chromosomal rearrangements, leading to an overly complex genomic picture. Its complexity is further aggravated when etiological factors that precede the tumor development for years are considered[60].

Some genes stand out for playing a fundamental role in cancer development, which is why they appear more frequently as TP53, MYC, WNT, CTNNB1. Also highlighted are genes related to the cell cycle, such as CCND1 and CDKN2A[60].

A study integrating RNA-seq, DNA sequencing, TCRseq, and SNP array was carried out to investigate the space-time interactions between cancer and immune cells. A difference in the interaction of the adaptive immune system was detected in different regions of the same tumor. The TP53 and CTNNB1 genes expressed clonal mutations. High-level amplifications have been reported for CCND1, FGF19, and VEGFA. Mutations related to environmental risk factors such as smoking and alcohol were found in TERT, CTNNB1, TP53, AXIN1, and ARID1A. There were also mutations without an apparent etiological factor in TERT, KMTB2, CCNA2, and CCNE1[61].

HCC is the result of a multistep process involving genetic, epigenetic, and transcriptomic interactions. Among these interactions, epigenetics is among the most affected pathways, leading to profound changes in gene expression that can facilitate the formation of the tumor. Hypermethylation of DNA is the most common form of epigenetic silencing of tumor suppressor genes and generally occurs in CpG islands of gene-promoting regions such as DLC1, TFPI-2, CDKN2A, and PTEN[31].

Pancreatic cancers

The etiology of PDAC is mainly related to genetic predisposition, environmental factors such as smoking, obesity, and poor nutritional diet that can lead to chromosomal instability affecting cell cycle pathways, chromatin remodeling, WNT, MYC, and NOTCH signaling, and DNA damage repairs[36,62]. Among the mutated genes, the one that appears most frequently is KRAS[63]. It is also possible to highlight mutations in

MLH1, MSH2, PMS2, and MSH6 responsible for Lynch Syndrome and mutations in the germ-lines of PALB26,11,12 and ATM7,12,13[36].

The pancreatic cancer genome analyses showed a homogenous profile with somatic mutations in few genes shared KRAS, TP53, CDKN2A, and SMAD4. However, other less frequent genes are also involved, like MAP2K4, KDM6A, RNF43, ARID1A, TGFβR2, GNAS, RREB1, and PBRM1[64–66]. These mutations can vary, and it is observed non-silent mutations, gene amplification (> 8 copies, deletions, and structural variants)[66]. The set of genes that appear often mutated in pancreatic cancer play a role in oncogenes, DNA damage repair, and chromatin modification[64,67]. The pancreatic cancer genome has chromosomal rearrangements classified into four subtypes: stable, locally rearranged, scattered, and unstable[64]. The mutation event more frequent is non-silent single nucleotide variants and copy number change (loss)[64]. The pancreatic cancer stable subtype was found in 20% of samples and had very few structural rearrangements (< 50 structural rearrangements) and more chromosomal mutations (aneuploidy). The locally rearranged subtype was found in 30% of samples with a high number of structural rearrangements (> 200) in few chromosomes (three or fewer chromosomes), and there is more gene amplification. The scattered subtype is the most frequent (36% of samples) and has 50-200 structural rearrangements. Besides that, the mutation type gene amplification is more frequent than in the other subtypes. The unstable subtype is less frequent (14% of the samples) has the highest number of structural rearrangements (> 200 structural rearrangements) such as intrachromosomal, translocations, inversion, deletions, and duplication. Besides the frequent mutation described in pancreatic cancer, the unstable subtype is also associated with BRCA pathway mutations (BRCA1, BRCA2, and PALB2)[64].

Colorectal cancer

The majority of CRC cases are sporadic (70%), and only 30% are inherited/family[39] with mutations in the DNA mismatch-repair genes, APC gene, or mutY DNA

glycosylase (MUTYH) gene[40,41]. The DNA mismatch-repair proteins malfunctioning creates the condition of genetic mutation accumulation and tumor cells rising.

Like GC, CRC is also classified into three genetic subtypes based on their genomic alterations. The CRC has the chromosome instability, microsatellite instability, and hypermethylation of CpG islands (CpG island methylator phenotype - CIMP)[68,69]. The chromosomal instability is the most frequent in CRC, present in 71%-85%[68,69]. The genetic differences also lead to overall survival differences in CRC. The CIMP subtype is associate with poor prognosis, followed by chromosome instability, and microsatellite instability showed the best survival[69–71]. The poor prognosis in the CIMP subtype indicates the importance of CpG methylation dysregulation in the CRC tumorigenesis. The methylation dysregulation might affect the proto-oncogenes and tumor-suppressor genes. The worst prognosis in the CIMP subtype indicates that a different approach is necessary to deal with molecular modifications. The epigenetic modifications can also be therapeutic targets to improve the treatment.

The genetic/genomic diversity in gastrointestinal cancers shows the importance of molecular characterization to improve the treatment and prognosis.

PATHWAYS

The cellular pathways show the main activities and functions present in a cell when proteins work together. The cancer pathways are responsible for the cell's behavior that allows unlimited cell replication, survival, and tissue invasion. The pathways also are responsible for the molecular changes driving tumorigenesis. Understanding how a set of proteins work together to develop a cancer cell might point to the target proteins to block these processes.

The pathways most present among the gastrointestinal cancers discussed here are apoptosis, focal adhesion, PAK pathway, PI3K-Akt signaling pathway, TGF-β pathway, and Toll-like receptor signaling pathway (Table 2).

Apoptosis plays a role in maintaining the balance in cell division and death during development and life. The unbalance of apoptosis leads to survival and

uncontrolled division in tumorigenesis[72]. The apoptosis pathway is triggered by irreparable DNA damage, and it has many proteins that can fail and be blocked to inhibit cell death. The intrinsic process is mediated by mitochondria releasing cytochrome C after BH3 proteins activate BAX and BAK. The cytochrome C together with Apaf-1 and caspase-9 creates the apoptosome to continue the apoptosis process. The extrinsic process has death receptor ligands (CD95L, TRAIL, and TNFα), death receptors, associated proteins (FADD and TRADD) that transduce the death signal until caspases-8. Both intrinsic and extrinsic processes act on caspase-3/6/7 to induce the apoptosis cascade. The cell death by apoptosis results in a non-inflammatory process, and this attracts the research to the development of therapies that use apoptosis to treat cancer[73–75].

The PAK1 signaling pathway (p21-activated kinase) has six members divided into two groups and induces proliferation, survival, and motility[76]. PAK1 participates in cancer tumorigenesis after being highly expressed. The crosstalk of PAK1 with MAPK/ERK and PI3K-AKT pathway induces proliferation and survival, respectively[77]. The PAK1 also connects with the Wnt signaling pathway through CTNNB1 (β-catenin) and continues to stimulate growth and metastasis[76]. The PAK1 expression protects the cell from apoptosis after interaction with Raf that inactivates Bcl-2 family member (Bad) in mitochondria[76,78].

Toll-like receptors (TLRs) are part of the family of pattern knowledge receptors (PRRs) and operate on innate immunity, participating in the body's first line of defense against invasion of microbial pathogens, tissue damage, and cancer. Its signaling pathway controls the activation, maturation, and immune functions of immune cells, especially the secretion of cytokines, influencing the tumor's metabolism, proliferation, and spread[79]. They are expressed by several immune cells such as macrophages, dendritic cells, B lymphocytes, natural killer cells, and non-immune cells such as epithelial cells, fibroblasts tumor cells[80]. When expressed in the tumor microenvironment, TLRs can release cytokines and chemokines into the tumor

environment to recruit other immune cells to release more proinflammatory cytokines, pro-angiogenic factors, and growth factors[79].

The TGF-β signaling pathways are pleiotropic, regulating multiple functions, such as cell growth, differentiation, apoptosis, angiogenesis, motility, invasion, and immune response. Modifications in this pathway might play an essential role in the development of tumors and metastasis. These modifications can affect not only the tumor cells but also the microenvironment. At this level, the TGF-β generates a microenvironment conducive to tumor growth and metastasis at all carcinogenesis stages. TGF-β has a contradictory behavior at the cellular level, acting as a suppressor and a tumor promoter[81,82]. Initially, the TGF-β pathway promotes cell cycle arrest and apoptosis. In advanced stages, it promotes cancer cells' motility, invasion, tumor progression, and metastasis. Thus, the accumulation of mutations is responsible for guiding the evolution from a suppressor pathway to a tumor promoter[83].

The HCC RNASeq study identified four subtypes of HCC using 212 samples. The pathway analyses using the expression data reveal the enriched pathways metabolism RNA processes such as RNA processing, binding, splicing. Although all the samples are from HCC, this result indicates different gene expressions, cell activity, and behaviors. These enriched processes are not all shared by the four HCC groups funded. However, translation, ribosome, metabolism of proteins, and cytoplasm ribosomal proteins were shared with at least three groups[84]. The microarray analysis using 25 HCC samples identified thousands of genes differentially expressed, and the pathways cell cycle response, DNA damage response, cell survival, and apoptosis were identified. Besides that, it was also linked to pathway terms and the poor prognosis clinical parameters. In part, these results also agree with the RNASeq study point transcriptional regulation, RNA processing, and cell cycle regulation. The single-cell RNASeq analysis indicates 119 genes associated with HCC. The pathways analyses using GO showed an acute inflammatory response, oxidative stress, and humoral response. Simultaneously, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways indicate IL17 and TNF signaling pathways, infectious disease, and

rheumatoid arthritis[85]. These samples present more immunological functions activated. According to the OncoVar database, the KEGG pathways associated with HCC are mainly cancer pathways, viral infection, cell longevity (growth and death), antineoplastic drug resistance, and transduction signaling pathways (Wnt and Hippo signaling pathways)[86]. The molecular pathways in HCC are not completed understood, and these results showed a notable variation of response in the differentially expressed genes working together to express a function.

Analysis combining colorectal and endometrial cancer microarrays samples identified 139 genes upregulated in both studies. These genes showed enriched the GO terms cell proliferation, Wnt signaling pathway, fatty acid β-oxidation, transcription, exocytosis, dopaminergic neuron differentiation, and platelet degranulation. The KEGG pathways enriched were tight junctions, rheumatoid arthritis, renal cell carcinoma, and cancer pathways signaling. The rheumatoid arthritis pathway was enriched in more than one study with the genes (ATP6V0D1, ATP6V1D, CD28, CTLA4, CTSK, FOS, IL18, and JUN)[87]. Other microarray metanalysis study using CRC samples point to also the KEGG pathways related to cell cycle, pathways in cancer, and Wnt signaling pathway. These pathways are linked; as a result, they shared proliferation and block apoptosis[68]. These processes are together inducing the normal cell to convert to a tumor cell.

THE PROTEIN-PROTEIN INTERACTION IN CANCER

The number of gastrointestinal cancer projects in different OMIC levels found many genes working in the tumorigenesis. The gastrointestinal cancers discussed here sum 178 different genes with mutations associated. The number of genes with mutations associated with gastrointestinal cancers ranges from 41 to 89 genes in HCC and GC.

Each of these cancers has variation and can be classified into subtypes according to cell origin, chromosomal structural rearrangements, gene expression, and cell behaviors. However, there are 46 genes shared at least by two types of cancers. These genes should be investigated to understand better how they assist in the cell transformations to tumors, biomarkers of tumor cells, and potential drug or therapy

targets. The genes present in all five types of cancers are ACVR2A, APC, ARID1A, and CTNNB1.

We used information from STRINGDB to check the protein-protein interaction (PPI) from these 178 genes. We used the experimental information only to build this PPI network. The PPI investigation allows building a network with 111 genes connected (Figure 1). The number of nodes in the PPI network indicates that these genes work together in gastrointestinal cancers' tumorigenesis.

We analyzed the gastrointestinal cancers network to identifies the most connected protein (high degree) as central nodes in this PPI. The protein CTNNB1, AKT1, TP53, EP300, and TP53BP1 are the central nodes with the highest degree.

The CTNNB1 gene encodes a β-catenin protein located in the adherents junctions[56]. The β-catenin is a cytoplasm protein that works in the adhesion between cells. The β-catenin binds the actin in the cytoskeleton and the E-cadherin protein in the cell membrane, connecting neighbor cells. The E-cadherin between protein create the cell linkage[116]. The β-catenin is also a mediator in the Wnt signaling pathway. The Wnt signaling pathway, when activated, induces the accumulation of β-catenin in the nucleus, and this, activates the transcription of target genes[56]. The WNT protein binds the receptor in the membrane and induces β-catenin to accumulate, promoting cell survival and proliferation[68]. The mutations in CTNNB1 gene are frequently found in HCC (13%)[117,118], CRC (6%)[119], and it is mutated in 4% of GC[50].

The AKT1 is a central protein in cell transduction signaling, which, when induced by PI3K, induces process cell proliferation, survival, and angiogenesis. The activation of the mTOR complex by ATK is investigated as a drug target to treat PDAC[120–122]. The Epstein-Barr virus (EBV) and *H. pylori* induce inflammation and the expression of AKT in GC, resulting in cell proliferation and telomerase activation[123,124]. The investigation of blockage of AKT in GC resulted in suppression of growth and metastasis[125]. The investigation of critical proteins in HCC PPI identified several functions crucial in the tumorigenesis cell proliferation, anti-apoptosis, and metastasis.

The PPI network showed AKT1 as a potential drug target[100]. These results indicate AKT1 central position in tumorigenic and a potential drug target.

The 53BP1 protein (gene TP53BP1) has a role in DNA damage response, cycle arrest, and trigger the expression of P53; the malfunctioning of this protein might lead to the development of genomic instability and molecular diseases. The lack of function of 53BP1 is associated with poor prognosis, angiogenesis, and metastasis[126]. The decreased expression of 53BP1 in CRC induces radiotolerance and chemoresistance. Moreover, the CRC cells with lower expression of 53BP1 are associated with a higher proliferating rate, decreased apoptosis, and poor prognosis[127–129]. The 53BP1 also interacts with P53, as indicated in CRC and EC, when the downregulation of 53BP1 induces the downregulation of P53[127,128,130]. The 53BP1 is expressed as soon as DNA damage treatment occurs in human pancreatic cells[131]. The 53BP1 might also influence tumor outcome in pancreatic cancer, as shown when the variation of 53BP1 expression changes the association of CA19-9, a well-known pancreatic cancer marker, and overall survival[132].

The p300 protein (encoded by EP300 gene) is a histone acetyltransferase and participates in chromatin remodeling and interacts with basal transcriptional machinery to improve DNA-binding, affecting the gene transcription in normal and cancer cells[133]. The EP300 mutations are common in CRC and GC by frameshift in microsatellite regions[134]. The mutation in EP300 is frequent in EC (10%), and it correlates with poor prognosis, associated with cell proliferation, migration, and invasion (metastasis)[135,136]. The role of p300 in remodeling the chromatin makes it appropriate to investigate epigenetic therapies, and the use of natural nutrients as potential prevention and treatment was already discussed to GC[137].

ESSENTIAL GENES AND KAPLAN-MEIER SURVIVAL ANALYSIS

All gastrointestinal cancers discussed here have a low five-year survival rate, but CRC (Table 1). The esophagus, liver, and pancreas have the lowest five-year survival rate. The late diagnosis, metastasis, and aggressive behavior are associated with a low five-

year survival rate. The poor prognosis is described associated with gene expression in many studies[105,127,135,139–142].

The expression levels are crucial information that might work as a prognosis in gastrointestinal cancers. Association between TP53BP1 expression and overall survival analyses in CRC indicated a connection with low expression and low survival in the I-IIA stage, T3-T4, and N0[127]. Again, indicating this protein has an essential role in CRC, not only to a high degree but also as a prognostic marker. The EP300 gene has high expression associated with poor survival in ESCC[135]. The lncRNAs has a critical role in cancer development, and the high expression of ANRIL and HOXA11-AS lncRNA is associated with poor survival in GC[139,140]. The overexpression of lncRNA ANRIL is significantly associated with gastric cancer progression and can serve as an independent predictor of patient survival[143]. The high expression of ANRIL combined with PRC2 significantly silences miR-99a and miR-449a at the transcriptional level, which increases the expression mTOR, CDK6, and E2F1[139]. The HOXA11-AS gene reduces the expression of suppressor tumor genes KLF2 and PRSS8 at the transcriptional level[140]. The KLF2 downregulation is associated with migration, invasion, and poor survival[144,145]. Besides, KLF2 inhibits growth, migration and induces pancreatic cancer cells to senescence.

The ESCC has a poor survival when low ECRG4 expression occurs compared to the high-expression group[146]. The EAC has worse overall survival when IL11 expression increases. Poor survival is also observed in a low expression of NPTX1, ITPR1, and PDGFD[147].

PDAC analyses show that high expression of the CENPF, SCEL, SERPINB5, SLC2A1, SLC6A14, TMC7, and TMRSS4 is associated with a lower probability of survival compared to the same genes in low expression[148].

We investigated the gene expression and overall survival from the central genes present in the PPI network (Figure 1). We used information from Kaplan Meier plotter (https://kmplot.com)[149] to investigate the potential prognosis of the central genes.

Among the five genes investigated, three have gene expression associated with survival (AKT1, TP53, and CTNNB1) (Figure 2).

The high expression of AKT1 and TP53 in GC is associated with a poor prognosis. Whereas CTNNB1 Low expressed is correlated with reduced survival in GC. The expression values and survival curves for TP53 (mRNA) in the Kaplan Meier plotter agree with tumor protein p53 expression in GC[150,151]. The TP53 expression is low and has a short half-life in normal cells, while in tumor cells, this gene has high expression and long half-file[152]. The higher expression of TP53 is indicative of the worst prognosis. From reference[153] we know that they did not find AKT1 expression and overall survival association. However, they found that EGFR and AKT1 expression are mutually exclusive associated with poor survival. This result might be due to the two proteins acting in the same pathway. The AKT1 phosphorylated and CTNNB1 high expression are associated with poor survival[154,155].

In liver cancer, there is no significant difference between AKT1 or CTNNB1 high and low expression groups. In the TP53 gene, the differences in expression are not significant in the initial stage of carcinoma. However, this high expression predicts a poor prognosis and a higher mortality rate than a low expression. The results are not according to the TP53 gene expression for HCC, where TP53 high expression is present in poor prognosis groups [156].

However, the prognosis markers based on expression have limitations, and the result must be taken together with other markers.

CONCLUSION

The OMIC information about gastrointestinal cancer is very complex, and each organ/region has subtypes and particularities. We presented here information about and bring to light the genomic changes most common among these cancers. The pathways shared by these molecular diseases also point to the standard functions and the crosstalk of these pathways and the PAK1 pathway centrality, connecting to MAPK/ERK, PI3K-AKT, apoptosis, and Wnt signaling pathway. The PPI network

pointed to five central genes, and the literature corroborates the crucial role in gastrointestinal cancer with expression and poor prognosis association. This information might help in target choice to drug and therapy research.

65760_Auto_Edited - 副本.docx

ORIGINALITY REPORT

2%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.oncotarget.com Internet	40 words — 1%
2	Foyzur Rahman, Prince Mahmud, Rezaul Karim, Tofazzal Hossain, Farhadul Islam. "Determination of novel biomarkers and pathways shared by colorectal cancer and endometrial cancer via comprehensive bioinformatics analysis", Informatics in Medicine Unlocked, 2020 Crossref	18 words — < 1%
3	mcf.bidmc.harvard.edu Internet	14 words — < 1%
4	cancerres.aacrjournals.org Internet	13 words — < 1%
5	cancerpreventionresearch.aacrjournals.org Internet	12 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON
EXCLUDE MATCHES < 12 WORDS