Informatics in Medicine Unlocked 21 (2020) 100492

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Comorbidity network analysis and genetics of colorectal cancer

Branimir Ljubic [a], Martin Pavlovski [a], Jumanah Alshehri [a], Shoumik Roychoudhury [a], Vladimir Bajic [b], Christophe Van Neste [b,c], Zoran Obradovic [a,*]

[a] Temple University, Center for Data Analytics and Biomedical Informatics (DABI), Philadelphia, PA, USA
[b] King Abdullah University of Science and Technology, Computational Bioscience Research Center, Thuwal, Saudi Arabia
[c] Center for Medical Genetics (CMGG), Ghent University, Ghent, Belgium

ARTICLE INFO

Keywords:
Colorectal cancer
Comorbidities
Genes
Social networks
Biomedical informatics

ABSTRACT

Background: Colorectal cancer (CRC) is the third most common cancer in the United States and the second leading cause of cancer death. The goal was to identify comorbidities and genes associated with CRC.
Methods: A novel social network model was developed on the Healthcare Cost and Utilization Project (HCUP) - State Inpatient Databases (SID) California database to study comorbidities of CRC. Ranked lists of comorbidities and comorbidity networks were created, and the prevalence of comorbidities in different stages of CRC was calculated. Ranked lists of comorbidities were utilized for text mining of PubMed and DisGeNET to extract genes associated with CRC.
Results: 5,786 comorbidities were identified in females and 5,607 in males in early stages and 5,609 comorbidities in females and 5,427 in males in advanced stages of CRC. Associations between 1,937 different genes and CRC were extracted from PubMed. 150 genes are associated with CRC in DisGeNET. The most mentioned genes associated with CRC were: TP53 (241 abstracts in PubMed), APC (115), and KRAS (106). These 3 genes as well as MLH1 (98) and TGFBR2 (18) had DisGeNET scores of 0.5. PPARG gene (43) had DisGeNET score of 0.6.
Conclusions: The results of comorbidity network analyses suggest which comorbidities of CRC are highly expected. Discovered genes could be used to recruit more individuals who would benefit from genetic consultations. Identified associations between comorbidities, CRC, and shared genes can have important implications on early discovery, and prognosis of CRC. Prevention and treatment of discovered comorbidities would potentially lead to improved quality of life and better outcomes of CRC.

1. Introduction

There are more than 140,000 new cases and 51,000 deaths from colorectal cancer (CRC) in the U.S. each year [1,2]. CRC is the third most common cancer in the U.S. and the second leading cause of cancer death [2]. It affects men and women almost equally. Incidence, mortality, and survival rates for CRC have regional variations and change over time [3]. CRC is mainly a disease of developed countries where it accounts for over 63% of all cases. It ranges from more than 40 per 100,000 people in the U.S., Australia, and Western Europe to less than 5 per 100,000 in Africa and parts of Asia [3]. The incidence of CRC is gradually decreasing in the U.S., mostly due to cancer screening and early detection of precancerous polyps. CRC survival is highly dependent on stage of cancer at diagnosis, and ranges from a 90% 5-year survival rate for early stages, to 10% for cancer with metastases [3]. The treatment of CRC has improved considerably in recent years. Better therapies have resulted in prolonged survival for patients with CRC [4]. Patients with CRC usually present in the older age group, with multiple comorbidities. About 59% of patients with CRC had one comorbidity, and about 19% of patients had 4 or more comorbidity conditions [5]. Comorbidities are strong prognostic factors of survival in CRC patients in addition to sociodemographic and cancer characteristics. Early identification and management of comorbidities could help to optimize care for CRC patients [6].

Comorbidity network analyses can help to understand illness progression [7–9]. A social network analysis method is proposed to represent the progression of cardiovascular diseases (CVD) in patients with Diabetes Mellitus Type 2 (T2DM) [10]. A social network was developed to help understand which comorbidities have a higher influence on T2DM progression [8]. Many genetic networks are readily available [9]. Hidalgo and Barabasi described the use of networks to integrate different genetic, proteomic, and metabolic datasets as a viable path

* Corresponding author. Temple University, SERC 035-02, Center for Data Analytics and Biomedical Informatics, 1925 N 12th St, Philadelphia, PA, 19122, USA.
E-mail address: Zoran.Obradovic@temple.edu (Z. Obradovic).

https://doi.org/10.1016/j.imu.2020.100492
Received 7 August 2020; Received in revised form 16 November 2020; Accepted 25 November 2020
Available online 1 December 2020

toward elucidating the origins of specific diseases [7]. Numerous factors suggest a genetic contribution to CRC such as a family history of CRC or polyps [11]. A significant expansion of the genetic understanding of colonic carcinogenesis in the last 30 years occurred [12–14]. The American College of Medical Genetics and Genomics has published guidelines for evaluating patients with CRC, to identify individuals whose clinical findings require referral for genetics consultations [15].

The objective of our research was to discover the most common comorbidities in different stages of CRC on the HCUP SID California database using network science. We used discovered comorbidities to identify genes associated with CRC, and comorbidity diseases. The text mining tool BeFree was employed to extract relationships between CRC and genes from PubMed [16]. BeFree consists of the Biomedical Named Entity Recognition (BioNER) module based on dictionaries using fuzzy and pattern matching methods to find and uniquely identify entity mentioned in the literature, and a module for Relation Extraction (RE) based on Support Vector Machine (SVM) [16].

The next objective of our study was to analyze associations between CRC and comorbidities to genes in abstracts indexed in the PubMed database. Identification of relationships between comorbidity diseases and shared genes can have important implications on early discovery and outcomes of cancer. We view genes and comorbidities as interconnected risk factors for CRC. Findings from the PubMed text mining were compared with results from expert curated sources. We used DisGeNET as the expert source to validate the findings of the text mining of PubMed. This is one of the largest collections of genes involved in human diseases. DisGeNET integrates data from expert curated repositories, the GWAS catalog, animal models, and scientific publications [17].

2. Material and METHODS

We performed analyses of comorbidities associated with CRC on the HCUP, SID California inpatient database (ICD-9/10 codes format) which includes the largest collection of longitudinal hospital discharge data in the U.S [18]. We analyzed data for a period of 9 years (2003–2011). Diseases were represented with ICD-9 codes in this period. We used ICD-9 codes 153 and 154 to designate presence of CRC. SQL queries and Python code were created to extract comorbidities of CRC. We share our Python code on this link: https://github.com/martinpavlovski/cancer-comorbidity-analysis. We completed analyses separately for patients older than age 50, and patients younger than age 50. In the group of patients older than 50, the comorbidities associated with CRC were downloaded and divided into 2 groups. We included all patients who had CRC without diagnosed metastases into the early stages of cancer and patients with already diagnosed metastases into advanced stages. We used ICD-9 codes 196, 197, and 198 to designate presence of metastases. The HCUP database contains only diagnoses of CRC without specifications of TNM stages, and it contains the information about the presence of metastases, which limited our comorbidity analysis to two groups: without and with diagnosed metastases. We could not analyze CRC and comorbidities by TNM staging system, because that information is not contained in the HCUP database. Regarding the group of patients age 50 and younger, the number of patients identified in the HCUP database, for the studied period, was relatively small and we couldn't perform separate analyses for patients with and without metastases. We calculated prevalence of comorbidities separately for males and females for this age group.

In the group of patients older than age 50, data were analyzed separately for the early and late stages of CRC and separately for males and females. Our comprehensive approach comprises the creation of ranked lists of comorbidities using frequencies of their occurrence, analysis of prevalence of comorbidities, construction of comorbidity networks, and calculation of centrality measures to estimate the significance of comorbidities. We used the following formula to calculate the prevalence of development of comorbidities of CRC:

$$P_i = \frac{n_i}{T} \tag{1}$$

P_i is the prevalence, n_i is the number of patients with a comorbidity, and T is the total number of patients with CRC in the HCUP dataset.

In order to construct comorbidity networks, we had to determine the strength of comorbidity relations among diseases. Two types of comorbidity networks were constructed, one based on φ-correlation and the other based on Relative Risk (RR). These two comorbidity measures were used to quantify the strength of relations between the two comorbidity diseases [7]. The strength of relations was calculated between each pair of top comorbidities. CRC was not included in these calculations. The RR of observing a pair of diseases i and j affecting the same patient is given by the following formula:

$$RR_{ij} = \frac{C_{ij}N}{P_iP_j} \tag{2}$$

where C_{ij} is the number of patients affected by both diseases, N is the total number of patients in the population, and P_i and P_j are the prevalence of diseases i and j.

φ-correlation is computed as:

$$\varphi_{ij} = \frac{C_{ij}N - P_iP_j}{\sqrt{P_iP_j(N-P_i)(N-P_j)}} \tag{3}$$

RR overestimates relationships involving rare diseases and underestimates highly prevalent comorbidities. φ-correlation overestimates comorbidities between diseases of similar prevalence but underestimates the comorbidity between rare diseases [7]. We constructed two types of networks, separately for each measure. In both models, the top 100 comorbidities from the rank lists were used to construct adjacency matrix $A[ij]$, which encodes whether and how a pair of nodes is connected. 86% comorbidity conditions from the ranked list in females and 87% comorbidities in males in early stages, as well as 84% comorbidity conditions in females, and 85% comorbidities in males in advanced stages of CRC were present in only 1 patient. It is difficult to consider any disease as a comorbidity of CRC if that disease appears in only 1 patient out of 30,000 patients. We opted to involve comorbidities that appear in about 5% of patients diagnosed with CRC in our network analysis, which turned out to be approximately 100 patients in each of the experimental settings. We created a power-law type of network. A Kolmogorov–Smirnov statistical test, at the significance level of 0.05, was conducted to determine if the networks follow power-laws. We constructed a signed correlation network from the adjacency matrix and performed power transformation (normalization) of correlations [19]. For normalization, we used a signed network normalization $|(\text{corMatrix}+1)/2|^{\beta}$ for a 100 × 100 adjacency matrix. By raising the absolute value of the correlation to a power $\beta \geq 1$ (soft threshold), signed correlation networks emphasize higher correlations versus lower correlations. We tested β-values between 1 and 10 and chose $\beta = 2$ for φ–correlation and $\beta = 5$ for RR. These β-values were selected because they provide the best visualization of networks. We used the correlation matrix after the transformation (normalization) as adjacent matrices to plot the network. Comorbidity networks were constructed in R program, using WGCNA, Statnet, and ggplot2 packages. Centrality measures (degrees of nodes and betweenness) were calculated to analyze the significance of each of the comorbidity. The significance of a comorbidity disease in the network is characterized by its node degree or its betweenness. The node degree represents the number of links the node has to other nodes in the network. The betweenness is the number of the shortest paths that pass through the node. We used the following formula to calculate the betweenness [20].

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{4}$$

Where $g(v)$ is the betweenness of node v, σ_{st} is the total number of

shortest paths from node *s* to node *t*, and $\sigma_{st}(v)$ is the number of those paths that pass through *v*.

Furthermore, we leveraged lists of ranked comorbidities of CRC for text mining of PubMed, to identify genes associated with CRC and comorbidities. The BeFree text mining system was utilized to extract associations between CRC, ranked comorbidity diseases, and genes by using morpho-syntactic information of the text. We also extracted relationships between genes and diseases from expert curated sources (DisGeNET).

We compared a proportion of gene-comorbidity disease associations discovered in PubMed with data collected in DisGeNET. Genes most often described in association with CRC were further matched with comorbidities from the ranking list. The number of abstracts in PubMed where these associations were described was counted. Lists of numbers of PubMed abstracts that show associations of the six most common genes and comorbidities of CRC were created to show the significance of these associations.

3. RESULTS

3.1. Comorbidity analysis

In the group of patients older than age 50, the total number of females in the HCUP SID California database (2003–2011) with early stages of CRC was 31,503, and males 30,870. The total number of patients with the advanced stage of CRC was: females – 26,906 and males – 27,454. We conducted analyses of comorbidities in patients with early stages of CRC and identified 5,786 different comorbidity conditions in females and 5,607 in males. In patients with the advanced stages of CRC, we discovered 5,609 different comorbidity conditions in females and 5,427 in males. Some of the top comorbidity conditions are specific to patients with CRC, but most conditions are similar to those in the general population.

Table 1 shows ranked comorbidities in females, older than age 50, in early stages (Table 1a) and in advanced stages (Table 1b) of CRC. The table also shows the prevalence of particular comorbidities according to the HCUP data. The top-ranked comorbidity in both stages is Essential hypertension, with very similar prevalence in both stages. Hyperlipidemia, history of malignant disease, anemia diabetes mellitus type 2, coronary artery sclerosis, paralytic ileus, malnutrition, acute kidney diseases, encounter of palliative care, the absence of parts of large or small intestines have a few percentage points higher prevalence in advanced stages than in early stages of cancer, which can be attributed to aging.

In Table 2 we present ranked comorbidities of CRC in males, in the early stages (Table 2a) and in the advanced stages (Table 2b). The top comorbidity in both stages of cancer is Essential hypertension, which is the same finding as in female patients. The prevalence of this disease is similar in both stages in females and males and varies between 0.58 and 0.63. Anemia, smoking history, hyperlipidemia, coronary atherosclerosis, congestive heart failure, atrial fibrillation and other cardiac dysrhythmias, esophageal reflux, history of malignant neoplasms of prostate, lungs, colon, and other cancers, acute kidney failure, dehydration, colostomy status, pneumonia, and acute respiratory failure happened more often in advanced stages of CRC. We can notice that majority of diseases have a few points higher prevalence in advanced stages than in early stages of CRC, which can be explained by aging and the progress of CRC.

In Table 3, we present ranked comorbidities of CRC and their prevalence for patients age 50 and younger, separately for males (3a) and females (3b). In this age group, we identified 3794 male patients with CRC and 3640 female patients. HCUP contains very small numbers of patients without metastases in this age group, in the studied period, so we analyzed all patients together, regardless of metastases. Males in this age group had a higher prevalence of essential hypertension, tobacco use disorder, hyperlipidemia, diabetes mellitus, and sepsis than females. On the other side, female patients had a higher prevalence of anemia, hypopotassemia, urinary tract infections, obesity, and depressive disorder than male patients.

Table 1
Comorbidities in early and advanced stages of CRC in females, older than age 50. 1a: Ranked comorbidities occurred in the early stages. 1b: Ranked comorbidities occurred in the advanced stages of CRC. The prevalence of the occurrence of comorbidities is shown.

	1a		1b	
	Females – Early stages		Females – Advanced stages	
	Diagnosis	Prevalence	Diagnosis	Prevalence
1	Essential hypertension	0.63	Essential hypertension	0.60
2	Anemia	0.33	Hyperlipidemia	0.41
3	Hyperlipidemia	0.32	Personal history of tobacco use	0.38
4	History of malignant Neo of large bowel	0.30	History of malignant Neo of large bowel	0.33
5	Urinary tract infection	0.26	Anemia	0.33
6	Hypopotassemia	0.24	Diabetes mellitus type II	0.27
7	Diabetes mellitus type II	0.23	Coronary atherosclerosis	0.26
8	Esophageal reflux	0.22	Paralytic ileus	0.23
9	Personal history of tobacco use	0.20	Atrial fibrillation	0.22
10	Hypothyroidism	0.20	Congestive heart failure	0.22
11	Congestive heart failure	0.19	Esophageal reflux	0.21
12	Paralytic ileus	0.19	Hypothyroidism	0.20
13	Atrial fibrillation	0.18	Acute kidney failure	0.19
14	Coronary atherosclerosis	0.18	Pure hypercholesterolemia	0.18
15	Dehydration	0.17	Digestive system complications	0.18
16	Diverticulosis of colon	0.17	Chronic airway obstruction	0.17
17	Pure hypercholesterolemia	0.16	Hypopotassemia	0.17
18	Osteoporosis without path. fracture	0.16	Dehydration	0.16
19	Osteoarthrosis	0.15	Urinary tract infection	0.16
20	Pneumonia	0.14	Pneumonia	0.15
21	Hyposmolality and/or hyponatremia	0.14	Diverticulosis of colon	0.15
22	Acute posthemorrhagic anemia	0.14	Tobacco use disorder	0.15
23	Iron deficiency anemia	0.14	Hypertensive chronic kidney disease	0.15
24	Chronic airway obstruction	0.14	Benign neoplasm of colon	0.15
25	Acute kidney failure	0.14	Acute posthemorrhagic anemia	0.14
26	Overweight, obesity	0.14	Old myocardial infarction	0.14
27	Iron deficiency anemia due to blood loss	0.14	Overweight, obesity	0.14
28	Depressive disorder	0.13	Hyposmolality and/or hyponatremia	0.14
29	Digestive system complications	0.13	Cardiac dysrhythmias	0.14
30	Benign neoplasm of colon	0.12	Iron deficiency anemia due to blood loss	0.14

3.2. Comorbidity networks

Comorbidity networks were constructed for patients older than age 50, separately for early and advanced stages of CRC, for males and

Table 2
Comorbidities in early and advanced stages of CRC in males, older than age 50. 2a: Ranked comorbidities occurred in the early stages. 2b: Ranked comorbidities occurred in the advanced stages of CRC. The prevalence of the occurrence of comorbidities is shown.

	2a		2b	
	Males – Early stages		Males – Advanced stages	
	Diagnosis	Prevalence	Diagnosis	Prevalence
1	Essential hypertension	0.60	Essential hypertension	0.58
2	Anemia	0.34	History of malignant Neo of large bowel	0.38
3	History of malignant Neo of large bowel	0.31	Anemia	0.37
4	Personal history of tobacco use	0.30	Hyperlipidemia	0.33
5	Hyperlipidemia	0.29	Personal history of tobacco use	0.32
6	Pneumonia	0.26	Diabetes mellitus type II	0.25
7	Diabetes mellitus type II	0.24	Acute kidney failure	0.25
8	Paralytic ileus	0.24	Dehydration	0.25
9	Dehydration	0.20	Chronic airway obstruction	0.24
10	Hypopotassemia	0.19	Coronary atherosclerosis	0.24
11	Hyposmolality and/or hyponatremia	0.19	Tobacco use disorder	0.20
12	Esophageal reflux	0.18	Hyposmolality and/or hyponatremia	0.20
13	Diverticulosis of colon	0.18	Urinary tract infection	0.20
14	Encounter for palliative care	0.18	Atrial fibrillation	0.19
15	Acute kidney failure	0.17	Hypopotassemia	0.18
16	Chronic airway obstruction	0.16	Esophageal reflux	0.18
17	Benign neoplasm of colon	0.16	Congestive heart failure	0.17
18	Hypertrophy (benign) of prostate	0.16	Encounter for palliative care	0.17
19	Anemia in neoplastic disease	0.15	Anemia in neoplastic disease	0.17
20	Congestive heart failure	0.15	Malignant neoplasm of prostate	0.16
21	Constipation	0.14	Unspecified protein-calorie malnutrition	0.16
22	Irradiation, presenting hazards to health	0.14	Malignant neo of bronchus and lung	0.16
23	Tobacco use disorder	0.13	Acute respiratory failure	0.16
24	Unspecified protein-calorie malnutrition	0.13	Pure hypercholesterolemia	0.16
25	Other postoperative infection	0.13	Constipation	0.15
26	Atrial fibrillation	0.12	Irradiation, presenting hazards to health	0.15
27	Acute respiratory failure	0.12	Hypertrophy (benign) of prostate	0.15
28	Pure hypercholesterolemia	0.12	Hearing loss	0.15
29	Coronary atherosclerosis	0.11	Pneumonia	0.15
30	Urinary tract infection	0.11	History of malignant neopl of prostate	0.14

females. Nodes represent diseases (comorbidities) and links connect comorbidities according to distance measures described above (RR, φ-correlation). Calculated node degrees and betweenness show the most connected (significant) comorbidities in each of eight networks. In Fig. 1 we present two networks for the early stages of CRC in females.

Considering φ-correlation, the following comorbidity conditions had the highest node degrees in the network 1a: urinary tract infection (UTI)-65, congestive heart failure (CHF)-63, coronary artery disease (CAD)-52, anemia-50, and acute kidney failure (AKF)-49. Comorbidities that had the highest betweenness in the network 1a were: UTI-329, CHF-275, CAD-254, colostomy status-248, anemia-153, and dehydration-140. Based on RR measures, comorbidities with the highest node degrees in the network 1b were: anemia-61, obstructive chronic bronchitis (OCB)-57, hypoxemia-56, hyperkalemia-55, and peripheral vascular disease (PVD)-53. The following conditions had the highest betweenness in the network 1b: postoperative infection-301, colostomy status-263, gastro-duodenitis-256, chest pain-242, and OCB-221.

Fig. 2 shows two networks (φ-correlations and RR) for the advanced stages of CRC in females.

Comorbidities with the highest node degrees in the network 2a (φ-correlation) were: UTI-57, anemia-52, CHF-47, and hypokalemia-40. The highest betweenness were: UTI-467, anemia-437, dehydration-420,

Table 3
Ranked comorbidities in CRC patients, age 50 and younger: Males (3a), and females (3b). The prevalence of the occurrence of comorbidities is shown.

	3a		3b	
	Males younger than 50		Females younger than 50	
	Diagnosis	Prevalence	Diagnosis	Prevalence
1	Essential hypertension	0.32	Anemia	0.32
2	Anemia	0.28	History of malignant neo of large intestine	0.30
3	History of malignant neo of large intestine	0.27	Essential hypertension	0.27
4	Paralytic ileus	0.22	Hypopotassemia	0.23
5	Tobacco use disorder	0.21	Paralytic ileus	0.21
6	Hypopotassemia	0.17	Urinary tract infection	0.20
7	Dehydration	0.16	Dehydration	0.18
8	Personal history of malignant neo of rectum	0.15	Anemia in neoplastic disease	0.15
9	Hyposmolality and/or hyponatremia	0.15	Tobacco use disorder	0.15
10	Esophageal reflux	0.14	Esophageal reflux	0.14
11	Colostomy status	0.14	Acquired absence of intestine (large, small)	0.14
12	Diabetes mellitus	0.14	Obesity	0.14
13	Anemia in neoplastic disease	0.14	Personal history of malignant neo of rectum	0.14
14	Acute kidney failure	0.14	Depressive disorder	0.14
15	Hyperlipidemia	0.14	Personal history of irradiation	0.13
16	Personal history of irradiation	0.13	Peritoneal adhesions	0.13
17	Acquired absence of intestine (large, small)	0.12	Iron deficiency anemia	0.13
18	Peritoneal adhesions	0.12	Nausea with vomiting	0.13
19	Intestinal obstruction	0.12	Hyposmolality and/or hyponatremia	0.12
20	Urinary tract infection	0.11	Constipation	0.12
21	Protein-calorie malnutrition.	0.11	Intestinal obstruction	0.12
22	Postoperative wound infection	0.11	Diabetes mellitus	0.11
23	Iron deficiency anemia	0.11	Colostomy status	0.11
24	Obesity	0.10	Anxiety	0.11
25	Pneumonia	0.10	Protein-calorie malnutrition.	0.10
26	Hearing loss	0.10	Diarrhea	0.10
27	Constipation	0.10	Hyperlipidemia	0.10
28	Chronic blood loss anemia	0.10	Chronic blood loss anemia.	0.10
29	Diarrhea	0.10	Family history of malignant neo of GI tract	0.10
30	Sepsis	0.10	Asthma	0.10

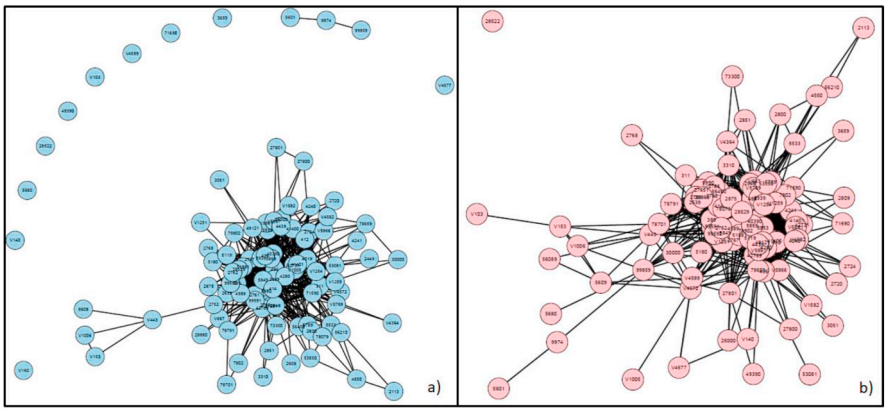

Fig. 1. Networks of comorbidities in the early stages of CRC in females older than age 50. Comorbidity networks, based on: a) φ-correlation and b) RR in females. Nodes represent the top 100 comorbidities (ICD9 codes). Correlations greater than 0.32 ($\beta = 2$) and RR greater than 5.99 ($\beta = 5$) were applied for construction of edges.

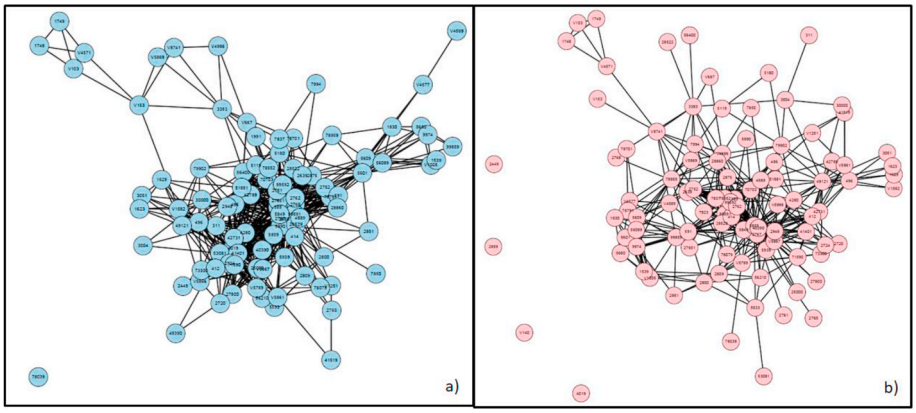

Fig. 2. Networks of comorbidities in the advanced stages of CRC in females older than age 50. Comorbidity networks, based on: a) φ-correlation and b) RR in females. Nodes represent the top 100 comorbidities (ICD9 codes). Correlations greater than 0.30 ($\beta = 2$) and RR greater than 8.99 ($\beta = 5$) were applied for construction of edges.

history of irradiation-405, and paralytic ileus-329. Considering RR measure, the highest node degrees in the network 2b were: chronic kidney disease-56, pressure ulcer lower back-50, hypertensive chronic kidney disease-48, hyperkalemia-42, sepsis-42, and long-term use of insulin-41. The highest betweenness in the network 2b were: chronic kidney disease-494, OCB-421, history of chemotherapy-368, and hypoxemia-307.

Next, we present two networks for the early stages of CRC in males (Fig. 3). We used the same metrics as in females, φ-correlations Fig. 3a and RR – Fig. 3b.

Comorbidities with the highest node degrees in the network 3a (φ-correlation) were: CHF-64, AKF-59, anemia-57, UTI-55, and CAD-50. The highest betweenness in the network 3a were: UTI-492, anemia-431, CHF-394, AKF-350, and hearing loss-277. The highest node degrees in

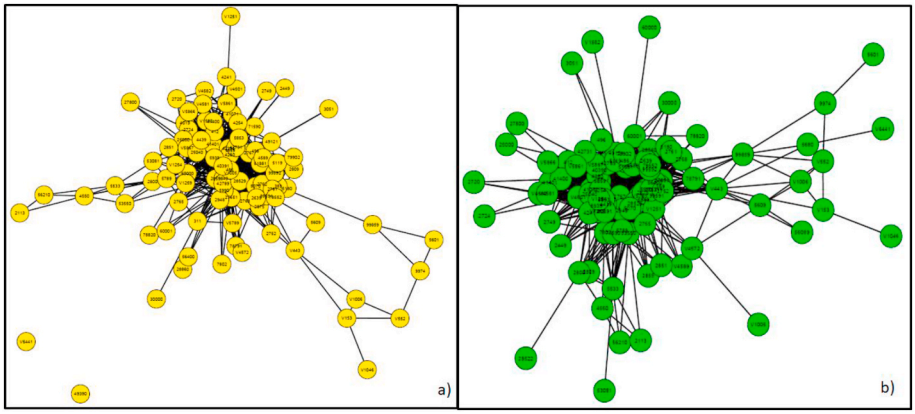

Fig. 3. Networks of comorbidities in the early stages of CRC in males older than age 50. Comorbidity networks, based on: a) φ-correlation and b) RR. Nodes represent the top 100 comorbidities (ICD9 codes). Correlations greater than 0.32 ($\beta = 2$) and RR greater than 8.99 ($\beta = 5$) were applied for edges.

the network 3b (RR) were: anemia-63, hyperkalemia-60, OCB-58, chronic kidney disease-55, primary cardiomyopathies-55, PVD-53, and sepsis-50. The highest betweenness in the network 3b were: colostomy status-418, anemia-294, and postoperative infection-267.

We created two networks for advanced stages of CRC in males (Fig. 4). We used the same metrics as in females, φ-correlation (Fig. 4a), and RR (Fig. 4b).

Conditions with the highest node degrees in the network 4a (φ-correlation) were: anemia-49, CHF-48, AKF-44, and CAD-37. The highest betweenness in the network 4a: anemia-414, dehydration-401, CHF-386, AKF-310, UTI-301, and protein-calorie malnutrition-287. The highest node degrees in the network 4b (RR): Primary cardiomyopathies-42, chronic kidney disease-40, PVD-39, and persistent mental disorders-39. The highest betweenness in the network 4b: Dysphagia-480, primary cardiomyopathies-410, OCB-384, hypoxemia-294, and do not resuscitate status-293.

3.3. Genes associated with CRC and comorbidities

Ranked lists of comorbidities of CRC were used for text mining of PubMed and expert curated sources, specifically DisGeNET, which is a comprehensive platform integrating information on human disease-associated genes. The BeFree text mining system was applied for the extraction of associations between genes, comorbidities and CRC from PubMed.

PubMed indexes more than 170,000 publications on CRC (MeSH Major topic search). The Befree data mining system extracted 1,937 different genes associated with CRC from PubMed. (The list of 1,937 provided as Suppl.1) We found 150 different genes associated with CRC in DisGeNET. Ninety-six genes are present in both sources. Most gene–CRC associations in PubMed were described in only one abstract (1160). The most often mentioned genes associated with CRC were: TP53 (241 abstracts in PubMed), APC (115), and KRAS (106). All 3 genes had DisGeNET scores of 0.5. The DisGeNET score for gene-disease associations takes into account the number and type of sources (level of curation), and the number of publications supporting the association. The scores range from 0 to 1.

Two more genes had a DisGeNET score of 0.5: MLH1 (98 abstracts) and TGFBR2 (18). One gene (PPARG) had a DisGeNET score of 0.6, and it's mentioned 43 times in PubMed abstracts.

In Table 4, we present associations of the APC gene with comorbidities of CRC. The APC gene is located on the long arm of chromosome 5. This gene signals the production of the APC protein, a tumor suppressor, which slows down the growth and division of cells, and controls how cells attach and move [21]. Mutations in the APC gene have been associated with several cancers including familial adenomatous polyposis and CRC. APC gene has also been described in people with a benign desmoid tumor, primary macronodular adrenal hyperplasia, Turcot syndrome, brain neoplasm (medulloblastoma), stomach (gastric) cancers, prostatic cancer, etc. Our text mining of PubMed extracted these tumors, as well as a few types of leukemia associated with the APC gene. BeFree text mining system also extracted a few comorbidity conditions such as Amyloidosis, AD, Systemic Lupus erythematosus, Diabetes Mellitus, and few more conditions as potentially associated with the APC gene.

The TP53 gene is located on the short arm of chromosome 17. The TP53 gene regulates the production of a protein p53, a tumor suppressor, which prevents cells from growing and dividing without control [22]. Mutations in the TP53 gene are the most common genetic changes in human cancer. Mutations in the TP53 gene are associated with the risk of developing breast cancer, bladder cancer, cholangiocarcinoma, squamous cell carcinomas, bone cancer, sarcomas, lung cancer, melanoma, CRC, etc.

Table 4
APC genes associated with CRC and comorbidities. Presented are numbers of abstracts in PubMed where these associations were described in 3 or more abstracts.

No	Diseases	PubMed abstracts
1	Malignant neoplasm of stomach	20
2	Malignant neoplasm of prostate	18
3	Amyloidosis	15
4	Barrett's esophagus	8
5	Malignant neoplasm of pancreas	8
6	AD	6
7	Malignant neoplasm of ovary	6
8	Malignant neoplasm of esophagus	5
9	Malignant neoplasm of bladder	5
10	Brain neoplasm	4
11	Malignant neoplasm of intestinal tract, part unspecified	4
12	Systemic lupus erythematosus	4
13	Ulcerative colitis	4
14	Rheumatoid arthritis	3
15	Burkitt's tumor or lymphoma	3
16	Megakaryocytic leukemia	3
17	Septicemia	3
18	Hypogammaglobulinemia	3
19	Diabetes mellitus	3
20	Sebaceous cyst	3
21	Graft-versus-host disease	3
22	Secondary and unspecified malignant neoplasm of lymph nodes	3
23	Candidiasis of mouth	3

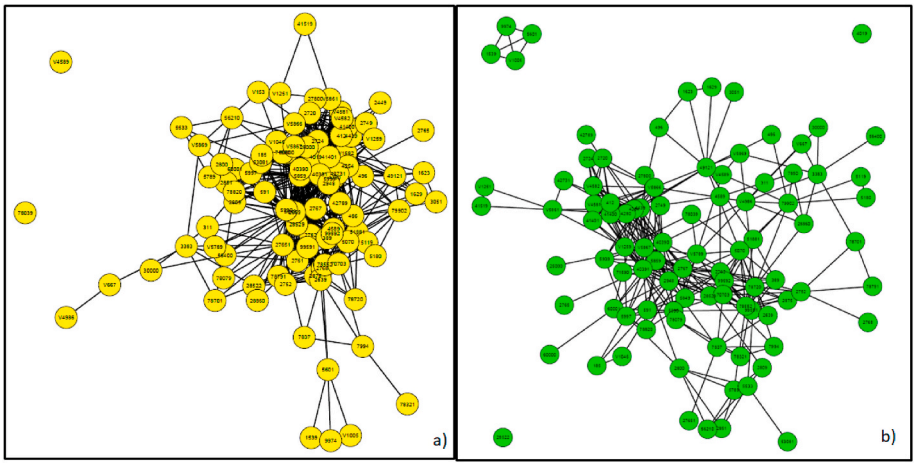

Fig. 4. Networks of comorbidities in the advanced stages of CRC in males older than age 50. Comorbidity networks, based on: a) φ-correlation and b) RR. Nodes represent the top 100 comorbidities (ICD9 codes). Correlations greater than 0.30 (β = 2) and RR greater than 8.99 (β = 5) were used for edges.

Our text mining findings contain tumors mentioned above, as well as few types of leukemia. Several comorbidity diseases are also described in associations with this gene, such as: Hodgkin's disease, ulcerative colitis, rheumatoid arthritis, Alzheimer's disease, etc. Associations between CRC and comorbidities with gene TP53 are shown in Table 5.

The KRAS gene is located on the short arm of chromosome 12. It regulates synthesis of a K-Ras protein [23]. Mutations in the KRAS gene are associated with autoimmune lymphoproliferative syndrome, cholangiocarcinoma, acute myeloid leukemia, epidermal nevus, lung cancer, pancreatic cancer, CRC, as well as conditions such as intellectual disability, distinctive facial features, short stature, macrocephaly, heart defects, skin abnormalities, Noonan syndrome, etc. In Table 6, associations between CRC, KRAS genes, and comorbidities are presented. Text mining of PubMed extracted associations of the KRAS gene with multiple tumors (predominantly pancreatic cancer) and several non-tumorous diseases. The KRAS gene was also extracted in associations with unspecified viral hepatitis C, chronic pancreatitis, ulcerative colitis, dengue, atrial premature beats, endometriosis, etc.

The MLH1 gene is located on the short arm of chromosome 3. The MLH1 gene encodes a protein that has a crucial role in repairing DNA errors [24]. Mutations in the MLH1 gene have been associated with constitutional mismatch repair deficiency syndrome, CRC, brain cancer, leukemia, lymphoma, Lynch syndrome, neurofibromatosis type 1, cancers of the endometrium, ovaries, stomach, etc. In Table 7, associations between CRC, MLH1 genes, and comorbidities found in PubMed abstracts are presented.

The TGFBR2 gene is positioned on the short arm of chromosome 3. It encodes transforming growth factor-beta (TGF-β) receptor type 2, which transduces a signal that prevents cells from growing and dividing too rapidly [25]. Mutations in the TGFBR2 gene have been identified in people with Marfan's syndrome (familial thoracic aortic aneurysm and dissection), with Loeys-Dietz syndrome type II, various cancers (CRC, and others). Our text mining of PubMed discovered these conditions in addition to Alzheimer's disease, asthma, ulcerative colitis, etc. In Table 8 we present associations between CRC, TGFBR2 genes, and comorbidities.

The PPARG gene is located on the short arm of chromosome 3. This gene encodes the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors [26]. PPARs have been involved in the pathology of numerous diseases including obesity, atherosclerosis, prostatic cancer, lipodystrophy, glioma, obesity, insulin resistance, diabetes mellitus, dyslipidemia, hypertension and CRC. Our text mining findings correspond to this description. In Table 9 we show associations between CRC, PPARG genes, and comorbidities.

In Table 10 we present 96 genes identified in both sources: PubMed and DisGeNET.

Table 5
TP53 genes associated with CRC and comorbidities. Presented are numbers of abstracts in PubMed where these associations were described in 20 or more abstracts.

No	Disease	PubMed abstracts
1	Malignant neoplasm of prostate	263
2	Lymphoid leukemia, chronic	248
3	Malignant neoplasm of ovary	239
4	Myeloid leukemia, acute	192
5	Malignant neoplasm of stomach	167
6	Malignant neoplasm of bladder	159
7	Malignant neoplasm of pancreas	134
8	Malignant neoplasm of cervix uteri	130
9	Leukemia of unspecified cell type	112
10	Brain neoplasm	87
11	Secondary and unspecified malignant neoplasm of lymph nodes	82
12	Myelodysplastic syndrome, unspecified	76
13	Malignant neoplasm of esophagus	76
14	Multiple myeloma	68
15	Myeloid leukemia, chronic	62
16	Burkitt's tumor or lymphoma	53
17	Lymphoid leukemia, acute	51
18	Malignant neoplasm of mouth, unspecified	51
19	Hodgkin's disease	50
20	Barrett's esophagus	38
21	Ulcerative colitis	36
22	Malignant neoplasm of liver, secondary	31
23	Mantle cell lymphoma	29
24	Rheumatoid arthritis	29
25	Malignant neoplasm of liver	27
26	Malignant neoplasm of thyroid gland	27
27	Actinic keratosis	25
28	Unspecified viral hepatitis C	23
29	Leukemia of unspecified cell type, acute	22
30	Neurofibromatosis, type 1 [von Recklinghausen's disease]	21
31	Alzheimer's disease	20

Table 6
KRAS genes associated with CRC and comorbidities. Presented are numbers of abstracts in PubMed where these associations were described in 4 or more abstracts.

No	Disease	PubMed abstracts
1	Malignant neoplasm of pancreas	123
2	Malignant neoplasm of liver, secondary	27
3	Malignant neoplasm of stomach	26
4	Unspecified viral hepatitis C	25
5	Malignant neoplasm of ovary	17
6	Secondary malignant neoplasm of lung	14
7	Secondary and unspecified malignant neoplasm of lymph nodes	14
8	Myeloid leukemia, acute	10
9	Malignant neoplasm of prostate	8
10	Chronic pancreatitis	7
11	Patent ductus arteriosus	7
12	Ulcerative colitis	7
13	Dengue	6
14	Atrial premature beats	5
15	Endometriosis	5
16	Leukemia of unspecified cell type	5
17	Malignant neoplasm of thyroid gland	5
18	Malignant neoplasm of biliary tract, part unspecified site	5
19	Multiple myeloma	5
20	Pleural effusion	4
21	Uterine leiomyoma	4
22	Lymphoid leukemia, acute	4

Table 7
MLH1 genes associated with CRC and comorbidities. Presented are numbers of abstracts in PubMed where these associations were described in 3 or more abstracts.

No	Disease	PubMed abstracts
1	Malignant neoplasm of stomach	38
2	Malignant neoplasm of ovary	24
3	Leukemia of unspecified cell type	10
4	Malignant neoplasm of bladder	7
5	Malignant neoplasm of prostate	7
6	Ulcerative colitis	6
7	Secondary and unspecified malignant neoplasm of lymph nodes	5
8	von Recklinghausen's disease	5
9	Huntington's chorea	3
10	Nodular lymphoma	3
11	Malignant neoplasm of uterus, part unspecified	3
12	Malignant neoplasm of pancreas	3
13	Endometrial hyperplasia with atypia	3

4. Discussion

Chronic diseases such as CRC are associated with many

Table 8
TGFBR2 genes associated with CRC and comorbidities. Presented are numbers of abstracts in PubMed where these associations were described in 2 or more abstracts.

No	Disease	PubMed abstracts
1	Marfan's syndrome	30
2	Malignant neoplasm of stomach	10
3	Malignant neoplasm of prostate	8
4	Malignant neoplasm of pancreas	7
5	Aneurysm	4
6	Alzheimer's disease	2
7	Asthma	2
8	Ulcerative colitis	2
9	Patent ductus arteriosus	2
10	Ehlers-Danlos syndrome	2
11	Systemic sclerosis	2
12	Migraine	2
13	Malignant neoplasm of mouth, unspecified	2
14	Malignant neoplasm of gallbladder	2
15	Dissection of aorta, unspecified site	2
16	Secondary and unspecified malignant neoplasm of lymph nodes	2

Table 9
PPARG genes associated with CRC and comorbidities on PubMed. Comorbidities that are associated with CRC and PPARG in 10 or more abstracts are given.

No	Disease	PubMed abstracts
1	Obesity	208
2	Diabetes mellitus	100
3	Atherosclerosis	37
4	Dysmetabolic syndrome x	35
5	Alzheimer's disease	34
6	Malignant neoplasm of prostate	33
7	Hypertensive disease	31
8	Heart failure	27
9	Huntington's chorea	21
10	Polycystic ovaries	17
11	Coronary atherosclerosis	14
12	Lipodystrophy	14
13	Malignant neoplasm of thyroid gland	13
14	Cardiovascular Diseases	12
15	Diabetic retinopathy	12
16	Malignant neoplasm of stomach	12
17	Diabetes with renal manifestations	11
18	Osteoarthrosis, unspecified whether generalized or localized	11
19	Parkinson's disease	11
20	Amyotrophic lateral sclerosis	10
21	Malignant neoplasm of pancreas	10

Table 10
Ninety-six genes associated with CRC and present in both sources: PubMed and DisGeNET.

ALDH1B1	CHEK2	IGF2	MTHFR	PTP4A3	SULT1A1
ALOX5	CTNNB1	IL1B	MUTYH	PTPN12	TAGLN
APC	CXCL10	IL32	MYC	PTPRJ	TCF7L2
ASCL2	DMBT1	JUN	NDRG2	RAD54B	TGFBR2
ATP7A	DNMT1	KCNH2	NFKB1	RCOR1	TNF
AXIN2	DPYD	KDM1A	NOS2	RECK	TNFRSF10A
BAX	EGFR	KRAS	NOTCH1	S100A4	TNFSF10
BCL2	ERBB2	LEF1	NOX1	SERPINB5	TP53
BECN1	FBP1	LEP	NOX4	SERTAD1	TP63
BIRC5	GAST	LGR5	ODC1	SLC2A1	TPM3
BRAF	GCG	MCM2	PCNA	SLC5A8	TYMP
BRD4	HES1	MECOM	POLB	SLCO1B3	TYMS
CBR1	HMGCS2	MIR98	PPARG	SOD2	VEGFA
CCAT1	HSPB1	MLH1	PRKN	SPARC	WT1
CCND1	ICAM1	MMP7	PROM1	SRC	XAF1
CDKN1A	IFNG	MMP9	PTGS2	STAT3	YBX1

comorbidities and complications, which affect the quality of life and the prognosis of CRC. Certain comorbidities (dementia, depressive disorder, alcoholism) are associated with a delayed CRC diagnosis, while some chronic comorbidities (T2DM, CVD, Hypertension) that require frequent medical care are associated with earlier CRC detection [6,27]. Patients with comorbidity often receive different CRC treatments (surgery, chemotherapy, radiation therapy) than patients without comorbidity [6,27].

Most published studies presented social network methods that include multiple diseases (phenotypes) not focusing on any particular disease as the main subject [7,28,29]. In our research we implemented a social network analysis of comorbidities of one disease (CRC). This cancer is mostly diagnosed at older age, when comorbidities are commonly present and when they can be an important risk and prognostic factor of CRC. Our prevalence and network analyses of comorbidities of CRC show which specific comorbidities are more prevalent and significant. The results of our study could have a practical implementation in medicine, by providing the information on which comorbidity conditions should be looked for as highly expected, and consequently prevented or treated. The results of our study show that the prevalence of some characteristic comorbidities, such as diverticulosis of colon, history of malignant neoplasms of intestines, benign neoplasms of intestines, paralytic ileus, and few others are higher in CRC than in the general population. Comparing findings between groups of patients younger or older than age 50, we can notice that patients younger than age 50 have significantly lower prevalence of essential hypertension, than patients older than age 50. Younger groups of patients don't have cardiovascular conditions such as coronary atherosclerosis, congestive heart failure, atrial fibrillation and other cardiac dysrhythmias, in the top 30 ranked comorbidities. They also don't have history of malignant neoplasms of prostate, lungs, colon, and other cancers, in the top 30 the most prevalent comorbidities. These results show that many comorbidities are influenced by aging of patients. Certain comorbidities such as history of malignant neoplasms of large bowel and rectum, paralytic ileus, intestinal obstruction, colostomy status, and peritoneal adhesions are present in both groups, younger and older than age 50, which confirms that some of comorbidities are influenced by development of CRC.

Constructed scale-free networks and calculated centrality measures show that comorbidities are interlinked beyond simple coincidence [7,28]. Centrality measures (node degree, betweenness) discovered the presence of highly connected comorbidities (hubs), presented in Figs. 1–4. Highly connected comorbidities, such as CHF, CAD, OCB, CKD, AKF, anemia, hypoxemia, and a few others should be followed and treated. Some of the conditions revealed as important by RR metric were: anemia, hypoxemia, colostomy status, gastro-duodenitis, chest pain, pressure ulcer-lower back, hyperkalemia, sepsis, etc. φ-correlation revealed comorbidities that are highly prevalent and expected. Evaluation of centrality measures discovered that both betweenness and node degrees come as valuable yet different measures considering findings of the most significant comorbidities associated with CRC.

Results of the comorbidity study were used to carry out analysis of associations between genes, CRC, and comorbidities. We extracted 627 more genes (described in 2 or more abstracts) on PubMed than on DisGeNET. Genetic findings could be used to recruit more individuals who would benefit from genetic testing and consultations [13–16]. Recent studies incorporated comorbidities in addition to the genomic data to identify new disease-genes associations [9,10,28,29]. If the same gene is linked to two different diseases, this is often an indication that the two diseases have a common genetic origin. Comorbidity networks and text mining of big data such as PubMed could help to find previously unknown genes associated with CRC [29]. CRC arises as the cumulative effect of multiple genetic mutations. In Table 4, we presented associations of the APC genes with CRC and comorbidities. The APC genes mutations are responsible for Familial adenomatous polyposis and hereditary non-polyposis CRC [30]. Mutations in the APC gene play a

pivotal role in CRC pathogenesis. The APC became one of the most frequently mutated, known driver genes in CRC [31]. The APC tumor suppressor is mutated or hyper-methylated in some breast cancers [32], and may also be associated with the development and progression of bladder cancer and prostate cancer [33,34].

In Table 5, the most often cancers and comorbidities associated with TP53 gene are listed. The TP53 has an important role in several fundamental processes such as cancer, aging, senescence, and DNA repair. Mutations in the TP53 gene are common in CRC [35,36], brain tumors, leukemia, and lymphomas [36]. BeFree text-mining discovered associations between the TP53 gene and different cancers as well as Hodgkin's disease, ulcerative colitis, rheumatoid arthritis, Alzheimer's disease, etc.

Table 6, presents the most often conditions associated with KRAS genes. Ucar and colleagues found that the presence of multiple mutations in KRAS indicates better overall survival than a single mutation [37]. Perdyan and colleagues studied the presence of KRAS mutation as a prognostic factor of CRC [38]. KRAS, NRAS, and BRAF mutations are found in half of myeloma patients and contribute to proteasome inhibitor (PI) resistance [39]. RAS genes (HRAS, KRAS, and NRAS) are seen as some of the top causes of cancer deaths in the U.S. (lung, colorectal, and pancreatic cancer) which influenced that anti-RAS therapies became a major field for cancer research [40]. Our text mining identified associations between the KRAS gene and viral hepatitis C, chronic pancreatitis, ulcerative colitis, dengue, atrial premature beats, endometriosis, leiomyoma, etc.

The most often conditions associated with MLH1 genes are shown in Table 7. Researchers demonstrated that MLH1 deficiency and tumor progression and metastasis are in close relation [41]. A defective DNA mismatch repair (MMR) of genes especially MLH1 and MSH2 is frequently involved in the carcinogenesis of various tumors including gastric cancer [42].

Associations of CRC, comorbidities, and TGFBR2 gene are presented in Table 8. In CRC with microsatellite instability (MSI), the majority of cases are affected by inactivating mutations of TGFBR2 [43]. Genetic variants in TGFBR1 and TGFBR2 genes have been associated with hereditary connective tissue disorders including thoracic aortic aneurysm and dissection, Marfan syndrome, and Loeys-Dietz syndrome [44]. BeFree system discovered associations between the TGFBR2 gene and Alzheimer's disease, asthma, ulcerative colitis, migraine, etc.

In Table 9, PPARG genes in associations with CRC comorbidities are listed. Studies described how epigenetic modifications influence PPARG gene expression in CRC [45]. PPARG has been analyzed in the regulation of metabolism, inflammation, atherosclerosis, cell differentiation, and proliferation, linking PPARG with conditions such as obesity and diabetes, cardiovascular disease, and cancer [46].

Network data analysis used in our research provides insight into comorbidities and genes associated with CRC, which could help medical experts to formulate appropriate preventive health measures to address genes and high-risk comorbidities associated with CRC. Our study has several limitations. The social network approach may underestimate rare diseases, that don't show as hubs, but they could also affect outcomes. Sometimes testing for comorbidity diseases could be expensive, which could limit confirmation of those conditions.

5. Conclusion

Numerous studies point out that the presence of comorbidities weighs as a negative factor that contributes to faster deterioration of one's health. Our findings suggest which comorbidities should be highly expected along the course of the cancer disease. The results of this study contribute to a better understanding of risk factors (genes, comorbidities) to the development of CRC. Genes associated with CRC and comorbidities, found only in PubMed abstracts should be evaluated as a potential risk or predictive factor for the development of comorbidity conditions. Subsequent genetic research is necessary to evaluate genes-CRC-comorbidities associations and incorporate findings into expert domain-specific curated databases.

Authors' contributions

BL, MP, ZO, SR, JA, VB, CVN designed the project. BL wrote the manuscript. BL, MP, JA searched and downloaded data from HCUP. CVN, VB text mined PubMed and DisGeNet. BL performed comorbidity network analysis. BL, SR, VB, and CVN completed genes-diseases associations analysis. ZO performed writing-review and editing. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that there are no competing interests regarding the publication of this paper.

Availability of data and material

The datasets generated and analyzed in the comorbidity part of this study are available in the HCUP SID database: https://www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp.

The genes were extracted by the text mining tool BeFree from the PubMed: https://www.ncbi.nlm.nih.gov/pubmed/.

Funding

This research is partly sponsored by the King Abdullah University of Science and Technology's Center Partnership Fund Program.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge the support of the King Abdullah University of Science and Technology's Center Partnership Fund Program. Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality provided data used in this study. The authors also thank Stefan Obradovic for proofreading and editing the language of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100492.

References

[1] Kumar V, Abbas A, Fausto N, et al. Robbins and Cotran. Pathologic basis of diseases. ninth ed. 2014.
[2] Survival rates for colorectal cancer. by Stage, https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/survival-rates.html.
[3] Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg 2009;22(4):191–7. https://doi.org/10.1055/s-0029-1242458.
[4] Gallagher DJ, Kemeny N. Metastatic colorectal cancer: from improved survival to potential cure. Oncology 2010;78(3–4):237–48.
[5] Hahn EE, Gould MK, Munoz-Plaza CE, et al. Understanding comorbidity profiles and their effect on treatment and survival in patients with colorectal cancer. J Natl Compr Canc Netw 2018;16(1):23–34. https://doi.org/10.6004/jnccn.2017.7026.
[6] Boakye D, Rillmann B, Walter V, et al. Impact of comorbidity and frailty on prognosis in colorectal cancer patients: a systematic review and meta-analysis. Canc Treat Rev 2018;64:30–9. https://doi.org/10.1016/j.ctrv.2018.02.003.
[7] Hidalgo CA, Blumm N, Barabasi AL, al at. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol 2009;5(4):e1000353.

[8] Khan A, Uddin S, Srinivasan U. Comorbidity network for chronic disease: a novel approach to understand type 2 diabetes progression. Int J Med Inf 2018;115:1–9. https://doi.org/10.1016/j.ijmedinf.2018.04.001.
[9] Chen Y, Zhang X, Zhang G, et al. Comparative analysis of a novel disease phenotype network based on clinical manifestations. J Biomed Inf 2015;53:113–20. https://doi.org/10.1016/j.jbi.2014.09.007.
[10] Hossain ME, Uddin S, Khan A, et al. A framework to understand the progression of cardiovascular disease for type 2 diabetes mellitus patients using a network approach. Int J Environ Res Publ Health 2020;17(2):E596. https://doi.org/10.3390/ijerph17020596.
[11] Genetics of colorectal cancer. https://www.cancer.gov/types/colorectal/hp/colorectal-genetics-pdq#link/_89.
[12] Kanth P, Grimmett J, Champine M, et al. Hereditary colorectal polyposis and cancer syndromes: a primer on diagnosis and management. Am J Gastroenterol 2017;112(10):1509–25.
[13] Rustgi AK. The genetics of hereditary colon cancer. Genes Dev 2007;21(20):2525–38.
[14] Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354(26):2751–63.
[15] Hampel H, Bennett RL, Buchanan A, et al. A practice guideline from the American College of medical genetics and genomics and the national society of genetic counselors: referral indications for cancer predisposition assessment. Genet Med 2015;17(1):70–87.
[16] Bravo A, Piñero J, Queralt-Rosinach N, et al. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. BMC Bioinf 2015;16:55. https://doi.org/10.1186/s12859-015-0472-9.
[17] Pinero J, Bravo A, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2017;45:833–9. https://doi.org/10.1093/nar/gkw943.
[18] https://www.hcup-us.ahrq.gov/databases.jsp.
[19] Ljubic B, Gligorijevic Dj, Gligorijevic J, et al. Social network analysis for better understanding of influenza. J Biomed Inf 2019;93:103161. https://doi.org/10.1016/j.jbi.2019.103161.
[20] https://en.wikipedia.org/wiki/Betweenness_centrality.
[21] https://ghr.nlm.nih.gov/gene/APC.
[22] https://ghr.nlm.nih.gov/gene/TP53.
[23] https://ghr.nlm.nih.gov/gene/KRAS.
[24] https://ghr.nlm.nih.gov/gene/MLH1.
[25] https://ghr.nlm.nih.gov/gene/TGFBR2.
[26] https://ghr.nlm.nih.gov/gene/PPARG.
[27] Sogaard M, Thomsen RW, Bossen KS, et al. The impact of comorbidity on cancer survival: a review. Clin Epidemiol 2013;5(Suppl 1):3–29.
[28] Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12(1):56-68. https://doi.org/10.1038/nrg2918.
[29] Wang X, Gulbahce N, Yu H, et al. Network-based methods for human disease gene prediction, Briefings Funct. Genom., https://doi.org/10.1093/bfgp/elr024.
[30] Fearnhead NS, Wilding JL, Bodmer WF. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. Br Med Bull 2002;64:27–43. https://doi.org/10.1093/bmb/64.1.27.
[31] Wachsmannova L, Mego M, Stevurkova V, et al. Novel strategies for comprehensive mutation screening of the APC gene. Neoplasma 2017;64(3):338–43. https://doi.org/10.4149/neo_2017_303.
[32] Stefanski CD, Keffler K, McClintock S, et al. APC loss affects DNA damage repair causing doxorubicin resistance in breast cancer cells. Neoplasia 2019;21(12):1143–50. https://doi.org/10.1016/j.neo.2019.09.002. Epub 2019 Nov 20. PMID: 31759252; PMCID: PMC6872841.
[33] Bai ZJ, Liu Q, Wang XS, et al. APC promoter methylation is correlated with development and progression of bladder cancer, but not linked to overall survival: a meta-analysis. Neoplasma 2019;66(3):470–80. https://doi.org/10.4149/neo_2018_181009N753.
[34] Richiardi L, Fiano V, Vizzini L, et al. Promoter methylation in APC, RUNX3, and GSTP1 and mortality in prostate cancer patients. J Clin Oncol 2009;27(19):3161–8. https://doi.org/10.1200/JCO.2008.18.2485.
[35] Porcelli B, Frosi B, Terzuoli L, et al. Expression of p185 and p53 in benign and malignant colorectal lesions. Histochem J 2001;33(1):51–7. https://doi.org/10.1023/a:1017543930661.
[36] Tolomeo D, L'Abbate A, Lonoce A, et al. Concurrent chromothripsis events in a case of TP53 depleted acute myeloid leukemia with myelodysplasia-related changes. Canc. Genet. 2019;237:63–8. https://doi.org/10.1016/j.cancergen.2019.06.009.
[37] Ucar G, Ergun Y, Aktürk Esen S, et al. Prognostic and predictive value of KRAS mutation number in metastatic colorectal cancer. Medicine (Baltim) 2020;99(39):e22407. https://doi.org/10.1097/MD.0000000000022407.
[38] Perdyan A, Spychalski P, Kacperczyk J, et al. Circulating Tumor DNA in KRAS positive colorectal cancer patients as a prognostic factor - a systematic review and meta-analysis. Crit Rev Oncol Hematol 2020;154:103065. https://doi.org/10.1016/j.critrevonc.2020.103065.
[39] Shirazi F, Jones RJ, Singh RK, et al. Activating KRAS, NRAS, and BRAF mutants enhance proteasome capacity and reduce endoplasmic reticulum stress in multiple myeloma. Proc Natl Acad Sci U S A 2020;117(33):20004–14. https://doi.org/10.1073/pnas.2005052117.
[40] Waters AM, Der CJ. KRAS: the critical driver and therapeutic target for pancreatic cancer. Cold Spring Harb. Perspect. Med. 2018;8(9):a031435. https://doi.org/10.1101/cshperspect.a031435.
[41] Ackermann A, Schrecker C, Bon D, et al. Downregulation of SPTAN1 is related to MLH1 deficiency and metastasis in colorectal cancer. PloS One 2019;14(3):e0213411. https://doi.org/10.1371/journal.pone.0213411.
[42] Haron NH, Mohamad Hanif EA, Abdul Manaf MR, et al. Microsatellite instability and altered expressions of MLH1 and MSH2 in gastric cancer. Asian Pac J Cancer Prev APJCP 2019;20(2):509–17. https://doi.org/10.31557/APJCP.2019.20.2.509.
[43] Fricke F, Mussack V, Buschmann D, et al. TGFBR2-dependent alterations of microRNA profiles in extracellular vesicles and parental colorectal cancer cells. Int J Oncol 2019;55(4):925–37. https://doi.org/10.3892/ijo.2019.4859.
[44] De Cario R, Sticchi E, Lucarini L, et al. Role of TGFBR1 and TGFBR2 genetic variants in Marfan syndrome. J Vasc Surg 2018;68(1):225–33. https://doi.org/10.1016/j.jvs.2017.04.071. e5.
[45] Sabatino L, Fucci A, Pancione M, et al. PPARG epigenetic deregulation and its role in colorectal tumorigenesis. PPAR Res 2012;2012:687492. https://doi.org/10.1155/2012/687492.
[46] Bandera Merchan B, Tinahones FJ, Macías-González M. Commonalities in the association between PPARG and vitamin D related with obesity and carcinogenesis. PPAR Res 2016;2016:2308249. https://doi.org/10.1155/2016/2308249.