Deep Next Generation Sequencing Identifies Somatic Mutational Signature in Egyptian Colorectal Cancer Patients

Amira Youssef
National Cancer Institute Cairo University https://orcid.org/0000-0003-3091-691X
Mohamed A. Abdel-Fattah
Ain Shams University Faculty of Agriculture
Ahmed O. Touny
NCI: National Cancer Institute Cairo University
Zeinab K. Hassan
National Cancer Institute Cairo University
Auhood Nassar
National Cancer Institute Cairo University
Mai M. Lotfy
National Cancer Institute Cairo University https://orcid.org/0000-0001-8282-8773
Ahmed Moustafa
AUC: The American University in Cairo
Mohammed Mohey Eldin
Cairo University Kasr Alainy Faculty of Medicine
Abeer Bahnassy
National Cancer Institute Cairo University
Abdel-Rahman N. Zekri (ncizekri@yahoo.com)
Cancer Biology Department, National Cancer Institute, Cairo University, Cairo, Egypt

Research

Keywords: Next generation sequencing, Egyptian colorectal cancer, Somatic mutations

Posted Date: December 18th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-125305/v1

Abstract

Background: Colorectal cancer (CRC) incidence is progressively increasing in Egypt. Unfortunately, there is inadequate knowledge of the acquired somatic mutations in Egyptian CRC patients which limit our understanding of its progression. To the best of our knowledge, our study is the first to sequence multiple-gene panel to identify the somatic mutation pattern associated with CRC disease progression in a cohort of Egyptian patients. Custom **72** genes, which are frequently associated with CRC, were sequenced using Qiaseq UMI-based targeted DNA panel in **120** fresh tissues classified into; inflammatory bowel disease (IBD; **n=20)**, colonic polyp **(**CP; **n=38)** and CRC **(n=62)** as well as **20** biopsies with non-specific colitis served as a control group **(n=20).**

Results: Using Ingenuity Variant Analysis (IVA), we revealed that ***APC, TP53 & ATM*** genes harbored the highly frequent CRC-specific somatic mutations **(15, 11 & 6, respectively)**. We also identified common somatic mutations (predictors) that were associated with disease progression from colitis to CRC; ***APC*** **(c.1742delA (65%)),** ***TP53*** **(c.121delG (58%), c.215C>G (52%)),** ***ATM*** **(c.640delT (16%)),** ***IGF2*** **(c.677delG (56%)),** ***RET*** **(c.2071G>A (37%)),** ***ACVR2A*** **(c.1310delA (26%)),** ***PIK3CA (*****c.1173A>G (16%))** **&** ***KIT*** **(c.1621A>C (13%)).** Furthermore, pathway analysis using Ingenuity Pathway Analysis (IPA) showed that Wnt/βcatenin, ATM signaling, RTK-RAS and TGF-β were the most altered pathways in the CRC group **(73%. 72%, 40% & 36%, respectively).**

Conclusion: In this data set, we shed the light on the most frequent somatic mutations and the most altered pathways that are crucial for understanding colorectal cancer predisposition and developing personalized therapies for the Egyptian CRC patients.

Background

Colorectal cancer (CRC) is one of the leading causes of mortality and morbidity worldwide. It is the third most common neoplasm and the second leading cause of cancer-related death world-wide according to **Globocan 2018** statistics [1]. In Egypt, the CRC is the 7th most common malignant tumor in Egypt; representing 4% of the totally diagnosed cancers and 53% of the GIT cancers [2, 3].

The majority of the CRC cases (60–65%) arise sporadically through a cascade of acquired somatic genomic alterations, while around 35–40% of the CRCs are hereditary [4]. The pathogenesis of the CRC is very heterogeneous and influenced by multiple factors related to dietary habits, genetic predisposition, long-standing inflammatory bowel disease and presence of colorectal polyps [4].

The newly applied technique of next generation sequencing (NGS) has enabled accurate characterization of mutations in patients' genome. Also, it permitted considerable insights into the mutational processes functioning in various types of cancers including the CRC [5]. The NGS also enhanced our understanding regarding the biology of diseases and their contribution to genetic screening and patients' management [6]. Targeted sequencing is a good clinical application of NGS technology. Advantage of the targeted approach is the increased coverage depth compared to whole exome approach due to reducing the number of genes analyzed with similar number of base pairs sequenced [7]. Moreover, using of modern molecular barcoding techniques could overcome the PCR errors generated from targeted PCR-based library techniques as well as detect true positive variants at minimal allele frequencies (MAFs) up to 0.1% [8]. This enables generation of reliable data with sufficient sequencing depth and minimal error rate in the targeted genes of interest. Therefore, we established a targeted PCR-based NGS approach with integrated unique molecular indices (UMIs) and high coverage depth to investigate somatic mutational signature in the Egyptian CRC patients using custom Qiaseq DNA targeted panel. To identify; **a)** cancer driver genes in colorectal carcinogenesis process, **b)** colorectal cancer specific variants that can be used for establishing routine mutation detection assay, **c)** personalized targeted therapy for the Egyptian colorectal cancer patients.

Materials And Methods

Patient samples

Fresh colonscopic biopsy samples were collected from the patients classified into; inflammatory bowel disease **(IBD; n = 20)**, colonic polyp **(CP; n = 38)** and colorectal carcinoma **(CRC; n = 62)** patients as well as participants with chronic non-specific colitis who served as a control group **(n = 20)**. The clinicopathological features of the studied groups including age, gender histological type and grade, recurrence and metastasis were collected from the NCI clinical records.

DNA preparation

The DNA was isolated from the collected biopsies using the QIAamp® DNA mini kit **(Cat. No. 51304, Qiagen, Germany)** according to the manufacturer's instructions. The purified DNA was measured using Qubit® 3.0 Fluorometer **(Cat. No, Q33216, Thermofischer Scientific Inc, USA)** with Qubit™ dsDNA HS assay kit **(Cat. No. Q32854, Thermofischer Scientific Inc, USA).**

Qiaseq unique molecular index (UMI)- based targeted DNA library preparation, template preparation and sequencing

The Qiaseq UMI- based DNA Library was constructed using QIAseq Targeted DNA Panel **(Cat. No. EDHS-10082-002Z-3002, Qiagen, Germany**) according to manufacturer instructions. The QIAseq Targeted DNA Panels are designed to enrich selected genes and regions using 10–40 ng fresh DNA. After library preparation, the QIAxcel **(Cat No. 900194 Qiagen, Germany)** was used to check the fragment size and concentration with QIAxcel DNA high resolution kit **(Cat No. 929002, Qiagen, Germany).** The prepared libraries showed a size distribution between **200–1000 bp.** The prepared libraries were quantified using QIAseq Library Quant Assay Kit (**Cat No. 333304, Qiagen, Germany**) according to manufacturer's instructions. Libraries with a different sample index were combined in equimolar amounts as a similar sequencing depth was needed for each library. The combined libraries were preceded to template preparation on the ion chef using the Ion PI Hi-Q Chef Kit (Cat. No. A27198, Life Technologies, USA) and loaded into an Ion PI Chip (Cat. No. A26770, Life Technologies, USA) in order to be sequenced on the Ion proton using the Ion Proton Sequencing 200 Kit v2 **(Cat. No. 4485149, Life Technologies, USA).**

Bioinformatics analysis

The Data were initially analyzed using QIAGEN GeneGlobe Data Analysis Center. The bioinformatics analysis started with Quality Control (QC) checking step by examining the reads of each NGS run and trimming out the reads with low quality. The unaligned bam (UBAM) files were aligned to the human reference genome (version hg19). The run was excluded if the aligned reads was less than 100X depth of coverage, the UMI coverage was lower than 400, and if the target region was covered with less 95% with at least 5% of the mean UMI coverage as explained by **Kupe et al.** [8].The Catalogue of Somatic Mutations in Cancer (COSMIC) was used to identify pathogenic somatic mutations. Variants that exist in the COSMIC database with MAFs less than 0.5 were retained, whereas variants that didn't exist in the COSMIC database or variants that were flagged as SNP were filtrated out. The identified variants were classified into benign and pathogenic according to ClinVar database [9]. The functional consequences were predicted using Sift [10], PolyPhen-2 [11], and CADD [12] tools. The QIAGEN Ingenuity Pathway Analysis (IPA; QIAGEN) and the Ingenuity Variant Analysis plugin (IVA; QIAGEN) were used for further filtering, annotation as well as interpretation of variants detected in the UMI-based analysis and pathway analysis [8]. Data visualization was performed using R package **(version 3.6).** The oncoplot and the Lollipop plots were visualized using maftools [13], while the chord diagram was visualized using circlize tools [14].

Statistical analysis

The clinicopathological features of the assessed patients were analyzed using SPSS software package (**version 22**). Continuous variables were expressed as mean ± SD, and range, while categorical variables were expressed as percentages. Comparisons between groups were analyzed by $\chi 2$ test or Fisher's exact test when appropriate for the categorical variables, and by Mann-Whitney test or Student's t-test when appropriate for the continuous variables. P-value was considered significant when P-value **≤ 0.05.**

Results

Clinical data of the studied participants

The patients were classified according to age, gender, histological type, grade, recurrence and metastasis **(**Table 1**).**

Table 1
Clinical features of the studied groups

	CRC	Polyp	IBD	Control	*P-value*
	n = 62	n = 38	n = 20	n = 20	
Age:	51 ± 12a	54 ± 12a	45 ± 15a	51 ± 16a	0.151
Mean ± SD (Range)	(24–76)	(26–74)	(22–74)	(21–77)	
Gender	28 (45%) a	20 (53%) a	10 (50%) a	10 (50%) a	0.59
Male	34 (55%) a	18 (47%) a	10 (50%) a	10 (50%) a	
Female					
Histological type	Adenocarcinoma (82%) [a]	Typical lesion (40%) [a]	-	-	< 0.001
	Mucinous adenocarcinoma (6%) [b]	Atypical lesion (47%) [a]			
	Signet Ring adenocarcinoma (5%) [b]	Mixed lesions (13%) [a]			
	Neuroendocrine carcinoma (5%) [b]				
	Squamous cell carcinoma (2%) [b]				
Site	20 (33%) [a]	30 (79%) [a]	11 (55%) [a]	16 (80%) [a]	< 0.001
Colon	34 (54%) [a]	6 (16%) [b]	9 (45%) [a]	2 (10%) [b]	
Rectum	8 (13%) [a]	2 (5%) [b]	0 (0%) [b]	2 (10%) [b]	
Sigmoid					
Grade	1 (2%) [b]	-	-	-	< 0.001
In situ	40 (64%) [a]				
II	3 (5%) [a]				
III	18 (29%) [a]				
NA					
Metastasis	1 (2%) [a]	-	-	-	< 0.001
Yes	61 (98%) [b]				
No					
Recurrence	2 (3%) [a]	-	-	-	< 0.001
Yes	60 (97%) [b]				
No					

Distribution of non-synonymous variants in the studied Groups

The variant mean depth of coverage ranged from 500 to 1000X in the all studied groups **(Figure S1 a).** The heatmap showed the pattern of the non-synonymous mutations in each group **(Figure S1 b).**

The frequency of the detected somatic mutations

Our results revealed that CRC group has the highest somatic mutation burden compared to the CP, IBD and control groups **(168 mutations vs. 124, 88, and 42, respectively).** The changes from reference alleles to alternative alleles including transition, transversion, insertion or deletion in each group were illustrated in Fig. 1a.

It has been showed that TP53, APC & ATM genes harbored the most frequently detected somatic mutations **(100, 65 & 52 mutations, respectively)** in the CRC group while; the TP53, TCERG1 & ATM genes harbored the most frequently detected somatic mutations in the CP group **(50, 47& 46 mutations, respectively)**. Regarding IBD and CP groups; TP53, ATM & TCERG1 genes harbored the most frequently detected somatic mutations in both of them **(35, 31 & 21 mutations in the IBD group, respectively)** and **(20, 11 & 12 mutations in the CP group, respectively).** The control group harbored the least frequently detected somatic mutations when compared to the other groups (Fig. 1b**).**

Somatic mutations associated with disease progression (Predictors)

Data analysis showed that **9 out of 24** somatic mutations were identified in all the studied groups and associated with disease progression with increasing frequencies from colitis to CRC; ACVR2A **(c.1310delA)**, APC **(c.640delT)**, ATM **(c.5557G > A)**, IGF2 **(c.677delG)**, KIT **(c.1621A > C)**, PIK3CA **(c.1173A > G)**, RET **(c.2071G > A), &** TP53 **(c.121delG, c.215C > G)** (Table 2). The identified somatic mutations which were not associated with disease progression were shown in **Table S1.**

Table 2
Common somatic mutations associated with disease progression detected in the studied groups (predictors)

Gene _Name	CHROM	ID	Annotation	IVA	Rank	HGVS.c	HGVS.p	Control (n = 20)	IBD (n = 20)	CP (n = 38)	CRC (n = 62)	Occurren in cosmi database
ACVR2A	chr2	COSM252949	Framshift	Loss	10/11	c.1310delA	p.Lys437fs	2 (10%)	4 (20%)	8 (21%)	16 (26%)	Large intestine 210
APC	chr5	COSM9101916	Framshift	Loss	16/16	c.3754delT	p.Ser1252fs	4 (20%)	6 (30%)	19 (50%)	40 (65%)	Large intestine 1
ATM	chr11	COSM41596*	Missense	Loss	37/63	c.5557G > A	p.Asp1853Asn	1 (5%)	2 (10%)	4 (10%)	10 (16%)	Breast & Soft tissu =10
IGF2	chr11	COSM6955852	Framshift	Loss	5/5	c.677delG	p.Gly226fs	8 (40%)	10 (50%)	22 (57%)	35 (56%)	Large intestine 2
KIT	chr4	COSM28026**	Missense	Gain	10/21	c.1621A > C	p.Met541Leu	1 (5%)	2 (10%)	4 (10%)	8 (13%)	Breast r = 13 & Larg intestine 5
PIK3CA	chr3	COSM328028	Missense	Gain	7/21	c.1173A > G	p.Ile391Met	1 (5%)	1 (5%)	4 (10%)	10 (16%)	Large intestine 12
RET	chr10	COSM1666596*	Missense	Gain	11/20	c.2071G > A	p.Gly691Ser	4 (20%)	5 (25%)	14 (36%)	23 (37%)	Soft tissu = 7
TP53	chr17	COSM46244*	Framshift	Loss	4/11	c.121delG	p.Asp41fs	5 (25%)	6 (30%)	21 (55%)	36 (58%)	Head &neck = 1
	chr17	COSM2745057**	Missense	Loss	4/11	c.215C > G	p.Pro72Arg	7 (35%)	7 (35%)	20 (52%)	32 (52%)	Large intestine 35

Rank indicates affected exon/ total exons of the specified gene.

***Somatic mutations found in other tissues rather than large intestine.**

****Somatic mutations found previously in the Egyptian CRC patients.**

The frequency of the variants detected in the studied groups was illustrated in the chord diagram and the heatmap **(**Figs. 2a&2b**)**. The diagnostic accuracy of the variants detected in all the studied groups was demonstrated by the random forest plot (Fig. 3**).** It showed that the variants detected in CRC group had the highest diagnostic accuracy in discriminating the CRC group from the other groups. Also, the CRC model had the lowest error rate compared to the models of the CP, IBD and control groups **(0.174 vs. 0.7, 1.0 & 1.0, respectively).**

Pathogenic somatic mutations identified in the CRC group

The somatic mutational burden per sample in the CRC group was displayed in **Figure S2 a**, the median was **7** somatic mutations per sample with frame shift deletion predominance. The most predominate single nucleotide variant (SNVs) type was the transition **C > T** (**Figure S2 b).** The top 12 highly mutated genes were displayed in the oncoplot **(**Fig. 4**)**; conceiving that the TP53 and APC were the most frequently mutated genes in CRC group (**73% and 69%** respectively).

The schematic representations of the APC & TP53 genes at the genetic and protein levels were shown in Figs. 5& 6. We found that exon 16 and exon 4 were the most frequently mutated exons in the APC & TP53 genes, respectively. The β-catenin binding and down regulation sites were the most affected regions in the **APC** protein, whereas the transactivation and the proline rich sites were the most affected regions in the **TP53** protein.

The Loss-of-function and gain of function cancer driver single nucleotide variants (SNVs) detected in the CRC patients compared to the catalogued SNVs in COSMIC database were summarized in Figs. 7&8, respectively.

Our data showed that the APC gene had **15** pathogenic somatic mutations in the CRC group only. They were categorized into **9** frame shift deletions **(c.4393_4394delAG, c.4474delG, c.4666delA, c.4666dupA, c.6747dupA, c.1847delT, c.6579dupA, c.6149dupA & c.5288delA)** and **6** stop-gained somatic mutations **(c.1495C > T, c.2055G > A, c.2309C > G, c.4588G > T, c.8446C > T & c.3856G > T).** On the other hand, two missense mutations were found in AXIN2 gene **(c.2347G > T& c.1975C > T)** with loss of function as well as one missense mutation in DKK3 **(c.995G > A)** and another one was framshift mutation **(c.966delC**) in TCF7L (Table 3**).**

Table 3
Specific somatic mutations related to Wnt-Beta catenin pathway detected in the CRC group

Gene_ Name	CHROM	ID	VMF	Annotation	IVA	Rank	HGVS.c	HGVS.p	Occurrence in the current study	Occurrence in cosmic database
APC	chr5	COSM29364	0.46	stop_gained	Loss	12/16	c.1495C > T	p.Arg499*	CRC = 1	Large intestine = 45
	chr5	COSM13124	0.43	stop_gained		16/16	c.2055G > A	p.Trp685*	CRC = 1	Large intestine = 1
	chr5	COSM6602057	0.35	stop_gained		16/16	c.2309C > G	p.Ser770*	CRC = 1	Large intestine = 1
	chr5	COSM18873	0.07	Framshift		16/16	c.4393_4394delAG	p.Ser1465fs	CRC = 1	Large intestine = 100
	chr5	COSM19340	0.11	Framshift		16/16	c.4474delG	p.Ala1492fs	CRC = 1	Large intestine = 8
	chr5	COSM1432429*	0.09	stop_gained		16/16	c.4588G > T	p.Glu1530*	CRC = 1	Pancreas = 8
	chr5	COSM19000	0.5	Framshift		16/16	c.4666delA	p.Thr1556fs	CRC = 1	Large intestine = 9
	chr5	COSM19695	0.16	Framshift		16/16	c.4666dupA	p.Thr1556fs	CRC = 1	Large intestine = 250
	chr5	COSM1432471	0.03	stop_gained		16/16	c.8446C > T	p.Arg2816*	CRC = 1	Large intestine = 1
	chr5	COSM18772	014	stop_gained		16/16	c.3856G > T	p.Glu1286*	CRC = 1	Large intestine = 23
	chr5	COSM6444668	0.15	Framshift		16/16	c.6747dupA	p.Gly2250fs	CRC = 1	Large intestine = 2
	chr5	COSM4169170	0.41	Framshift		16/16	c.1847delT	p.Leu616fs	CRC = 1	Large intestine = 1
	chr5	COSM5010257	0.63	Framshift		16/16	c.6579dupA	p.Val2194fs	CRC = 1	Large intestine = 1
	chr5	COSM6601990	0.36	Framshift		16/16	c.6149dupA	p.Lys2051fs	CRC = 2	Large intestine = 1
	chr5	COSM6341833*	0.13	Framshift		16/16	c.5288delA	p.Asn1763fs	CRC = 1	Liver = 1
AXIN2	chr17	COSM317040 *	0.47	Missense	Loss	10/11	c.2347G > T	p.Ala783Ser	CRC = 1	Lung = 3
	chr17	COSM357732*	0.56	Missense		8/11	c.1975C > T	p.Arg659Trp	CRC = 3	Liver = 2 & Lung = 1
DKK3	chr11	COSM923908	0.54	Missense	Normal	8/8	c.995G > A	p.Arg332His	CRC = 1	Large intestine = 1 & Endometrium = 1
TCF7L	chr2	COSM1409764	0.51	Framshift	Loss	8/12	c.966delC	p.Ser323fs	CRC = 1	Large intestine = 3

Rank indicates affected exon/ total exons of the specified gene.

*Somatic mutations found in other tissues rather than the large intestine.

On the other hand, the TP53 gene harbored **10** pathogenic mutations. They were classified into **4** missense mutations **(c.743G > A, c.581T > A, c.581T > C & c.524G > A), two** protein-protein contact mutations **(c.844C > T & c.766A > C), one** splice-acceptor and intron-variant **(c.673-1G > C)** and **3** stop-gained variants **(c.1024C > T, c.916C > T & c.281C > G)**. Moreover, the ATM gene harbored **8** mutations; categorized into **5** missense mutations **(c.1010G > A, c.2803A > G, c.8138G > A, c.9007A > G & c.8138G > A), two** frameshift deletions **(c.3510delA & c.6908delA)** and **one** stop-gained variant **(c.8977C > T)** (Table 4).

Table 4
Specific somatic mutations related to ATM & P53 signaling pathway detected in the CRC group

Gene_ Name	CHROM	ID	VMF	Annotation	IVA	Rank	HGVS.c	HGVS.p	Occurrence in the current study	Occurrence in cosmic database
ATM	chr11	COSM1158322	0.51	missense	Loss	8/63	c.1010G > A	p.Arg337His	CRC = 1	Large intestine = 8
	chr11	COSM428365	0.05	stop-gained		63/63	c.8977C > T	p.Arg2993Ter	CRC = 1	Large intestine = 8
	chr11	COSM3733315*	0.5	missense		63/63	c.9007A > G	p.Asn3003Asp	CRC = 1	Hemapioteic and Lymphoid = 1
	chr11	COSM1350832	0.26	frameshift		24/63	c.3510delA	p.Lys1170fs	CRC = 1	Large intestine = 2
	chr11	COSM6933909	0.2	frameshift		47/63	c.6908delA	p.Lys2303fs	CRC = 1	Large intestine = 1
	chr11	COSM7160489*	0.42	missense		55/63	c.8138G > A	p.Arg2713Lys	CRC = 3	hemapioteic and lymphoid = 1
TP53	chr17	COSM11073**	0.29	stop_gained	Loss	10/11	c.1024C > T	p.Arg342*	CRC = 2	Large intestine = 55
	chr17	COSM99947**	0.42	stop_gained		8/11	c.916C > T	p.Arg306*	CRC = 1	Large intestine = 70
	chr17	COSM99925**	0.40	protein_protein_ contact		8/11	c.844C > T	-	CRC = 1	Large intestine = 240
	chr17	COSM6492144	0.13	protein_protein _contact		7/11	c.766A > C	-	CRC = 1	Large intestine = 1
	chr17	COSM99602**	0.47	missense		7/11	c.743G > A	p.Arg248Gln	CRC = 1	Large intestine = 240
	chr17	COSM562647	0.54	splice_acceptor & intron_variants		6/10	c.673-1G > C	-	CRC = 1	Large intestine = 2
	chr17	COSM43623	0.25	missense		6/11	c.581T > A	p.Leu194His	CRC = 1	Large intestine = 2
	chr17	COSM43827	0.55	missense		6/11	c.581T > C	p.Leu194Pro	CRC = 1	Large intestine = 4
	chr17	COSM10648**	0.4	missense		5/11	c.524G > A	p.Arg175His	CRC = 3	Large intestine = 500
	chr17	COSM45653*	0.16	stop_gained		4/11	c.281C > G	p.Ser94*	CRC = 1	Osephagous = 2
	chr17	COSM45766	0.46	frameshift		4/11	c.304delA	p.Thr102fs	CRC = 1	Breast = 4 & Large intestine = 1

Rank indicates affected exon/ total exons of the specified gene.

*Somatic mutations found in other tissues rather than the large intestine.

**Somatic mutations found previously in the Egyptian CRC patients.

Additionally, four mutations were identified in the SMAD4 gene; one frame shift mutation **(c.692delG)** and **3** missense mutations **(c.1064A > G, c.1081C > T & c.1088G > A)**. Another **4** frame shift mutations were detected in ARIDA1 **(c.3216delA, c.4555delC, c.5548delG & c.5548dupG).** Furthermore, three mutations were detected in EP300; categorized into a single splice-donor and intron-variant **(c.3671 + 1G > A)**, a single frame shift variant **(c.832delA)** and another single missense variant **(c.1058G > A).** Also, 2 somatic mutations were found in FBXW7 gene; classified into one frameshift variant **(c.2001delG)** and another stop-gained variant **(c.4900A > G)** (Table 5**).**

Table 5
Specific somatic mutations related to Apoptosis and TGF-beta signaling pathways detected in the CRC group

Gene Name	CHROM	ID	VMF	Annotation	IVA	Rank	HGVS.c	HGVS.p	Occurrence in the current study	Occurrence in cosmic database
EP300	chr22	COSM84765*	0.05	splice_donor & intron_variants	loss	20/31	c.3671 + 1G > A	-	CRC = 1	Stomach & Thyroid = 1
	chr22	COSM6654391	0.2	Framshift		3/31	c.832delA	p.Thr278fs	CRC = 3	Large intestine = 1
	chr22	COSM6566095*	0.12	Missense		4/31	c.1058G > A	p.Arg353His	CRC = 1	Breast = 1
FBXW7	chr4	COSM1427622	0.5	Framshift	loss	12/12	c.2001delG	p.Ser668fs	CRC = 1	Large intestine = 16
	chr4	COSM1052102	0.42	stop_gained		16/24	c.4900A > G	p.Ser1634Gly	CRC = 1	Large intestine = 14
ARIDA1	chr2	COSM51427	0.33	Framshift	Loss	12/20	c.3216delA	p.Lys1072fs	CRC = 1	Large intestine = 5
	chr2	COSM1639819	0.18	Framshift		18/20	c.4555delC	p.Gln1519fs	CRC = 1	Stomach & Large intestine = 5
	chr2	COSM133001	0.14	Framshift		20/20	c.5548delG	p.Asp1850fs	CRC = 2	Large intestine = 32
	chr2	COSM133004	0.32	Framshift		20/20	c.5548dupG	p.Asp1850fs	CRC = 3	Stomach = 17 & Large intestine = 10
ACVR2A	chr2	COSM1399822	0.27	Missense	Loss	2/11	c.217A > G	p.Ile73Val	CRC = 1	Large intestine = 1
SMAD4	chr18	COSM1389042	0.11	Framshift	loss	6/12	c.692delG	p.Gly231fs	CRC = 1	Large intestine = 2
	chr18	COSM14232	0.10	Missense		9/12	c.1064A > G	p.Asp355Gly	CRC = 1	Large intestine = 5
	chr18	COSM14140	0.56	Missense		9/12	c.1081C > T	p.Arg361Cys	CRC = 2	Large intestine = 60
	chr18	COSM1152201	0.31	Missense		9/12	c.1088G > A	p.Cys363Tyr	CRC = 1	Large intestine = 3

Rank indicates affected exon/ total exons of the specified gene.

*Somatic mutations found in other tissues rather than the large intestine

Moreover, a single missense mutation was found in each of the following genes; KRAS **(c.35G > T)**, NRAS **(c.35G > T)** and PIK3CA **(c.3140A > G)**. Also, two missense mutations with identified gain of function were detected in ERBB2 **(c.922G > A & c.2690G > A)** and BRAF **(c.1781A > G & c.1799T > A)** (Table 6**)**.

Table 6
Specific somatic mutations related to anti-EGFR treatment response detected in the CRC group

Gene_ Name	CHROM	ID	VMF	Annotation	IVA	Rank	HGVS.c	HGVS.p	Occurrence in the current study	Occurrence in cosmic database
BRAF	chr7	COSM467**	0.19	Missense	Gain	15/18	c.1781A > G	p.Asp594Gly	CRC = 1	Large intestine = 1
	chr7	COSM476**	0.25	Missense	Gain	15/18	c.1799T > A	p.Val600Glu	CRC = 2	Thyroid = 14000 & Large intestine = 5000
ERBB2	chr17	COSM6080928	0.03	Missense	Gain	8/27	c.922G > A	p.Val308Met	CRC = 1	Large intestine = 1
	chr17	COSM978678	0.01	Missense		22/27	c.2690G > A	p.Arg897Gln	CRC = 1	Large intestine = 2
KRAS	chr12	COSM520**	0.25	Missense	Gain	2/6	c.35G > T	p.Gly12Val	CRC = 5	Large intestine = 5000
NRAS	chr1	COSM566	0.18	Missense	Gain	2/7	c.35G > T	p.Gly12Val	CRC = 1	Large intestine = 15 & Lymphoid = 100
PIK3CA	chr3	COSM775**	0.15	Missense	Gain	21/21	c.3140A > G	p.His1047Arg	CRC = 1	Large intestine = 400 & Breast = 100
Rank indicates affected exon/ total exons of the specified gene.										
**Somatic mutations found previously in the Egyptian CRC patients.										

The most commonly altered pathways in the CRC patients

The Ingenuity Variant Analysis (IVA) revealed that the following pathways were commonly altered in the CRC group; Wnt/βcatenin pathway was up-regulated in **73%**, ATM signaling pathway was down regulated in **72%**, RTK-RAS pathway was up regulated in **40%**, and TGF- β pathway was down regulated in **36%**. The most commonly altered pathways and their involved genes were illustrated in Fig. 9a. The proportion of the cancer driver genes and the most altered pathways in the CRC patients is illustrated in Fig. 9b. The proposed Egyptian model of the most commonly altered pathways in the Egyptian CRC patients, using the IPA software, was illustrated in Fig. 10.

Discussion

CRC is one of the leading causes of mortality and morbidity world-wide [1]. To the best of our knowledge, our study is the first to sequence multiple-gene panel to identify the somatic mutation pattern associated with colon cancer disease progression in a cohort of Egyptian patients to help understanding colorectal carcinogenesis process.

In the current study, the somatic mutational burden was higher in the CRC patients when compared to the other groups. The TP53, APC and ATM genes were the most frequently mutated genes in the CRC group. Matching with Cancer Genome Atlas Network, the TP53 and the APC were the most frequently mutated two genes in the CRC patients [15]. So, this finding validates the reliability of our sequencing results.

As for the TP53 which defined as the 'guardian of the genome'; its alteration is one of the tumor hallmarks and its mutational status is associated with the progression and outcome of sporadic CRC [16]. The TP53 mutation prevalence rate in Arab population is **52.5%** in comparison with **47.5%** in matched Western population [17].

Our study showed that the TP53 was the 1st rank and highly mutated gene that has been detected in **73%** of the CRC patients; indicating its role in the transition from an adenoma to carcinoma [18].

Eleven mutations out **13** were pathogenic with identified loss of function and have been detected only in the CRC patients. Interestingly, the most affected exon was **exon4** as well as the most frequent TP53 mutations **(c.121delG (58%), c.215C > G (52%))** were located in **exon4.** So, sequencing of TP53 **exon4** could be used for CRC prediction in our population.

Matching with a recent study by **Kassem et al.** [19] on the Egyptian CRC patients, we found five TP53 somatic mutations; **c.1024C > T**, **c.844C > T**, **c.743G > A, c.524G > A** and **215C > G** in our CRC group which might reveal that such variants are Egyptian specific and explain their contribution in colon cancer disease progression as a driver mutation in tumor development.

Of interest, the TP53 drug response variant **(c.215C > G)** [20], was detected in more than half of our CRC group. Also, it has been recently observed in **17%** of the Egyptian breast cancer patients [21]. Thus, this variant might serve as an efficient predictive marker for chemotherapy response in the Egyptian cancer patients.

Our pathway analysis revealed that the **P53** signaling pathway was inactivated in **72%** of the CRC patient due to either loss of the TP53 wild type or oncogenic gain of the TP53 mutant. Several new therapies specifically target p53-mutant cells, while others correct the p53 mutations directly or restore the integrity of the **p53** pathway [22]. Therefore, the deficient p53 signaling pathway in our CRC group may arise as an attractive therapeutic target.

Mutation of the APC **gene**, a multi-functional tumor-suppressor gene, is an early event in the development of CRC and result in activation of Wnt/β-catenin signaling pathway [23]. Mutant APC, Axin2 along with AMER1 **(APC-recruitment protein)** disrupt the formation of β-catenin destruction complex leading to stabilization and accumulation of β-catenin protein which in turn induces overactivation Wnt/β-catenin signaling and promote the proliferation, invasion and metastasis of cancerous cells [24, 25].

We have found that the APC gene was the 2nd rank highly mutated gene in **69%** of the CRC patients. **Fifteen** mutations out of **17** were pathogenic somatic mutations with identified loss of function detected only in the CRC group. Interestingly, **exon 16** was the most affected exon and it was found to harbor **14** out of **15** detected mutations in the CRC group. Moreover, the most frequently detected mutation in the APC gene that found to be associated with disease progression **(c.3754delT (65%))** was also in exon 16. So, sequencing of **exon 16** could be used as genetic test assay for CRC diagnosis. Of interest, most of our identified somatic mutations were located in the β-catenin and downregualtion site.

According to COSMIC database, we detected two APC variants (**c.4588G > T (2%) and c.5288delA (2%))** that were previously reported to be found respectively in pancreatic cancer and hepatocellular carcinoma. Interestingly, this is the first study to report the presence of such variants in the CRC and to report their involvement in the colorectal carcinogenesis process. However, further studies are needed to validate our findings.

We also detected three AXIN2 mutations in **27%** of the CRC patients. Two out of them (**c.2347G > T & c.1975C > T)** were reported in **2%** and **5%** respectively in the CRC group only. It was previously reported by **Imielinski et al.** and **George et al.** that these two mutations are associated with small cell lung cancer [26, 27]. So, this study is the first to report their association with CRC.

In this study and according to the IPA software, we found that Wnt/βcatenin pathway was upregulated in **73%** of the CRC patients; revealing that Wnt/βcatenin pathway plays a major role in sporadic colorectal carcinogenesis and therefore it is an attractive target for therapeutic intervention [28].

Somatic mutations of the ATM gene, as a DNA repair gene, occur in many tumor types including colorectal cancer. In the colorectal cancer, loss of ATM protein expression is associated with worse prognosis [29]. So, we are in need for such targeted sequencing studies to help monitoring the prognosis in Egyptian CRC patients.

Our data revealed that the ATM has been mutated in **44%** of the CRC patients. We detected **9** somatic mutations in the ATM gene, 6 out of them were detected only in the CRC group. The identified ATM mutations were previously reported to be associated with CRC **(Zehir et al., 2017).** Also, we detected two other ATM polymorphisms (**c.9007A > G & c.8138G > A)**, however they were reported to be associated with NHL lymphoma [30].

Moreover, this study showed that the ATM signaling pathway was downregulated in **72%** of the CRC patients; revealing the critical role of the defective DNA repair mechanism in colorectal carcinogenesis process [31].

Nowadays, novel therapies have been developed to selectively target patients with ATM-deficient cancers. Those therapies induce synthetic lethality due to lacking efficient repair mechanism such as platinum drugs [32]. Thus, the ATM mutational status could be used to help in the clinical decision-making for those patients along with the development of specific targeted strategies [33]. Thus, it is important to conduct targeted sequencing studies on the Egyptian CRC patients to evaluate drug efficacy and treatment protocols.

Matching with two studies that reported the association of SMAD4 mutations with the CRC, we detected four somatic mutations **(c.692delG, c.1064A > G, c.1081C > T & c.1088G > A)** only in the CRC group [34]. The SMAD4 gene acts as an intracellular mediator of TGF-β superfamily signals. TGF-β/SMAD4 signaling maintains DNA damage response (DDR) and DNA damage repair [35]. In this study, the TGF beta pathway was downregulated in **36%** of the CRC patients. It was suggested that loss or downregulation of the SMAD4 promotes malignant progression via acquiring resistance to TGF- β superfamily growth inhibition [36]. Moreover, its loss shifts TGF-β signaling pathway to a tumor promoter instead of a tumor suppressor [37]. Our findings revealed that the SMAD4 mutations had prominent role in colorectal carcinogenesis. **Isaksson-Mettavainio et al.** reported that loss of the SMAD4 occurs in the CRC in frequencies ranging from 9 to 67% [38]. Moreover, the SMAD4 loss was also associated with worse clinical outcome and resistance to fluoropyrimidine-based chemotherapy [39]; implicating its use as a prognostic marker in the CRC patients [40]. Thus, we propose that the Egyptian CRC patients carrying SMAD4 mutations may not benefit from fluoropyrimidine-based treatment.

Functional loss of the putative tumor suppressor EP300 gene has been previously observed in gastric, breast, pancreatic, and colorectal cancers. Also, **Gayther et al** reported a great relevance of the EP300 loss in the colorectal carcinogenesis [41]. Our study found that the EP300 gene harbored 3 somatic mutations. Out of them, one frameshift mutation **(c.832delA)** was found to be associated with CRC [42], while the other ones, the splice donor & missense mutations **(c.1058G > A &c.3671 + 1G > A)**, were respectively detected in breast and gastric cancers [43, 44]. Moreover, **Huh et al.** reported that p300 overexpression was an indicator of good prognosis in the CRC patients [45]. Therefore, the identified somatic mutations in the EP300 gene might serve as predictor of bad prognosis in the Egyptian CRC patients.

One of the most frequently detected somatic mutations in the CRC is in the tumor suppressor FBXW7 gene. Loss of the FBXW7 was reported to promote epithelial-mesenchymal transition (EMT) and metastasis in the CRC cells [46]. The present study reported two somatic mutations with functional loss in FBXW7 gene (**c.2001delG & c.4900A > G**). These mutations were previously reported to be found mainly in the CRC patients [47, 34]. Moreover, a recent study reported the presence of an association between the FBXW7 mutations and resistance to anti-epidermal growth factor receptor (EGFR) immunotherapy treatment that commonly used to manage metastatic CRC [48]. So, identifying the mutational status of the FBXW7 gene in Egyptian CRC patients may serve as a good diagnostic biomarker to determine the appropriate individualized therapy [49].

The functional loss of the tumor suppressor ARIDA1 gene has been previously reported as a frequent event in the colorectal carcinogenesis [50]. In agreement with a previous study by **Erfani et al.** who reported a relatively high mutation rate of the ARID1A in the CRC ranging from 10% up to 40%, we also reported a

high mutation rate of the ARIDA1 around **29%** in our CRC group [51]. Matching with a previous study which showed that the ARIDA1 mutations were more likely to be frameshift or nonsense, we detected **5** framshift mutations in the ARIDA1 [50]. Additionally, four mutations out of them were detected only in the CRC group. Thus, our study showed the prominent role of the ARIDA1 gene in the colorectal carcinogenesis.

Matching with several previous studies [52–54], we detected many gain of function somatic mutations that have been previously reported to be commonly detected in the CRC; BRAF **(c.1781A > G, c.1799T > A)**, KRAS **(c.35G > T, 38G > A)**, NRAS **(c.35G > T)** &PIK3CA **(c.3140A > G)**, ERBB2 **(c.922G > A & c.2690G > A).** Additionally, other reports revealed that these mutations contribute to the acquired resistance to the anti-EGFR therapy [55, 56]. Thus, genetic testing of those genes provides beneficial information that help in the clinical management of the CRC patients.

Regarding the common somatic mutations detected in all the studied groups; we have found that nine out of twenty four somatic mutations were the most frequent **(c.1310delA in** ACVR2A**), (c.640delT in** APC**), (c.5557G > A in** ATM**), (c.677delG in** IGF2**), (c.1621A > C in** KIT**), (c.1173A > G in** PIK3CA**), (c.2071G > A in** RET**), & (c.121delG, c.215C > G in** TP53**).** The frequency of those mutations increased from the colitis to finally the CRC, so they could be used as predictors for disease progression. Most of the above identified somatic mutations were observed mostly with the CRC [47], except for **(c.640delT in** ATM**), (c.1621A > C in** KIT**), (c.2071G > A in** RET**) & (c.121delG in** TP53**)** which were observed respectively in breast [57], soft tissues [58] & head and neck cancers [59]. This is the first study to report their association with the colorectal carcinogenesis in Egyptian CRC patients.

Conclusion

Our data showed that the genetic makeup of the Egyptian CRC patients is different from other population. Also, the identified somatic mutations are crucial for understanding cancer predisposition and developing personalized therapies for the Egyptian colorectal cancer patients.

Recommendations

We recommend; 1) Patients follow up for response to treatment based on our genetic analysis, 2) Microsatellite instability sequencing analysis at different CRC stages, 3) Post- translational research at the protein level for our targeted genes.

Abbreviations

CRC
Colorectal cancer; **IBD**:Inflammatory bowel disease; **CP**:Colonic polyp; **IVA**:Ingenuity Variant Analysis; **IPA**:Ingenuity Pathway Analysis; **NGS**:Next generation sequencing; **UMIs**:Unique molecular indices; **UBAM**:Unaligned bam; **QC**:Quality Control; **COSMIC**:Catalogue of Somatic Mutations in Cancer; **MAF**:Minimal allele frequency; **SNVs**:Single nucleotide variants; **EGFR**:Epidermal growth factor receptor; **EMT**:Epithelial-mesenchymal transition.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board **(IRB number: IRB00004025)** of the National Cancer Institute (NCI), Cairo University, Egypt. The study was conducted in accordance with the ICH-GCP guidelines **(approval number: 201617011.3).** A written informed consent was obtained from each patient enrolled in the study.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study and its supplementary information files are included in this published article

Competing interests

The authors declare that they have no competing interest.

Funding

This study was funded by the Science and Technology Development Fund **(STDF)**, Project ID: **5193.**

Author contributions

AZ and **AY** designed the study. **AT** and **MM** recruited patients and collected clinical data. **AY, ML** and **AN** performed the library preparation and NGS workflow. **MA** and **AM** performed the bioinformatics analysis. **AY** and **AZ** performed the data analysis and interpretation**.** drafted the manuscript. **AN**, **ZK,** and **AB** revised the manuscript. All authors read and approved the final version.

Acknowledgment

The authors thank the enrolled participants from the NCI and El Kasr Al-aini for the fresh tissues donated for research.

References

1. **International Agency for Research on Cancer (IARC). Available from: http://gco.iarc.fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet.pdf. (2018).**
2. **Metwally I H, Shetiwy M, Elalfy AF, et al.** Epidemiology and survival of colon cancer among Egyptians: a retrospective study. JCOL 2018; 38(1):24-29.
3. **Ibrahim A, Khaled H, Nabiel Mikhail K, et al.** Cancer Incidence in Egypt: Results of the National Population-Based Cancer Registry Program. J Cancer Epidemiol 2014; 2014:437971.
4. **Yamagishi H, Kuroda H, Imai Y and Hiraishi H.** Molecular pathogenesis of sporadic colorectal cancers. Chin J Cancer 2016; 6:35-39.
5. **Kim TM, Lee SH and Chung YJ.** Clinical applications of next-generation sequencing in colorectal cancers. World J Gastroenterol 2013; 19: 6784-6793.
6. **Moorcraft SY, Gonzalez D and Walker BA.** Understanding next generation sequencing in oncology: A guide for oncologist. Crit Rev Oncol Hematol 2015; 96:469-474.
7. **Han SW, Kim HP, Shin JY, et al.** Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing. PloS One 2013; 8: e64271.
8. **Keup C , Benyaa K , Hauch S, et al.** Targeted deep sequencing revealed variants in cell-free DNA of hormone receptor-positive metastatic breast cancer patients. Cell Mol Life Sci 2020; 77(3):497-509.
9. **Landrum MJ, Lee JM, Benson M, et al.** ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018; 46(D1):D1062-D1067.
10. **Vaser R, Adusumalli S, Leng SN, et al.** SIFT missense predictions for genomes. Nat Protocols 2016; 11: 1-9.
11. **Adzhubei IA, Schmidt S, Peshkin L, et al.** A method and server for predicting damaging missense mutations. Nat Methods 2010; 7: 248-9.
12. **Rentzsch P, Witten D, Cooper GM, et al.**
CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 2018; 47(D1):D886-D894.
13. **Mayakonda A, Lin DC, Assenov Y, et al.** Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018; 28(11):1747-1756.
14. **Zuguang Gu, Lei Gu, Roland Eils, et al.** Circlize implements and enhances circular visualization in R. Bioinformatics 2014;30(19):2811-2.
15. **Cancer Genome Atlas Network.** Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012 ;487(7407):330-7.
16. **Li N, Lorenzi F, Kalakouti E, et al.** FBXW7-mutated colorectal cancer cells exhibit aberrant expression of phosphorylated-p53 at serine-15. Oncotarget. 2015; 6(11):9240–56.
17. **Al-Shamsi HO, Jones J, Fahmawi Y, et al.** Molecular spectrum of KRAS, NRAS, BRAF, PIK3CA, TP53, and APC somatic gene mutations in Arab patients with colorectal cancer: Determination of frequency and distribution pattern. J Gastrointest Oncol 2016; 7: 882-902.
18. **Takayama T, Miyanishi K, Hayashi T, et al.** Colorectal cancer: Genetics of development and metastasis. J Gastroenterol. 2006; 41: 185-192.
19. **Kassem NM , Emera G, Kassem HA , et al.** Clinicopathological features of Egyptian colorectal cancer patients regarding somatic genetic mutations especially in KRAS gene and microsatellite instability status: a pilot study. Egypt. J Med Hum Genet. 2019; 20:20.
20. **Bergamaschi D, Gasco M, Hiller L, et al.** p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. Cancer Cell 2003 ;3(4):387-402.
21. **Nassar A, Abouelhoda M, Mansou O, et al.** Targeted next generation sequencing identifies somatic mutations in a cohort of Egyptian breast. JAR 2020; 24:149-157.
22. **Hamid O, Varterasian ML, Wadler S, et al.** Phase II trial of intravenous CI-1042 in patients with metastatic colorectal cancer. J Clin Oncol .2003; 21:1498.
23. **Powell SM, Zilz N, Beazer-Barclay Y, et al.** APC mutations occur early during colorectal tumorigenesis. Nature. 1992; 359: 235-237.
24. **Nguyen HT and Duong HQ.** The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy. Oncol Lett.2018; 16(1):9-18.
25. **Sanz-Pamplona R, Lopez-Doriga A, Paré-Brunet L, et al.** Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. Clin Cancer Res. 2015;21(20):4709-4718.
26. **Imielinski M, Berger AH, Hammerman PS, et al.** Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. 2012; 150(6):1107-20.
27. **George J, Lim JS, Jang SJ, et al.** Comprehensive genomic profiles of small cell lung cancer. 2015;524 (7563):47-53.
28. **Cheng X, Xu X, Chen D, et al.** Therapeutic potential of targeting the Wnt/β-catenin signaling pathway in colorectal cancer. Biomed Pharmacother.2019; 110: 473-481.
29. **Beggs AD, Domingo E, McGregor M, et al.** Loss of expression of the double strand break repair protein ATM is associated with worse prognosis in colorectal cancer and loss of Ku70 expression is associated with CIN. Oncotarget. 2012; 3: 1348–55.
30. **Gumy-Pause F, Wacker P, Maillet P, et al.**. ATM alterations in childhood non-Hodgkin lymphoma. Cancer Genet Cytogenet.2006;166(2):101-11.
31. **Macheret M and Halazonetis TD.** DNA replication stress as a hallmark of cancer. Annu Rev Pathol .2015;10:425–48.
32. **Choi M, Kipps T and Kurzrock R.** ATM Mutations in Cancer: Therapeutic Implications. Mol Cancer Ther.2016 ;15(8):1781-91.
33. **Randon G, Fucà G, Rossini D, et al.** Prognostic impact of ATM mutations in patients with metastatic colorectal cancer. Sci Rep.2019 ;9(1):2858.
34. **Zehir A, Benayed R, Shah RH, et al.** Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med.2017;23(6):703-713.
35. **Zhao M, Mishra L and Deng CX.** The role of TGF-β/SMAD4 signaling in cancer. Int J Biol Sci.2018;14(2):111-123.

36. **Levy L and Hill CS.** Smad4 dependency defines two classes of transforming growth factor {beta}(TGF-{beta}) target genes and distinguishes TGF-{beta}-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. Mol Cell Biol. 2005; 25: 8108–25.
37. **Zhang B, Halder SK, Kashikar ND, et al.** Antimetastatic role of Smad4 signaling in colorectal cancer. Gastroenterology. 2010; 138:969-980.
38. **Isaksson-Mettavainio M, Palmqvist R, Forssell J, et al.** SMAD4/DPC4 expression and prognosis in human colorectal cancer. Anticancer Res. 2006; 26:507–10.
39. **Zhang B, Zhang B, Chen X, et al.** Loss of Smad4 in colorectal cancer induces resistance to 5-fluorouracil through activating Akt pathway. Br J Cancer. 2014; 110:946-957.
40. **Wasserman I, Lee LH, Ogino S , et al.** SMAD4 Loss in Colorectal Cancer Patients Correlates with Recurrence, Loss of Immune Infiltrate and Chemoresistance. Clin Cancer Res.2019 ; 25(6):1948-1956.
41. **Gayther SA, Batley SJ, Linger L, et al.** Mutations truncating the EP300 acetylase in human cancers. Nat Genet.2000; 24: 300-303.
42. **Giannakis M, Mu XJ, Shukla SA, et al.** Genomic correlates of immune- cell infiltrates in colorectal carcinoma. Cell Rep. 2016; 15:857–865.
43. **Lefebvre C, Bachelot T, Filleron T, et al.** Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. PLoS Med.2016;13(12):e1002201.
44. **Wang K, Yuen ST, Xu J, et al.** Whole-genome sequencingand comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014; 46(6):573-82.
45. **Huh JW, Kim HC, Kim SH, et al.** Prognostic impact of p300 expression in patients with colorectal cancer. J Surg Oncol. 2013; 108: 374-377.
46. **Wang Y, Liu Y, Lu J, et al.** Rapamycin inhibits FBXW7 loss-induced epithelial-mesenchymal transition and cancer stem cell-like characteristics in colorectal cancer cells. Biochem Biophys Res Commun. 2013; 434: 352–356.
47. **Oliveira DM, Laudanna C, Migliozzi S, et al.** Identification of different mutational profiles in cancers arising in specific colon segments by next generation sequencing. 2018; 9(35):23960-23974.
48. **Lupini L, Bassi C, Mlcochova J, et al.** Prediction of response to anti-EGFR antibody-based therapies by multigene sequencing in colorectal cancer patients. BMC Cancer 2015;15 :808.
49. **Cao J, Ge MH and Ling ZQ.** Fbxw7 Tumor Suppressor: A Vital Regulator Contributes to Human Tumorigenesis. Medicine. 2016; 95: e2496.
50. **Sarshekeh AM, Loree JM, Manyam, GC, et al.**The characteristics of ARID1A mutations in colorectal cancer.2018; 36(15): 3595-3595.
51. **Erfani M, Hosseini SV, Mokhtari M, et al**. Altered ARID1A expression in colorectal cancer.BMC Cancer. 2020;20 (1):350.
52. **Jin J, Shi Y, Zhang Sand Yang S.***PIK3CA*mutation and clinicopathological features of colorectal cancer: a systematic review and Meta-Analysis. Acta 2020; 59(1):66-74.
53. **Guo F, Gong H, Zhao H, et al.** Mutationstatus and prognostic values of KRAS, NRAS, BRAF and PIK3CA in 353 Chinese colorectal cancer patients. Sci Rep. 2018; 8(1):6076.
54. **Loree JM, Bailey AM, Johnson AM, et al.** Molecular Landscape of ERBB2/ERBB3 Mutated Colorectal Cancer. J Natl Cancer2018 ; 110(12):1409-1417.
55. **Ross JS, Fakih M, Ali SM, et al.** Targeting HER2 in colorectal cancer: The landscape of amplification and short variant mutations in ERBB2 and ERBB3. Cancer.2018 ; 124(7):1358-1373.
56. **Xu JM, Wang Y, Wang YL, et al.***PIK3CA*Mutations Contribute to Acquired Cetuximab Resistance in Patients with Metastatic Colorectal Cancer. Clin Cancer 2017; 23(16):4602-4616.
57. **De Luca F, Rotunno G, Salvianti F, et al.** Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer. 2016 ; 7(18):26107-19.
58. **Osoegawa A, Hashimoto T, Takumi Y, et al.** Acquired resistance to an epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) in an uncommon G719S EGFR mutation. Invest New Drugs.2018 ; 36(6):999-1005.
59. **Cabelguenne A, Blons H, de Waziers I, et al.** p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma: a prospective series. J Clin Oncol.2000; 18(7):1465-73.
60. **Lawrence MS, Stojanov P, Polak P, et al.** Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499: 214-8.

Figures

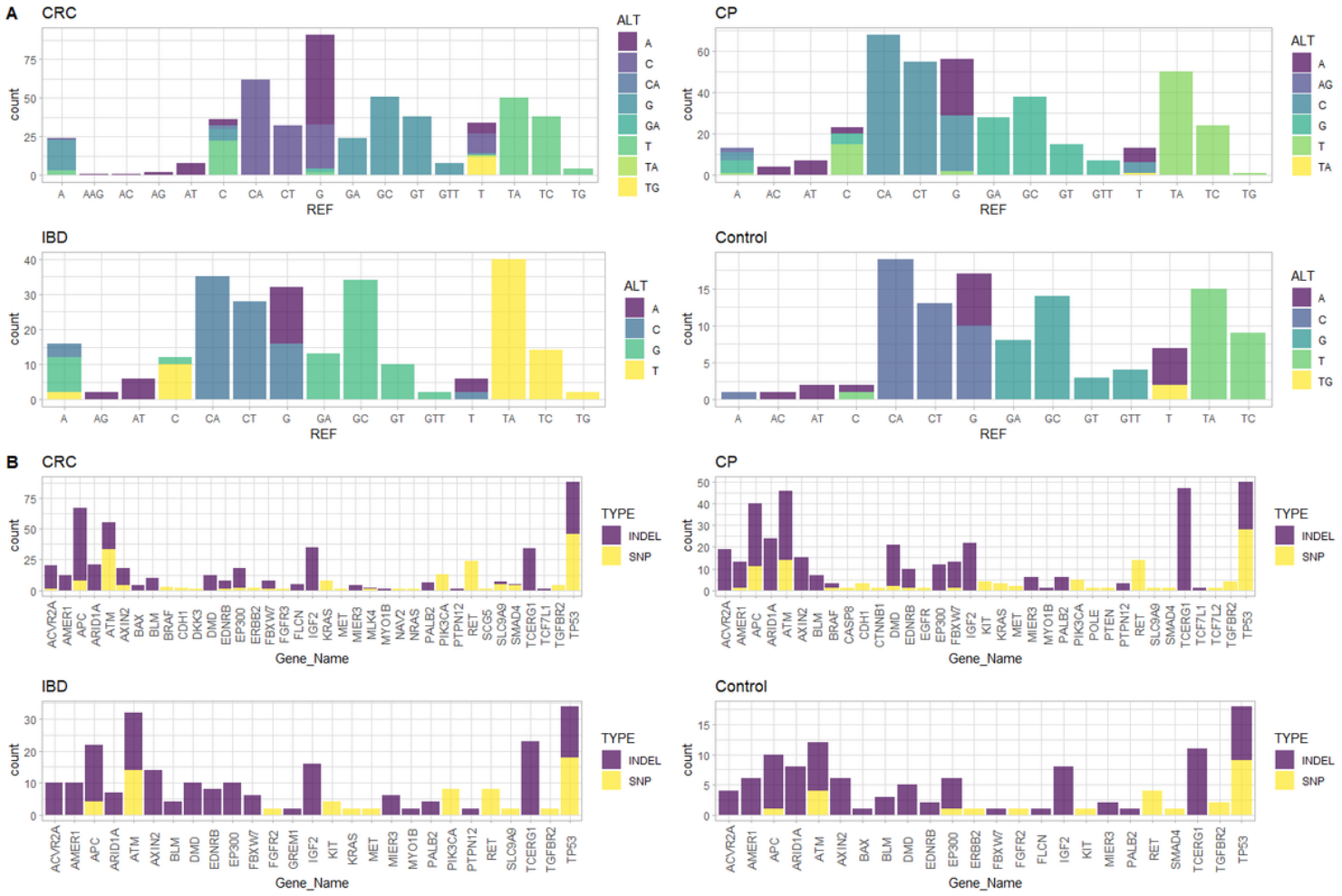

Figure 1

Bar charts showing a) the changes of the reference alleles to the alternative ones in each studied group, b) the highly mutated genes with classification in each studied group.

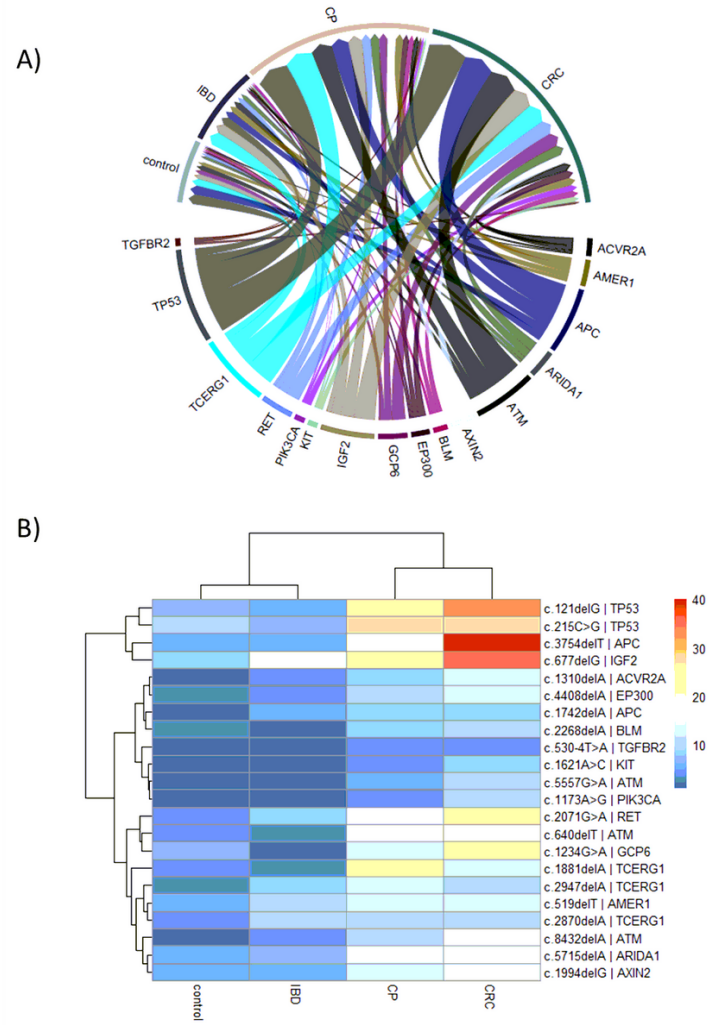

Figure 2

A) Chord diagram; the upper half circle represented the studied groups, whereas the lower half circle represented the common genes & B) Heatmap showing the frequency of the common mutations detected in all groups. The x-axis represented the studied groups, while the y-axis represented the common genes with its variants.

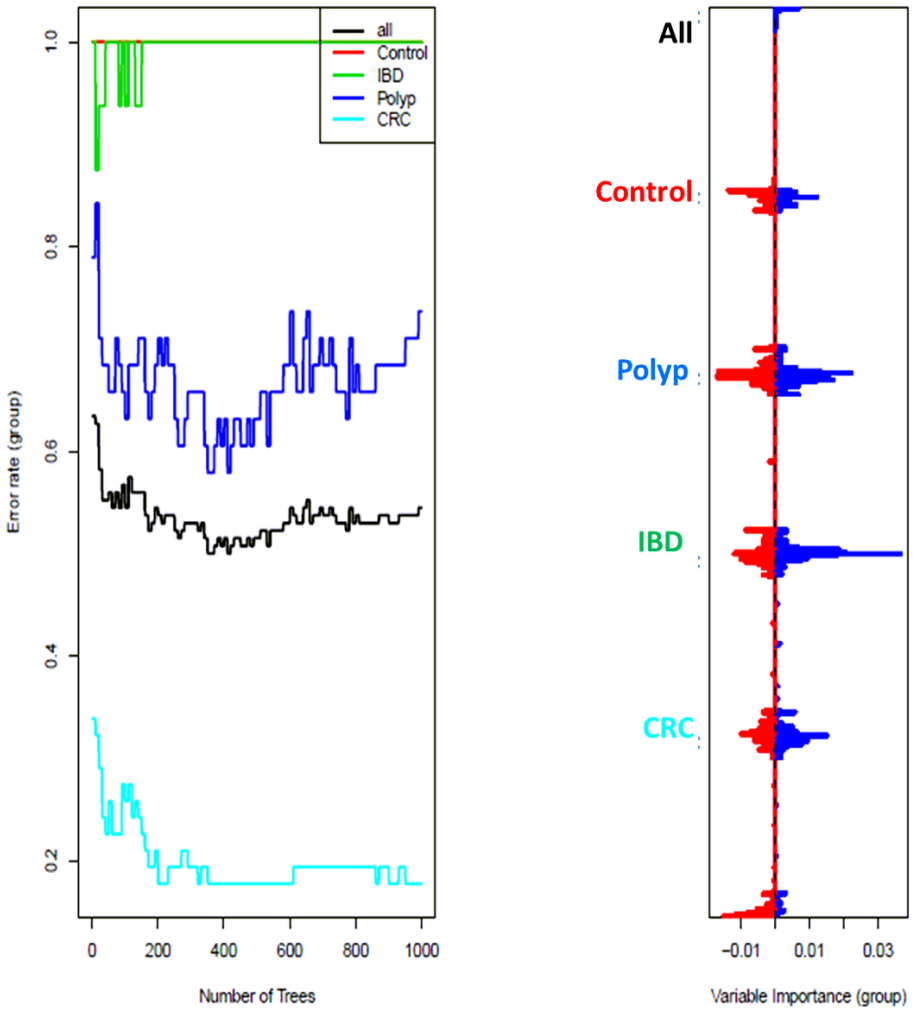

Figure 3

The random forest plot of the detected non-synonymous variants showing the diagnostic error rate of each model in the studied groups. The CRC model has the lowest error rate (0.2) compared to the other groups.

Figure 4

Oncoplot displayed the somatic mutations distribution of the top highly mutated genes in the CRC group. Each column represented a sample; it can be classified according to sex into male (blue color) or female (red color) and according to pathology as showed above. Each raw represented a particular gene with different variant classification. The plot showed a total positive for 57 (91.9%) out of 62 CRC samples.

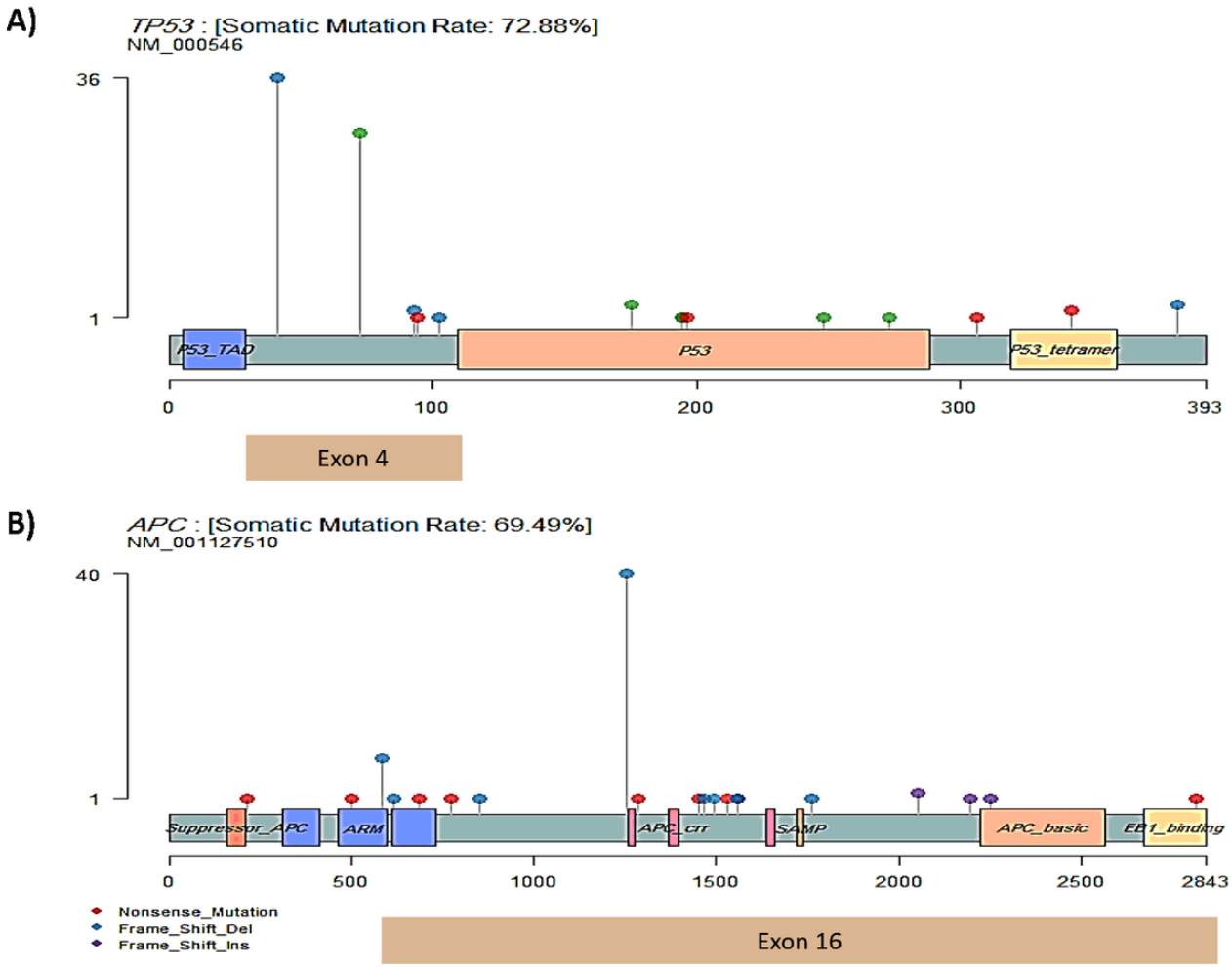

Figure 5

Lollipop representation displayed the detected somatic mutations in (a) the TP53 gene indicating a highly frequent mutations in exon 4 , (b) the APC gene indicating a highly frequent mutations in exon 16.

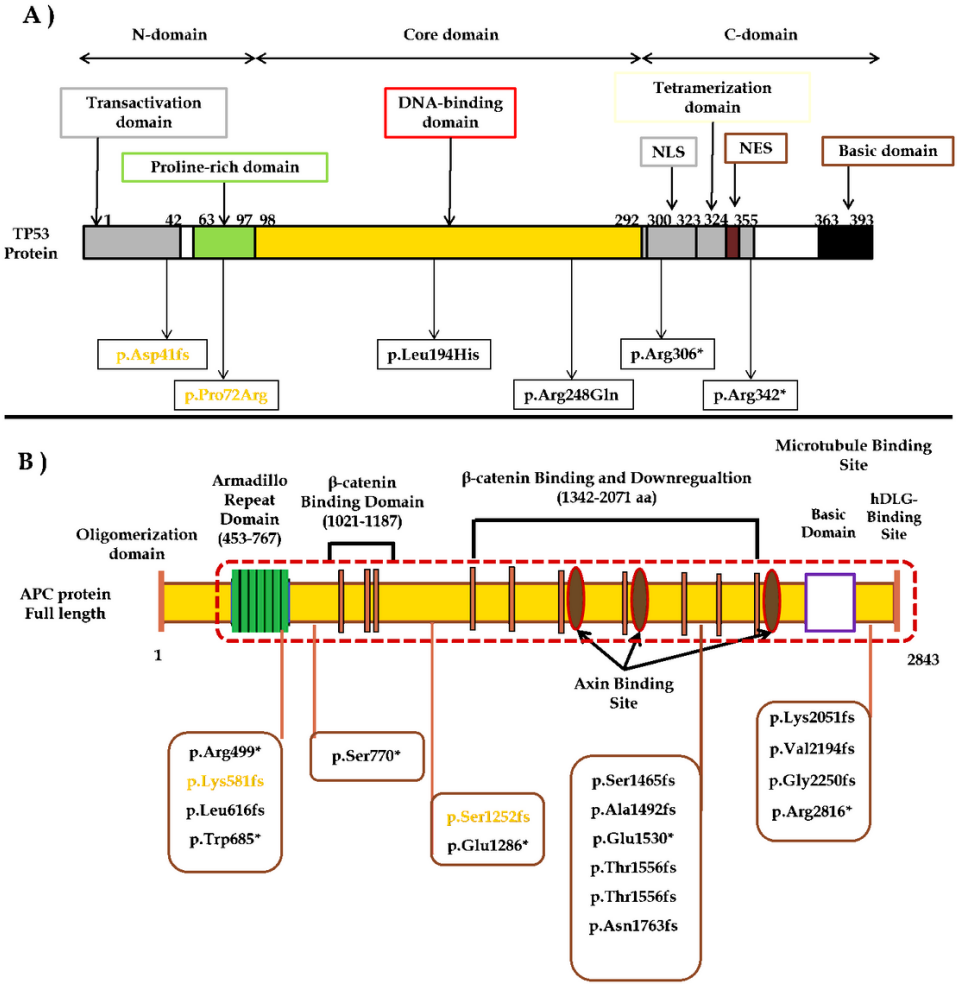

Figure 6

Schematic representation of detected somatic mutations in (a) the TP53 protein showed highly frequent mutations at both transactivation and prolin rich region, (b) the APC protein showed highly frequent mutations at β-catenin binding and down regulation site. The mutations with yellow color indicated common mutations associated with disease progression.

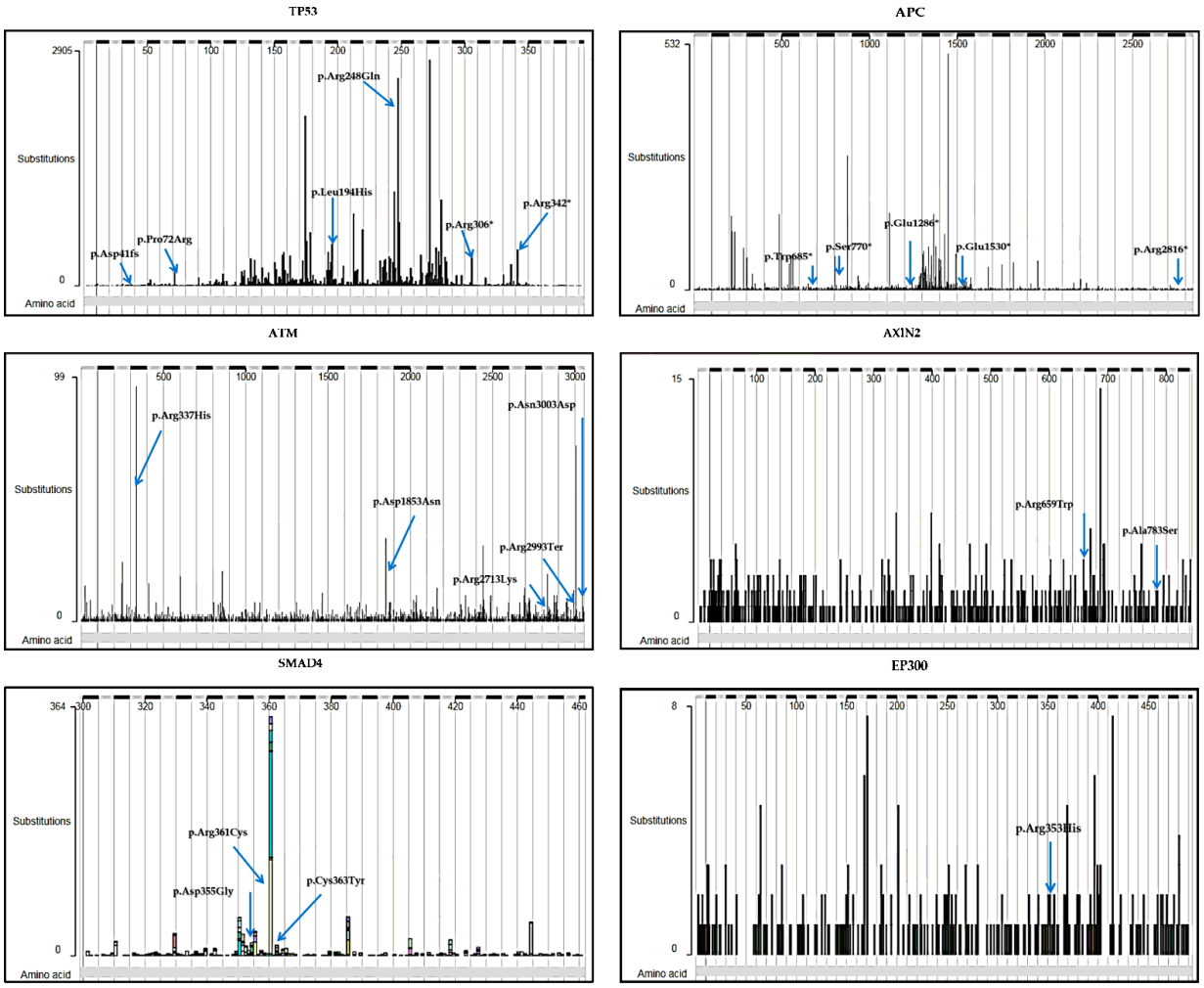

Figure 7

Loss-of-function cancer driver mutations in the CRC group; putative driver missense mutations in tumor suppressor hotspots. The number of substitutions catalogued in the COSMIC database was shown on the y axis at each position along the gene, with the mutations that were observed in the Egyptian CRC patients indicated with blue arrows.

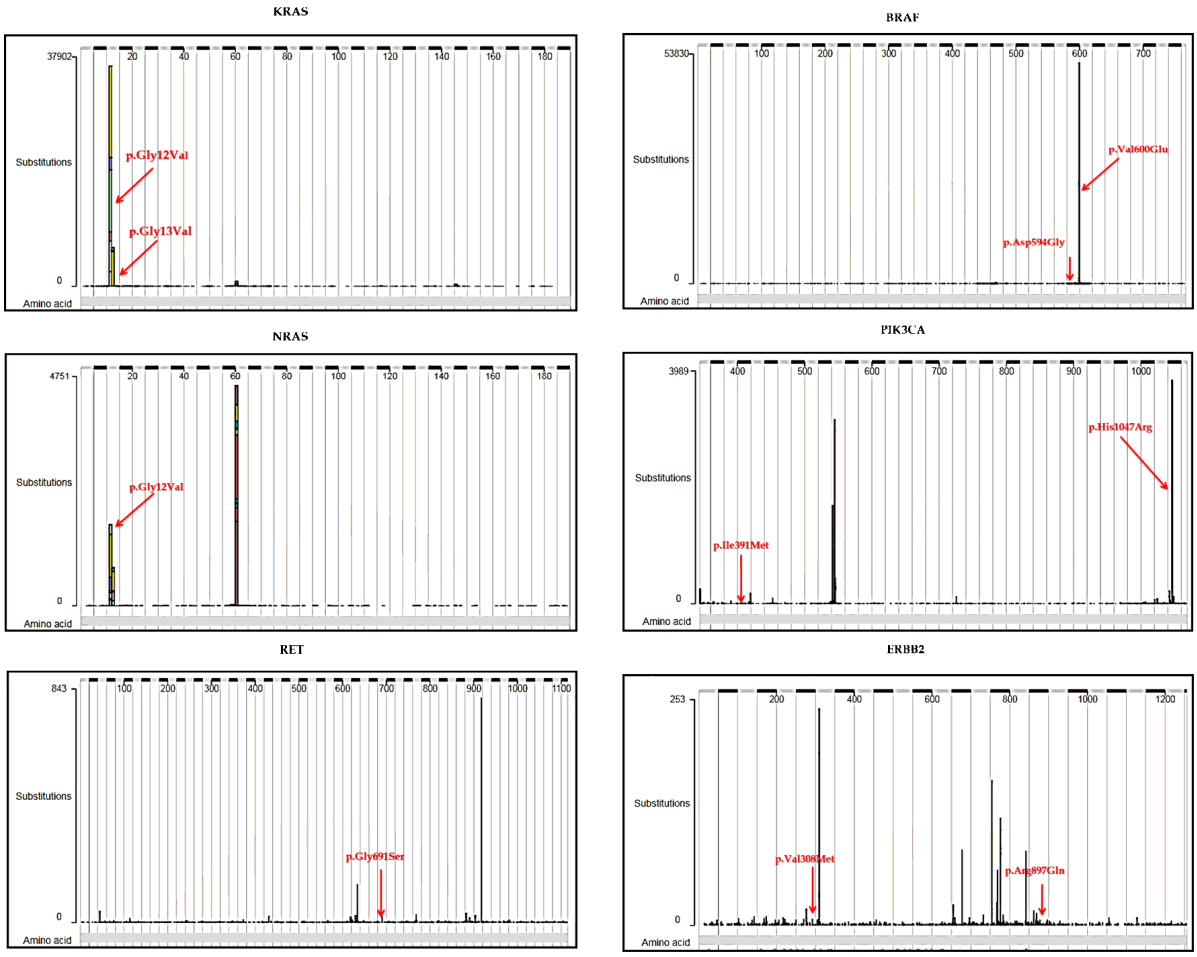

Figure 8

Gain-of-function cancer driver mutations in the CRC group; putative driver missense mutations in oncogene hotspots. The number of substitutions catalogued in the COSMIC database was shown on the y axis at each position along the gene, with the mutations that were observed in the Egyptian CRC patients indicated with red arrows.

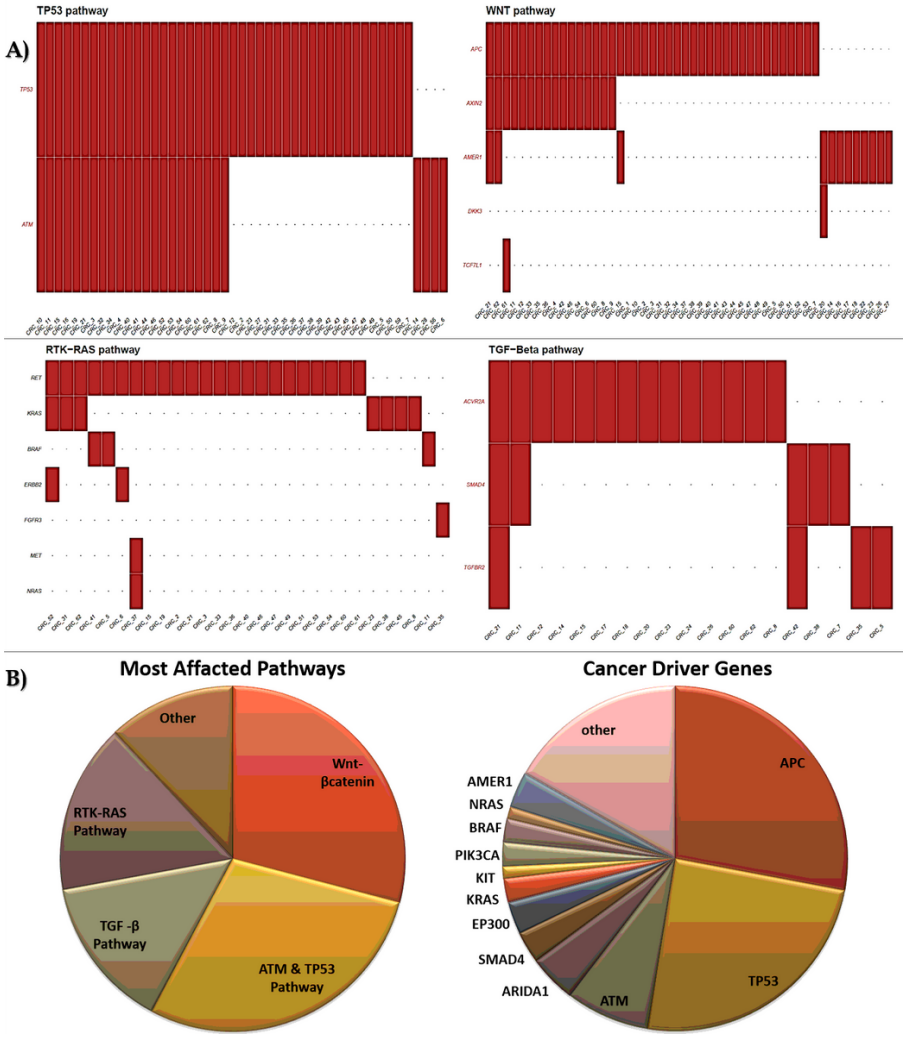

Figure 9

A) Pathways analysis displayed the somatic mutations distribution of the genes involved in the Wnt/βcatenin, P53, RTK-RAS, TGF-β pathways, B) Pie charts displayed the proportion of the cancer driver genes and the most altered pathways in CRC group.

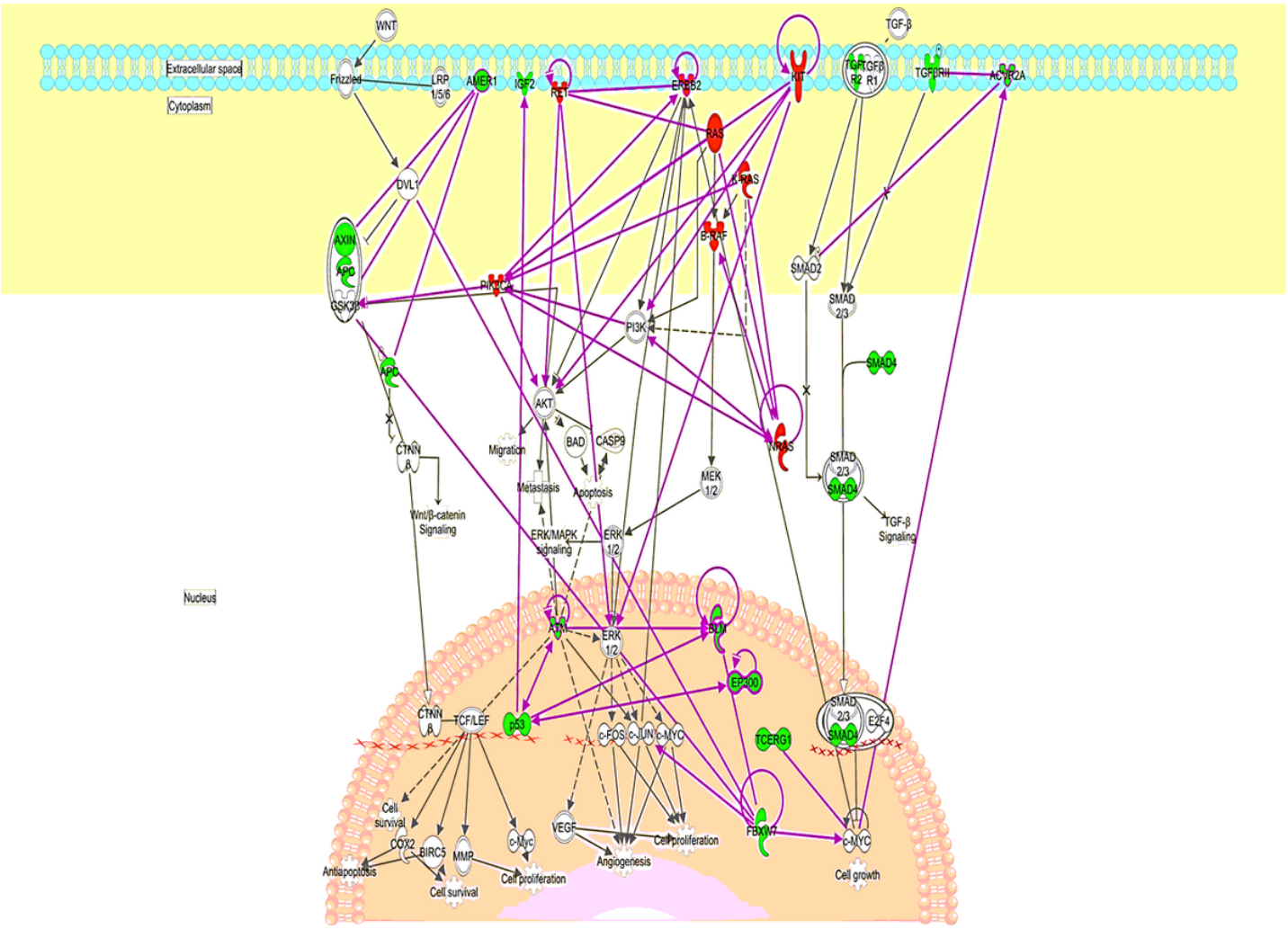

Figure 10

Proposed Egyptian model of the most commonly altered pathways in the Egyptian CRC patients due to somatic mutations. This model showed the interplay between the Wnt/βcatenin, RTK-RAS, TGF-β and P53 signaling pathways. The green color indicated loss of function, while the red one indicated gain of function. (www.ingenuity.com/products/pathway-analysis).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif
- FigureS2.tif
- TableS1.docx