



Kent Academic Repository

Hussian Mangi, Fayaz, Ahmed Shaikh, Tanweer, Soria, Daniele, Muhammad Waryah, Ali, Din Ujjan, Ikram, Naeem Qureshi, Jawaid and Manzoor Syed, Binafsha (2022) *Novel Molecular classification of colorectal cancer and correlation with survival*. Saudi Journal of Biological Sciences . ISSN 1319-562X.

Downloaded from

<https://kar.kent.ac.uk/93613/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.sjbs.2022.03.014>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Journal Pre-proofs

Original article

Novel Molecular classification of colorectal cancer and correlation with survival

Fayaz Hussian Mangi, Tanweer Ahmed Shaikh, Daniele Soria, Ali Muhammad Waryah, Ikram Din Ujjan, Jawaid Naeem Qureshi, Binafsha Manzoor Syed

PII: S1319-562X(22)00160-7
DOI: <https://doi.org/10.1016/j.sjbs.2022.03.014>
Reference: SJBS 3264

To appear in: *Saudi Journal of Biological Sciences*

Received Date: 13 September 2021
Revised Date: 5 March 2022
Accepted Date: 7 March 2022

Please cite this article as: F. Hussian Mangi, T. Ahmed Shaikh, D. Soria, A. Muhammad Waryah, I. Din Ujjan, J. Naeem Qureshi, B. Manzoor Syed, Novel Molecular classification of colorectal cancer and correlation with survival, *Saudi Journal of Biological Sciences* (2022), doi: <https://doi.org/10.1016/j.sjbs.2022.03.014>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier B.V. on behalf of King Saud University.



Novel Molecular classification of colorectal cancer and correlation with survival

Fayaz Hussian Mangi^{1&2}, Tanweer Ahmed Shaikh^{1&2}, Daniele Soria³, Ali Muhammad Waryah⁴, Ikram Din Ujjan², Jawaid Naeem Qureshi⁴, Binafsha Manzoor Syed¹

¹Departments of Clinical Research Medical Research Centre, and ²Pathology, Liaquat University of Medical & Health Sciences, Jamshoro, Pakistan ³School of Computing, University of Kent, Canterbury, CT2 7NF, United Kingdom, ⁴ Department of Molecular Biology and Genetics, Medical Research Centre, Liaquat University of Medical & Health Sciences, Jamshoro, Pakistan, ⁵Departments of Surgery, Indus Medical College, Tando Mohammad Khan, Pakistan

Corresponding author

Dr. Binafsha Manzoor Syed, PhD
Principal Investigator NRPU 3883
Associate Professor, Clinical Research,
Medical Research Centre,
Liaquat University of Medical & Health Sciences, Jamshoro
Tele: 0092 333 2749201
Email: binafsha.syed@lumhs.edu.pk

Conflict of Interest

Authors declare no conflict of interest for the submitted manuscript title “**Novel Molecular classification of colorectal cancer and correlation with clinical outcome**”

Abstract

Introduction

Colorectal cancer (CRC) is one of the most common cancers worldwide. This study was designed to evaluate biological patterns, explore molecular classification and correlate with survival outcome in treatment naïve CRC patients.

Methods

Over 11 years consecutive series of 435 CRC patients were operated on as primary surgical therapy. A total of 201 CRC patients were included, whose complete set of clinical information was available, and their good quality tumour blocks were retrieved. Immunohistochemistry was used for tumour analysis, and partitional clustering was performed using R software for cluster analysis.

Results

The median age was 43 (range 10-85) years; adenocarcinoma was the most commonly seen histological type. The great majority had positive CK20, CEA, E-Cadherin, Ki67, CDX2, and p53 expression. There were four distinct molecular classes found, whereas Ki67, CDX2, and p53 play

the main role in partitioning. Younger age negatively impacted survival; overall and disease-specific survival was 26 months only with 50 months' longest survival.

Conclusion

Colorectal cancer is a biologically heterogeneous disease with at least four distinct molecular patterns, where cell proliferation and gene repair mechanisms appear to play the key role.

1. Introduction:

Colorectal cancer (CRC) is a heterogeneous disease, the second most common cancer in women and the third most common in men worldwide(1). According to Globocan 2020, it is the 4th top cancer reported around the world and 7th most common cancer in Pakistan with a rate of 5.3/100,000 population. Globocan stated to have ~14000 new cases of colorectal cancer in 2020 globally, and further rise is expected in the upcoming years where an absolute increase of 79% was predicted by 2040. However, with a geographical variation highest rise was predicted in Africa at 95%, followed by Latin America at 74%, Asia at 71%, Oceania at 57%, North America at 35%, Europe at 25% rise in the new cases by 2040. A similar rise in mortality is suspected, reported to be 915880 in 2020(1).

Given the heterogeneous nature of CRC, at least three different pathogenesis pathways have been reported to be involved(2). Each pathway involves several markers, and these markers can be used as potential therapeutic targets to improve clinical outcome. Major prognostic and predictive factors routinely used include the clinical stage of the disease with the uncertain role of other potential factors. Molecular classification of colorectal cancer based on gene mutation (i.e.,

KRAS, BRAF) has reported at least five genetically distinct molecular classes(3). KRAS mutant tumours have been reported not to respond to anti-EGFR therapy; thus, it can be suspected that CRC has low EGFR expression, while other classes were suggested to be related to DNA instability. Therefore studying gene repair mechanisms would be an essential aspect to be looked at, which can be represented by p53 protein in cancer cells(3).

Similarly Carcinoembryonic antigen (CEA) has shown positive expression in a great majority of CRC patients, and there has been associated metastatic potential(4). Based on the heterogeneity of the disease, different patterns of molecular classes were reported; thus CRC subtyping consortium finally developed a consensus molecular classification with four consensus classes. However, their translational aspect is still far from being used in clinical practice(4).

The lesson learned from the molecular signature of breast cancer based on genetic signature and validated by immunohistochemistry showed a potential to bring predictive and prognostic factors in clinical practice with promising results in improving clinical outcome. Also, the combination of the biomarkers in the molecular class suggests more about pathogenesis and potentially predicting response to therapy than a single marker alone. Therefore, this study was conducted, including treatment naïve CRC patients who underwent primary surgery, and their tumours were analyzed using Immunohistochemistry (IHC). Given their role in colorectal cancer development, progression, or prognosis, the biological markers were chosen. A list of potential biomarkers was identified and finally included Oestrogen receptor(5,6), Progesterone receptor (5,6), Human Epidermal Growth Factor Receptor- 2 (HER2) (7), Cell proliferation marker (Ki67)(8), B-Cell Lymphoma -2 (Bcl2)(9), E-Cadherin(10), p53(11), CEA(11), Epidermal Growth Factor Receptor (EGFR)(12), Vascular Endothelial Growth Factor (VEGF)(13), Programmed Death Ligand-1 (PD-L1)(14), Caudal Type Homeobox-2 (CDX2)(15) and Cytokeratin 20 (CK20)(16) were selected given their IHC protocol and their role in prognosis and progression of colorectal cancer. The aims were to evaluate:

1. The pattern of biomarkers, i.e.,ER, PR, HER2, Ki67, Bcl2, E-Cadherin, p53, CEA, EGFR, VEGF, PD-L1, CDX2 and CK20 in CRC by using IHC.
2. Partitional Clustering of CRC by using K-means and Partitioning Around Medoids (PAM) methods
3. Correlating biological characteristics with survival outcome

2. Patients and Methods

2.1. Patients

Over 11 years (i.e., 2008 to 2018), a total of 435 consecutive patients with colorectal cancer were diagnosed and treated at a single center at Liaquat University of Medical & Health Sciences (LUMHS), Jamshoro, Pakistan, and their clinical information available from diagnosis until death/last follow-up at NIMRA cancer hospital, LUMHS, Jamshoro. These patients received primary surgical therapy without neo-adjuvant systemic or radiotherapy (n=201). Those who received neoadjuvant chemotherapy or radiotherapy or incomplete clinical information and missing tumour blocks were excluded (n=234). After their diagnosis, they received treatment as per hospital policy following International guidelines. The patients with good quality tumor samples available and a complete set of clinical information and follow-up were included in this study. All the patients were treatment naïve, including those who underwent emergency surgery due to obstruction without any prior diagnosis and those who were operated on after diagnosis and staging but without receiving any neo-adjuvant (both systemic or radiotherapy) therapy.

2.2. Tumour analysis

Formalin-Fixed Paraffin-embedded (FFPE) tumor blocks were retrieved from the cancer tissue archive. Hematoxylin and Eosin (H& E) staining was done to identify the most representative tumor block. Whole tumour block were used to analyze biomarkers, including ER, PR, HER 2-neu, Ki-67, Bcl-2, E-Cadherin, P53, CEA, EGFR, and VEGF. PDL1, CDX-2, and CK 20 by using indirect IHC. For IHC one, cut section of 3 to 5 μm thickness was used for each block. The preparation process was done by using PT-Link while primary antibodies were ready to use (RTU) by DAKO. Incubation time, dilution, and temperature are summarized in Supplementary Table 1. Envision Flex High pH (Link) secondary antibody was used. Finally treated with chromogen DAB and counterstained with Haematoxylin.

2.3. **Scoring:** Immunohistochemistry staining of biomarkers assessed by the percentage of cells stained, as well as McCarty's immunohistochemical scoring (H-score) was done (range 0–300)(17). The cutoff of the percentage of cells was used to define positivity/negativity. The scoring was done by FM, each section was scored three times, and an average of the scores was taken as the biomarker's final score. For inter-observer concordance, 25% of slides were randomly scored

by BMS. Kappa statistics was performed for all markers for intra and inter- observer concordance. The Kappa score was 0.9-1.0, 0.8-1.0, respectively. The scoring was done using a Euromax simple microscope at 40x magnification size.

2.4. **Cluster analysis:** The biological patterns were characterized by partitional clustering methods as described in(18), using R, a data analysis software. The H-score of the biomarkers was used for cluster analysis. K-means and PAM clustering algorithms were run over the data, varying the number of clusters between 2 and 20. Validity indices (external validation criteria) were used to suggest the best number of clusters to consider. When running K-means between 2 to 20 clusters, the algorithm stopped after the split in 4 groups, and clusters with 0 elements were returned. This suggests that a split in more than four groups may not be ideal.

2.5. **Statistical analysis:** The X-tile Bio-informatics software was used to define cutoffs(19). The Statistical Package for Social Sciences (SPSS, version 21.0, Chicago) was used for data collection and analysis. Chi-square test used for comparisons of biomarker expression between groups. Survival analysis was performed using Kaplan–Meier methods with the application of log-rank and generalized Wilcoxon tests as appropriate. A p-value of <0.05 was considered significant. Overall survival was calculated as the time from date of diagnosis till death from any cause. In contrast, disease-specific survival was calculated as the time from the date of diagnosis till death from metastases due to colorectal cancer.

3. Results:

A total of 201 patients with histopathologically confirmed colorectal cancer were included in this study. The median age of the patients was 43 (range= 10 to 85) years. Out of which 110 (54.7%) were males, and 91 (45.3%) were females. All patients underwent primary surgery without any prior intervention. 33 (16%) had radiotherapy postoperatively, while 95(47%) had adjuvant chemotherapy. Adenocarcinoma was the most commonly seen cancer (61.5%), followed by mucinous variety and a small proportion of other subtypes. Most of the cancers were well differentiated. A summary is presented in Table 2. Colon was involved in 119 (59.5%) including cecum (n=6), ascending colon (n=16), hepatic flexure (n=15), transverse colon (n=8), splenic

flexure (n=6), descending colon (n=11), sigmoid colon (n=31), recto-sigmoid junction (n=26). Rectum was involved in 81 (40.5%) patients.

3.1. Biological Characteristics

CK20, CEA, and E-Cadherin were found positive in ~99% of cases, while Ki67 was positive in 82%, CDX2 in 85% of patients, and p53 was positive in 43%. However, ER (3%), PR (2.5%), HER2(4.5%), EGFR (8.5%), Bcl2 (2.5%), PDL1(14.4%) and VEGF (5.5%) were found positive in a small proportion of patients (Table 2). Age between 51 to 75 showed a differing pattern in males and females where a higher rate of male patients of CRC was seen (Figure 1a). With advancing age, poorly differentiated cancers appear to decline (Figure 1b). Grade showed significant association with p53 positivity (Figure 1c).

3.2. Molecular classification

Four distinct molecular classes were identified (Figure 2). The key differentiating markers were Ki67, CDX2 and p53. Cluster 4 was characterized by younger patients (mean age 37 years), while relatively older patients fell into cluster 2 (mean age 48 years). Clusters 3 and 4 were predominantly observed in the rectum and sigmoid region. Table 3 describes the characteristics of molecular classes.

3.3. Survival

The median overall survival was 26 months, and disease-specific survival was 26 months. The longest survival was seen at 50 months in one patient. Younger age (<50) was significantly associated with poor disease-specific survival, and Ki67 positive status showed borderline significance (p=0.06) with inferior disease-specific survival. Histological grade, nodal status, ER, PR, HER 2-neu, Bcl-2, E-Cadherin, P53, CEA, EGFR, VEGF, PDL1, CDX-2, and CK 20 did not show any association with survival (p-value >0.05). Molecular classes did not show any significant association with clinical outcome (Figure 3).

4. Discussion

The study presented a novel molecular classification of colorectal cancer where Ki67, CDX2, and p53 IHC expression plays a crucial role in partitioning. Overall survival was observed to be poor, and the same was observed with disease-specific survival. Our study showed that younger patients

tend to have more age preponderance in clusters such that the mean age for cluster 4 was 37 years while for cluster 2 was 48 years. That was an interesting finding linked with younger age as a poor prognostic factor. This was previously reported in a study that included two age groups according to the age of onset of CRC. The classification was done based on MSI and BRAF mutation. More younger patients tend to have more left colon cancers and there was a comparison made based on a mutation in CpG island Methylator Phenotype (CIMP)(20). More younger patients tend to have CIMP- high type tumours with a mean age of 36 years, while combined classes showed MSI/CIMP- high type with a mean of 29 years(20). However, the clinical correlation was not reported in this study.

Another reported classification included the tumour microenvironment, and 167 gene signature identified four distinct molecular classes(21). This model of CRC classification microsatellite instability, histological types, high stromal content, β -Catenin, and BRAF significantly influenced hierarchical clustering. MSS subtypes, BRAF, and KRAS mutations were associated with the worst survival(22).

A previously published IHC based molecular classification of CRC including four independent cohorts (including AMC-AJCC-II, LUMC, CAIRO, and CAIRO2) including CDX2, FRMD6, HTR2B, and ZEB1 by using IHC(23) and validated the presence of four distinct molecular classes where there were two basic classes were seen as epithelial-like and mesenchymal-like. Another model analyzed gene patterns mainly looking at RAS where four molecular classes were reported with distinct response to FOLFIRI(24). However, a previously reported study included paraffin-embedded tissue sections and analyzed genes, and classification of CRC was made on the expression pattern of KRAS, BRAF, and CIMP, which reported five molecular classes. There was a significant association of change of pattern of CRC molecular classes with advancing age, gender, family history of CRC, and the tumour site (25). Similar findings were reported when only stage 2 CRC was evaluated using consensus molecular classification(26). The molecular classes showed a significant influence of age, gender, site of the tumour, and stage of the disease.

P53 is a tumour suppressor gene located on chromosome 17 (27). Its mutations have been linked with many cancers, including CRC. In our study, p53 appears to be a key governing factor in classifying molecular patterns. P53 mutation was previously associated with depressed neoplasms (DNs) more than other phenotypes(28). Similarly, p53 was also significantly associated with the

worst prognosis(29). A study that used Polymerase Chain Reaction (PCR) for molecular classification of CRC showed that p53 mutation was associated with subclass 2 out of five subclasses(30). P53 mutation was present in 44% of patients out of 753 tested CRC(31). Thus our findings are consistent with already existing literature. Previously reported data on HER2 showed membranous expression in 3% (3 out of 100 cases), which is relatively consistent with our data where around 4% showed membranous expression(32).

Programmed Cell Death Protein – 1(PD-L1) has been associated with BRAF mutations and linked with poor differentiation(33). High expression of PD-L1 was reported to be 20%, while low expression was 35%. It was significantly associated with consensus molecular classification over-expressed in CMS1 and under-expressed in CMS4. It was also reported to be concomitantly found with BRAF mutation(34). Thus protein expression differentiating molecular classification can correspond to the BRAF mutant class. In our study, the novel molecular classification pattern was not directly dependent on PD-L1, but in novel CRC class 2, it was not positive in any patient.

Given the experience of breast cancer research, IHC based molecular classification can potentially play a significant role in identifying therapeutic targets and providing proper precision medicine to improve clinical outcome. As the incidence of CRC is feared to rise in upcoming years, it is of utmost importance to control disease survival.

This study was a single-center study with a consecutive series of CRC patients, including treatment naïve tumours; thus, a natural pattern of biomarkers without the interference of chemo and radiotherapy has been presented. The method of molecular classification using R software is also a well-established method in breast cancer, as reported previously. However, there are limited biomarkers, and smaller sample size is appreciated as a limitation of the study. Correlation with gene signature is also recommended in the future, given that the immunohistochemical classification did not show any significant association with survival outcome. However, it has pointed out distinct pathways of colorectal cancer that might have a relationship with the development of colorectal cancer and genetic relationships. Therefore, further studies on risk factors and genetic exploration of these pathways are recommended.

5. Conclusion

In conclusion, CRC is a heterogeneous disease with at least four distinct molecular types. Gene repair mechanism and cell proliferation markers (i.e., p53 and Ki67). Intestine-specific transcription factor (CDX2) has also shown an association with the molecular classification of CRC. Thus, this is now observed that multiple intracellular mechanisms are working together, taking part in tumorigenesis and disease progression. Other associated markers can be traced by following these key pathways.

Ethical consideration: The study was approved by the local research ethics committee of Liaquat University of Health & Sciences Jamshoro Sindh Pakistan under Ethical approval Number LUMHS/REC/641 dated 26-12-2017 The clinical data were retrospectively collected from the Institutional database and tumour blocks were retrieved from the Institutional tissue archive, where tissues are preserved for research purpose. The dataset includes cancer patients from 2008 and a considerable number is not alive; thus, informed consent from individual patients was waived.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Funding:

The study was funded by the Higher Education Commission of Pakistan under **NRPU 3883** to PI Binafsha Manzoor Syed and Co-PI Jawaid Naeem Qureshi. Fayaz Hussain Mangi worked as a Ph.D. scholar under the supervision of PI.

References

1. Globocan [Internet]. 2020. Available from: <https://www.uicc.org/news/globocan-2020-new-global-cancer-data#:~:text=GLOBOCAN 2020 is an online,for all cancer sites combined>.
2. Álvaro E, Cano JM, García JL, Brandáriz L, Olmedillas-López S, Arriba M, et al. Clinical and molecular comparative study of colorectal cancer based on age-of-onset and tumor location: Two main criteria for subclassifying colorectal cancer. *Int J Mol Sci.* 2019;20(4).
3. Alwers E, Bläker H, Walter V, Jansen L, Kloor M, Arnold A, et al. External validation of molecular subtype classifications of colorectal cancer based on microsatellite instability, CIMP, BRAF and KRAS. *BMC Cancer.* 2019;19(1):1–10.
4. Belov L, Zhou J, Christopherson RI. Cell surface markers in colorectal cancer prognosis. Vol. 12, *International Journal of Molecular Sciences.* 2011. 78–113 p.
5. Salehi far S, Soltani M, Zardast M, Ghasemian Moghaddam MR. Investigating the Factors Associated with the Level of Expression of Estrogen and Progesterone Receptors in Patients Suffering from Colorectal Cancer. Antwi S, editor. *J Cancer Epidemiol* [Internet]. 2021 Oct 15;2021:1–6. Available from: <https://www.hindawi.com/journals/jce/2021/4478155/>
6. Qasim BJ. Immunohistochemical Expression of Estrogen and Progesterone Receptors in Human Colorectal Adenoma and Carcinoma Using Specified Automated Cellular Image Analysis System: A Clinicopathological Study. *Oman Med J* [Internet]. 2011 Sep 20;26(5):307–14. Available from: http://www.omjournal.org/fultext_PDF.aspx?DetailsID=139&pdf=images/139_M_Deatials_Pdf_.pdf&type=pdf
7. Hasan R, Bhatt D, Khan S, Khan V, Verma AK, Anees A, et al. Association of Her-2 Expression and Clinicopathological Parameters in Colorectal Carcinoma in Indian Population. *Open Access Maced J Med Sci* [Internet]. 2018 Dec 22;7(1):6–11. Available from: <https://spiroski.migration.publicknowledgeproject.org/index.php/mjms/article/view/oamjms.2019.008>
8. Imaizumi K, Suzuki T, Kojima M, Shimomura M, Sakuyama N, Tsukada Y, et al. Ki67

- expression and localization of T cells after neoadjuvant therapies as reliable predictive markers in rectal cancer. *Cancer Sci* [Internet]. 2020 Jan 18;111(1):23–35. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/cas.14223>
9. Tukaram Patil S, Wilfred Devadass C, Shetty Badila P. Histopathological Evaluation and Analysis of Immunohistochemical Expression of Bcl-2 Oncoprotein in Colorectal Carcinoma. *Iran J Pathol* [Internet]. 2019 Sep 22;14(4):317–21. Available from: http://ijp.iranpath.org/article_36634.html
 10. KIM HK, YANG Y, BYEON S, JEONG Y, KWON J, LEE KH, et al. E-Cadherin and Angiopoietin-2 as Potential Biomarkers for Colorectal Cancer With Peritoneal Carcinomatosis. *Anticancer Res* [Internet]. 2021 Sep;41(9):4497–504. Available from: <http://ar.iarjournals.org/lookup/doi/10.21873/anticancer.15260>
 11. Lakpa SC, Kumar RV, Lilly M. Role of Immunohistochemistry Markers (p53 and CEA), in Study of Colorectal Tumours. *J Pharm Res Int* [Internet]. 2021 Apr 6;94–106. Available from: <https://journaljpri.com/index.php/JPRI/article/view/31362>
 12. Uhlyarik A, Piurko V, Vizkeleti L, Pápai Z, Rásó E, Lahm E, et al. EGFR Protein Expression of KRAS Wild-Type Colorectal Cancer: Predictive Value of the Sidedness for Efficacy of Anti-EGFR Therapy. *Pathol Oncol Res* [Internet]. 2020 Jul 5;26(3):1429–34. Available from: <http://link.springer.com/10.1007/s12253-018-00572-2>
 13. Mohamed SY, Mohammed HL, Ibrahim HM, Mohamed EM, Salah M. Role of VEGF, CD105, and CD31 in the Prognosis of Colorectal Cancer Cases. *J Gastrointest Cancer* [Internet]. 2019 Mar 15;50(1):23–34. Available from: <https://link.springer.com/10.1007/s12029-017-0014-y>
 14. Pyo J-S, Ko SH, Ko YS, Kim NY. Clinicopathological significance of PD-L1 expression in colorectal cancer: Impact of PD-L1 expression on pFOXO1 expression. *Pathol - Res Pract* [Internet]. 2020 Feb;216(2):152764. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0344033819320047>
 15. Aasebø K, Dragomir A, Sundström M, Mezheyeuski A, Edqvist P-H, Eide GE, et al. CDX2: A Prognostic Marker in Metastatic Colorectal Cancer Defining a Better BRAF Mutated and a Worse KRAS Mutated Subgroup. *Front Oncol* [Internet]. 2020 Feb 11;10. Available from: <https://www.frontiersin.org/article/10.3389/fonc.2020.00008/full>
 16. Al-Maghrabi J, Emam E, Gomaa W. Immunohistochemical staining of cytokeratin 20 and cytokeratin 7 in colorectal carcinomas: Four different immunostaining profiles. *Saudi J Gastroenterol* [Internet]. 2018;24(2):129. Available from: <http://www.saudijgastro.com/text.asp?2018/24/2/129/229487>
 17. Howell A, Harland RNL, Bramwell VHC, Swindell R, Barnes DM, Redford J, et al. Steroid-Hormone Receptors and Survival After First Relapse in Breast Cancer. *Lancet*. 1984 Mar;323(8377):588–91.
 18. Syed BM, Green AR, Paish EC, Soria D, Garibaldi J, Morgan L, et al. Biology of primary breast cancer in older women treated by surgery: With correlation with long-term clinical outcome and comparison with their younger counterparts. *Br J Cancer*. 2013;108(5).

19. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: A new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res.* 2004;10(21):7252–9.
20. Perea J, Rueda D, Canal A, Rodríguez Y, Álvaro E, Osorio I, et al. Age at onset should be a major criterion for subclassification of colorectal cancer. *J Mol Diagnostics* [Internet]. 2014;16(1):116–26. Available from: <http://dx.doi.org/10.1016/j.jmoldx.2013.07.010>
21. Perez Villamil B, Romera Lopez A, Hernandez Prieto S, Lopez Campos G, Calles A, Lopez Asenjo JA, et al. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. *BMC Cancer* [Internet]. 2012;12(1):1. Available from: *BMC Cancer*
22. Alwers E, Jia M, Kloor M, Bläker H, Brenner H, Hoffmeister M. Associations Between Molecular Classifications of Colorectal Cancer and Patient Survival: A Systematic Review. *Clin Gastroenterol Hepatol* [Internet]. 2019;17(3):402-410.e2. Available from: <https://doi.org/10.1016/j.cgh.2017.12.038>
23. Trinh A, Trumpi K, De Sousa E Melo F, Wang X, De Jong JH, Fessler E, et al. Practical and robust identification of molecular subtypes in colorectal cancer by immunohistochemistry. *Clin Cancer Res.* 2017;23(2):387–98.
24. Stintzing S, Wirapati P, Lenz HJ, Neureiter D, Fischer von Weikersthal L, Decker T, et al. Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KKR-0306) trial. *Ann Oncol.* 2019;30(11):1796–803.
25. Phipps AI, Limburg PJ, Baron JA, Burnett-Hartman AN, Weisenberger DJ, Laird PW, et al. Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology.* 2015;148(1):77-87.e2.
26. Purcell R V., Schmeier S, Lau YC, Pearson JF, Frizelle FA. Molecular subtyping improves prognostication of Stage 2 colorectal cancer. *BMC Cancer.* 2019;19(1):1–9.
27. Nasierowska-Guttmejer A, Trzeciak L, Nowacki MP, Ostrowski J. P53 Protein Accumulation and P53 Gene Mutation in Colorectal Cancer. *Pathol Oncol Res.* 2000;6(4):275–9.
28. Konda K, Konishi K, Yamochi T, Ito YM, Nozawa H, Tojo M, et al. Distinct molecular features of different macroscopic subtypes of colorectal neoplasms. *PLoS One.* 2014;9(8).
29. Xu F, Wang F, Di M, Huang Q, Wang M, Hu H, et al. Classification based on the combination of molecular and pathologic predictors is superior to molecular classification on prognosis in colorectal carcinoma. *Clin Cancer Res.* 2007;13(17):5082–8.
30. Sugai T, Eizuka M, Takahashi Y, Fukagawa T, Habano W, Yamamoto E, et al. Molecular subtypes of colorectal cancers determined by PCR-based analysis. *Cancer Sci.* 2017;108(3):427–34.
31. Domingo E, Ramamoorthy R, Oukrif D, Rosmarin D, Presz M, Wang H, et al. Use of multivariate analysis to suggest a new molecular classification of colorectal cancer. *J Pathol.* 2013;229(3):441–8.

32. Osako T, Miyahara M, Uchino S, Inomata M, Kitano S, Kobayashi M. Immunohistochemical study of c-erbB-2 protein in colorectal cancer and the correlation with patient survival. *Oncology*. 1998;55(6):548–55.
33. Azcue P, Encío I, Setas DG, Alecha JS, Galbete A, Mercado M, et al. Pd-11 as a prognostic factor in early-stage colon carcinoma within the immunohistochemical molecular subtype classification. *Cancers (Basel)*. 2021;13(8):1–18.
34. Shiovitz S, Grady WM. Molecular Markers Predictive of Chemotherapy Response in Colorectal Cancer. *Curr Gastroenterol Rep*. 2015;17(2):1–19.

Tables

Table 1. Demographic and Biological characteristics of colorectal cancer

Factor	N(%)
Age	
≤25	37(18.4)
26-50	95(47.3)
51-75	63(31.3)
≥75	6(3.0)
Tumour stage	
T1	2(1.1)
T2	50(27.5)
T3	89(48.9)
T4	15(8.2)
Lymph node stage	

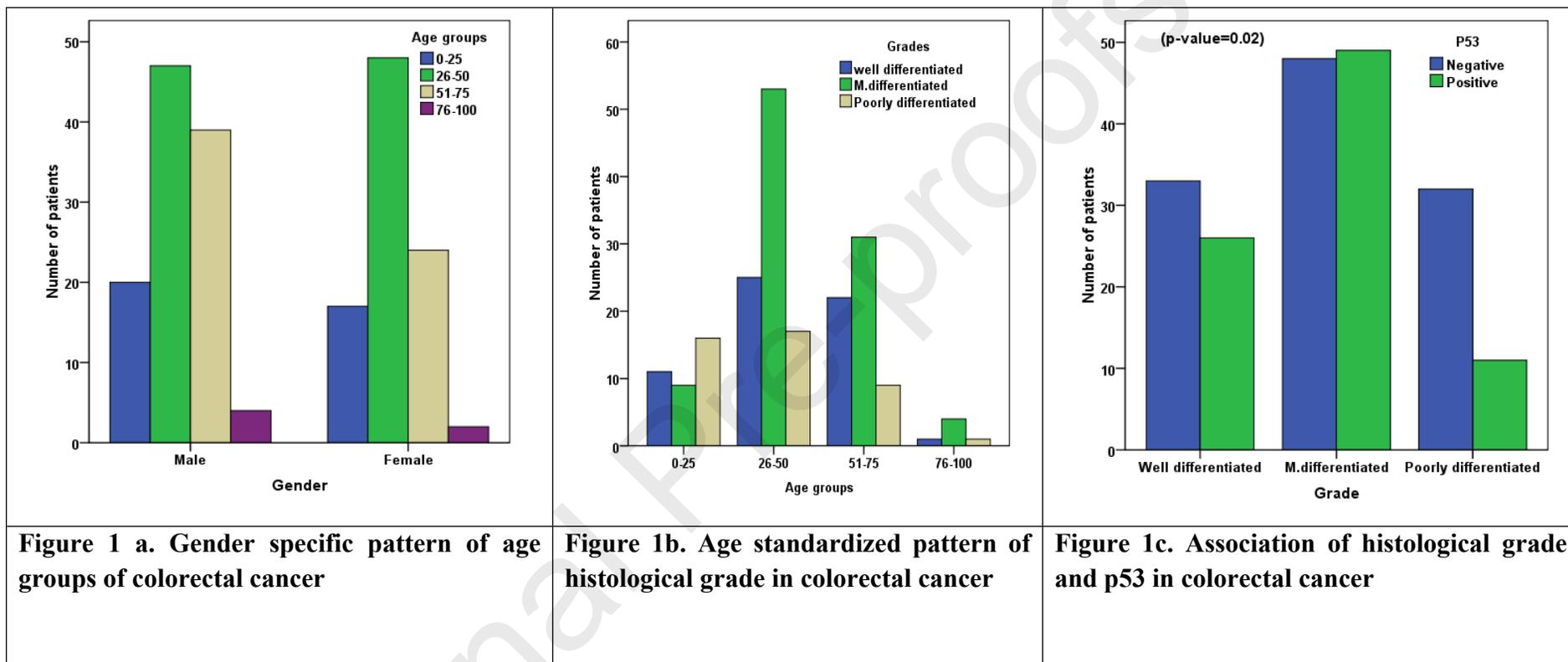
N0	41(22.5)
N1	49(26.9)
N2	9(4.9)
Metastatic status	
M0	169 (92.4)
M1	14 (7.7)
Grade	
Well differentiated	59(29.6)
Moderately differentiated	97(48.7)
Poorly differentiated	43(21.6)
Histological types	
Adenocarcinoma	123(61.5)
Mucinous adenocarcinoma	50(25.0)
Micro-papillary carcinoma	14(7.0)
Signet ring cell carcinoma	7(1.5)
Other types	6(3.0)
Biological markers	
Ki67 Positive	165(82.1)
CDX2 Positive	171(85.1)
ER positive	6(3.0)
PR positive	5(2.5)
HER2 positive	9(4.5)
EGFR positive	17(8.5)
Bcl2 positive	5(2.5)
P53 positive	87(43.3)
CK20 positive	199(99.0)
CEA positive	198(98.5)
E-Cadherin positive	199(99.0)
PDL-1 positive	29(14.4)
VEGF positive	11(5.5)

Table 2. Clinical and biological characteristics of novel molecular classes of Colorectal cancer

Characteristics	CRC Novel Cluster 1 (n=67)	CRC Novel Cluster 2 (n=38)	CRC Novel Cluster 3 (n=28)	CRC Novel Cluster 4 (n=50)
Mean age (range)	42.8 (10-78)	48.6 (10-85)	42.8 (12-65)	37.6 (10-77)
Gender n(%)				
Male	39(58.2)	22(57.9)	14(50.0)	27(54.0)

Females	28(41.8)	16(42.1)	14(50.0)	23(46.0)
Tumour location n(%)				
Right Colon	8(11.9)	9(24.3)	3(10.7)	11(22.0)
Transverse	4(6.0)	2(5.4)	2(7.1)	0
Left colon	10(14.9)	4(10.8)	2(7.1)	0
Sigmoid	10(14.9)	7(18.9)	4(14.3)	6(12.0)
Rectum	33(49.3)	11(29.7)	10(35.7)	26(52.0)
Recto-sigmoid	8(11.9)	4(10.8)	7(25.0)	7(14.0)
Grade n(%)				
1	19(28.4)	12(32.4)	11(39.3)	12(24.5)
2	31(46.3)	20(54.1)	13(46.4)	25(51.0)
3	17(25.4)	5(13.5)	4(14.3)	12(24.5)
Lymphocytic infiltration n(%)				
Present	52(77.6)	32(84.2)	26(92.9)	36(72.0)
Not present	15(22.4)	6(15.8)	2(7.1)	14(28.0)
Median disease specific survival (months)	30	25	26	23

Figures



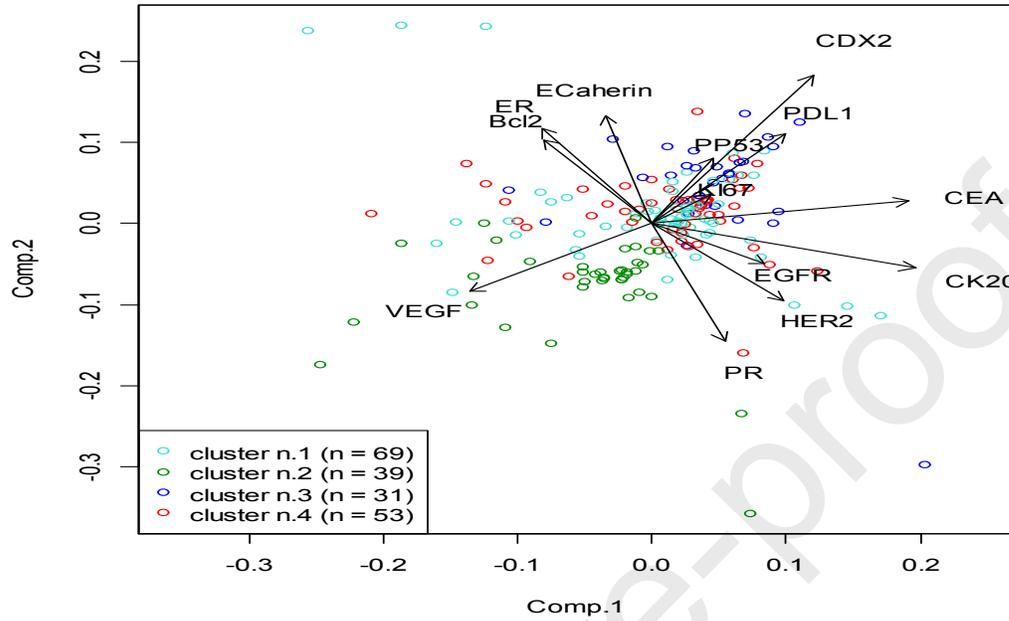
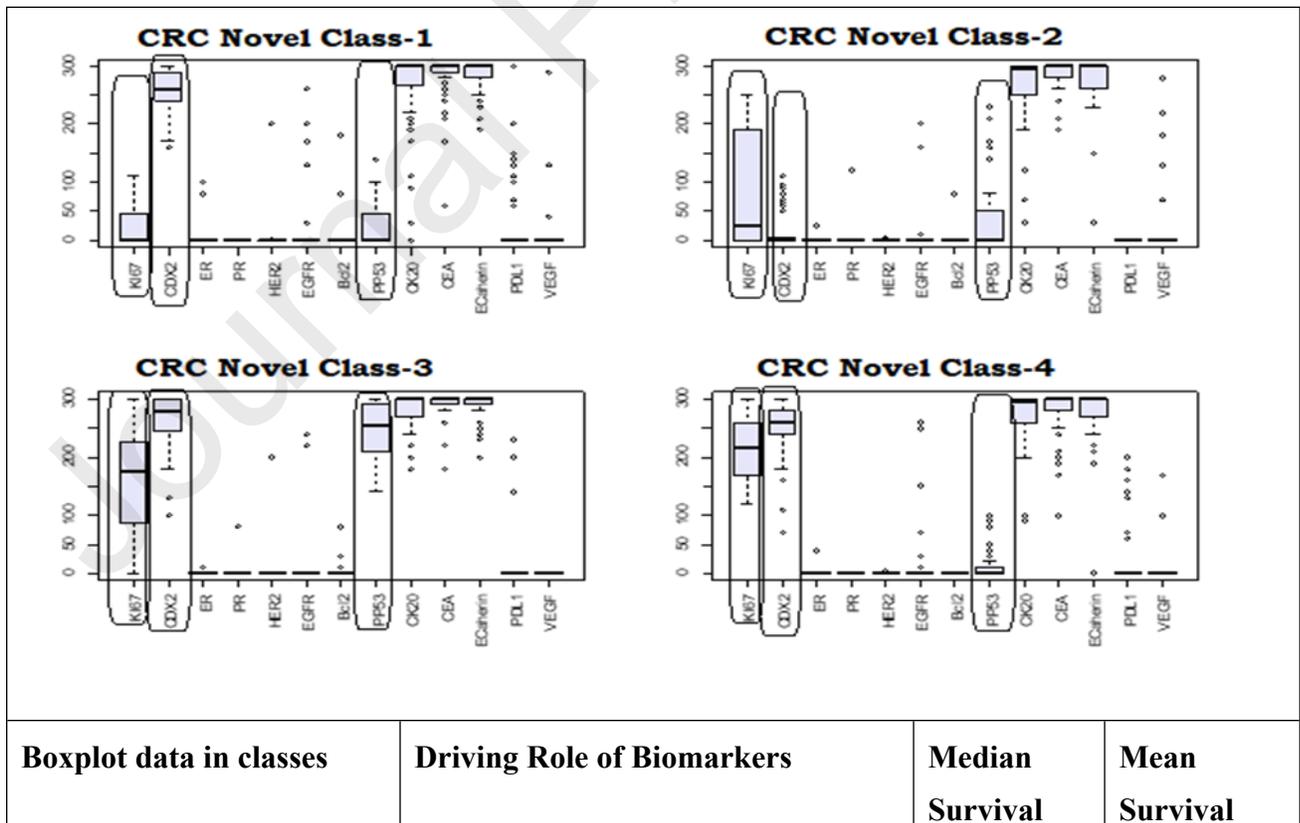
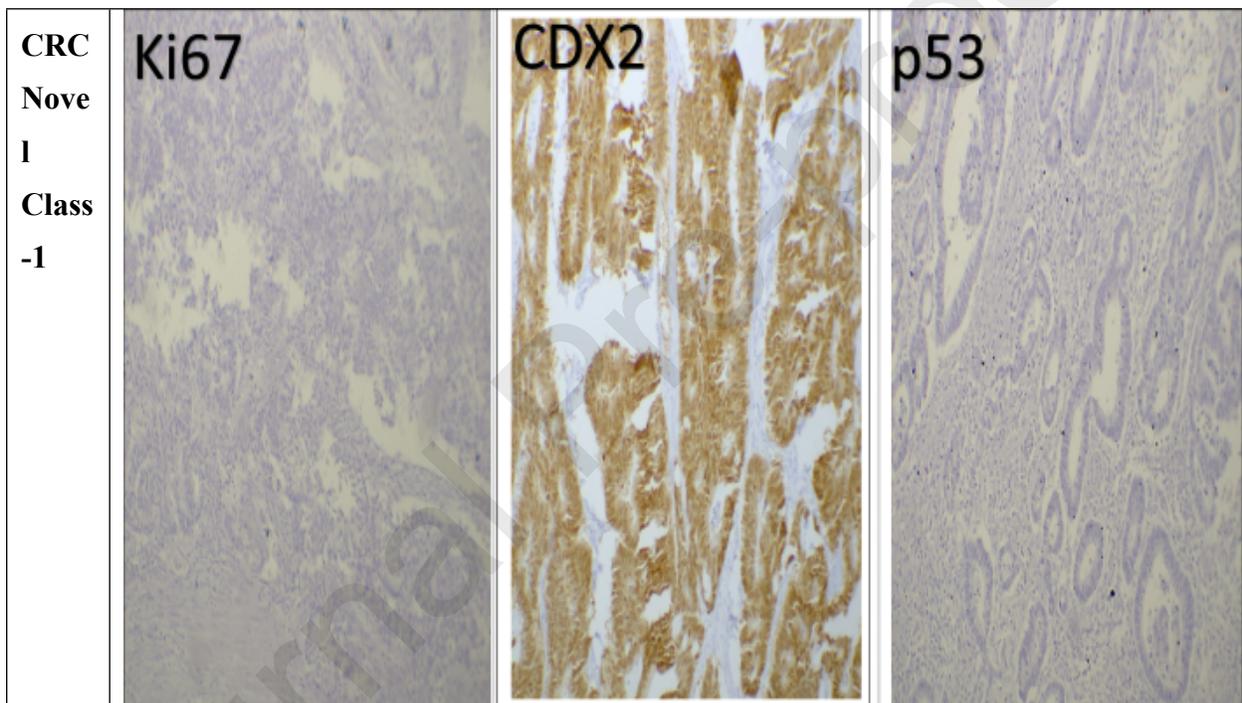


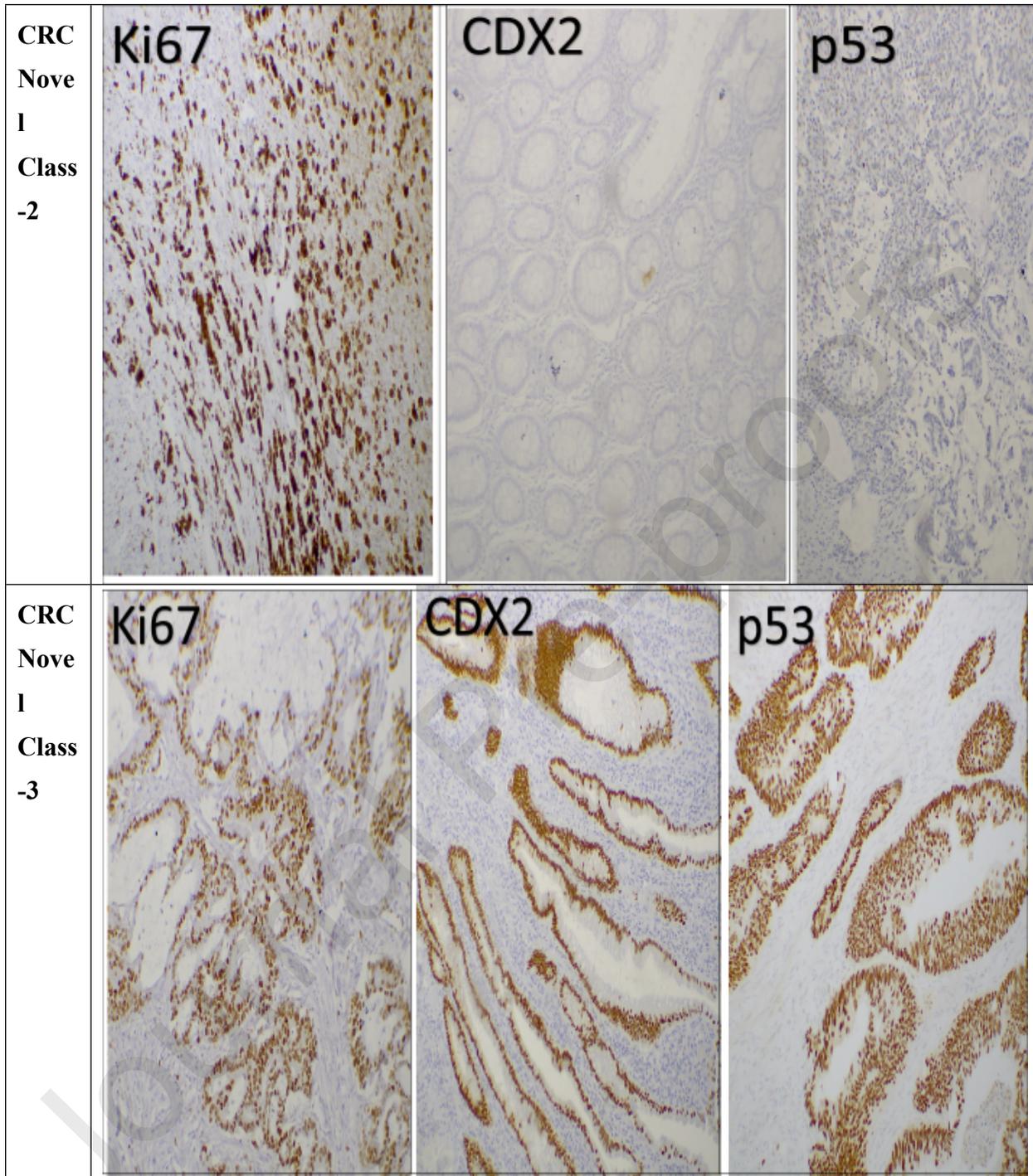
Figure: 2a. Biplots K-means of molecular classification of colorectal cancer



CRC Novel Class-1	low KI67 , high CDX2 &low P53	30	22.256
CRC Novel Class-2	High ki67,low CDX2 &low P53	25	24.472
CRC Novel Class- 3	High ki67,high CDX2 &high P53	26	27.483
CRC Novel Class- 4	High ki67,high CDX2 &low P53	23	26.873

Figure 2b. Novel molecular classification of colorectal cancer in Pakistani population





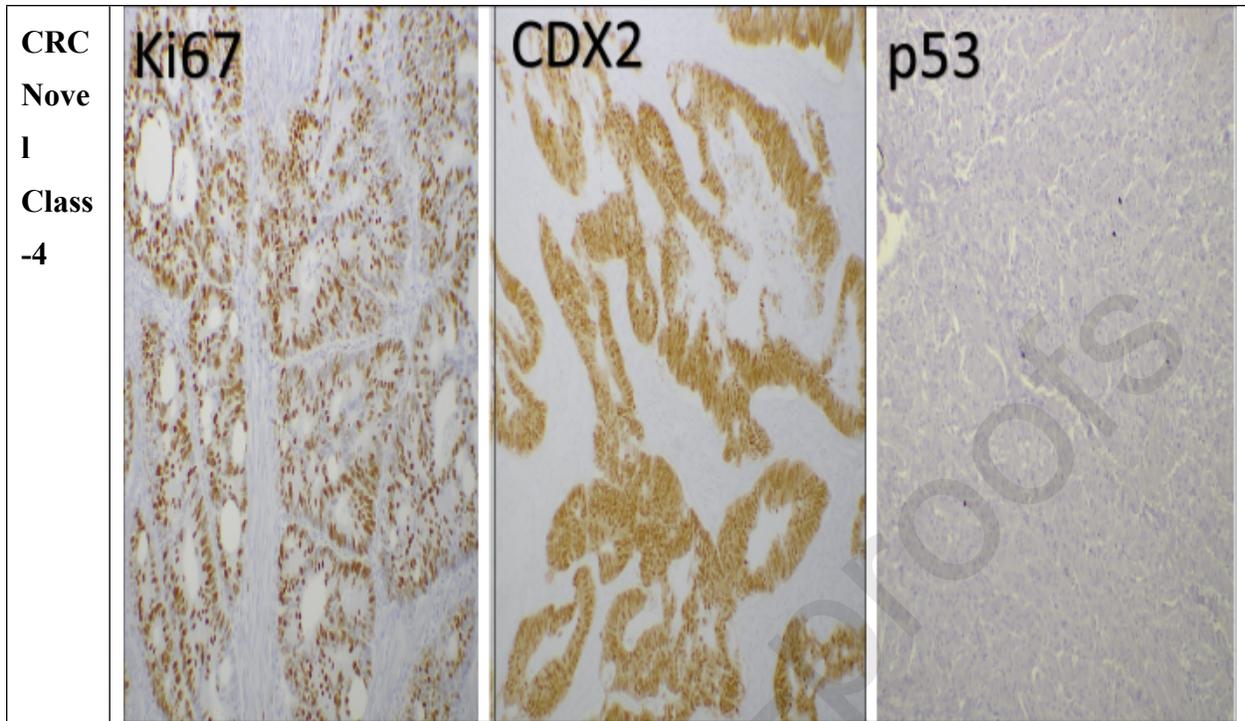


Figure 2c. Immunohistochemical pattern of novel molecular classification of colorectal cancer in Pakistani population(Magnification size -10x using Euromax Microscope with camera)

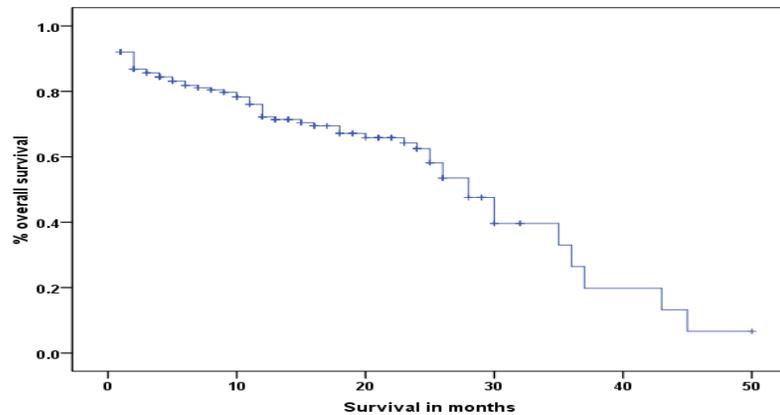


Figure 3a. Overall survival in patients with colorectal cancer

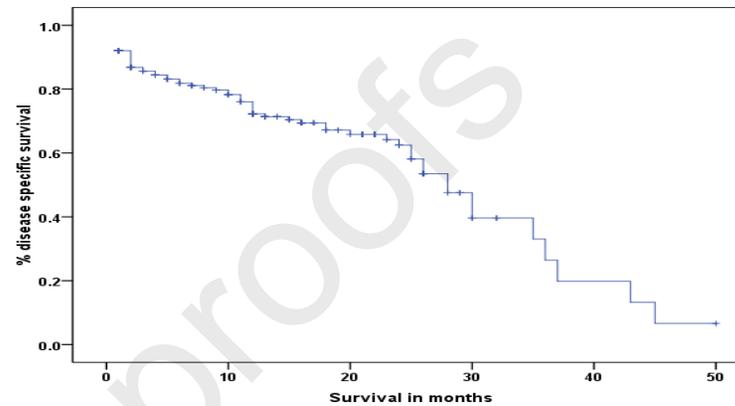


Figure 3b. Disease specific survival in patients with colorectal cancer

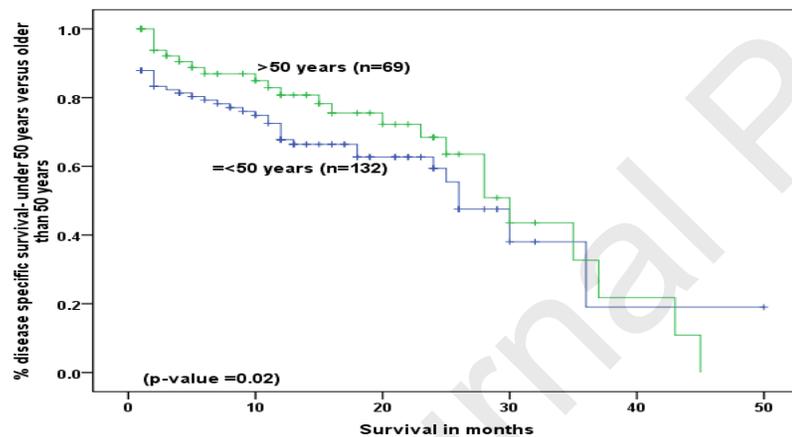


Figure 3c. Disease specific survival of patients with colorectal cancer – a comparison of <50 and older then 50 years

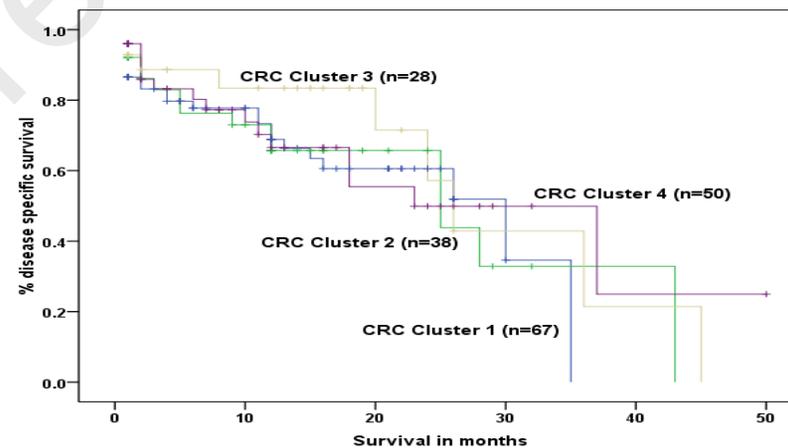


Figure 3d. Disease specific survival of patients with colorectal cancer- Comparison of novel molecular classes

Supplementary Tables

Supplementary Table- S1. Summary of Immunohistochemical markers: primary antibody, incubation time, temperature and cutoff percentage of positive cells to define positivity

S.NO	Name of Antibody	Incubation Time	Temperature	Cutoff % for Positive expression
01	ER (Ready to Use)	30 minutes	Room Temperature	1
02	PR (Ready to Use)	30 minutes	Room Temperature	1
03	HER2-neu Dilution 1:500	40 minutes	Room Temperature	3+
04	EGFR Dilution 1:100	40 minutes	Room Temperature	5
05	VEGF Dilution 1:25	40 minutes	Room Temperature	5
06	BCL2 (Ready to Use)	30 minutes	Room Temperature	10
07	Ki-67 (Ready to Use)	30 minutes	Room Temperature	10
08	P-53 (Ready to Use)	30 minutes	Room Temperature	5
09	E-Cadherin (Ready to Use)	30 minutes	Room Temperature	1
10	CEA (Ready to Use)	30 minutes	Room Temperature	1
11	CDX-2 (Ready to Use)	20 minutes	Room Temperature	1
12	PDL-1 Dilution 1:50	40 minutes	Room Temperature	1

13	CK-20 (Ready to Use)	20 minutes	Room Temperature	1
-----------	----------------------	------------	---------------------	---

Journal Pre-proofs