ABSTRACT
of the dissertation by Nurlan Baltayev
«Optimizing the Early Diagnosis of Colorectal Cancer in Individuals with a Positive Family History»
submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD) in the specialty «8D10103 – Medicine»

Scientific Advisor
Dilyara Kaidarova
Doctor of Medical Sciences, Professor
Academician of the National Academy of Sciences
of the Republic of Kazakhstan
Foreign Scientific Advisor
Arsen Rasulov
Doctor of Medical Sciences, Professor

Almaty, 2026

1. Relevance of the Research Topic

According to the National Cancer Registry of the Republic of Kazakhstan, colorectal cancer (CRC) ranked second among all malignant neoplasms in terms of incidence and third in cancer-related mortality in 2025, with more than 4,000 new cases diagnosed annually [1]. A substantial proportion of CRC cases are familial or hereditary and are characterized by the clustering of cases within families, an earlier age at disease onset than in the general population, and an increased incidence of multiple primary malignancies. Published evidence indicates that familial and hereditary forms of colorectal cancer, including both syndromic and non-syndromic variants, account for approximately 25–30% of all CRC cases [1]. Blood relatives of patients with colorectal cancer constitute a high-risk population. However, these clinically unaffected individuals are not routinely included in personalized surveillance or early diagnostic programs and typically seek medical care only after clinical manifestations of CRC have developed, often at a locally advanced stage. Importantly, most unaffected relatives are young or middle-aged adults, many of whom do not meet the age eligibility criteria for the national colorectal cancer screening program [3].

Multigene DNA testing using next-generation sequencing (NGS) enables the identification of pathogenic germline variants that either directly contribute to CRC development or significantly increase disease susceptibility [4]. This creates a strong rationale for establishing a personalized system for genetic risk assessment, early detection, and surveillance of clinically unaffected individuals in order to reduce the incidence of colorectal cancer. To date, however, no NGS-based system for the early diagnosis of colorectal cancer in high-risk individuals has been implemented in the Republic of Kazakhstan. Furthermore, the limited number of NGS studies conducted in the country have primarily been exploratory rather than clinically oriented [5].

Personalized oncology relies on clinically relevant genomic information from both the patient and the tumor to support accurate risk assessment, prognosis, and treatment selection. This is particularly important in colorectal cancer, where the identification of pathogenic germline variants in clinically unaffected relatives provides a clear indication for targeted colonoscopic surveillance aimed at the early detection of precancerous lesions and early-stage colorectal neoplasia [6].

The present dissertation addresses this important clinical challenge. For the first time in the Republic of Kazakhstan, it provides a scientific basis for optimizing the early diagnosis of colorectal cancer in individuals with a positive family history through the combined use of multigene testing and expert colonoscopy. The proposed approach facilitates individualized risk stratification, personalized surveillance, and the timely detection of colorectal pathology, thereby contributing to a reduction in CRC incidence among genetically predisposed individuals.

The topic of this dissertation is consistent with current international priorities in both fundamental and translational cancer research. It is also aligned with the strategic healthcare and research priorities of the Republic of Kazakhstan, including:

1. The Code of the Republic of Kazakhstan "On Public Health and the Healthcare System" (Article 5, Paragraph 16), which promotes state support for medical and pharmaceutical research and the implementation of advanced scientific

and technological achievements in disease prevention, diagnosis, and treatment, as well as the adoption of international best practices in healthcare [7];

2. The Comprehensive Cancer Control Plan of the Republic of Kazakhstan for 2023–2027, particularly Priority Area 2, *Highly Effective Early Diagnosis* [8];

3. The List of Priority Areas for Biomedical Research approved by the Scientific and Technical Council of the Ministry of Health of the Republic of Kazakhstan, specifically Section 1, *Nation's Health: Disease Prevention and Management*, Priority 2, *Oncology: Early Detection and Treatment*.

2. Study Goal

To provide a scientific basis for optimizing the early diagnosis of colorectal cancer in individuals with a positive family history through the implementation of multigene testing.

3. Objectives of the Dissertation Research

1. To investigate the patterns of familial clustering of colorectal cancer (CRC) among the families of patients included in the study.
2. To determine the spectrum and frequency of pathogenic genetic variants associated with CRC in patients and their clinically unaffected relatives using next-generation sequencing (NGS).
3. To evaluate the colonoscopic findings in clinically unaffected relatives carrying pathogenic genetic variants associated with CRC.
4. To assess the diagnostic value of the serum tumor markers CA19-9 and CEA in clinically unaffected relatives of patients with colorectal cancer.
5. To develop and scientifically validate a multigene testing-based system for optimizing the early diagnosis of colorectal cancer in individuals with a positive family history.

4. Materials and methods of the Dissertation Research

This study was designed as an open-label, prospective, comparative cohort study involving human participants and biological specimens. The study protocol was approved by the Scientific Council of the Kazakhstan Medical University of Continuing Education (KazMUCE) (Protocol No. 5, December 3, 2020), the Local Ethics Committee of KazMUCE (Protocol No. 2, February 16, 2021), and the Local Ethics Committee of the Kazakh Institute of Oncology and Radiology (Protocol No. 1, May 12, 2021). A total of 247 participants were enrolled, including 155 patients with colorectal cancer aged 21–78 years (mean age, 46.1 ± 3.5 years) who underwent examination and treatment at the Almaty Oncology Center, and 92 clinically unaffected relatives (mean age, 31.7 ± 1.3 years). The inclusion criteria for patients were: (1) histologically confirmed colorectal adenocarcinoma of any degree of differentiation; (2) age ≥18 years; and (3) provision of written informed consent. The exclusion criteria were: (1) colorectal malignancies other than adenocarcinoma; (2) squamous cell carcinoma of the rectum or anal canal; (3) age <18 years; and (4) inability to provide informed consent. The inclusion criteria for relatives were: (1) first- or second-degree biological relationship to a patient with colorectal cancer; (2) no history of malignant disease; (3) absence of inflammatory bowel disease or acute infectious conditions affecting the colon at the time of colonoscopy; and (4) provision of written informed

consent. The exclusion criteria were: (1) previously diagnosed colorectal neoplasia; (2) a history of any malignant disease; and (3) inability to provide informed consent. For participants younger than 18 years of age, questionnaires were completed and blood samples were collected in the presence of their legal guardians.

This dissertation employed the following research methods: clinical-genealogical (family history analysis), instrumental (colonoscopy), laboratory (histopathological examination), molecular genetic (next-generation sequencing [NGS]-based DNA analysis), and bioinformatic (annotation and interpretation of genetic variants).

The object of this study is the personalized early diagnosis of colorectal cancer in apparently healthy relatives of patients with colorectal cancer through multigene testing.

The subject of this study comprises germline pathogenic genetic variants and colorectal pathological lesions identified in patients with colorectal cancer and their apparently healthy relatives.

Statistical significance was assessed using the *p*-value, which represents the probability of obtaining the observed results under the null hypothesis (i.e., the hypothesis that no effect or difference exists). Descriptive statistics for categorical variables were presented as frequency distributions, whereas continuous variables were analyzed using the Wilcoxon rank-sum test. Additional statistical analyses included Fisher's exact test, the Mann–Whitney U test, calculation of relative frequencies, Pearson's correlation coefficient (*r*), Cramer's V, and the correlation ratio (η).

5. Description of the Main Results of the Dissertation Research

Analysis of familial aggregation of colorectal cancer (CRC) demonstrated that none of the investigated cases were associated with known hereditary cancer syndromes caused by pathogenic variants in high-penetrance genes, indicating that all cases were non-syndromic. The substantial proportion of patients with a family history of CRC (9.7%) or other malignancies (24.5%) highlights the need for risk stratification and multigene testing to identify clinically significant pathogenic variants among individuals at high and moderate risk, in accordance with the criteria of the British Society of Gastroenterology (BSG) and the UK Cancer Genetics Group (UKCGG).

Germline pathogenic variants were identified in 19.4% of patients. More than half (53.3%) of all variant carriers were young adults (18–44 years of age), indicating a significantly higher prevalence of pathogenic variants in younger patients ($p = 0.0184$). The most frequently affected genes included the tumor suppressor gene APC, genes involved in cell cycle regulation and DNA repair (*CHEK2* and *BRCA1*), and mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *MUTYH*).

Combined assessment of serum carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) levels showed no diagnostic value in apparently healthy relatives of patients with CRC who carried pathogenic germline variants.

Pathogenic germline variants were detected in 21.7% of apparently healthy relatives of patients with CRC. The identified variants were identical to those detected in the corresponding patients, both with respect to the affected genes and the specific DNA sequence alterations. Colonoscopic examination revealed colorectal lesions in

33.3% of variant carriers and early-stage colorectal cancer in 9.5%. The comparison group consisted of 32 first-degree relatives (20–75 years of age) in whom no pathogenic variants were detected. Colonoscopy identified no colorectal lesions, including polyps or malignant neoplasms, in this group (p = 0.05). These findings indicate that although family history is an important predictor for selecting individuals for genetic testing, it should not be considered the sole criterion. When pathogenic germline variants are identified in apparently healthy relatives, active surveillance is warranted. Colonoscopic surveillance should be individualized according to the clinical phenotype, the pattern of familial CRC aggregation, and the functional significance of the identified genetic variants.

For the first time in the Republic of Kazakhstan, the frequency and spectrum of pathogenic germline variants associated with colorectal cancer and other malignancies were investigated in patients with CRC and their apparently healthy relatives. NGS analysis identified six previously unreported pathogenic variants in the *APC, PMS1, MLH1, NBN,* and *EPCAM* genes that had not been described in the scientific literature or catalogued in international genomic databases. The relatively high prevalence of pathogenic germline variants (19.4%), particularly among young patients with CRC, suggests population-specific features of the genetic landscape of colorectal cancer in Kazakhstan and highlights the limitations of previously used phenotype-based testing strategies. The findings support the implementation of NGS-based multigene testing for patients with CRC and their relatives using a panel that includes both established colorectal cancer susceptibility genes and additional clinically relevant genes with moderate penetrance. This prospective cohort study demonstrated that multigene panel testing provides substantial clinical benefit for both patients and their family members. From a practical perspective, the principal contribution of this study is the demonstration of the clinical utility of combining multigene testing with colonoscopy in apparently healthy relatives of patients with CRC. This integrated approach formed the basis for the development and scientific validation of algorithms for personalized early detection and surveillance of colorectal cancer in individuals with a significant family history of the disease, representing the first such evidence-based framework developed in the Republic of Kazakhstan.

6. Scientific Novelty of the Dissertation Research

1. For the first time in the Republic of Kazakhstan, a comparative analysis of the frequency and spectrum of germline pathogenic variants associated with colorectal cancer (CRC) and other malignancies was conducted in cohorts of patients with CRC and their apparently healthy relatives.

2. Six previously unreported pathogenic variants were identified in the *APC, PMS1, MLH1, NBN,* and *EPCAM* genes associated with CRC and other cancers. These variants had not been previously described in the scientific literature or catalogued in international genomic databases.

3. For the first time in the Republic of Kazakhstan, the effectiveness of the combined use of multigene testing and colonoscopy in individuals with a significant family history of cancer was evaluated for the purposes of early detection and risk assessment of colorectal cancer.

4. For the first time in the Republic of Kazakhstan, a system for optimizing early detection of colorectal cancer in individuals with a significant family history, based on multigene testing, was developed and scientifically validated.

5. For the first time in the Republic of Kazakhstan, diagnostic algorithms for early detection of colorectal cancer in individuals with a significant family history were developed and implemented.

7. Key Findings of the Dissertation Research

1. The prevalence of germline pathogenic variants in the study cohort was 19.4%, representing a relatively high frequency among patients with non-syndromic colorectal cancer.

2. Pathogenic variants were most frequently identified in the *APC* tumor suppressor gene, genes involved in cell cycle regulation and DNA repair (*CHEK2* and *BRCA1*), and mismatch repair (MMR) genes (*MLH1, MSH2, MSH6,* and *MUTYH*).

3. More than half (53.3%) of all carriers of pathogenic variants were young adults (18–44 years of age), highlighting the importance of genetic screening in younger individuals at risk of colorectal cancer.

4. Rare germline pathogenic variants were detected in 21.7% of apparently healthy relatives of patients with colorectal cancer, underscoring the clinical value of molecular genetic testing for family members from families with clustering of colorectal cancer.

5. The combined use of multigene testing and high-quality colonoscopy improves the early detection of colorectal cancer and facilitates personalized surveillance of individuals at increased risk of familial and hereditary colorectal cancer.

6. The proposed algorithms for the early diagnosis of individuals with a family history of colorectal cancer provide a framework for improving risk stratification and may contribute to reducing the burden of colorectal cancer.

8. Practical Significance of the Results Obtained

1. Algorithms for the early diagnosis of colorectal cancer in individuals with a significant hereditary predisposition have been implemented in the clinical practice of regional oncology centers of the Republic of Kazakhstan, including the Multidisciplinary Medical Center (Astana City Health Department), the East Kazakhstan Regional Multidisciplinary Center for Oncology and Surgery (East Kazakhstan Region Health Department), the Center for Nuclear Medicine and Oncology (Abai Region Health Department), and Multidisciplinary Hospital No. 3 of Karaganda City.

2. The clinical implementation of the early diagnosis algorithms developed within this dissertation research enables assessment of individual genetic risk for colorectal cancer development and supports personalized diagnostic strategies and surveillance of individuals with a significant family history of the disease. To standardize the application of these approaches, methodological guidelines describing the proposed algorithms have been published: Afonin G.A., Kaidarova D.R., Baltayev N.A., Goncharova T.G., Khozhaev A.A. *Multigene Testing in the Early Diagnosis and Genetic Screening of Colorectal Cancer.* Almaty: Kazakh Institute of Oncology and

Radiology; Asfendiyarov Kazakh National Medical University, 2023. 88 p. ISBN 978-601-246-743-7.

3. Optimization of early colorectal cancer diagnosis through the implementation of multigene testing provides a basis for improving risk assessment, enabling targeted surveillance, and potentially reducing the burden of colorectal cancer in the Republic of Kazakhstan.

9. Personal Contribution of the Author to the Dissertation Research

The author's personal contribution to the dissertation research included the study design and development of the research protocol; patient recruitment and clinical management; recruitment of apparently healthy relatives of patients; collection of biological samples; development of questionnaires and individual case report forms; creation and management of databases; and analysis and interpretation of the results of molecular genetic (DNA sequencing) and instrumental (colonoscopy) investigations. Bioinformatic analysis and interpretation of NGS sequencing data for cancer-associated target genes were performed with the support of geneticists from the Institute of Genetics and Physiology of the Ministry of Science and Higher Education of the Republic of Kazakhstan. The author independently developed algorithms for the early diagnosis of colorectal cancer in individuals with a significant hereditary predisposition, supervised their implementation in clinical practice, analyzed the obtained results, and prepared the dissertation manuscript.

10. Conclusions

1. Analysis of familial aggregation of colorectal cancer (CRC) demonstrated that none of the investigated cases were associated with known hereditary cancer syndromes caused by pathogenic variants in high-penetrance genes, indicating their non-syndromic nature. The considerable proportion of patients with a family history of CRC (9.7%) and other malignancies (24.5%) highlights the need for risk stratification and multigene testing to identify clinically significant germline variants.

2. Germline pathogenic variants were identified in 19.4% of patients, with young patients (18–44 years of age) accounting for 53.3% of variant carriers. Pathogenic variants were most frequently detected in the tumor suppressor gene *APC*, genes involved in cell cycle regulation and DNA repair (*CHEK2* and *BRCA1*), and mismatch repair (MMR) genes (*MLH1, MSH2, MSH6,* and *MUTYH*). Germline pathogenic variants were identified in 21.7% of apparently healthy relatives of patients with CRC, and the variant profiles demonstrated concordance with those of the corresponding patients at both the gene and DNA sequence levels.

3. Colonoscopic evaluation revealed colorectal lesions in 33.3% of relatives carrying pathogenic germline variants, including early-stage CRC in 9.5% of these individuals.

4. Combined assessment of serum tumor markers, including carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA), demonstrated no diagnostic value for identifying colorectal cancer in apparently healthy relatives of patients with pathogenic germline variants.

5. The developed system for optimizing early colorectal cancer diagnosis enables the implementation of personalized genetic screening and colonoscopic

surveillance protocols for apparently healthy individuals at hereditary risk, providing a framework for improving early detection and reducing the clinical burden of CRC among genetically predisposed individuals.

11. Validation of the Dissertation Results

The results and principal findings of the dissertation research were presented at the following international and national scientific and scientific-practical conferences:

1. International Scientific and Practical Conference "Oncology in Kazakhstan: From Past to Future," dedicated to the 60th anniversary of the Kazakh Research Institute of Oncology and Radiology. Almaty, Republic of Kazakhstan, December 10–11, 2020.
2. VIII Congress of Oncologists and Radiologists of the Republic of Kazakhstan. Turkistan, Republic of Kazakhstan, October 16, 2021.
3. International Scientific and Practical Conference "Modern Schools of Oncology: Current Perspectives on All Aspects." Almaty, Republic of Kazakhstan, September 22–23, 2022.
4. International Scientific and Practical Conference "Ecology. Radiation. Health," dedicated to the 70th anniversary of Semey Medical University. Semey, Republic of Kazakhstan, August 28–29, 2023.
5. IX Congress of Oncologists and Radiologists of the Republic of Kazakhstan. Astana, Republic of Kazakhstan, October 26–27, 2023.
6. International Scientific and Practical Conference "Onco-Renaissance 2024." Almaty, Republic of Kazakhstan, November 14–16, 2024.
7. Scientific and Practical Conference "New Frontiers in the Molecular Genetic Aspects of Clinical Oncology." Almaty, Republic of Kazakhstan, December 5–6, 2024.
8. International Conference "Personalized Medicine: From Tumor Molecular Genetic Characteristics to Therapy Selection." Astana, Republic of Kazakhstan, November 27 – 28, 2025.

12. Publications

Based on the results of the dissertation research, 10 scientific works have been published, including: 3 articles in journals recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (*Oncology and Radiology of Kazakhstan, Eurasian Journal of Oncology and Radiology*); 2 articles in international peer-reviewed journals indexed in the Scopus and Web of Science databases (*Genes* [Q2, impact factor 3.1, percentile 70] and *Scientific Reports* [Q1, impact factor 4.9, percentile 87]); 4 abstracts in the proceedings of international scientific conferences and congresses; and 1 set of methodological guidelines.

13. Scope and Structure of the Dissertation

The dissertation comprises 152 pages and includes an introduction, four main sections (literature review, materials and methods, results, and discussion), a conclusion, a list of references, and eight appendices. The dissertation contains 8 tables and 26 figures. The reference list includes 226 sources.