ANNOTATION

on PhD thesis of Georgiy Afonin entitled **“Molecular genetic analysis of the development of colorectal cancer in patients under the age of 50”** presented as an application for PhD degree on the specialty 6D110100 – Medicine

The relevance of research

According to the World Health Organization (WHO) and the International Agency for Research on Cancer (IARC), there is an increase in the incidence of colorectal cancer (CRC) in the world, and by 2030 the number of new cases is expected to increase to 2.2 million per year (by 77% compared to 2018) and mortality from this pathology up to 1.1 million cases (by 80% compared to 2018) (Colorectum. Estimated age-standardized incidence rates (World) in 2018; Arnold M., 2016; Bray F., 2018).

In the Republic of Kazakhstan (RK), there is also an annual increase in the incidence of CRC, which is in 3rd place in terms of frequency of occurrence in the general structure of oncological pathology. According to the data of the National Cancer Registry, the country has consistently high mortality rates from CRC in men and women, despite the functioning of the national CRC screening program since 2009 (Indicators of the oncological service of the Republic of Kazakhstan for 2018). Also in the Republic of Kazakhstan there is a tendency for an annual increase in the number of cases of CRC in people under the age of 50; for the period 2008-2017 3121 CRC patients under the age of 50 (inclusive) were registered. Despite the fact that the growth rate of the absolute number of diseases in people under the age of 50 years is different, there is a pattern of increasing cases of CRC, on average by 2.3% annually (Indicators of the oncological service of the Republic of Kazakhstan for 2008-2017). In this contingent, the disease manifests itself before the threshold age of screening (50 years), and by the time of diagnosis, there is usually a widespread tumor process (O'Connell J.B., 2004).

As a cohort studies result, it was proved that young patients are characterized by tumor localization in the distal colon, "late" stages of the disease, the proportion of which, according to some authors, is more than 70%, as well as the aggressive nature of the course and low degree of tumors differentiation (Gupta S., 2010; Negri F.V., 2005). In turn, such phenotypic characteristics of "aggressiveness" as relatively rapid tumor growth and metastasis, low degree of differentiation and the response to therapy with epidermal growth factor receptor inhibitors are due to specific genotypic differences (Semyanikhina A.V., 2020; Vladimirova L.Yu., 2016).

Multicenter studies in the USA, European countries, South Korea, Japan, China on representative cohorts of the corresponding ethnic subpopulations have demonstrated the associativity of individual genotypes with the risk of developing CRC at a young age, taking into account the multifactorial nature of the disease (Johansen Taber K.A., 2014; Abecasis G.R., 2010; Lin P.H., 2016). Molecular genetic studies have revealed such fundamental properties of colorectal neoplasms in young patients as a low level of p53 protein expression, a higher level of DNA replication errors, expressed as microsatellite instability (MSI), specific mutations in the *KRAS* gene, and a higher ploidy index

compared with tumors from patients in general population (Negri F.V., 2005; Belyayeva A.V., 2012).

Young patients' genome analysis based on next generation sequencing (NGS) allows identifying cases of colorectal cancer of syndromic (~5%) and sporadic (~95%) nature (Semyanikhina A.V., 2020). This makes it possible to stratify the level of risk for a patients and their blood relatives, to modify diagnostic and therapeutic approaches and follow up observation. Given the evidence for the genes functional significance and phenotypes overlapping in hereditary syndromes associated with CRC, as well as cases where cancer can be caused by mutations in more than one gene, multigenic testing (MGT) is a cost-effective method of molecular genetic analysis, because it allows simultaneously analyze multiple genes, including those related to prognosis and response to therapy (Stanislaw C., 2016; Gallego S.J., 2015).

To date, the results of studying this area in the Republic of Kazakhstan are presented in single publications.

Thus, an increase in the incidence of CRC in people under the age of 50, the lack of early diagnosis in familial and hereditary forms of CRC, high rates of advanced cancers and low survival rates (according to the pilot studies data stage III and IV, cumulatively and 5-year survival rate is 59.4 % and 39.5%, respectively), necessitate the study of the disease molecular genetic basis.

Aim of the study is optimization of colorectal cancer early diagnosis and prevention in persons under the age of 50 and with a genetically burdened anamnesis.

Research objectives

1. To study the epidemiological aspects of CRC in people under the age of 50 in the Republic of Kazakhstan.
2. To conduct a comparative analysis of CRC clinical and phenotypic features in groups of patients under the age of 50 and over 65 years.
3. Form a register of patients under the age of 50 with the DNA bank creation.
4. To conduct a molecular genetic study of the DNA of patients under the age of 50 based on next generation sequencing.
5. To develop algorithms for colorectal cancer early diagnosis and prevention in people under the age of 50 and with a genetically burdened anamnesis based on molecular genetic analysis.

Objects of the study are patients with CRC under the age of 50 with a verified diagnosis, with various stages of the disease.

Subjects of the study are hereditary, familial and sporadic variants of colorectal cancer in patients under the age of 50.

Study methods: clinical, clinical and genealogical (charting the inheritance of the disease), endoscopic (colonoscopy), morphological (histological examination), molecular genetic (next generation sequencing), bioinformatic analysis, statistical analysis.

Scientific novelty of research:

For the first time, a comprehensive epidemiological assessment of CRC incidence in people under the age of 50 over a 10 years period was carried out on a national scale and by regions, taking into account gender and age characteristics.

For the first time in the Republic of Kazakhstan, CRC clinical and phenotypic features were studied in people under the age of 50 and over 65 years, taking into account gender, stage of the disease, histology and tumor localization.

For the first time in the Republic of Kazakhstan, based on next-generation sequencing, a molecular genetic analysis of pathogenic mutations and variants with uncertain functional significance associated with CRC in patients under the age of 50 years was carried out.

For the first time, 17 new, previously undescribed mutations associated with colorectal cancer have been identified.

For the first time in the Republic of Kazakhstan, algorithms for colorectal cancer early diagnosis and prevention in persons under the age of 50 and with a genetically burdened anamnesis based on molecular genetic analysis have been developed.

The main provisions for defense

Epidemiological analysis revealed an annual increase in the number of colorectal cancer cases by an average of 2.3% among people under the age of 50 with a trend towards an increase in the widespread tumor process proportion (III, IV stages).

Pathogenic mutations frequency in genes with high penetrance (*APC, MLH1, MSH2, MSH6, BRCA1, BRCA2*) is significantly higher in patients with a family history (21.1% vs. 3.1%; p=0.0002) and multiple primary tumors (20.0% vs. 3.1%; p=0.0004) compared with patients with sporadic CRCs (3.1%).

Pathogenic mutations that cause colorectal cancer development have been identified in gene regions (coding for functionally significant regions of the corresponding proteins, in promoter and regulatory sequences) characterized by a high level of conservatism. Mutations in the most conserved regions were found in genes that provide genome stability through the mechanisms of tumor growth suppression (*APC*), mutation level control, DNA repair and maintenance of genetic stability (*BRCA2*), cell cycle control (*CHEK2*), regulation of gene expression (*DICER1*), maintenance of chromosomal stability (*FANCI*), DNA mismatch repair (*MSH2, PMS2*), and regulation of apoptosis and response to oxidative stress (*MUTYH*).

9 new pathogenic mutations not previously described in the literature in the *FANCI, APC, BMPR1A, ATM, DICER1, NBN* genes and 8 new variants with a deleting effect according to *in silico* analysis in the *ATM, NSD1, RB1, FANCD2, BLM, MSH2, DICER1, PMS1* genes were identified. This determines the necessity for its further study as causal molecular genetic changes leading to an earlier colorectal cancer development compared to the general population.

On the basis of molecular genetic analysis, algorithms have been developed and implemented in practice that allow colorectal cancer early diagnosis and prevention in young people and those with a hereditary burdened anamnesis with the implementation of therapeutic measures and individualized follow up observation.

Practical significance of the study

Algorithms for colorectal cancer early diagnosis and prevention in persons under the age of 50 and with a genetically burdened anamnesis based on molecular genetic analysis have been developed and implemented in practice.

Algorithms for individualized dispensary observation of patients with polyposis and non-polyposis syndromes and their blood relatives have been developed and implemented in practice.

Multigene testing panel to assess the CRC risk at a young age have been improved.

Algorithms for colorectal cancer early diagnosis and prevention in people under the age of 50 and with a genetically burdened anamnesis can be further used in all institutions providing oncological care in the Republic of Kazakhstan.

Conclusions:

1. In the Republic of Kazakhstan, there is an increase in the incidence of colorectal cancer among people under the age of 50 years. The annual increase in the number of cases is 2.3%. The highest standardized incidence rates were noted in North Kazakhstan, Pavlodar and East Kazakhstan regions, the lowest – in South Kazakhstan, Zhambyl and Kyzylorda regions. At the same time, the standardized indicator in North Kazakhstan region is 2.2 times higher than this indicator in the Turkestan region and 1.9 times – in Kyzylorda region.

2. In patients under the age of 50 years, compared with patients over 65 years, 15% more cases of locally advanced tumor process (stage III) were noted. In patients under the age of 50, stage IV disease is 27.3% more common in men than in women. A trend towards an increase in the frequency of tumors localized in the right half of the colon with increasing age of patients was revealed.

3. In the course of the study, a register of patients with colorectal cancer under the age of 50 and a DNA bank was created on the basis of the Institute of Genetics and Physiology of the Ministry of Education and Science of the Republic of Kazakhstan.

4. The frequency of pathogenic mutations in genes with high penetrance (*APC, MLH1, MSH2, MSH6, BRCA1, BRCA2*) is significantly higher in patients with a family history (21.1% vs. 3.1%; p=0.0002) and primary multiple tumors (20.0% vs. 3.1%; p=0.0004) compared with patients with sporadic cancers (3.1%). Most often, pathogenic mutations and variants with a deleting effect were noted in genes associated with Lynch syndrome (13%), Li-Fraumeni syndrome (13%), familial adenomatous polyposis (9.67%), familial breast and ovarian cancer (6.4%) and in the genes of the complementation group of Fanconi anemia (6.4%).

5. 17 new mutations in genes with high and moderate penetrance not previously described in the literature have been identified, which need to be validated in case-control studies as causal changes leading to an earlier development of colorectal cancer compared to the general population.

6. Algorithms for early diagnosis and prevention of colorectal cancer in individuals under the age of 50 and with a genetically burdened anamnesis have been developed and implemented, which should be used in the presence of a family history of CRC and CRC-associated tumors in order to identify causative mutations and early cancer detection in patient's blood relatives.

Personal contribution of the author

The author analyzed intensive and standardized incidence rates of CRC in the world and in the Republic of Kazakhstan, taking into account the age aspect. The data from electronic resources of the oncological service (hospital and population cancer registries,

electronic register of cancer patients) also have been analyzed by the investigator. The author's personal contribution is also in creating study design and protocol, patients recruiting and treating, collecting blood samples, developing a patient's questionnaire and an individual registration card, forming a database and a DNA bank, analyzing, summarizing and interpreting the results of a molecular genetic study (DNA sequencing), statistical processing of research results.

Connection with the main scientific programs

The dissertation study was carried out within the framework of the following scientific programs:

1. "Development of an epigenetic test system for the diagnosis of colorectal cancer in Kazakhstan" within the framework of the state order under the budget program 055 "Scientific and/or scientific and technical activities". Financing Ministry of Education and Science of the Republic of Kazakhstan. Agreement for the implementation of research work: "Life Sciences" No. 196 dated 04/29/2016 for the implementation of grant projects and "Intellectual potential" No. 24 dated 02/12/2015. The research period is 2015-2017.

2. "New molecular genetic methods for pre-symptomatic diagnosis and treatment of a number of significant diseases (Molecular genetic analysis of colorectal cancer based on sequencing of a new generation)" under the budget program 013 "Applied scientific research in health care". Financing Ministry of Health of the Republic of Kazakhstan. Agreement for the performance of research work No. 164/07.17 dated July 10, 2017, No. 026-17-mu dated July 17, 2017. The research period is 2017-2019.

Approbation of the research

The results and main provisions of the dissertation were presented at the following events:

1. VI Congress of Oncologists and Radiologists of the Republic of Kazakhstan "Integration of technologies and knowledge" (Republic of Kazakhstan, Almaty, April 28, 2017);

2. VII Congress of Oncologists and Radiologists of the Republic of Kazakhstan (Republic of Kazakhstan, Nur-Sultan, October 18, 2019);

3. International Congress Molecular Analysis for Personalized Therapy Congress 2019 (UK, London, November 8, 2019);

4. Republican meeting of the heads of specialized clinical diagnostic departments of oncological centers (Republic of Kazakhstan, Shymkent, February 4, 2020);

5. International scientific and practical conference "Forum Ural Onco 2020" (Russian Federation, Yekaterinburg, March 12, 2020);

6. International scientific and practical conference "Cancer care in Kazakhstan. From past to the future" (Republic of Kazakhstan, Almaty, December 11, 2020);

7. VIII with international participation Congress of Oncologists and Radiologists of the Republic of Kazakhstan (Republic of Kazakhstan, Turkestan, October 15, 2021).

Publications:

Based on the dissertation materials, 28 papers were published, including 7 articles in journals recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan, 3 articles in foreign peer-reviewed publications indexed in the scientific information databases Scopus

and Web of Science – Tumor Biology (40% percentile, Q2, impact factor 2.926), South African Medical Journal (44% percentile, Q2, impact factor 1.500), Frontiers in Oncology (46% percentile, Q2, impact factor 4.848), 2 abstracts in foreign peer-reviewed publications indexed in the Scopus and Web of Science – FEBS Journal (Q2, impact factor 3.986), Annals of Oncology (98% percentile, Q1, impact factor 18.279), 3 abstracts in the international scientific and practical conferences for young scientists, 2 articles and 7 abstracts in the materials of international scientific conferences and congresses.

The volume and structure of the thesis

The thesis is presented on 220 pages. Dissertation consists of an introduction, 4 sections (literature review, materials and methods, results, discussion), conclusion, list of references, 7 applications. The text contains 31 tables and 64 figures. The list of references includes 267 sources.