UNIVERSITY OF MEDICINE AND PHARMACY CRAIOVA

DOCTORAL SCHOOL

THE ROLE OF microRNA POLYMORPHISMS IN GASTRIC CANCER PATHOGENESIS

PhD THESIS ABSTRACT

SCIENTIFIC ADVISOR:

PROF. UNIV. DR. ION ROGOVEANU

PhD student:

CÎMPEANU RALUCA ALEXANDRA

CRAIOVA

2017

TABLE OF CONTENTS

KEY WORDS

gastric cancer, pathogenesis, Helicobacter pylori, polymorphism, micro-ARN.

INTRODUCTION

Gastric cancer is a major health problem worldwide, due to its increased incidence and high related mortality and morbidity.

Gastric carcinogenesis is a complex, multistage process that begins with chronic inflammation and proceeds to chronic atrophic gastritis, intestinal metaplasia, dysplasia and invasive adenocarcinoma. Emerging evidences suggest that Helicobacter pylori infection and the individual genetic susceptibility are key factors in the initiation and progression of gastric cancer.

Genetic variability, through changes in biological pathways controlled by genes involved in processes of cell growth and differentiation, plays an important role in determining individual cellular response under the influence of environmental factors.

Recents studies have reported the role of small noncoding RNA molecules, called micro-RNA, in the pathogenesis of human cancers. More than 50% of genes coding for miARN are located in chromosomal regions associated with neoplastic processes. Multiple research suggests that miRNA regulates a wide range of biological processes including cellular development, differentiation, proliferation and apoptosis and the lose of normal function plays an important role in neoplastic processes.

Single nucleotide polymorphism are a type of common genetic variation associated with population diversity, disease susceptibility, drug metabolism and genome evolution. SNPs may affect the expression and function of miRNAs, either by overexpression of mRNA and their action as oncogenes, or by miRNA subexpression and their action as tumor suppressor genes. Thus, miRN plays an important role in the occurrence and progression of human cancers, including gastric cancer, and the changes in their expression are associated with the multistage progression of malignant gastric tumors, diagnosis, prognosis and sensitivity to cytotoxic treatments.

I. CURRENT STATE OF KNOWLEDGE

I.1. EPIDEMIOLOGY

Gastric adenocarcinoma was the leading cause of cancer related death worldwide in the 20th century. Nowadays, is the fourth more common cause of cancer after lung cancer, breast cancer and colorectal cancer, with an estimated 951,600 new cases diagnosed annually, representing 6.8% of all malignant tumors. A total of 723,100 deaths due to this malady were recorded annually worldwide. Of these, approximately 140,000 new cases annually and 107,000 deaths occur in Europe (Torre et al., 2012).

The incidence increases with age, with a peack of incidence observed between 60 and 80 years. The mean age of diagnosis of neoplasia is 69 years for men and 73 years for women. In younger patients under 30 years, this type of cancer is rarely encountered. The disease has a significantly higher prevalence in males in almost all countries, with rates of two to four times higher incidence in males than in women (Forman et al., 2013).

In Romania gastric cancer ranks eighth in the incidence of malignant tumors, with an incidence of 26-28 cases per 100,000 inhabitants, slightly decreasing in the last years.

I.2. ANATOMICAL PATHOLOGY

Gastric cancer defines any malignant neoplasia that develops in the region between the gastroesophageal junction and the pylorus. Approximately 95% of all stomach cancers are adenocarcinomas: diffuse, papillary, tubular, mucinous or intestinal. The term of gastric cancer mainly reffers to this type of tumor. The most common localization of gastric cancer is in the antro-pyloric region (50-60% of cases), followed by the small curve (20%), the gastric body, the cardio-tuberosity region, and the large curvature (DeVita et al. 2008)

The most widely used classification of gastric cancer is the Lauren classification published in 1965 and reviewed in 1995 by Carneiro et al., which divides gastric cancer into two distinct histological subtypes: diffuse type and intestinal type.

I.3. RISK FACTORS IVOLVED IN GASTRIC CANCER PATHOGENESIS

Gastric cancer can be considered a multifactorial disease due to involvement in carcinogenesis of inherited or acquired risk factors, including host genetic profile, infectious agents such as H. pylori, or eating habits. Multiple studies have shown a strong association between H. pylori infection and the increased incidence of gastric cancer in areas where the infection is endemic. Two thirds of gastric cancers are attributed to H. pylori infection, the bacterium being classified as a class I carcinogen (Moss et al., 2017).

Nutritional factors such as a diet poor in fruits and vegetables, but rich in preserved, salted and smoked foods along with obesity, smoking and alcohol consumption are environmental factors considered to play an important role in the pathogenesis of gastric cancer.

Gastric cancer is sporadic in 90% of cases, but 10% of affected individuals have a family history of gastric cancer. A series of genetic syndromes are associated in varying proportions with the risk of developing this type of neoplasia: hereditary diffuse gastric cancer, Lynch syndrome, familial adenomatous polyposis, Li-Fraumeni syndrome, Peutz-Jeghers syndrome or juvenile polyposis.

I.4. DIAGNOSTIC

Early gastric cancer is asymptomatic in 80% of cases, leading to diagnosis mainly in advanced stages of the disease. When symptoms are present, they tend to mimic the ulcer disease. In advanced gastric neoplasms the most common symptoms are: abdominal pain, weight loss, nausea and vomiting, fatigue, physical asthenia, anorexia, dysphagia, early satiety, occult gastrointestinal bleeding.

The study of the epigenetic changes involved in the initiation and progression of the gastric carcinogenesis process led to new discoveries regarding pathogenesis, the identification of new biomarkers and potential therapeutic targets, with significant benefits in patient prognosis through the development of individualized therapeutic strategies.

miRNA molecules represent a biomarker that can be determined from serum or plasma, whose role in screening, diagnosis and prognosis in various neoplasias is extensively

studied lately. A substantial number of miARN molecules have been shown to have different expression in patients with gastric cancer. Genes encoding these molecules act as protooncogenes or tumor suppressor genes, and changes in miRNA expression lead to cellular proliferation, confer resistance to apoptosis, and facilatate the process of invasion and metastasation (Jiang et al., 2015). By the post-transcriptional regulatory function, miRNA molecules are involved in a series of cellular processes including cell development, differentiation, proliferation and apoptosis.

II. PERSONAL CONTRIBUTIONS

II.1. MOTIVATION, PURPOSE AND STUDY DESIGN

Gastric cancer (CG) is one of the most common malignancies and a major cause of morbidity and mortality worldwide. In the countries of Eastern Europe, including Romania, the incidence and mortality due to gastric cancer continues to be high.

Genetic variability, by altering the biological pathways controlled by genes involved in cell growth and differentiation processes, plays an important role in determining individual cellular response under the influence of environmental factors.

Recent studies have reported the involvement of small, noncoding RNA molecules called microRNAs, in the pathogenesis of human cancers, more than 50% of the miRNA-encoding genes being located in the chromosomal regions associated with cancer. miRNAs can function either as oncogenes which are usually overexpressed, or as tumor suppressor genes that are under-expressed. miRNA plays an important role in the development and progression of human cancers, including CG, changes in their expression being associated with multistage progression of gastric tumors (initiation, progression, invasion, metastasis), GC diagnosis and prognosis (Tong et al. 2014).

According to published literature, the most studied SNPs located in miRNA-encoding genes to evaluate the miRNA-CG risk association are miR-27a rs895819, miR-149 rs2292832, miR-146a rs2910164, miR-196a2 rs11614913 and miR- 499 rs3746444. A large number of studies have analyzed these genetic variants primarily in Asian populations and only few data is available in Caucasian populations.

Our aim was to investigate the association of the five most studied polymorphisms located in genes coding for miRNA and the risk of gastric cancer in a population in Eastern Europe / Romania, an area where these SNPs have not been analyzed, as well as the identification of new polymorphisms possibly associated with gastric cancer, in order to identify new potential markers useful in the early diagnosis of gastric cancer.

In order to achieve this goal, the following objectives were pursued:

- determining the genotype frequency of the main genes which encodes for miRNA,
- analysis of mononucleotide polymorphisms localized in genes encoding miRNA that are involved in malignant pathologies and establishing possible associations between selected genes and CG risk,
- correlation of the investigated gene genotypes with gastric tumor location: cardiac or non-cardiac and histological type: intestinal or diffuse.

II.2. PATIENTS AND METHODS

The study was conducted on a group of 430 Romanian subjects, comprising 142 patients with gastric cancer: histopathologically confirmed adenocarcinoma and 288 healthy patients who represented the control group. The biological material used was peripheral venous blood harvested in anticoagulant tubes.The samples were encoded and anonymized according to the confidentiality rules. From all harvested venous blood samples, genomic DNA was isolated to analyze the polymorphisms: miR-27a rs895819, miR-149 rs2292832, miR-146a rs2910164, miR-196a2 rs11614913 and miR-499 rs3746444.

The evaluation of the selected SNPs was carried out in the Human Genomics Laboratory, part of the Center for Research in Gastroenterology and Hepatology of the University of Medicine and Pharmacy of Craiova.

The steps taken in the genotyping protocol were:

• isolation and purification of genomic DNA from venous blood

• quantitative and qualitative analysis by the spectrophotometric method of isolated DNA

• Real Time PCR with TaqMan probes to identify allelic variants

• analyzing and interpreting results in a clinical context.

The isolation of genomic DNA was performed from peripheral venous blood samples using the manual work protocol provided by the Wizard® Genomic DNA Purification Kit. This protocol has four main steps in obtaining genomic DNA. The first step is lysis of cells and the nucleus. The second step involving RNA digestion with the enzymes called RN-aze is optional. In the third step, the proteins released from the cells are removed by precipitation with a saline solution, the DNA is concentrated and desalted by precipitation in isopropanol. The last step is washing and dehydration in ethanol, and finally the DNA is rehydrated.

Genotyping was performed by Real time PCR reaction with Taqman probes by a technique that follows a three-step sequence: denaturation, primer attachment, and extension. The protocol used to identify mononucleotide polymorphisms in this study used TaqMan® Universal Master Mix and TaqMan® SNP Genotyping Assays specific for each of the studied polymorphisms. They identify specific polymorphisms by using 5 'nuclease activity by amplifying and detecting specific SNP alleles in genomic DNA samples. The presence or absence of SNPs is determined by changing the fluorescence of the dyes associated with the probes. When only one fluorescent signal is present, the state is homozygous either for allele 1 or allele 2. The appearance of fluorescent signals for both dyes defines the heterozygous state.

The identification of miRNA mononucleotide polymorphisms - miR-27a rs895819, miR-149 rs2292832, miR-146a rs2910164, miR-196a2 rs11614913 and miR-499 rs3746444 was performed using the Real Time ViiA7 system.

Reactions were run in 384-well plates, the total volume per reaction was 5 µl: 2.5 µl of DNA was transferred to 2.5 µl reaction mixture composed of: Universal Master Mix and TaqMan® SNP Genotyping Assays 40x.

Interpretation of the results was achieved with the software software - ViiA ™ 7 Software v1.0 and the Allelic Discrimination option.

Data analysis was performed with the Data Analysis module of Microsoft Excel 2007 and the GraphPad Prism 6 software.

II.3. RESULTS

A total of 430 subjects in the SouthWestern region of Romania were enrolled in this study, 142 patients were diagnosed with CG and formed the study group and 288 healthy subjects (no CG or other neoplasias) of the same ethnicity and geographical origin made the control group. CG cases were diagnosed according to standard procedures (clinical examination, laboratory assessments, superior digestive endoscopy, imaging evaluations) and confirmed by the anatomo-pathological examination of the tissue sample taken either by biopsy or by surgical resection. The selection of control cases was based on the inclusion and exclusion criteria set and so that there were no significant differences in age and male / female ratio with the CG group.

The first stage of the bistatistic analysis of genotypes identified in the two study groups was the HWE Deviation Test. By comparing the frequencies of the observed and expected genotypes by the scanned test, we evaluated the Hardy-Weinberg equilibrium relationship. No relevant HWE deviations were found for any of the polymorphisms tested in the control group ($p > 0.05$, $X2 < 3.84$).

Polymorphism miR-27aT> C rs895819

The analysis of the fluorescence signal emitted by the TaqMan probe allowed identification of the two alleles: allele C - VIC signal and allele T - FAM signal and determination of the three genotypes: homozygote CC, heterosigot CT and homozygote TT.

Statistical analysis has shown that the miR-27aT> C, rs895819 polymorphism is associated with an increased risk of gastric cancer when one genotype was compared to another (the most numerous genotype served as a reference - TT) or when compared carriers of the C-allele. Thus, CC genotype carriers have an increased risk of gastric cancer of about 1.98 times higher than those with the TT genotype ($p = 0.036$); Also, C-allele carriers have an increased risk of gastric cancer (OR 1.95, 95% CI: 1.07-3.55). Separate analysis based on tumor localization revealed that CC genotype carriers have an increased risk of 2.06 times for non-cardiac gastric adenocarcinoma. Separate analysis according to the histological type revealed that CC genotype carriers have an increased risk of 2.27 fold for intestinal gastric adenocarcinoma.

Polymorphism miR-149 C> T rs2292832

The analysis of the fluorescent signal emitted by Taqman probe allowed identification of the two alleles: allele C - VIC signal and the allele T - FAM signal and the setting of the 3 genotypes: homozygous CC, heterozygous CT, and homozygous TT. By comparative analysis of genotypes (according to international databases we took genotype CC as reference) and the statistical data obtained, it resulted that the miR-149 C> T rs2292832 polymorphism is not associated with an increased risk of developing gastric malignancies when one genotype was compared to another (the most numerous served as a reference - CC) or when the T-allele carriers were compared. Statistical analysis based on tumor location or histopathological type revealed no statistically significant differences.

Polymorphism miR-146a G> C rs2910164

According to the result obtained based on the fluorescence signal interpretation, specifically emitted by the TaqMan probe, the 2 alleles for the polymorphism miR-146a G> C rs2910164: the G allele - the FAM signal and the VIC allele C-allele were identified. Thus, the three genotypes were established: homozygous GG, heterozygous GC and homozygous CC. The GG genotype is best represented in both the study and the control group, which is confirmed in both the literature and international databases (NCBI). From the comparative analysis of the genotypes (according to international databases we took genotype GG as reference) and the statistical data obtained it was found that the miR-146a G> C, rs2910164 polymorphism is not associated with an increased risk of developing CG when one genotype was compared to another (the most numerous served as a reference - GG), or when the carriers of the risk allele C were compared. Statistical analysis based on tumor location or histopathological type did not reveal statistically significant differences.

Polymorphism miR-499 A> G rs3746444

Following the interpretation of the fluorescence signal specifically emitted by the TaqMan probe, the 2 alleles for the miR-499 A> G rs3746444 polymorphism were identified: the G allele - the VIC signal and the FAM allele A - allele. Thus the three genotypes were established: homozygous AA, heterozygous AG and homozygous GG. From the distribution of genotypes identified among patients diagnosed with CG and subjects in the group, the AA

genotype is best represented in both the study group and the control group, which is confirmed in both the literature and the international databases. Comparative analysis of genotypes (according to international databases we took genotype GG as reference) and the statistical data obtained showed that the miR-499 A> G, rs3746444 polymorphism is not associated with an increased risk of developing CG when one genotype was compared to another (the most numerous served as a reference - AA), or when compared to carriers of the risk allele G. Statistical analysis based on tumor location or histopathological type did not reveal statistically significant differences.

Polymorphism miR-196a2 C> T rs11614913

According to the result obtained on the basis of the fluorescence signal interpretation, specifically emitted by the TaqMan probe, the two alleles for the miR-196a2 C> T rs11614913 polymorphism were identified: the T allele - the FAM signal and the CIC - the VIC signal. Thus, the three genotypes were established: homozygous CC, CT heterozygous and homozygous TT. From the analysis of the distribution of genotypes identified among the patients diagnosed with CG and the subjects in the control group, the CC genotype is best represented in both the study and the control group, a fact confirmed both in the literature and in the databases The comparative analysis of the genotypes and statistical data obtained showed that the miR-196a2 C> T rs11614913 polymorphism is not associated with an increased risk of developing CG when one genotype was compared to another (the most numerous served as reference - CC) or when the T-allele carriers were compared. The statistical analysis based on the location of the tumor or the histopathological thymus did not reveal statistically significant differences.

II.4. DISCUSSIONS

In this study, we evaluated whether five mononucleotide polymorphisms located in functional regions of genes encoding miARN (miR-27a rs895819, miR-149 rs2292832, miR-146a rs2910164, miR-196a2 rs11614913 and miR-499 rs3746444) influence the susceptibility to GC in a roumanian population. In the literature, there are very few studies conducted on Caucasian populations regarding the association between these SNPs and the risk of developing GC in these individuals.

This study identified a significant association between mir-27a T>C polymorphism, rs895819 (OR 2.27, 95% IC: 1.05-4.92, p = 0.034) and GC susceptibility; carriers of the CC genotype were at a higher risk of developing CG, mainly those with intestinal type. miR-27a seems to act as an oncogene in CG, with BTG2 and prohibitin molecules as biological targets. Regarding the other four miRNA polymorphisms analyzed, we did not identify any statistically significant association between them and susceptibility to CG.

According to data in the literature, this is the first study showing the association between mir-27A rs895819 T>C polymorphism and the risk of developing GC in a European population. The results obtained from our study support the hypothesis that the presence of the C allele in rs895819 can lead to the overexpression of the mir-27a oncogene in the cell and, consequently, to an increased risk of developing CG. A possible explanation for the inconsistency between the results obtained in our study and those already published on populations of different geographic origins could be related to the haplotype. Furthermore, our data may reflect the role of genetic heterogeneity in CG pathogenesis.

This is also the first study that attempted to identify the existence of a correlation between miR-149 rs2292832 C>T polymorphism and the risk of CG occurrence in an Eastern European population (the population of Romania) and the second in a Caucasian population. The results obtained did not identify any association between the miR-149 rs2292832 polymorphism and the susceptibility to CG both in the dominant, recessive or codominant model, nor in the stratified statistical analysis based on localization of the tumor or histologically diagnosed type. The data obtained from this study is similar to the results of several meta-analyses of studies conducted mainly in Asian populations where no significant association between this SNP and the risk of CG has been detected.

II.5. FINAL CONCLUSIONS

This is the first study to evaluate the association between mononucleotide polymorphisms located in genes coding for miRNA molecules and the risk of gastric cancer in Eastern Europe (Romania).

mir-27a T-C polymorphism, rs895819 correlated with an increased risk of gastric cancer, CC genotype carriers showed a 1.98-fold higher risk of developing gastric cancer than TT genotype carriers (p = 0.036); Also, C-allele carriers have an increased risk of gastric

cancer (OR 1.95, 95% CI: 1.07-3.55). Separate analysis based on histological type and localization revealed that CC genotype carriers have an increased risk of 2.06 fold for non-cardiac gastric adenocarcinoma and 2.27 fold for intestinal type. Further studies performed on different ethnic cohorts are needed to clarify the role of this polymorphism in gastric carcinogenesis.

For polymorphisms miR-146a G>C rs2910164, miR-149 C> T rs292832, miR-196a2 C>T rs11614913 and miR-499 A>G rs3746444, no association with gastric cancer was found in the Romanian population in any of the genetic analysis models used (dominant, codominant or recessive). More extensive studies are needed on different ethnic groups to improve the level of knowledge about the role of this miRNA polymorphism in gastric carcinogenesis. This finding may be related to genetic heterogeneity in the pathogenesis of gastric cancer.

The results of our study in the South-West of Romania should be a starting point in conducting multi-center studies on East-European (Caucasian) populations to clarify the role of these polymorphisms in carcinogenesis.