Alfa satelitska RNA i gen MSH2:potencijalni kandidati za dijagnostički/prognostički krvni biomarker raka prostate

Matić, Zrinka

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:382815

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-03-29**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

University of Zagreb

Faculty of Science

Department of Biology

Zrinka Matić

Alpha (α) satellite RNA and *MSH2* gene: potential candidates for diagnostic/prognostic blood's biomarker of prostate cancer

Graduation Thesis

Zagreb, 2019.

This thesis was made at Ruđer Bošković Institute, under supervision of Dr. Đurđica Ugarković and Dr. Isidoro Feliciello from University of Naples Federico II, and in collaboration with prof. dr. sc. Ana Fröbe from University Clinical Hospital Centre (UHC) Sestre Milosrdnice. It was submitted for evaluation to Department of Biology, Faculty of Science in order to acquire the title Master of molecular biology.

BASIC DOCUMENTATION CARD

University of Zagreb

Faculty of Science

Department of Biology

Graduation Thesis

ALPHA (α) SATELLITE RNA AND *MSH2* GENE: POTENTIAL CANDIDATES FOR DIAGNOSTIC/PROGNOSTIC BLOOD'S BIOMARKER OF PROSTATE CANCER

Zrinka Matić

Roosevelt Square 6, 10 000 Zagreb, Croatia

Prostate cancer is the second most common cancer among men in Croatia. The clinically most accepted biomarker for prostate cancer diagnosis is PSA (prostate-specific antigen), but due to the lack of its specificity intensive efforts are currently directed towards a search for alternative diagnostic and prognostic prostate cancer biomarkers. Biomarker is defined as a biological molecule found in blood, other body fluids, or tissues that can be objectively measured and evaluated as a sign of a normal or abnormal biological process and a pathogenic condition/disease. The aim of this research was to check alteration of expression of two proposed biomarker candidates in the blood of prostate cancer samples compared with healthy controls. As observed downregulation was statistically significant, alpha satellite RNA and *MSH2* gene are proposed as candidates for non-invasive diagnostic/prognostic biomarkers of prostate cancer.

(55 pages, 17 figures, 5 tables, 58 references, original in: English)

Thesis deposited in the Central Biological Library

Key words: blood, molecular markers, downregulation, non-invasive method

Supervisor: Dr. Đurđica Ugarković, Prof.

Reviewers: Dr. Đurđica Ugarković, Prof.; Dr. Sven Jelaska, Prof.; Dr. Nada Oršolić, Prof.

Thesis accepted: 13/2/2019

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek Diplomski rad

ALFA (α) SATELITSKA RNA I GEN *MSH2*: POTENCIJALNI KANDIDATI ZA DIJAGNOSTIČKI/PROGNOSTIČKI KRVNI BIOMARKER RAKA PROSTATE

Zrinka Matić

Rooseveltov trg 6, 10 000 Zagreb, Croatia

Rak prostate drugi je najčešći oblik raka kod muškaraca u Hrvatskoj. Klinički najprihvaćeniji biomarker za dijagnozu raka prostate je PSA (prostata-specifični antigen), ali zbog nedostatka njegove specifičnosti intezivni napori trenutno su usmjereni na otkrivanje alternativnih dijagnostičkih i prognostičkih biomarkera za rak prostate. Biomarker se definira kao biološka molekula iz krvi, drugih tjelesnih tekućina ili tkiva koja može biti objektivno izmjerena i evaluirana kao znak normalnog ili abnormalnog biološkog procesa i patoloških stanja/bolesti. Cilj istraživanja bio je provjeriti promjenu ekspresije dvaju predloženih kandidata iz krvi pacijenata s rakom prostate uspoređujući ih s zdravim kontrolama. Budući da je uočena smanjena ekspresija statistički značajna, alfa satelitska RNA i gen *MSH2* predloženi su kao kandidati za neinvazivni dijagnostički/prognostički biomarker raka prostate.

(55 stranica, 17 slika, 5 tablica, 58 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: krv, molekularni markeri, smanjenje ekspresije, neinvazivna metoda

Mentor: Dr. sc. Đurđica Ugarković, red. prof.

Ocjenitelji: Dr. sc. Đurđica Ugarković, red. prof.; Dr. sc. Sven Jelaska, red. prof.; Dr. sc. Nada Oršolić, red. prof.

Rad prihvaćen: 13. 2. 2019.

Contents

1 Introduction 1
1.1 Prostate cancer statistics 1
1.2 Prostate cancer biomarkers 2
1.3 Prostate-specific antigen 2
1.4 Other prostate cancer biomarkers 3
1.5 Satellite DNA 5
1.6 Alpha satellite DNA 5
1.7 Hypothesised mechanisms of amplification and distribution of satellite DNA through genome 6
1.8 Alpha satellite function: relationship with centromere and kinetochore proteins .. 8
1.9 Transcription of alpha satellite DNA 8
1.10 Noncoding alpha satellite RNAs in centromeric and pericentromeric heterochromatin assembly 9
1.11 Satellite DNA transcripts in si-RNA mediated establishment of heterochromatin and gene regulation 10
1.12 Satellite DNAs and RNAs association with cancer 12
1.13 Mismatch repair pathway 13
1.14 Mechanism of mismatch repair in *Escherichia coli* 13
1.15 Mechanism of mismatch repair in human cells 15
1.16 Mismatch repair genes in colon cancer 18
1.17 Mismatch repair genes in prostate cancer 19
1.18 Aims 20
2 Materials and methods 21
2.1 Materials 21
2.1.1 Solutions, reagents and other materials 21
2.1.2 Primers for quantitative polymerase chain reaction (qPCR) analysis 22
2.1.3 Blood samples 23
2.2 Methods 25
2.2.1 RNA isolation 25
2.2.2 Measurement of RNA concentration 26
2.2.3 Reverse transcription (RT) 26
2.2.4 Measurement of cDNA concentration 27

2.2.5 Quantitative polymerase chain reaction (qPCR) 27
2.2.6 Statistical methods 32
3 Results 34
3.1 Expression of *MSH2* gene 34
3.2 Expression of alpha satellite DNA 39
4 Discussion 44
5 Conclusion 48
6 References 49
7 Curriculum vitae 55

List of abbreviations

AMACR - alpha-methylacyl-CoA racemase

ANOVA - one-way analysis of variance

ASR - age-standardized rate

A-T - adenine-thymine

bp - base pair

CEN - centromere

CENP-A - centromere protein A

CENP-B - centromere protein B

CENP-C - centromere protein C

CI - confidential interval

CTCs - circulating tumor cells

dGATC - daughter guanine, adenine, thymine, cytosine

dH2O - distilled water

DMSO - dimethyl sulfoxide

DRE - digitcal rectal exam

e. g. - *exemplī grātiā* (lat.); „for example"

ERG - ETS-related gene

et al - *et alii* (lat.); „and others"

ETS - E26-transformation specific

EU - European Union

EXO - exonuclease

FDA - Food and Drug Administration

gDNA - genomic DNA

GUS - β-glucuronidase

H3K9 - lysine 9 of histone 3

HAC - human artificial chromosome

HNPCC - hereditary nonpolyposis colorectal cancer

HOR - high order repeats

HP1 - heterochromatin protein 1

i. e. - *id est* (lat.); „that is"

indel - insertion/deletion

LS - Lynch syndrome

miRNA - micro RNA

MLH1 - MutL protein homolog 1

MLH3 - MutL protein homolog 3

MMR - mismatch repair

mRNA - messenger RNA

MSH2 - MutS protein homolog 2

MSH3 - MutS protein homolog 3

MSH6 - MutS protein homolog 6

MSI - microsatellite instability

MSS - microsatellite stable

NCBI - National Center for Biotechnology Information

NCCN - National Comprehensive Cancer Network

NCI - National Cancer Insitute

nt - nucleotide

NTC - no template control

PAP - prostatic acid phosphatase

PC - prostate cancer

PCA - prostate cancer antigen

Pol δ- DNA polymerase δ

PSA - prostate-specific antigen

qPCR - quantitative polymerase chain reaction

Q-Q - quantile-quantile

RER - ribonucleotide excision repair

RISC - RNA-induced silencing complex

RITS - RNA-induced transcriptional silencing complex

RNAi - RNA interference

RNAPII - RNA polymerase II

RT - reverse transcription

satDNAs - satellite DNAs

siRNA - small interfering ribonucleic acid

SNP - single-nucleotide polymorphism

SSB - single-stranded DNA binding protein

ssDNA - single-stranded DNA

SUV39H1 - suppressor of variegation 3-9 homolog 1

TCAST - *Tribolium castaneum* satellite DNA

TE - Tris-EDTA

Tm - melting temperature

TMPRS22 - transmembrane serine protease isoform 2

UHC - University Clinical Hospital Centre

US - United States

W-o-L - Window-of-Linearity

1 Introduction

1.1 Prostate cancer statistics

Prostate cancer is the second most common cancer in men worldwide and the fifth cause of cancer death. Prostate tumors can be indolent or very aggressive, often metastasizing to bone and other organs, thereby causing significant morbidity and mortality (Saini, 2016).

According to estimates for 2012, there were more than a million newly diagnosed cases of prostate cancer worldwide (ASR = 30,6/100 000), and 307 481 deaths (ASR = 7,8/100 000). Generally, more developed countries have higher incidence. Only 30 % of prostate cancer cases worldwide are diagnosed in less developed countries. Increasing use of prostate-specific antigen (PSA) affected the observed incidence of prostate cancer in many more developed countries (Wong et al, 2016). Compared to GLOBOCAN 2012 estimates for European countries, Croatia has an intermediate incidence rate of prostate cancer and its ranked 26th of 40 European countries (ASR = 46,2/ 100 000) (Reljić et al, 2018).

The most recent available published data for prostate cancer incidence in Croatia from the Croatian Cancer Registry show that were 1708 new cases of prostate cancer in Croatia in 2014 (ASR = 64,8/100 000) (Reljić et al, 2018).

The data for 2014, available at EUROSTAT web page, also indicate that Croatia is among top ten European Union (EU) countries in prostate cancer mortality. The latest available mortality data for Croatia show that there were 769 deaths resulting from prostate cancer in 2016 (ASR = 38,2/100 000) (Reljić et al, 2018).

Recent data on prostate cancer survival from two major international studies show that survival in Croatia is below the European average. The study showed a 5-year relative survival of 71.2 % in Croatia, and the European average is 83.4 % (Reljić et al, 2018).

Although, there are no reliable data on the uptake of PSA testing in Croatia, the combination of basic epidemiological parameters indicates that there is room for improvement regarding the overall prognosis and risks associated with prostate cancer in Croatian population. The combination of adopting new approaches in finding prostate

cancer earlier in patients who would benefit most from it and following European trends in the diagnostics and treatment of detected prostate cancer should be the way to move forward (Reljić et al, 2018).

1.2 Prostate cancer biomarkers

The National Cancer Institute (NCI) defines biomarker as a biological molecule found in blood, other body fluids, or tissues that can be objectively measured and evaluated as a sign of a normal/abnormal biological process and a pathogenic condition/disease. A biomarker may be used for screening purposes, for disease diagnosis and prognosis, for the evaulation of disease disposition and for the prediction/monitoring of treatment reponses to various therapeutic interventions (Saini, 2016).

The ideal biomarker for clinical use should have three major characteristics: 1) a safe and easy means of measurement, preferably non-invasively; 2) high sensitivity, specificity and positive and negative predictive values for its intended outcome; and 3) improves decision making abilities in conjunction with clinicopathological parameters (Prensner et al, 2012).

The purpose of biomarker testing for prostate cancer is to inform the selection of men who should undergo initial biopsy (American Medical Association, 2017).

The first prostate cancer biomarker, prostatic acid phosphatase (PAP), was described in the 1930s and it has been used as a clinical marker for prostate cancer progression. Serum PAP levels were found to be elevated in metastatic cases (Saini, 2016).

1.3 Prostate-specific antigen

PAP was replaced by prostate-specific antigen (PSA). PSA is a kallikrein-related serine protease produced by the epithelial cells of the prostate gland that is present in normal prostatic secretions and is often elevated in prostate cancer. Currently, PSA is the most prevalently used prostate cancer biomarker by clinicians. In 1994, US Food and Drug Administration (FDA) approved PSA testing for prostate cancer screening in conjunction with a digital rectal exam (DRE) (Saini, 2016).

The introduction of PSA into the prostate cancer diagnostics community also led to its widespread uses as a screening test among asymptomatic men. Subsequently, the proportion of men with metastatic prostate cancer at the time of diagnosis decreased dramatically, a major feat for the prostate cancer community that altered disease management. More men were being diagnosed with prostate cancer, with the majority having early-stage, clinically indolent disease. More men with benign conditions such as inflammation or hyperplasia were also being biopsied. PSA therefore enables the early detection of many latent prostate cancers, the majority of which may never have led to harm. This discrepancy between decreasing disease aggressivness and increasing treatment has led to widespread criticism that prostate cancer is now an „overdiagnosed" and „overtreated" cancer. A subsequent analysis suggested an overdiagnosis between rate of 20 to 42 % (Prensner et al, 2012).

Moreover, treatment of indolent cancer may cause a patient more harm than benefit. Biopsies and prostate cancer treatments have been associated with psychological distress, loss of bodily function, pain and suffering for patients (Prensner et al, 2012).

However, despite its lack of specificity, PSA is still the most clinically accepted biomarker, but because of its limitations intensive efforts are currently directed towards a search for alternative prostate cancer biomarkers (Saini, 2016).

1.4 Other prostate cancer biomarkers

Prostate cancer antigen (PCA) 3 is a prostate cancer-specific antigen, which is located on chromosome 9q21-22. PCA 3 RNA is not detected in normal prostate tissue and exhibits a low expression level in prostatic hyperplastic tissues. PCA 3 RNA serves as a useful biomarker and it was approved by the US FDA in 2012 (Saini, 2016).

Gene fusions involving ETS (E26-transformation specific) transcription factor family member genes (most common *ERG*) with the androgen regulated gene *TMPRSS2* (transmembrane serine protease isoform 2) are frequently encountered in prostate tumors. These gene fusions serve as potential non-invasive urinary biomarkers with high specificity and positive predictive value, but low frequency in some populations (Saini, 2016).

Enzyme alpha-methylacyl-CoA racemase (AMACR) has demonstrated high sensitivity and specificity when tested as a diagnostic biomarker on prostate biopsy tissue. The main problem is that is not suitable for non-invasive detection in urine (Prensner et al, 2012).

Circulating tumor cells (CTCs) are found in the bloodstream in locally aggressive or metastatic disease. Several groups have demostrated an increased abundance of CTCs in the blood of castration-resistant prostate cancer. But, detecting CTCs is currently labor-intensive and expensive (Prensner et al, 2012).

For now, PSA screening still remains the most widely used method, but identification and validation of novel biomarkers to „rule out" lethal prostate cancer at the point of screening is the greatest clinical need, as this may reduce unnecessary interventions that may cause more harm than good (Prensner et al, 2012).

1.5 Satellite DNA

Satellite DNAs (satDNAs) are tandemly repeated non-coding DNA sequences that make large part of almost all higher eukaryote chromosomes, e. g. 50 % of human sequences is made of repetitive sequences (Lander et al, 2001). Genomic regions which are extremely enriched in satellite DNAs are (peri)centromeric regions of chromosomes. Genome sequencing analysis revealed that satDNAs also represent a substantial fraction of noncentromeric regions (Warburton et al, 2008).

1.6 Alpha satellite DNA

Alpha satellite DNA was originally isolated as a highly repetitive component of the African green monkey genome and so far is found in all primate species (Maio, 1971). Alpha satellite DNA is composed of 171-bp AT-rich monomeric repeat units organized in head to tail orientation. Hierarchicaly organized monomers create complex called higher order repeats (HOR) (Figure 1.) (Warburton & Willard, 1995; Rudd et al, 2003). Alpha satellite DNA is present either in the form of HORs or as tandemly repeated 171-bp divergent monomers that lack any organizational pattern. These two types of organization of alpha satellite DNA are typically located near one another, with unordered monomeric alpha satellite often sandwiched between a large block of HOR alpha satellite (Rudd et al, 2003). High order repeats are array- and chromosome-specific and they are repeated hundreds to thousands of times to create homogenous arrays in which the HORs within a given array are 97-100 % identical (Aldrup-MacDonald et al, 2016).

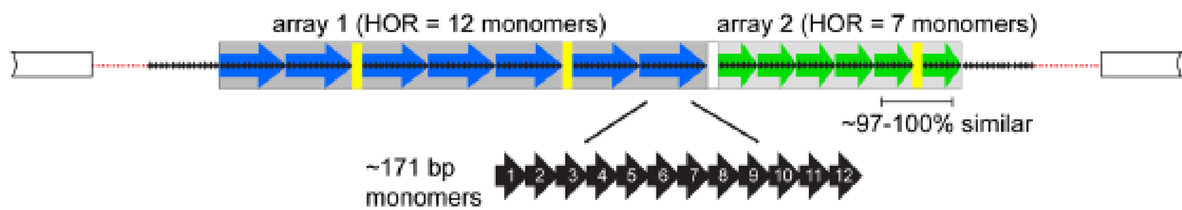


Figure 1. Schematic of the general organization of alpha satellite DNA arrays at human centromere regions. A defined number of individual monomers (black arrows) are arranged tandemly to form a HOR unit, shown here as either 12 monomer HOR (blue array) or 7 monomer HOR (green array) (taken and modified from McNulty & Sullivan, 2018).

1.7 Hypothesised mechanisms of amplification and distribution of satellite DNA through genome

As already mentioned, two different organizations are adopted by alpha satellite DNA. First one is HOR organizational pattern, highly conserved repeat, with 97-100 % sequence identity and the second is monomeric, which is less conserved with 70-90 % sequence identity (Cucheux et al, 2016). Although the mechanisms that gave rise to this diversity and organization are not precisely known, it is commonly accepted that so-called concerted evolution of repetitive sequences is based on different mechanisms of non-reciprocal transfer occurring within or between chromosomes, such as unqual crossover, gene conversion, rolling circle replication and transposon-mediated exchange (Plohl et al, 2008).

Feliciello et al (2006) demostrated for the first time that the structure and evolution of a satellite DNA family is not due to crossing-over and gene conversion, but to a mechanism that maintains the ability of the satellite DNA to form constitutive heterochromatin by replacing altered satellite segments with new arrays generated by rolling circle amplification. The mode of action of this repair process (Figure 2.) not only directly explains the variability of the structure and organization of the satellite repeats, but also accounts for all general features of satellite DNA in eukaryotes, including its discontinuous evolution. This repair mechanism can maintain the satellite structure in a species indefinitely, but also promote a rapid generation of new variants or types of satellite DNA when environmental conditions favor the formation of new species. Thus,

the demostration that satellite DNA is under specific control of a DNA repair mechanism, most probably based on protein-RNA/DNA interactions, indicate that satDNA sequence homogenity is crucial for centromere and more generally, genome stability. It is expected that alteration of this mechanism could potentially give rise to cancerous cells (Feliciello et al, 2006). This specific mechanisms has also been proposed to be the origin of dispersion through the rest of genome of satellite DNA in the form of monomer, dimer and thereof was recently demostrated (Brajković et al 2012; Feliciello et al, 2015).

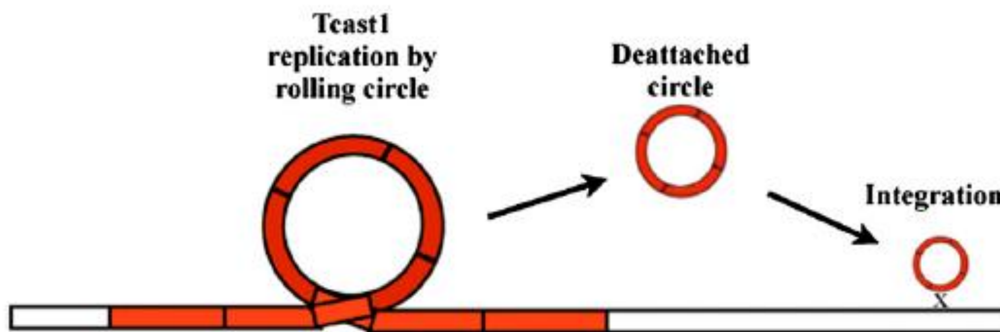


Figure 2. The mechanism of amplification and distribution of TCAST (*Tribolium castaneum* satellite DNA) through genome. Rolling circle replication of TCAST sequences excised from their heterochromatin loci via intrastrand recombination, followed by reintegration into different genome locations by homologous recombination (taken from Brajković et al, 2012).

1.8 Alpha satellite function: relationship with centromere and kinetochore proteins

The chromosomal loci known as centromeres (CEN) mediate the equal distribution of the duplicated genome between both daughter cells. Specifically, centromeres recruit kinetochore, protein complex, that bi-orients the replicated chromosome to the spindle structure (Smurova & De Wulf, 2018).

Satellite repeats contain sequence motifs recognized by protein components for the assembly and organization of centromeres. The best known is CENP-B box, a 17-bp long sequence motif in human alpha satellite DNA and the binding site for centromere protein B (CENP-B), which is suggested to facilitate kinetochore formation (Plohl et al, 2008).

1.9 Transcription of alpha satellite DNA

Transcripts of satellite DNAs have been reported in several organisms including vertebrates, invertebrates and plants. In most species, satellite DNAs are temporally transcribed at particular developmental stages or are differentially expressed in some cell types, tissues or organs (Ugarković, 2005).

In mammals, accumulation of centromeric and pericentromeric transcripts occurs at transcriptional level in the course of proliferation and cell cycle, differentiation of myoblast and in cancer (Pezer et al, 2012).

Given their relatively simple sequence and the lack of any significant open reading frame, previously reported transcription of satellite DNA has been ascribed to read-through from upstream genes and transposable elements. However, promotor elements and transcription start sites, as well as binding motifs for transcription factors, have been mapped within some satellites (Pezer et al, 2012). It was shown that CEN are actively transcribed by RNAPII (RNA polymerase II), that has been detected in yeast *Schizosaccharomyces pombe*, flies and human cells and at centromeric human artificial chromosomes (HAC) (Smurova & DeWulf, 2018).

Two types of alpha satellite transcripts were identified: one corresponds to alpha-repeat RNA, while other is group of mRNAs that contain an alpha-like satellite sequence in their 5' and 3' untranslated regions (Ugarković, 2005).

1.10 Noncoding alpha satellite RNAs in centromeric and pericentromeric heterochromatin assembly

Initial reports of alpha satellite RNA localization suggested that transcripts were confined to the nucleolus until their relocalization to the centromere at the onset of mitosis via CENP-C (Wong et al, 2007). Alpha satellite has also been reported to localize to centromeres in both interphase and metaphase (Ideue et al, 2009).

Firstly, proteins Sgo1 and Aurora B are both key players in mitosis and were identified as alpha satellite RNA-binding partners. In early G1 phase, alpha satellite RNAs produced from the active array help load new CENP-A at the centromere. The presence of alpha transcripts is required for normal cell cycle progression through S and G2 phase. In G2 phase, these RNAs are required for SUV39H1 localization at pericentomere. SUV39H1 enzyme catalyze the addition of methyl groups to lysine 9 in histone 3 (H3K9), which are then binding sites for HP1 (heterochromatin protein 1). HP1 is necessary for establishment and spreading of heterochromatin (McNulty & Sullivan, 2018).

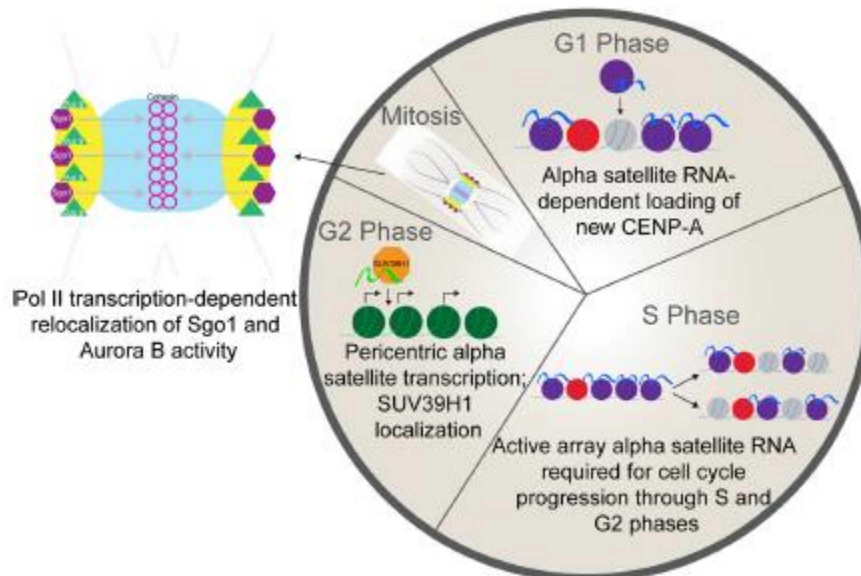


Figure 3. Alpha satellite transcription and noncoding RNAs play distinct roles at the centromere and pericentromere throughout the cell cycle (taken from McNulty & Sullivan, 2018).

1.11 Satellite DNA transcripts in si-RNA mediated establishment of heterochromatin and gene regulation

Epigenetic research over the past years revealed that satellite DNAs are target of silencing mechanisms, such as RNA interference (RNAi) and involving epigenetic modifications. Tandem arrangment of satellite DNAs seems to be responsible for triggering RNAi-based mechanisms and formation of heterochromatin, thereby regulating expression of nearby genes. Expression of nearby genes can as well be influenced by the presence of functional promoters and transcription factor binding sites within satellite sequences. Based on this observations, it is hypothesized that influence of satellite DNAs is epigenetic in its nature and could be modulated by environmental changes (Pezer et al, 2012).

The expression of the satellite DNA of the insect *Tribolium castaneum* TCAST1 is strongly induced by heat shock and proceeds in the form of long double-strand transcripts which are rapidly processed into 21-30 nt siRNAs. Enhanced TCAST1 satellite DNA expression is accompained by an increase in repressive epigenetic modifications of histones at satellite DNA regions in heterochromatin. It was showed that TCAST1 satellite DNA which is dispersed in the close vicinity of protein-coding genes within euchromatin is also subjected to such epigenetic changes induced by TCAST1 siRNAs after exposure to heat stress. So, the role of satellite DNA in the modulation of gene expression after long term heat stress is established through heterochromatization by RNAi inducing satellite DNA-mediated silencing of genes (Feliciello et al, 2015).

Volpe et al (2002) first showed that transcripts derived from tandemly repeated centromeric DNA of *Schizosaccharomyces pombe* exist in the form of small, 20-25 bp long RNAs and are involved in specific chromatin modification through RNA interference (RNAi). The mechanism is initiated by long double stranded RNA that arises from bidirectional transcription of repeated centromeric DNA and is further processed by the RNAse III-like ribonuclease Dicer into siRNAs. siRNAs are then loaded into the RNA-induced transcriptional silencing complex (RITS) through their association with the Argonaute protein. RITS associates with the nascent non-coding centromeric RNA transcript, and binding to the RITS is probably achieved through the base-pairing of siRNA molecules with nascent RNA and by direct contact with the RNA Pol II elongation complex. The association of RITS with chromatin also requires a histone methyltransferase, which methylates histone H3 at lysine 9 (H3K9). That leads to

recruitment of heterochromatin protein 1 (HP1), which is the intial step in formation of heterochromatin.

In summary, siRNAs originating from satellite repeats seem to have an extensive role in gene expression and heterochromatin formation (Figure 4.) (Ugarković, 2005).

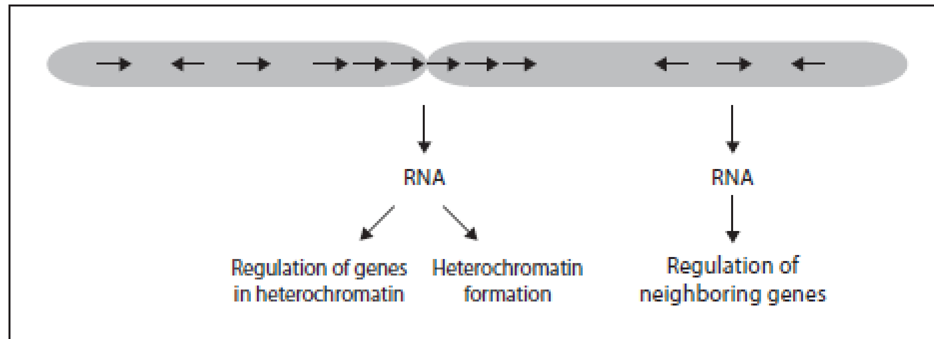


Figure 4. Regulatory role of satellite DNAs and their transcripts (taken from Pezer et al, 2012).

Recently, dispersed elements of alpha satellite DNA were charachterized in human genome. NCBI database research revealed 31 element of alpha satellite DNA elements in euchromatin of which 10 elements was found in introns of genes and 21 elements in intergenic regions (Feliciello, unpublished results). The exact mechanism about the regulatory role of alpha elements in human gene regulation and impact is under investigation in the laboratory where this thesis was made.

1.12 Satellite DNAs and RNAs association with cancer

Proper establishment of the centromere/kinetochore complex is a prerequisite for regular chromosomal segregation in mitosis and meiosis. In addition, pericentromeric heterochromatin, in which the centromere is embedded, is necessary for sister chromatid cohesion and accurate chromosome segregation in mitosis. In the processes of centromere/kinetochore establishment as well as heterochromatin formation, non-coding RNA are playing an important role, so defects in these components or changes in gene expression of satellite RNAs lead to genome instability, chromosome mis-segregation and aneuploidy which is proposed as one of the major inducers of tumourigenesis (Pezer & Ugarković, 2008).

1.13 Mismatch repair pathway

DNA mismatch repair (MMR) pathway is a higly conserved biological pathway that plays a key role in DNA replication fidelity, mutation avoidance and in maintaining genomic stability. One source of DNA damage are exogenous chemicals and endogenous reactive metabolites (reactive oxygen and nitrogen species), while another source of DNA damage are errors that occur during DNA replication, recombination and repair (Li, 2008). The specificity of MMR is primarily for base-base mismatches and insertion/deletion mispairs generated during DNA replication and recombination (Liu et al, 2017).

1.14 Mechanism of mismatch repair in *Escherichia coli*

MMR in *E. coli* was discovered by Meselson group (1976) and it requires the following protein components: MutS, MutL, MutH, DNA helicase II (MutU/UvrD), four exonucleases (ExoI, ExoVII, ExoX and RecJ), single-strand DNA binding protein (SSB), DNA polymerase III holoenzyme and DNA ligase (Li, 2008).

MutS, MutL and MutH initiate MMR and play specialized biological roles in *E. coli.* MutS recognizes base-base mismatches and small nucleotide insertion/deletion mispairs and it possesses intrisic ATPase activity. MutL interacts physically with MutS and activates MutH and also possesses ATPase activity. In *Escherichia coli* DNA is methylated at the N6 position of adenine in dGATC sequences, but in replicating DNA, the daughter strand is transiently unmethylated. Hemimethylated dGATC sites determine the strand specificity of repair and are recognized by MutH, which after activation by MutS and MutL, incises the unmethylated daughter strand. This nick is the initation site for mismatch-provoked excision (Li, 2008).

The three main steps in MMR are recognition, excision and gap-filling DNA synthesis. The mismatch site is recognised by MutS homodimer, then in the presence of MutL, helicase II (UvrD) unwinds the duplex in order to gain single-stranded DNA. SSB bounds to methylated strand (mother strand) in order to protect it from the nucleases. The second step is excision, which is performed by exonucleases, ExoVII or RecJ. The gap

undergoes DNA resynthesis and ligation by DNA polymerase III holoenzyme (Li, 2008). All steps are shown in Figure 5.

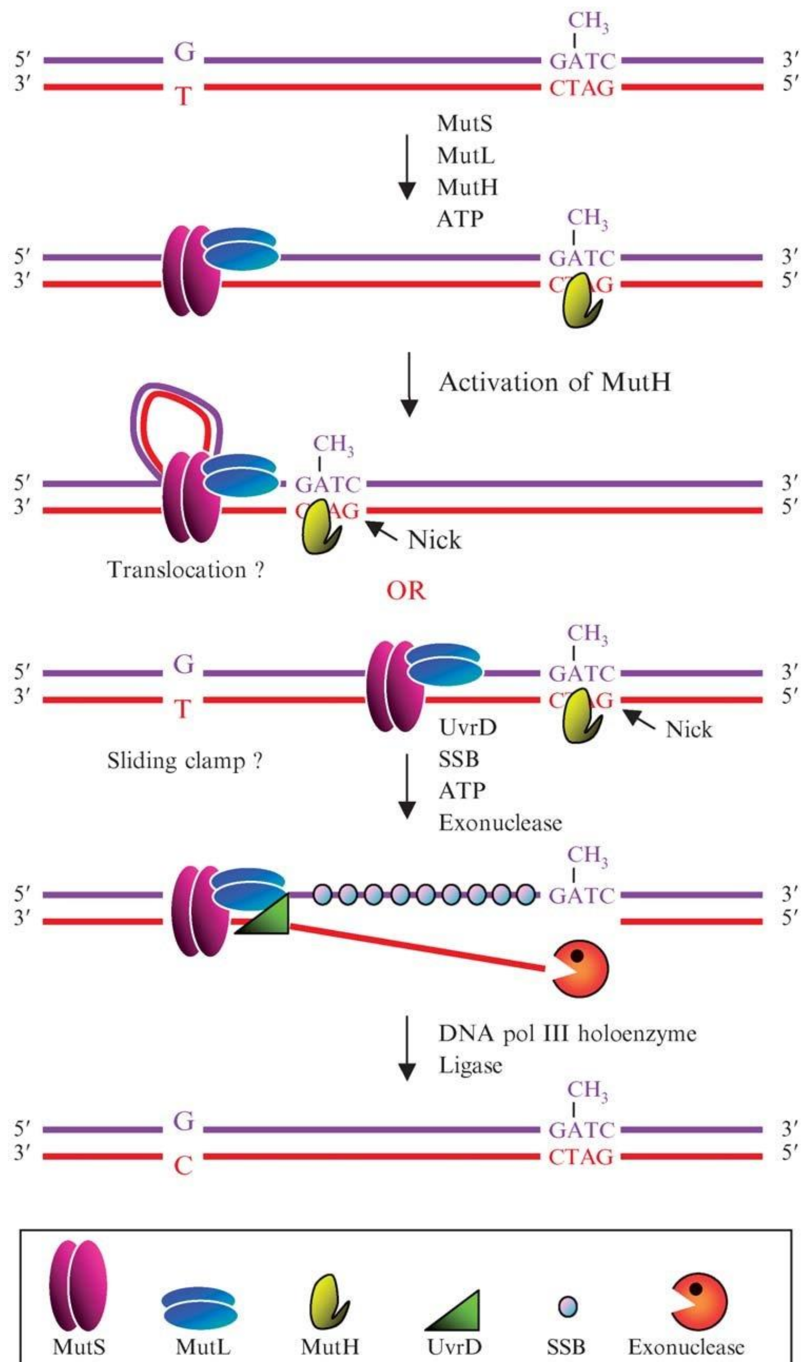


Figure 5. Mechanism of mismatch repair in *Escherichia coli* (taken from Nimesh et al, 2006).

1.15 Mechanism of mismatch repair in human cells

Strong similarities exist between MMR pathway in human cells and prototypical *E. coli*. Several humans MMR proteins have been identified based on their homology to the *E. coli* MMR proteins (Figure 6.) (Li, 2008).

E. coli	Human	Function
$(MutS)_2$	hMutSα (MSH2-MSH6)[a]	DNA mismatch/damage recognition
	hMutSβ (MSH2-MSH3)	
$(MutL)_2$	hMutLα (MLH1-PMS2)[a]	Molecular matchmaker; endonuclease, termination of mismatch-provoked excision
	hMutLβ (MLH1-PMS1)	
	hMutLγ (MLH1-MLH3)	
MutH	?[b]	Strand discrimination
UvrD	?[b]	DNA helicase
ExoI, ExoVII, ExoX, RecJ	ExoI	DNA excision; mismatch excision
Pol III holoenzyme	Pol δ	DNA re-synthesis
	PCNA	Initiation of MMR, DNA re-synthesis
SSB	RPA	ssDNA binding/protection; stimulating mismatch excision; termination of DNA excision; promoting DNA resynthesis
	HMGB1	Mismatch-provoked excision
	RFC	PCNA loading; 3' nick-directed repair; activation of MutLα endonuclease
DNA Ligase	DNA ligase I	Nick ligation

Figure 6. MMR proteins identified based on homology to the *E. coli* (taken from Li, 2008).

In humans, base-base mismatches are recognized by MutSα homolog, heterodimer of MSH2/MSH6 or MutSβ heterodimer of MSH2/MSH3. Heterodimer of MSH2/MSH6 binds to base mismatches or small insertions/deletions (indels) up to 3 nucleotides (nt) in size, while heterodimer of MSH2/MSH3 binds to large indels up to 13 nt in size. At the excision step, MutLα, heterodimer of MLH1/PMS2 is recruited by MSH2/MSH6 or MSH2/MSH3 heterodimer. The subsequent steps of the MMR pathway in human cells are not that understood as in the *E. coli*. MutH protein homolog is not expressed in human cells, but MLH1/PMS2 heterodimer has the latent nuclease acitivity (Erie & Weninger, 2014).

DNA replication is asymmetric and proceeds by different mechanisms on the leading-strand, where DNA synthesis is continuous, and the lagging strand, where DNA synthesis is carried out in short discontinuous segments, known as Okazaki fragments (Kunkel & Erie, 2015).

First proposition is that discontinuities in the lagging strand act as strand discrimination signal in human cells. Second proposition is that asymmetric proliferating cell nuclear antigen (PCNA) directs the strand specific incision. The third proposal suggests that nicks introduced during ribonucleotide excision repair (RER) are used for strand discrimination MMR, because the nicks during RER occur ony in nascent DNA (i.e. daughter strand) (Kunkel & Erie, 2015).

After recognition followed by DNA incision step, exonuclease 1 (EXO1) excises the newly-synthesized DNA strand, and the gap is filled by DNA polymerase δ (Polδ) (Kunkel & Erie, 2015). All steps are shown in Figure 7.

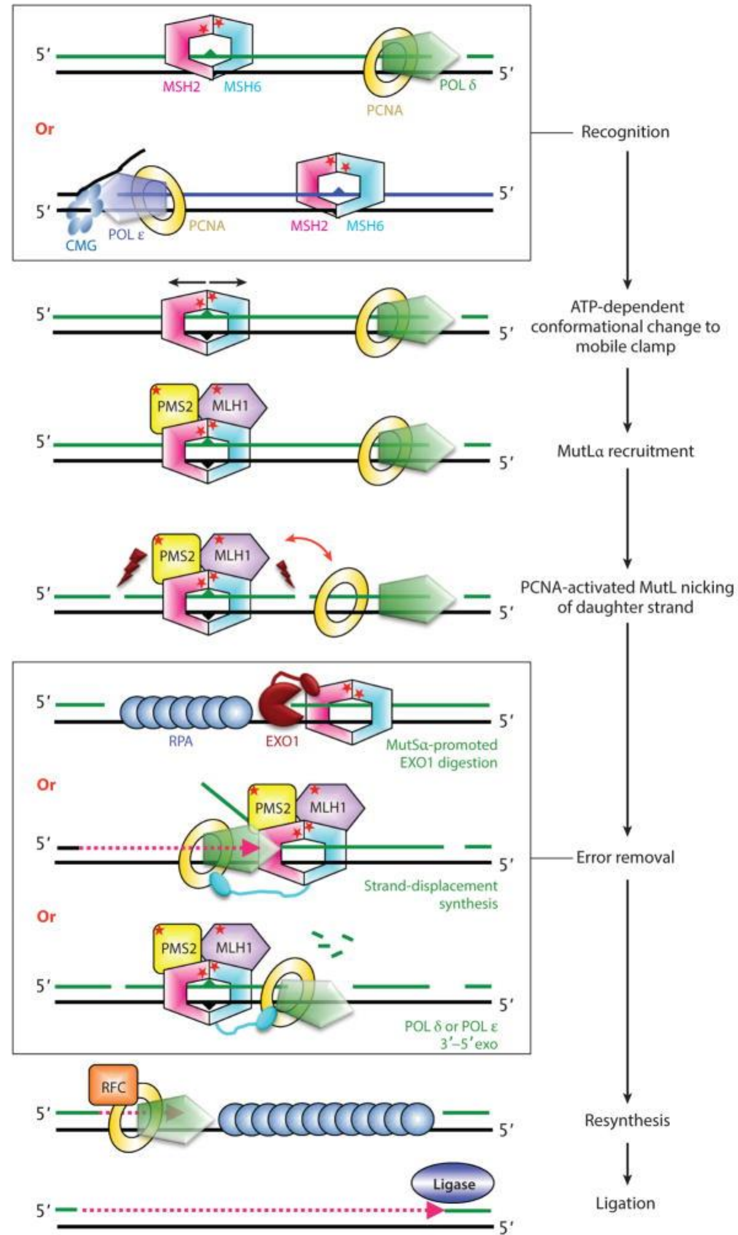


Figure 7. Mechanism of mismatch repair pathway in human cells (taken from Kunkel & Erie, 2015).

1.16 Mismatch repair genes in colon cancer

Mismatch repair genes defects are firstly identified and the most investigated in hereditary nonpolyposis colorectal cancer (HNPCC), also known as Lynch syndrome (LS) and sporadic colon cancer (Li, 2008). LS is an autosomal dominant condition, characterized by early-onset colorectal and endometrial tumor, but also by an increased risk of other tumors, including prostate cancer (Gelsomino et al, 2016; Soravia et al, 2003).

Microsatellite instability (MSI) represents a molecular hallmark of HNPCC. Microsatellites are short DNA motifs of 1-6 bases repeated and distributed through the genome, both in coding and non-coding regions. Because of their repeated structure, they are prone to replication errors, which are normally corrected by the MMR (Gelsomino et al, 2016).

Loss of function of one of the MMR proteins causes MMR system not working correctly and leads to the accumulation of mistakes in microsatellites, which results in a genetic instability. Germline mutations in MLH1, MSH2, MSH6, and PMS2 have been characterized to cause the MSI development (Gelsomino et al, 2016). Sporadic colon cancer is more common type of colon cancer then hereditary colon cancer, with frequency of 85 %. Most of the sporadic colon cancer are microsatellite stable (MSS). Epigenetic silencing of MLH1 and MSH2 via hypermethylation of theirs promoters is demostrated in sporadic colon cancer (Sinicropet & Yang, 2013).

Microsatellite instability and methylation of *MLH1* promoter were also detected in squamous carcinoma and sebaceous carcinoma (Cheung et al, 2018). As well, reduced expression of *MLH1* and *MSH2* was observed in breast cancer with MSI. In sporadic breast tumors hypermethylation of *MSH2* was detected, which is in concordance with the results obtained in colon cancer (Murata et al, 2002).

1.17 Mismatch repair genes in prostate cancer

Similarly to what has been observed in colorectal cancer, defects in MMR genes were also observed in prostate cancer. Firstly, homozygous deletion of *MSH2* gene was reported in LNCaP (androgen-sensitive human prostate adenocarcinoma) cells. Secondly, the loss of MLH1 protein expression was described in DU145 (human prostate cancer cell line) cells (Chen et al, 2001).

After discovery of MSI and lack of expression of *MSH2* and *MSH6*, Soravia et al (2003) were first one to propose prostate cancer as a part of Lynch syndrom. In the same year, Grindedal et al (2003) associated germline mutations in MLH1, MSH2, MSH6 and PMS2 with increased risk of prostate cancer. The defect of the same genes is a cause of Lynch syndrom. In correlation with that, Chen et al (2001) showed that reduction of PMS1 and PMS2 expression was more common than loss of MLH1 and MSH2 expression in sporadic prostate cancer, which is not the case commonly seen in Lynch syndrome. These findings put MMR genes in the spot of prostate cancer's related investigations. Two SNPs were associated with increase risk in overall prostate cancer: rs9852810 and rs1799977 in *MLH1* gene (Langerberg et al, 2010). Moreover, complex structural rearrangments of MSH2 and MSH6 were observed in hypermutated subtype of advanced prostate cancer and deficiency of MLH1 and MSH6 was established as hallmark in differentiating prostate cancer from benign prostatic hyperplasia (Pritchard et al, 2014; Basu et al, 2015). Recently, MSH2 protein loss was identified in primary prostate tumors caused by *MSH2* inactivation (Guedes et al, 2017).

1.18 Aims

The aim of this study is to experimentally analyze the expression of proposed candidates, alpha satellite RNA and *MSH2* gene, in order to identify new diagnostic and/or prognostic blood's biomarker of prostate cancer.

2 Materials and methods

2.1 Materials

2.1.1 Solutions, reagents and other materials

- Tubes for blood collection (PAXgene Blood RNA Tubes, QIAGEN)
- RNA isolation kit (PAXgene Blood RNA Kit, IVD, QIAGEN)
- Reverse transcription kit (PrimeScript™ RT Reagent Kit with gDNA Eraser (Perfect Real Time), Takara)
- Primers (Table 1.), (Macrogen, Republic of Korea)
- Fluorescence or double-strand DNA binding dye (SYBR® PCR Master Mix, Applied Biosystems)
- Qubit reagents (Qubit® RNA BR Buffer ,Qubit® RNA BR Standard #1 (0 ng/μL in TE buffer), Qubit® RNA BR Standard #2 (100 ng/μL in TE buffer), Qubit® RNA BR Reagent 200X concentrate in DMSO, Qubit® ssDNA Reagent 200X concentrate in DMSO, Qubit® ssDNA Buffer, Qubit® ssDNA Standard #1 (0 ng/μL in TE buffer), Qubit® ssDNA Standard #2 (20 ng/μL in TE buffer), Invitrogen)
- Water (AccuGENE, Lonza)
- MicroAmp™ Optical 96-Well Reaction Plate (Applied Biosystems)
- MicroAmp™ Optical Adhesive Film (Applied Biosystems)

2.1.2 Primers for quantitative polymerase chain reaction (qPCR) analysis

Transcripts of alpha satellite RNA, *MSH2* and *GUS* were measured by qPCR, using primers pairs that enable their specific amplification. Primers were designed using Primer3Plus software. Specificity of primers was assessed using PrimerBlast by blasting chosen primers against NCBI Genome (chomosomes from all organisms) database limiting organism field to *Homo sapines*. Primers pairs sequences, temperature of melting (Tm) and expected size of amplicon are listed in Table 1. All primers were ordered from Macrogen, Republic of Korea.

Table 1. Primers used for qPCR analysis.

Primer name	**Primer sequence**	**Tm/ ° C**	**Product size / bp**
GUSBfw	5'-GAAAATATGTGGTTGGAGAGAGCATT-3'	63,2	101
GUSBrev	5'-CCGAGTGAAGATCCCCTTTTTA-3'	60,3	
MSH2fw	5'-AAGCCCAGGATGCCATTGTT-3'	58,3	130
MSH2rev	5'-GCTCCATTTGACACGTGAGC-3'	60,5	
ALPHAfw	5'-CACTCTTTTTGTAGAATCTGC-3'	55,5	126
ALPHArev	5'-AATGCACACATCACTATGTAC-3'	55,5	

2.1.3 Blood samples

Blood samples were provided and collected by prof. dr. sc. Ana Fröbe from University Clinical Hospital Centre (UHC) Sestre Milosrdnice using PAXgene Blood RNA Tubes (QIAGEN) and stored at -20 °C.

This study comprised a total of 25 subjects with prostate cancer diagnosis and 10 control subjects (Table 2.).

As control subjects, we used 10 male blood donors with no personal history of prostate cancer at the time of blood collection.

Ethical approval for this study was obtained by Ethical Committee of UHC Sestre Milosrdnice (approval number: EP-9251/17).

Table 2. Samples used in study. RNA and cDNA concentration were measured on Qubit™ Fluorometer; PC: prostate cancer.

sample	**group**	**RNA (ng/µL)**	**cDNA (ng/µL)**
A	control	27	101
B	control	170	104
E2	control	45	200
K2	control	80	87
K5	control	42	169
K6	control	47	154
78	control	107	33
79	control	158	38
80	control	335	30
81	control	87	45
32	PC	82	150
33	PC	100	135
34	PC	267	187
35	PC	130	52
36	PC	170	124
37	PC	135	136
39	PC	100	151
40	PC	140	146
58	PC	98	93
59	PC medium-risk	96	92
60	PC	70	69
62	PC medium-risk	88	144
62x	PC medium-risk	68	112
64	PC	52	121
65	PC	163	165
66	PC	132	169
67	PC	70	60
68	PC	104	122
69	PC	81	51
70	PC	128	84
71	PC	76	115
73	PC	56	200
75	PC medium-risk	114	185
76	PC medium-risk	95	67
77	PC medium-risk	80	186

2.2 Methods

2.2.1 RNA isolation

RNA was isolated from blood using RNA PAXgene Blood RNA Kit, IVD (QIAGEN).

Before isolation, PAXgene Blood RNA Tubes with blood samples must be incubated for at least 2 hours at room temperature to ensure complete lysis of blood cells. After 2 hours, PAXgene Blood RNA Tubes were centrifuged for 10 minutes at 3000xg. After the centrifugation, supernatant is removed by pipetting. 4 mL RNase-free water was added to the pellet, and tube was closed using a fresh secondary BD Hemogard closure. The pellet was vortexed until it was visibly dissolved and then centrifuged for 10 minutes at 3000xg. The supernatant was removed. 350 µL of buffer BR1 was added and pellet was vortexed until it was visibly dissolved. The samples were pipetted into a 1.5 mL microcentrifuge tubes. 300 µL of buffer BR2 and 40 µL were added to each tube. Tubes were vortexed for 5 seconds and then incubate for 10 minutes at 55 °C using a shaker incubator at 400-1400 rpm. The lysate was pipetted directly into a PAXgene Shredder spin column placed in a 2 mL processing tube, and centrifuged for 3 minutes at 10000xg. The entire supernatant of the flow-through fraction was transferred to a fresh 1.5 mL microcentrifuge tube. 350 µL of 96 % ethanol was added and tubes were centrifuged for 2 seconds at 500xg in order to remove drops from the inside of the tube. 700 µL of samples was pipetted into PAXgene RNA spin column placed in a 2 mL processing tube, and centrifuged for 1 minute at 10000xg. 350 µL of buffer BR3 was added at samples were centrifuged for 1 minute at 10000xg. Spin columns were placed into a new 2 mL processing tubes. DNA incubation mix I of 10 µL DNase I stock solution and 70 µL buffer RDD was prepared for a number of samples plus 2, added in 1.5 mL microcentrifuge tube and mixed. 80 µL of DNA incubation mix was pipetted onto the PAXgene RNA spin column membrane and left for 15 minutes. 350 µL of buffer BR3 was pipetted into PAXgene RNA spin column and centrifuged for 1 minute at 10000xg. RNA spin column was placed in new 2 mL processing tube. 500 µL of buffer BR4 was added and columns were centrifuged for 1 minute at 10000xg. The spin column was placed into the new 2 mL processing tube. Another 500 µL of buffer BR4 was added and columns were centrifuged for 3 minutes at 10000xg. The columns were placed in new 1.5 mL microcentrifuge tubes

and 40 µL of buffer BR5 was added. Columns were centrifuged for 1 minute at 10000xg to elute RNA. Eluates were incubated for 5 minutes at 65 °C and then stored at -20 °C.

2.2.2 Measurement of RNA concentration

For all isolated RNA samples, RNA concentration was measured using Qubit™ RNA BR Assay Kit (Invitrogen). The measurement is performed on Qubit™ Fluorometer using 1 µL of RNA sample mixed with 199 µL of solution made by mixing 199 µL of BR Buffer and 1 µL of fluorescent reagent Qubit® RNA BR Reagent 200X. Before every measurement run the flurometer was calibrated using two standard solutions: 10 µL of Qubit® RNA BR standard #1 (0 ng/µL in TE buffer) and 10 µL of Qubit® RNA BR standard #2 (100 ng/µL in TE buffer) mixed with 190 µL of solution made by mixing 199 µL of BR Buffer and 1µL of fluorescent reagent.

2.2.3 Reverse transcription (RT)

Reverse transcription for all isolated RNA samples was performed using PrimeScript™ RT Reagent Kit with gDNA Eraser (Perfect Real Time), (Takara) in order to synthesize cDNA.

Genomic DNA elimination reaction

The total amount of each RNA sample in reaction was decided to be 500 ng, so an appropriate volume of RNA was added in the reaction, followed by adding an appropriate volume of RNase Free dH2O up to 7 µL. Master mix for components other than the RNA sample and RNase Free dH2O was prepared in a volume sufficient for the number of reactions plus 2. Per reaction, 2 µL of 5X gDNA Eraser Buffer and 1 µL of gDNA Eraser are added in master mix. 3 µL of master mix were added in each reaction and the reactions are put at 42 °C for 2 minutes.

Reverse - transcription reaction

Reverse - transcription reaction solution was prepared on ice. Master mix was prepared in a sufficient volume for the number of reactions plus 2. Per each reactions, 4 µL of 5X

PrimeScript Buffer 2, 1 μL of PrimeScript RT Enzyme Mix 1, 1 μL of RT Primer Mix and 4 μL of RNase Free dH2O were added. 10 μL of master mix was added to each Genomic DNA elimination reaction and then put at 37° C for 15 minutes, followed by 5 seconds at 85 °C. Reverse transcription minus control was done per each RNA sample by adding water instead of reverse transcriptase enzyme in reaction. Samples were stored at -20 °C.

2.2.4 Measurement of cDNA concentration

For all cDNA samples, cDNA concentration was measured using Qubit™ ssDNA Assay Kit (Invitrogen). The measurement is performed on Qubit™ Fluorometer using 1 μL of cDNA sample mixed with 199 μL of solution made by mixing 199 μL of ssDNA Buffer and 1 μL of fluorescent reagent Qubit® ssDNA Reagent 200X. Before every measurement run the flurometer was calibrated using two standard solutions: 10 μL of Qubit® ssDNA Standard #1 (0 ng/μL in TE buffer) and 10 μL of Qubit® ssDNA Standard #2 (20 ng/μL in TE buffer) mixed with 190 μL of solution made by mixing 199 μL of ssDNA Buffer and 1μL of fluorescent reagent.

2.2.5 Quantitative polymerase chain reaction (qPCR)

qPCR also known as real-time PCR is a method used to quantify specific nucleic acid sequences in a sample of interest. The method relies on coupling of DNA amplification by PCR to a fluorescent signal that increases in intensity as more DNA is synthesized. The inclusion of a DNA-binding dye generates a fluorescent signal that is directly proportional to the amount of amplified target DNA. A range of fluorescent markers can be used for quantitive PCR, including DNA binding dyes such as SYBR green that intercalate with all double-stranded DNA (Figure 8. a), or sequence-specific fluorescent probes that only bind to a particular target sequence. Double-stranded DNA binding dyes have the advantage of being compatible with multiple primer sets, but the possibility of interference by off-target amplification or primer-dimer formation is higher. Sequence-specific probes provide greater specificity, e.g. Taqman. Such probes consist of

a short nucleotide sequence with a reporter and quencher attached to either end. The quencher prevents the emission of a fluorescent signal while the probe reamins unbound, but when the probe is bound to its target gene during DNA amplification, it is cleaved by the Taq polymerase to free the reporter and emit a detectable fluorescent signal (Miller & Andre, 2015).

Primer design and validation is critical for qPCR. The optimal annealing temperature can be determined experimentally using melting curve analysis. A single peak indicates the presence of specific products, whereas multiple peaks indicate additional off-target amplification or primer dimer formation. Dissociation stage is necessary to perfrom for constructing dissociation or melting curve (Figure 8. c). A quantitative PCR experiment should include sufficient biological and technical replicates, with at least three technical replicates being preferable so that outliners can be eliminated if one of the samples is not concordant. A no-template control should also be run to test for contamination with exogenous DNA, which means water inestad of cDNA. Also, reverse transcriptase-minus control is preferable and checks for contamination and purity of RT reaction (Miller & Andre, 2015).

Realtime PCR (or qPCR) is a widely used gold of standard for quantitative analysis of gene expression, but also for miRNA expression profiling and quantitating of noncoding RNA transcripts. So, to conclude RNA preparation and cDNA synthesis, together with primer design are crucial moves in qPCR experiment (Miller & Andre, 2015).

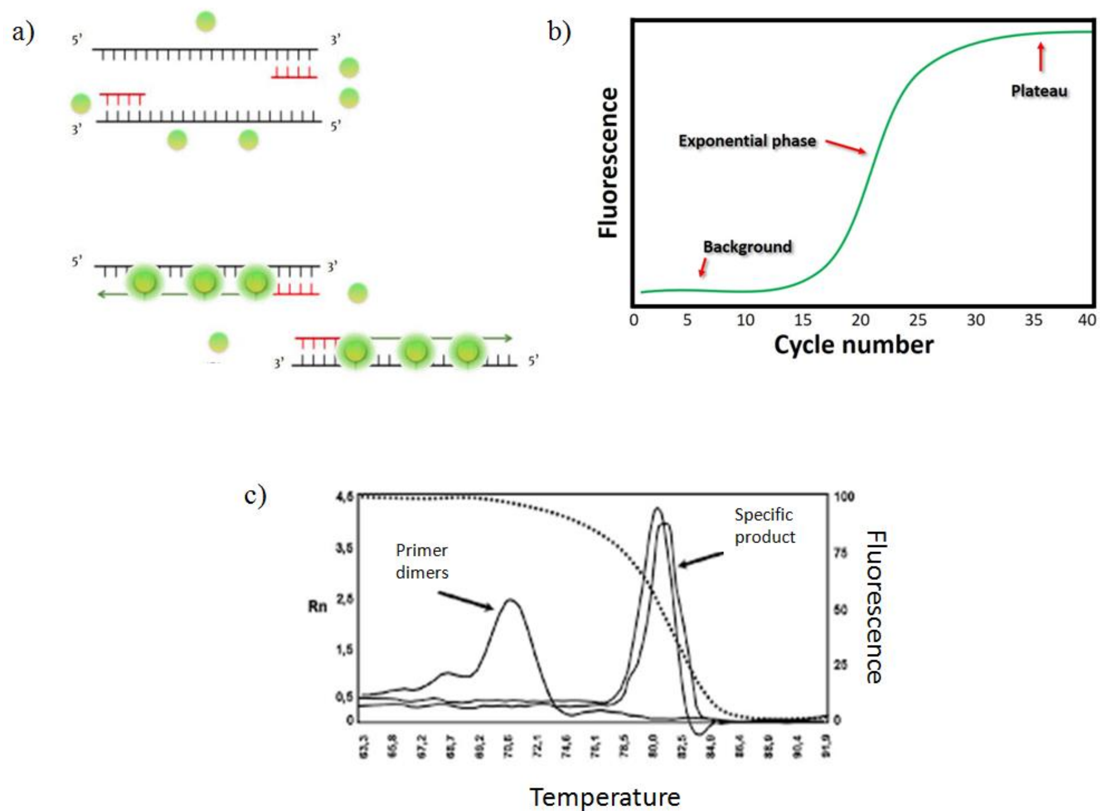


Figure 8. a) SYBR green dye incorporating into double stranded DNA; b) At the start of the run, when the amount of PCR product is low, this produces very little fluorescence. This section of cycles (between cycles 0 and 15 in the above figure) is known as the background signal. Once the amount of PCR product is amplified enough, it will enter the exponential phase. This is when the amount of PCR product doubles for every PCR cycle. This can be seen in cycles 15 and 25 in the above amplification plot. Once all of the reagents, such as nucleotides, have been used up in the PCR reaction, the amplification will slow and ultimately plateau. This is the region where no more PCR products cannot be produced. This can be seen in cycles 30 and 40 in the above amplification plot; c) dissociation curve showing primer dimer formation (taken and modified from Miller & Andre, 2015; Yibing, 2012).

Firstly, all cDNA samples were diluted to 30 ng/μL by adding an appropriate volume of water (AccuGene, Lonza). Master mix for components other than the cDNA was prepared in a volume sufficient for the number of reactions plus 2. Per each reaction, 37,5 μL of 2X SYBR® PCR Master Mix, 7,5 μL of 2mM primers (Table 1.) (final concentration of primers is 0,2 mM) and 28,5 μL of water was mixed and vortexed. 1,5 μL of each cDNA sample was added in the 1.5 mL microcentrifuge tube, with the total amount

of 45 ng of cDNA. Subsequently, 73,5 μL of master mix was mixed with cDNA and then solution was vortexed. Experiment was performed in triplicates, so reaction solution of 75 μL was splitted and volume of 20 μL was added in three wells of 96-well plate (MicroAmp™ Optical 96-Well Reaction Plate, Applied Biosystems) with the total amonut of 15 ng of cDNA in each well. Also, no template control (NTC) (adding of 1,5 μL of water instead of cDNA) was included in each qPCR run. 96-well plate was covered with MicroAmp™ Optical Adhesive Film (Applied Biosystems) and put in the real-time thermocycler (Real-Time PCR 7300 System, Applied Biosystems), under conditions shown in Table 3. Amplification specificity was confirmed by dissociation curve analysis. Specificity of amplified product was additionally tested on agarose gel.
Each qPCR was run three times.

Table 3. Conditions used for qPCR run.

<table>
<tr><th>Step</th><th>Temperature</th><th>Duration</th><th>Number of cycles</th></tr>
<tr><td>1.</td><td>50 ° C</td><td>2 minutes</td><td rowspan="4">40</td></tr>
<tr><td>2.</td><td>95 ° C</td><td>7 minutes</td></tr>
<tr><td>3.</td><td>95 ° C</td><td>15 seconds</td></tr>
<tr><td>4.</td><td>60 ° C</td><td>1 minute</td></tr>
<tr><td rowspan="4">Dissociation stage</td><td>95 ° C</td><td>15 seconds</td><td rowspan="4">15</td></tr>
<tr><td>60 ° C</td><td>1 minute</td></tr>
<tr><td>95 ° C</td><td>15 seconds</td></tr>
<tr><td>60 ° C</td><td>15 seconds</td></tr>
</table>

The amount of cDNA used in PCR reactions gave Ct value which is defined as the threshold PCR cycle at which measured fluorescence is significantly higher than background levels (Figure 8. b) (Miller & Andre, 2015).

Data during the run were collected and processed for baseline after the run using 7300 system SDS version 1.4 sofware (Applied Biosystems). Baseline represents level of fluorescence measured before any specific amplification can be detected. Baseline-corrected data were exported from SDS sofware and processed using LinRegPCR software v.2014.5 (Remakers et al, 2003).

The calculation of starting concentrations in qPCR analysis requires an estimate of the PCR efficiency, the setting of a fluorescence threshold and determination of the Ct value. The basic equation for PCR kinetics states that the amount of amplicon after c cycles (Nc) is the starting concetration of amplicon (No) times amplification efficiency (E) to the power c (Equation 1.) In this definition PCR efficiency is defined as the fold increase of the DNA amount per cycle, ranging from 1 (no amplification) to 2 (complete doubling). Equation 1 can be rearranged to calculate starting concentration (Equation 2.) In this calculation as Nc value is used-user defined fluorescence threshold (Nt) located in the exponential phase of qPCR and fractional nuber of cycles (Ct) needed to reach that flourescence threshold. Calculated No value is expressed in arbitrary fluorescence units (Akrap, 2013).

$$Nc = No \times E^{c}$$ Equation 1.

$$No = Nt/E^{Ct}$$ Equation 2.

Analysis of qPCR data with LinRegPCR is based on the PCR efficiency that can be derived from the slope of the regression line through the data points in the exponential phase when log (fluorescence) data are plotted against cycle number. Points in the exponential phase used for estimation of the PCR efficiency are reffered to as data points in the Window-of-Linearity (W-o-L). It was shown that the amplicon sequence is the main contributor to the efficiency and mean efficiency values of all cDNA samples of a given amplicon shows less variation then individual efficiencies of those samples. Based on this consideration, the algorithm to set W-o-L searches for the window with the least variation is minimal. From the slope of the line plotted through 4 to 6 data points in this window, the mean efficiency is determined. For each amplicon group, a fluorescence threshold (Nt) is

set a 1 cycle below the top border of the W-o-L and the Ct value is determined for each reaction. For the calculation of No value for all samples setting of W-o-L and efficiency for that amplicon group was used (Akrap, 2013).

To correct for the differences in sample composition and the yield of the reverse transcription reaction, No values of tested transcripts were normalized using β-glucuronidase (GUS). Selection of reference gene (housekeeping gene) during relative quantification experiments is a crucial part of qPCR experiment. The expression of a reference gene should be equal between samples, and its expression should be unaffected by which-ever factor or disease process is being investigated (Miller & Andre, 2015).

To calculate relative expression, firstly average No value in target wells of replicates per sample was taken. The same average was taken for the reference wells. The ratio of these two averages gives relative expression of a target which can then be compared between different samples.

2.2.6 Statistical methods

A t-test is also known as Student's t-test. It is a statistical analysis technique that was developed by William Sealy Gosset in 1908 for comparing the means of two groups. Unpaired t-test is used for comparing means between two independent samples (Kim, 2015). I performed unpaired t-test with the purpose of checking if level of supression of each individual PC sample is statistically significant when compared with control samples. The p-value per each sample was calculated using No values of triplicates from three independent qPCR runs. The null hypothesis of unpaired t-test was that there is no difference in means between No values of individual PC sample and No values of control samples with the significant level of 0.05. If obtained p-value is < 0.05, null hypothesis is rejected. The alternative hypothesis is that difference in means between No values of individual PC sample and control samples exists. The normality test was not performed, because it was assumed that No values of the same sample are distributed normally. The t-test was done in Microsoft Office Excel 2007.

In order to check, if the No values of control group and PC group of samples are normally distributed, I performed Shapiro-Wilk test. The null hypothesis of this test is that the data are normally distributed. If the p-value is lower than significance level, which was 0.05 in this test, the null hypothesis is rejected and there is evidence that the data tested are not normally distributed. If the p-value is > 0.05 the null hypothesis is accepted and the data is normally distributed. Shapiro-Wilk test was performed in GraphPad Prism 8.

In order to compare, if two groups statistically significant differ from each other I used Mann-Whitney test. The Mann-Whitney non-parametric test is often presented as an alternative to a parametric t-test when the data are not normally distributed and when at least one group has small number of samples. Whereas a t-test is a test of population means, the Mann-Whitney test is commonly regarded as a test of population medians. It tests whether one variable tends to have values higher than the other (Hart, 2001).

In order to compare, if more then two groups differ statistically significant from each other, I used Kurskal-Wallis test. The Kruskal-Wallis test or one-way analysis of variance (ANOVA) on ranks is a non-parametric method which does not assume a normal distribution of the samples, unlinke the analogous parametric ANOVA. It extends the Mann-Whitney test, which is used for comparing only two groups. A significant Kruskal-Wallis test indicates that at least one sample stochastically dominates one other sample (Siegel & Castellan, 1988).

To statistically test the significance of the observed differences in expression in different groups, the Mann-Whitney and Kruskal-Wallis tests were performed in GraphPad Prism 8.

The null hypothesis of both Mann-Whitney and Kruskal-Wallis tests is that there is no difference in medians between groups. If obtained p-value (the critical value) is < 0.05, the null hypothesis is rejected. Then, we would accept the alternative hypothesis with the 95 % of confidence level that there is difference in medians between groups. We can also say that observed difference is statistically significant at a significance level of 0.05. In line with that, if obtained p-value is > 0.05, the null hypothesis is accepted, that is, we are 95 % confident that there is no difference in medians among tested groups.

3 Results

3.1 Expression of *MSH2* gene

The expression in blood of prostate cancer and control samples of *MSH2* gene was measured using qPCR in the interest of checking for the possible differences between theirs expression profiles. In order to calculate relative expression, firstly average No value in target wells of replicates per sample was taken. The same average was taken for the reference wells. The ratio of these two averages gave relative expression of *MSH2* gene for each sample, which is No value shown in Figure 9. No values of controls and No values of PC samples were used in statistical tests, in order to check for the difference between them, respectively the difference between *MSH2* expression of control and PC samples.

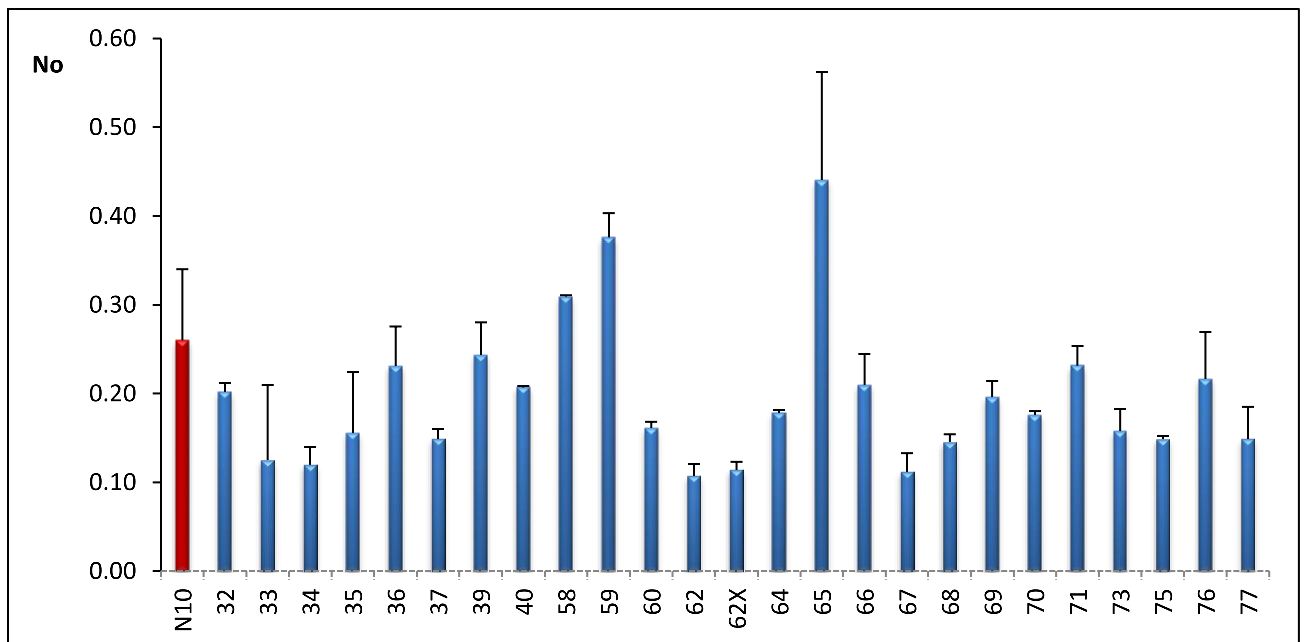


Figure 9. Expression of *MSH2* gene was analysed from cDNA amplified by qPCR using specific primers (listed in Table 1.) in blood of prostate cancer and control samples (listed in Table 2.). Results are presented as No values, normalized and averaged for each sample. N10 column shows average of 10 control samples of 3 different qRT-PCR experiments performed in triplicates and error bars represent standard deviations (red column). Other columns (32-77) show average of 3 different qRT-PCR experiments performed in triplicates of each individual prostate cancer sample and error bars represent standard deviation (blue columns).

Table 4. Level of supression of *MSH2* gene in the blood of prostate cancer patients when compared with average of 10 control samples. The p-values were calculated using unpaired t-test; PC: prostate cancer.

PC sample	**The level of supression with respect to average of 10 controls**	**Significance (p-value)**
32	1,29	0,07440231
33	2,08	0,23552483
34	2,17	0,00150226
35	1,67	0,24050516
36	1,13	0,59791127
37	1,74	0,00274918
39	1,07	0,76430378
40	1,26	0,09248558
58	0,84	0,06419686
59	0,69	0,01016425
60	1,61	0,00524084
62	2,43	0,00030173
62X	2,28	0,00035605
64	1,46	0,01524936
65	0,59	0,25477721
66	1,24	0,27481751
67	2,32	0,00126035
68	1,79	0,00203636
69	1,33	0,06854526
70	1,48	0,01297387
71	1,12	0,45486835
73	1,65	0,01905585
75	1,75	0,00243944
76	1,20	0,47712434
77	1,74	0,04813248

The results of expression analysis of *MSH2* gene are summarized in Table 4. The unpaired t-test is used for examination which PC samples have statistically significant level of supression. The test was performed in Microsoft Office Excel 2007 (the p-values are shown in Table 4.). Statistically significant downregulation (p-value < 0.05) is observed in 11/22 (50 %) of downregulated PC samples.

Shapiro-Wilk test was performed in the interest of checking if the data of control and PC group of samples are distributed normally. The results showed that the data are not distributed normally, because p-value of PC data was less than 0.05 (Figure 10.).

Normality and Lognormality Tests Tabular results	A control	B PC
1 Test for normal distribution		
2 Shapiro-Wilk test		
3 W	0.9538	0.8343
4 P value	0.7131	0.0009
5 Passed normality test (alpha=0.05)?	Yes	No
6 P value summary	ns	***
7		
8 Number of values	10	25

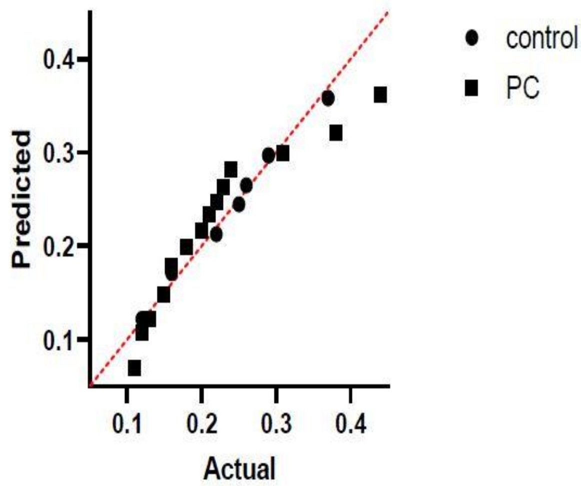
Figure 10. Results of Shapiro-Wilk test. Test was perfomed using *MSH2* expression data obtained by qPCR from blood of 10 control and 25 PC samples; PC: prostate cancer.

The results from Shapiro-Wilk test are also presented graphically, by using Q-Q (quantile-quantile) plot, which compares two probability distributions by plotting their quantiles againt each other. If the distributions being compared are similar, the points in the Q-Q plot would lie on the line y = x (Figure 11. a).

The Mann-Whitney test is used for checking for the differences between two groups, when the data are not normally distributed. The test was performed in GrapPad Prism 8. The results of analysis are shown in Figure 12. a. Since, the P- value was < 0.05, the difference in expression between control and patients was statistically significant. Graphical representation of this test is shown in Figure 11. b.

Six PC samples represent medium-risk prostate cancer (Table 2.), so we decided to performed statistical test in order to check if differences between medium risk prostate cancer exist with respect to the rest of prostate cancer samples and control samples. The Kruskal-Wallis test was performed in GrapPad Prism 8. The results of analysis are shown in Figure 12. b. Since, the P- value was > 0.05, the difference in expression between these groups was not statistically significant.

a) b)



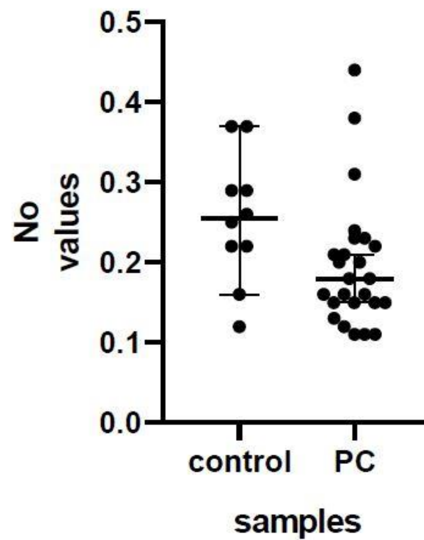


Figure 11. a) Q-Q plot of expression in *MSH2* gene from blood of control and PC samples. If the data is normally distributed, the points would lie on red line (y = x); b) Graphical representation of Mann-Whitney test of *MSH2* gene. Medians with CI 95 % for both groups are designated; PC: prostate cancer.

a)

	Mann-Whitney test	
1	Table Analyzed	Data 2
2		
3	Column B	PC
4	vs.	vs.
5	Column A	control
6		
7	**Mann Whitney test**	
8	P value	0.0270
9	Exact or approximate P value?	Exact
10	P value summary	*
11	Significantly different (P < 0.05)?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	240 , 390
14	Mann-Whitney U	65
15		
16	**Difference between medians**	
17	Median of column A	0.2550, n=10
18	Median of column B	0.1800, n=25
19	Difference: Actual	-0.07500
20	Difference: Hodges-Lehmann	-0.07000

b)

	Kruskal-Wallis test ANOVA results	
1	Table Analyzed	Data 2
2		
3	**Kruskal-Wallis test**	
4	P value	0.0671
5	Exact or approximate P value?	Approximate
6	P value summary	ns
7	Do the medians vary signif. (P < 0.05)?	No
8	Number of groups	3
9	Kruskal-Wallis statistic	5.402
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of values (total)	35

Figure 12. Statistical analysis of *MSH2* expression data obtained by qPCR in control and PC samples. a) Mann-Whitney test was performed for comparing PC samples with respect to control samples. The results are statistically significant if obtained p-value is < 0.05; b) Kruskal-Wallis test was performed for comparing medium risk PC samples with respect to the rest of PC samples and control samples. The results are statistically significant if obtained p-value is < 0.05; PC: prostate cancer.

3.2 Expression of alpha satellite DNA

The expression in blood of prostate cancer and control samples of alpha satellite transcripts was measured using qPCR in the interest of checking for the possible differences between theirs expression profiles. In order to calculate relative expression, firstly average No value in target wells of replicates per sample was taken. The same average was taken for the reference wells. The ratio of these two averages gave relative expression of alpha satellite for each sample, which is No value shown in Figure 13. No value of controls and No values of PC samples were used in statistical tests, in order to check for the difference between them, respectively the difference between alpha satellite expression of control and PC samples.

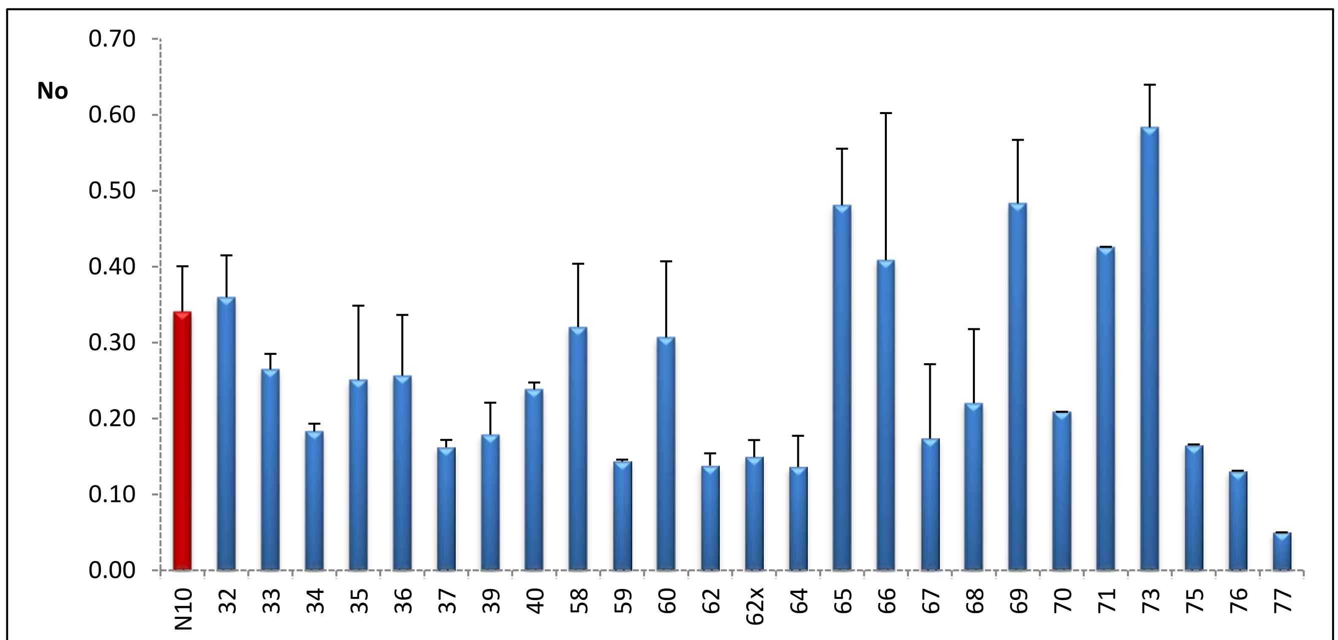


Figure 13. Expression of alpha satellite DNA was analysed from satellite transcripts amplified by qPCR using specific primers (listed in Table 1.) in blood of prostate cancer and control samples. (listed in Table 2.). Results are presented as No values, normalized and averaged for each sample. N10 column shows average of 10 control samples of 3 different qRT-PCR experiments performed in triplicates and error bars represent standard deviations (red column). Other columns (32-77) show average of 3 different qRT-PCR experiments performed in triplicates of each individual prostate cancer sample and error bars represent standard deviation (blue columns).

Table 5. Level of supression with level of significance of alpha satellite RNA in the blood of prostate cancer patients when compared with average of 10 control samples. The p-values were calculated using unpaired t-test; PC: prostate cancer.

PC sample	The level of supression with respect to average of 10 controls	Significance (p- value)
32	0,95	0,61268528
33	1,29	0,02220976
34	1,86	0,00000318
35	1,36	0,41206913
36	1,33	0,20685683
37	2,10	0,00000066
39	1,91	0,00288655
40	1,43	0,00041928
58	1,06	0,81163816
59	2,38	0,00000025
60	1,11	0,64607436
62	2,47	0,00000535
62x	2,29	0,00029548
64	2,50	0,02169577
65	0,71	0,18617860
66	0,83	0,70002192
67	1,96	0,22933619
68	1,55	0,15766064
69	0,71	0,21610635
70	1,63	0,00002751
71	0,80	0,00109157
73	0,58	0,04810150
75	2,07	0,00000107
76	2,61	0,00000012
77	6,82	0,00000001

The results of alpha satellite RNA expression are summarized in Table 5. The unpaired t-test is used for examination which PC samples have statistically significant supression. The test was performed in Microsoft Office Excel 2007 (the p-values are shown in Table 5.). In 13/19 (68 %) of downregulated samples, downregulation is statistically significant (p-value < 0.05).

Shapiro-Wilk test was performed in the interest of checking if the data of control and PC group of samples are distributed normally. The results showed that the data are not distributed normally, because p-value of PC data is less than 0.05 (Figure 14).

	Normality and Lognormality Tests Tabular results	A control	B PC
1	**Test for normal distribution**		
2	**Shapiro-Wilk test**		
3	W	0.9613	0.9157
4	P value	0.8007	0.0410
5	Passed normality test (alpha=0.05)?	Yes	No
6	P value summary	ns	*
7			
8	**Number of values**	10	25

Figure 14. Results of Shapiro-Wilk test. Test was perfomed using alpha satellite expression data obtained by qPCR from blood of 10 control and 25 PC samples; PC: prostate cancer.

The results from Shapiro-Wilk test are also presented graphically, by using Q-Q (quantile-quantile) plot, which compares two probability distributions by plotting their quantiles againt each other. If the distributions being compared are similar, the points in the Q-Q plot would lie on the line y = x (Figure 15. a).

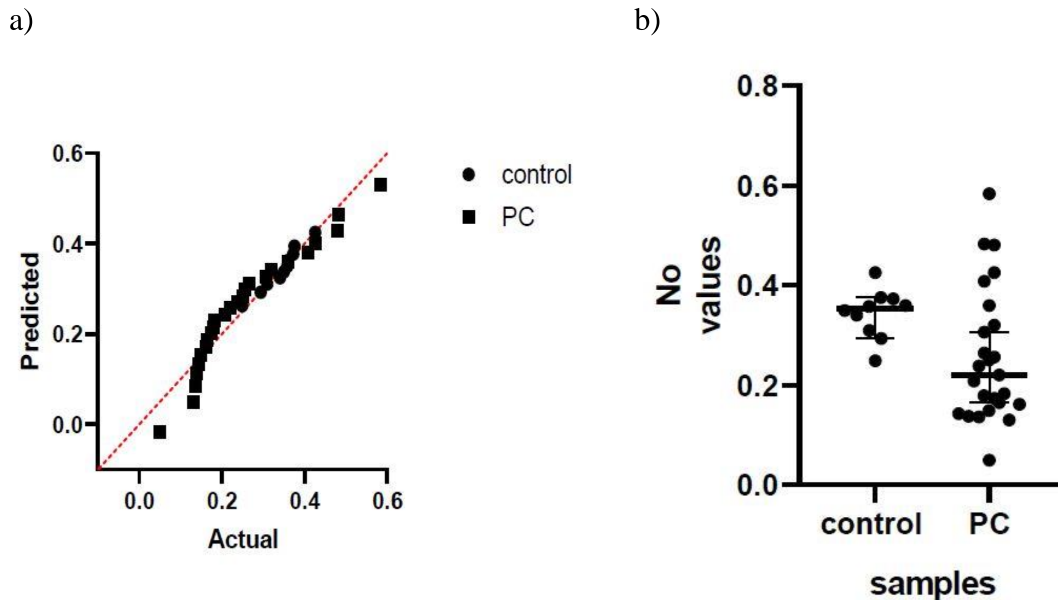


Figure 15. a) Q-Q plot of expression of alpha satellite from blood of control and PC samples. If the data is normally distributed, the points would lie on red line (y = x); b) Graphical representation of Mann-Whitney test of alpha satellite. Medians with CI 95 % for both groups are designated; PC: prostate cancer.

The Mann-Whitney test is used for checking for the differences between two groups, when the data are not normally distributed. The test was performed in GrapPad Prism 8. The results of analysis are shown in Figure 16. a. Since, the P- value was < 0.05, the difference in expression between control and patients was statistically significant. Graphical representation of this test is shown at Figure 15. b. We also performed Kruskal-Wallis test in order to check if samples from medium risk prostate cancer are different when comparing both with control samples and the rest of prostate cancer samples. The results of analysis are shown in Figure 16. b. P- value was < 0.05, so the statistically significant difference exists.

a)

	Mann-Whitney test	
1	Table Analyzed	Data 1
2		
3	Column B	PC
4	vs.	vs.
5	Column A	control
6		
7	**Mann Whitney test**	
8	P value	0.0207
9	Exact or approximate P value?	Exact
10	P value summary	*
11	Significantly different (P < 0.05)?	Yes
12	One- or two-tailed P value?	Two-tailed
13	Sum of ranks in column A,B	243 , 387
14	Mann-Whitney U	62
15		
16	**Difference between medians**	
17	Median of column A	0.3540, n=10
18	Median of column B	0.2201, n=25
19	Difference: Actual	-0.1338
20	Difference: Hodges-Lehmann	-0.1174

b)

	Kruskal-Wallis test	
1	Table Analyzed	Data 1
2		
3	**Kruskal-Wallis test**	
4	P value	0.0007
5	Exact or approximate P value?	Approximate
6	P value summary	***
7	Do the medians vary signif. (P < 0.05)?	Yes
8	Number of groups	3
9	Kruskal-Wallis statistic	14.63
10		
11	**Data summary**	
12	Number of treatments (columns)	3
13	Number of values (total)	35

Figure 16. Statistical analysis of alpha satellite expression data obtained by qPCR in control and PC samples. a) Mann-Whitney test was performed for comparing PC samples with respect to control samples. The results are statistically significant if obtained p-value is < 0.05; b) Kruskal-Wallis test was performed for comparing medium risk PC samples with respect to the rest of PC samples and control samples. The results are statistically significant if obtained p-value is < 0.05; PC: prostate cancer.

4 Discussion

Loss of DNA mismatch repair (MMR) function has been linked to genetic instability that results in high mutation rates, especially among microsatellites. Microsatellite sequences are highly unstable in tumor cells with defective MMR. Microsatellite instability (MSI) was firstly reported in hereditary nonpolyposis colorectal cancer (HNPCC), which is part of Lynch syndrome. After prostate cancer was suggested to be a part of Lynch syndrome, MMR genes in prostate cancer cells and tissues were target of numerous studies (Chen et al, 2001).

In our study, for the first time, expression of any of seven MMR genes - *MSH2* was investigated in the blood of the prostate cancer patients.

Our results indicate that observed statistically significant downregulation of *MSH2* gene in 44 % (11/25) of prostate cancer samples when comparing with control samples can serve as a diagnostic tool for prostate cancer (results shown in Figure 9; Table 4.). These results are in strong correlation with results observed in prostate cancer tissues and cell lines. Recently, a hypermutated subtype of advanced prostate cancer was described, which is associated with mutations, structural rearrangments and loss of expression of the MMR genes: *MSH2* and *MSH6* (Pritchard et al, 2014). Also, several other studies reported altered expression of MLH1, MSH2, MSH6 and PMS2 (Wilczak et al, 2017). In 2017, loss of MSH2 protein was identified in primary prostate cancer and its correlated with *MSH2* gene inactivation and hypermutation (Guedes et al, 2017).

So, *MSH2* expression in blood of prostate cancer patients can serve as potential diagnostic marker. This diagnostic method would be safe and simple for measurement, and most important - non-invasive. In future studies, we suggest investigating on mechanisms of *MSH2* downregulation (e.g. checking for methylation level of *MSH2*), what can lead to a better understanding of disease biology and reliable identification of clinically applicable marker.

Moreover, we checked for association between medium risk prostate cancer and level of *MSH2* downregulation, but results showed no statistically significant difference between levels of expression of medium risk PC samples, the rest of PC samples and

control samples (Figure 12. b). Based on these results *MSH2* can not be used in predictive or prognostic purposes, but analysis should be repeated with the higher number of samples.

In our study, for the first time, the level of alpha satellite transcripts in blood of PC was measured. Kinetochore formation is a prerequsite for chromosomal segregation and defects in satellite alpha transcription are leading to theirs disruption. Also, alpha satellite transcripts are important elements for regulating cell division and heterochromatin formation. The cause of alterate expression of satellite DNA are not completely clear, but the consequences could originate genome instability (Ichida et al, 2018).

Our results (Figure 13.; Table 5.) are showing significant downregulation of alpha satellite expression in 13/25 (52 %) of the samples. The results from the other investigations strongly associated hypomethylation of alpha satellite DNA and cancerogenesis. Hypomethylation leads to overexpression of alpha satellite transcripts and results in abnormal chromosomal segregation and chromosomal instability (Ichida et al, 2018).

The overexpression of satellite RNA has been found to induce abnormal segregation of chromosome in breast cancer (Zhu et al, 2011).

In osteosarcoma overexpression of repetitive elements was observed with respect to normal tissue (Ho et al, 2017). Also, aberrant overexpression of satellite repeats was reported in epithelial cancers (Ting et al, 2011).

Alpha satellite transcripts are also required for localization of SUV39H1, enzyme which catalyzes addition of methyl groups to lysine 9 in histone 3 (H3K9). These groups are binding sites for heterochromazin protein 1 (HP1) which leads to spreading of heterochromatin (McNulty & Sullivan, 2018).

Although, our results are not in concordance with existing data, which show increase of satellite transcripts in different types of cancer, they seem to be statistically powerful (Figure 16. a). Also, we are the first one to report expression of alpha satellite DNA in blood of cancer patients. Alpha satellite transcripts seems to have important role in spreading heterochromatin as already mentioned above, so we propose that either downregulation or upregulation would cause alteration of that process.

Our results reveal that expression profile of alpha satellite RNA in the blood of prostate cancer patients could serve as a diagnostic biomarker. Moreover, the differences of alpha satellite RNA expression profile of medium risk prostate cancer samples is

significantly different relative to the rest PC and control samples (Figure 16. b). The National Comprehensive Cancer Network (NCCN) classification system stratifies men into very-low, low, medium, high or very high prostate cancer risk groups. The NCCN system classification is based on tumor stage, Gleason score and PSA level. Medium risk prostate cancers are those which are unlikely to grow or spread in a few years (Nicholas et al, 2016). These results put alpha satellite RNA as a potential candidate for prognostic biomarker of medium risk prostate cancer. Prognostic biomarkers are used to measure the progress of a disease and help medicians to select the appropriate treatment. The evaluation of new prostate cancer prognostic biomarker is huge clinical need in PC management (Terada et al, 2017).

Lot of challenges and conditions should be obtained from promising experimental results and biomarker discovery to clinical validation of biomarker. Some of the questions which should be answered are about dynamic range and absolute level of biomarker expression, as low-expressed biomarkers are often less reliable.

It is important that clinical assay is non-inasive, reproducible and easy to perform. Moreover, control and experimental cohorts should matched uniformly. Bias in cohort design account for many false-positive biomarker reports (Terada et al, 2017).

Even though, long journey is ahead of us, we propose two new potential blood's biomarkers for prostate cancer: *MSH2* gene as diagnostic and alpha satellite RNA as prognostic biomarker. The assay for both proposed candidates is shown in Figure 17.

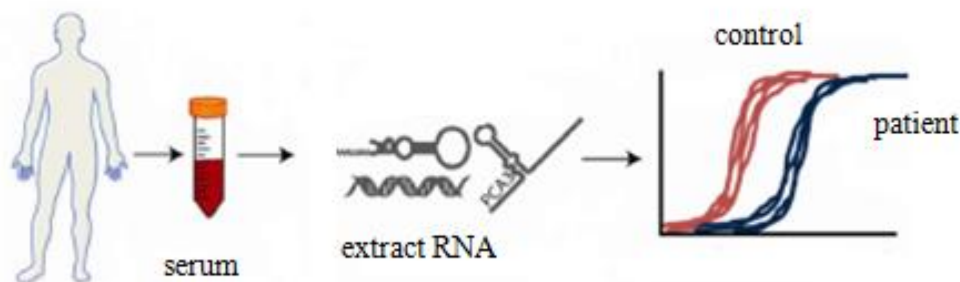


Figure 17. The assay for our potential biomarkers measurement (taken and modified from Terada et al, 2017).

It is important to mention the limitations with this study design, which include the inability to confirm all prostate cancer diagnoses with medical records. By knowing medical history of patients we would be able to distinguish between patients diagnosed with hereditary prostate cancer and those diagnosed with sporadic prostate cancer. Potentially, we could associate expression profile of investigated candidates between men with or without family history of PC. Additionally, with the available data we were unable to use clinical features such as prostate-specific antigen (PSA) or Gleason score to differentiate between low-risk, medium or high-risk prostate cancer. The only information about differences between prostate cancer patients we got is which six PC samples belong to medium risk prostate cancer type. Furthermore, we could also associate upregulated samples with theirs clinicopathological features.

Also, control group of samples should be screened more detailed. Our condition was that control samples do not have diagnosis of prostate cancer at the time of blood collection, but more conditions should be evaluated. For example, testing by using some of the previously mentioned biomarkers should be done, e. g. PSA or PCA 3. Although, those biomarkers have their limitations, the tests results would give us a larger picture.

Most importantly, the higher number of samples must be used. The higher number of samples will provide us with more realible data and will reduce uncertainties. Sample size determines our precision or level of confidence.

Despite the study limitations, our finding of two new potential blood's biomarkers for prostate cancer by non-invasive method of detection is valuable. Additional studies to quantify benefits and cost effectiveness of this screening will offer us larger picture about these candidates.

5 Conclusion

This study reveals two new potential blood's biomarker candidates of prostate cancer: *MSH2* gene as a diagnostic and alpha satellite RNA as a prognostic biomarker. Expression of proposed candidates were analyzed in blood samples of 10 controls and 25 PC cases. Downregulation was observed in 44 % of PC samples in *MSH2* gene and in 52 % of PC samples in alpha satellite RNA. In order to confirm this results, further analysis needs to be carried out with the higher number of samples. Also, to understand the mechanisms of downregulation of both candidates in prostate cancer, additional experiments must be performed.

6 References

1. Akrap I. (2013): Study of satellite DNA-mediated gene regulation in red flour beetle *Tribolium castaneum* (Herbst, 1797) (graduation thesis). Retrieved from Croatian Digital Theses Repository, National and University Library in Zagreb. https://zir.nsk.hr/islandora/object/pmf%3A769
2. Aldrup-MacDonald M. E., Kuo M. E., Sullivan L. L., Chew K., Sullivan B. A. (2016): Genomic variation within alpha satellite DNA influences centromere location on human chromosomes with metastable epialleles. Genome Research **26(10)**: 1301-1311.
3. Basu S., Majumder S., Bhowal A., Ghoash A., Naskar S., Nandy S., Mukherjee S., Slnha R. K., Basu K., Kamakar D., Banerjee S., Sanghamitra B. (2015): A study of Molecular Signals Deregulating Mismatch Repair Genes In Prostate cancer compared to Benign Prostatic Hyperplasia. PLoS ONE **10(5):** e0125560.
4. Brajković J., Feliciello I., Bruvo-Madarić B., Ugarković Đ. (2012): Satellite DNA-Like Elements Associated With Genes Within Euchromatin of the Beetle *Tribolium castaneum*. G3 **2**: 931-41.
5. Chen Y., Wang J., Fraig M. M., Metcalf J., Turner W. R., Bissada N. K., Watson D. K., Schweinfest C. W. (2001): Defects of DNA mismatch repair in human prostate cancer. Cancer research **61**: 4112-4121.
6. Cheung A. H., Chow C., Yu M. Y., Law W. W., Law P. P., Choi P. C., Kang W., To K. F. (2018): Mismatch repair deficiency is implicated in carcinoma arising from ovarian teratoma. Pathology **51(1)**: 67-73.
7. Cucheux L., Ponger L., Gerbault-Seureau M., Richard F. A., Escude C. (2016): Diversity and distribution of alpha satellite DNA in the genome of and Old World monkey: *Cercopithecus solatus*. BMC Genomics **17**: 916.
8. Current Procedural Terminology © American Medical Association. (2017): Genetic and Protein Biomarkers for the Diagnosis and Cancer Risk Assessment of Prostate Cancer. Blue Cross BlueShield of Kansas 1-41.
9. Erie D. A., Weninger K. R. (2014): Single molecule studies of DNA mismatch repair. DNA repair **20**: 71-81.

10. Feliciello I., Akrap I., Ugarković Đ. (2015): Satellite DNA Modulates Gene Expression in the Beetle Tribolium castaneum after Heat Stress. PLoS Genetics **11(8):** e1005466.
11. Feliciello I., Picariello O., Chinali G. (2006): Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: Molecular evidence of a new mechanism of DNA repair acting on satellite DNA. Gene **383**: 81-92.
12. Gelsomino F., Barbolini M., Spallanzani A., Pugliese G., Cascinu S. (2016): The evolving role of microsatellite instability in colorectal cancer: A review. Cancer Treatment Review **51**: 19-26.
13. Grindedal E. M., Moller P., Eeles R., Stormoken A. T., Bowitz-Lothe I. M., Landro S. M., Clark N., Kvale R., Shanley S., Maehle L. (2009): Germ-line mutations in mismatch repair genes associated with prostate cancer. Cancer Epidemiology, Biomarkers & Prevention **18(9)**: 2460-2467.
14. Guedes L. B., Antonarakis E. S., Schweizer M. T., Mirkhesti N., Almutairi F., Park J. C., Glavaris S., Hicks J., Eisenberger M. A., De Marzo A. M., Epstein J. I., Isaacs W. B., Eshleman J. R., Pritcahrd C. C., Lotan T. L. (2017): MSH2 loss in primary prostate cancer. Clinical cancer research **23(22)**: 6863-6874.
15. Hart A. (2001): Mann-Whitney test is not just a test of medians: - differences in spread can be important. British Meadical Journal **323 (7309)**: 391-393.
16. Ho X. D., Nguyen H. G., Trinh L. H., Reinmann E., Prans E., Koks G., Maasalu K., Le V. Q., Nguyen V. H., Le N. T. N., Phung P., Martson A., Lattekivi F., Koks S. (2017): Analysis of the Expression of Repetitive DNA Elements in Osteosarcoma. Frontiers in Genetics **8**: 193.
17. Hou J., Zhao L., Zhang D., Lv D., He W., Chen B., Li H., Li P., Chen L., Chen X. (2018): Prognostic value of mismatch repair genes for patients with colorectal cancer: meta analysis. Technology in cancer research & treatment **17**: 1-11.
18. Ichida K., Suzuki K., Fukui T., Takayama Y., Kakizawa N., Watanabe F., Ishikawa H., Muto Y., Kato T., Saito M., Futsuhara K., Miyakura Y., Noda H., Ohmori T., Konishi F., Rikiyama T. (2018): Overexpression of satellite alpha transcripts lead to chromosomal instability via segregation errors at specific chromosomes. International journal of oncology **52**: 1685-1693.

19. Ideue T., Hino K., Kitao S., Yokoi T., Hirose T. (2009): Efficient oligonucleotide-mediated degradation of nuclear noncoding RNAs in mammalian cultured cells. RNA **15(8):** 1578-87.
20. Kim T. K. (2015): T test as a parametric statistics. Korean Journal of Anesthesiology **68(6):** 540-546.
21. Kunkel T. A., Erie D. A. (2015): Eukaryotic Mismatch Repair in Relation to DNA replication. Annual review of genetics **49**: 291-13.
22. Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., ... Morgan M. J. (2001): Initial sequencing and analysis of the human genome. Nature **409 (6822)**: 860-921.
23. Langeberg W. J., Kwon E. M., Koopmeiners J. S., Ostrander E. A., Stanford J. L. (2010): Population-Based Study of the Association of Variants in Mismatch Repair Genes with Prostate Cancer Risk and Outcomes. Cancer Epidemiology, Biomarkers & Prevention **19(1)**: 258-264.
24. Li G. (2008): Mechanisms and functions of DNA mismatch repair. Cell research **18**:85-98.
25. Liu D., Keijzers G., Rasmussen L. J. (2017): DNA-mismatch repair and its many roles in eukaryotic cells. Mutation Research **773**: 174-187.
26. Maio J. J. (1971): DNA strand reassociation and polyribonucleotide binding in the African green monkey *Cercopithecus aethiops*. Journal of Molecular Biology **56**: 579-595.
27. Manuelidis L. (1978): Chromosomal localization of complex and simple repeated human DNAs. Chromosoma **66**: 23-32.
28. McNulty S. M., Sullivan B. A. (2018): Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosome Research **26 (3)**: 115-138.
29. Miller J. R. C., Andre R. (2015): Quantitative polymerase chain reaction. British Journal of Hospital Medicine **75**: 188-191.
30. Murata H., Khattar N. H., Kang Y., Gu L., Li G. (2002): Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. Oncogene **21(37)**: 5696-5703.
31. Nicholas A., Serrano M. D., Mitchell S. A. (2016): Favorable vs Unfavorable Intermediate-Risk Prostate Cancer: A Review of the New Classification System and Its Impact on Treatment Recommendations. Oncology Journal **30(3)**: 229-236.

32. Nimesh J., Viswanadham D., Deasirazu N. R. (2006): Prokaryotic DNA Mismatch repair. Progress in Nucleic Acid Research and Molecular Biology **81**: 1-49.
33. Pezer Ž., Brajković J., Feliciello I., Ugarković Đ. (2012): Satellite DNA - Mediated Effects on Genome Regulation. Genome dynamics **7**:153-69.
34. Pezer Ž., Ugarković Đ. (2008): Role of non-coding RNA and heterochromatin in aneuploidy and cancer. Seminars in Cancer Biology **18**: 123-130.
35. Plohl M., Luchetti A., Meštrović N., Mantovani B. (2008): Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene **409**: 72-82.
36. Prensner J. R., Rubin M. A., Wei J. T., Chinnaiyan A. M. (2012): Beyond PSA: The next generation of prostate cancer biomarkers. Science Translation Medicine **4(127)**: 127rv3.
37. Pritchard C. C., Morrissey C., Kumar A., Zhang X., Smith C., Coleman I., Salipante S. J., Milbank J., Yu M., Grady W. M., Tait J. F., Corey E., Vessella R. L., Walsh T., Schendure J., Nelson P. S. (2014): Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer. Nature communications **5**: 4988.
38. Ragner J. R., Meselson M. (1976): Repair tracts in mismatched DNA heteroduplexes. Proceedings of the National Academy of Sciences of the United States of America **73(11)**: 4135-4139.
39. Ramakers C., Ruijter J. M., Deprez R. H., Moorman A. F. (2003): Assumption-free analysis of quantitative real-time PCR data. Neuroscience Letters **339(1)**: 62-66.
40. Reljić A., Čukelj P., Tomašković I., Ružić B., Šekerija M. (2018): Epidemiology of prostate cancer in Croatia - situation and perspectives. Acta Clinica Croatica **57**: 27-34.
41. Rudd K., Mays R.W., Schwartz S., Willard H. F. (2003): Sequence organization and functional annotation of human centromeres. Cold Spring Harbor Symposia on Quantitative Biology **68**: 141-149.
42. Saini S. (2016): PSA and beyond: alternative prostate cancer biomarkers. Cell Oncology **39(2)**: 97-106.
43. Shepelev V. A., Alexandrov A. A., Yurov Y. B., Alexandrov I. A. (2009): The evolutionary origion of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genetics **5**: e1000641.

44. Siegel S., Castellan N. J. (1988): Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, New-York. ISBN 0070573573.
45. Sinicropet F. A., Yang Z. J. (2013): Prognostic and predictive impact of DNA mismatch repair in the management of colorectal cancer. Future oncology **7(3)**: 467-474.
46. Smurova K., De Wulf P. (2018): Centromere and Pericentromere Transcription: Roles and Regulation… in Sickness and in Health. Frontiers in Genetics **9**: 674.
47. Soravia C., van der Klift H., Brundler M., Blouin J., Wijnen J., Hutter P., Fodde R., Delozier-Blanchet C. (2003): Prostate cancer is part of the hereditary non-polyposis colorectal cancer (HNPCC) tumor spectrum. American Journal of Medical Genetics **121A**: 159-162.
48. Terada N., Akamatsu S., Kobayashi T., Inoue T., Ogawa O., Antonarakis E. S. (2017): Prognostic and predictive biomarkers in prostate cancer: latest evidence and clinical implications. Therapeutic Advances in Medical Oncology **9(8)**: 565-573.
49. Ting D. T., Lipson D., Paul S., Brannigan B. W., Akhavanfard S., Coffman E. J., Contino G., Deshpande V., Iafrate A. J., Letovsky S., Rivera M. N., Bardeesy N., Maheswaran S., Haber D. A. (2011): Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science **331(6017)**: 593-596.
50. Ugarković Đ. (2005): Functional elements residing within satellite DNAs. EMBO reports **6**: 1035-1039.
51. Volpe T. A., Kidner C., Hall I. M., Teng G., Grewal S. I. S., Martienssen R. A. (2002): Regulation of heterochromatic silencing and histone H3 lysine-9 methlation by RNAi. Science **297**: 1833-1837.
52. Warburton P. E., Hasson D., Guillem F., Lescale C., Jin X., Abrusan G. (2008): Analysis of the largest tandemly repeated DNA families in the human genome. BMC Genomics **9**: 533.
53. Warburton P. E., Willard H. F. (1995): Interhomologue sequence variation of alpha satellite DNA from human chromosome 17: evidence for concerted evolution along haplotypic lineages. Journal of Molecular Evolution **41**: 1006-1015.
54. Wilczak W., Rashed S., Hube-Magg C., Kluth M., Simon R., Buscheck F., Clauditz T. S., Grupp K., Minner S., Tsourlakis M. C., Koop C., Graefen M., Adam M., Hase A., Wittmer C., Sauter G., Izbicki J. R., Huland H., Schlomm T., Steurer S., Krech T., Lebok P. (2017): Up-regulation of mismatch repair genes MSH6, PMS2 and MLH1 parallels development of genetic instability, and is linked to tumor

aggressiveness and early PSA recurrence in prostate cancer. Carcinogenesis **38(1)**: 19-27.

55. Wong L. H., Brettingham-Moore K. H., Chan L., Quach J. M., Anderson M. A., Northrop E. L., Hannan R., Saffery R., Shaw M. L., Williams E., Choo K. H. (2007): Centromere RNA is a key component for the assembly of nuceloproteins at the nucleolus and centromere. Genome Research **17**: 1146-1160.
56. Wong M. C. S., Goggins W. B., Wang H. H. X., Fung F. D. H., Leung C., Wong S. Y. S., Ng C. F., Sung J.J. Y. (2016): Global Incidence and Mortality for Prostate Cancer: Analysis of Temporal Patterns and Trends in 36 Countries. European Urology **70(5)**: 862-874.
57. Yibing J. (2012): Methods in Cell Biology. In: Conn M. P. (Ed.). Real-Time PCR Oxford, Academic Press, pages 55-68. ISBN: 978-0-12-405914-6.
58. Zhu Q., Pao G. M., Huynh A. M., Suh H., Tonnu N., Nederlof P. M., Gage F. H., Verma I. M. (2011): BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature **477**: 179-184.

7 Curriculum vitae

NAME AND SURNAME: Zrinka Matić

DATE AND PLACE OF BIRTH: January 4th ,1995; Šibenik, Croatia

EDUCATION:
2016 – 2019: Master studies in molecular biology, Faculty of Science, University of Zagreb, Croatia
2013 – 2016: Bachelor studies in molecular biology, Faculty of Science, University of Zagreb, Croatia
2009 – 2013: High school education, Antun Vrančić Gymnasium, Šibenik, Croatia

RESEARCH EXPERIENCE
03/2018 – present: Master student, Ruđer Bošković Institute, Zagreb, Croatia; Laboratory of Evolutionary Genetics (LEG)
03/2017 – 06/2017: Traineeship, Institute for Medical Research and Occupational Health, Biochemistry and Organic Analytical Chemistry Unit, Zagreb, Croatia
03/2016 – 06/2016: Traineeship, Transfusion medicine and transplatation biology department, University Hospital Centre Rebro, Zagreb, Croatia

LANGUAGES
Croatian (mother tongue), English (independent user)

HONOURS, AWARDS, GRANTS, FELLOWSHIPS
2013 – 2018: Grant for oustanding students, Šibenik
10/2017 – 2/2018: ERASMUS + Student Mobility of Placement (SMP); Adam Mickiewicz University, Poznan, Poland

MEMBERSHIPS
2016 - present: Speleological department CMS Željezničar, Zagreb, Croatia