Charles University

Faculty of Science

Study program: Genetics, molecular biology and virology

Branch of study: Molecular biology and genetics of eukaryotes

Bc. Eliška Jandová

Sequence variants in metastatic colorectal cancer and their evolution during the disease course and treatment

Sekvenční varianty u metastatického kolorektálního karcinomu a jejich vývoj v průběhu nemoci a léčby

Diploma thesis

Supervisor: MUDr. Ludmila Boublíková, Ph.D.

Prague, 2023

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne 9.8. 2023

…….

Bc. Eliška Jandová

Acknowledgment

I would like to thank Martina Z., Lucie W., Laura M. and Kristýna Š. for their unlimited help and great ideas, without which I would not be able to finish this thesis. I would like to thank my supervisor, Dr. Ludmila Boublíková, for her guidance. I thank my family, my mother, my brother and my partner for their support and delicious food that they provided me during the time of finishing this thesis.

Abstrakt

Lepší pochopení molekulárně-genetických charakteristik kolorektálního karcinomu (CRC) může pomoci vysvětlit rozvoj onemocnění i rezistence k léčbě, předpovědět progresi nádoru a predikovat odpověď na léčbu. V současné době jsou součástí klinické praxe minimální molekulární testy, které pomáhají určit, zda pacient bude profitovat z konkrétní terapie, nicméně studie ukazují stále více a více nových mutací zodpovídajících za rezistenci k léčbě. Zlatým standardem pro molekulární testování jsou tkáňové biopsie, konstantně ale roste poptávka po méně invazivní metodě, jako je například tekutá biopsie (liquid biopsy). S využitím sekvenování nové generace jsme analyzovali sekvenční varianty přítomné v tkáni primárního nádoru, metastáz a volné cirkulující DNA nádoru – ctDNA u pacientů s metastatickým kolorektálním karcinomem. Cílem práce bylo analyzovat sekvenční varianty přítomné v primárním tumoru a identifikovat varianty patogenní, porovnat sekvenční varianty primárního tumor a metastáz a zjistit, zda se nějakým způsobem liší, zhodnotit, jestli ctDNA může být použita v diagnostice a zda se v ní mohou nacházet tumor specifické markery, které by se dali použít k monitorování progrese. Výsledky naší práce indikují možné využití ctDNA pro podrobnou molekulární analýzu nádorového onemocnění i pro monitorování progrese nemoci, zároveň ale poukazují na nutnost optimalizace této metody a její možné limitace, především kvůli existujícímu prahu sensitivity.

Klíčová slova: sekvenční variant; mutace; kolorektální karcinom; cirkulující volná nádorová DNA, sekvenování nové generace

Abstract

A deeper understanding of the molecular background of colorectal cancer (CRC) can help explain the development of the disease and its resistance to treatment, predict disease progression, and improve treatment prognosis. Some minimal molecular testing has been incorporated into standard clinical management to determine if a particular patient will benefit from a particular therapy, but more and more new genetic alterations are being discovered that appear to be associated with the development of resistance. Tissue biopsy of the tumor is the gold standard in terms of molecular testing, but there is an increasing demand for more non-invasive methods such as liquid biopsy. Using targeted next-generation sequencing, we analyzed sequence variants present in primary tumor, metastases, and cell-free tumor DNA - ctDNA of patients with metastatic CRC. The objectives were to analyze sequence variants of the primary tumor and identify possible pathogenic variants, to analyze differences between DNA of the primary tumor and metastases, to evaluate the use of ctDNA as a diagnostic tool, and to identify potential tumor-specific markers in ctDNA that can be used to monitor disease progression. Our results suggest the feasibility of using ctDNA for diagnostic purposes or even to monitor disease progression, but at the same time we found that this method needs to be optimized and have its limits due to sensitivity issues.

Keywords: sequence variants; mutation; colorectal cancer; circulating free tumor DNA; next generation sequencing

Contents

1. Introduction

According to the World Health Organization's GLOBOCAN database, colorectal cancer (CRC) is the third most common cancer in men and the second most common in women. Although the understanding of CRC and it's treatment has improved greatly in recent years, CRC still remains the second leading cause of cancer death in the world (Xi a Xu 2021). Early stages of CRC can be treated by surgical resection, however, 25 % of all CRC cases are diagnosed in more advanced stages and another 20 % are metastatic, in these cases surgical removal alone is not curative (Hossain et al. 2022). There are more treatment strategies that can be used to treat CRC, such as chemotherapy, targeted therapy (e.g., anti-EGFR antibodies (Mendelsohn et al. 2015)), or immunotherapy, unfortunately, the problem with metastatic CRC is that most patients develop resistance to these treatments, likely due to tumor heterogeneity and clonal evolution of cancer cells (Dang et al. 2020) under the selective pressure of chemotherapy inducing genetic changes in cancer cells (Russo et al. 2019). Currently, some minimal molecular testing is used in standard clinical practice (Cervantes et al. 2023) to determine whether or not the patient will benefit from a particular therapy, but more and more new genetic alterations are being discovered that are responsible for some form of resistance. There are many studies analyzing the gene alterations of the tumor at the time of diagnosis, but there are fewer studies analyzing the changes in these genetic alterations during treatment and disease progression. One of the most promising methods to monitor the evolution of the genetic background of CRC is liquid biopsy, more specifically circulating free tumor DNA – ctDNA in plasma, urine and serum (M. Stroun et al. 1989; Chan, Chiu, a Lo 2003; Salvianti et al. 2021) . It is a non-invasive method that is easily reproducible, ctDNA can be obtained at any time during disease progression and provides information about the molecular background of the tumor as well as metastases. However, there are still some challenges such as low sensitivity compared to tissue biopsy (Parikh et al. 2019) and the need to optimize the used assays in terms of sample quality and quantity requirements, sequencing depth, isolation of ctDNA and sample collection procedures (Pathak 2022).

2. Aims of the Thesis

The aims of this thesis are: to analyze gene variants present in primary tumor DNA of patients with metastatic CRC and using available databases to identify those that are pathogenic - likely responsible for tumor formation; to identify specific gene alterations in the metastatic samples and in ctDNA; to evaluate and compare the applicability of the ctDNA and tissue biopsy as a diagnostic tool; and to try to find tumor-specific markers present in the ctDNA that can indicate the development of treatment resistance and could be used to monitor the progression of the disease.

3. Literature Review

As mentioned earlier, even though the understanding of CRC development and progression has advanced, CRC still remains at the top of the cancer statistics in terms of incidence and mortality (Xi a Xu 2021). Understanding of the molecular background of CRC can help with prediction of the disease course as well as with the treatment prognosis and development of new therapies (Reimers et al. 2013; Coppedè et al. 2014).

3.1 Molecular background of colorectal cancer

Mutations in oncogenes or tumor suppressor genes have been identified to be involved in colorectal cancer ethiopatogenesis. Depending on the origin and type of the mutation, colorectal cancer can be classified as sporadic, hereditary, or familial. The sporadic CRC accounts for 60 to 65 % of all CRC cases (Xi a Xu 2021) and is caused by point somatic mutations or epigenetic events (Fischer et al. 2019), there is no family history of CRC and no germline mutation is present. Hereditary CRC, that accounts for 6 to 10 % of all CRC cases, is characterized by germline mutations, presence of family history of CRC, and is inherited in true Mendelian fashion (Brosens, Offerhaus, a Giardiello 2015). The second most common form of CRC is familial CRC, which accounts for 10 to 20 % of all CRC cases (Armelao a de Pretis 2014). There is a family history of CRC, but there is no particular single gene aberration associated with CRC present; polygenic patterns and/or specific gene polymorphisms are supposed to be involved.

The molecular and biologic features further cluster CRC into four consensus molecular subtypes – CMSs (Guinney et al. 2015): CMS1 (MSI-immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal) (further explained in section 3.2.1.2.1).

There are three main genetic pathways associated with CRC (Nguyen a Duong 2018): CIN (chromosomal instability pathway), MSI (microsatellite instability pathway), and CIMP (CpG island methylator phenotype - epigenetic pathway). MSI is associated mostly with hereditary CRC – namely Lynch syndrome (Aaltonen et al. 1993 as cited in Lynch et al. 2015), but MSI can be present in sporadic CRC as well. CIN and CIMP are mostly associated with sporadic CRC (Fischer et al. 2019). A graphical representation of the different pathways associated with the formation of CRC is presented in the picture below (Fig. 1).

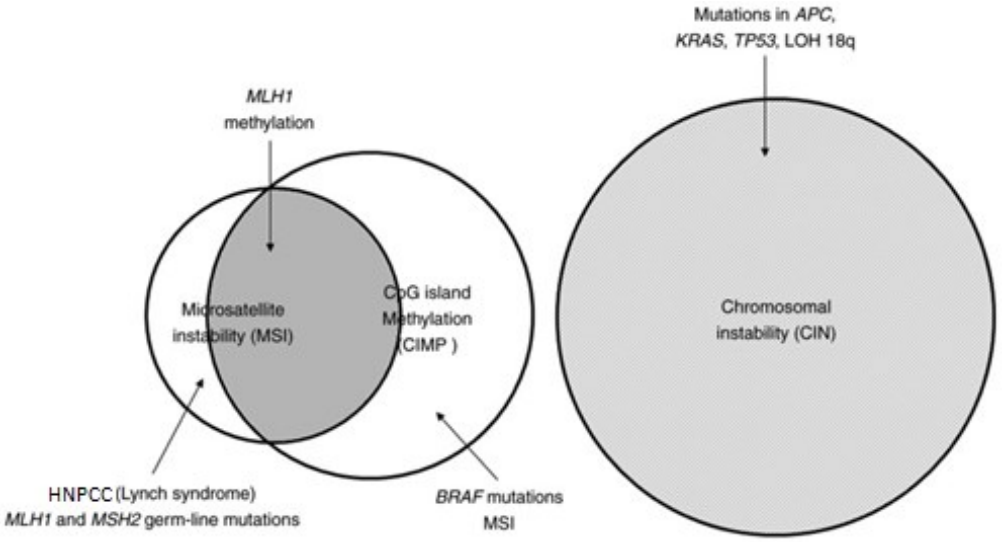

Fig. 1.: Graphic display of different pathways associated with CRC (Nguyen a Duong 2018). HNPCC - Hereditary Non-Polyposis Colorectal Cancer*, MLH1 - MutL Homolog 1*, *MSH2 - MutS Homolog 2*, *BRAF - B-Raf-Oncogene, Serine Theronine Kinase*, *APC - Adenomatosis Polyposis Coli Tumor Suppressor, KRAS - Kirsten Rat Sarcoma Viral Oncogene Homolog, TP53 - Tumor Protein 53*, LOH – Loss Of Heterozygosity

3.1.1 Genetic pathways associated with CRC

3.1.1.1 CIN pathway

Chromosomal instability is typical for the majority of sporadic colorectal cancer (Nguyen a Duong 2018). CIN leads to numerical or structural changes in chromosomes, formation of aneuploid karyotype, gains or losses of chromosomal segments and loss of heterozygosity (LOH) of some tumor suppressor genes (Tanaka a Hirota 2016). Commonly affected genes are *Adenomatosis Polyposis Coli Tumor Suppressor - APC* (with loss of function resulting in promotion of cell proliferation, playing role in the WNT (Wingless-Type MMTV Integration Site Family*)* signaling pathway) (Filippo et al. 2002); *Kirsten Rat Sarcoma Viral Oncogene Homolog – KRAS* (with activation that promotes apoptosis suppression, cell growth, cell transformation, migration and differentiation) (Ye et al. 2020); *PhosphatidylInositol-*

4,5-bisphosphate 3-Kinase, Catalytic subunit Alpha - *PIK3CA* (with activation resulting in hyperactivation of AKT (AKT Serine/Threonine Kinase 1) kinase influencing many signaling pathways, including mTOR (Mechanistic Target Of Rapamycin Kinase) activating growth and metabolism) (Samuels et al. 2004); *B-Raf-Oncogene, Serine Theronine Kinase – BRAF* (with activation resulting in support of MAPK ((EGFR)-mediated mitogen-activated protein kinase) pathway initiating cell growth, proliferation, differentiation, cell migration and cell survival) (Davies et al. 2002); *Mothers Against Decapentaplegic Homolog 4* - *SMAD4* (with inactivation influencing TGFβ (Transforming Growth Factor Beta) signaling pathway controlling proliferation, differentiation, apoptosis, migration of epithelial cell) (MacGrogan et al. 1997 as cited in Tarafa et al. 2000); *Tumor Protein 53* - *TP53* (with inactivation resulting mainly in activation of proliferation) (Baker et al. 1989).

3.1.1.2 MSI pathway

Mutations in genes involved in the mismatch repair pathway (MMR) cause microsatellite instability (Ionov et al. 1993 as cited in Velho et al. 2014) (microsatellites are short non-coding repeats with naturally higher frequency of mutations). Defects in MMR genes are the main cause of the Lynch syndrome (Ionov et al. 1993 as cited in Velho et al. 2014) which falls into the category of hereditary colorectal cancer. Sporadic CRC can also exhibit MSI, but unlike Lynch syndrome, mutations in MMR genes in sporadic form cannot be found in the germline DNA and are caused predominantly by promoter hypermethylation leading to inactivation of these genes, especially *MutL Homolog 1* - *MLH1* (Herman et al. 1998). Sporadic tumors with high microsatellite instability frequently present frameshift mutations in genes such as *Transforming Growth Factor-β Receptor 2 – TGFBR2* (with inhibition influencing positively epithelial cell growth) (Markowitz et al. 1995); *Insulin-like Growth Factor 2 Receptor – IGF2R* (with loss of function resulting in higher levels of IGF2 (insulin growth factor 2) and influencing proliferation and cell growth) (Souza et al. 1996 as cited in Yamagishi et al. 2016)*; E2F transcription factor 4, p107/p130-binding – E2F4* (involved in regulation of pro-proliferating genes) (Yoshitaka et al. 1996); *MutS Homolog 6* - *MSH6* and *MutS Homolog 3* - *MSH3* genes (involved in the miss match repair pathway) (Yamamoto, Sawai, a Perucho 1997 as cited in Yamagishi et al. 2016); *Caspase 5 – CASP5* (with inactivation

resulting in inactivation of pro-apoptotic genes) (Schwartz et al. 1999 as cited in Yamagishi et al. 2016).

3.1.1.3 CIMP phenotype

CIMP phenotype is an epigenetic alteration that can be characterized by widespread CpG island methylation (Toyota et al. 1999). This aberrant hypermethylation can cause some tumor suppressor genes to become inactivated, leading to tumorigenesis (Baylln et al. 1997). CIMP and MSI are closely linked, CpG island hypermethylation in the promoter of the gene *MLH1* is associated with inactivation leading to MSI (Herman et al. 1998) (as described in section 3.1.1.2). Some of the genes found to be aberrantly methylated in the promoter region and inhibited in this way are: *Cyclin Dependent Kinase Inhibitor 2A* - *CDKN2A* (coding for p16 tumor suppressor, cyclin inhibitor, signaling antiproliferation) (Gonzalez-Zulueta et al. 1995; Spagnol et al. 2022) and *Cadherin 1* - *CDH1* (coding for E-cadherin, with inhibition resulting in increased proliferation, invasion and/or metastasis) (Graff et al. 1995 as cited in Lao a Grady 2011).

3.1.2 Sporadic CRC

Sporadic CRC is the most common form of CRC, accounting for 60 to 65 % of all CRC cases (Xi a Xu 2021). Patients with sporadic CRC have no significant family history of CRC and no germline mutation is present. Sporadic CRC is mainly characterized by CIN and CIMP (Fischer et al. 2019). MSI may also be present, in sporadic CRC it is caused by epigenetic alteration of MMR genes, mainly by hypermethylation of the promoter of *MutL Homolog 1* - *MLH1* gene (Herman et al. 1998), however MSI is more typical for hereditary CRC (as will be described in section 3.1.3).

3.1.3 Hereditary CRC

Although up to 20 to 30 % of all colorectal cancer patients have a family history of the disease, only 6 to 10% of all colorectal cancers are inherited in a true Mendelian fashion (Brosens, Offerhaus, a Giardiello 2015). The inherited form is called hereditary CRC and is often classified as a series of diseases with specific mutations that predispose to colorectal cancer (Valle et al. 2019). There are two main types of hereditary form: Hereditary Non-Polyposis Colorectal Cancer (HNPCC) and Hereditary Polyposis Colorectal Cancer (HPCC). The number and histology of colorectal polyps are used to characterize HNPCC and HPPC (Valle et al. 2019).

3.1.3.1 HNPCC

The most common form of HNPCC is Lynch syndrome (C. R. Boland a Troncale 1984 as cited in Lynch et al. 2015; C. Richard Boland 2005). HNPCC colorectal cancers are caused by germline mutations that inactivate genes involved in mismatch repair pathway (MMR) (Aaltonen et al. 1993 as cited in Lynch et al. 2015), especially in genes *MLH1* (Bronner et al. 1994; Papadopoulos et al. 1994), *MutS Homolog 2* - *MSH2* (Liu et al. 1994 as cited in Taylor et al. 2003), *MSH6* (Miyaki et al. 1997; Muzny et al. 2012) and *PMS1 Homolog 2, Mismatch Repair System Component* - *PMS2* (Haraldsdottir et al. 2017; Nicolaides et al. 1994). Due to mutations in MMR pathway, HNPCC syndromes are characterized by microsatellite instability, usually resulting in increased immunogenicity of the tumor with important clinical and therapeutical implications (Peltomäki et al. 1993 as cited in Müller et al. 2004).

3.1.3.2 HPPC

Familial Adenomatous Polyposis (FAP), Polymerase Proofreading-associated Polyposis (PPAP), MUTYH-Associated Polyposis (MAP) and Juvenile Polyposis Coli (JPS) are just a few examples of HPCC (Valle et al. 2019). FAP is associated with pathogenic germline mutations in *APC* tumor suppressor gene, that plays role in regulation of WNT signaling pathway (Groden et al. 1991), PPAP, on the other hand, is characterized by germline mutations in exonuclease (proofreading) genes encoding domains of polymerase epsilon *POLE* and delta *POLD1* (Flohr et al. 1999; Yoshida et al. 2011). Both forms are characterized by dozens to hundreds of colorectal adenomas, thus falling into the category of adenomatous variants of HPCC, and both are caused by autosomal dominant mutations (Valle et al. 2019). MUTYH-Associated Polyposis - MAP also belongs to the category of adenomatous variants of HPCC, but is caused by a recessive autosomal mutation, namely of the *MUTYH* - *MutY DNA Glycosylase* gene of the base excision repair pathway (Jones et al. 2002). JPS is characterized by development of haramatous polyps in gastrointestinal tract (Larsen Haidle, MacFarland, a Howe 1993). Most of the time this form is benign, but it can also spread by metastases (Larsen Haidle, MacFarland, a Howe 1993). This form is characterized by germline mutations in *Bone Morphogenetic Protein Receptor Type 1A* - *BMPR1A* (coding receptor of TGF-β pathway, playing role in the cell homeostasis) and *SMAD4* genes (James R. Howe et al. 1998; J. R. Howe et al. 2001).

A summary of the hereditary forms mentioned above can be seen in the picture below (Fig. 2). There are many more genes that are directly or indirectly involved in the formation of the hereditary CRC (Valle et al. 2019), but for the simplicity of this literature review, only the most common ones have been described. Hereditary CRC are not only those described above, but there are many more (Valle et al. 2019) and the work to identify new syndromes is still in progress. The category of polyposis syndromes is further subdivided into mixed (Rohlin et al. 2016) and serrated (Gala et al. 2014) forms, which will not be covered in this work.

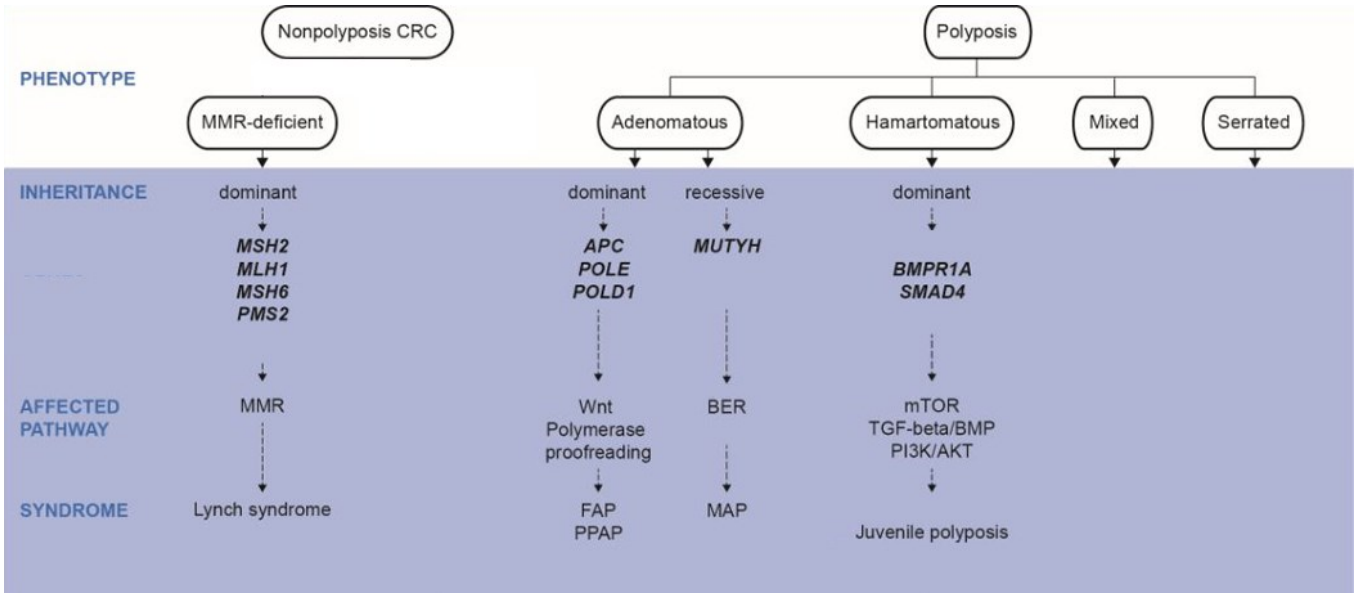

Fig. 2: Simplified guide to hereditary forms of CRC (Valle et al. 2019 modified). ***MSH2***: *MutS Homolog 2*; ***MLH1***: *MutL Homolog 1;* ***MSH6****: MutS Homolog 6;* ***PMS2:*** *PMS1 Homolog 2, Mismatch Repair System Component*; **MMR**: Miss Match Repair pathway; ***APC***: *Adenomatosis Polyposis Coli Tumor Suppressor;* ***POLE****: DNA Polymerase Epsilon, Catalytic Subunit;* ***POLD1****: DNA Polymerase Delta 1, Catalytic Subunit;* WNT: Wingless-Type MMTV Integration Site Family, Member 1 pathway; **FAP**: Familiar Adenomatous Polyposis; **PPAP**: Polymerase Proofreading-associated Polyposis, ***MUTYH***: *MutY DNA Glycosylase*; **BER**: Base Excision Repair; **MAP**: MUTYH-Associated Polyposis; *BMPR1A: Bone Morphogenetic Protein Receptor Type 1A* ; SMAD4: *Mothers Against Decapentaplegic Homolog 4;* mTOR: Mechanistic Target Of Rapamycin Kinase; TGF-beta: Transforming Growth Factor Beta; BMP: Bone Morphogenetic Protein 1, PI3K: PhosphoInositide-3 Kinase, AKT: AKT Serine/Threonine Kinase 1

3.1.4 Familial CRC

Twenty to thirty percent of all CRCs are inherited but cannot not be diagnosed as one of the syndromes described in section 3.1.3 (Armelao a de Pretis 2014). There is a family history, but the cancer does not meet the criteria of the syndromes and usually there is no particular genetic aberration associated with CRC present. It is a heterogeneous group consisting of undiagnosed syndromes and tumors that appear to originate from the same genetic background as the sporadic form, but are inherited (Armelao a de Pretis 2014). The diagnosis of patients with the familial form is problematic due to the great genetic heterogeneity and unknown origin (probably a combination of environmental and inherited genetic factors including polygenic inheritance, gene polymorphisms, etc..) (Armelao a de Pretis 2014).

3.2 Clinical application of the molecular genetics of CRC

Although there is great interest in understanding the tumorigenesis of CRC, developing new or improving existing treatments, and focusing on early detection, the mortality rate of CRC remains the second highest in the world (in terms of cancer) (Xi a Xu 2021). Comprehensive genomic analyses performed by Wood et al. in 2007 revealed a median of 76 non-silent mutations in patients with CRC, demonstrating the heterogeneity of CRC and the molecular uniqueness of molecular background of CRC of individuals (Wood et al. 2007). Since then, many studies have analyzed the mutational landscape of CRC, demonstrating its heterogeneity and discovering new gene alterations likely responsible for tumorigenesis (Cornish et al. 2022). A deeper understanding of the molecular background of CRC can help predict progression, metastasis, and even response to certain therapies (Reimers et al. 2013; Coppedè et al. 2014).

3.2.1 Molecular testing in standard patient care

Molecular genetic testing of some of the genes mentioned in previous sections has become part of the standard patients care with colorectal cancer - to predict response to targeted therapy, for prognostic purposes, and to diagnose inherited syndromes.

The testing involves: 1. genetic testing of germline mutations, to identify hereditary forms of CRC, predict risks for the patient and their family members, which sometimes also predicts a response to certain modalities of therapy. The most widely used panel for this type of genetic testing in our country is so called Czecanca panel (Soukupová et al. 2016); 2. tumor testing for somatic aberrations, to identify tumor-specific mutations that may have

prognostic or predictive impact or could be used as tumor markers for disease follow up or as targets for tumor-specific therapy.

3.2.1.1 Molecular testing for prediction of response to treatment

3.2.1.1.1 The minimal testing panel

In clinical practice, it is recommend to test for MSI status, and the most common mutations: *KRAS, NRAS* exon 2, 3 and 4 mutations and *BRAF* V600E mutations, at the time of diagnosis of an advanced disease for future decisions on therapy (Cervantes et al. 2023). Other additional molecular markers can be tested, such as *HER2* (*Human Epidermal Growth Factor 2*) amplification, *NTRK* (*Neurotrophic Receptor Tyrosine Kinase*) fusions, *ALK* (*Anaplastic Lymphoma Kinase/ALK Receptor Tyrosine Kinase*) fusions or *ROS1* (*ROS Proto-Oncogene 1*, Receptor Tyrosine Kinase) fusions, but they are applied mostly in clinical studies (Cervantes et al. 2023).

Apart from gene mutations, overexpression of growth factors, e.g. VEGF (Mohamed et al. 2019) or growth factor receptors like EGFR (Gross et al. 1991; Spano et al. 2005) or HER2 (Richman et al. 2016) may be present. In some cases, it is associated with a gene amplification (e.g. HER2), in other cases the underlying molecular alteration is not well characterized.

These aberrations may be targeted by specific antibodies or small molecular inhibitors, representing a novel modality within oncologic therapy – targeted or also called biologic therapy. One of the most widely used therapies of this kind for CRC is anti-EGFR therapy, which uses antibodies that target the epidermal growth factor receptor (EGFR) signaling pathway, such as cetuximab and panitumumab (Divgi et al. 1991 as cited in Mendelsohn et al. 2015; Reynolds a Wagstaff 2004; Keating 2010). EGFR is a type I membrane receptor tyrosine kinase that is frequently overexpressed in CRC (Cunningham et al. 2005 as cited in Francoual et al. 2006). EGFR binds multiple ligands (e.g. transforming growth factor alpha - TGF-alpha (Oda et al. 2005)) and regulates cell growth, proliferation, differentiation and survival. One of the downstream players activated by the EGFR signaling pathway is the RAS family of genes (Oda et al. 2005). It has been shown that *KRAS*, when mutated, is activated independently of the EGFR pathway (Mancl, Kolesar, a Vermeulen 2009). Before this discovery, it was already noticed that CRC patients with mutations in *KRAS* gene did not respond well to anti-EGFR therapy (Benvenuti et al. 2007). After these discoveries,

KRAS mutations were established as a predictive marker of poor response to anti-EGFR therapy (Lièvre et al. 2006) and are now recommended to be routinely tested before choosing the therapy (Cervantes et al. 2023). Detection tests are set to detect not only *KRAS* mutations but also mutations of the *NRAS* gene (another downstream player of the EGFR pathway (Oda et al. 2005)) (Peeters et al. 2015).

Another mutation that is recommended to be tested in routine clinical practice is the V600E of *BRAF* (Cervantes et al. 2023). Patients harboring this mutation may be successfully treated with a combination of the anti-EGFR antibody cetuximab and a small molecular inhibitor targeting *BRAF* V600E encorafenib (van Brummelen et al. 2017; Kopetz et al. 2019).

Understanding the MSI status of the tumor is another factor that is recommended to be tested prior to treatment (Cervantes et al. 2023). Several studies have shown that patients with tumors that have a high level of MSI (instability at two or more loci) are more immunogenic and benefit from immune checkpoint anti-PD1/PDL1 therapy (Le et al. 2015; André et al. 2020). The PD1/PDL1 (Programmed Death 1/ Programmed Death Ligand 1) pathway negatively regulates the Th1 cytotoxic immune response and is upregulated in many tumors (Le et al. 2015). High levels of MSI result in a high number of tumor neoantigens that can be recognized by the immune system (Le et al. 2017). In addition, MSI status may serve as a prognostic marker for stage II patients with high MSI tumors who are considered for fluorouracil chemotherapy, as these patients have been shown not to benefit from this chemotherapy (Sargent et al. 2010). In relation to MSI level, tumor mutational burden - the total number of mutations per coding region of a tumor genome - may be recognized as another promising prognostic marker for anti-PD1/PDL1 therapy (Yarchoan, Hopkins, a Jaffee 2017).

3.2.1.1.2 Extended and advanced testing panel

There are other genes that may serve as prognostic markers for the outcome of anti-EGFR therapy (Cervantes et al. 2023).

Fusions of the genes *ALK, ROS1* or *NTRK* are very rare in metastatic CRC, with only 0,2 to 2,4% of patients having these alterations (Pietrantonio et al. 2017). However, Pietrantonio et al. found in their study, that patients with these fusions present in the tumor have an extremely poor prognosis and do not benefit from the

anti-EGFR therapy, suggesting the use of specific inhibitors of these kinases (Pietrantonio et al. 2017). Resistance to anti-EGFR in *ALK* fusion-positive tumors, *ROS1* fusion positive-tumors and *NTRK* fusion-positive tumors results from the crosstalk of these pathways (Acquaviva, Wong, a Charest 2009; Pietrantonio et al. 2014; Cocco, Scaltriti, a Drilon 2018).

In several studies it was shown that presence of mutation in *PIK3CA* (another downstream gene regulated by EGFR (Oda et al. 2005)) in patients with wild-type *KRAS* is associated with poor outcome when using anti-EGFR therapy (Perrone et al. 2009). However, another study suggested that a mutation of *PIK3CA* does not play a role in resistance to anti-EGFR (Prenen et al. 2009). Later it was shown that it depends on the exon affected by the mutation, suggesting that only exon 20 mutations are associated with the treatment resistance (De Roock et al. 2010). Patients with mutations in *PIK3CA* may be indicated for adjuvant aspirin therapy, which has been shown to be associated with longer survival and lower recurrence rates in CRC patients with *PIK3CA* mutations (Liao et al. 2012; Domingo et al. 2013). Because of the rareness of these mutations and often limited amount of tumor samples available, these mutations are usually not tested separately but within a cancer panel by next generation sequencing approach.

Another marker of poor outcome of anti-EGFR therapy is *HER2* gene amplification (Sawada et al. 2018). This is based on the fact, that HER2 and EGFR are receptors of the same family and affect similar signaling pathways (Yarden a Sliwkowski 2001). A recent study has suggested that *HER2* may be a new potential target for targeted therapy (Siena et al. 2021).

3.2.1.2 Molecular testing for prognosis of the disease

Currently only clinicopathologic features and cancer staging at the time of diagnosis are used to determine disease prognosis (Koncina et al. 2020). However, some molecular biomarkers can be used as a prognostic tools. For example, several studies have shown that patients in early stage of the disease with MSI-positive tumors have a better survival rate compared to those with microsatellite-stable tumors (Thibodeau, Bren, a Schaid 1993; Lim et al. 2004). In addition MSI tumors appear to be associated with younger age (Gryfe et al. 2000). *KRAS* and *BRAF* mutations were found to be negative prognostic markers, as patients with these

mutations had a lower overall survival rate than those without these mutations (Richman et al. 2009; Yoon et al. 2014; Modest et al. 2016). *HER2* amplification was thought to be a marker of poor survival as well as of higher postoperative recurrence (Park et al. 2007), but some studies did not confirm this (Ochs et al. 2004). CIMP-positive status seems to be associated with worse survival and recurrence after resection in stage III CRC, but there are not enough studies to confirm this theory, so additional studies are needed (Ahn et al. 2011). For metastatic colorectal cancer, the presence of *TP53* mutation is thought to be a predictive marker for worse survival, but again more studies are needed to evaluate this (Pilat et al. 2015)

3.2.1.2.1 Consensus Molecular Subtypes

Another useful tool for determining disease progression can be the Consensus Molecular Subtyping (CMS) (Guinney et al. 2015). The CRC Subtyping Consortium established four groups of CRC, known as the Consensus Molecular Subtypes, because there was no uniform guide for gene expression-based subtyping due to differences in experimental design and data processing (Guinney et al. 2015). This classification was made using transcriptome-wide analyses merged with clinically relevant anatomical, genetic, and oncogenic classification systems. The classification is as follows: CMS1 (MSI-immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal) (Guinney et al. 2015). A brief summary of the molecular characteristics of all groups is given in the Table 1.

	CMS1	CMS2	CMS3	CMS4
Name	MSI, Immune	Canonical	Metabolic	Mesenchymal
Primary characteristics	Hypermutated, MSI, strong immune activation	Epithelial, WNT and MYC signaling activation	Epithelial and evident metabolic dysregulation	Prominent TGF-β activation, stromal invasion and angiogenesis
Incidence	14 %	37 %	13 %	23 %
Genomic associations	MSI, high mutation count, low copy number variations	CIN, low-moderate mutation count and copy number variations	CIN, moderate mutation count, low to moderate copy number variations	CIN, low mutation count, high copy number variations
Epigenetics	High methylation	Low methylation	Moderate methylation	Loew methylation
Affected pathways	Immune activation, *JAK-STAT* activation	WNT and MYC activation, EGFR or SRC activation, VEGF or VEGFR activation, Integrin activation, TGFβ activation, IGF and IRS2 activation, HNF4α, HER2 and cyclin upregulation	DNA damage repair, Glutaminolysis, Lipidogenesis, Cell cycle	Mesenchymal activation, complement activation, immunosuppressi-on
Affected genes	*MSH6, RNF43, ATM, TGFBR2, BRAF, PTEN*	*APC, KRAS, TP53, PIK3CA*	*APC, KRAS, TP53, PIK3CA*	*APC, KRAS, TP53, PIK3CA*

Table 1: Summarization of molecular characteristic of the four CMSs (Guinney et al. 2015). Needles to be said, that not a single group can be defined by specific mutation, because no mutations were found to be specific to one group (Guinney et al. 2015). MYC *(*MYC Proto-Oncogene, BHLH Transcription Factor), *SRC* (SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase), *VEGF* or *VEGFR* (Vascular Endothelial Growth Factor or Receptor), *VEGF* or *VEGFR* (Vascular Endothelial Growth Factor or Receptor), *HNF4α* (Hepatocyte Nuclear Factor 4 Alpha)

CMS classification can not only help in treatment decisions, but also provide information leading to novel treatment strategies (Lenz et al. 2019). Furthermore, a study of patients with metastatic CRC who underwent first-line chemotherapy in combination with anti-VEGF (targeting the angiogenesis pathway) or anti-EGFR therapy found, that CMS2 patients had a longer median survival than other groups, followed by CMS4, CMS3, and finally CMS1 (Lenz et al. 2019), which was also confirmed in a more recent study (ten Hoorn et al. 2022). Sveen et al. and later Lenz et al. also showed differences in response to different treatments (including anti-VEGF, anti-EGFR, HSP90 (Heat Shock Protein 90) inhibitors) between CMS groups (Sveen et al. 2018; Lenz et al. 2019).

3.2.1.3 Molecular testing to identify hereditary syndromes

MSI testing, routinely used for tumor molecular subtyping, may suggest the presence of Lynch syndrome (Cervantes et al. 2023). This must be then confirmed by testing for germline mutations in MMR genes (Cervantes et al. 2023). Also, testing for *BRAF* mutation can be helpful in distinguishing Lynch syndrome from a sporadic form of CRC with MSI, as it has been shown that *BRAF* mutation does not accompany MSI with mutations in MMR genes associated with Lynch syndrome (Deng et al. 2004). Other hereditary syndromes can be tested in case of genetic indications based on the tumor subtype, age of the patient, accumulation of malignancies in their family history etc.,, for the mutations described in section 3.1.3.

3.3 Cell free tumor DNA

Tumor biopsy is the standard for CRC diagnosis and molecular testing, but there is an increasing demand for more non-invasive methods, such as liquid biopsies. Circulating tumor DNA, ctDNA, appears to be a novel promising source to meet this demand, as the information obtained from tumor biopsy and ctDNA are complementary.

Circulating free DNA in the plasma of healthy individuals was first discovered by Mandel and Metais in 1948 (Mandel a Metais 1948), and it was later shown in the plasma of cancer patients that this cell-free DNA can contain cancer-specific alterations (M. Stroun et al. 1989).

ctDNA can be found in the plasma of cancer patients as 180-200 bp fragments (Fan et al. 2010). In this project we focus on ctDNA in plasma, but ctDNA can also be found in other body fluids such as urine or serum (Chan, Chiu, a Lo 2003). The sources of ctDNA

can be the primary tumor as well as metastases and circulating tumor cells (Haber a Velculescu 2014), but there is still much unknown about the mechanisms of release of ctDNA into body fluids. One of the possible mechanisms of release is apoptosis or necrosis, referred to as passive release (Jahr et al. 2001). Release by apoptosis is supported by the size of the ctDNA fragments, 180 bp, as they are approximately the size of the DNA wrapped around the nucleosome, although larger fragments of ctDNA can also be found, supporting the idea of release by necrosis (Jahr et al. 2001). Another possible model of release is active release. This phenomenon was demonstrated using the cancer cell line HL-60, in which Stroun et al. showed to be spontaneously releasing their DNA (M Stroun et al. 2001). Circulating tumor cells also contribute to the amount of ctDNA found in plasma, but are not the main source as the amount of circulating tumor cells is very low (Punnoose et al. 2012). It has also been suggested that all living cells release DNA into the circulation (Maurice Stroun et al. 2001).

Several studies have shown that the amount of ctDNA found in plasma depends on tumor type, size, stage, and metastases present. More advanced tumors with metastases show higher levels of ctDNA in plasma compared to localized tumors (Bettegowda et al. 2014). Differential ctDNA levels were shown in different CRC stages, with significantly higher levels in stage IV (Yang et al. 2018). Franttini et al. showed that the level of ctDNA was significantly higher in patients with CRC than in tumor-free patients during follow-up and that the level of ctDNA increased with recurrence or metastasis (Frattini et al. 2006). This indicates, that ctDNA levels can be used as a prognostic marker or marker of disease recurrence. For example, ctDNA levels may indicate the potential for disease recurrence in patients with stage II CRC following surgical removal of the tumor, and thus indicate the need for adjuvant chemotherapy (Tie et al. 2016). Changes in ctDNA concentration levels are primarily detected using qPCR (Iizuka et al. 2006).

Another way to use ctDNA as a biomarker is to test it for qualitative changes, such as gene mutations or DNA copy number variations. Wang et al. confirmed that tumor mutations found in ctDNA correspond to those found in tumor tissue. They simultaneously detected *APC, TP53*, and *KRAS* mutations in the plasma and tumor tissue of CRC patients and additionally discovered that the detection rate was higher in patients with more advanced disease or metastases (J.-Y. Wang et al. 2004). The concordance between data obtained from ctDNA and tumor tissue was described as 89,5 % for the presence of any *KRAS* mutation in a recent study (Salvianti et al. 2021), but in another study it was also shown that the

concordance can vary, ranging from 14 % to 100 % (Molnár et al. 2019). Analysis of mutations present in ctDNA has been shown to be important in addressing treatment resistance. In one study, ctDNA analysis helped explain resistance to anti-EGFR therapy in the absence of *KRAS, NRAS* or *BRAF* mutations in tumor tissue (Siravegna et al. 2015). In the same study, ctDNA analysis was used to explain secondary resistance to anti-EGFR therapy, when emerging *KRAS* mutations were found in plasma samples (Siravegna et al. 2015). Monitoring of ctDNA may also support rechallenge of treatment, as *KRAS* mutations have been shown to decrease after withdrawal of anti-EGFR therapy, suggesting that rechallenge of this treatment may be possible (Siravegna et al. 2015). It has been shown that ctDNA can also be used to detect copy number variations, but this discovery was accompanied by some limitations, such as the problem of detecting small copy number gains or losses and the inability to detect the mutation status or methylation phenotype of affected genes (J. Li et al. 2017).

All these findings suggest that ctDNA is a valuable biomarker with the advantage of being non-invasive and timely and spatially representative. However, there are still areas that need improvement, such as optimization of assay methods, standardization of sample detection procedures, and quantitative standards. Needless to say that detection ability with ctDNA is dependent on the clinical stage of CRC and with that linked the amount of ctDNA and detection technologies.

4. Materials and Methods

4.1 Materials

4.1.1 Patients

Eighteen patients with metastatic colorectal cancer were enrolled into this project, from whom FFPE (formalin-fixed paraffine-embeded) samples from the primary tumor and metastases, and blood samples were collected at different time points (at the diagnosis, during and after the treatment, at relapse or disease progression); in total 133 samples were obtained and analyzed. Samples were collected after ethics committee approval and informed consent signed by the patients. Patients included in this project were treated at the Thomayer University Hospital, Prague.

Each patient also had to have some plasma samples from different treatment periods. In the table below (Table 2), the cohort description is given including the patients age, gender, location of the primary tumor, histological type of the tumor, clinical stage at the time of diagnosis - based on the TNM staging (evaluation of tumor´s size (T), lymph node spread (N) and classification of distant metastases (M)), grade - based on the degree of differentiation of the tumor cells, the location of the metastasis, the treatment, the length of time the patients have been in our study - time from diagnosis to the last follow-up, the survival status, and the number of ctDNA samples collected. In some patients, the metastatic samples were not available for the analysis because these patients have not undergone their surgical removal, but all patients had metastasis during the course of their disease.

Total number of patients included		18
Gender (%)	Female	7 (39)
	Male	11 (61)
Age	Median	65 years
	Range	30-76 years
Localization of the primary tumor (%)	colon	13 (72)
	rectum	4 (22)
	rectosigmoid junction	1 (6)
Histopathological nature (%)	adenocarcinoma	14 (78)
	mucus-forming adenocarcinoma	4 (22)
Stage (%)	stage IV – primary metastatic	13 (72)
	stage I-III with subsequent relapse	5 (28)
Grade (%)	G2	18 (100)
Localisation of metastases (%)*)	liver	14 (78)
	lungs	4 (22)
	lymph nodes	3 (17)
	gynecologic	1 (5)
Therapy (%)	1 line	4 (22)
	2 lines	7 (39)
	3 lines	6 (33)
	5 lines	1 (6)
Number of ctDNA samples (%)	Median	5,5
	Range	3-11
Time from diagnosis till last check-up (in months)	Median	25
	Range	8-57
Last known condition (%)	Died	13 (72)
	Alive	3 (17)
	Unknown	2 (11)

Table 2: CRC patients cohort description. ***)** – more than one site of metastases present in some of the patients

Seven out of 18 enrolled patients were female (39 %) and 11 were male (61 %), these figures are well in line with the epidemiology of the disease (White et al. 2018). The median age of these patients was 65 years, with the range from 30 years to 76 years. The majority of cases were colon cancer (13 patients, 73 %). 72 % of patients (13) were already in stage 4 disease at the time of diagnosis, 5 patients (28 %) were diagnosed with limited stage disease who subsequently relapsed with distant metastases. The most common organ where metastases occurred was the liver, found in 78 % of patients (14 out of 18). Liver is the most common organ in colorectal cancer where metastasis occurs (Filip et al. 2020). At the end of my project, 13 patients had already died, 3 were still alive at that time, and about two

patients we do not have any information about the condition (one of them decided to leave the study).

4.1.2 DNA sources

DNA was isolated from two types of sources - FFPE primary tumor and metastasis samples and blood.

FFPE (formalin-fixed paraffin-embedded) samples are preserved tissue sections of the tumor and serve as a direct source of tumor DNA. Where possible, FFPE samples from both the primary tumor and the metastasis were taken from the patient. It is a great source of information about mutations that may be responsible for the tumor's formation or progression, but it is considered an invasive method because it requires surgical removal of the tumor or metastasis. Another issue with FFPE samples is the quality of the DNA, which has been reported to be dependent on how the tissue was collected and processed (Arreaza et al. 2016).

From the peripheral blood, plasma was separated by centrifugation (2 200 × g for 20 minutes with slow ramping) immediately after the blood collection and stored at -80 °C. The buffy-coat rich interlayer was also collected and frozen down.

Plasma samples were used to isolate ctDNA. It was collected over an extended period of time at different stages of treatment – at the beginning of a new line of chemotherapy/targeted therapy, every 3 months during the therapy and at the end of the treatment. This is supposed to serve as a non-invasive method of monitoring disease progression and as a predictive marker for relapse, as research has shown that levels of tumor ctDNA are low to non-existent in patients without disease recurrence and elevated in those with disease recurrence (Frattini et al. 2006).

In order to analyze the mutations present in the tumor, it is necessary to eliminate germline mutations, which are more likely to be just polymorphisms. Blood leucocytes from the collected buffy coat were used to isolate germline DNA which served as a background control. By comparing sequencing data from germline DNA with sequencing data from ctDNA and DNA from FFPE samples, we were able to distinguish between somatic and germinal mutations.

4.2 Methods

4.2.1 Isolation of DNA

4.2.1.1 Isolation of ctDNA

The QIAamp® Circulating Nucleic Acid Kit (Qiagene) was used to isolate ctDNA. This kit allows silicate membrane purification of DNA using QIAamp Mini Columns on a vacuum manifold. The use of a vacuum pump instead of collection tubes and centrifugation to collect liquid allows safe and clean extraction without the risk of sample cross-contamination and isolation of up to 24 samples at a time.

Tubes containing plasma samples were first centrifuged for 3 minutes at 10000 × g at 4 ºC to ensure that everything was collected from the tube and that only plasma was used for the next step, without any residual cells that could carry contamination from genomic DNA. A 50ml centrifuge tube was filled with 100 µl of Proteinase K and 1 ml of the patient's plasma. To this mixture, 0,8 ml of ACL Buffer Mix was added (well mixed by pulse vortexing) and the 50ml tube was immediately mixed by pulse vortexing. ACL buffer was prepared according to the table below (Table 3) based on the number of samples isolated..

	Buffer ACL (ml)			
Number of samples	■	▲	●	Carrier RNA in Buffer AVE (µl)
1	0.9	1.8	2.6	5.6
2	1.8	3.5	5.3	11.3
3	2.6	5.3	7.9	16.9
4	3.5	7.0	10.6	22.5
5	4.4	8.8	13.2	28.1
6	5.3	10.6	15.8	33.8
7	6.2	12.3	18.5	39.4
8	7.0	14.1	21.1	45.0
9	7.9	15.8	23.8	50.6
10	8.8	17.6	26.4	56.3
11	9.7	19.4	29.0	61.9
12	10.6	21.1	31.7	67.5
13	11.4	22.9	34.3	73.1
14	12.3	24.6	37.0	78.8
15	13.2	26.4	39.6	84.4
16	14.1	28.2	42.2	90.0
17	15.0	29.9	44.9	95.6
18	15.8	31.7	47.5	101.3
19	16.7	33.4	50.2	106.9
20	17.6	35.2	52.8	112.5
21	18.5	37.0	55.4	118.1
22	19.4	38.7	58.1	123.8
23	20.2	40.5	60.7	129.4
24	21.1	42.2	63.4	135.0

Table 3: Preparation of ACL Buffer Mix. The green square marks the volume when 1 ml of plasma is used, the blue triangle marks the volume when 2 ml of plasma is used and the red circle marks the volume when 3 ml of plasma is used. In our case, the ACL buffer was prepared by selecting the volume of ACL in the column marked with green squares (since 1 ml of plasma was used) based on the number of samples isolated. Same with the carrier RNA. (QIAamp Circulating Nucleic Acid Handbook 10/2019)

The whole mixture was then incubated for 30 minutes at 60 ºC in a water bath. Meanwhile, the vacuum pump was assembled (as shown in the picture below (Fig. 3)).

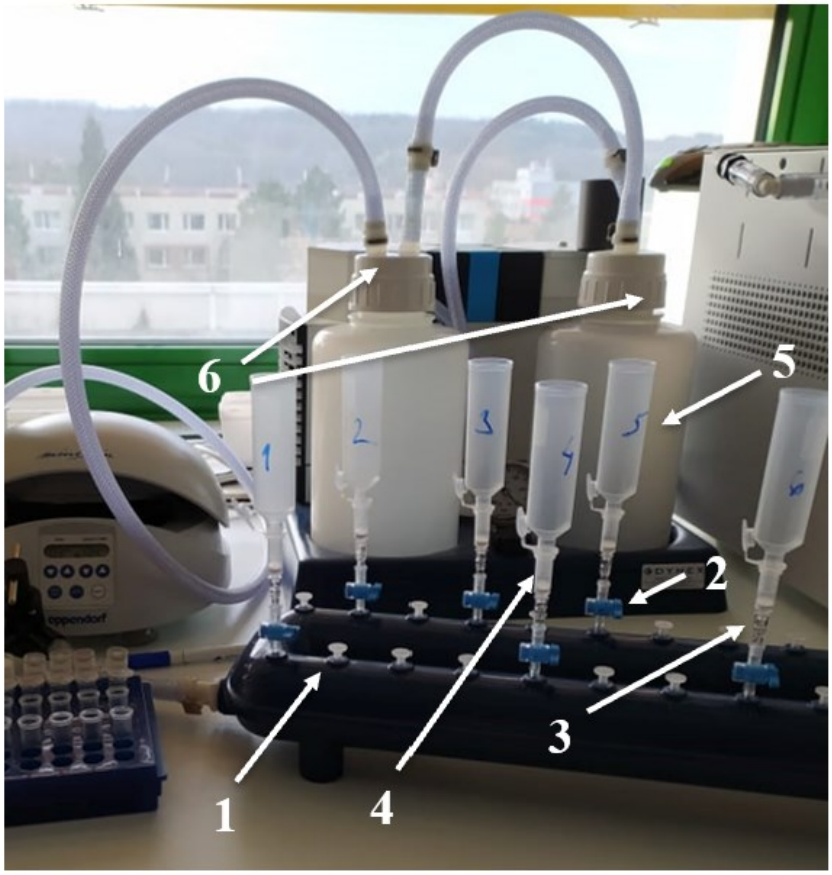

Fig. 3: Assembly of the vacuum pump. Assembly has been carried out in accordance with the manufacturer's instructions. The picture shows the Luer slot of the QIAvac 24 Plus (1) (closed with a Luer plug - blocking unused slots), VacValve (2) - closing and opening the VacValve changes the flow of the liquid, VacConnector (3) - keeps the VacValve and column separated, preventing contamination, QIAamp Mini column (4), Tube Extender (5), which allows larger volumes of liquid to be transferred. Unwanted liquid was collected in one of the collection bottles (6).

After incubation, 1,8 ml of ACB buffer was added to the mixture and the tube was immediately mixed by pulse vortexing and placed on ice for 5 min. Using the tube extender, the mixture was transferred to the QiAamp Mini column, which was already set up in the vacuum pump system. The pump was started and after all the liquid had passed through, the washing step was performed by adding 600 µl of ACW1 buffer, 750 µl of ACW2 buffer and finally 750 µl of 100% ethanol to the membrane. To ensure that all the ethanol was collected, the QiAamp Mini column was placed in a clean collection tube and centrifuged at 20000 × g for 3 min at room temperature. The QiAamp Mini column was then placed in a new clean collection tube and incubated in the heating block for 10 min at 56 ºC with the cap open to evaporate the last remaining ethanol from the membrane. After incubation, the column was transferred to a new clean 1,5ml microcentrifuge tube and 35 µl of AVE buffer was added to the membrane to elute the DNA. After incubation for 3 minutes

at room temperature, the isolated DNA was collected in the 1,5ml microcentrifuge tube by centrifugation for 1 minute at 20000 × g. The concentration was measured after isolation using the Qubit™ Flurometer 4.0 as described in section 4.2.2.

4.2.1.2 Isolation of DNA from FFPE samples

DNA from FFPE samples was isolated using the GeneRead™ DNA FFPE Kit (Qiagene). The first step of the isolation protocol using this kit is to remove the preservative paraffin and reverse the formalin cross-linking of the DNA. The DNA is then bound to a membrane of the QIAamp MinElute column provided by the manufacturer.

FFPE samples were prepared at the pathology department of the hospital as 10 µm sections of the tumor samples in 1,5ml microcentrifuge tubes. As the first step, 160 µl of deparaffinization solution was added to the FFPE samples in a 1,5ml microcentrifuge tube and the mixture was immediately mixed by pulse vortexing to ensure that the slices were broken up and covered as much as possible by the solution. The samples were then incubated for 3 minutes at 56ºC in a heating block. After the three minute incubation, 55 µl RNA-free water, 25 µl FTB solution and 20 µl Proteinase K were added to the 1,5 ml microcentrifuge tube, the tube was mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all the liquid. To successfully remove all paraffin and reverse crosslinking, incubation was performed at 56 ºC for 1 hour and then at 90 ºC for 1 hour in a heating block. After the incubation, the DNA was dissolved in the transparent part of the liquid, the blue part of the liquid should contain paraffin, so the clear liquid with the desired DNA was transferred to a new 1,5ml microcentrifuge tube and 115 µl of RNA-free water was added, the tube was then mixed by pulse vortexing. The next step was to add 35 µl of UNG (Uracil N-Glycosilase enzyme). This step is essential for removal of deaminated cytosine residues. Before the next 1 hour incubation at 50 ºC in a heating block, the tube was mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all the liquid. After incubation, the tube was centrifuged again at 2000 × g for 15 s to collect all the liquid. 2 µl of RNase A was added to the mixture, the tube was mixed by pulse vortexing, centrifuged for 15 s at 2000 × g to collect all liquid and incubated for 2 min at room temperature. Then 250 µl of Lysis Buffer AL was added, the tube was mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all the liquid. Finally, 250 µl of 100% ethanol was added to the mixture. To ensure

that everything was well mixed, the tube was again mixed by pulse vortexing and then centrifuged at 2000 × g for 15 s to collect all the liquid before transferring it to the QIAamp MinElute column (in a 2ml collection tube). The QIAamp MinElute column with all the transferred liquid was centrifuged for 1 min at 14000 × g, the collected liquid was discharged and the collection tube was reused. A first step of washing off unwanted residues was then performed by adding 500 µl of AW1 buffer. The column was centrifuged at 14000 × g for 1 min to collect the liquid, which was later discharged and the collection tube reused. The second washing step was performed with 500 µl of AW2 buffer, the column was centrifuged at 14000 × g for 1 min, the collected liquid was discharged and the collection tube was reused once again. The final wash step was performed with 250 µl of 100% ethanol, the column was centrifuged for 1 min at 14000 × g, the collected liquid was removed with the collection tube and the QIAamp MinElute column was placed in a new collection tube. To ensure that all ethanol was collected, the column was centrifuged again for 1 min at 14000 × g in a new collection tube. To elute the DNA, the QIAamp MinElute column was transferred to a clean 1,5ml microcentrifuge tube and 30 µl of Elution Buffer ATE was added to the membrane. Incubation was then performed for 1 min at room temperature. The DNA was collected by centrifugation at 14000 × g for 1 min. The amount of DNA recovered was measured using the Qubit™ Flurometer 4.0 as described in section 4.2.2.

4.2.1.3 Isolation of germ-line DNA

Germline DNA was isolated using the QIAamp® DNA Blood Mini Kit (Qiagen). This kit purifies DNA through a silicate membrane in a QIAamp Mini Spin Column (in a 2ml collection tube) supplied by the manufacturer.

As a first step, 20 µl of Proteinase K, 200 µl of patient´s blood enriched with leucocytes and 200 µl of AL buffer was mixed in a clean 1,5ml microcentrifuge tube and the whole mixture was then mixed by pulse vortexing. To successfully lyse the proteins with Proteinase K, the samples were incubated for 10 minutes at 56 ºC in a heating block. After incubation, 200 µl of 100% ethanol was added and the mixture was mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all the liquid from the walls of the tube. The entire mixture was transferred to the QIAamp Mini Spin Column (in a 2ml collection tube) and the column was then centrifuged at 6000 × g for 1 min. The collection tube containing the collected liquid was discharged

and the column was transferred to a new collection tube. 500 µl of buffer AW1 was added and the whole tube was centrifuged again for 1 min at 6000 × g. This was done as a first step to wash away unwanted residues. The collected liquid was removed with the collection tube and the column was placed in a new collection tube. As the next washing step, 500 µl of AW2 buffer was added and the tube was centrifuged at 6000 × g for 3 min. The collection tube was discharged and the column was transferred to a clean 1,5ml microcentrifuge tube. 200 µl of AE buffer was added to elute the isolated DNA. The column in the microcentrifuge tube was incubated for 1 min at room temperature and then centrifuged for 1 min at 6000 × g to collect the isolated DNA. The concentration of the isolated DNA was measured using Qubit™ Flurometer 4.0 as described in section 4.2.2.

4.2.2 Quantity control after isolation

After each isolation step, the amount of DNA isolated was checked using the Qubit™ 4.0 Fluorometer (Thermo Fisher Scientific). The principle of this method is the detection of fluorescent signals emitted by a fluorescent dye bound to double-stranded DNA. In this project, the Qubit™ dsDNA HS Assay Kit (Invitrogen) and 1x dsDNA HS (high sensitivity) assay were used to measure the concentration. This kit can detect concentrations in the range of 0,005 to 120 ng/µl. Due to the nature of ctDNA found in plasma, a more sensitive detection kit of concentration is needed, this kit allows us to detect low concentration of ctDNA as well as the higher concentration of germline DNA and DNA from FFPE.

Before measuring the concentration of the sample, two standards supplied by the manufacturer were measured to ensure that the instrument is measuring correctly. The measured fluorescence units (RFUs) of the standards are also used to calculate the concentration of the sample by the instrument. 10 µl of Qubit™ dsDNA HS Standard #1 or #2 were mixed with 1 µl of fluorescent dye Qubit™ dsDNA HS Reagent*200X and with 190 µl of Qubit™ dsDNA HS Buffer in a Qubi™ Assay Tube and measured. To measure the concentration of the sample, 1 µl of the sample was mixed with 1 µl of the fluorescent dye Qubit™ dsDNA HS Reagent*200X and with 199 µl of Qubit™ dsDNA HS Buffer in the Qubit™ Assay Tube.

The average concentration of DNA from FFPE samples was approximately 30 ng/µl, ctDNA from plasma samples 1,2 ng/µl and of the germline DNA from blood 13,7 ng/µl,

which represents the total DNA amount of 30 ng from FFPE samples, 42 ng from plasma, and 2740 ng from blood respectively, obtained from the whole processed samples.

4.2.3 Fragmentation of DNA

The first step required prior to the preparation of sequencing libraries is the fragmentation of the isolated DNA. ctDNA occurs in plasma in fragments of 180-200 bp in length (Fan et al. 2010) so there is no need to fragment it. However, DNA from FFPE and blood samples is isolated in larger fragments, so fragmentation is required.

The samples were fragmented by sonication or enzymatically. FFPE samples were fragmented by sonication using the Covaris S220 Focus ultrasonicator (Covaris). This method can be used for samples of higher concentration - for successful fragmentation it is needed at least 50 ng of DNA in 50µl volume, but the optimum is 200 ng od DNA in 50 µl. The germline DNA was fragmented differently to ensure better fragmentation due to its lower concentration and faster workflow by enzymatic fragmentation using SureSelect XT HS and XT low Input Enzymatic Fragmentation Kit (Agilent).

4.2.3.1 Fragmentation by sonication with Covaris S220 Focus ultrasonicator

Prior to sonication, samples had to be of the desired concentration. Ideally, the concentration should be 200 ng in 50 µl, but not every sample had this concentration, so concentrations of 100 ng in 50 µl and 50 ng in 50 µl were also allowed. The amount of DNA was calculated based on the concentration measured with the Qubit™ 4.0 Fluorometer (Thermo Fisher Scientific) (see section 4.2.2). The concentration of DNA was measured in ng/µl and the total volume after isolation was 30 µl, which means that the volume of DNA was calculated as follows: $V_1 = \frac{c_2 * V_2}{c_1} = \frac{200\ or\ 100\ or\ 50\ ng/50\mu l * 50\mu l}{c1}$; where c_2 is the desired concentration of 50/100/200 ng in 50 µl; V_2 is the desired volume of 50 µl; c_1 is the concentration measured after isolation using Qubit™ Flurometer 4.0 (Thermo Fisher Scientific). The volume was completed using Nuclease Free Water. The dilution was done using 1,5ml Eppendorf DNA LoBind® tube (this tube, according to the manufacturer, reduces sample-surface binding so that DNA recovery is maximized) before transferring the total volume to a special microtube designed for sonication.

After transfer, the microTUBE was centrifuged at 2000 × g for 15 s to ensure that there was no liquid on the tube´s walls. The 200 bp cardio program was selected

to produce fragments around 200 bp (settings were set by a lab technician as follows: 175 W (peak incident power), duty factor of 10 %, 200 cpb (cycles per burst), and treatment time was 170 s), and the microTUBE was inserted into the Covaris machine. After sonication, all liquid was collected from the walls of the microTUBE by centrifugation for 15 s at 2000 × g, and the fragmented DNA was transferred to a clean 1,5ml Eppendorf DNA LoBind® tube.

4.2.3.2 Enzymatic fragmentation

The SureSelect XT HS and XT low Input Enzymatic Fragmentation Kit (Agilent) was used for enzymatic fragmentation of germline DNA. Prior to starting this protocol, the vial of 5X SureSelect Fragmentation Buffer must be thawed, mixed by pulse vortexing, and stored on ice until use. Enzymatic fragmentation was performed using 0,2ml thin-walled strips with caps.

Germline DNA was diluted to a concentration of 10-200 ng in a volume of 7 µl using Nuclease Free Water. The amount of DNA and water was calculated using the following formula: $V_1 = \frac{c_2 * V_2}{c_1} = \frac{10-200ng/7\mu l * 7\mu l}{c1}$; where c_2 is the desired concentration of 10-200 ng in 7 µl; V_2 is the desired volume of 7 µl; c_1 is the concentration measured after isolation using Qubit™ Flurometer 4.0. The diluted sample was prepared in a 0,2ml thin-walled strip with a cap. 2 µl of 5X SureSelect Fragmentation Buffer was then mixed with 1 µl of SureSelect Fragmentation Enzyme, the solution was mixed by pipetting up and down and the entire 3µl volume was added to the diluted sample. Again, the mixture was mixed by pipetting up and down and the capped strip was transferred to the thermocycler where a three-step program was run, the protocol is shown in the table below - Table 4.

Temperature	Time	
37 ºC	15 min	**Preheated lid on at 100 ˚C**
65 ºC	5 min	
4 ºC	hold	

Table 4: Program for enzymatic fragmentation with SureSelect XT HS and XT low Input Enzymatic Fragmentation Kit by Agilent in thermocycler.

After enzymatic fragmentation, 40 µl of Nuclease Free Water was added to the sample and the strip was placed on ice. Sequencing libraries must be prepared immediately after, as this is not a stopping point in the workflow.

4.2.4 Preparation of sequencing libraries

Preparation of sequencing libraries using the SureSelectXT HS Target Enrichment System Kit for Illumina Paired-End Multiplexed Sequencing Library (Agilent) is a two-step process. For both steps, 0,2ml thin-walled, capped strips are used instead of 1,5ml microcentrifuge tubes because the libraries are prepared in smaller volumes. Prior to use, all reagents except the polymerases and the ClearSeq Comprehensive Cancer Panel (Agilent) must be mixed by pulse vortexing and centrifuged at 2000 × g for 15 seconds to ensure reagent quality after thawing. Polymerases and ClearSeq Comprehensive Cancer Panel (Agilent) can only be centrifuged for 15 s at 2000 × g to collect all the liquid, pulse-vortexing could damage them. All reagent mixes were prepared with a 0,5% reserve due to pipetting bias. Also, each time the cap of the strip was removed, it was replaced with a new one to prevent contamination.

First, there is the repair of ends of DNA fragments. During manipulation, isolation and also by the nature of fragment´s, ends can be damaged and may have unwanted overhangs - this is problematic for ligation of adapter oligos that are ligated after reparation and are necessary for clonal amplification and sequencing. Two identical duplexes of adapters are ligated to both ends of the DNA fragments; these adapter oligos are typically 50 to 60 nucleotides long and contain a unique molecular identifier that is necessary for distinguishing PCR duplicates and sequencing primers (Schiemer 2011). In the second step, the DNA is hybridized to target-specific probes, in our case to the ClearSeq Comprehensive Cancer Panel (Agilent). This ensures that only regions (exons, exon-intron boundaries and selected introns) of the 151 genes most associated with cancer are sequenced. After each step, quality and quantity control is performed by on-chip capillary electrophoresis using an 2100 Bioanalyzer Instrument (Agilent) (described in section 4.2.5) and a Qubit™ 4.0 Fluorometer (Thermo Fisher Scientific) (described in section 4.2.2).

4.2.4.1 End repair and ligation of adaptor oligos

Library preparation requires 10-200 ng of DNA. Due to the overall lower concentration of ctDNA, the total amount of isolated DNA was used (25 µl). However, 50 µl of sample volume is required for the first step of the library preparation, so the DNA volume was completed to 50 µl by adding 25 µl Nuclease Free H_2O. For DNA from FFPE samples fragmented by sonication on the Covaris

machine, the whole volume was used, since it was already diluted to 10-200 ng DNA in 50 µl with Nuclease Free Water. With DNA from blood samples it was proceeded from the end step of enzymatic fragmentation.

First, DNA ends were repaired. During manipulation and fragmentation, DNA ends can be damaged (Head et al. 2014), and they usually have 5' or 3' end overhangs, so to ensure that the 5' end of the DNA is phosphorylated, the 3' end of the DNA is adenylated, and the ends are blunt (this is needed for ligation of the adapters in the next step), end repair must be performed. EndRepair Mix was prepared by mixing 16 µl of EndRepair-A Tailing Buffer with 4 µl of EndRepair-A Tailing Enzyme Mix. 20 µl of End Repair Mix was mixed with 50 µl of DNA. The mixture was mixed by pulse vortexing, centrifuged at 2000 × g for 15 s to collect all the liquid, and incubated in a thermocycler according to the table below (Table 5).

Temperature	Time	
20 ºC	15 min	**Preheated lid on at 100 ˚C**
72 ºC	15 min	
4 ºC	hold	

Table 5: Incubating program for reparation of ends of DNA with End Repair mix in thermocycler.

Meanwhile, the ligation mix was prepared. 23 µl Ligation Buffer was mixed with 2 µl of T4 DNA Ligase. This mixture should be kept at room temperature prior to use. At the end of the repair program, samples were transferred on ice and 25 µl of ligation mix was added. The strip was capped with a new cap, mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all liquid. Then 5 µl of Adaptor Oligo Mix was added and the strip was again mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all liquid before incubation. Incubation was performed in the thermocycler according to the program in Table 6.

Temperature	Time	
20 ºC	30 min	**Preheated lid on at 100 ˚C**
4 ºC	hold	

Table 6: Incubation program for ligation of adaptor oligos in thermocycler.

After ligation, 80 µl of well mixed AMPure XP Beads (Beckman Coulter) (allowed to stand at room temperature for 30 minutes prior to use) were added to the

DNA to purify the sample. The mixture was vortexed, centrifuged at 2000 × g for 15 seconds and incubated at room temperature for 5 minutes. During these 5 minutes, the DNA binds to the beads so unwanted molecules can be easily washed away. After 5 minutes, the strip was transferred to a magnetic separation device. When the liquid cleared, approximately after 5 minutes, it was carefully removed and discharged, leaving the strip in the magnetic stand. Then 200 µl of 70% fresh ethanol was added to the beads. The strip was held in the magnetic stand and the beads were washed with the ethanol by pipetting up and down. This washing step was repeated twice. After the liquid cleared, approximately after 2 minutes, the ethanol was removed. After the second round of washing, it is critical to remove all ethanol so that elution with Nuclease Free Water can be performed in the next step. To ensure that the DNA was truly completely ethanol free, incubation was performed without a cap in an open thermocycler at 37 ºC for 2 minutes. After 2 minutes, 35 µl of Nuclease Free Water was added to elute the DNA from the beads. Before removing the liquid containing the DNA, the strip was incubated at room temperature for 2 minutes. After incubation, the strip was placed in the magnetic stand and when the liquid had cleared (approximately after 3 minutes), the liquid was transferred to a new strip and the beads were discharged.

The next step was PCR to amplify the repaired and ligated DNA. PCR mix was prepared according to the Table 7. 13,5 µl of this PCR mix was added to the DNA.

Reagent	Volume
Herculase II Fusion DNA polymerase	1 µl
5x Herculase II Reaction Buffer	10 µl
100 mM dNTP Mix	0,5 µl
Forward Primer	2 µl

Table 7: Volumes of reagents of PCR mix for amplifying the repaired and ligated DNA.

After the PCR mix was added to the DNA, 2 µl of SureSelect XT HS Index Primers (A01-H04) were added to the sample. These primers allow pooling of up to 32 samples for sequencing. The strip was mixed by pulse vortexing, centrifuged at 2000 × g for 15 s to collect all liquid, and loaded into the thermocycler where the PCR program was run according to the settings in the Table 8 below.

Segment	Temperature	Time	Number of cycles	
1	98 ºC	2 min	1	**Preheated lid on at 100 ˚C**
2	98 ºC 60 ºC 72 ºC	30 s 30 s 1 min	14	
3	72 ºC	5 min	1	
4	4 ºC	hold	1	

Table 8: Program for PCR of repaired and ligated DNA in thermocycler.

After PCR, a purification with AMPure XP Beads (Beckman Coulter) was performed. 50 µl of well mixed AMPure XP Beads (Beckman Coulter) was added to the amplified DNA and the mixture was incubated for 5 minutes at room temperature. The strip was then transferred to a magnetic stand and allowed to stand for approximately 5 minutes until the liquid had cleared. The liquid was removed, and two washes were performed with 200 µl of 70% fresh ethanol. The strip was left in the magnetic stand and the beads were washed by pipetting the ethanol up and down. When the liquid was clear (approximately after 2 minutes), the ethanol was removed. After the second wash, it is necessary to remove all ethanol to ensure that the beads are completely free of ethanol. For that incubation was performed at 37 ºC for 2 minutes in an open thermocycler without a cap. To elute the DNA, 15 µl of Nuclease Free Water was added to the beads, the strip was mixed by pulse vortexing, centrifuged at 2000 × g for 15 s to collect all liquid, and incubated for 2 minutes at room temperature. After incubation, the strip was transferred to the magnetic stand. The clear liquid (after approximately 3 minutes in the magnetic stand) was transferred to a new strip and the beads were discharged. After this last step, the DNA was ready for a second step of library preparation. As a control step before hybridizing the DNA to the capture library, the concentration was measured using the Qubit™ Flurometer 4.0 (Thermo Fisher Scientific) (section 4.2.2) and the quality was checked using capillary electrophoresis on the 2100 Bioanalyzer Instrument (Agilent) (section 4.2.5).

4.2.4.2 Hybridization of DNA samples to the Capture library

In the second step of library preparation, DNA is hybridized to the Capture Library - ClearSeq Comprehensive Cancer Panel (Agilent). This ensures that we collect only DNA fragments that contain the sequences of interest, in our case exons, exon-intron boundaries and selected introns of 151 genes most commonly associated with cancer. Sequencing only a small portion of the genome is cost-effective based on clinical use.

First, 5 µl of SureSelect XT HS and XT Low Input Blocker Mix was added to the sample from the first round of sequencing library preparation. The amount of DNA should be 500-1000 ng in 12 µl, in our case not all samples had this desired concentration, so the entire amount from the previous step was used so that no DNA was lost. The sample was mixed by pulse vortexing, centrifuged for 15 s at 2000 × g to ensure that all the liquid was collected and transferred to the thermocycler where the program was started according to the Table 9.

Segment	**Temperature**	**Time**	**Number of cycles**	
1	95 ºC	5 min	1	
2	65 ºC	10 min	1	**Preheated lid on at 100 ˚C**
3	65 ºC	1 min **(PAUSE here)**	1	
4	65 ºC 37 ºC	1min 3 s	60	
5	65 ºC	hold	1	

Table 9: Program in thermocycler for hybridization of DNA to Capture Library.

When the program enters segment 3 (1 min at 65 ºC), it must be stopped and 13 µl of Capture Library hybridization mix should be added to the sample. This mix is prepared according to the table below (Table 10). RNase Block is supplied at 100% concentration, to obtain the 25% concentration, 0,5 µl of RNase Block was mixed with 1,5 µl of Nuclease Free Water. The RNase Block solution was prepared in a separate 1,5ml microcentrifuge tube prior to addition to the Capture Library hybridization mix.

Reagent	Volume
Nuclease Free Water	3 µl
SureSelect Fast Hybridization Buffer	6 µl
Capture Library - ClearSeq Comprehensive Cancer panel	2 µl
25% RNase Block solution	2 µl

Table 10: Volumes of reagents of Capture Library hybridization mix.

When adding the 13 µl of Capture Library Hybridization Mix, the strip with the sample should be kept in the thermocycler and the mixture must be mixed first by pipetting the solution up and down, then by pulse vortexing and finally centrifuged at 2000 × g for 15 s to collect all the liquid. After all this, the strip should be put back into the thermocycler and the program could be resumed.

During hybridization, Dynabeads MyOne Strepavidin T1 (Thermo Fisher Scientific) magnetic beads were prepared. 50 µl of streptavidin beads (well mixed and resuspended) were added to a new strip. 200 µL of SureSelect Binding Buffer was added to the beads, the mixture was mixed by pipetting up and down, and the strip was placed in a magnetic stand. When the liquid cleared (approximately after 2 minutes), it was removed and another 200 µL of Binding Buffer was added while the strip remained in the magnetic stand. The beads were again washed by pipetting the buffer up and down. This was done a total of 3 times. After the last wash, the beads were resuspended in 200 µl of Binding Buffer and stored at room temperature until needed.

After the hybridization program, the sample (approximately 30 µl) was transferred to a new strip containing Dynabeads MyOne Strepavidin T1 (Thermo Fisher Scientific) magnetic beads and the strip was incubated at room temperature for 30 minutes at 1800 rpm on a 96-well plate mixer MixMate® (Eppendorf). During this time, 6 strips were prepared with 200 µl of Wash Buffer 2. The strips were transferred to a thermocycler where they were heated to 70 ºC until needed. After mixing, the strip was centrifuged at 2000 × g for 15 s to ensure that all the liquid was collected and transferred to a magnetic stand. After the liquid cleared (approximately after 5 minutes), the supernatant was removed and discharged. The beads were washed by adding 200 µl Wash Buffer 1, the strip was mixed by pipetting the liquid up and down and then transferred to a magnetic stand. The clear supernatant was

removed and discharged (approximately after 3 minutes in the magnetic stand). The beads were then washed with 200 µl of pre-warmed Wash Buffer 2. After mixing the buffer with the beads by pipetting the liquid up and down, the beads with the buffer were incubated in the thermocycler at 70 ºC for 5 minutes. After these 5 minutes, the strip was transferred to a magnetic stand and when the liquid cleared (approximately after 5 minutes), the supernatant was discharged and the beads were resuspended in 200 µl of preheated Wash Buffer 2. A total of 6 washes with Wash Buffer 2 and 5 minutes of incubation were performed. After the last wash, when the Wash Buffer 2 was completely removed from the beads using the magnetic stand, 25 µl of Nuclease Free Water was added to the beads and the strip was transferred to ice.

The next step was to amplify the captured library by PCR. The PCR mix was prepared according to the following Table 11.

Reagent	Volume
Nuclease Free Water	12,5 µl
5x Herculase II Reaction Buffer	10 µl
100 dM dNTP Mix	0,5 µl
Herculase II Fusion DNA Polymerase	1 µl
SureSelect Post-Capture Primer Mix	1 µl

Table 11: Volumes of reagents of PCR mix for amplification of captured library.

25 µl of PCR mix was added to the beads with the sample, the strip was mixed by pulse vortexing and centrifuged at 2000 × g for 15 s to collect all the liquid, then the PCR reaction was run in the thermocycler according to the program as shown in the Table 12.

Segment	Temperature	Time	Number of cycles	
1	98 ºC	2 min	1	
2	98 ºC 60 ºC 72 ºC	30 s 30 s 1 min	12	**Preheated lid on at 100 ˚C**
3	72 ºC	5 min	1	
4	4 ºC	hold	1	

Table 12: Program in thermocycler for PCR amplification after hybridization.

It is important to allow the AMPure XP Beads (Beckman Coulter) to equilibrate to room temperature for at least 30 minutes prior to use.

After PCR, the strip was centrifuged at 2000 × g for 15 s to collect all liquid and transferred to a magnetic stand. When the liquid cleared (approximately after 5 minutes), the supernatant was transferred to a new 0,2ml thin wall strip and 50 µl of well mixed AMPure XP Beads (Beckman Coulter) were added to the sample. The mixture was incubated for 5 minutes at room temperature and the strip was transferred to a magnetic stand. When the liquid cleared (approximately after 4 minutes), the supernatant was removed, and the beads were washed twice with 200 µl of 70% ethanol. During the ethanol washing, the strip was held in the magnetic stand and the liquid was mixed by pipetting up and down. After the last wash, the ethanol was removed so that no ethanol was left in the tube, which is necessary for elution with Nuclease Free Water in the next step. Next, the beads were incubated in an open thermocycler without a cap for 2 minutes at 37ºC to ensure that the beads were completely free of ethanol. To elute the DNA from the beads, 25 µl of Nuclease Free Water was added, the strip was mixed by pulse vortexing, centrifuged at 2000 × g for 15 s, and the mixture was incubated for 2 minutes at room temperature. The strip was transferred to a magnetic stand and the eluted DNA (clear supernatant) was transferred to a new 1,5ml Eppendorf DNA LoBind® tube (after approximately 3 minutes in the magnetic stand) and the beads were discharged. The concentration of the prepared libraries was measured using a Qubit™ Flurometer 4.0 (Thermo Fisher Scientific) (section 4.2.2) and the quality was checked using an 2100 Bioanalyzer Instrument (Agilent) (section 4.2.5).

4.2.5 Quality control

After both steps of sequencing library preparation, the quality was checked. To check the quality and also to determine the size of fragments chip capillary electrophoresis by 2100 Bioanalyzer Instrument (Agilent) was used. For this measurement, High Sensitivity DNA kit (Agilent) was used each time, chip for this assay allows measurement of 11 samples at once and can detect samples of size from 50 to 7000 bp.

The chip was prepared according to the manufacturer's instructions. For use, the reagents must equilibrate to room temperature for at least 30 minutes. First, the gel dye mix was prepared by mixing 15 µl of the well mixed High Sensitivity DNA Dye Concentrate

with all the liquid in the High Sensitivity DNA Gel Matrix tube. The mixture was mixed by pulse vortexing and transferred to a tube with a spin filter (provided by the manufacturer) and centrifuged for 10 minutes at room temperature at 2240 × g. After centrifugation, the filter was removed and the Gel-Dye Mix was ready for use..

Next, 9 µl of the gel-dye mix was pipetted into the well marked with G (A, Fig. 4) and the chip was transferred to the Priming Station. The plunger of the syringe of the Priming Station was positioned at 1 ml, the station was closed and the plunger was pressed down until it was held by the clip. The chip was held for 1 minute, then the syringe plunger was released to the 0,5 ml mark, from this position it was pulled back to the 1 ml position, the station was opened and the chip was removed. Then 9 µl of the dye-gel mixture was pipetted into the other wells marked with G (B, Fig. 4). 5 µl of High Sensitivity DNA Marker was pipetted into the 11 sample wells and the well marked with the ladder symbol (C, Fig. 4). 1 µl High Sensitivity DNA Ladder was pipetted into the well marked with the ladder symbol (D, Fig. 4) and then 1 µl of DNA samples was pipetted into the 11 sample wells (E, Fig. 4).

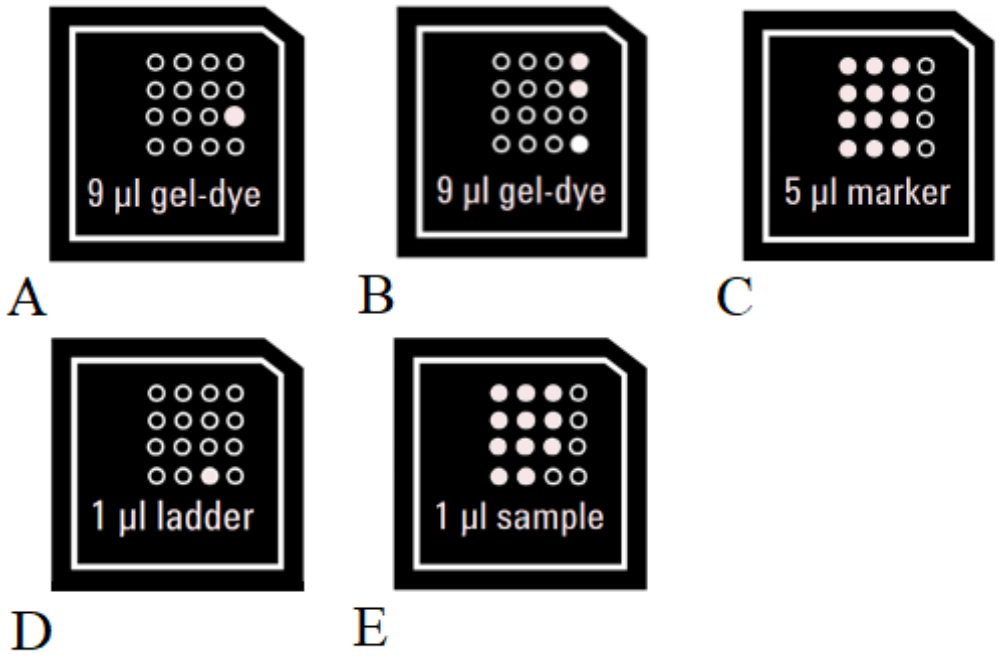

Fig. 4: **A**: position of the well marked by G; **B**: positions of the wells marked by G; **C**: positions of the wells of the 11 samples and the well for the ladder where 5 µl of High Sensitivity DNA marker was pipetted; **D**: position of the well for ladder marked by ; **E**: positions of the wells for 11 samples ("Agilent High Sensitivity DNA Kit Guide," n.d.)

Finally, the chip was transferred to an adapter of the IKA® MS3 Basic Vortex Mixer and mixed for 1 minute at 2400 rpm. The chip was then ready for analysis.

The High Sensitivity Assay program was selected for the analysis. An example of the results of the analysis can be seen on a picture below (Fig. 5).

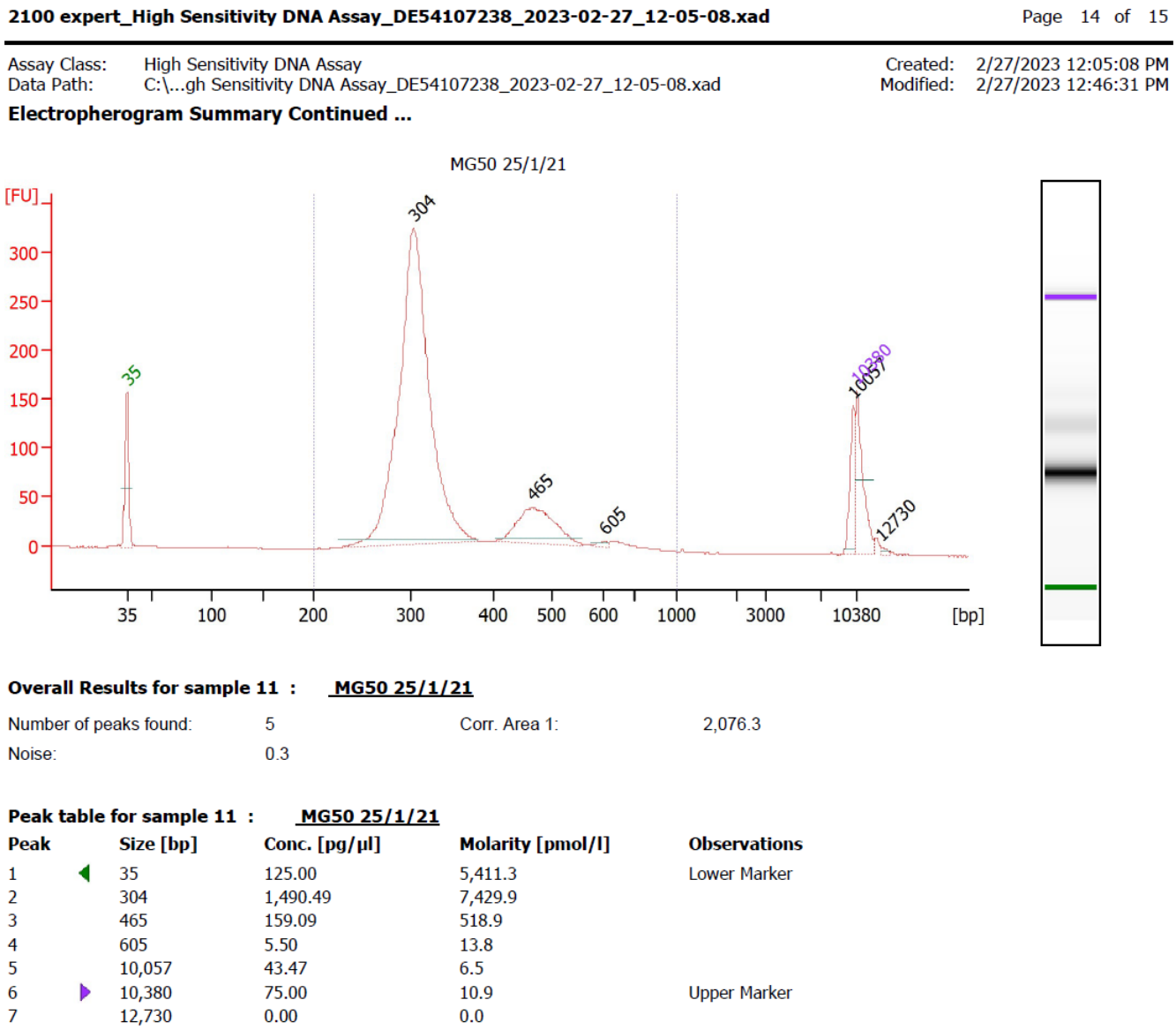

Fig. 5: Example result from 2100 Bioanalyzer Instrument (Agilent) on-chip capillary electrophoresis. The program generates a graph (1) showing the distribution of different sized fragments (on the y-axis is the fluorescence intensity in fluorescence units (FU), on the x-axis is the size of the fragment in base pairs (bp)) as well as the electrophoretic gel (2). Below is a table of all the fragments of different sizes detected and their concentration. This particular graph shows the quality control of the library prepared from one of the ctDNA samples, the peaks represent the expected size of the ctDNA, the most represented peak of size 304 bp (ctDNA fragment + adapters) is of 1,49 ng/µ concentration.

4.2.6 Sequencing

Our DNA samples were sequenced in the form of amplicon/targeted sequencing using ClearSeq Comprehensive Cancer panel (Agilent) covering 151 cancer-related genes (details available at https://www.agilent.com/en/product/next-generation-sequencing/hybridization-based-next-generation-sequencing-ngs/ngs-catalog-target-enrichment-probes/clearseq-comprehensive-cancer-232814#productdetails) and NextSeq 500/550 Mid Output Kit v2.5. The sequencing was performed on the NextSeq 500 Instrument (Illumina, San Diego, CA, USA), in a multiplexed manner with pair-end runs (2 x 150 bases). The first clusters of fragments on the flow cell are made by binding to oligos on the flow cell that are complementary to the adapters ligated to our fragments (section 4.2.4.1). After binding, all fragments are amplified by clonal bridge amplification. Illumina sequencing uses sequencing by synthesis technology - labeled nucleotides are detected while the DNA chain is copied in a massively parallel manner.

Prior to sequencing, DNA samples were diluted to a concentration of 4mM by mixing it with ddH2O in a new 1,5ml microcentrifuge tube. The volume was calculated as follows: first, the concentration obtained in ng/µl from the Qubit™ Flurometer 4.0 (Thermo Fisher Scientific) was converted to nM units, and then the volume of DNA and ddH2O was calculated using the formula c1*V1=c2*V2. When converting ng/µl to nM, we estimated that 1 mole of a base pair is 660 g. The entire calculation process is shown in the following figure (Fig. 6).

[1] $weight\ of\ the\ sample\ m_s = size\ in\ bp * 660g$

[2] $\frac{ng}{\mu l} = \frac{1}{10^9}\frac{g}{\mu l} = \frac{1}{m_s}\frac{mole}{\mu l} = \frac{10^6}{1}\frac{mole}{l} = M = \frac{1}{10^{-9}}nM = c_1$

[3] $c_1 * V_1 = c_2 * V_2 \rightarrow V_1 = \frac{4\ nM * 10\ \mu l}{c_1 nM}$

Fig. 6: Equations for calculating the volume of DNA to be sequenced. (1) Calculation of sample weight. Sample size was measured using an 2100 Bioanalyzer Instrument (Agilent) capillary electrophoresis. (2) Conversion of ng/µl units to nM. (3) Calculation of the volume of DNA required. c_2 is the desired concentration of 4nM, V_2 is the desired volume of 10 µl, c_1 is the calculated concentration of the sample in nM, V_1 is the volume of DNA required for sequencing.

After dilution, all samples ready for sequencing were mixed together in a new 1,5ml microcentrifuge tube to form a library pooling mixture of 30 samples. The selected flow cell allows sequencing of 30 samples at once thanks to the different index primers added in step

4.2.4.1. As a control, the PhiX Control v3 control library (Illumina) was added to the library mixture of 30 samples. According to the manufacturer's manual, 1-2% of this control library should be added in our case, so 3 µl of 0,4nM control library was added (PhiX Control v3 comes in a stock concentration of 10nM, so it needs to be diluted - by mixing 10 µl of 10nM PhiX with 15 µl of RBS buffer, the total volume of 25 µl of 4nM PhiX can be stored at -25 ˚C for up to three months). PhiX Control v3 is used as a quality control for clustering, sequencing and alignment.

After preparation of the library-control mix of all samples and the control library, DNA denaturation was performed by mixing 5 µl of the library-control mix and 5 µl of 0,2M NaOH in a new 1,5ml microcentrifuge tube. The mixture was mixed by pulse vortexing, centrifuged at 280 × g for 1 min and incubated at room temperature for 5 min. Then 5 µl of 200mM Tris HCl was added to stop the denaturation process, the mixture was mixed by pulse vortexing and centrifuged at 280 × g for 1 min. The total volume of the denatured library-control mix (15 µl) was then mixed with 985 µl of pre-chilled HT1 buffer to produce a library-control mix with a final concentration of 20pM, which is required for sequencing. The final mix was kept on ice until use. Meanwhile, the sequencing machine was prepared by loading a premade sample sheet table in .csv format of our samples and their index primers. Next, the machine was prepared according to the program instructions, the Mid Output Flow Cell Cartridge and Buffer Cartridge were inserted into the machine. Finally, the library was transferred to the Mid Output Reagent Cartridge into the well marked 10. 98 µl of the 20pM library-control mix was pipetted into the well with 1202 µl HT1 buffer to a final volume of 1,3 ml, resulting in a final concentration of the sequencing library of 1,5pM. As a last step, 3 ml of ddH2O with 30 µl of sodium hypochlorite solution was pipetted into the well numbered 28 as a wash solution.

The read length using the NextSeq 500 (Illumina) and the NextSeq 500/550 Mid Output Kit v2.5 (300 cycles) (Illumina) was 2 x 150 bp and the coverage per sample was 4 to 5 million pairs of reads.

4.2.7 Data analysis

R-Studio was used to process the sequencing data using a custom pipeline developed by one of the lab's bioinformaticians.

The analysis began by processing the FASTQ files generated by the sequencer with the FastQC command. This function performs a quality control of the raw reads. Based on the data obtained from the FastQC report, trimming of technical sequences (adapters) and low quality ends was performed. For control of the trimming, another FastQC was performed on the trimmed data.

The trimmed data were then processed using the custom pipeline. The pipeline includes the BWA command (alignment/mapping to a reference genome); Picard tools, namely the sorting tool that allows processing of BAM files generated after alignment; removing duplicates, and then variant calling. Variant calling includes the samtools mpileup command, which creates a text format from an alignment, and the VarScan tool for off-sequence variant detection. As a last step, the Qualimap command was used to perform quality control of the alignment. The identified single nucleotide variants (SNVs) were exported to excel files and then filtered using several sets of filtration criteria.

We were looking for nonsynonymous somatic gene alterations that are either significant for tumorigenesis or can be used as tumor specific markers.

The following rules were followed in the processing of the sequence data: the detected mutation could not be present in germline DNA; the mutated locus had to be read at least 3 times and the locus alone at least 5 times in primary tumor DNA, DNA of metastasis and ctDNA; the mutation had to be present with a frequency of at least 5% in tumor or metastatic DNA. Synonymous SNVs were excluded from further analysis.

5. Results

5.1 Analysis of the sequencing

The expected coverage per sample was 4 to 5 million pairs of reads, in reality the median of coverage per primary tumor samples was 9 million pairs of reads, with minimum of 1 million pairs of reads and maximum of 15 million pairs of reads; the median of coverage per metastatic samples was 10 million pairs of reads, with the range of 10 to 12 million pairs of reads; for ctDNA samples the median was 6 million pairs of reads, with the range of 5 to 12 million pairs of reads; the median for germline DNA samples was 9 million pairs of reads, with the range of 5 to 19 million pairs of reads. The plan was to sequence 30 samples at once, but one sequencing run was not run at full capacity, so the coverage may be higher. The coverage for primary tumor, metastatic, ctDNA and germline samples for each patient is presented in Table 13 with the additional information of the number of reads mapped to target sequence of Comprehensive Cancer Panel (Agilent).

At minimum 92,54 % of reads from all types of samples were mapped to the target sequence of Comprehensive Cancer Panel (Agilent), at maximum 99,93 % of reads were mapped, with median of 99,79 %. 3

ID	Number of reads PT	Mapped reads PT	Number of reads M (average)	Mapped reads M (average)	Number of reads ctDNA (average)	Mapped reads ctDNA (average)	Number of reads germline	Mapped reads germline
1	9,77E+06	9,75E+06	-	-	8,85E+06	8,83E+06	7,48E+06	7,47E+06
2	9,24E+06	9,20E+06	9,62E+06	9,60E+06	6,92E+06	6,89E+06	7,63E+06	7,63E+06
3	6,70E+06	6,68E+06	-	-	6,15E+06	6,12E+06	8,27E+06	8,27E+06
4	8,32E+06	8,30E+06	-	-	4,78E+06	4,76E+06	8,96E+06	8,96E+06
5	8,16E+06	8,14E+06	-	-	5,56E+06	5,54E+06	8,19E+06	8,19E+06
6	1,41E+07	1,41E+07	-	-	5,94E+06	5,90E+06	8,72E+06	8,72E+06
7	1,01E+07	1,01E+07	-	-	8,14E+06	8,12E+06	7,83E+06	7,82E+06
8	8,85E+06	8,83E+06	9,56E+06	9,54E+06	5,72E+06	5,68E+06	5,14E+06	5,13E+06
9	8,76E+06	8,75E+06	-	-	7,61E+06	7,59E+06	8,63E+06	8,62E+06
10	5,87E+06	5,85E+06	-	-	7,10E+06	7,06E+06	8,58E+06	8,57E+06
11	6,22E+06	6,21E+06	-	-	5,65E+06	5,62E+06	6,22E+06	6,21E+06
12	6,80E+06	6,76E+06	1,22E+07	1,21E+07	5,88E+06	5,85E+06	1,71E+07	1,70E+07
13	9,73E+06	9,72E+06	-	-	6,61E+06	6,59E+06	1,51E+07	1,51E+07
14	7,37E+06	7,35E+06	9,90E+06	9,88E+06	6,48E+06	6,46E+06	1,15E+07	1,15E+07
15	1,14E+07	1,14E+07	-	-	6,31E+06	6,29E+06	1,72E+07	1,72E+07
16	1,35E+07	1,34E+07			7,69E+06	7,66E+06	1,85E+07	1,85E+07
17	1,33E+06	1,23E+06	9,85E+06	9,84E+06	4,69E+06	4,66E+06	1,69E+07	1,69E+07
18	1,52E+07	1,51E+07	-	-	1,20E+07	1,19E+07	1,20E+07	1,20E+07

Table 13: Number of reads – coverage, and number of reads mapped to hybridization library in primary tumor DNA, metastatic DNA, ctDNA and germline DNA for each patient. PT= primary tumor; M= metastasis

The average coverage of variants found in primary tumor DNA from all patients was 582, with minimum of 22 and maximum of 982. The average coverage of variants found in metastatic DNA from all patients was 636, with minimum of 170 and maximum of 851. The coverage of variants in primary tumor DNA and metastatic DNA from each patient is shown in the Table 14.

ID		Average coverage of variants
1	primary tumor	700
2	primary tumor	415
	local relapse	538
3	primary tumor	544
4	primary tumor	457
5	primary tumor	700
6	primary tumor	962
7	primary tumor	556
8	primary tumor	585
	metastasis of liver	731
	metastasis of sigmoid colon	838
9	primary tumor	591
10	primary tumor	407
11	primary tumor	580
12	primary tumor	98
	metastasis of liver	507
	metastasis of lungs	170
	metastasis of liver	785
13	primary tumor	742
14	primary tumor	366
	metastasis of gynaecological origin	666
15	primary tumor	860
16	primary tumor	917
17	primary tumor	22
	metastasis of liver	851
18	primary tumor	982

Table 14: The average coverage of variants from primary tumor DNA and metastatic DNA for each patient enrolled in this project.

5.2 DNA mutations in primary tumors

A total of 141 nonsynonymous gene variants were found in the primary tumor samples, of which 122 were unique (87 %) – meaning they were found only once in the whole cohort of patients with CRC. The rest 13 % of found gene alterations were repeats of 6 gene variants. The most common unique variants were missense exon variants, accounting for 59 %. Second most common were variants leading to an early stop codon in the exon, accounting for 14 %, then frame shift exon variants - 7 %, followed by intron variants – 3 %, upstream gene variants and non-coding exon variants were each accounting for 2 % and the least common were disruptive inframe deletions account for only 1 %. A summary of the variants is displayed in the Table 15. The visual representation of variants found in primary tumor DNA can be seen on the graph below (Fig. 7).

	Variant type		**Count (Patients, N=18)**
Total variants found			141
Unique variants			122
	Upstream gene variants		2
	Intron variants		5
	Exon variants	Frame shift	10
		Missense	83
		Stop gained	19
		Non-coding exon variants	2
		Disruptive inframe deletions	1

Table 15: A summary of the variants found in the DNA of primary tumors. Unique variants are further divide by counts to upstream gene variants, intron variants and exon variants (which are further divided to frame shift, missense, stop gained, non-coding exon variants and disruptive inframe deletions).

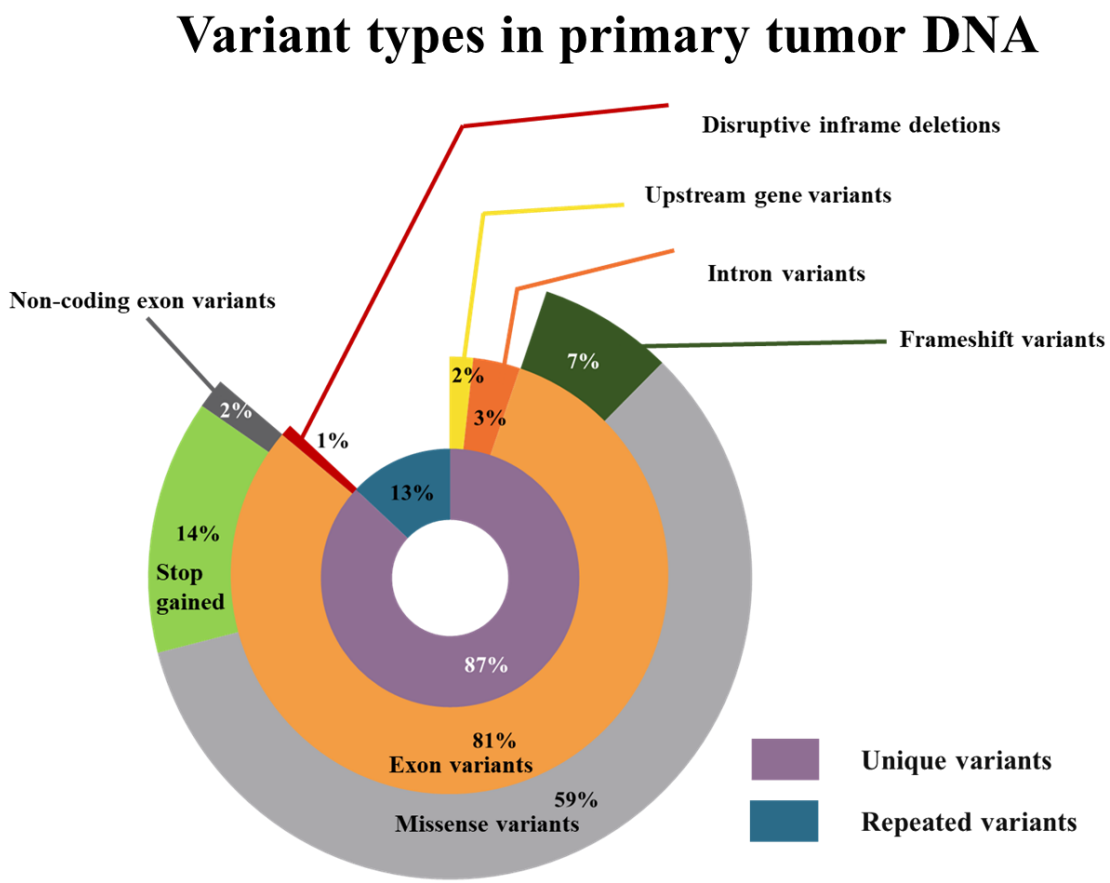

Fig. 7: Visual representation of variants found in DNA of primary tumor. 87 % of all variants found were unique – meaning there were found only once in the whole cohort of patients with CRC. 13 % were variants of 6 genes, that were found repeatedly in the whole cohort of patients with CRC. Out of the 87 % unique variants 81 % were exon variants, 3 % intron variants, 2 % upstream gene variants and 1 % disruptive inframe deletions. Out of the 81 % of exon variants 59 % were missense variants, 14 % were stop gained variants, 7 % were frameshift variants and 2 % were non-coding exon variants.

The number of genes mutated in the primary tumor DNA ranged from 2 to 24, with a median of 6. The three most commonly mutated genes were *TP53*, which was found in 13 samples, *APC*, which was found in 11 primary tumor samples, and *KRAS*, which was found in 8 primary tumor samples. The list of genes found to be mutated in the DNA of primary tumors is shown in the Table 16. Table 17 shows the gene variants found to be mutated more than once in the whole cohort of patients with metastatic CRC.

Gene	Patients, N=18	Frequency in cohort	Gene	Patients, N=18	Frequency in cohort
TP53	13	72 %	*CYP2D7*	1	6 %
APC	11	61 %	*ERBB2*	1	6 %
KRAS	8	44 %	*ERBB3*	1	6 %
DDX3X	5	28 %	*FGFR2*	1	6 %
PIK3CA	5	28 %	*FGFR3*	1	6 %
SMAD4	4	22 %	*FGFR4*	1	6 %
YES1	4	22 %	*FLT3*	1	6 %
FBXW7	3	17 %	*FLT4*	1	6 %
ZMYM3	2	11 %	*FSTL5*	1	6 %
DPYD	2	11 %	*HNF1A*	1	6 %
FLT1	2	11 %	*CHD7*	1	6 %
IL2RB	2	11 %	*IDH2*	1	6 %
KDR	2	11 %	*IKZF1*	1	6 %
LCK	2	11 %	*INPP4B*	1	6 %
LTK	2	11 %	*JAG2*	1	6 %
MPL	2	11 %	*LAMA2*	1	6 %
NOTCH1	2	11 %	*LINC00273*	1	6 %
PDGFRA	2	11 %	*NELL2*	1	6 %
ROS1	2	11 %	*NRAS*	1	6 %
TAS2R38	2	11 %	*PSMB1*	1	6 %
ABCB1	1	6 %	*PSMD1*	1	6 %
ABCC2	1	6 %	*PSMD2*	1	6 %
ALK	1	6 %	*PTENP1*	1	6 %
ATM	1	6 %	*RARA*	1	6 %
ATRX	1	6 %	*RUNX1*	1	6 %
BRCA1	1	6 %	*RXRA*	1	6 %
BRCA2	1	6 %	*RXRG*	1	6 %
CDH1	1	6 %	*SLC22A1*	1	6 %
CDKN2B	1	6 %	*SLCO1B3*	1	6 %
CREBBP	1	6 %	*SMARCA4*	1	6 %
CYP19A1	1	6 %	*SMARCB1*	1	6 %
CYP2A13	1	6 %	*TET2*	1	6 %
CYP2A7	1	6 %	*TRRAP*	1	6 %
CYP2D6	1	6 %	*TYK2*	1	6 %

Table 16: The list of genes found to be mutated at a frequency of 5 % or more in the DNA of primary tumor, with the number of patients with mutations in that particular gene and the frequency in our cohort of patients with CRC.

Gene	Effect	protein HGVS	HGVS	Count (Patients, N=18)
DDX3X	intron variant	-	c.1025+116A>G	4
KRAS	missense variant	p.Gly12Val	c.35G>T	4
YES1	missense variant	p.Glu156Gly	c.467A>G	4
PIK3CA	missense variant	p.Glu545Lys	c.1633G>A	3
IL2RB	missense variant	p.Asp391Glu	c.1173C>A	2
TP53	missense variant	p.Arg282Trp	c.844C>T	2

Table 17: Gene alterations that were found more than once in the whole cohort of patients with CRC. HGVS: Human Genome Variation Society nomenclature

For each patient, the pathogenicity of the variants found was evaluated using the ClinVar and OncoKB™ databases. Some gene alterations were found to be likely oncogenic in the OncoKB™ database when they were not found in ClinVar at all. This may be due to the fact that ClinVar evaluates gene variants in germline context, while OncoKB™ currently only contains information in somatic context. Those marked as pathogenic or likely pathogenic are listed in the Table 18.

Gene	Effect	protein HGVS	HGVS	Count (Patients, N=18)	ClinVar	OncoKB™
APC	stop gained	p.Ser688*	c.2063C>G	1		
	frameshift variant	p.Arg1435fs	c.4305delA	1		
	stop gained	p.Glu1379*	c.4135G>T	1		
	stop gained	p.Ser1356*	c.4067C>G	1		
	stop gained	p.Glu1521*	c.4561G>T	1		
	frameshift variant	p.Ile1307fs	c.3921_3925delAAAAG	1		
	stop gained	p.Gln1242*	c.3724C>T	1		
	stop gained	p.Arg805*	c.2413C>T	1		
	stop gained	p.Ser1327*	c.3980C>G	1		
	stop gained	p.Tyr935*	c.2805C>A	1		
	frameshift variant	p.Thr1218_Ser1219fs	c.3652_3653insA	1	×	
	stop gained	p.Cys995*	c.2985C>A	1	×	

Table 18 continued

Gene	Effect	protein HGVS	HGVS	Count (Patients, N=18)	ClinVar	OncoKB™
APC	frameshift variant	p.Glu1573fs	c.4717_4727delGAA ATACTAGA	1	×	
	frameshift variant	p.Ile1579_Ile1580fs	c.4735_4736insC	1	×	
	frameshift variant	p.Ala1492fs	c.4475delC	1	×	
	frameshift variant	p.Thr1301fs	c.3901delA	1	×	
ATM	missense variant	p.Trp57Ser	c.170G>C	1		×
FBXW7	stop gained	p.Arg106*	c.316C>T;c.478C>T	2		
	stop gained	p.Ser164*	c.491C>A	1	×	
KRAS	missense variant	p.Gly12Val	c.35G>T	5		
	missense variant	p.Ala146Thr	c.436G>A	1		
	missense variant	p.Gly13Asp	c.38G>A	1		
NRAS	missense variant	p.Gln61Lys	c.181C>A	1		
PIK3CA	missense variant	p.Glu545Lys	c.1633G>A	3		
	missense variant	p.His1047Leu	c.3140A>T	1		
SMAD4	missense variant	p.Arg361His	c.1082G>A	1		
	missense variant	p.Arg496Gly	c.1486C>G	1	×	
	frameshift variant	p.Lys50_Lys51fs/c.148_149insA	c.148_149insA	1	×	
	missense variant	p.Glu526Lys	c.1576G>A	1	×	
SMARCA4	frameshift variant	p.Thr353fs	c.1058delC	1	×	
TP53	missense variant	p.Arg175His	c.524G>A	1		
	missense variant	p.His179Arg	c.536A>G	1		
	stop gained	p.Arg213*	c.637C>T	1		
	missense variant	p.Ala138Pro	c.412G>C	1		

Table 18 continued						
Gene	**Effect**	**protein HGVS**	**HGVS**	**Count (Patients, N=18)**	**ClinVar**	**OncoKB™**
TP53	missense variant	p.Arg282Trp	c.844C>T	2		
	missense variant	p.Arg273Cys	c.817C>T	1		
	missense variant	p.Tyr236Cys	c.707A>G	1		
	missense variant	p.Arg248Gln	c.743G>A	1		
	missense variant	p.Glu285Lys	c.853G>A	1		
	missense variant	p.Tyr163His	c.487T>C	1	×	
	missense variant	p.Gly262Val	c.785G>T	1	×	
	missense variant	p.Val217Gly	c.650T>G	1	×	

Table 18: Pathogenic or likely pathogenic variants found in primary tumor DNA. The last two columns contain information on the number of patients with this particular gene variant and whether the mutation was marked as pathogenic/likely pathogenic in the ClinVar or OncoKBTM databases - × meaning that it was not considered pathogenic in that particular database. HGVS: Human Genome Variation Society nomenclature

In one patient, we were not able to identify any pathogenic mutations in the primary tumor DNA at all (patient with ID 17). The sequencing data of primary tumor DNA were not of a good quality, this might be due to a lower quality and concentration of DNA libraries.

5.3 Comparison of the DNA mutations found in primary tumors and metastases

For four patients only, samples of metastasis were available for sequencing. The fundamental information about these four patients (ID 2, 8, 12 and 14) is presented in the Table 19.

ID	sex	age	diagnosis	histol.	clinical stage at diagnosis	grade	metastases	treatment	time from diagnosis till last check-up (in months)	last known condition
2	F	67	rectum	AC	1	G2	locoregional relapse	3 lines	37	died
8	M	52	colon (sigmoid)	AC	4	G2	liver + sigma	2 lines	30	died
12	F	44	rectum	AC	4	G2	liver (2) + lungs	5 lines	44	alive
14	F	56	colon	AC	4	G2-3	liver (NC) + gynecological	3 lines	20	died

Table 19: Fundamental information about patients with available metastases samples for sequencing. AC = adenocarcinoma; NC = not collected, histol.: histological nature

Visual representation of the comparison of mutations found in the DNA of the primary tumor and metastases of the 4 patients for whom FFPE samples of metastases were available is on Fig. 8. The minimum of common mutations for primary tumor and metastasis was 1, the maximum 5 and median was 2. The minimum of unique metastatic variants was 0, maximum 19, with median of 2.

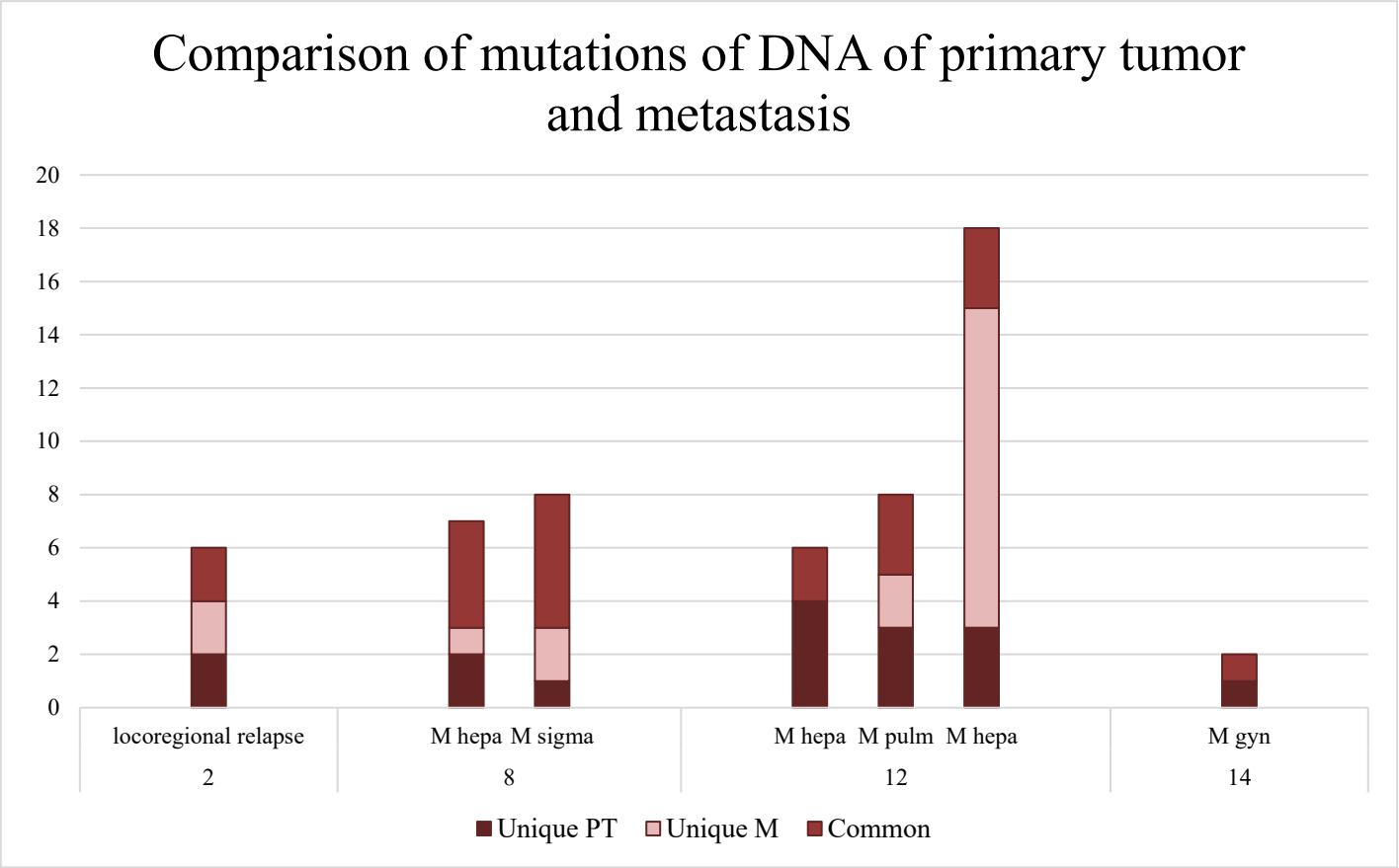

Fig. 8: Visual representation of comparation of mutations found in DNA of primary tumor and metastases of 4 patients ID 2, 8, 12 and 14 from whom FFPE samples of metastases were available. M hepa = DNA of liver metastasis; M sigma = NDA of sigmoid colon metastasis; M pulm = DNA of lung metastasis; (here of unspecified origin); M gyn = DNA of metastasis of gynecological origin; Unique PT = unique mutations for primary tumor DNA; unique M = unique mutations for metastatic DNA

5.3.1 Patient ID 2

Four genes were found to be mutated in the DNA of primary tumor of patient ID 2, with two identified as pathogenic – *APC* stop gained (p.Glu1379*/c.4135G>T), *TP53* missense variant (p.Val217Gly/c.650T>G). Only two of these four mutation were present in the DNA of locoregional relapse. Two mutations were unique for locoregional relapse – *RUNX1* intron variant (c.58+10245G>C) and *APC* stop gained (p.Ser713*/c.2138C>G), the *APC* variant was evaluated as pathogenic by ClinVar and OncoKB^{TM} databases. Table of genes mutated with specific mutations and frequencies in the primary tumor DNA and the metastatic DNA is below – Table 20.

Gene	Mutation (protein HGVS/ HGVS)	F (PT)	F (M)	ClinVar + OncoKB™
APC	Stop gained (p.Glu1379*/c.4135G>T)	28 %	50 %	Pathogenic
TP53	Missense variant (p.Val217Gly/c.650T>G)	23 %	65 %	Pathogenic
DDX3X	Intron variant (c.1025+116A>G)	17%	0 %	Not found
RXRA	Upstream gene variant (NM_001291921.1)	5 %	0 %	Not found
RUNX1	Intron variant (c.58+10245G>C)	0 %	8 %	Not found
APC	Stop gained (p.Ser713*/c.2138C>G)	0 %	18 %	Pathogenic

Table 20: Genes mutated in primary tumor and metastasis of patient ID 2 with frequencies and ClinVar + OncoKB™ evaluation. F(PT) = frequency in the DNA of primary tumor; F(M)= frequency in the DNA of metastasis, HGVS: Human Genome Variation Society nomenclature

5.3.2 Patient ID 8

Patient ID 8 had two metastases in liver and sigmoid colon. Six genes were found to be mutated in the primary tumor DNA. Five mutations were evaluated as pathogenic or likely pathogenic by ClinVar and OncoKB™ databases – *APC* frame shift variant (p.Ile1307fs/c.3921_3925delAAAAG), *TP53* missense variant (p.Gly262Val/c.785G>T), *SMAD4* missense variant (p.Glu526Lys/c.1576G>A), *KRAS* missense variant (p.Gly12Val/c.35G>T) and *PIK3CA* missense variant (p.Glu545Lys/c.1633G>A). In liver metastasis four of the six genes found to be mutated in the DNA of primary tumor were also mutated, in sigmoid colon metastasis five of the six mutated genes in the DNA of primary tumor were also found to be mutated. One mutation was found to be specific for liver metastasis - *RARB* missense variant (p.Arg359Thr/c.1076G>C) and two to be specific for metastasis of sigmoid colon – *CYP2D7* missense variant (p.Asp169Glu/c.507T>G) and *LTK* missense variant (p.Gly213Ala/c.638G>C). Table 21 summarizes genes found to be mutated in the DNA of primary tumor and both metastases.

Gene	Mutation (protein HGVS/ HGVS)	F (PT)	F (M hepa)	F (M sigma)	ClinVar + OncoKB™
APC	Frame shift variant (p.Ile1307fs/c.3921_3925delAAAAG)	42 %	27 %	77 %	Pathogenic
TP53	Missense variant (p.Gly262Val/c.785G>T)	42 %	28 %	63 %	Pathogenic
SMAD4	Missense variant (p.Glu526Lys/c.1576G>A)	39 %	30 %	55 %	Pathogenic
KRAS	Missense variant (p.Gly12Val/c.35G>T)	29 %	36 %	41 %	Pathogenic
PIK3CA	Missense variant (p.Glu545Lys/c.1633G>A)	11 %	0 %	23 %	Pathogenic
CYP2A13	Missense variant (p.Asp169Glu/c.507T>G)	5 %	0 %	0 %	Not found
CYP2D7	Non coding exon variant (n.796G>A)	0 %	0 %	24 %	Not found
RARB	Missense variant (p.Arg359Thr/c.1076G>C)	0 %	5 %	0 %	Not found
LTK	Missense variant (p.Gly213Ala/c.638G>C)	0 %	0 %	7 %	Not found

Table 21: Genes mutated in primary tumor and metastases of patient ID 8 with frequencies and ClinVar + OncoKB™ evaluation. F(PT) = frequency in the DNA of primary tumor; F(M hepa)= frequency in the DNA of metastasis of liver; F(M sigma) = frequency in the DNA of metastasis of sigmoid colon, HGVS: Human Genome Variation Society nomenclature

5.3.3 Patient ID 12

Six genes were found to be mutated in the DNA of the primary tumor of patient ID 12. *APC* stop gain variant (p.Gln1242*/c.3724C>T) and *TP53* missense variant (p.Tyr236Cys/c.707A>G) were evaluated by ClinVar and OncoKB™ databases as pathogenic. In metastatic DNA of liver 2 mutations from primary tumor DNA were found, in metastatic DNA of lungs and DNA of another liver metastasis, three mutations from primary tumor DNA were found. In all metastases pathogenic variants from primary tumor were found. The second, later, metastasis had 12 new mutations that were not reported in the primary tumor DNA before. Table 22 summarizes genes found to be mutated in the DNA of primary tumor and all metastases.

Gene	Mutation (protein HGVS/ HGVS)	F (PT)	F (M hepa)	F (M pulm)	F(M hepa)	ClinVar + OncoKB™
TP53	Missense variant (p.Tyr236Cys/c.707A>G)	14 %	83 %	19 %	78 %	Pathogenic
FLT1	Missense variant (p.Ala660Pro/c.1978G>C)	14 %	0 %	0 %	0 %	Not found
JAG2	Missense variant (p.Glu501Lys/c.1501G>A)	13 %	0 %	11 %	11 %	Not found
APC	Stop gained (p.Gln1242*/c.3724C>T)	13 %	76 %	15 %	65 %	Pathogenic
RXRA	Upstream gene variant (NM_001291921.1)	0 %	0 %	10 %	0 %	Not found
SMAD4	Missense variant (p.Glu134Lys/c.400G>A)	0 %	0 %	5 %	0 %	Not found
PDGFRA	Missense variant (p.Arg500Gln/c.1499G>A)	6 %	0 %	0 %	0 %	Uncertain significance
KDR	Stop gained (p.Arg1032*/c.3094C>T)	6 %	0 %	0 %	0 %	Not found
MAP3K1	Missense variant (p.Asp806Asn/c.2416G>A)	0 %	0 %	0 %	18 %	Benign
MAP3K1	Missense variant (p.Val906Ile/c.2716G>A)	0 %	0 %	0 %	15 %	Benign
TET2	Missense variant (p.His1778Arg/c.5333A>G)	0 %	0 %	0 %	8 %	Benign
ALK	Missense variant (p.Asp1529Glu/c.4587C>G)	0 %	0 %	0 %	8 %	Benign
SMARCB1	Intron variant (c.473+305C>A)	0 %	0 %	0 %	7 %	Benign
TET2	Missense variant (p.Val218Met/c.652G>A)	0 %	0 %	0 %	6 %	Benign
ALK	Missense variant (p.Lys1491Arg/c.4472A>G)	0 %	0 %	0 %	6 %	Benign

Table 22 continued						
Gene	**Mutation (protein HGVS/ HGVS)**	**F (PT)**	**F (M hepa)**	**F (M pulm)**	**F(M hepa)**	**ClinVar + OncoKB™**
PDGFRB	Missense variant (p.Thr140Met/c.419C>T)	0 %	0 %	0 %	6 %	Likely benign
TET2	Missense variant (p.Leu34Phe/c.100C>T)	0 %	0 %	0 %	6 %	Benign
PTENP1	Non coding exon variant (n.1325G>C)	0 %	0 %	0 %	5 %	Not found
CYP2C19	Missense variant (p.Glu92Asp/c.276G>C)	0 %	0 %	0 %	5 %	Not found

Table 22: Genes mutated in primary tumor and metastases of patient ID 12 with frequencies and ClinVar + OncoKB™ evaluation. F(PT) = frequency in DNA of primary tumor; F(M hepa)= frequency in DNA of metastasis of liver; F(M pulm) = frequency in DNA of metastasis of lungs, HGVS: Human Genome Variation Society nomenclature

5.3.4 Patient ID 14

Only two genes were found to be mutated in the DNA of the primary tumor of patient ID 14 (see table Table 23), of which only *KRAS* missense variant (p.Gly12Val/c.35G>T) was evaluated as pathogenic by the ClinVar and OncoKB™ databases. This patient had gynecological metastasis in which the *KRAS* missense variant found in the DNA of primary tumor was also found.

Gene	Mutation (protein HGVS/ HGVS)	F (PT)	F (M gyn)	ClinVar + OncoKB™
KRAS	Missense variant (p.Gly12Val/c.35G>T)	9 %	31 %	Pathogenic
YES1	Missense variant (p.Glu156Gly/c.467A>G)	7 %	0 %	Not found

Table 23: Genes mutated in primary tumor and metastases DNA of patient ID 14 with frequencies and ClinVar + OncoKB™ evaluation. F(PT) = frequency in the DNA of primary tumor; F(M gyn)= frequency in the DNA of metastasis of gynecological origin, HGVS: Human Genome Variation Society nomenclature

5.4 Comparison of the DNA mutations found in ctDNA and primary tumor at the time of diagnosis

To assess the quality of ctDNA as a diagnostic tool, sequencing data from primary tumor DNA and ctDNA at the time of diagnosis were compared. The same filtering rules were applied to ctDNA as to primary tumor DNA, including that the variant must be present at a frequency of 5 % or greater. Two patients (ID 12 and 17) did not have ctDNA available

at the time of diagnosis. Only one patient had all the same pathogenic variants in the ctDNA at the time of diagnosis as in the DNA of the primary tumor. In total, we detected pathogenic variants from primary tumor DNA in the ctDNA of five patients. On average, the percentage of pathogenic variants detected in ctDNA that were similar to those found in primary tumor DNA was 15 %. The average coverage of variants in ctDNA found in primary tumor DNA was 225. In terms of all variants, an average of 14 % of all variants found in primary tumor DNA were also found in ctDNA. Table 24 summarizes how many variants and pathogenic variants were found in primary tumor DNA, how many of them were found in ctDNA, the time period between tissue biopsy and ctDNA collection, and the coverage of primary tumor variants in ctDNA samples for each patient. In addition, there is whether new variants were found in the ctDNA that were not previously found in the primary tumor DNA.

ID	Count of PT variants	Count of pathogenic PT variants	Similar variants in ctDNA (%)	Similar pathogenic variants in ctDNA (%)	New variants found in ctDNA (pathogenic)	Timespan between PT biopsy and ctDNA collection (days)	Average of coverage of variants in ctDNA
1	10	3	8 (80)	3 (100)	0	27	727
2	4	2	1 (25)	1 (50)	38 (0)	919	223
3	22	8	5 (23)	2 (25)	0	349	518
4	4	2	1 (25)	0 (0)	0	899	58
5	6	2	0 (0)	0 (0)	0	23	212
6	3	1	0 (0)	0 (0)	0	41	4
7	7	3	0 (0)	0 (0)	0	545	0
8	6	5	0 (0)	0 (0)	1 (0)	-	-
9	2	1	0 (0)	0 (0)	1 (0)	153	0
10	16	3	0 (0)	0 (0)	0	545	50
11	9	6	3 (33)	2 (33)	1 (0)	504	364
12	6	2	-	-		-	-
13	3	3	1 (33)	1 (33)	1 (0)	70	818
14	2	1	0 (0)	0 (0)	0	94	376
15	5	2	0 (0)	0 (0)	0	54	0
16	5	3	0 (0)	0 (0)	0	493	254
17	24	0	-	-		89	0
18	7	5	0 (0)	0 (0)	0	544	0

Table 24: Summarization of number of variants and pathogenic variants in primary tumor DNA and number of variants and pathogenic variants that were simultaneously found in primary tumor DNA and ctDNA. Last two columns contain the timespan between primary tumor biopsy and ctDNA collection and approximal coverage of variants found in DNA of primary tumor in ctDNA at time of diagnosis.

The total of variants found in the DNA of primary tumor from all patients was 141, compared to 55 found in ctDNA at the time of diagnosis. From those variants 92 were unique to primary tumor, 36 to ctDNA and 19 were common (see Fig. 9).

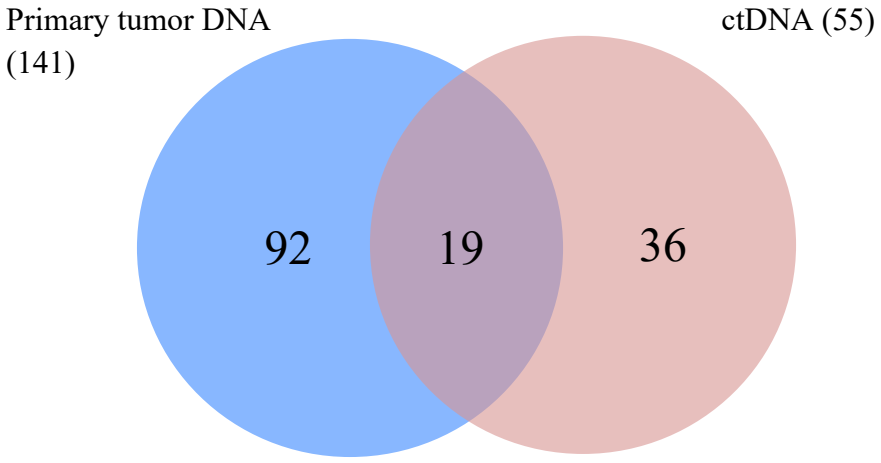

Fig. 9: The distribution of variants unique to primary tumor (blue, left), unique to ctDNA (red, right), and common to both (purple, middle) from whole cohort of patients with CRC.

5.5 Tumor-specific markers in ctDNA

To determine whether ctDNA could be used to monitor cancer progression, we attempted to find tumor-specific variants in ctDNA samples from different time points during treatment and disease progression. In only five patients we were able to detect tumor-specific variants in ctDNA samples from diagnosis to the last follow-up. In ctDNA samples from the remaining patients, we were unable to detect variants from primary tumor DNA or metastases due to low detection sensitivity or in some cases due to low quality and quantity of ctDNA sequencing libraries. We kept quite stringent criteria to avoid equivocal findings and false positivity at the cost of sensitivity. Lower threshold for filtering in the variants in ctDNA (e.g. 1 % instead of 5 %) may improve the sensitivity of cfDNA-based analyses. Fundamental information about patients with possible tumor-specific markers for cancer progression is shown in Table 25.

ID	sex	age	diagnosis	histol.	clinical stage in time of diagnosis	grade	metastases	treatment (time month/year)	time from diagnosis till last check-up (in months)	last known condition
1	M	71	colon (sigmoid)	mAC	4	G2	liver + lungs (NC)	2 lines (11/20 – 2/21; 3/21)	8	died
2	F	67	rectum	AC	1	G2	locoregional relapse	3 lines (10-11/19; 4-7/20; 7-9/20)	37	died
3	M	76	Colon	AC	4	G2-3	liver (NC)	3 lines (1-2/20; 2-4/20; 5/20-2/21)	15	died
10	M	65	Colon	AC	1	G2	liver (NC)	3 lines (3-5/20; 6/20-7/21; 7-11/21)	33	died
11	F	51	Colon	AC	4	G2	liver (NC)	3 lines (8/19-1/20; 1-4/20; 6/20-4/21)	25	n
12	F	44	rectum	AC	4	G2	liver (2) + lungs	5 lines (12/19-1/20; 5-9/20; 12/20-5/21; 8-9/22; 10/22-3/23)	44	alive

Tab. 25: Fundamental information about patients with tumor-specific variants found in ctDNA samples from different time points of disease progression and treatment. AC: adenocarcinoma; grade G1-3 refers to the degree of tumor dedifferentiation, reflecting its aggressiveness; NC: not collected; histol.: histological nature; n: not know

Note: The patients were treated by combined chemotherapy, sometimes with targeted therapy added. The chemotherapy of the same composition (= the line) was applied for 3 – 6 months, sometimes more, and the response was evaluated every 3 – 4 months. Once the resistance to therapy developed and progression occurred, this chemotherapy was stopped and chemotherapy of different composition (next line) was started. Patients thus may undergo chemotherapy of several lines (1^{st} to 5^{th} line)

5.5.1 Patient ID 1

Patient ID 1 had seven gene variants (Tab. 26) that we were able to detect in ctDNA from different time points during disease progression and treatment, which can serve as potential tumor-specific markers for monitoring disease progression. Three of them were found to be pathogenic by ClinVar and OncoKB^{TM} databases. Patient ID 1 underwent first line treatment from November 2020 to February 2021, which correlates with the lower frequency of possible tumor-specific markers in ctDNA from December 8, 2020. The frequency of possible tumor-specific markers increased in December 2020 and remained quite high until January 2021, indicating further disease progression, which supports the fact that this patient died in April 2021.

Gene	F PT 1/9/20	variant		F ctDNA diagnosis 27.10.20	F ctDNA 24.11.20	F ctDNA 8.12.20	F ctDNA 29.12.20	F ctDNA 26.01.21
SMAD4	65%	missense variant (p.Trp509Gly/c.1525T>G)		14%	11%	9%	38%	19%
FLT1	45%	missense variant (p.Arg1112Cys/c.3334C>T)		18%	11%	11%	24%	17%
ROS1	38%	stop gained (p.Gln930*/c.2788C>T)		15%	11%	15%	33%	17%
KRAS	37%	missense variant (p.Gly12Val/c.35G>T)	pathogenic	13%	8%	6%	32%	12%
APC	37%	stop gained (p.Ser688*/c.2063C>G)	pathogenic	19%	12%	13%	30%	17%
IKZF1	35%	missense variant (p.Arg473His/c.1418G>A)		11%	7%	6%	16%	10%
APC	33%	frameshift variant (p.Arg1435fs/c.4305delA)	pathogenic	17%	9%	7%	24%	14%

Table 26: Gene variants of primary tumor DNA detected in ctDNA of patient ID 1, with detected frequencies.

F: frequency; PT: primary tumor

5.5.2 Patient ID 2

Patient ID 2 first received therapy after resection of locoregional relapse from October to November 2019, that's probably why there is a decrease in the frequency of which we detect potential tumor-specific markers in liquid biopsy (see Table 27). In April 2020, the two potential tumor-specific markers were not found at all in plasma, which could be due to lower plasma ctDNA concentration after successful second line therapy that the patient underwent from April to July 2020. In July 2020, there was a slight increase in the frequency with which the two gene variants were found in plasma, probably due to the deterioration of the patient's condition. Another, third, line of therapy was prescribed in July 2020 and continued until September 2020. The patient died in October 2020.

Gene	PT 2017	variant		loc. rel. 8/19	ctDNA diagnosis 3.4.20	ctDNA 17.4.20	ctDNA 30.4.20	ctDNA 10.6.20	ctDNA 20.7.20
TP53	23%	stop gained (p.Glu1379*/c.4135G>T)	pathogenic	65%	4%	NA	NA	NA	10%
APC	28%	missense variant (p.Val217Gly/c.650T>G)	pathogenic	50%	6%	NA	NA	NA	6%

Table 27: Frequencies of which the two potential tumor-specific markers, evaluated by ClinVar and OncoKB^{TM} as pathogenic, were found in primary tumor DNA and ctDNA from different time periods of disease course. PT: primary tumor; loc. rel.: locoregional relapse

5.5.3 Patient ID 3

We found four possible tumor-specific markers in the DNA of the primary tumor of patient ID 3 (see Table 28), which were also detected in the ctDNA. Patient ID 3 received three lines of therapy, the first from January to February 2020, the second from February to April 2020, and the third from May 2020 to February 2021. There were no plasma samples available from the time of the first and second lines of treatment. We can see an increase in the frequency with which we detect the four possible tumor-specific markers from the first available ctDNA to the last. This could be due to a deterioration of the patient's condition and ineffectiveness of the last line of treatment, this patient died in February 2021.

Gene	PT 11/19	variant		ctDNA diagnosis 9.11.20	ctDNA 18.1.21	ctDNA 1.2.21
KRAS	42%	missense variant (p.Ala146Thr/c.436G>A)	pathogenic	7%	13%	18%
NOTCH1	33%	missense variant (p.Val2110Met/c.6328G>A)		5%	10%	19%
ERBB3	27%	missense variant (p.Gly325Arg/c.973G>A)		7%	10%	20%
SMAD4	21%	missense variant (p.Arg496Gly/c.1486C>G)	pathogenic	5%	6%	12%

Table 28: Gene variants that can serve as possible tumor-specific markers with frequencies in primary tumor DNA (PT) and ctDNA from different time periods. Two variants were evaluated as pathogenic by ClinVar and OncoKB^{TM} databases.

5.5.4 Patient ID 10

Patient ID 10 underwent the first line of therapy before the first plasma collection, from March to May 2020, the second line of therapy started in June 2020 and lasted until July 2021, in this period the first plasma sample was collected, that is probably why we do not detect any gene variants in ctDNA from October 2020 (see table Tab. 29), due to low sensitivity (low concentration of ctDNA in plasma). The frequency of the five potential tumor-specific markers increased in January 2021, which means that the therapy was probably ineffective. In April, the frequency remained fairly high, but lower than in January. In June, we did not detect any gene variants, but the patient received the last line of therapy from July to October 2021, which means that his condition must have worsened. The patient died in December 2021.

Gene	PT 5/19	variant	ctDNA diagnosis 27.10.20	ctDNA 7.1.21	ctDNA 19.4.21	ctDNA 24.6.21
BRCA1	14%	missense variant (p.Glu1038Gly/c.3113A>G)	NA	14%	9%	NA
ABCC2	8%	missense variant (p.Val1188Glu/c.3563T>A)	NA	10%	9%	NA
IL2RB	8%	missense variant (p.Asp391Glu/c.1173C>A)	NA	12%	10%	NA
FLT4	6%	missense variant (p.Thr494Ala/c.1480A>G)	NA	15%	12%	NA
INPP4B	6%	missense variant (p.Gly554Ser/c.1660G>A)	NA	7%	7%	NA

Table 29: Gene variants that can serve as possible tumor-specific markers for monitoring the cancer development, with frequencies in primary tumor DNA and ctDNA from different time periods. PT= primary tumor

5.5.5 Patient ID 11

We detected only one possible tumor-specific marker, a *TP53* missense variant (see Table 30), in the ctDNA of patient ID 11. The first two lines of therapy this patient underwent before the first plasma collection, from August 2019 to January 2020 and from January to April 2020. The third line started in June 2020 and lasted until April 2021. During this time, the first plasma sample was collected, in November 2020, which showed a slight decrease in the frequency of the possible tumor-specific marker compared to the frequency in DNA of the primary tumor. The next two plasma samples, from February and April 2021, show an increase in the frequency of the *TP53* gene variant. Since we do not know anything about the last condition of this patient, we can only speculate that their condition has worsened.

Gene	PT 7/19	variant		ctDNA diagnosis 16.11.20	ctDNA 22.2.21	ctDNA 16.4.21
TP53	16%	missense variant (p.Arg273Cys/c.817C>T)	pathogenic	12%	32%	40%

Table 30: Possible tumor-specific marker that was found in ctDNA of patient ID 11, with frequencies in ctDNA and DNA of primary tumor. PT= primary tumor

5.5.6 Patient ID 12

For patient ID 12, there are two possible tumor-specific markers (see table Tab. 30) that can be found in both primary tumor and metastases, as well as in ctDNA. After tumor biopsy, the patient had the first line of therapy (December 2019 - January 2020). During the treatment she developed liver metastasis, the treatment was stopped and she underwent resection of the primary tumor and liver metastasis. In the resected tumor higher DNA frequencies of the two possible tumor-specific markers were found comparing with the primary biopsy. This suggests a worsening of the disease, which supports the appearance of metastasis in the liver, in which DNA these two tumor-specific markers were found with high frequency. ctDNA from April 2020 also shows high frequencies of these two possible tumor-specific markers. The patient underwent further treatment from May to September 2020. ctDNA from this period did not show the presence of any tumor-specific markers, but new metastasis appeared in the lungs, suggesting further deterioration of the disease. The tumor-specific markers were probably not found in the ctDNA due to the lower concentration in plasma. The patient underwent third line of treatment from December 2020 to May 2021, ctDNA from this period did not show any markers, either due to low detection sensitivity or slight improvement of the patient's condition. Unfortunately, another metastasis appeared in the liver in July 2021. The two tumor-specific markers were detected in the DNA of this metastasis as well as in the ctDNA from June 2021, indicating

a further deterioration of the disease. The patient underwent two additional treatments, from August to September 2022 and from October 2022 to March 2023. ctDNA from this period did not show either of the possible tumor-specific markers. This could be due to a lower ctDNA concentration that was not detectable or improvement of the disease. This patient is still alive, in partial remission.

Gene	PT 11/19	variant	neo 4/20	M hepa 4/20	M pulm 10/20	M hepa 7/21	ctDNA 9.4.20	ctDNA 18.5.20 ctDNA 14.10.20 ctDNA 10.11.20 ctDNA 22.2.21	ctDNA 16.6.21	ctDNA 5.10.21 ctDNA 30.5.22
TP53	14%	missense variant (p.Tyr236Cys /c.707A>G)	35%	83%	19%	78%	38%	NA	19%	NA
APC	13%	stop gained (p.Gln1242* /c.3724C>T)	15%	76%	15%	65%	33%	NA	16%	NA

Table 31: The two possible tumor-specific markers for patient ID 12, with frequencies in primary tumor, metastases and ctDNA. PT= primary tumor; M hepa= metastasis of liver; M pulm= metastasis of lungs

6. Discussion

6.1 DNA mutations in primary tumors

Molecular testing of patients with CRC may aid in diagnosis, prediction of treatment response and disease progression (Coppedè et al. 2014). Minimal molecular testing is part of standard patient care - testing for *RAS* and *BRAF* mutations, as well as MSI status, can help guide treatment selection, but more and more new gene alterations are being discovered as responsible for treatment resistance. Many studies have highlighted the heterogeneity of CRC and the uniqueness of the molecular background of individuals with CRC (Wood et al. 2007; Cornish et al. 2022) . Here, it was shown that in 18 patients with metastatic CRC, 141 gene variants of 68 genes were found in the DNA of primary tumors, of which 122 were detected only once in the entire cohort. These variants were evaluated for pathogenicity using ClinVar and OncoKB^{TM} databases. Forty-three variants of nine genes - *APC, ATM, FBXW7, KRAS, NRAS, PIK3CA, SMAD4, SMARCA4*, and *TP53* - were evaluated as pathogenic. These genes were previously reported to be mutated in CRC.

6.1.1 APC

As mentioned in the literature review *APC* mutations are responsible for Familial Adenomatous Polyposis (Half, Bercovich, a Rozen 2009) as well as for some sporadic cases

of CRC (Filippo et al. 2002). *APC* is a tumor suppressor gene located on the long arm of chromosome 5 (5q22.2) (Béroud a Soussi 1996). In the cytoplasm, APC protein negatively regulates the WNT signaling pathway, by inducing the degradation of β-catenin by forming a complex with Glycogen Synthase Kinase 3β (GSK-3β), AXIN1 or AXIN2 and other kinases and phosphatases (Munemitsu et al. 1995). Phosphorylated β-catenin cannot serve as a licensing factor for TCF/LEF family transcription factors (Behrens et al. 1996). Without APC-induced degradation, β-catenin and TCF/LEF transcription factors activate transcription of genes such as *c-MYC* (He et al. 1998) *and Cyclin D1* (Shtutman et al. 1999), resulting in activation of proliferation, survival and maintenance of an undifferentiated state in progenitor cells of the colorectal epithelium (Fevr et al. 2007). Typically, somatic and germline mutations of *APC* result in a shorter protein, and the mutations tend to occur in the region called the mutation cluster region, MCR, between codons 1250 and 1464 (Kohler et al. 2008).

6.1.2 ATM

The *ATM* tumor suppressor gene is located on the long arm of chromosome 11 (11q22-23) and has been reported to be mutated in several cancers, including gastric cancer (Kim et al. 2013) and CRC (Sundar et al. 2018). The *Ataxia Telangiectasia Mutated* gene (*ATM*) encodes a Ser/Thr kinase of the PI3K (PhosphoInositide-3 Kinase) - related protein kinase family (Abraham 2004), which is activated by DNA double-strand breaks and signals the cell cycle to slow down to facilitate DNA repair via the Mre11/Rad50/Nbs1 (MRN) complex (Lee a Paull 2005). Activation of ATM by binding to the MRN complex leads to phosphorylation of a number of downstream targets involved in DNA damage repair, apoptosis, cell cycle arrest and cell cycle checkpoints (Phan a Rezaeian 2021). Mutations in *ATM* result in a disease called Ataxia Telangiectasia, which is characterized by, among others, a predisposition to various cancers (Rothblum-Oviatt et al. 2016). *ATM*-mutated tumors have deficient homologous recombination systems and may response well to the targeted therapy with PARP (Poly-ADP ribose polymerase)-inhibitors (C. Wang et al. 2017).

6.1.3 FBXW7

The *FBXW7* gene is located on chromosome 4 (4q31.3), and mutations in this gene have been identified as one of the causes of cancer development (Mao et al. 2004). FBXW7 is an F-box protein that is a part of the SCF (Skip1-Cull-F-box) ubiquitin ligase complex, which plays an important role in the degradation of cell cycle progression proteins such as

Cyclin E1 and c-MYC (Minella a Clurman 2005). Previously, it was reported that expression of FBXW7 gene was downregulated in patients with CRC (Iwatsuki et al. 2010).

6.1.4 *KRAS* and *NRAS*

KRAS (chromosome 12p12.1) and *NRAS* (chromosome 1p13.2) are isoforms of the RAS oncogene family that play a role in cell growth - influencing cell proliferation, apoptosis, migration, and differentiation, mainly through MAPK (Mitogen-Activated Protein Kinase) and PI3K (PhosphoInositide-3 Kinase) pathways (Qi a Elion 2005; Hemmings a Restuccia 2012). *KRAS* accounts for 86 % of all RAS mutations associated with cancer, followed by *NRAS* at 8% (Bamford et al. 2004). *NRAS* is more rare in CRC, mutations in this gene are more typical for melanoma, while *KRAS* is more typical for CRC (Hobbs, Der, a Rossman 2016). *RAS* genes encode membrane-bound proteins with GTPase activity (Zenonos a Kyprianou 2013). Signaling from a nearby transmembrane receptor (growth factor receptor/G-protein coupled receptor/toll-like receptor, etc.) activates the RAS proteins by activating a signaling cascade that results in GTP exchange, RAS with bound GTP is activated and can signal downstream (Zenonos a Kyprianou 2013). Signaling through RAS is terminated when GAP proteins are activated and GTP is exchanged for GDP, rendering RAS inactive. When mutated, RAS proteins are constitutively bound to GTP, resulting in unregulated cell growth signaling (Hobbs, Der, a Rossman 2016). Mutation of *KRAS* G12C may be targeted by a selective molecular inhibitor sotorasib that may be used for treatment of patients harboring this specific mutation (Hong et al. 2020).

6.1.5 PIK3CA

PIK3CA gene is located on chromosome 3 (3q26.32) and encodes for the alpha catalytic subunit of PI3K. PI3K is a lipid kinase that, when activated by receptor tyrosine kinases (such as EGFR), plays a positive role in the mTOR signaling pathway, resulting in the activation of cell growth and protein synthesis (Voutsadakis 2021). *PIK3CA* was found to be mutated in several solid tumors, including CRC (Millis et al. 2016).

6.1.6 SMAD4

SMAD4 gene is part of the TGFβ signaling pathway, regulating proliferation, differentiation, apoptosis, and migration of epithelial cells (Zhao, Mishra, a Deng 2018), and it is located on the chromosome 18 (18q21-22). TGFβ signaling has a dual function in the context of tumorigenesis, at the early stage of tumor formation it induces cell cycle arrest and apoptosis, but at more advanced stage tumor cells become insensitive to this signaling and secreted TGFβ protein starts to play role in tumor immunosuppression, angiogenesis,

invasion and metastasis (Kubiczkova et al. 2012). SMAD4 seems to be responsible for the switch of TGFβ function on tumorigenesis, since its inactivation seems to be linked with more advanced stage tumors (Kubiczkova et al. 2012; Dardare et al. 2020). *SMAD4* mutations are associated with CRC, namely with Juvenile Polyposis Syndrome (James R. Howe et al. 1998; Tarafa et al. 2000)

6.1.7 TP53

TP53 is a tumor suppressor gene located on chromosome 17 (17p13.1) that initiates DNA repair, cell-cycle arrest, senescence and apoptosis (Borrero a El-Deiry 2021). *TP53* is mutated in various types of cancer including CRC (Olivier, Hollstein, a Hainaut 2010). In most tumors combination of missense mutation of one allele and 17p deletion of second allele is present, resulting in elimination of both alleles (Baker et al. 1989). Missense mutations are responsible for production of abnormal p53 protein (Gannon et al. 1990). *TP53*- mutated tumors are highly resistant to chemotherapy (Huang et al. 2019) and no specific targeted therapy is available for them. Thus, these mutations are associated with poor prognosis and shorter survival of the patients with *TP53*-mutated tumors (Bazan et al. 2005).

6.2 Comparison of the DNA mutations found in primary tumors and metastases

Metastastatic disease is the major cause of CRC-related death. We were interested in whether there are significant molecular differences between primary tumors and metastases in patients with CRC. Only four patients had metastatic DNA available for analysis and a total of seven metastatic samples were available from these four patients. The remaining patients enrolled in this project did not have metastatic DNA available for analysis because these patients did not undergo surgical resection, but all patients had metastases during the course of their disease. By comparing the data obtained from sequencing DNA from primary tumors and metastases, we found that three out of four patients had some unique metastatic variants, of which only one, the *APC* gene stop gained (p.Ser713*/c.2138C>G), was evaluated as pathogenic by the ClinVar and OncoKB^{TM} databases. Interestingly, this patient had another APC mutation in the primary tumor. The sequencing could not distinguish if the mutations are on the same or different alleles. Sequential inactivation of both alleles of tumor-suppressor genes occurs typically during the tumor progression and may confer the resistance to the applied therapy. In total, we found 19 (median 2) unique metastatic variants from seven metastatic DNAs. Dang et al. recently observed that the cells that form

metastases are not the most abundant clones from the primary tumor, so we can see some changes in the genetics of metastatic cells. (Dang et al. 2020). In another study, it was observed that the variants found in the metastatic DNA correlated with the variants found in the primary tumor (J. Liu et al. 2020). Both these studies and our observations demonstrate the heterogeneity of CRC and clonal evolution. Our data correlate more with the observations of Liu et al., in three metastatic DNAs (43 %) only two unique mutations were found, in one (14 %) metastatic DNA one unique mutation was found and in two (29 %) metastatic DNAs no new mutations were found. One metastatic sample (14 %) was unique in our cohort, containing twelve new mutations, this might be more in correlation with data of Dang et al.. Our data and both these studies correspond with heterogeneity and clonal evolution of CRC cells.

6.3 Comparison of the DNA mutations found in ctDNA and primary tumor at the time of diagnosis

In the past, circulating cell free DNA has been shown to contain cancer specific alterations (M. Stroun et al. 1989). Wang et al. demonstrated that variants found in ctDNA from serum correspond to those found in primary tumor DNA (J.-Y. Wang et al. 2004). We wanted to test whether ctDNA from plasma could serve as a non-invasive diagnostic tool, by comparing our data obtained from ctDNA with data from primary tumor DNA. Ideally, ctDNA from diagnosis should be compared with the primary tumor. Unfortunately, it has not been available in many of the patients,the average time between tissue biopsy and liquid biopsy was 334 days. This was due to the later time of enrollment of these patients in our project, mostly after their first tissue biopsy. Nonetheless, we were able to detect several pathogenic variants from primary tumor DNA in ctDNA of five patients, in one patient we were able to detect all pathogenic variants identified in primary tumor. Overall, the similarity between ctDNA variants and primary tumor DNA variants was 14% in the whole cohort of patients with CRC. Recently published study compared genomic data obtained from tissue biopsy and liquid biopsy and found a high fidelity between these two types of samples (G. Li et al. 2019). Problem with the use of ctDNA is its detection sensitivity (Parikh et al. 2019), which depends on the amount of ctDNA present in the plasma or other body fluids, since cell-free DNA - cfDNA - from healthy cells is also present (Maurice Stroun et al. 2001). Levels of ctDNA are influenced by tumor size, cancer stage, metastases, and even therapy (as described in section 3.3).

6.4 Tumor specific markers in ctDNA

Here, we sought to demonstrate whether ctDNA can be used to monitor disease progression in patients with CRC by looking for tumor-specific markers in ctDNA. ctDNA has previously been shown to be a valuable tool for monitoring CRC progression when it helped explain secondary resistance and supported rechallenge of treatment (Siravegna et al. 2015). In a recent study, ctDNA was successfully used for postoperative and post-adjuvant chemotherapy risk stratification, treatment response monitoring, and early recurrence detection using tumor-specific multiplex PCR-based NGS in patients with stage I to III CRC (Reinert et al. 2019). We were able to detect primary tumor-specific gene variants in ctDNA from only five patients. In ctDNA samples from the remaining patients, we were unable to detect variants from primary tumor DNA or metastases this might be due to low detection sensitivity or in some cases due to low quality and quantity of ctDNA sequencing libraries. The sensitivity could be also influenced by the pre-set criteria for variant filtration. We kept quite stringent criteria to avoid equivocal findings and false positivity at the cost of sensitivity. Lower threshold for filtering in the variants in ctDNA (e.g. 1 % instead of 5 %) may improve the sensitivity of cfDNA-based analyses.

In all of these patients, there was a correlation between disease progression and a higher frequency of tumor-specific markers in ctDNA. In one patient, sequencing data from metastatic DNA were also available, and there we saw another correlation between the occurrence of metastasis and a higher frequency of tumor-specific markers in ctDNA. The problem with using ctDNA for monitoring disease progression remains, as with using it for diagnosis, detection sensitivity (Maurice Stroun et al. 2001; Parikh et al. 2019).

7. Conclusions

In this study, we sought to demonstrate the molecular heterogeneity and evolution of CRC cells using targeted next-generation sequencing of DNA from primary tumors and metastases. A total of 141 nonsynonymous mutations of 68 genes were found in primary tumor DNA. Forty-three mutations in the genes *APC, ATM, FBXW7, KRAS, NRAS, PIK3CA, SMAD4*, and *TP53* were evaluated as pathogenic using the ClinVar and OncoKB™ databases. No major changes in the molecular background of metastases were found, except for one out of seven metastatic samples, where 12 new mutations unique to metastasis were detected. In the literature, it has been shown that metastasis may or may not be formed by the most abundant clones. This and our findings support the concept of heterogeneity and clonal evolution of CRC cells. The problem in analyzing the molecular background of metastasis was the lack of metastatic samples. Most of the patients enrolled in this study did not undergo metastatic resection.

Another goal was to demonstrate the potential use of ctDNA in the diagnosis and monitoring of disease progression. Overall, the similarity between mutations found in primary tumor DNA and ctDNA at the time of diagnosis was 14 %, and the similarity in pathogenic variants was 15 %. In terms of disease monitoring, the frequencies of possible tumor-specific markers in ctDNA showed a high correlation with disease improvement and worsening, even with the appearance of metastases. Problems with the use of ctDNA is limited detection sensitivity, as its concentration in plasma is influenced by tumor size, cancer stage and metastases, and there is still some cfDNA from healthy cells present in plasma. Change in filtration criteria might increase the detection sensitivity. Nonetheless, ctDNA appears to be a promising source of information on CRC development during the disease course and treatment.

8. References

Aaltonen, L. A. et al. 1993. „Clues to the Pathogenesis of Familial Colorectal Cancer". *Science (New York, N.Y.)* 260(5109): 812–16.

Abraham, Robert T. 2004. „PI 3-Kinase Related Kinases: ‚big′ Players in Stress-Induced Signaling Pathways". *DNA repair* 3(8–9): 883–87.

Acquaviva, Jaime, Ricky Wong, a Al Charest. 2009. „The Multifaceted Roles of the Receptor Tyrosine Kinase ROS in Development and Cancer". *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1795(1): 37–52.

Ahn, Joong Bae et al. 2011. „DNA METHYLATION PREDICTS RECURRENCE FROM RESECTED STAGE 3 PROXIMAL COLON CANCER". *Cancer* 117(9): 1847–54.

André, Thierry et al. 2020. „Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer". *New England Journal of Medicine* 383(23): 2207–18.

Armelao, Franco, a Giovanni de Pretis. 2014. „Familial colorectal cancer: A review". *World Journal of Gastroenterology : WJG* 20(28): 9292–98.

Arreaza, Gladys et al. 2016. „Pre-Analytical Considerations for Successful Next-Generation Sequencing (NGS): Challenges and Opportunities for Formalin-Fixed and Paraffin-Embedded Tumor Tissue (FFPE) Samples". *International Journal of Molecular Sciences* 17(9): 1579.

Baker, Suzanne J. et al. 1989. „Chromosome 17 Deletions and p53 Gene Mutations in Colorectal Carcinomas". *Science* 244(4901): 217–21.

Bamford, S. et al. 2004. „The COSMIC (Catalogue of Somatic Mutations in Cancer) Database and Website". *British Journal of Cancer* 91(2): 355–58.

Baylln, Stephen B. et al. 1997. „Alterations in DNA Methylation: A Fundamental Aspect of Neoplasia". In *Advances in Cancer Research*, ed. George F. Vande Woude a George Klein. Academic Press, 141–96. https://www.sciencedirect.com/science/article/pii/S0065230X08607022 (9. červenec 2023).

Bazan, V. et al. 2005. „Specific TP53 and/or Ki-Ras Mutations as Independent Predictors of Clinical Outcome in Sporadic Colorectal Adenocarcinomas: Results of a 5-Year Gruppo Oncologico Dell'Italia Meridionale (GOIM) Prospective Study". *Annals of Oncology* 16: iv50–55.

Behrens, J. et al. 1996. „Functional Interaction of Beta-Catenin with the Transcription Factor LEF-1". *Nature* 382(6592): 638–42.

Benvenuti, Silvia et al. 2007. „Oncogenic Activation of the RAS/RAF Signaling Pathway Impairs the Response of Metastatic Colorectal Cancers to Anti–Epidermal Growth Factor Receptor Antibody Therapies". *Cancer Research* 67(6): 2643–48.

Béroud, C, a T Soussi. 1996. „APC gene: database of germline and somatic mutations in human tumors and cell lines." *Nucleic Acids Research* 24(1): 121–24.

Bettegowda, Chetan et al. 2014. „Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies". *Science Translational Medicine* 6(224): 224ra24.

Boland, C. R., a F. J. Troncale. 1984. „Familial Colonic Cancer without Antecedent Polyposis". *Annals of Internal Medicine* 100(5): 700–701.

Boland, C. Richard. 2005. „Evolution of the Nomenclature for the Hereditary Colorectal Cancer Syndromes". *Familial Cancer* 4(3): 211–18.

Borrero, Liz J. Hernández, a Wafik S. El-Deiry. 2021. „Tumor Suppressor p53: Biology, Signaling Pathways, and Therapeutic Targeting". *Biochimica et biophysica acta. Reviews on cancer* 1876(1): 188556.

Bronner, C. Eric et al. 1994. „Mutation in the DNA Mismatch Repair Gene Homologue HMLH 1 Is Associated with Hereditary Non-Polyposis Colon Cancer". *Nature* 368(6468): 258–61.

Brosens, Lodewijk A.A., G. Johan A. Offerhaus, a Francis M. Giardiello. 2015. „Hereditary Colorectal Cancer". *Surgical Clinics of North America* 95(5): 1067–80.

van Brummelen, Emilie M. J., Anthonius de Boer, Jos H. Beijnen, a Jan H. M. Schellens. 2017. „BRAF Mutations as Predictive Biomarker for Response to Anti-EGFR Monoclonal Antibodies". *The Oncologist* 22(7): 864–72.

Cervantes, A. et al. 2023. „Metastatic Colorectal Cancer: ESMO Clinical Practice Guideline for Diagnosis, Treatment and Follow-Up". *Annals of Oncology* 34(1): 10–32.

Cocco, Emiliano, Maurizio Scaltriti, a Alexander Drilon. 2018. „NTRK fusion-positive cancers and TRK inhibitor therapy". *Nature reviews. Clinical oncology* 15(12): 731–47.

Coppedè, Fabio, Angela Lopomo, Roberto Spisni, a Lucia Migliore. 2014. „Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer". *World Journal of Gastroenterology : WJG* 20(4): 943–56.

Cornish, Alex J. et al. 2022. „Whole genome sequencing of 2,023 colorectal cancers reveals mutational landscapes, new driver genes and immune interactions". https://dx.doi.org/10.1101/2022.11.16.515599.

Cunningham, Matthew P. et al. 2005. „Coexpression, prognostic significance and predictive value of EGFR, EGFRvIII and phosphorylated EGFR in colorectal cancer". *International Journal of Oncology* 27(2): 317–25.

Dang, Ha X. et al. 2020. „The clonal evolution of metastatic colorectal cancer". *Science Advances* 6(24): eaay9691.

Dardare, Julie et al. 2020. „SMAD4 and the TGFβ Pathway in Patients with Pancreatic Ductal Adenocarcinoma". *International Journal of Molecular Sciences* 21(10): 3534.

Davies, Helen et al. 2002. „Mutations of the BRAF gene in human cancer". *Nature* 417(6892): 949–54.

De Roock, Wendy et al. 2010. „Effects of KRAS, BRAF, NRAS, and PIK3CA Mutations on the Efficacy of Cetuximab plus Chemotherapy in Chemotherapy-Refractory Metastatic Colorectal Cancer: A Retrospective Consortium Analysis". *The Lancet Oncology* 11(8): 753–62.

Deng, Guoren et al. 2004. „BRAF Mutation Is Frequently Present in Sporadic Colorectal Cancer with Methylated hMLH1, But Not in Hereditary Nonpolyposis Colorectal Cancer". *Clinical Cancer Research* 10(1): 191–95.

Ding, Yi, a Guoli Chen. 2022. „Molecular Testing Panel in Colorectal Cancer". *Human Pathology Reports* 28: 300632.

Divgi, C. R. et al. 1991. „Phase I and Imaging Trial of Indium 111-Labeled Anti-Epidermal Growth Factor Receptor Monoclonal Antibody 225 in Patients With Squamous Cell Lung Carcinoma". *JNCI: Journal of the National Cancer Institute* 83(2): 97–104.

Domingo, Enric et al. 2013. „Evaluation of PIK3CA Mutation as a Predictor of Benefit from Nonsteroidal Anti-Inflammatory Drug Therapy in Colorectal Cancer". *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 31(34): 4297–4305.

Fan, H. Christina et al. 2010. „Analysis of the Size Distributions of Fetal and Maternal Cell-Free DNA by Paired-End Sequencing". *Clinical Chemistry* 56(8): 1279–86.

Fevr, Tea, Sylvie Robine, Daniel Louvard, a Joerg Huelsken. 2007. „Wnt/β-Catenin Is Essential for Intestinal Homeostasis and Maintenance of Intestinal Stem Cells". *Molecular and Cellular Biology* 27(21): 7551–59.

Filip, Stanislav et al. 2020. „Distant Metastasis in Colorectal Cancer Patients—Do We Have New Predicting Clinicopathological and Molecular Biomarkers? A Comprehensive Review". *International Journal of Molecular Sciences* 21(15): 5255.

Filippo, C. De et al. 2002. „Mutations of the APC Gene in Human Sporadic Colorectal Cancers". *Scandinavian Journal of Gastroenterology* 37(9): 1048–53.

Fischer, Jesse et al. 2019. „Clinical Implications of the Genetics of Sporadic Colorectal Cancer". *ANZ Journal of Surgery* 89(10): 1224–29.

Flohr, T. et al. 1999. „Detection of Mutations in the DNA Polymerase Delta Gene of Human Sporadic Colorectal Cancers and Colon Cancer Cell Lines". *International Journal of Cancer* 80(6): 919–29.

Francoual, M. et al. 2006. „EGFR in Colorectal Cancer: More than a Simple Receptor". *Annals of Oncology* 17(6): 962–67.

Frattini, Milo et al. 2006. „Quantitative Analysis of Plasma DNA in Colorectal Cancer Patients". *Annals of the New York Academy of Sciences* 1075(1): 185–90.

Gala, Manish K. et al. 2014. „Germline Mutations in Oncogene-Induced Senescence Pathways are Associated with Multiple Sessile Serrated Adenomas". *Gastroenterology* 146(2): 520–29.

Gannon, J. V., R. Greaves, R. Iggo, a D. P. Lane. 1990. „Activating Mutations in P53 Produce a Common Conformational Effect. A Monoclonal Antibody Specific for the Mutant Form". *The EMBO journal* 9(5): 1595–1602.

Gonzalez-Zulueta, M. et al. 1995. „Methylation of the 5' CpG Island of the P16/CDKN2 Tumor Suppressor Gene in Normal and Transformed Human Tissues Correlates with Gene Silencing". *Cancer Research* 55(20): 4531–35.

Graff, J. R. et al. 1995. „E-Cadherin Expression Is Silenced by DNA Hypermethylation in Human Breast and Prostate Carcinomas". *Cancer Research* 55(22): 5195–99.

Groden, Joanna et al. 1991. „Identification and Characterization of the Familial Adenomatous Polyposis Coli Gene". *Cell* 66(3): 589–600.

Gross, M. E. et al. 1991. „Cellular Growth Response to Epidermal Growth Factor in Colon Carcinoma Cells with an Amplified Epidermal Growth Factor Receptor Derived from a Familial Adenomatous Polyposis Patient1". *Cancer Research* 51(5): 1452–59.

Gryfe, Robert et al. 2000. „Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer". *New England Journal of Medicine* 342(2): 69–77.

Guinney, Justin et al. 2015. „The Consensus Molecular Subtypes of Colorectal Cancer". *Nature medicine* 21(11): 1350–56.

Haber, Daniel A., a Victor E. Velculescu. 2014. „Blood-Based Analyses of Cancer: Circulating Tumor Cells and Circulating Tumor DNA". *Cancer discovery* 4(6): 650.

Half, Elizabeth, Dani Bercovich, a Paul Rozen. 2009. „Familial adenomatous polyposis". *Orphanet Journal of Rare Diseases* 4: 22.

Haraldsdottir, Sigurdis et al. 2017. „Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2". *Nature Communications* 8: 14755.

He, T. C. et al. 1998. „Identification of C-MYC as a Target of the APC Pathway". *Science (New York, N.Y.)* 281(5382): 1509–12.

Head, Steven R. et al. 2014. „Library Construction for Next-Generation Sequencing: Overviews and Challenges". *BioTechniques* 56(2): 61–64, 66, 68, passim.

Hemmings, Brian A., a David F. Restuccia. 2012. „PI3K-PKB/Akt Pathway". *Cold Spring Harbor Perspectives in Biology* 4(9): a011189.

Herman, James G. et al. 1998. „Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma". *Proceedings of the National Academy of Sciences of the United States of America* 95(12): 6870–75.

Hobbs, G. Aaron, Channing J. Der, a Kent L. Rossman. 2016. „RAS isoforms and mutations in cancer at a glance". *Journal of Cell Science* 129(7): 1287–92.

Hong, D.S. et al. 2020. „KRASG12C Inhibition with Sotorasib in Advanced Solid Tumors". *The New England journal of medicine* 383(13): 1207–17.

ten Hoorn, Sanne, Tim R de Back, Dirkje W Sommeijer, a Louis Vermeulen. 2022. „Clinical Value of Consensus Molecular Subtypes in Colorectal Cancer: A Systematic Review and Meta-Analysis". *JNCI: Journal of the National Cancer Institute* 114(4): 503–16.

Hossain, Md Sanower et al. 2022. „Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies". *Cancers* 14(7): 1732.

Howe, J. R. et al. 2001. „Germline Mutations of the Gene Encoding Bone Morphogenetic Protein Receptor 1A in Juvenile Polyposis". *Nature Genetics* 28(2): 184–87.

Howe, James R. et al. 1998. „Mutations in the SMAD4/DPC4 Gene in Juvenile Polyposis". *Science* 280(5366): 1086–88.

Huang, Yuan et al. 2019. „Mutant p53 drives cancer chemotherapy resistance due to loss of function on activating transcription of PUMA". *Cell Cycle* 18(24): 3442–55.

Chan, Allen K. C., Rossa W. K. Chiu, a Y. M. Dennis Lo. 2003. „Cell-Free Nucleic Acids in Plasma, Serum and Urine: A New Tool in Molecular Diagnosis". *Annals of Clinical Biochemistry* 40(2): 122–30.

Iizuka, Norio et al. 2006. „Elevated Levels of Circulating Cell-Free DNA in the Blood of Patients with Hepatitis C Virus-Associated Hepatocellular Carcinoma". *Anticancer Research* 26(6C): 4713–19.

Ionov, Y. et al. 1993. „Ubiquitous Somatic Mutations in Simple Repeated Sequences Reveal a New Mechanism for Colonic Carcinogenesis". *Nature* 363(6429): 558–61.

Iwatsuki, Masaaki et al. 2010. „Loss of FBXW7, a Cell Cycle Regulating Gene, in Colorectal Cancer: Clinical Significance". *International Journal of Cancer* 126(8): 1828–37.

Jahr, Sabine et al. 2001. „DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells1". *Cancer Research* 61(4): 1659–65.

Jones, Siân et al. 2002. „Biallelic Germline Mutations in MYH Predispose to Multiple Colorectal Adenoma and Somatic G:C-->T:A Mutations". *Human Molecular Genetics* 11(23): 2961–67.

Keating, Gillian M. 2010. „Panitumumab". *Drugs* 70(8): 1059–78.

Kim, Hee Sung et al. 2013. „Mutation at Intronic Repeats of the Ataxia-Telangiectasia Mutated (ATM) Gene and ATM Protein Loss in Primary Gastric Cancer with Microsatellite Instability". *PLOS ONE* 8(12): e82769.

Kohler, Eva Maria et al. 2008. „Functional Definition of the Mutation Cluster Region of Adenomatous Polyposis Coli in Colorectal Tumours". *Human Molecular Genetics* 17(13): 1978–87.

Koncina, Eric, Serge Haan, Stefan Rauh, a Elisabeth Letellier. 2020. „Prognostic and Predictive Molecular Biomarkers for Colorectal Cancer: Updates and Challenges". *Cancers* 12(2): 319.

Kopetz, Scott et al. 2019. „Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer". *New England Journal of Medicine* 381(17): 1632–43.

Kubiczkova, Lenka, Lenka Sedlarikova, Roman Hajek, a Sabina Sevcikova. 2012. „TGF-β – an Excellent Servant but a Bad Master". *Journal of Translational Medicine* 10(1): 183.

Lao, Victoria Valinluck, a William M. Grady. 2011. „Epigenetics and Colorectal Cancer". *Nature reviews. Gastroenterology & hepatology* 8(12): 686–700.

Larsen Haidle, Joy, Suzanne P. MacFarland, a James R. Howe. 1993. „Juvenile Polyposis Syndrome". In *GeneReviews®*, ed. Margaret P. Adam et al. Seattle (WA): University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1469/ (10. červenec 2023).

Le, Dung T. et al. 2015. „PD-1 Blockade in Tumors with Mismatch-Repair Deficiency". *New England Journal of Medicine* 372(26): 2509–20.

———. 2017. „Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade". *Science* 357(6349): 409–13.

Lee, Ji-Hoon, a Tanya T. Paull. 2005. „ATM Activation by DNA Double-Strand Breaks through the Mre11-Rad50-Nbs1 Complex". *Science (New York, N.Y.)* 308(5721): 551–54.

Lenz, Heinz-Josef et al. 2019. „Impact of Consensus Molecular Subtype on Survival in Patients With Metastatic Colorectal Cancer: Results From CALGB/SWOG 80405 (Alliance)". *Journal of Clinical Oncology* 37(22): 1876–85.

Li, Gerald et al. 2019. „Genomic Profiling of Cell-Free Circulating Tumor DNA in Patients with Colorectal Cancer and Its Fidelity to the Genomics of the Tumor Biopsy". *Journal of Gastrointestinal Oncology* 10(5): 831–40.

Li, Jian et al. 2017. „Cell-free DNA copy number variations in plasma from colorectal cancer patients". *Molecular Oncology* 11(8): 1099–1111.

Liao, Xiaoyun et al. 2012. „Aspirin Use, Tumor PIK3CA Mutation, and Colorectal-Cancer Survival". *New England Journal of Medicine* 367(17): 1596–1606.

Lièvre, Astrid et al. 2006. „KRAS Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer". *Cancer Research* 66(8): 3992–95.

Lim, Seok-Byung et al. 2004. „Prognostic significance of microsatellite instability in sporadic colorectal cancer". *International Journal of Colorectal Disease* 19(6): 533–37.

Liu, B. et al. 1994. „HMSH2 Mutations in Hereditary Nonpolyposis Colorectal Cancer Kindreds". *Cancer Research* 54(17): 4590–94.

Liu, Jiangang et al. 2020. „Molecular Dissection of CRC Primary Tumors and Their Matched Liver Metastases Reveals Critical Role of Immune Microenvironment, EMT and Angiogenesis in Cancer Metastasis". *Scientific Reports* 10(1): 10725.

Lynch, Henry T. et al. 2015. „Milestones of Lynch Syndrome: 1895–2015". *Nature Reviews Cancer* 15(3): 181–94.

MacGrogan, Donal, Mark Pegram, Dennis Slamon, a Robert Bookstein. 1997. „Comparative Mutational Analysis of DPC4 (Smad4) in Prostatic and Colorectal Carcinomas". *Oncogene* 15(9): 1111–14.

Mancl, Erin E., Jill M. Kolesar, a Lee C. Vermeulen. 2009. „Clinical and economic value of screening for Kras mutations as predictors of response to epidermal growth factor receptor inhibitors". *American Journal of Health-System Pharmacy* 66(23): 2105–12.

Mandel, P., a P. Metais. 1948. „[Nuclear Acids In Human Blood Plasma]". *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales* 142(3–4): 241–43.

Mao, Jian-Hua et al. 2004. „Fbxw7/Cdc4 Is a P53-Dependent, Haploinsufficient Tumour Suppressor Gene". *Nature* 432(7018): 775–79.

Markowitz, S. et al. 1995. „Inactivation of the Type II TGF-Beta Receptor in Colon Cancer Cells with Microsatellite Instability". *Science (New York, N.Y.)* 268(5215): 1336–38.

Mendelsohn, John, Marie Prewett, Patricia Rockwell, a Neil I. Goldstein. 2015. „CCR 20th Anniversary Commentary: A Chimeric Antibody, C225, Inhibits EGFR Activation and Tumor Growth". *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 21(2): 227–29.

Millis, Sherri Z. et al. 2016. „Landscape of Phosphatidylinositol-3-Kinase Pathway Alterations Across 19 784 Diverse Solid Tumors". *JAMA Oncology* 2(12): 1565–73.

Minella, Alex C., a Bruce E. Clurman. 2005. „Mechanisms of Tumor Suppression by the SCF(Fbw7)". *Cell Cycle (Georgetown, Tex.)* 4(10): 1356–59.

Miyaki, Michiko et al. 1997. „Germline Mutation of MSH6 as the Cause of Hereditary Nonpolyposis Colorectal Cancer". *Nature Genetics* 17(3): 271–72.

Modest, D. P. et al. 2016. „Outcome according to KRAS-, NRAS- and BRAF-mutation as well as KRAS mutation variants: pooled analysis of five randomized trials in metastatic colorectal cancer by the AIO colorectal cancer study group". *Annals of Oncology* 27(9): 1746–53.

Mohamed, Salem Y. et al. 2019. „Role of VEGF, CD105, and CD31 in the Prognosis of Colorectal Cancer Cases". *Journal of Gastrointestinal Cancer* 50(1): 23–34.

Molnár, Béla et al. 2019. „Circulating cell-free nucleic acids as biomarkers in colorectal cancer screening and diagnosis - an update". *Expert Review of Molecular Diagnostics* 19(6): 477–98.

Müller, Annegret et al. 2004. „MSI-Testing in Hereditary Non-Polyposis Colorectal Carcinoma (HNPCC)". *Disease Markers* 20(4–5): 225–36.

Munemitsu, S et al. 1995. „Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein." *Proceedings of the National Academy of Sciences of the United States of America* 92(7): 3046–50.

Muzny, Donna M. et al. 2012. „Comprehensive Molecular Characterization of Human Colon and Rectal Cancer". *Nature* 487(7407): 330–37.

Nguyen, Ha Thi, a Hong-Quan Duong. 2018. „The molecular characteristics of colorectal cancer: Implications for diagnosis and therapy". *Oncology Letters* 16(1): 9–18.

Nicolaides, Nicholas C. et al. 1994. „Mutations of Two P/WS Homologues in Hereditary Nonpolyposis Colon Cancer". *Nature* 371(6492): 75–80.

Oda, Kanae, Yukiko Matsuoka, Akira Funahashi, a Hiroaki Kitano. 2005. „A comprehensive pathway map of epidermal growth factor receptor signaling". *Molecular Systems Biology* 1: 2005.0010.

Ochs, Ann-Margaret et al. 2004. „Expression of Vascular Endothelial Growth Factor and HER2/Neu in Stage II Colon Cancer and Correlation with Survival". *Clinical Colorectal Cancer* 4(4): 262–67.

Olivier, Magali, Monica Hollstein, a Pierre Hainaut. 2010. „TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use". *Cold Spring Harbor Perspectives in Biology* 2(1): a001008.

Papadopoulos, Nickolas et al. 1994. „Mutation of a mutL Homolog in Hereditary Colon Cancer". *Science* 263(5153): 1625–29.

Parikh, Aparna R. et al. 2019. „Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers". *Nature medicine* 25(9): 1415–21.

Park, Dong Il et al. 2007. „HER-2/Neu Overexpression Is an Independent Prognostic Factor in Colorectal Cancer". *International Journal of Colorectal Disease* 22(5): 491–97.

Pathak, Yashwant, ed. 2022. : *: Nanotechnology and Therapeutic Applications*. Boca Raton: CRC Press.

Peeters, Marc et al. 2015. „Analysis of KRAS/NRAS Mutations in a Phase III Study of Panitumumab with FOLFIRI Compared with FOLFIRI Alone as Second-line Treatment for Metastatic Colorectal Cancer". *Clinical Cancer Research* 21(24): 5469–79.

Peltomäki, P. et al. 1993. „Microsatellite Instability Is Associated with Tumors That Characterize the Hereditary Non-Polyposis Colorectal Carcinoma Syndrome". *Cancer Research* 53(24): 5853–55.

Perrone, F. et al. 2009. „PI3KCA/PTEN Deregulation Contributes to Impaired Responses to Cetuximab in Metastatic Colorectal Cancer Patients". *Annals of Oncology: Official Journal of the European Society for Medical Oncology* 20(1): 84–90.

Phan, Liem Minh, a Abdol-Hossein Rezaeian. 2021. „ATM: Main Features, Signaling Pathways, and Its Diverse Roles in DNA Damage Response, Tumor Suppression, and Cancer Development". *Genes* 12(6): 845.

Pietrantonio, Filippo et al. 2014. „Gain of ALK Gene Copy Number May Predict Lack of Benefit from Anti-EGFR Treatment in Patients with Advanced Colorectal Cancer and RAS-RAF-PI3KCA Wild-Type Status". *PLoS ONE* 9(4): e92147.

———. 2017. „ALK, ROS1, and NTRK Rearrangements in Metastatic Colorectal Cancer". *JNCI: Journal of the National Cancer Institute* 109(12). https://dx.doi.org/10.1093/jnci/djx089.

Pilat, N. et al. 2015. „Assessing the TP53 Marker Type in Patients Treated with or without Neoadjuvant Chemotherapy for Resectable Colorectal Liver Metastases: A P53 Research Group Study". *European Journal of Surgical Oncology* 41(5): 683–89.

Prenen, Hans et al. 2009. „PIK3CA Mutations Are Not a Major Determinant of Resistance to the Epidermal Growth Factor Receptor Inhibitor Cetuximab in Metastatic Colorectal Cancer". *Clinical Cancer Research* 15(9): 3184–88.

Punnoose, Elizabeth A. et al. 2012. „Evaluation of Circulating Tumor Cells and Circulating Tumor DNA in Non–Small Cell Lung Cancer: Association with Clinical Endpoints in a Phase II Clinical Trial of Pertuzumab and Erlotinib". *Clinical Cancer Research* 18(8): 2391–2401.

Qi, Maosong, a Elaine A. Elion. 2005. „MAP kinase pathways". *Journal of Cell Science* 118(16): 3569–72.

Reimers, Marlies S. et al. 2013. „Biomarkers in precision therapy in colorectal cancer". *Gastroenterology Report* 1(3): 166–83.

Reinert, Thomas et al. 2019. „Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer". *JAMA Oncology* 5(8): 1124–31.

Reynolds, Neil A, a Antona J Wagstaff. 2004. „Cetuximab". *Drugs* 64(1): 109–18.

Richman, Susan D. et al. 2009. „KRAS and BRAF Mutations in Advanced Colorectal Cancer Are Associated With Poor Prognosis but Do Not Preclude Benefit From Oxaliplatin or Irinotecan: Results From the MRC FOCUS Trial". *Journal of Clinical Oncology* 27(35): 5931–37.

Richman, Susan D et al. 2016. „HER2 Overexpression and Amplification as a Potential Therapeutic Target in Colorectal Cancer: Analysis of 3256 Patients Enrolled in the QUASAR, FOCUS and PICCOLO Colorectal Cancer Trials". *The Journal of Pathology* 238(4): 562–70.

Rohlin, Anna et al. 2016. „GREM1 and POLE Variants in Hereditary Colorectal Cancer Syndromes". *Genes, Chromosomes & Cancer* 55(1): 95–106.

Rothblum-Oviatt, Cynthia et al. 2016. „Ataxia telangiectasia: a review". *Orphanet Journal of Rare Diseases* 11(1): 159.

Russo, Mariangela et al. 2019. „Adaptive mutability of colorectal cancers in response to targeted therapies". *Science* 366(6472): 1473–80.

Salvianti, Francesca et al. 2021. „Circulating tumour cells and cell-free DNA as a prognostic factor in metastatic colorectal cancer: the OMITERC prospective study". *British Journal of Cancer* 125(1): 94–100.

Samuels, Yardena et al. 2004. „High Frequency of Mutations of the PIK3CA Gene in Human Cancers". *Science* 304(5670): 554–554.

Sargent, Daniel J. et al. 2010. „Defective Mismatch Repair As a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer". *Journal of Clinical Oncology* 28(20): 3219–26.

Sawada, Kentaro et al. 2018. „Prognostic and Predictive Value of HER2 Amplification in Patients With Metastatic Colorectal Cancer". *Clinical Colorectal Cancer* 17(3): 198–205.

Shtutman, Michael et al. 1999. „The cyclin D1 gene is a target of the β-catenin/LEF-1 pathway". *Proceedings of the National Academy of Sciences of the United States of America* 96(10): 5522–27.

Schiemer, James. 2011. „Illumina TruSeq DNA Adapters De-Mystified".

Schwartz, S. et al. 1999. „Frameshift Mutations at Mononucleotide Repeats in Caspase-5 and Other Target Genes in Endometrial and Gastrointestinal Cancer of the Microsatellite Mutator Phenotype". *Cancer Research* 59(12): 2995–3002.

Siena, Salvatore et al. 2021. „Trastuzumab Deruxtecan (DS-8201) in Patients with HER2-Expressing Metastatic Colorectal Cancer (DESTINY-CRC01): A Multicentre, Open-Label, Phase 2 Trial". *The Lancet Oncology* 22(6): 779–89.

Siravegna, Giulia et al. 2015. „Monitoring clonal evolution and resistance to EGFR blockade in the blood of metastatic colorectal cancer patients". *Nature medicine* 21(7): 795–801.

Soukupová, Jana et al. 2016. „CZECANCA: CZEch CAncer paNel for Clinical Application – Design and Optimization of the Targeted Sequencing Panel for the Identification of Cancer Susceptibility in High-risk Individuals from the Czech Republic". *Klinicka onkologie* 29(Suppl 1): S46–54.

Souza, Rhonda F. et al. 1996. „Microsatellite Instability in the Insulin–like Growth Factor II Receptor Gene in Gastrointestinal Tumours". *Nature Genetics* 14(3): 255–57.

Spagnol, Luigi Wolkmer et al. 2022. „P16 Gene Promoter Methylation Is Associated with Oncogenesis and Progression of Gastric Carcinomas: A Systematic Review and Meta-Analysis". *Critical Reviews in Oncology/Hematology* 180: 103843.

Spano, J. P. et al. 2005. „Epidermal Growth Factor Receptor Signaling in Colorectal Cancer: Preclinical Data and Therapeutic Perspectives". *Annals of Oncology* 16(2): 189–94.

Stroun, M. et al. 1989. „Neoplastic Characteristics of the DNA Found in the Plasma of Cancer Patients". *Oncology* 46(5): 318–22.

Stroun, M et al. 2001. „About the Possible Origin and Mechanism of Circulating DNA: Apoptosis and Active DNA Release". *Clinica Chimica Acta* 313(1): 139–42.

Stroun, Maurice et al. 2001. „Alu Repeat Sequences Are Present in Increased Proportions Compared to a Unique Gene in Plasma/Serum DNA". *Annals of the New York Academy of Sciences* 945(1): 258–64.

Sundar, Raghav et al. 2018. „Ataxia Telangiectasia Mutated Protein Loss and Benefit From Oxaliplatin-Based Chemotherapy in Colorectal Cancer". *Clinical Colorectal Cancer* 17(4): 280–84.

Sveen, Anita et al. 2018. „Colorectal Cancer Consensus Molecular Subtypes Translated to Preclinical Models Uncover Potentially Targetable Cancer Cell Dependencies". *Clinical Cancer Research* 24(4): 794–806.

Tanaka, Kozo, a Toru Hirota. 2016. „Chromosomal Instability: A Common Feature and a Therapeutic Target of Cancer". *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1866(1): 64–75.

Tarafa, G et al. 2000. „DCC and SMAD4 alterations in human colorectal and pancreatic tumor dissemination". *Oncogene* 19(4): 546–55.

Taylor, C.F. et al. 2003. „Genomic deletions inMSH2 orMLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: Identification of novel and recurrent deletions by MLPA". *Human Mutation* 22(6): 428–33.

Thibodeau, S. N., G. Bren, a D. Schaid. 1993. „Microsatellite Instability in Cancer of the Proximal Colon". *Science* 260(5109): 816–19.

Tie, Jeanne et al. 2016. „Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer". *Science translational medicine* 8(346): 346ra92.

Toyota, Minoru et al. 1999. „CpG island methylator phenotype in colorectal cancer". *Proceedings of the National Academy of Sciences of the United States of America* 96(15): 8681–86.

Valle, Laura, Eduardo Vilar, Sean V. Tavtigian, a Elena M. Stoffel. 2019. „Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine." *The Journal of pathology* 247(5): 574–88.

Velho, Sérgia et al. 2014. „Causes and consequences of microsatellite instability in gastric carcinogenesis". *World Journal of Gastroenterology : WJG* 20(44): 16433–42.

Voutsadakis, Ioannis A. 2021. „The Landscape of PIK3CA Mutations in Colorectal Cancer". *Clinical Colorectal Cancer* 20(3): 201–15.

Wang, Chen et al. 2017. „ATM-Deficient Colorectal Cancer Cells Are Sensitive to the PARP Inhibitor Olaparib". *Translational Oncology* 10(2): 190–96.

Wang, Jaw-Yuan et al. 2004. „Molecular Detection of APC, K-Ras, and P53 Mutations in the Serum of Colorectal Cancer Patients as Circulating Biomarkers". *World Journal of Surgery* 28(7): 721–26.

White, Alan et al. 2018. „A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK". *BMC Cancer* 18: 906.

Wood, Laura D. et al. 2007. „The Genomic Landscapes of Human Breast and Colorectal Cancers". *Science (New York, N.Y.)* 318(5853): 1108–13.

Xi, Yue, a Pengfei Xu. 2021. „Global colorectal cancer burden in 2020 and projections to 2040". *Translational Oncology* 14(10): 101174.

Yamagishi, Hidetsugu, Hajime Kuroda, Yasuo Imai, a Hideyuki Hiraishi. 2016. „Molecular pathogenesis of sporadic colorectal cancers". *Chinese Journal of Cancer* 35: 4.

Yamamoto, H., H. Sawai, a M. Perucho. 1997. „Frameshift Somatic Mutations in Gastrointestinal Cancer of the Microsatellite Mutator Phenotype". *Cancer Research* 57(19): 4420–26.

Yang, Ying-Chi et al. 2018. „Circulating Tumor DNA Detectable in Early- and Late-Stage Colorectal Cancer Patients". *Bioscience Reports* 38(4): BSR20180322.

Yarden, Yosef, a Mark X. Sliwkowski. 2001. „Untangling the ErbB signalling network". *Nature Reviews Molecular Cell Biology* 2(2): 127–37.

Yarchoan, Mark, Alexander Hopkins, a Elizabeth M. Jaffee. 2017. „Tumor Mutational Burden and Response Rate to PD-1 Inhibition". *The New England journal of medicine* 377(25): 2500–2501.

Ye, Jun et al. 2020. „Tissue gene mutation profiles in patients with colorectal cancer and their clinical implications". *Biomedical Reports* 13(1): 43–48.

Yoon, Harry H. et al. 2014. „KRAS codon 12 and 13 mutations in relation to disease-free survival in BRAF-wild type stage III colon cancers from an adjuvant chemotherapy trial (N0147 Alliance)". *Clinical cancer research : an official journal of the American Association for Cancer Research* 20(11): 3033–43.

Yoshida, Rintaro et al. 2011. „Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer". *European Journal of Human Genetics* 19(3): 320–25.

Yoshitaka, Tomoya et al. 1996. „Mutations of E2F-4 Trinucleotide Repeats in Colorectal Cancer with Microsatellite Instability". *Biochemical and Biophysical Research Communications* 227(2): 553–57.

Zenonos, Kypros, a Katy Kyprianou. 2013. „RAS signaling pathways, mutations and their role in colorectal cancer". *World Journal of Gastrointestinal Oncology* 5(5): 97–101.

Zhao, Ming, Lopa Mishra, a Chu-Xia Deng. 2018. „The role of TGF-β/SMAD4 signaling in cancer". *International Journal of Biological Sciences* 14(2): 111–23.