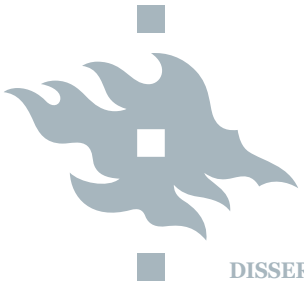

DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

84/2018

OMAR YOUSSEF

Analysis of Gene Mutations in Exhaled Breath Condensate from Healthy and Lung Cancer Individuals and Profiling of Mutations and Gut Microbiota in Stools from Patients with Gastrointestinal Neoplasms

MEDICUM
DEPARTMENT OF PATHOLOGY
FACULTY OF MEDICINE
DOCTORAL PROGRAMME IN BIOMEDICINE
UNIVERSITY OF HELSINKI

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis No. 84/2018

ISSN 2342-3161 (print)
ISSN 2342-317X (online)
ISBN 978-951-51-4644-1 (paperback)
ISBN 978-951-51-4645-8 (PDF)
https://ethesis.helsinki.fi/

Hansaprint
Vantaa 2018

Medicum
Department of Pathology
University of Helsinki
Helsinki, Finland

Doctoral Program in Biomedicine (DPBM)
Doctoral School in Health Sciences (DSHealth)
University of Helsinki
Helsinki, Finland

Analysis of gene mutations in exhaled breath condensate from healthy and lung cancer individuals and profiling of mutations and gut microbiota in stools from patients with gastrointestinal neoplasms

Omar Youssef

Academic dissertation

To be presented, with the permission of the Faculty of Medicine, University of Helsinki, for public examination in auditorium XII, Päärakennus, Fabianinkatu 33, Helsinki, on November 16th, 2018, at 12 noon

Helsinki 2018

Supervised by

Professor Sakari Knuutila, PhD
Medicum
Department of Pathology
University of Helsinki
Helsinki, Finland

Virinder Kaur Sarhadi, PhD
Medicum
Department of Pathology
University of Helsinki
Helsinki, Finland

Reviewed by

Antti Jekunen, MD, PhD
Professor, Chief Physician
Vasa Oncology Clinic
Turku University, Vaasa Central Hospital
Vaasa, Finland

Anne Kallioniemi, MD, PhD
Professor of Biotechnology and Medical Technology
Faculty of Medicine and Life Sciences
University of Tampere
Tampere, Finland

Official Opponent

Tarja Laitinen, MD, PhD
Chief Physician, Professor
Faculty of Medicine, University of Turku
Turku, Finland

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations

CONTENTS

ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numeral (I-IV):

I. **Omar Youssef,** Paivi Piirila, Aija Knuuttila, Tom Bohling, Virinder Kaur Sarhadi, and Sakari Knuutila. Presence of cancer-associated mutations in exhaled breath condensates of healthy individuals by next generation sequencing. Oncotarget. 2017; 8: 18166-18176.

II. **Omar Youssef,** Aija Knuuttila, Paivi Piirila, Tom Bohling, Virinder Kaur Sarhadi, and Sakari Knuutila. Hotspot mutations detectable by next generation sequencing in exhaled breath condensates from patients with lung neoplasms. Anticancer Res. 2018; 38: 5627-5634.

III. **Omar Youssef,** Virinder Kaur Sarhadi, Homa Ehsan, Tom Böhling, Monika Carpelan-Holmström, Selja Koskensalo, Pauli Puolakkainen, Arto Kokkola and Sakari Knuutila. Stool mutations in gastric and colorectal neoplasia patients by next generation sequencing. World J Gastroenterol. 2017; 23: 8291-8299.

IV. **Omar Youssef,** Leo Lahti, Arto Kokkola, Tiina Karla, Milja Tikkanen, Homa Ehsan, Monica Carpelan-Holmström, Selja Koskensalo, Tom Böhling, Hilpi Rautelin, Pauli Puolakkainen, Sakari Knuutila, and Virinder Sarhadi. Stool microbiota composition differs in patients with stomach, colon and rectal neoplasms. Digest Dis Sci. doi: 10.1007/s10620-018-5190-5. [Epub ahead of print]

ABBREVIATIONS

16S rRNA	rRNA 16S ribosomal RNA
AAH	atypical adenomatous hyperplasia
ADC	adenocarcinoma
ATP	adenosine triphosphate
BSC	best supportive care
CpG	cytosine being 5 prime to the guanine base.
CNV	copy number variation
COSMIC	the Catalogue of Somatic Mutations in Cancer
dbSNP	the Single Nucleotide Polymorphism Database
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
FDA	Food and Drug Administration
GIT	gastrointestinal tract
GTPase	small guanosine triphosphatase
HPV	human papilloma virus
HR	hazard ratio
IGV	Integrative Genomics Viewer
IHC	immunohistochemistry
LCC	large cell carcinoma
LC-CRT	long-course chemoradiotherapy
miRNA	micro-RNA
mRNA	messenger ribonucleic acid
MSI	microsatellite instability
NGS	next-generation sequencing
NSCLC	non-small cell lung cancer
OS	overall survival
PCR	polymerase chain reaction
PFS	progression-free survival
PGM	Personal Genome Machine
RAF	rapidly accelerated fibrosarcoma
RR	response rate
RTK	receptor tyrosine kinase
SBRT	stereotactic body radiation therapy
SCC	squamous cell carcinoma
SCLC	small cell lung cancer

SC-RT	short-course radiotherapy
SNP	single nucleotide polymorphism
SNV	single-nucleotide variant
TKI	tyrosine kinase inhibitor
TNM	tumor, node, metastasis
WHO	World Health Organization

Gene symbols are in *italics* within the text following the guidelines of the Human Genome Organization nomenclature committee (HGNC). Gene names used in the thesis can be found at http://www.genenames.org/.

ABSTRACT

Lung and gastrointestinal (GIT) cancers are two types of malignancies in which early diagnosis has a significant impact on prognosis and better survival rate. Exhaled breath condensate (EBC) from lung cancer patients and stool from patients with GIT tumors can represent non-invasive sources for diagnosis of malignancy at an early stage. These materials contain DNA from cells exfoliated from malignant or pre-malignant lesions and consequently could reflect all genetic alterations occurring during development of the cancer. Stool samples are also a good source to study gut bacterial composition. Changes in the gut bacterial profile are linked to many diseases including GIT cancers. The aim of the study was to explore gene mutations in these samples, and to test their feasibility for the detection of malignancy in different tumor stages, including both early and advanced stages. A further aim was to investigate differences in the gut microbiota profile in stool samples of GIT cancer patients based on the location of the tumor.

The study material consisted of EBC samples from 29 lung cancer patients and 20 healthy individuals and stool specimens collected from 87 GIT neoplasia patients and 13 healthy individuals included as controls. DNA was isolated from both the EBC and stool samples. Targeted amplicon next generation sequencing (NGS) and 16S rRNA sequencing, using the Ion Torrent platforms, were performed to study gene mutations and stool bacterial profiling, respectively.

In study I, the methodology was optimized for applying NGS to study gene mutations in the EBC DNA from healthy individuals. The results revealed 15 subjects showing a total of 35 hotspot mutations in their EBC samples. The most frequent hotspot mutations occurred at *TP53, KRAS, NRAS,* and *SMAD4* genes. A codon 12 *KRAS* G12V mutation was detected in one control EBC sample with a mutant allele fraction of 6.8%. In the follow-up, study II, the same methodological steps were applied to the DNA isolated from EBC samples of patients with lung neoplasms. The success rate was 67.9% with 17 patients revealing a total of 39 hotspot mutations in their

EBC. The most frequent hotspot mutations occurred in the following genes: *TP53, SMAD4, PIK3CA*, and *KRAS*. A codon 13 *KRAS* G13D mutation was detected in one patient's EBC sample with a mutant allele fraction of 17%. The average mutant allele fraction for the gene mutations seen in patients were higher compared to that in controls; e.g. for *TP53,* the average mutant allele fraction was 22.9% and 13.6% and for *KRAS*, 11.4% and 4.3% in the patients and controls, respectively.

In study III, a cancer hotspot gene panel together with colon and lung cancer gene panels were used to study mutations in stool DNA from 87 patients with gastric and colorectal neoplasms. The success rates were 78% and 87% for gastric and colorectal neoplasia, respectively. Stools from patients with gastric neoplasms revealed 5 hotspot mutations, while from colorectal neoplasms 20 hotspot mutations were found. *APC, TP53,* and *KRAS* were the most frequently mutated genes in colorectal neoplasms. However, *APC, CDKN2A,* and *EGFR* were the only genes that showed hotspot mutations in gastric neoplasms. Hotspot mutations could also be detected in stool DNA from benign (8 mutations) and early malignant (9 mutations) GIT neoplasms.

In study IV, bacterial profiling in stool samples from patients with GIT neoplasms revealed variations in abundance according to the site of the GIT neoplasm. Two families, Lactobacillaceae and Bifidobacteriaceae, showed lower relative abundance while Enterobacteriaceae showed higher relative abundance when compared with control samples. The observed bacterial diversity could serve as an indicator in GIT neoplasms and help in disease monitoring.

To conclude, EBC and stool specimens are easily accessible non-invasive samples that could be used for studying different genetic alterations in neoplasms. Our studies revealed that NGS is a sensitive molecular technique that can be successfully applied to study gene mutations in multiple cancer genes from a very small amount of input DNA.

INTRODUCTION

Cancer is a disease in which the genome is always modified by various genetic and epigenetic alterations. Genetic alterations include mutations and chromosomal aberrations (structural and numerical), while epigenetic alterations include DNA methylation, non-coding micro-RNAs (miRNA), and histone modifications. Genetic changes in cancer have added many molecular details to tumor classification. Recently, with the revolution of precision medicine, the need for genetic data to act as a guide in clinical diagnostic and therapeutic decisions is crucial. Likewise, assessment of these genetic changes is essential for understanding the mechanisms of resistance to several cancer targeted therapies.

One of the major challenges is that the standard method for analysis of these molecular changes is tissue biopsies, which are not always easy to obtain, especially in early tumor stages. To overcome these obstacles, researchers have started to look for alternative novel methods for assessment of tumor molecular alterations. Non-invasive specimens represented the ideal solution by providing simple and easily accessible materials that could be obtained at different disease stages. There are several examples of non-invasive specimens, but in this thesis, I focused only on exhaled breath condensate (EBC) and stool samples from lung and gastrointestinal (GIT) neoplasia patients, respectively. However, there are a number of challenges when analyzing the non-invasive samples, for instance, the inconsistency in the quality and/or quantity of circulating tumor DNA, the variability in circulating tumor cells, and the lower frequency and volume of gene alterations occurring at the very early malignancy stages.

With the evolution of advanced and high throughput next generation sequencing (NGS), it is now possible to analyze different kinds of non-invasive samples with small amounts if DNA as an input. NGS enables researchers to analyze multiple distinct alterations simultaneously in a time and cost-efficient way. Moreover, NGS analysis of the conventional tissue specimens has been approved for clinical use especially in cancer diagnostics.

Lung and GIT neoplasms are two groups of malignancies in which clinically relevant mutations have been described. Currently, the role of genetic mutations as diagnostic and predictive markers has been established, and their detection and assessment has become crucial for early diagnosis and targeted therapy in these two types of malignancies. Moreover, the involvement of gut microbiota in promoting growth of GIT neoplasms is well acknowledged. Therefore, this thesis focuses on investigating EBC and stool as non-invasive samples for detection of clinically significant gene mutations, and assessment of gut microbiota composition in various GIT neoplasms, by using the targeted NGS molecular technique.

REVIEW OF LITERATURE

1. Cancer genetics, epigenetics, and development

Cancer is a continuous proliferative process in which a cell starts abnormal uncontrolled growth that occurs in a multistep manner. The process starts with tumor initiation because of certain genetic and epigenetic alterations. These alterations can happen in the same gene in different types of cancers [1]. The principal genes involved in the tumorigenesis process are proto-oncogenes and tumor suppressor genes, which include DNA mismatch repair genes. On the one hand, proto-oncogenes code for proteins that regulate cell growth and differentiation, and they can be transformed to oncogenes by mutations or increased expression leading to increased cell proliferation. On the other hand, tumor suppressor genes code for proteins that have an inhibitory effect on cell proliferation. In general terms, in tumor suppressor genes, both alleles that code for a particular protein must be altered before an effect is noticed, which is better known as the "two hit theory", whereas, a proto-oncogene alteration in one allele is sufficient for gaining function and transforming to an oncogene [2].

It is not known exactly why one person develops cancer and another person does not. There are several predisposing factors that contribute to cancer development including both non-genetic and genetic factors. Non-genetic risk factors include carcinogens such as smoking, air pollution, chemicals, radiation, chronic inflammation, microorganisms, e.g. viruses, and sun exposure. These factors also include non-controllable factors like age and family history [3]. Genetic factors include mutations in cancer predisposition genes which constitute approximately 10% of hereditary cancers. Subjects with a mutant allele have an increased susceptibility to develop cancer [4,5].

.

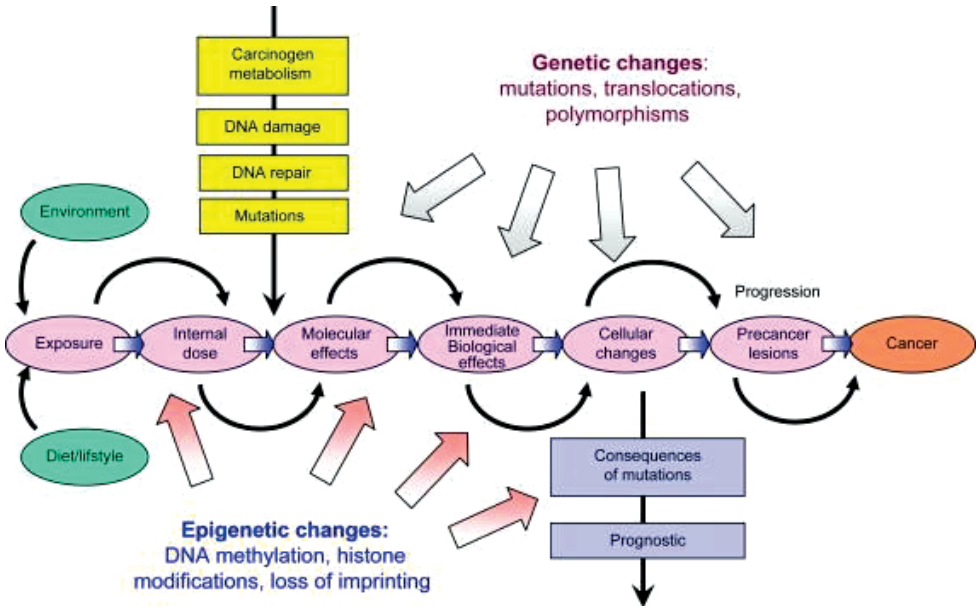

Figure 1. Genetic and epigenetic changes induced by different environmental factors contributing to carcinogenesis. Figure reproduced with permission from Wiley. Herceg et al., 2007 [6], copyright 2007.

It has been reported that different types of cancers share the same underlying characteristics, known collectively as "hallmarks of cancer". These hallmarks include: 1) self-sufficiency in growth signals, 2) insensitivity to antigrowth signals, 3) evasion of apoptosis, 4) limitless replication potential, 5) sustained angiogenesis, 6) tissue invasion. An additional four hallmarks were added to the previous ones including: 1) abnormal metabolic pathways, 2) evasion of the immune system, 3) genome instability, and 4) inflammation [7].

One of the fundamental features of cancer is tumor heterogeneity in which tumor cells are genetically and morphologically distinct and heterogeneous, and it can be inter-tumor and/or intra-tumor heterogeneity. Intra-tumor heterogeneity is an essential consideration when performing genome-wide analysis on a single tumor biopsy. It is also a major challenge in biomarker detection and in optimization of personalized medicine [8]. Inter-tumor heterogeneity refers to altered genotype and phenotype between cancer patients with the same cancer type and is usually induced by different etiological and environmental factors. Accordingly, genomic profiling of cancer

patients revealed diverse molecular subtypes and different protocols for targeted therapy [9].

1.1. Genetic alterations in cancer

Genetic alterations include small DNA alterations as well as large numerical and structural aberrations of the chromosomes. These alterations can lead to altered protein products that can serve as targets in guided gene therapy. Two to eight alterations need to occur before the process of tumorigenesis can be driven [8].

1.1.1 Small DNA alterations

Changes to short DNA nucleotide sequences are called gene level mutations, which include base substitution, base insertions, and base deletions. Substitutions occur when one base is replaced by another, known also as a single nucleotide variant (SNV) or point mutation. They are classified as transitions (purine to purine or pyrimidine to pyrimidine base) or transversions (pyrimidine to purine or purine to pyrimidine base). An effect of a point mutation can be reversed either by true reversion (another mutation reversing the original nucleotide status) or by a second site reversion (another mutation elsewhere causing regain of normal gene function). Point mutations occurring in exons can lead to three different kinds of mutations depending on the affected codon. The mutation might be: 1) a silent mutation, in which there is no change in the amino acid or resulting protein, 2) a missense mutation, in which there is a change in amino acid sequence resulting in an altered abnormal protein, 3) a non-sense mutation, in which there is an early stop codon leading to a shortened truncated protein. Changes affecting amino acid sequences are called non-synonymous mutations [10].

Insertions add one or more nucleotides into the DNA sequence. They can be caused by errors during replication. Insertion of nucleotides or frameshift in the exons may cause a shift in the reading frame (frameshift mutation), or disrupt splicing of mRNA (splice site mutation). Both types have significant effects on gene products. Similarly, deletions remove one or more nucleotides from a DNA sequence and can also cause

frameshift mutations. Both insertions and deletions are collectively known as indels [10].

1.1.2 Structural and numerical aberrations in chromosomes

Chromosomal aberrations are generally classified into numerical and structural abnormalities. Numerical abnormalities are whole chromosome changes that occur either by loss or gain of extra chromosomes, causing aneuploidy. Aneuploidy can cause disruption to the genome stability, loss of tumor suppressor genes, or activation of proto-oncogenes [11].

Structural aberrations can result from breakage or incorrect rejoining of chromosomal segments. They can be balanced (if the whole chromosomal set is still preserved, though rearranged), or unbalanced. Unbalanced structural abnormalities include deletions and insertions affecting the normal chromosomal copy number leading to copy number variations (CNV) [12,13]. Balanced abnormalities include inversions, and translocations, which can result in the production of fused genes such as the *ALK* gene fusion with *EML4* in non-small cell lung carcinoma (NSCLC) [14]. Other structural abnormalities include a ring chromosome, which occurs when there are two chromosomal breaks and the broken ends join each other giving rise to a ring. An isochromosome can be produced through duplication of the same chromosomal arm when the other arm is missing [15].

1.2. Epigenetic alterations in cancer

Epigenetic alterations are changes in the gene function that are not related to gene DNA sequences. They include micro RNA (miRNA), DNA methylation, and histone modifications.

1.2.1 Small non-coding miRNAs

MiRNAs are small non-coding RNA molecules (around 22 nucleotides long) that are implicated in RNA silencing and post-transcriptional regulation of gene expression. MiRNAs are located mainly intracellularly, but some types are also found to be extracellular. MiRNAs interact with mRNA leading to its silencing either by mRNA cleavage, shortening of the poly A tail leading to mRNA destabilization, or by

causing altered mRNA translation into proteins [16]. Altered expression of MiRNA, either as downregulation or upregulation, is associated with cell proliferation and cancer development. Several types of miRNAs have been reported to be associated with different types of neoplasms. Down regulation of miRNA let-7 is associated with several malignancies, and restoring its normal expression is reported to inhibit tumor growth [17]. Expression of miRNA-21 is an example of the involvement of miRNA as a diagnostic and prognostic marker in a number of cancer types [18]. Other miRNAs have been demonstrated as indicators for poor survival in cancer, e.g. miRNA-324a in non-small cell lung carcinoma (NSCLC) [19]. Additionally, many studies have illustrated that miRNAs can be targets for several therapeutic agents such as antisense oligonucleotides and mRNA sponges. They all act by inhibiting the oncogenic miRNAs [20].

1.2.2 DNA methylation

DNA methylation is a process in which a methyl group is added to the cytosine base of a CpG dinucleotide by a DNA methyltransferase enzyme (DNMTs). Also, the adenine base can be methylated. Normally, DNA methylation is an essential element in a number of cellular processes such as genomic imprinting, inactivation of chromosome X, and aging. CpG islands are CpG rich sequences that are normally unmethylated and are usually located in the gene promoter regions and act as major regulatory units [21].

In cancer, CpG islands in gene promoters acquire abnormal hypermethylation causing transcriptional silencing with the possibility to be inherited by the dividing cells. Hypermethylation can result in inactivation of tumor suppressor genes causing tumorigenesis, while hypomethylation of some normally methylated genes can cause chromosomal instability [22]. Reports have demonstrated hypermethylation of tumor suppressor genes and hypomethylation of oncogenes [23].

An example that demonstrates the role of DNA methylation in the diagnosis of cancer is illustrated by Tahara et al. 2015, who showed that hypermethylation of CpG islands can distinguish between different subtypes of gastric cancer and can be used as a molecular biomarker in a variety of non-invasive samples including serum, plasma, and gastric wash [24]. Inhibitors of DNMTs can be used in the treatment of cancer as

they can increase the expression of tumor suppressor genes and decrease tumor load. So far, only two DNMT inhibitors, azacytidine and decitabine, are approved by the Food and Drug Administration (FDA) for treatment of myelodysplastic syndrome [25].

1.2.3 Histone modifications

Histone proteins are responsible for packaging nuclear DNA into the chromatin structure within the nucleosome. Histone proteins include two copies each of H2A, H2B, H3, and H4. The N terminal tails of histones are the sites where post-transcriptional modifications occur. These modifications can be methylation, acetylation, deamination, and phosphorylation. They can alter histones affecting chromatin structure and gene expression [26]. For instance; monomethylation of H3K9 and H4K20 were reported to be linked to gene activation, whereas trimethylation of H3K9 and H3K27 were found to be linked to gene repression [27]. Additionally, around 3000 genes that are highly expressed in human $CD4^+$ cells show high level of 17 histone modifications in their promotors [28]. In cancer, histone modifications are involved in tumor development such as hyperacetylation of H4 K5 and H4 K8 and hypoacetylation of H4 K12 and H4 K16, in NSCLC cells [29].

2. Lung cancer

2.1. Epidemiology and risk factors

In global terms, lung carcinoma is one of the most common malignancies worldwide and a major cause of cancer related mortalities. According to the 2018 report, lung carcinoma is the most frequently diagnosed cancer comprising approximately 11.6% of total cases, and the most common cause of cancer related mortality (18.4% of total cancer deaths) [30]. Lung cancer has a very poor prognosis and survival rate, 17.8% lower than other cancers [31]. Smoking is the major risk factor accounting for more than 85% of the of lung cancer cases. Cigarette smoke has more than 70 types of carcinogens that can cause severe DNA damage and affect the repair mechanism [32]. Many studies have also revealed the increasing frequency of lung cancer among passive smokers [33,34]. Other risk factors include asbestos exposure, air pollution, radon gas exposure, and viral infection, e.g. human papilloma virus (HPV) [35,36]. Moreover, the genetic component also contributes to the pathogenesis of lung cancer.

Genetic susceptibility to different carcinogens includes high-penetrance rare genes in familial aggregation of lung cancer, and low-penetrance common genes involved in the tobacco smoke metabolism and DNA repair pathways [37]. Polymorphisms in genes responsible for DNA repair can also contribute to lung cancer development [38]. Moreover, genetic susceptibility to lung cancer is affected by various genetic variants especially in genes related to carcinogen metabolism, DNA repair pathways, cell cycle checkpoint control, inflammatory genes, and cell microenvironment genes [39]. Genome-wide association studies have additionally identified haplotypes/SNPs at the nicotinic acid/acetylcholine receptor at 15q25 and TERT-CLPTM1L at 5p15.33 associated with increased risk of lung cancer [40].

2.2. Histopathological classification and staging

Lung cancer is classified histologically into two main categories: non-small cell lung carcinoma (NSCLC) and small cell lung carcinoma (SCLC). SCLC originates from neuroendocrine cells and often is an aggressive tumor that spreads easily to the lymph nodes and distant organs [41]. The three main subtypes of NSCLC are adenocarcinoma (ADC), squamous-cell carcinoma (SCC) and large-cell carcinoma (LCC). Adenocarcinoma is the most common type accounting for approximately 40% and it tends to occur in the peripheral pulmonary tissues. Most of the non-smokers and ex-smokers cancer cases develop adenocarcinoma [42]. Squamous-cell carcinoma accounts for about 30% of lung neoplasms and typically occurs in the center of the lung close to the large airways [43]. Large-cell carcinoma constitutes around 9% of all lung cancers [44].

The World Health Organization (WHO) updated the previous classification of NSCLC by adding immunohistochemistry (IHC) and tumor genetics to the morphological features of the tumor [45]. Five IHC markers are added for NSCLC classification including thyroid transcription factor 1 (TTF-1) and napsin-A, with a sensitivity of 80 % for ADCs for both markers. Also, the P40 marker has the best sensitivity and specificity, followed by P63 and cytokeratin 5/6 (CK5/6) for SCCs. Furthermore, adding tumor gene alterations, such as EGFR and KRAS mutations and ALK rearrangements in ADCs and FGFR1 amplification and DDR2 mutations in SCCs, to guide targeted therapy improve the prognosis and survival rate [46]. The

precise and correct molecular classification for NSCLC is extremely vital for therapeutic decisions. Examples for this include patients with SCC who reveal a poor response to pemetrexed [47], while treating with the anti-VEGF monoclonal antibody bevacizumab can cause severe bleeding[48].

NSCLC staging follows the tumor, node, metastasis (TNM) staging system in which T refers to the primary tumor size, and N refers to regional lymph node involvement, while M refers to distant metastasis. The stages are defined as Ia, Ib, IIa, IIb, IIIa, IIIb and IV. Stages from Ia to IIIb represent local tumor and locally advanced cancer with spread to lymph nodes, whereas stage IV describes tumor metastasis to other organs. Staging is a key factor during diagnosis as lung malignancies are often diagnosed at a late stage [41].

2.3. Genetic alterations and markers in lung carcinoma

Lung carcinoma development occurs due to the accumulation of pathological and molecular events known as pre-neoplastic lesions. Exposure of the cytologically normal epithelium lining the airways to smoking causes molecular alterations that predispose to the onset of cancer (Figure 2), a paradigm known as “airway field of injury” [49]. These pre-neoplastic molecular changes can provide comprehensive insight into the process of tumorigenesis and better opportunities for early detection of cancer [50].

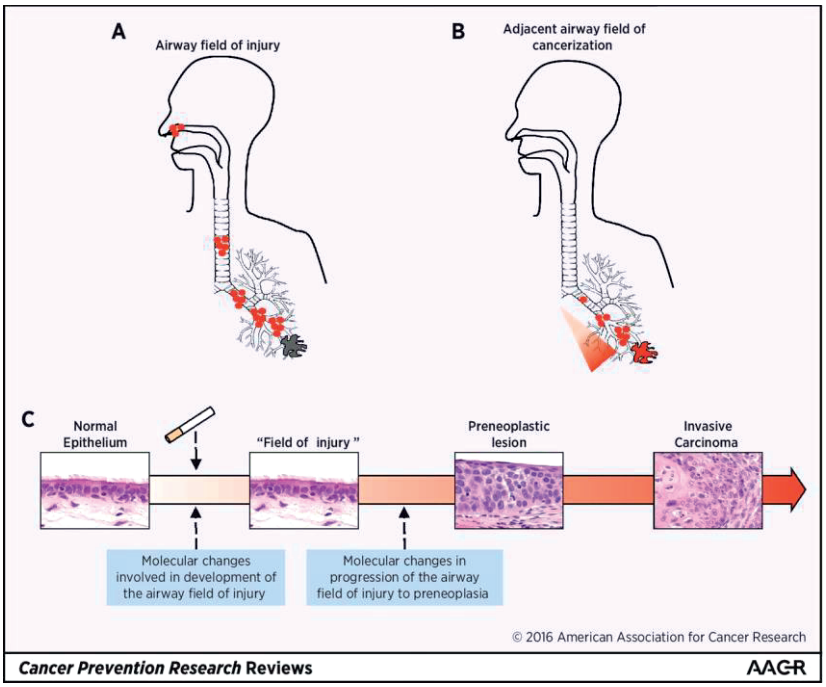

Figure 2. Airway field of injury. Figure reproduced with permission from American Association for Cancer Research. Kadara et al., 2016 [50] copyright 2016.

2.3.1. Non-small cell lung adenocarcinoma (ADCs)

Atypical adenomatous hyperplasia (AAH) is considered the only pathological alteration identified prior to development of ADCs. Similar molecular changes have been reported in both AAHs and ADCs. Earlier studies reported *KRAS* and *EGFR* mutations in AAHs and adjacent normal epithelium in ADCs, which are two major molecular pathways affected in ADCs [51,52].

The most commonly altered pathways in ADCs, according to comprehensive molecular profiling performed by the Cancer Genome Atlas, include the RTK/RAS/RAF pathway (76 %), p53 pathway (63 %), cell cycle regulators pathway (64 %), PI3K-mTOR pathway activation (25 %), and oxidative stress pathways (22 %) [53].

Whole exome sequencing data from the Cancer Genome Atlas revealed frequent somatic mutations in the following proto-oncogenes genes: *KRAS* (33%), *EGFR* (14%), *BRAF* (10%), *MET* (7%), and *PIK3CA* (7%) and the following tumor suppressor genes: *TP53* (46%), *KEAP1* (17%), *STK11* (17%), *NF1* (11%), *RB1* (4%),

and *CDKN2A* (4%) (Figure 3). Mutations in *EGFR,* occurring either in the intracellular or extracellular parts of the receptor, are among the most clinically important mutations that occurs in NSCLC as predictive markers for tyrosine kinase inhibitors. Studies reported *EGFR* mutations in the tyrosine kinase domain in approximately 25% of cases [54,55]. More than 90% of these mutations occur as a deletion in exon 19 or mutations in exon 21 (L858R). Consequently, signal transduction pathways are continuously activated leading to cell proliferation and escape from apoptosis [56].

Copy number alterations commonly seen in ADCs include amplifications of: *EGFR, CCNE1, CCND1, KRAS, MDM2, MECOM, MET, NKX2-1, TERC*, and *TERT*. Additional copy number alterations are reported in a chromosomal region of 8q24 near *MYC,* a novel peak containing *CCND3,* and deletions in *CDKN2A* [53].

Gene fusion in ADCs are of significance as drug targets, especially those involving *ALK*. The most common fusion partner of *ALK* is *EML4*, although many other genes are also reported including *ROS1*, *RET*, *PRKCB*, *MET* [57]. ALK fusion genes occur in approximately 3-7% of NSCLC especially ADCs. They tend to occur at a younger age in light and/or never smokers. In a majority of cases, ALK fusion are non-overlapping with other oncogene mutations reported in NSCLC such as *EGFR* and *KRAS* [58,59].

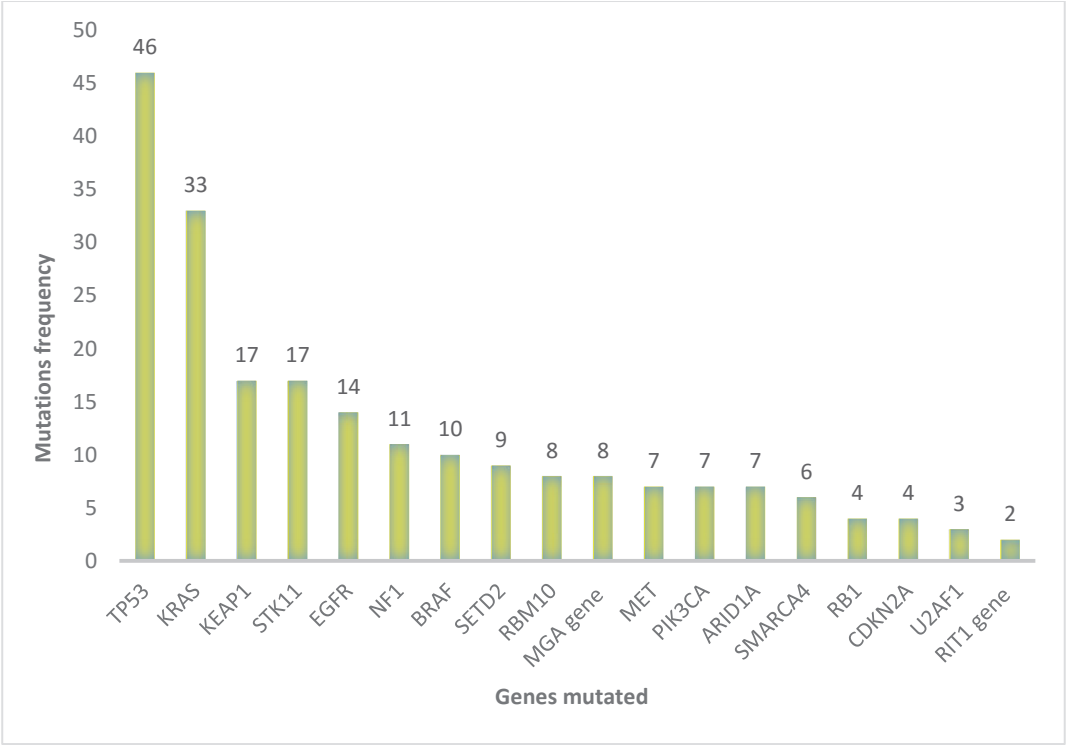

Figure 3. Somatic mutations frequently reported in ADCs. Original graph by the author, based on data from the Cancer Genome Atlas [53].

2.3.2. Non-small cell lung squamous cell carcinoma (SCCs)

Comprehensive molecular analysis performed by the Cancer Genome Atlas showed alterations in pathways related to oxidative stress *NFE2L2/KEAP1* (34%), and squamous differentiation *SOX2/p63/NOTCH1* (44%) [60]. *NOTCH1*and *NOTCH2* truncating mutations have been reported in both cutaneous and lung squamous cell carcinomas. Other affected pathways in SCCs included PI3K/AKT (47%), and CDKN2A/RB1 (72 %) [60].

Whole exome sequencing identified the most common mutated genes including: *BRAF*, *EGFR*, *HRAS,* and *PIK3CA* as proto-oncogenes, and *APC*, *CDKN2A, KEAP1, PTEN, RB1, TSC1* and *TP53* as tumor suppressor genes. Other mutated genes include *FAM123B (WTX), HRAS, FBXW7, SMARCA4, NF1, SMAD4, EGFR, APC, TSC1, BRAF, TNFAIP3* and *CREBBP*. Tumors are also characterized by a chromosomal 3q gain [60]. Fusion genes in SCCs are reported in *FGR, FGFR1, FGFR2, FGFR3, PKN1, PRKCA,* and *PRKCB* [60]. SOX2 amplification and consequent SOX2 protein

overexpression are important mechanisms of cancer initiation and progression in SCCs. Amplification of the genomic region 3q are frequent in SCCs, and SOX2 is the primary amplification target within the common 3q amplicon. It has been reported that SOX2 correlates with squamous differentiation in SCCs such as TP63 and Keratin 6A [61,62].

2.3.3. Small cell lung carcinoma (SCLC)

In SCLC, the most frequent mutated genes are *TP53* and *RB1, KIAA1211* and *COL22A1*, as well as *RGS7* and *FPR1*. Additionally, significant mutation clustering occurs in genes that have functional roles in the centrosome, *ASPM, ALMS1* and *PDE4DIP,* and the RNA-regulating gene *XRN1*. The *TP73* gene, a homologue for *TP53,* also shows clustered mutations. Mutations in *BRAF, KIT* and *PIK3CA* have a potential therapeutic role in SCLC [63].

2.4. Treatment

Lung cancer comprises two types; NSCLC (85%) and SCLC (15%). In this section we will focus on treatment options for the common type, NSCLC. There are two major treatment categories; the conventional standard protocol and the targeted therapy protocol [64]. According to National Comprehensive Cancer Network (NCCN) guidelines, the principal aim in treatment is to give optimal therapy for each patient individually and ensure the best chance of reaching long PFS and OS with limited side effects. NSCLC are categorized into subgroups which benefited from defined treatment, e.g. targeted therapies and check point inhibitors. While therapies in Stage I-III are intended to be curative, Stage IV therapy is non-curative with the intention of stabilizing and inhibiting cancer as much as possible [65].

2.4.1. Conventional treatment

Patients with NSCLC stages I, II, IIIA typically undergo surgical operations for tumor removal provided that the tumor is resectable and the patient is operable. Currently, video assisted thoracoscopic surgery (VATS) is an advanced technique that is used in many thoracic surgical operations [66].

Stage IV NSCLC patients (40% of total newly diagnosed cases) receive a combination of cytotoxic chemotherapeutic agents to improve overall survival and

reduce tumor adverse events [67]. Combinations are decided on an individual basis and are guided by frequencies of drug side effects. However, lung adenocarcinoma patients may benefit from pemetrexed [68]. Platinum-containing regimens have been established as the cornerstone of treatment for the majority of patients. Current guidelines ensure selection of patients suitable for gene targeted therapy and showing PDL-1 expression for immunotherapy. The remaining patients receive platinum-based agents. In second line therapy, combinations with check pint inhibitors can be clinically decided [69].

Radiotherapy is indicated in patients with localized tumors in the lungs. High dose or lower doses are chosen mainly based on the size of the planned radiation fields. Modern radiotherapy techniques, like stereotactic approaches are allowing higher doses, because margins toward normal tissue could be minimized. Stereotactic body radiation therapy (SBRT) can precisely locate the tumor through an advanced coordinate system and is associated with higher overall survival rate. Radiotherapy could also be given to reduce symptoms of patients, if symptoms are caused by tumor lesions, e.g. pain relief by local radiation [70].

2.4.2. Targeted therapy

Currently, different advanced molecular techniques allow further classification of lung neoplasms into subtypes according to gene mutations and alterations that can be targeted using either monoclonal antibodies (mAb) or tyrosine kinase inhibitor (TKI) molecules. Non-squamous NSCLC patients harboring *EGFR* mutations [56], *ALK* fusions [58] or *ROS1* [57] rearrangements could derive benefit from these targeted regimens with improved outcome [71].

2.4.2.1. *EGFR* and ALK inhibitors

The first TKIs developed were gefitinib and erlotinib. Both have reversible action through competitive binding with ATP for the tyrosine kinase domain leading to inhibition of the downstream pathways. Molecular analysis of tumor tissues from patients who responded to TKIs revealed activating mutations in *EGFR*. *EGFR* mutation incidence varies with ethnicity, accounting for 50% occurrence in the Asian population [71]. These findings clearly intensify the importance of molecular testing

for carcinoma patients before starting treatment. Several studies have declared the superiority of first line EGFR TKIs to chemotherapy in *EGFR*-mutated NSCLC with regard to the overall response rate (ORR), PFS and quality of life [72–74].

An alternative method targeting EGFR is by using monoclonal antibodies (mABs). Available mABs include cetuximab, necitumumab, panitumumab and matuzumab. A combination of mABs with chemotherapy has been tested in some trials (FLEX and BMS099). The FLEX trial revealed a slight increase in the median overall survival (11.3 months with cetuximab combination compared to 10.1 months with chemotherapy alone) with HR of 0.871 [95% CI 0.762-0.996]; p=0.044) [75,76].

Resistance to EGFR targeted therapy is caused by some identified mechanisms in 70% of cases, whereas the exact mechanism in 30% of resistant cases remains unknown [77]. The most common mechanism for resistance (50%) is through the concurrent mutation in exon 20 of the *EGFR* gene causing T790M. The resulting altered protein has less affinity for first-line TKIs and simultaneously increased affinity for ATP [78,79]. The second common mechanism (20%) includes MET amplification overcoming EGFR inhibition via PI3K-AKT-mTOR signaling [80]. Other mechanisms involve mutations in *BRAF, PIK3CA, HER2*, and transformation into small cell lung carcinoma [77,81,82]. The LUX-Lung1 clinical trial demonstrated that second generation TKIs such as afatinib bind covalently to EGFR/HER1 and HER2 overcoming T790M mutations with a 7% ORR and 3.3 months PFS [83]. Third generation TKIs have a significantly greater activity in *EGFR* mutant cells than in *EGFR* wild type cells, making them mutant-selective. The only approved third generation TKI is osimertinib [84].

ALK inhibitors are small molecules that target and inhibit ALK. They include three drugs approved by the FDA; crizotinib, ceritinib and alectinib [85]. Crizotinib acts as an inhibitor of ALK, MET and ROS tyrosine kinase with a 57% ORR and a PFS of 9.7 months [86]. Compared with second line chemotherapy, crizotinib showed superior outcomes with a PFS of 7.7 months versus 3.0 months with chemotherapy, causing at the same time fewer adverse effects such as mild visual and gastrointestinal disturbances [87].

Resistance to ALK inhibitors occurs due to variable mechanisms including ALK amplification, EGFR/HER1, HER2 and HER3 up-regulation, cKIT amplification and L1196M ALK mutations [88].

2.4.2.2. Angiogenesis inhibitors and immunotherapy

Cutting the tumor blood supply is recognized as an anti-cancer treatment in several malignancies [89]. In a phase III trial, a combination of bevacizumab, an anti-VEGF monoclonal antibody, with platinum chemotherapy doublets reported a median overall survival (OS) of 12.3 months when compared to treatment with chemotherapy alone, which was 10.3 months (hazard ratio for death, 0.79; 95% CI, 0.67 to 0.92; P=0.003). However, severe adverse toxic effects have been documented in the form of bleeding, thromboembolism, and hypertension [90].

The immune system is partially inhibited from attacking cancer cells due to binding of CTLA-4 and PD-1 as co-receptors on T cells with their ligands on tumor cells, CD80 and PDL-1, respectively, known as checkpoints activation. Immunotherapy involves drugs that target these checkpoints; once released from their inhibition the immune cells are able to attack and kill cancer cells [91]. Nivolumab is a PDL-1 antibody first approved by the FDA for treatment of progressed SCC after chemotherapy [92]. Other FDA approved drugs blocking the PD-1/PDL-1 interaction include pembrolizumab and atezolizumab. However, there are numerous immunotherapeutic molecules that are currently under ongoing clinical trials [93]. Ipilimumab is a CTLA-4 checkpoint inhibitor exhibiting PFS benefit in late NSCLC compared to chemotherapy and placebo in phase II clinical trials [94].

3. Gastrointestinal neoplasms

3.1 Epidemiology and risk factors

3.1.1. Epidemiology and risk factors of gastric carcinoma

Gastric cancer (GC) is the third cause of cancer related mortalities worldwide [95]. Although the incidence has decreased in the last decades, it still accounts for about 989,600 new cases and 738,000 deaths annually [96]. Moreover, the incidence varies widely between different countries with the highest incidence in East Asia, Eastern Europe and South America, whereas, North America and most of Africa have a lower

incidence [97]. Risk factors include Helicobacter pylori (*H. pylori*) infection, dietary habits, smoking, obesity and others. It is well known that *H. pylori* is a major cause of GC development specially strains with a cytotoxin associated gene A [98]. Better hygiene and wider use of antibiotics caused a reduction in *H. pylori* prevalence leading to an obvious decline in GC incidence [99]. Dietary habits including higher consumption of salty and smoked food, and lower intake of fresh fruits and vegetables have a potential carcinogenic effect on the gastric epithelium. They might cause mucosal damage and increase the possibility of *H. pylori* infection [100]. Additionally, obesity predisposes to GC especially the adenocarcinoma type. By predisposing to gastroesophageal (GE) reflux leading to Barrett's esophagus, obesity along with GE and Barrett's esophagus are risk factors for gastric adenocarcinoma [101]. A significant relationship between higher body mass index and increased incidence of GE has been illustrated [102].

3.1.2. Epidemiology and risk factors of colorectal carcinoma

Colorectal carcinoma (CRC) is a principal cause of morbidity and mortality throughout the world. In western countries, the risk of incidence in a lifetime is around 5% [103]. Risk factors include non-modifiable risk factors such as age, family history, inherited genetic risk and modifiable environmental risk factors such as dietary habits, obesity, physical activity, smoking and alcohol consumption. About 95% of sporadic CRC arise from either tubular or villous types of colorectal adenomas[104]. Patients with inflammatory diseases have a 20 fold higher risk of developing CRC [105]. The most common inherited CRC conditions are familial adenomatous polyposis (FAP), which is caused by mutations in the *APC* gene, and hereditary non-polyposis colorectal cancer (Lynch syndrome), which is caused by mutations in DNA repair genes; *MLH1 and MSH2* [104]. Diet including a higher fat content, especially animal fat, is a major risk factor for CRC, as it may favor growth of certain bacteria which produce carcinogenic N-nitroso compounds [106]. Cigarette smoking promotes development and faster growth of adenomatous polyps predisposing to malignant transformation [107].

3.2. Histopathological classification and staging

3.2.1. Histopathological classification and staging of gastric carcinoma

There are several histopathological classifications for gastric carcinoma. The Lauren classification is the most commonly used system for classification. It includes the intestinal type, which is the most common type, followed by diffuse and intermediate types [108,109]. Intestinal metaplasia and *H. pylori* infection are common associations with the intestinal type with increased chances of vascular and lymphatic invasions. However, in the diffuse type, there is minimal cell to cell adhesions with mucosal infiltration occurring in the form of single cells or small subgroups. Signet ring cell carcinoma belongs to this type [110]. The World Health Organization (WHO) provides another common system for classification. In addition to gastric adenocarcinoma, the WHO classification includes all other less frequent types of gastric tumors, e.g. adenosquamous carcinoma. Gastric adenocarcinoma is divided into tubular (the most common), papillary, mucinous and mixed types [111]. Other classifications include Goseki, Ming, and Grundmann systems [112–114].

3.2.2. Histopathological classification and staging of colorectal neoplasms

The World Health Organization (WHO) classified colorectal neoplasms into epithelial and non-epithelial (mesenchymal) tumors. Epithelial adenomas include tubular, villous, tubulovillous and serrated, while epithelial carcinomas include adenocarcinoma, mucinous adenocarcinoma, signet-ring cell carcinoma, small cell carcinoma, squamous cell carcinoma and others. The non-epithelial tumors include Lipoma, Leiomyoma, Gastrointestinal stromal tumor, malignant lymphomas and others [115]. The American Joint Committee on Cancer (AJCC) has issued its seventh edition of TNM staging for CRC. Stage I is T1N0M0 and T2N0M0. Stage IIA is T3N0M0, IIB is T4AN0M0, and IIC is T4bN0M0. Stage IIIA is T1N1/1cM0, T2N1/1cM0 and T1N2aM0. Stage IIIB is T3N1M0, T4bN1M0, T1N2bM0, T2N2a-bM0 and T3N2aM0. Stage IIIC is T4aN2aM0, T3N2bM0, T4aN2bM0, T4bN2M0 and T4bN1M0. Stage IVA is any T any N M1a, and stage IVB is any T any N M1b [116]. Other staging systems include Dukes and Astler-Coller systems [117].

3.3. Genetic alterations and markers in gastrointestinal neoplasms

3.3.1. Genetic alterations and markers in gastric carcinoma

The Cancer Genome Atlas Network illustrated that gastric carcinoma exhibits a variety of molecular subtypes, such as microsatellite instability (MSI), and different histological phenotypes, such as gland forming intestinal type and highly infiltrating isolated cells in the diffuse type. The most frequently mutated genes are *TP53* in microsatellite stable (MSS) tumors, *ARID1A* in MSI tumors, Epstein–Barr virus (EBV) tumors and *CDH1* in diffuse type tumors [118]. Other genes are *MUC6* in 9.6% of MSS and in 18.2% of MSI tumors, *RNF43* in 4.8% of MSS and 54.6% of MSI tumors. Other genes mutated in MSS tumors include *CTNNA2* (6.4%), *GLI3* (6.9%), *ZIC4* (4.8%), *TGFBR2* (4.8%), *ACVR2A* (2.1%), *SMAD4* (4.3%), *ELF3* (3.7%), *DCLK1* (4.3%), and *THBS1* (4.8%). Well known genes, including *CTNNB1, TET1, TSC1, FBXW7* and *ATM,* are mutated at lower frequencies [118]. The *RHOA* gene was mutated in 14.3% of diffuse-type tumors and in 7.8% of indeterminate-type tumors, with no *RHOA* mutations reported in the intestinal-type. Additionally, in intestinal type tumors, there are frequent copy number gains on chromosomes 1q, 5p, 7, 8, 13 and 20 and frequent losses on chromosomes 1p, 3p, 4, 5q, 9p, 17p, 18q, 19p, 21 and 22, whereas, in MSI tumors, a gain of chromosome 8 is frequent [118]. All tumor subtypes revealed increased expression of AURKA/B and E2F, targets of MYC activation, FOXM1 and PLK1 signaling and DNA damage response pathways but to a lesser extent in genomically stable types. However, genomically stable types revealed increased expression of cell adhesion pathways, including the B1/B3 integrins, syndecan-1 mediated signaling, and angiogenesis-related pathways [118].

3.3.2. Genetic alterations and markers in colorectal carcinoma

CRC is divided into tumors with microsatellite instability (MSI), which are frequently associated with the CpG island methylator phenotype (CIMP), and tumors that are microsatellite stable but chromosomally unstable. A comprehensive molecular analysis done by The Cancer Genome Atlas Network revealed that the non-hypermutated tumor types exhibit frequent mutations in *APC, CTNNB1, FAM123B, FBXW7, KRAS, NRAS, PIK3CA, SMAD4, SMAD2, SOX9, TCF7L2* and *TP53.* Mutations in *KRAS* and *NRAS* occur at codons 12, 13, and 61. Mutations in tumor

suppressor genes *ATM* and *ARID1A* were also frequent. In the hypermutated tumor type, *ACVR2A*, *APC*, *BRAF* (V600E), *MSH3*, *MSH6*, *SLC9A9*, *TCF7L2, TGFBR2*, and *TP53* were frequent targets of mutation [119]. The hypermutated tumors show gains of 1q, 7p and q, 8p and q, 12q, 13q, 19q, and 20p and q, and chromosome arm deletions at 18p and q (including *SMAD4*), 17p and q (including *TP53*), 1p, 4q, 5q, 8p, 14q, 15q, 20p and 22q. Frequent translocations include *NAV2* - *TCF7L1* and *VTI1A* - *TCF7L2* fusion genes [119]. WNT pathway signaling is altered in 93% of all tumors, including inactivation of *APC* or activating mutations of *CTNNB1* in ~80% of cases. Genetic alterations in the PI3K and RAS–MAPK pathways are also common in CRC with mutations in *BRAF, KRAS, NRAS, PIK3R1* and *PIK3CA* as well as deletions in *PTEN*. The TGF-β signaling pathway is affected with involvement of genomic alterations in *TGFBR1*, *TGFBR2*, *ACVR2A*, *ACVR1B*, *SMAD2*, *SMAD3* and *SMAD4*. Moreover, the p53 pathway is altered including mutations in *TP53* and *ATM*. MYC transcriptional targets have been found to be altered in nearly 100% of all tumor types and have an important role in CRC [119].

3.4. Treatment

3.4.1. Treatment of gastric cancer

3.4.1.1. Conventional treatment of gastric cancer

Comprehensive treatment planning is crucial before taking the final clinical decision by surgeons, medical and radiation oncologists, radiologists and pathologists. It depends on whether the patient is operable with a regional tumor or inoperable with a huge metastatic tumor.

Surgical resection is curative especially in early stages. The extent of resection depends on the tumor stage and varies from endoscopic resection to radical gastrectomy with or without lymph nodes dissection. Laparoscopic surgery has advantages of decreased postoperative complications and recovery time [120].

Patients with resectable tumors stage ≥IB receive the cytotoxic agents platinum/fluoropyrimidine combination pre and post-operatively [121]. Similar combinations are used in advanced and metastatic tumors in the first line of therapy. Additionally, a second line chemotherapy , e.g. with a taxane, is given in advanced and metastatic tumors [122].

Adjuvant radiotherapy with 5-FU/leucovorin (Q28) plus conventionally fractionated RT resulted in 50% 3-year survival compared to 41% for patients treated with surgery alone [123].

3.4.1.2. Targeted therapy

3.4.1.2.1. HER2 inhibitors

Trastuzumab is the first molecular targeted drug that was used in the treatment of gastric carcinoma. It inhibits HER2 mediated signaling and prevents release of its extracellular domain. The Trastuzumab for Gastric Cancer (ToGA) clinical trial revealed a significantly better OS in patients receiving a combination of chemotherapy and trastuzumab. However, the FDA has restricted trastuzumab therapy for patients with HER2 overexpression [124]. Alternative approaches that target the HER2 receptor through monoclonal antibodies include pertuzumab, which binds to the extracellular domain preventing HER2 dimerization [125].

Secondary resistance to trastuzumab can develop due to molecular alterations. The PI3K/Akt/mTOR pathway is one of the factors that causes dysregulation of the HER2 downstream signal. The mTOR inhibitors such as everolimus inhibits the mTOR/S6K signal thus improving fluorouracil-induced apoptosis in gastric cancer cells with HER2 amplification [126]. Another agent, afatinib, is an irreversible inhibitor of EGFR, HER2, and HER4. Afatinib can be effective against receptors with secondary mutations resistant to trastuzumab.

3.4.1.2.2. Angiogenesis inhibitors

A monoclonal antibody, bevacizumab, acts by blocking the binding of VEGF to its receptors. A double-blind, phase III trial (REGARD) recruited 355 progressive gastric cancer patients, and investigated the combination of ramucirumab and chemotherapy, leading to a small but statistically significant prolonged median OS (3.8 to 5.2 month, $p = 0.0473$) [127]. Another trial (RAINBOW), which investigated ramucirumab as a second-line treatment in patients with advanced gastric cancer and disease progression after first-line chemotherapy, showed a significantly better OS in the ramucirumab plus chemotherapy group (median 9.6 vs. 7.4 month, $p = 0.017$) [128]. Finally, the FDA has approved ramucirumab in April 2014.

3.4.2. Treatment of colorectal cancer

3.4.2.1. Conventional treatment of colorectal cancer

Surgical removal of small and local tumors is usually a curative treatment. Endoscopic removal of precancerous polyps can be done and is called polypectomy. Operations vary from right, transverse or left hemicolectomy, subtotal colectomy, or total colectomy. Comprehensive geriatric assessment (CGA) is performed to assess the risk of post-operative complications [129].

Patients with stage II and III colorectal tumors are usual candidates for adjuvant chemotherapy. Addition of fluoropyrimidine or fluoropyrimidine plus oxaliplatin are the current standard of care, based on findings from the Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) and the National Surgical Adjuvant Breast and Bowel Project (NSAFDA) C-07 trials. Currently, numerous trials are evaluating oral fluoropyrimidines combined with oxaliplatin and the addition of targeted drugs to oxaliplatin-based regimens for use in colon cancer adjuvant treatment [130].

Short-course radiotherapy (SC-RT) and long-course chemoradiotherapy (LC-CRT) are recommended as preoperative radiotherapy because they can reduce the tumor size especially (LC-CRT) and the relapse risk postoperatively. For stage II/III rectal cancer, neoadjuvant radiotherapy shows superiority, and stereotactic body radiotherapy (SBRT) of the liver shows better local control in oligometastatic patients [131].

3.4.2.2. Targeted therapy

3.4.2.2.1. *EGFR* inhibitors

Cetuximab and panitumumab are anti-EGFR monoclonal antibodies and one of the first which are used in CRC targeted therapy. They act by binding to the EGFR receptor and blocking the intracellular signaling cascade, thus stopping cellular proliferation and growth. Cetuximab was FDA approved in 2004 as a combination with irinotecan for irinotecan-refractory CRC patients. However, experimental studies revealed that an exon 2 activating mutation of *KRAS* caused resistance to cetuximab, therefore, cetuximab was used only in CRC patients with wild type *KRAS* [132,133]. A second line treatment including cetuximab plus best supportive care

(BSC) was performed by the NCIC CO.17 trial with improvement of OS and PFS and preserved quality of life measures observed by adding cetuximab [134]. Panitumumab was tested as a first line treatment in combination with the FOLFOX-4 regimen in the phase III PRIME clinical trial. This combination resulted in significant improvement in PFS compared with the FOLFOX-4 alone (9.6 months vs 8.0 months; HR = 0.80, P = 0.002). There was no significant difference in terms of OS and RR (OS 23.9 months vs 19.7 months, HR = 0.83, P = 0.072; RR 55% vs 48%, OR = 1.35, P = 0.068) [135].

3.4.2.2.2. Angiogenesis inhibitors

Bevacizumab is a monoclonal antibody that selectively binds to VEGF-A and demonstrates anti-tumor activity by blocking VEGFR2. It had FDA approval in 2004 for treatment of CRC in combination with other cytotoxic agents. In the AVF2107g trial, bevacizumab was added to either IFL (irinotecan, fluorouracil, and leucovorin) and demonstrated significant improvements in overall survival [OS, 20.3 months *vs* 15.6 months, hazard ratio (HR) = 0.66, $P < 0.001$], progression-free survival (PFS, 10.6 months *vs* 6.2 months, HR = 0.54, $P < 0.001$), and RR (44.8% *vs* 34.8%, $P = 0.004$) [136].

Aflibercept is another anti-angiogenic agent which can bind to VEGF-A, VEGF-B, and PIGF preventing these ligands from binding to their receptors and inhibiting the VEGF pathway [137]. Regorafenib is an oral multi-kinase inhibitor which blocks the activity of several protein kinases related to the angiogenic pathway (VEGFR-1, VEGFR-2, VEGFR-3, TIE-2), the oncogenic pathway (KIT, RET, RAF1, BRAF), and the tumor microenvironment (PDGFR and FGFR). However, grade 3 side effects such as hand-foot skin reaction, fatigue, diarrhea, hypertension, and rash or desquamation were reported [138].

4. Gut microbiota in gastrointestinal neoplasms

The gut microbiota comprises approximately 3×10^{13} bacterial cells that colonize the human gut as commensals living in a balanced state with the host known as eubiosis. The gut homeostasis is continuously maintained by crosstalk between gut microbiota, immune cells and mucosal barriers. Disruption of this host/microbiota relationship

(dysbiosis) is often associated with different diseases, including cancer, by affecting oncogenesis, tumor progression and response to cancer therapy (Figure 4) [139,140]. Gastric carcinoma is an example of microbiota caused oncogenesis. Infection with *H. pylori* may cause a sequence of gastritis, ulcer, atrophy and finally cancer. *H. pylori* was classified as a carcinogen, however, it is associated with a lower risk of esophageal cancer, which in a way clarifies the effect of organ specificity in microbiota-induced oncogenesis [141–143]. A microbiota tumor promoting effect is clearly evident in CRC. A dysbiosis effect caused by long-term treatment with broad spectrum antibiotics and germ free status is remarkable leading to alteration in the host-microbiota relationship [144–147]. Moreover, numerous by-products from gut microbiota can target intestinal epithelial cells and cause either a tumorigenic effect, e.g. *Bacteroides fragilis* toxin, or a tumor suppressive effect, e.g. short-chain fatty acids [148].

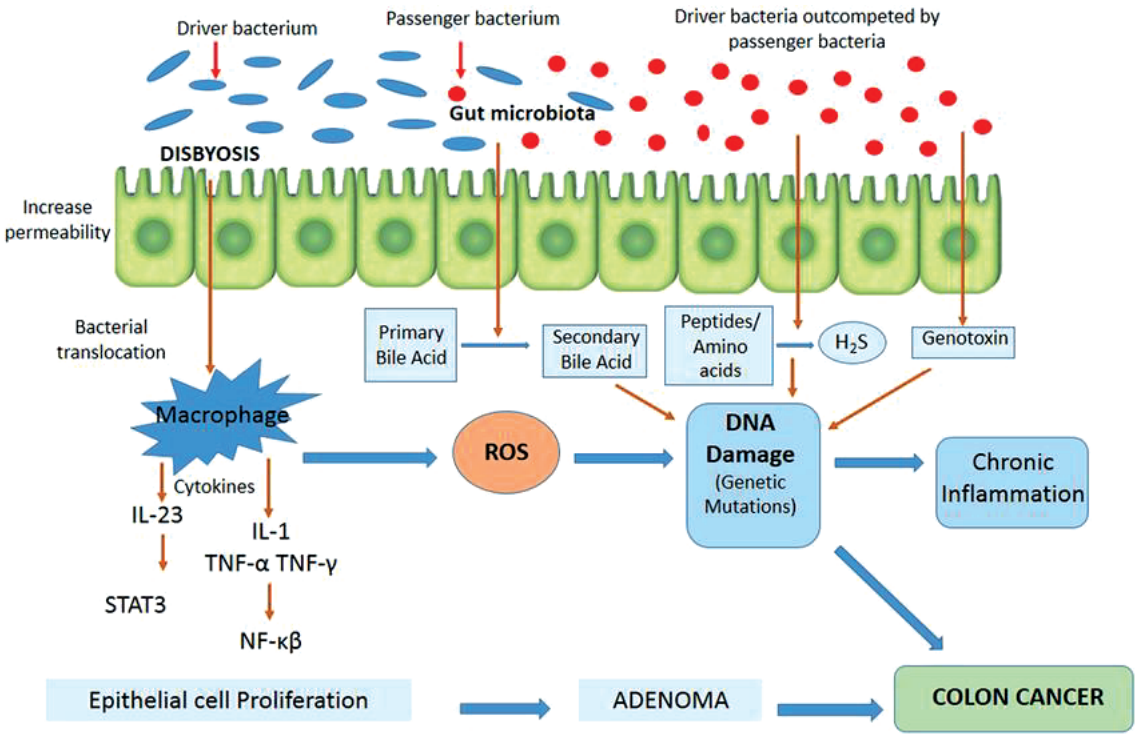

Figure 4. Contribution of gut microbiota to colorectal carcinogenesis. Nistal et al., 2015 [149], reprinted under the terms of the Creative Commons Attribution License (CC BY), 2015.

A principal element that maintains the balance between host and microbiota is the presence of well-established multilevel barriers. Disruption of these barriers causes inflammation and various diseases including cancer. Examples of these barriers include intact epithelial and mucosal lining, low PH (in stomach), special cell types such as goblet cells and gut associated lymphoid tissue (GALT) [150]. Studies revealed variations at the different taxonomic levels in the normal gut microbiota, and changes in the diet, immune system alterations, and infections can affect microbial richness, composition and metagenome [151]. Several mechanisms can explain the involvement of microbiota in carcinogenesis. Regulation of microbiota by the immune system can favor the growth of certain bacterial types. The microbiota are labeled with pattern recognition receptors (PRRs), which are recognized by toll like receptors (TLRs) on immune cells and can elicit a strong pro-inflammatory state [152]. Tumor cells express TLRs. Signaling pathways of TLRs, such as the myeloid differentiation primary response 88 (MYD88), usually exhibit multiple effects that alter tumor cells [153]. TLR4, which is the receptor for lipopolysaccharide on the cell wall of gram negative bacteria, can promote tumorigenesis in colon, liver and pancreas as evidenced by increased tumor load in mice expressing activated epithelial derived TLR4 [154–156]. TLR2 is a receptor for peptidoglycan in the bacterial cell wall and is shown to promote gastric cancer [157].

Another mechanism by which microbiota can contribute to carcinogenesis is through generation of metabolic activities that may affect carcinogenesis by regulating obesity and obesity-induced inflammation, metabolic activation and activation of carcinogens. Bile acid metabolism is regulated by gut microbiota enabling microorganisms to use secondary bile acids as an energy source. Recently, it was shown that a high-fat diet alters the gut microbiota and increases the levels of the secondary bile acid (produced by bacterial dehydroxylation), which could promote liver and colon carcinomas [158,159]. Carbohydrate fermentation by gut microbiota produces beneficial short chain fatty acids such as butyrate, which have a vital role in the control of inflammation and autophagy and consequently a protecting effect from cancer [160–162]. In contrast, protein metabolism produces toxic cancer promoting metabolites such as ammonia and nitrosamines. Protein fermentation

mainly occurs in the distal colon, and diets rich in protein and low in carbohydrate may change intestinal fermentation and can lead to increased levels of hazardous metabolites [163–165]. Furthermore, microbiota metabolites can be altered by inflammation, e.g. nitrate. Nitrate provides energy for facultative anaerobes such as Enterobacteriaceae, supporting them to thrive within a community dominated by obligate anaerobic bacteria that lack the proper electron transport chain to use nitrate as shown by the prevalence of Enterobacteriaceae in numerous inflammatory disease models and in patients with chronic inflammation [166–168].

Large-scale studies in an organ- and cancer-specific manner including metagenomic, metatranscriptomic and metabolomic analysis from large cohorts of patients and healthy controls are crucial for a better understanding of whether changes in microbial composition or richness, especially at the metagenomic level, affect cancer development, progression and treatment [150]. Assessment of human cancer microbiomes in preclinical sessions would help to assess the tumorigenic potential of the cancer-associated microbiota. In this thesis, we studied the differential microbiota taxonomic composition in a cohort of stool specimens from different GIT neoplasms including stomach, pancreas, small intestine, colon and rectum.

5. Non-invasive samples as cancer biomarkers

A diagnosis of cancer often occurs at late stages due to a lack of symptoms or the presence of vague non-specific symptoms in most cases. Moreover, when malignancy is suspected, the classical method to establish a definite diagnosis is to take a biopsy from the tumor tissues, which is not always available during the early stages of the disease or during treatment. Therefore, establishing a novel non-invasive method that could be used for early detection of malignancy or for cancer screening has been the main concern of several researchers over the past few decades leading to the appearance of the “liquid biopsy” concept [169–171].

The continuous process of the rapidly proliferating malignant cells causes exfoliation and shedding of tumor cells into the blood circulation and epithelial cavities. In malignancy, removal of these cells by macrophages is impaired leading to the increased amount of the tumor cells and their availability for detection in different kinds of body fluid samples. The exfoliated tumor cells can be a source for detection of different genetic and epigenetic alterations found in the original tumor biopsies [170].

5.1. Sources of non-invasive samples

There are several different sample materials that can serve as non-invasive samples. Blood and its derivatives such as plasma and serum represent the earliest liquid biopsies that have been investigated. Two types of cancer derived materials can be detected in blood, including circulating tumor cells (CTCs) and cell free circulating tumor DNA (ctDNA) [172]. It has been reported that release of ctDNA into the blood stream is affected by multiple factors such as tumor type, location, size, and vascularity [173]. Also, it has been found that the amount of circulating DNA in the plasma of serum of cancer patients is higher than in healthy individuals [174]. They have been studied in a variety of solid tumors, such as colon, breast and lung malignancies [175–177], and reported even to predict tumor location [178].

Another important source for liquid biopsies is urine. Patients with renal cell carcinoma, bladder cancer, and prostate neoplasms show detectable cell free DNA in approximately 70% of urine samples [179]. Additionally, mitochondrial DNA (mtDNA) mutations have been detected in urine from bladder cancer patients and to a lower extent from patients with prostate cancer [180]. In patients with NSCLC, urine has been tested for *EGFR* mutation analysis [181].

Like blood, sputum samples contain both cells that can be studied through cytological and molecular examinations, and cell free DNA that can be investigated by molecular approaches. The relatively low amount of tumor DNA and target cells in sputum impose particular requirements for sputum molecular analysis including the use of internal and external positive and negative controls [182]. Both *KRAS* mutations and promoter methylation have been reported relatively frequently in sputum while

reports on loss of heterozygosity (LOH) and RNA are far fewer to date [183,184]. In cases of non-small cell lung carcinoma (NSCLC) and mesothelioma, sputum samples have been investigated to search for tumor markers and the genetic and epigenetic changes that are commonly reported in the tumor tissues [185].

Saliva contains both human and microbial DNA. It also gives a very good yield of DNA regarding quantity and quality [186]. Cancer biomarkers in saliva have been reported in pancreatic, breast, lung, and oral malignancies [187–190].

5.2. Advantages and drawbacks of non-invasive samples

Non-invasive samples offer several advantages over conventional tissue specimens. In addition to being a patient friendly approach and easy to collect, they are easy to repeat; providing the opportunity to take serial samples and follow a tumor situation continuously and dynamically. Moreover, non-invasive samples can be used for screening of different neoplasms especially in high-risk individuals. They can help in diagnosis of neoplasms at an early or premalignant stage and act as a guide in the targeted therapy regimen.Thus development of resistance could be followed, e.g. *EGFR* resistance and appearance of T790M which then can be treated by a specific drug, osimertinib. Additionally, many of the tumor biomarkers that have been frequently detected in liquid samples can be used as predictive factors, as the liquid samples might represent tumor heterogeneity better than a tissue sample biopsy [190].

Limitations of liquid biopsies are associated with the components of liquid samples. Firstly, the ctDNA is a fraction of total cell free DNA that originates primarily from normal tissues and therefore is present in low concentrations, which may cause false negative results. Secondly, circulating tumor cells (CTCs) have complex heterogeneity in morphology and number as they undergo the process of epithelial to mesenchymal transition. Also, the percentage of CTCs detected varies between different tumors. Moreover, there is considerable molecular heterogeneity and it is difficult to specify the origin of circulating DNA [191].

5.3. Clinical applications of non-invasive samples

Being non-invasive, one of the principal applications of non-invasive samples is to be used as cancer biomarkers. They can be used for primary diagnosis of tumors at an early stage, which consequently results in a significant improvement in patients' survival and outcome [192]. Liquid biopsies can also monitor progression of the tumors as frequencies of genetic mutations in ctDNA are positively correlated with tumor grade, stage, and evolution [192,193]. The methylation pattern of a CpG in urine DNA samples is reported to diagnose prostate cancer with 94.6% sensitivity and 78.3% specificity [194]. One more essential application of liquid biopsies is in monitoring treatment effect and predicting drug resistance. It has been reported that lung cancer patients with an *EGFR* C797S mutation in ctDNA develop resistance to the drug AZD9291 [195]. Additionally, *PIK3CA* mutations in ctDNA can be used to detect minimal residual disease (MRD) in breast cancer patients after chemotherapy or surgery [196].

5.4. Diagnostic role of exhaled breath condensate (EBC) in lung cancer

One of the major challenges in the diagnosis of lung carcinoma is that symptoms are usually vague and might appear only in late stages of the disease. Although, chest radiographic examinations, such as computed tomography (CT) is the primary diagnostic approach, eventually a biopsy is taken for definitive diagnosis [197]. Genetic alterations occurring during the development of lung carcinoma and detected in tumor tissues is an essential element in molecular classification and identification of tumor type and subtype. Exhaled breath condensate contains cells that are exfoliated from the respiratory epithelium or nucleic acids that are shed from pulmonary cells into the airway lumen. Thus, EBC DNA could directly reflect the molecular and genetic alterations occurring in lung tumor tissues [198]. Somatic mutations in the *TP53* and *EGFR* genes are reported in EBC from NSCLC patients [199,200]. A large cohort study demonstrated the existence of *KRAS* mutations in EBC from NSCLC patients and the follow up of the patients revealed the disappearance of *KRAS* mutations from some patients after successful surgery [201]. Mutations in the mitochondrial D-loop have been detected with higher prevalence in EBC from lung cancer patients compared to healthy controls [202]. Furthermore,

miRNA-21 is reported to be upregulated while miRNA-486 is downregulated in the EBC of NSCLC patients [203]. Microsatellite alterations, gene promoter methylation, and human papilloma virus DNA (HPV) have all been detected in EBC from lung cancer patients [198]. It is obvious that EBC could have several applications in the lung cancer field being advantageous over other conventional methods in screening of high risk patients or monitoring patients after treatment. Although EBC is a non-invasive specimen, further studies are needed to assess its diagnostic applicability in pulmonary neoplasms [198].

5.5. Diagnostic role of stool specimens in gastrointestinal neoplasms

Gastrointestinal tract (GIT) neoplasms consist of a diverse group of tumors that have different pathological and molecular events. They usually begin with benign growth of cells which often proceed to malignancy without being detected in the benign phase. The classical method for diagnosis of GIT neoplasms is endoscopy with its common sequelae being an invasive procedure. Several attempts have been made over the last decades to develop a non-invasive tool for early detection of GIT neoplasms. Stool contains a mixture of cells including leucocytes, blood cells, inflammatory cells, as well as tumor cells. [204]. The process of tumor cell shedding from the neoplastic epithelium into feces in the intestinal lumen is continuously occurring and to a much greater extent (qualitative and quantitative) than under normal conditions. Furthermore, the cell free DNA level is elevated in stool of cancer patients [205]. These exfoliated tumor cells are believed to reflect the ongoing carcinogenesis, thus revealing several genetic alterations associated with malignancy. These observations encouraged the translational research community to apply these tests for cancer diagnostics. In 2014, the FDA approved the multi-target test Cologuard for screening and early detection of CRC from stool samples [206]. Gene mutations in stool from premalignant colorectal adenomas or early stage malignant GIT neoplasms have been detected by applying next generation sequencing (NGS), which opens possibilities for non-invasive diagnosis and screening [207]. The optimum clinical application of stool-based assays in GIT tumors need more investigation including different types and subtypes of GIT neoplasms.

AIM OF THE STUDY

The general aim of the study was to determine whether non-invasive specimens, stool in GIT tumors and EBC in lung carcinoma, offer a reliable source for detection of driver gene mutations that are commonly found in tumor tissue by application of NGS, and hence, can be used for early diagnosis of cancer.

Specific aims were:

- To assess the genetic alterations that are present in EBC from healthy individuals before application of EBC on lung cancer patients (Study I).
- To study tumor associated hotspot mutations in EBC samples from lung cancer patients and compare them with the detected mutations in the healthy controls (Study II).
- To investigate the detection of gene mutations in stool from patients with gastric or colorectal neoplasms, including benign and malignant tumors (Study III).
- To investigate the differences in taxonomic composition in stool samples from GIT neoplasia patients (stomach, pancreas, small intestine, colon, and rectum) based on the tumor location (Study IV).

MATERIALS AND METHODS

Sample features and analytical methods used in this thesis are described generally in the following section. Detailed information can be obtained from the original publications. Table 1 provides an overview of samples and methods used in studies I-IV.

1. Study samples

1.1. Exhaled breath condensate samples (I, II)

Exhaled breath condensate (EBC) samples were collected from 20 healthy adult subjects with an average age of 34.9 years in study I. The samples were collected in the Helsinki University Hospital. Two different samples were taken from one individual with a one-month interval. The smoking status of the healthy subjects was categorized into never smokers (n= 15), ex-smoker (n= 4), and current smoker (n= 1). In study II, EBC samples were collected from 26 patients, average age of 67.5 years, with different lung malignancies. Lung neoplasms included non-small cell lung carcinoma (NSCLC) patients (n=17), small cell lung carcinoma (SCLC) patients (n= 6), 2 patients with mesothelioma, and 1 patient suspected for malignancy. In studies I and II, EBC samples were collected by breathing into the EcoScreen instrument (Jaeger/Germany) for 15 minutes, while recording breathing frequency and mean breath volumes every 5 minutes, then finally storing the collected EBC at -70^{o}C (Figure 5).

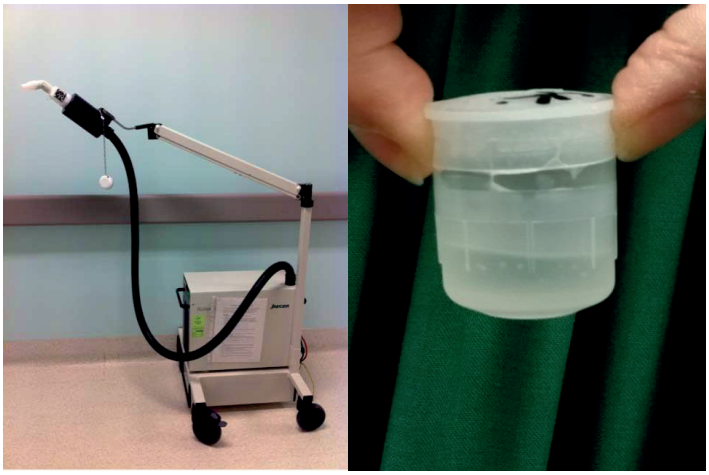

Figure 5. EcoScreen instrument (Jaeger/Germany) for EBC collection and the cup containing the exhaled breath condensate (EBC) sample (Original photo by the author).

1.2. Stool samples from GIT neoplasms (III, IV)

For study III, stool samples were collected from 87 patients with gastric (n= 41) and colorectal neoplasms (n= 46). Stool samples from 14 healthy adults were also collected. The samples were collected in the Helsinki University Hospital. A total of 21 patients received treatment in the form of chemotherapy, radiotherapy or antibiotics before sample collection. In study IV, stool samples from 63 GIT neoplasia patients (classified as gastric, pancreas, small intestine, colon and rectum) were collected before any kind of treatment, whereas samples from 20 patients were obtained after starting treatment, either chemotherapy and/or radiotherapy. Stool samples from 13 controls were also included in the study. In both studies, stool samples were collected in special tubes provided in the PSP Spin Stool DNA Plus Kit (Stratec Biomedical, Berlin, Germany). These tubes contain stool DNA stabilizer to allow collection, transport and storage of the samples without DNA degradation, and can be stored at -20^{o}C till further analysis.

Table 1. Samples and methods used in studies (I-IV)

	Study I	**Study II**	**Study III**	**Study IV**
Analytical group	Healthy individuals	Lung cancer patients	GIT neoplasia patients	GIT neoplasia patients
Sample material	EBC	EBC	Stool	Stool
Number of subjects	20 Healthy subjects	26 patients	87 patients	83 patients
Average of age (years)	34.9	67.6	69.7	69.6
Female **Male**	F=10 M=10	F=12 M=14	F=40 M=47	F=38 M=45
NGS gene panel used	Ion Ampliseq Colon and Lung Cancer panel v2 (22 genes)	Ion Ampliseq Colon and Lung Cancer panel v2 (22 genes)	Ion Ampliseq Colon and Lung Cancer panel v2 (22 genes) and Ion AmpliSeq Cancer Hotspot Panel v2 (50 genes)	Ion 16S Metagenomics kit (16S rRNA gene)

2. Ethical permissions

Studies on EBC samples (I, II) were approved by the Hospital and Uusimaa (HUS) review board (Ethical permission number 253/13/03/01/2015), while studies on stool samples (III, IV) were approved by the Hospital District of Helsinki and Uusimaa (HUS) review board (Ethical permission number 351/13/03/02/2014). Written informed consent was obtained from all participating subjects (patients and controls).

3. DNA extraction

In studies I and II, DNA was extracted from the whole EBC sample with the QIAamp circulating nucleic acid kit (Qiagen Cat No./ID 55114) according to the manufacturer's instructions using a vacuum pump. The kit has an RNA carrier for proper isolation of the DNA.

In studies III and IV, stool samples were well mixed with the stabilizer for proper homogenization, then a volume of 1.4mL of stabilized stool sample was extracted using the PSP® Spin Stool DNA Plus Kit (Stratec Biomedical) according to the manufacturer's instructions.

4. Next generation sequencing (NGS)

4.1. Targeted NGS

In studies I and II, approximately 10ng of EBC DNA was used for the preparation of libraries with an Ion AmpliSeq™ Library kit 2.0 (Thermo Fisher Scientific) and the Ion Ampliseq Colon and Lung Cancer panel v2 (Life Technologies, California, United States) was used. It consists of a primer pool for 92 amplicons from 504 hotspot regions in 22 genes frequently mutated in lung cancer. The genes included in this panel are *AKT1, ALK, BRAF, CTNNB1, DDR2, EGFR, ERBB2, ERBB4, FBX7, FGFR1, FGFR2, FGFR3, KRAS, MAP2K1, MET, NOTCH1, NRAS, PIK3CA, PTEN, SMAD4, STK11,* and *TP53*.

In study III, 20ng of stool DNA was used to prepare the libraries by using two different gene panels; the Ion AmpliSeq Cancer Hotspot Panel v2 (Life Technologies, California, United States) consisting of a primer pool for 207 amplicons from an average of 2800 mutational hotspot regions in 50 genes, including *ABL1, AKT1, ALK, APC, ATM, BRAF, CDH1, CDKN2A, CSF1R, CTNNB1, EGFR, ERBB2, ERBB4,*

EZH2, FBXW7, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAS, GNAQ, HNF1A, HRAS, IDH1, JAK2, JAK3, IDH2, KDR, KIT, KRAS, MET, MLH1, MPL, NOTCH1, NMP1, NRAS, PDGFRA, PIK3CA, PTEN, PTPN11, RB1, RET, SMAD4, SMARCB1, SMO, SRC, STK11, TP53, and *VHL*. This panel was used to study gastric neoplasia mutations. The other gene panel, the Ion Ampliseq Colon and Lung Cancer panel v2 (Life Technologies, California, United States), was used for studying mutations in colorectal neoplasia.

In study IV, 3 ng of stool DNA was used to prepare the libraries. An Ion 16S Metagenomics kit (Thermo Fisher Scientific, USA) was used according to the manufacturer's instructions. For each sample two primer pools were used to amplify six hypervariable regions (Primer set V2, V4, V8 and Primer set V3, V6-7, and V9) of the 16S rRNA gene.

In all studies (I-IV), purification of the amplified libraries was done by using Agencourt AMPure XP beads (Beckman Coulter Genomics, High Wycombe, UK), and the concentration of the purified libraries was measured on the Qubit® 2.0 Fluorometer using the Qubit® dsDNA HS Assay kit. Template preparation was performed with an Ion OneTouch 2 system, using the Ion PGM™ Hi-Q™ OT2 Kit (Thermo Fisher Scientific). The final step was sequencing carried out on the Ion Personal Genome Machine System (PGM™, Life Technologies, California, United States) using Ion 316™ chips (Ion 318™ chips in study IV) and the Ion PGM™ Sequencing Hi-Q kit v2.

4.2. Primary data analysis

In studies I, II, and III, the raw data obtained from the Ion Torrent PGM sequencer were analyzed with the Torrent Suite TM Software (v.5.2.2) (Thermo Fisher Scientific). The Variant Caller plug-in (v5.2) (Thermo Fisher Scientific) was used for variant calling with the default settings: a quality score ≥6, relative read quality ≥10, coverage of 100 for SNP/COSMIC variant and indel, and strand bias ≤95 % for SNP/COSMIC variant and 90 % for indel. The Coverage Analysis plug-in (v5.2) (Thermo Fisher Scientific) was used for coverage analysis.

In study IV, raw data were used to create OTU (Operational taxonomic unit) abundance tables. A Phyloseq R package was used to even the depth of the sequencing counts between samples converting the rarified read counts to relative abundances and generating taxa (105 families and 121 genera). Fecal microbiota community alpha diversity and observed richness were analyzed at the family and genus levels using the microbiome and vegan R packages. Community richness and diversity were quantified by the number of unique observed taxa and Shannon index, respectively.

4.3. Secondary data analysis

The secondary analysis was performed in multiple steps. In studies I, II and III, for the visualization of data, the Integrative Genomics Viewer (IGV) was used (IGV v 2.4 Broad Institute). Two analysis tools, PROVEAN and SIFT, were used to predict the effect of the non-synonymous variants on the encoded protein. Only SNVs resulting in a non-synonymous amino acid change, or a premature stop codon, and all short indels resulting in either a frameshift or insertion/deletion of amino acids were selected. All SNVs were analyzed for somatic mutations previously reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database and novel variations, i.e. new mutations detected by NGS but not reported in either COSMIC or dbSNP (build 151) databases. Threshold parameters were ≥3 % for the mutant allele frequency, and ≥20 for the quality score (≥15 in study III).

In study IV, data grouping was based on relative abundances of the taxonomic groups using hierarchical clustering with Bray-Curtis distance. Ordination with the unsupervised principal coordinates analysis (PCoA), as implemented in the phyloseq R package, was based on the Euclidean distance between Hellinger-transformed abundance profiles. Only the core genera or families that were detected in at least 20% of all samples were included in the analysis.

5. Statistical analysis

In study II, comparison between the median of two groups was performed using the Mann-Whitney test (non-parametric t-test) using the IBM SPSS advanced statistics version 24 (SPSS Inc., Chicago, IL). Independent samples T test was used to compare

the significant difference between the number of detected mutations in the before treatment and during/after treatment groups. All tests were two-sided and *P*-values ≤0.05 were considered statistically significant.

In study IV, the significance of the group-level differences was estimated with the Kruskal-Wallis test. Multiple testing correction was done separately for each group of analyses based on the Benjamini-Hochberg FDR correction. The significance of the community level differences between the groups was assessed with PERMANOVA with the R package compositions, while the significance for the differences in the abundance of individual taxa was assessed with ANCOM.

RESULTS AND DISCUSSION

1. Mutations in EBC from healthy individuals and lung neoplasia patients (I, II)

A total of 20 EBC samples from healthy controls and from 26 lung neoplasia patients were analyzed by NGS. The success rate for NGS application was 95.5% and 65.4% for controls and patients, respectively. Targeted NGS was performed using the Ion Ampliseq Colon and Lung Cancer panel v2 (Life Technologies, California, United States). The panel did not include all the exonic regions of the targeted 22 genes but only the mutational hotspot regions covering most of the driver mutations used in clinical practice. The panel consists of 22 genes frequently mutated in lung cancer including *AKT1, ALK, BRAF, CTNNB1, DDR2, EGFR, ERBB2, ERBB4, FBX7, FGFR1, FGFR2, FGFR3, KRAS, MAP2K1, MET, NOTCH1, NRAS, PIK3CA, PTEN, SMAD4, STK11,* and *TP53*. With cutoff parameters including threshold for quality score ≥ 20 and for mutant allele fraction ≥ 3% (tested previously for EGFR mutations in lung cancer patients [208]), EBC samples from healthy controls (study I) revealed a total number of 35 hotspot mutations, which were previously reported as somatic mutations in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Samples from 5 controls did not show any kind of hotspot mutations. Hotspot mutations occurred at the following genes: *BRAF, CTNNB1, DDR2, EGFR, ERBB2*, *FBXW7, FGFR3*, *KRAS, MET, NOTCHf1, NRAS, PIK3CA, PTEN, SMAD4,* and *TP53*. Simultaneously, a total of 106 novel mutations were found including all missense, nonsense and indels, which were not previously reported in COSMIC or dbSNP databases (build 150). By applying the same threshold criteria to patients' EBC samples (Study II), a total number of 39 hotspot mutations were detected in the following genes: *APC, BRAF, DDR2, EGFR, ERBB4, FBXW7, FGFR1, FGFR3, KRAS, MAP2K1, MET, NRAS, PIK3CA, PTEN, RET, SMAD4, STK11,* and *TP53*. Additionally, 98 novel mutations including non-synonymous mutations, which were not previously reported in COSMIC or dbSNP databases (build 151), were also detected. Figures 6a and 6b illustrate the genes which had hotspot mutations in both controls and patients.

In healthy controls' EBC samples (study I), 11 *TP53* and 4 *KRAS* hotspot mutations were reported in COSMIC databases. An earlier study reported similar findings through detection of *TP53* (3.2%) and *KRAS* (1%) mutations in the plasma of healthy individuals and illustrated that these mutations could be detected in the cell free DNA before cancer detection [209]. Another similar approach found *TP53* mutations in the cell free DNA in plasma from 11% of 205 non-cancerous control subjects, and in 35.7% from early stage and 54.1% from late stage SCLC patients [210]. Furthermore, in study I, one EBC control sample revealed a codon 12 G12V *KRAS* mutation with a mutant allele fraction of 6.8%. A similar study by other researchers detected a *KRAS* G12V mutation in the plasma in three out of six controls by using droplet digital PCR [211], and by using the Ion Torrent NGS, *KRAS* mutations have been found in the plasma of 3.7% healthy subjects and 4.3% of patients with chronic pancreatitis [212].

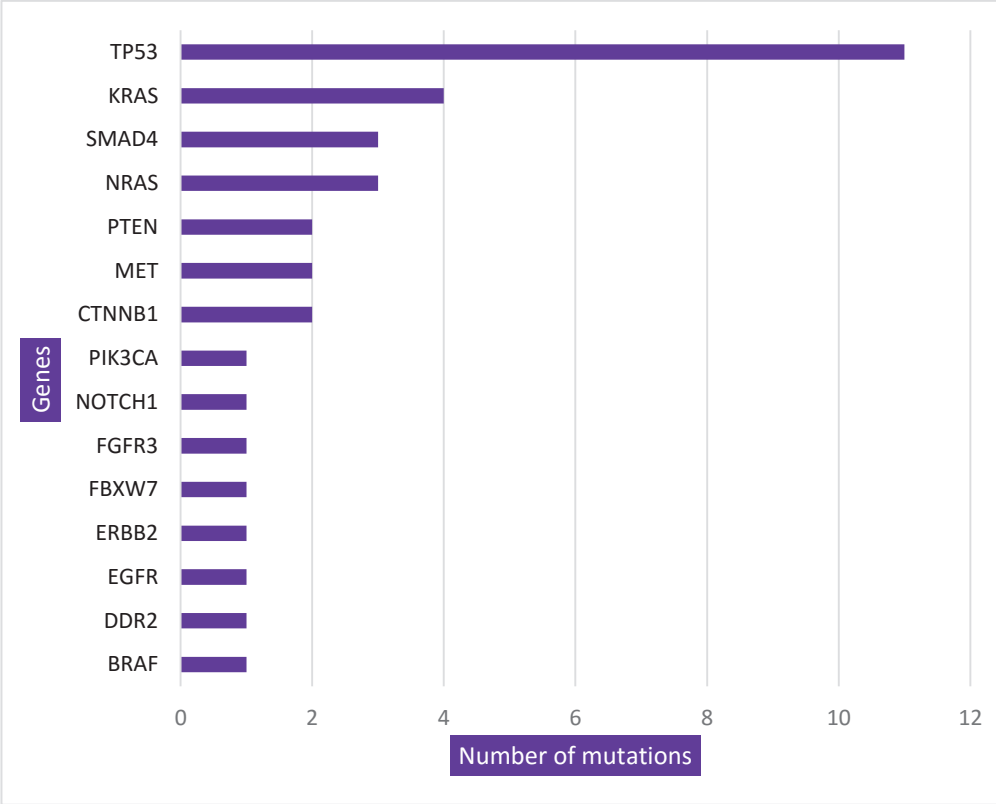

Figure 6a. Genes mutated in healthy controls' EBC along with number of hotspot mutations.

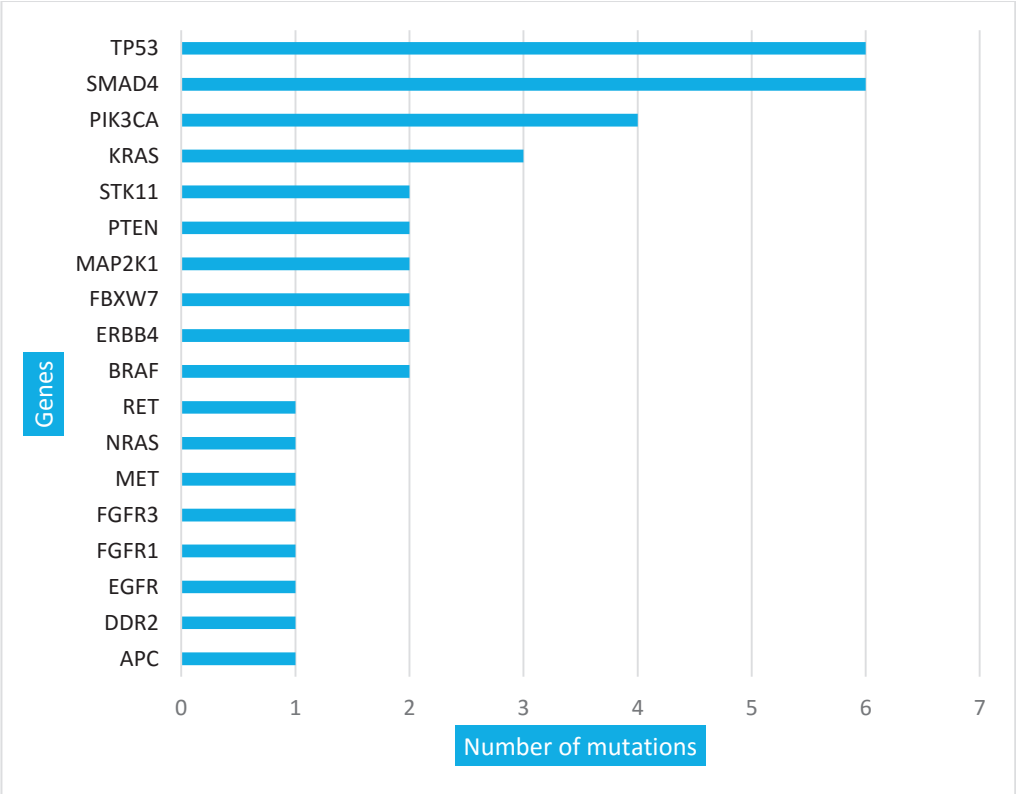

Figure 6b. Genes mutated in neoplasia patients' EBC along with number of hotspot mutations.

The presence of mutations in EBC of healthy subjects still needs to be clarified. The exact explanation could rely on two main entities. First, pulmonary cells are continuously exposed to different mutagens, e.g. asbestos that can cause DNA damage. Pulmonary cells with unrepaired DNA are directed towards apoptosis leading to release of the damaged nucleic acids into the alveolar spaces and consequently into EBC [213]. Second, gene mutations are happening in an accumulative manner, which might or might not proceed to malignancy later. Thus, mutations in controls could indicate either a homeostatic process to maintain a balanced cell cycle, or neoplastic alterations occurring at an early stage that has been detected by the highly sensitive NGS technique [214]. EBC genetic studies are still in an initial phase, and more investigations and comparison with a large cohort of patients' EBC samples is still needed before any firm conclusions can be drawn.

For validation of the NGS technique, two EBC samples were collected from the same control subject with a one-month gap. Most of the germline single nucleotide alterations were detected in both samples indicating the reproducibility of the technique. However, the sequencing depth from one of the EBC samples was not good, causing some discrepancies in the somatic alterations detected in both samples.

In lung cancer patients (study II), the average number of detected hotspot mutations was higher in their EBC samples than in the control samples (2.29 and 1.75 respectively), although this difference was not statistically significant (P=0.292). However, the average mutant allele fraction was higher in mutations seen in patients than those in controls, e.g. 22.9% and 13.6% in *TP53,* and 11.4% and 4.3% in *KRAS* in patients and controls, respectively. When comparing the EBC mutational analysis with the corresponding tissue analysis, four adenocarcinoma patients had NGS results from tumor tissue as a part of routine clinical diagnostics. One patient revealed the same *KRAS* exon 3 mutation in both EBC and tissue samples, although with low frequency in EBC (1%) on visual inspection of sequencing results in IGV. Another patient showed a *MET* exon14 (c.3028+ 3del) mutation in tissue which in our analysis was outside of the amplicon region and could not be studied. Two more patients showed *KRAS* exon2 mutations in their tissue but not in their EBC samples. Importantly, these two patients had their EBC samples taken during or after the chemotherapy treatment course. An early study reported a significant decrease in the mut/wild allele fraction of *KRAS* in EBC after tumor resection [201]. Additionally, the treated patients' EBC samples revealed a lower number of mutations than in the samples collected from patients before any treatment, however, this difference was statistically not significant (P=0.83). One of the drawbacks in study IV is the small number of patients with available sequencing data which makes it difficult to validate EBC sequencing results. Therefore, comparison of tissue and EBC mutations on a large scale is still needed before considering its application in cancer biomarkers and diagnostics. However, the principal and major advantage of investigating EBC for mutation analysis is that it is easy to obtain at different tumor stages, as most of the patients either do not have a tissue specimen or have a very small tumor tissue in the specimen which is not sufficient for sequencing. Since tissue NGS was not applicable

to 15 patients for several reasons such as insufficient tissue material, EBC could therefore represent a non-invasive material that could be used for assessment of genetic alterations and molecular profiling. Also, EBC can be an alternative for tumor re-biopsies in following the course of disease and it could be beneficial in the detection of new resistant mutations during the treatment. So far, the mutation studies on EBC are in the beginning stage, and in this thesis, we optimized the methodology to be applied in larger cohort studies.

2. Mutations screening in stool specimens in gastric and colorectal neoplasms (Study III)

DNA was successfully isolated from 77 stool specimens out of a total of 87 collected from Finnish patients, and from 13 out of 14 control specimens. The success rates of NGS from stool DNA samples for stomach and colorectal neoplasia were 78% and 87%, respectively. Targeted NGS was performed using the Ion AmpliSeq Cancer Hotspot Panel v2 (Life Technologies, California, United States). The panel consists of 50 genes, including *ABL1, AKT1, ALK, APC, ATM, BRAF, CDH1, CDKN2A, CSF1R, CTNNB1, EGFR, ERBB2, ERBB4, EZH2, FBXW7, FGFR1, FGFR2, FGFR3, FLT3, GNA11, GNAS, GNAQ, HNF1A, HRAS, IDH1, JAK2, JAK3, IDH2, KDR, KIT, KRAS, MET, MLH1, MPL, NOTCH1, NMP1, NRAS, PDGFRA, PIK3CA, PTEN, PTPN11, RB1, RET, SMAD4, SMARCB1, SMO, SRC, STK11, TP53,* and *VHL*. A total of 25 hotspot mutations (5 in gastric and 20 in colorectal) and 9 novel mutations were reported in the study. No hotspot mutations were detected in controls except for two novel mutations in *ALK* and *STK11* genes. Figure 7 illustrates the number of hotspot mutations in different genes in gastric and colorectal neoplasms.

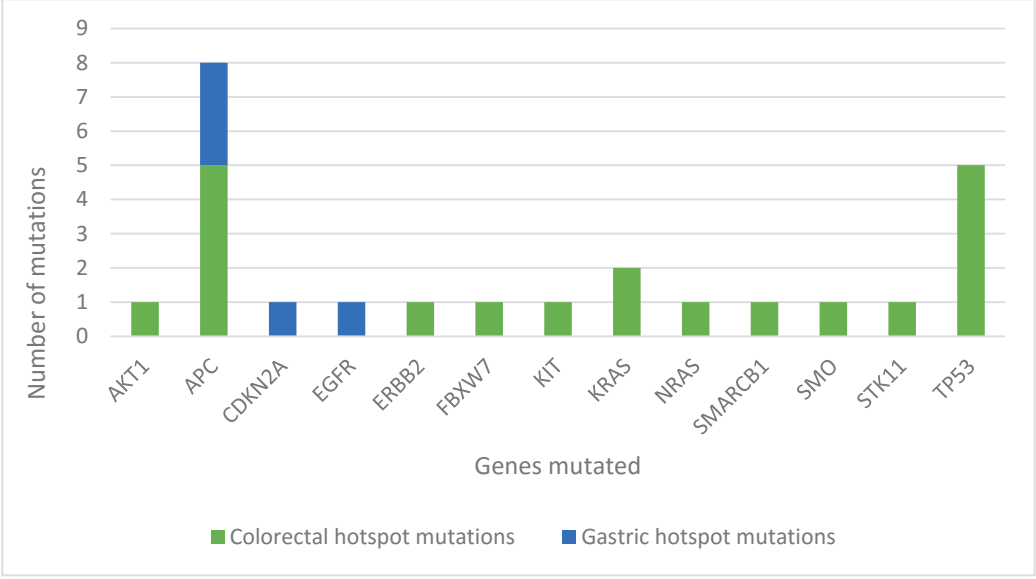

Figure 7. Number of hotspot mutations in both in gastric and colorectal neoplasms.

Sequencing data from colorectal cases revealed two codon 12 *KRAS* mutations (G12V and G12D) and one codon 61 *NRAS* mutation (Q61R). Mutations in the *KRAS* gene were also frequent in our previous study on colorectal tumors from Iranian patients [215] at codons 12, 13, 20, 63, 117, 146 and 43. The presence of *KRAS* G12V in the stool of one patient with a mutant allele fraction of 13% was confirmed in the corresponding tissue with a 20% mutant allele fraction. The corresponding tissue from another patient revealed no *KRAS* nor *NRAS* mutations with the same negative findings in the stool specimen [207]. Stool specimens from benign colorectal adenomas revealed hotspot mutations in *APC, ERBB2, FBXW7, NRAS,* and *SMARCB1* genes. The presence of *APC* mutations in colorectal adenomas increases the chance that it will become malignant [216,217].

In gastric neoplasms, the most frequent mutations occurred in the *APC* gene with three mutations encountered in gastric carcinoma and one in benign gastric dysplasia. Mutations in the *APC* gene have been reported in both histological types of gastric adenocarcinoma (intestinal and diffuse) as well as benign neoplasms with increased possibilities to convert to malignancy [218–220]. The *APC* A1582P mutation was seen in stool specimens from malignant cases and in stool specimens from benign gastric dysplasia as well. Previous studies reported *APC* mutations in tissues from gastric adenomas and flat dysplasia [221]. Mutations at *CDH1* (V82A) and *EGFR* (A750T) were also detected in the diffuse subtype of gastric carcinoma in our study [207]. It is known that these two mutations are commonly encountered in the diffuse gastric carcinoma [137,138].

In this study, we illustrated for the first time that NGS can be successfully applied to stool samples not only from patients with colorectal neoplasms, but also from patients with gastric neoplasms for investigation of gene mutations. Also, we demonstrated that mutations can be detected in stool from neoplasms at early malignant stages (stage I and II) or even in stool from benign neoplasms. The first step toward the application of stool DNA sequencing as a non-invasive biomarker in CRC neoplasms has been made. In 2014, the FDA approved a multi-target test called Cologuard for screening of adults with high risk for CRC from stool specimens. It tests for mutated

human DNA and hemoglobin [206]. However, with the application of targeted NGS, a larger number of genetic alterations in many genes could be assessed in a single test. One of the major limitations of stool DNA analysis is the low amount of human DNA, and with NGS, this issue can be overcome, and the test can be performed with only 10ng of human DNA. Another vital issue, is the lack of similar types of non-invasive tests for gastric carcinoma. In my study, NGS analysis also detected mutations in stool from gastric neoplasms, providing a promising tool for better diagnosis of these cancers, which are difficult to detect. However, application of stool DNA analysis for early detection of gastric cancer still needs future investigations. Figure 8 illustrates the total number of mutations (hotspot and novel) in both gastric and colorectal neoplasms of different tumor stages.

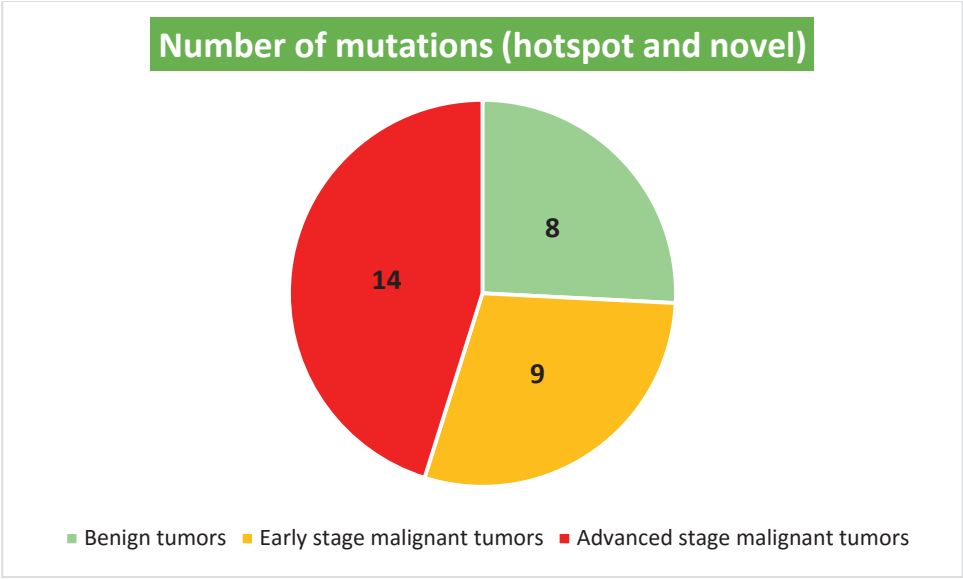

Figure 8. Mutations detected in stool DNA of patients with gastrointestinal neoplasms according to stage of tumor

Clinical applications and future prospects of EBC and stool analysis:

The reports of genetic alterations in EBC from healthy individuals and lung cancer patients are still scanty, whereas, the genetic changes in stool DNA has more investigations in CRC but not in gastric neoplasms. The suitability of NGS analysis for both EBC and stool DNA has been investigated in this thesis. However, the results show variations between healthy controls, EBC or stool samples, and tumor tissue specimens (Table 2), and stool analysis revealed hotspot mutations only in cancer patients and not in controls, thus, it has more specificity than EBC.

Table 2. Differences in mutation profile between EBC and stool in cancer patients along with findings in healthy controls and tumor tissue

	Hotspot mutations in controls	**Hotspot mutations in patients**	**Hotspot mutations in corresponding tumor tissue**
EBC	35 mutations	39 mutations	4 available specimens Confirmed in 1 specimen
Stool	No hotspot mutations	25 mutations (2 in CRC & 5 in gastric)	2 available specimens Confirmed in 2 specimens

The differences between the findings in both stool and EBC might be due to the composition of both samples. While it seems more likely that EBC is composed mainly of cell free DNA, stool has a larger cellular component [198,204]. Hotspot mutations in EBC samples were equally frequent in both controls and patients. EBC results probably describe a process of homeostasis whereby the pulmonary tissue gets rid of defective cells and damaged DNA [214]. Further EBC studies might clarify whether specific mutations in certain genes can differentiate between healthy and patients or if the mutant allele fraction of certain gene mutations could be of significance in diagnosis of malignancy. Somatic mutations from the Cancer Genome atlas database were used to customize a prediction model for immunotherapy in lung adenocarcinoma [222]. The high incidence of treatment resistance and side effects of immunotherapies revealed the need for novel biomarkers beyond PDL-1 expression [223]. Mutations in DNA repair genes can successfully predict response to PD-1

inhibitors in some cancers, e.g. CRC, but not yet in lung carcinoma. These mutations represent potential biomarkers for immunotherapies, as the high mutation frequency is associated with neo-antigens and tumor infiltrating lymphocytes [224]. In that sense, EBC mutations profiling by targeted NGS in lung cancer might provide a tool for predicting response to targeted treatment and check point inhibitor regimens.

In stool DNA based analysis, hotspot mutations were not detected in healthy controls and were seen mainly in cancer patients, thus they were more specific for the presence of GIT malignancy. In addition to stool cells exfoliated from digestive epithelium, there are also other components including extracellular human DNA and bacterial DNA [204]. This indicates the enhancement of stool DNA based studies, which can amplify only human or bacterial DNA efficiently, by applying sensitive amplicon based NGS that could screen up to 50 genes in a small DNA quantity [207].

The essential advantage of EBC and stool analysis approaches over other conventional diagnostic methods such as endoscopies and biopsies, is their suitability at all tumor stages and during treatment follow-up. Certain inquiries still to be clarified:

1- Whether EBC and stool genetic changes match perfectly with changes in the corresponding tumor tissues. This issue requires still large series of patients' samples.
2- Can EBC and stool biomarkers serve as tools for early cancer diagnosis or screening of high risk individuals?
3- How do these genetic events change after starting treatment, and how could they be used to monitor treatment and follow-up?

3. Stool microbiota abundance in different gastrointestinal neoplasms (Study IV)

The non-treated patients were classified into five categories according to the tumor location; stomach, pancreas, small intestine, colon and rectum. The treated patients were classified as a separate category regardless of the tumor site. At the family and genus levels, no significant differences were found in alpha diversity between groups (P=0.21), nor in the beta diversity between patients and controls (p>0.2). Each category was compared against controls. At the family level, Enterobacteriaceae showed higher abundances in stools of patients with neoplasms of the stomach or the small intestine, while Bifidobacteriaceae and Acidaminococcaceae had lower abundances in stools of patients with rectal and colonic neoplasms, respectively. Lactobacillaceae had a significantly lower abundance in stool from colonic and pancreatic neoplasms, although when comparing the whole treated group with the whole non-treated groups, it revealed a significantly higher abundance. (Table 3)

Table 3. Bacterial families with a significant difference in composition in patient groups compared to the control group.

Neoplasm location	High relative abundance	Low relative abundance
Stomach	Enterobacteriaceae	
Pancreas		Lactobacillaceae
Small Intestine	Enterobacteriaceae	
Colon		Lactobacillaceae
		Acidaminococcaceae
Rectum		Bifidobacteriaceae
Treated*	Lactobacillaceae	

*Treated group is compared against the whole non-treated patients group.

At the genus level, *Ruminococcus* and *Subdoligranulum* showed a higher relative abundance in stool samples from patients with stomach and colon neoplasms (*Ruminococcus* in stomach only), while *Lachnoclostridium* and *Oscillibacter* had a lower relative abundance in stools from patients with both stomach and colonic neoplasms. *Lachnoclostridium* also had a lower relative abundance in stools from

patients with small intestine neoplasms. In the rectal neoplasm group, *Bifidobacterium* showed lower relative abundance in comparison to controls. Similar to the findings at the family level, *Lactobacillus* showed higher relative abundance when compared to the whole non-treated group (Table 4).

Table 4. Bacterial genera with significant difference in all categories compared to the control group.

Tumor category	High relative abundance	Low relative abundance
Stomach	*Ruminococcus*	*Lachnoclostridium*
	Subdoligranulum	*Oscillibacter*
Small Intestine		*Lachnoclostridium*
Colon	*Subdoligranulum*	*Lachnoclostridium*
		Oscillibacter
Rectum		*Bifidobacterium*
Treated*	*Lactobacillus*	

*Treated group is compared against the whole non-treated patients group.

By focusing mainly on the genus level, similar microbiota were altered in both gastric and colonic neoplasms (*Subdoligranulum,* and *Lachnoclostridium*). They are reported to be associated with metabolic diseases and inflammation [225,226]. Moreover, *Subdoligranulum* has been found to inhibit inulin fermentation by bifidogenic bacteria which has a beneficial role in preventing colon carcinoma [227]. *Lachnoclostridium* has been reported to have a lower abundance in stools of Hashimoto's Thyroiditis patients, while *Oscillibacter* are reported to produce anti-inflammatory metabolites [228,229]. Stool samples from patients with rectal neoplasms showed a significantly lower abundance of *Bifidobacterium*. These bacteria have a role in inhibiting the growth of pathogens thus maintaining the balance of the healthy gut bacterial profile [230]. In stool samples from patients with neoplasms of the small intestine and pancreas, lower abundances of *Lachnoclostridum* and *Parabacteroides* were detected, respectively. However, due to

the small number of patients in these two groups, no robust conclusions could be drawn.

Furthermore, we compared the bacterial profiles in stool samples from patients collected before the start of any treatment (irrespective of tumor site) with those collected after treatment. This comparison showed a higher abundance of Lactobacillaceae at the family level and *Lactobacillus* at the genus level in the treated group compared to the non-treated group. Since Lactobacillaceae is considered part of the normal gut flora, its higher level in stools of the treated group could indicate restoration of the balance of the normal bacterial flora after treatment.

In study IV, significant differences in the abundances of gut bacterial taxa were found in stool specimens from patients with various GIT neoplasms according to the location of the neoplasm. These findings could be useful in assessment of neoplastic alterations in various parts of the GIT, with possible applications in monitoring disease status and treatment.

CONCLUSIONS

In this thesis, EBC and stool materials as non-invasive samples were investigated for detection of various gene mutations in lung cancer and GIT neoplasms, respectively. All mutations including hotspot and novel mutations were reported along with the mutant allele fractions. In stool samples, we also studied the composition of fecal gut microbiota in patients with GIT neoplasms grouped according to the location of the tumor in the GIT. The principal molecular technique used in this thesis was NGS (Ion Torrent PGM).

The results obtained from EBC showed the successful application of NGS on EBC DNA from both healthy individuals and lung cancer patients (Study I and II). Although 35 hotspot mutations were reported in EBC from normal controls (study I), their significance is thought to simply reflect the amount of mutagenic load to which normal pulmonary cells are exposed, e.g. smoking. One EBC sample from a healthy control revealed the clinically relevant codon 12 *KRAS* mutations with a 6.8% mutant allele fraction. To maintain cellular hemostasis, cells with unrepaired damaged DNA are eliminated through the physiological process of apoptosis. At the same time, these genetic alterations might represent very early neoplastic changes occurring in the pulmonary tissue detected by applying the highly sensitive NGS technique. By applying the same methodology to EBC from lung cancer patients, a total of 39 hotspot mutations were found (Study II). Importantly, the average mutant allele fraction was higher in patients than in controls, for instance 22.9% and 13.6% in *TP53* and 11.4% and 4.3% in *KRAS* in patients and controls, respectively. EBC could provide a helpful tool in analysis of the mutational status and molecular profiling in lung cancer patients, however, more investigations are required to test its applicability for diagnostic purposes.

Results obtained from stool samples revealed that NGS-based mutation analysis can be successfully applied to stool DNA from patients with different GIT neoplasms (Study III). With a success rate of 78% and 87% for samples from gastric and colorectal neoplasms, respectively, a total of 25 hotspot mutations (5 in gastric and 20 in colorectal) were detected. In this study, we demonstrated that gene mutations can be detected from stomach neoplasms as well as colorectal tumors. Additionally,

mutations were detected in stool from patients with benign tumors and neoplasms at early malignant stages. Indeed, these findings could have future implications in stool based diagnostic assays in different types of GIT neoplasms, and in follow up of treatment protocols.

The relative abundance of stool microbiota was compared in various GIT neoplasm locations (Study IV) against the relative abundance in control samples. The differences were variable depending on the location of the GIT neoplasm. The increased abundance of Enterobacteriace and lower abundance of two common families, Lactobacillaceae and Bifidobacteriace could provide indicators of altered balance in the gut bacterial microenvironment and potentially facilitate GIT disease monitoring. Moreover, Lactobacilli showed a higher relative abundance in stool from treated cancer patients at both the family and genus taxa levels when compared to the non-treated group. The main conclusion is that the composition of the gut microbiota varies according to the neoplasm location and depends on the treatment status of the patients.

The current status of applying these non-invasive samples in clinical practice is just at the beginning. In EBC, a few sporadic studies were reported in which single gene alteration such as *KRAS* or miRNA dysregulation in lung cancer was investigated. This thesis is the first study that tests amplicon-based NGS on EBC from healthy individuals and cancer patients. The situation is slightly different in stool samples. While fecal DNA-based analysis has taken a step forward, very little is known about applying the same methodology for gastric carcinoma. Before applying those non-invasive techniques in clinical situations such as targeted therapy decisions, a route map starting from the current stage needs to be established, and larger cohorts including larger patients' samples need to be tested.

The era of non-invasive samples in cancer diagnosis and management has taken one step further after introduction of the NGS technique, and NGS is gradually replacing the conventional molecular methods. Although the application of NGS to non-invasive cancer samples has opened a window of hope for earlier and better cancer detection, it is still at an initial stage and needs more studies and investigations before it can be fully ready for clinical use.

ACKNOWLEDGEMENT

This work was carried out at the Department of Pathology in the Haartman Institute, University of Helsinki during the period 2015-2018. I am grateful for the financial support received from EDUFI fellowships, the Finnish Cancer Society, the Otto A. Malm Foundation, the Sigrid Jusélius Foundation, the special governmental subsidy research funds provided to the Helsinki and Uusimaa Hospital District (HUS EVO), and the University of Helsinki Dissertation Completion grant.

First, many thanks to God who gave me the faith and the power to complete this work. Second, I sincerely express my gratitude and thanks to everyone who made this work achievable by the end of the doctoral journey, and namely for the following persons:

My supervisor, Professor Sakari Knuutila, for his limitless guidance and support over the years. Thanks for giving me the opportunity to be a member in your successful research group, and for teaching me the basics of molecular genetics and scientific research. I am grateful for his continuous enthusiastic and positive attitude and for teaching me how to keep challenging at a very high profile whatever the obstacles are.

My second supervisor, Dr. Virinder Kaur Sarhadi, for all her valuable support. Thanks for being ready for all types of questions all the time, guidance through my whole work, and for introducing me to academic research.

My thesis committee members, Associate Professor Sanna Lehtonen, and Dr. Panu Kovanen for their supportive and instructive role.

The official pre-examiners Dr. Antti Jekunen and Professor Anne Kallioniemi for their comments and valuable suggestions that helped to improve the quality of this thesis manuscript. Professor Tarja Laitinen is warmly acknowledged for accepting the opponent role in my thesis defense. Many thanks to Debbie Kaska for a thorough and careful language revision of the final manuscript.

The coauthors of my studies, Leo Lahti, Tiina Karla, Milja Tikkanen, Homa Ehsan, Monika Carpelan-Holmström, Selja Koskensalo, Hilpi Rautelin, Gemma Armengol, and Paivi Piirila. My sincere thanks to Dr.Arto Kokkola and Dr. Pauli Puolakkainen for their extremely helpful and essential role in the project of digestive disease. Special thanks to Dr. Aija Knuuttila for her continuous contribution to the EBC research project and help in collection of patients' clinical data. Exclusive gratefulness to Professor Tom Böhling for his endless encouragement and support at both professional and personal levels.

Former and present members of the CMG group, Kowan Jee, Gemma Armengol, Neda Mosakhani, Lauri Lehtimäki, Satu Mäki-Nevala, Milja Tikkanen, Tiina Wirtanen, Homa Ehsan, Farideh Saberi and all others for making a friendly and pleasant work environment.

All the other colleagues outside the laboratory especially Alhadi Almangush for his wonderful ideas and motivation to complete this work perfectly.

My deepest gratitude and passion goes to my lovely wife, Fardous, and my son Belal, for their love, emotional support, encouragement, and patience during the most stressful moments throughout our life. Finally, I am also very grateful to my dear parents, sisters, brothers, and the whole family for being my backbone during my life, and to whom I dedicate this work.

Helsinki, 2018

Omar Youssef

WEB-BASED RESOURCES

COSMIC	http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
NCBI	dbSNP http://www.ncbi.nlm.nih.gov/SNP/
SIFT	http://sift.jcvi.org/
PROVEAN	http://provean.jcvi.org

REFERENCES

1. Cooper GM. The development and causes of cancer. 2000; The Cell: A Molecular Approach. 2nd edition.

2. Lee EYHP, Muller WJ. Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol. 2010;2:a003236.

3. Danaei G, Vander Hoorn S, Lopez AD, Murray CJL, Ezzati M, Comparative Risk Assessment collaborating group (Cancers). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet Lond Engl. 2005;366:1784–93.

4. Garber JE, Offit K. Hereditary cancer predisposition syndromes. J Clin Oncol. 2005;23:276–92.

5. Wang Q. Cancer predisposition genes: molecular mechanisms and clinical impact on personalized cancer care: examples of Lynch and HBOC syndromes. Acta Pharmacol Sin. 2016;37:143–9.

6. Herceg Z, Hainaut P. Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis. Mol Oncol. 2007;1:26–41.

7. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.

8. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science. 2013;339:1546–58.

9. Liu J, Dang H, Wang XW. The significance of intertumor and intratumor heterogeneity in liver cancer. Exp Mol Med. 2018;50:e416.

10. Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J. Mutations: types and causes. 2000; Molecular Cell Biology. 4th edition.

11. Sen S. Aneuploidy and cancer. Curr Opin Oncol. 2000;12:82–8.

12. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet. 2007;39:S37-42.

13. Ren H, Francis W, Boys A, Chueh AC, Wong N, La P, et al. BAC-based PCR fragment microarray: high-resolution detection of chromosomal deletion and duplication breakpoints. Hum Mutat. 2005;25:476–82.

14. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448:561–6.

15. Alliance G, Screening Services. Chromosomal abnormalities. 2009; Understanding genetics: a New York, mid-Atlantic guide for patients and health professionals.

16. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem. 2010;79:351–79.

17. Barh D, Malhotra R, Ravi B, Sindhurani P. Microrna let-7: an emerging next-generation cancer therapeutic. Curr Oncol. 2010;17:70–80.

18. Nielsen BS, Jørgensen S, Fog JU, Søkilde R, Christensen IJ, Hansen U, et al. High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients. Clin Exp Metastasis. 2011;28:27–38.

19. Võsa U, Vooder T, Kolde R, Fischer K, Välk K, Tõnisson N, et al. Identification of miR-374a as a prognostic marker for survival in patients with early-stage non small cell lung cancer. Genes Chromosomes Cancer. 2011;50:812–22.

20. Monroig Pdel C, Chen L, Zhang S, Calin GA. Small molecule compounds targeting miRNAs for cancer therapy. Adv Drug Deliv Rev. 2015;81:104–16.

21. Jin B, Li Y, Robertson KD. DNA methylation. Genes Cancer. 2011;2:607–17.

222. Daura-Oller E, Cabre M, Montero MA, Paternain JL, Romeu A. Specific gene hypomethylation and cancer: new insights into coding region feature trends. Bioinformation. 2009;3:340–3.

23. Gonzalo S. Epigenetic alterations in aging. J Appl Physiol Bethesda Md 1985. 2010;109:586–97.

24. Tahara T, Arisawa T. DNA methylation as a molecular biomarker in gastric cancer. Epigenomics. 2015;7:475–86.

25. Gnyszka A, Jastrzebski Z, Flis S. DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer Res. 2013;33:2989–96.

26. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128:635–8.

27. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823–37.

28. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008;40:897–903.

29. Van Den Broeck A, Brambilla E, Moro-Sibilot D, Lantuejoul S, Brambilla C, Eymin B, et al. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. Clin Cancer Res. 2008;14:7237–45.

30. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. doi: 10.3322/caac.21492

31. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017;67:7–30.

32. Hecht SS. Lung carcinogenesis by tobacco smoke. Int J Cancer. 2012;131:2724–32.

33. Jaakkola MS, Jaakkola JJK. Impact of smoke-free workplace legislation on exposures and health: possibilities for prevention. Eur Respir J. 2006;28:397–408.

34. Parkin DM. 2. Tobacco-attributable cancer burden in the UK in 2010. Br J Cancer. 2011;105 Suppl 2:S6–13.

35. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med. 2011;32:605-44.

36. Ridge CA, McErlean AM, Ginsberg MS. Epidemiology of lung cancer. Semin Interv Radiol. 2013;30:93–8.

37. Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. Carcinogenesis. 2007;28:507–18.

38. Larsen JE, Minna JD. Molecular biology of lung cancer: clinical implications. Clin Chest Med. 2011;32:703–40.

39. Marshall AL, Christiani DC. Genetic susceptibility to lung cancer--light at the end of the tunnel? Carcinogenesis. 2013;34:487–502.

40. Shiraishi K, Honda T, Kohno T. An overview of genetic polymorphism and lung cancer risk. Adv Cancer Prev. 2016;1:1–5.

41. Lemjabbar-Alaoui H, Hassan OU, Yang Y-W, Buchanan P. Lung cancer: biology and treatment options. Biochim Biophys Acta. 2015;1856:189–210.

42. Subramanian J, Govindan R. Lung cancer in never smokers: a review. J Clin Oncol. 2007;25:561–70.

43. Gandara DR, Hammerman PS, Sos ML, Lara PN, Hirsch FR. Squamous cell lung cancer: from tumor genomics to cancer therapeutics. Clin Cancer Res. 2015;21:2236–43.

44. Sholl LM. Large-cell carcinoma of the lung: a diagnostic category redefined by immunohistochemistry and genomics. Curr Opin Pulm Med. 2014;20:324–31.

45. Inamura K. Update on immunohistochemistry for the diagnosis of lung cancer. Cancers. 2018; 10(3). pii: E72. doi: 10.3390/cancers10030072.

46. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10:1243–60.

47. Scagliotti G, Brodowicz T, Shepherd FA, Zielinski C, Vansteenkiste J, Manegold C, et al. Treatment-by-histology interaction analyses in three phase III trials show superiority of pemetrexed in nonsquamous non-small cell lung cancer. J Thorac Oncol. 2011;6:64–70.

48. Johnson DH, Fehrenbacher L, Novotny WF, Herbst RS, Nemunaitis JJ, Jablons DM, et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J Clin Oncol. 2004;22:2184–91.

49. Kadara H, Wistuba II. Field cancerization in non-small cell lung cancer: implications in disease pathogenesis. Proc Am Thorac Soc. 2012;9:38–42.

50. Kadara H, Scheet P, Wistuba II, Spira AE. Early events in the molecular pathogenesis of lung cancer. Cancer Prev Res (Phila Pa). 2016;9:518–27.

51. Westra WH, Baas IO, Hruban RH, Askin FB, Wilson K, Offerhaus GJA, et al. K-ras oncogene activation in atypical alveolar hyperplasias of the human lung. Cancer Res. 1996;56:2224–8.

52. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. Cancer Prev Res (Phila Pa). 2008;1:192–200.

53. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511:543–50.

54. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. Int J Cancer. 2006;118:257–62.

55. Suzuki M, Shigematsu H, Hiroshima K, Iizasa T, Nakatani Y, Minna JD, et al. Epidermal growth factor receptor expression status in lung cancer correlates with its mutation. Hum Pathol. 2005;36:1127–34.

56. Ladanyi M, Pao W. Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. Mod Pathol. 2008;21 Suppl 2:S16-22.

57. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun. 2014;5:4846.

58. Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N Engl J Med. 2010;363:1693–703.

59. Wong DW-S, Leung EL-H, So KK-T, Tam IY-S, Sihoe AD-L, Cheng LC, et al. The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS. Cancer. 2009;115:1723–33.

60. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489:519–25.

61. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet. 2009;41:1238–42.

62. Hussenet T, Dali S, Exinger J, Monga B, Jost B, Dembelé D, et al. SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. PloS One. 2010;5:e8960.

63. George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature. 2015;524:47–53.

64. Gadgeel SM, Ramalingam SS, Kalemkerian GP. Treatment of lung cancer. Radiol Clin North Am. 2012;50:961–74.

65. Ettinger DS, Aisner DL, Wood DE, Akerley W, Bauman J, Chang JY, et al. NCCN Guidelines Insights: Non–Small Cell Lung Cancer, Version 5.2018. J Natl Compr Canc Netw. 2018;16:807–21.

66. Howington JA, Blum MG, Chang AC, Balekian AA, Murthy SC. Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest. 2013;143:e278S-e313S.

67. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 2016;5:288–300.

68. Masters GA, Temin S, Azzoli CG, Giaccone G, Baker S, Brahmer JR, et al. Systemic therapy for stage IV non-small-cell lung cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol. 2015;33:3488–515.

69. Baxevanos P, Mountzios G. Novel chemotherapy regimens for advanced lung cancer: have we reached a plateau? Ann Transl Med [Internet]. 2018;6:139.

70. Amini A, Yeh N, Gaspar LE, Kavanagh B, Karam SD. Stereotactic body radiation therapy (SBRT) for lung cancer patients previously treated with conventional radiotherapy: a review. Radiat Oncol Lond Engl. 2014;9:210.

71. Chan BA, Hughes BGM. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. Transl Lung Cancer Res. 2015;4:36–54.

72. Mitsudomi T, Morita S, Yatabe Y, Negoro S, Okamoto I, Tsurutani J, et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol. 2010;11:121–8.

73. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012;13:239–46.

74. Wu Y-L, Zhou C, Hu C-P, Feng J, Lu S, Huang Y, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. Lancet Oncol. 2014;15:213–22.

75. Lynch TJ, Patel T, Dreisbach L, McCleod M, Heim WJ, Hermann RC, et al. Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. J Clin Oncol. 2010;28:911–7.

76. Pirker R, Pereira JR, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. Lancet Lond Engl. 2009;373:1525–31.

77. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med. 2011;3:75ra26.

78. Sun JM, Ahn MJ, Choi YL, Ahn JS, Park K. Clinical implications of T790M mutation in patients with acquired resistance to EGFR tyrosine kinase inhibitors. Lung Cancer. 2013;82:294–8.

79. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A. 2008;105:2070–5.

80. Bean J, Brennan C, Shih J-Y, Riely G, Viale A, Wang L, et al. MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. Proc Natl Acad Sci U S A. 2007;104:20932–7.

81. Ohashi K, Sequist LV, Arcila ME, Moran T, Chmielecki J, Lin Y-L, et al. Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. Proc Natl Acad Sci U S A. 2012;109:E2127-2133.

82. Takezawa K, Pirazzoli V, Arcila ME, Nebhan CA, Song X, de Stanchina E, et al. HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFRT790M mutation. Cancer Discov. 2012;2:922–33.

83. Miller VA, Hirsh V, Cadranel J, Chen YM, Park K, Kim SW, et al. Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. Lancet Oncol. 2012;13:528–38.

84. Tan CS, Kumarakulasinghe NB, Huang YQ, Ang YLE, Choo JRE, Goh BC, et al. Third generation EGFR TKIs: current data and future directions. Mol Cancer. 2018;17:29.

85. Sullivan I, Planchard D. ALK inhibitors in non-small cell lung cancer: the latest evidence and developments. Ther Adv Med Oncol. 2016;8:32–47.

86. Camidge DR, Bang YJ, Kwak EL, Iafrate AJ, Varella-Garcia M, Fox SB, et al. Activity and safety of crizotinib in patients with ALK-positive non-small-cell lung cancer: updated results from a phase 1 study. Lancet Oncol. 2012;13:1011–9.

87. Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. N Engl J Med. 2013;368:2385–94.

88. Doebele RC, Pilling AB, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, et al. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clin Cancer Res. 2012;18:1472–82.

89. Giantonio BJ, Catalano PJ, Meropol NJ, O'Dwyer PJ, Mitchell EP, Alberts SR, et al. Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern Cooperative Oncology Group Study E3200. J Clin Oncol. 2007;25:1539–44.

90. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med. 2006;355:2542–50.

91. Dine J, Gordon R, Shames Y, Kasler MK, Barton-Burke M. Immune checkpoint inhibitors: an innovation in immunotherapy for the treatment and management of patients with cancer. Asia-Pac J Oncol Nurs. 2017;4:127–35.

92. Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl J Med. 2015;373:123–35.

93. Minguet J, Smith KH, Bramlage P. Targeted therapies for treatment of non-small cell lung cancer--recent advances and future perspectives. Int J Cancer. 2016;138:2549–61.

94. Lynch TJ, Bondarenko I, Luft A, Serwatowski P, Barlesi F, Chacko R, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. J Clin Oncol. 2012;30:2046–54.

95. Recio-Boiles A, Babiker HM. Cancer, Gastric. 2018; StatPearls.

96. Ibrahim M, Gilbert K. Management of gastric cancer in Indian population. Transl Gastroenterol Hepatol. 2017;2:64.

97. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomark Prev. 2014;23:700–13.

98. Shiota S, Suzuki R, Yamaoka Y. The significance of virulence factors in Helicobacter pylori. J Dig Dis. 2013;14:341–9.

99. Blaser MJ, Falkow S. What are the consequences of the disappearing human microbiota? Nat Rev Microbiol. 2009;7:887–94.

100. Crew KD, Neugut AI. Epidemiology of gastric cancer. World J Gastroenterol. 2006;12:354–62.

101. Clark GW, Smyrk TC, Burdiles P, Hoeft SF, Peters JH, Kiyabu M, et al. Is Barrett's metaplasia the source of adenocarcinomas of the cardia? Arch Surg Chic Ill 1960. 1994;129:609–14.

102. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003;348:1625–38.

103. Øines M, Helsingen LM, Bretthauer M, Emilsson L. Epidemiology and risk factors of colorectal polyps. Best Pract Res Clin Gastroenterol. 2017;31:419–24.

104. Haggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009;22:191–7.

105. Janout V, Kollárová H. Epidemiology of colorectal cancer. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub. 2001;145:5–10.

106. Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies. Int J Cancer. 2006;119:2657–64.

107. Botteri E, Iodice S, Raimondi S, Maisonneuve P, Lowenfels AB. Cigarette smoking and adenomatous polyps: a meta-analysis. Gastroenterology. 2008;134:388–95.

108. Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol Microbiol Scand. 1965;64:31–49.

109. Leocata P, Ventura L, Giunta M, Guadagni S, Fortunato C, Discepoli S, et al. [Gastric carcinoma: a histopathological study of 705 cases]. Ann Ital Chir. 1998;69:331–7.

110. Ma J, Shen H, Kapesa L, Zeng S. Lauren classification and individualized chemotherapy in gastric cancer. Oncol Lett. 2016;11:2959–64.

111. Fléjou J-F. [WHO Classification of digestive tumors: the fourth edition]. Ann Pathol. 2011;31:S27-31.

112. Goseki N, Takizawa T, Koike M. Differences in the mode of the extension of gastric cancer classified by histological type: new histological classification of gastric carcinoma. Gut. 1992;33:606–12.

113. Ming SC. Gastric carcinoma. A pathobiological classification. Cancer. 1977;39:2475–85.

114. Grundmann E, Schlake W. Histological classification of gastric cancer from initial to advanced stages. Pathol Res Pract. 1982;173:260–74.

115. Aaltonen LA, Hamilton SR, World Health Organization, International Agency for Research on Cancer, editors. Pathology and genetics of tumours of the digestive system. Lyon : Oxford: IARC Press ; Oxford University Press; 2000.

116. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17:1471–4.

117. Akkoca AN, Yanık S, Özdemir ZT, Cihan FG, Sayar S, Cincin TG, et al. TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. Int J Clin Exp Med. 2014;7:2828–35.

118. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202–9.

119. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487:330–7.

120. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D. Gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2016;27:v38–49.

121. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJH, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. N Engl J Med. 2006;355:11–20.

122. Ford HER, Marshall A, Bridgewater JA, Janowitz T, Coxon FY, Wadsley J, et al. Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (COUGAR-02): an open-label, phase 3 randomised controlled trial. Lancet Oncol. 2014;15:78–86.

123. Macdonald JS, Smalley SR, Benedetti J, Hundahl SA, Estes NC, Stemmermann GN, et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. N Engl J Med. 2001;345:725–30.

124. Bang Y-J, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet Lond Engl. 2010;376:687–97.

125. Kang Y-K, Rha SY, Tassone P, Barriuso J, Yu R, Szado T, et al. A phase IIa dose-finding and safety study of first-line pertuzumab in combination with trastuzumab, capecitabine and cisplatin in patients with HER2-positive advanced gastric cancer. Br J Cancer. 2014;111:660–6.

126. Matsuoka T, Yashiro M. The role of PI3K/Akt/mTOR signaling in gastric carcinoma. Cancers. 2014;6:1441–63.

127. Fuchs CS, Tomasek J, Yong CJ, Dumitru F, Passalacqua R, Goswami C, et al. Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): an international, randomised, multicentre, placebo-controlled, phase 3 trial. Lancet Lond Engl. 2014;383:31–9.

128. Wilke H, Muro K, Van Cutsem E, Oh S-C, Bodoky G, Shimada Y, et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. Lancet Oncol. 2014;15:1224–35.

129. Mishra J, Dromund J, Quazi SH, Karanki SS, Shaw J, Chen B, et al. Prospective of colon cancer treatments and scope for combinatorial approach to enhanced cancer cell apoptosis. Crit Rev Oncol Hematol. 2013;86:232–50.

130. Carrato A. Adjuvant treatment of colorectal cancer. Gastrointest Cancer Res GCR. 2008;2:S42–6.

131. Häfner MF, Debus J. Radiotherapy for colorectal cancer: current standards and future perspectives. Visc Med. 2016;32:172–7.

132. De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, et al. 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. Lancet Oncol. 2010;11:753–62.

133. Ohhara Y, Fukuda N, Takeuchi S, Honma R, Shimizu Y, Kinoshita I, et al. Role of targeted therapy in metastatic colorectal cancer. World J Gastrointest Oncol. 2016;8:642–55.

134. Jonker DJ, O'Callaghan CJ, Karapetis CS, Zalcberg JR, Tu D, Au H-J, et al. Cetuximab for the treatment of colorectal cancer. N Engl J Med. 2007;357:2040–8.

135. Douillard JY, Siena S, Cassidy J, Tabernero J, Burkes R, Barugel M, et al. Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. J Clin Oncol. 2010;28:4697–705.

136. Hurwitz H, Fehrenbacher L, Novotny W, Cartwright T, Hainsworth J, Heim W, et al. Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. N Engl J Med. 2004;350:2335–42.

137. Van Cutsem E, Tabernero J, Lakomy R, Prenen H, Prausová J, Macarulla T, et al. Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. J Clin Oncol. 2012;30:3499–506.

138. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet Lond Engl. 2013;381:303–12.

139. Roy S, Trinchieri G. Microbiota: a key orchestrator of cancer therapy. Nat Rev Cancer. 2017;17:271–85.

140. Zitvogel L, Galluzzi L, Viaud S, Vétizou M, Daillère R, Merad M, et al. Cancer and the gut microbiota: an unexpected link. Sci Transl Med. 2015;7:271ps1.

141. Fox JG, Wang TC. Inflammation, atrophy, and gastric cancer. J Clin Invest. 2007;117:60–9.

142. Islami F, Kamangar F. Helicobacter pylori and esophageal cancer risk: a meta-analysis. Cancer Prev Res (Phila). 2008;1:329–38.

143. Peek RM, Blaser MJ. Helicobacter pylori and gastrointestinal tract adenocarcinomas. Nat Rev Cancer. 2002;2:28–37.

144. Chen GY, Shaw MH, Redondo G, Núñez G. The innate immune receptor Nod1 protects the intestine from inflammation-induced tumorigenesis. Cancer Res. 2008;68:10060–7.

145. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. Nature. 2012;491:254–8.

146. Klimesova K, Kverka M, Zakostelska Z, Hudcovic T, Hrncir T, Stepankova R, et al. Altered gut microbiota promotes colitis-associated cancer in IL-1 receptor-associated kinase M-deficient mice. Inflamm Bowel Dis. 2013;19:1266–77.

147. Vannucci L, Stepankova R, Kozakova H, Fiserova A, Rossmann P, Tlaskalova-Hogenova H. Colorectal carcinogenesis in germ-free and conventionally

reared rats: different intestinal environments affect the systemic immunity. Int J Oncol. 2008;32:609–17.

148. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol. 2014;12:661–72.

149. Nistal E, Fernández-Fernández N, Vivas S, Olcoz JL. Factors determining colorectal cancer: the role of the intestinal microbiota. Front Oncol. 2015;5:220.

150. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013;13:800–12.

151. Holmes E, Li JV, Marchesi JR, Nicholson JK. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. Cell Metab. 2012;16:559–64.

152. Moresco EMY, LaVine D, Beutler B. Toll-like receptors. Curr Biol CB. 2011;21:R488-493.

153. Kennedy CL, Najdovska M, Tye H, McLeod L, Yu L, Jarnicki A, et al. Differential role of MyD88 and Mal/TIRAP in TLR2-mediated gastric tumourigenesis. Oncogene. 2014;33:2540–6.

154. Dapito DH, Mencin A, Gwak G-Y, Pradere J-P, Jang M-K, Mederacke I, et al. Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4. Cancer Cell. 2012;21:504–16.

155. Fukata M, Chen A, Vamadevan AS, Cohen J, Breglio K, Krishnareddy S, et al. Toll-like receptor-4 promotes the development of colitis-associated colorectal tumors. Gastroenterology. 2007;133:1869–81.

156. Fukata M, Shang L, Santaolalla R, Sotolongo J, Pastorini C, España C, et al. Constitutive activation of epithelial TLR4 augments inflammatory responses to

mucosal injury and drives colitis-associated tumorigenesis. Inflamm Bowel Dis. 2011;17:1464–73.

157. Tye H, Kennedy CL, Najdovska M, McLeod L, McCormack W, Hughes N, et al. STAT3-driven upregulation of TLR2 promotes gastric tumorigenesis independent of tumor inflammation. Cancer Cell. 2012;22:466–78.

158. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, et al. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol. 2011;85:863–71.

159. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature. 2013;499:97–101.

160. Bindels LB, Porporato P, Dewulf EM, Verrax J, Neyrinck AM, Martin JC, et al. Gut microbiota-derived propionate reduces cancer cell proliferation in the liver. Br J Cancer. 2012;107:1337–44.

161. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell Metab. 2011;13:517–26.

162. Hu S, Dong TS, Dalal SR, Wu F, Bissonnette M, Kwon JH, et al. The microbe-derived short chain fatty acid butyrate targets miRNA-dependent p21 gene expression in human colon cancer. PloS One. 2011;6:e16221.

163. Carbonero F, Benefiel AC, Alizadeh-Ghamsari AH, Gaskins HR. Microbial pathways in colonic sulfur metabolism and links with health and disease. Front Physiol. 2012;3:448.

164. Russell WR, Gratz SW, Duncan SH, Holtrop G, Ince J, Scobbie L, et al. High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. Am J Clin Nutr. 2011;93:1062–72.

165. Windey K, De Preter V, Verbeke K. Relevance of protein fermentation to gut health. Mol Nutr Food Res. 2012;56:184–96.

166. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, et al. Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. Cell Host Microbe. 2007;2:204.

167. Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, et al. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science. 2010;328:228–31.

168. Winter SE, Winter MG, Xavier MN, Thiennimitr P, Poon V, Keestra AM, et al. Host-derived nitrate boosts growth of E. coli in the inflamed gut. Science. 2013;339:708–11.

169. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10:472–84.

170. Diaz LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32:579–86.

171. Siravegna G, Bardelli A. Blood circulating tumor DNA for non-invasive genotyping of colon cancer patients. Mol Oncol. 2016;10:475–80.

172. Haber DA, Velculescu VE. Blood-based analyses of cancer: circulating tumor cells and circulating tumor dna. Cancer Discov. 2014;4:650–61.

173. Jung A, Kirchner T. Liquid biopsy in tumor genetic diagnosis. Dtsch Ärztebl Int. 2018;115:169–74.

174. Salvi S, Gurioli G, De Giorgi U, Conteduca V, Tedaldi G, Calistri D, et al. Cell-free DNA as a diagnostic marker for cancer: current insights. OncoTargets Ther. 2016;9:6549–59.

175. Danese E, Minicozzi AM, Benati M, Montagnana M, Paviati E, Salvagno GL, et al. Comparison of genetic and epigenetic alterations of primary tumors and matched plasma samples in patients with colorectal cancer. PLoS One. 2015; 10(5): e0126417.

176. Chen X, Bonnefoi H, Diebold-Berger S, Lyautey J, Lederrey C, Faltin-Traub E, et al. Detecting tumor-related alterations in plasma or serum DNA of patients diagnosed with breast cancer. Clin Cancer Res. 1999;5:2297–303.

177. Hou H, Yang X, Zhang J, Zhang Z, Xu X, Zhang X, et al. Discovery of targetable genetic alterations in advanced non-small cell lung cancer using a next-generation sequencing-based circulating tumor DNA assay. Sci Rep. 2017;7:14605.

178. Kang S, Li Q, Chen Q, Zhou Y, Park S, Lee G, et al. CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome Biol. 2017;18:53. doi: 10.1186/s13059-017-1191-5.

179. Di Meo A, Bartlett J, Cheng Y, Pasic MD, Yousef GM. Liquid biopsy: a step forward towards precision medicine in urologic malignancies. Mol Cancer. 2017;16(1):80. doi: 10.1186/s12943-017-0644-5.

180. Goessl C, Müller M, Straub B, Miller K. DNA alterations in body fluids as molecular tumor markers for urological malignancies. Eur Urol. 2002;41:668–76.

181. Franovic A, Raymond VM, Erlander MG, Reckamp KL. Urine test for EGFR analysis in patients with non-small cell lung cancer. J Thorac Dis. 2017;9:S1323–31.

182. Thunnissen FBJM. Sputum examination for early detection of lung cancer. J Clin Pathol. 2003;56:805–10.

183. Somers VA, Pietersen AM, Theunissen PH, Thunnissen FB. Detection of K-ras point mutations in sputum from patients with adenocarcinoma of the lung by point-EXACCT. J Clin Oncol. 1998;16:3061–8.

184. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. 2002;62:2370–7.

185. Li R, Todd NW, Qiu Q, Fan T, Zhao RY, Rodgers WH, et al. Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. Clin Cancer Res. 2007;13:482–7.

186. Wang X, Kaczor-Urbanowicz KE, Wong DTW. Salivary Biomarkers in Cancer Detection. Med Oncol. 2017;34:7.

187. Saxena S, Sankhla B, Sundaragiri KS, Bhargava A. A review of salivary biomarker: A tool for early oral cancer diagnosis. Adv Biomed Res. 2017;6:90.

188. Li X, Yang T, Lin J. Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy. J Biomed Opt. 2012;17:037003.

189. Agha-Hosseini F, Mirzaii-Dizgah I, Rahimi A. Correlation of serum and salivary CA15-3 levels in patients with breast cancer. Med Oral Patol Oral Cirugia Bucal. 2009;14:e521-524.

190. Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park N-H, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology. 2010;138:949-957.e1-7.

191. Li G, Sun Y. Liquid biopsy: advances, limitations and clinical applications. 2017;6.

192. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6:224ra24.

193. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11:426–37.

194. Yao L, Ren S, Zhang M, Du F, Zhu Y, Yu H, et al. Identification of specific DNA methylation sites on the Y-chromosome as biomarker in prostate cancer. Oncotarget. 2015;6:40611–21.

195. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mediates resistance to AZD9291 in advanced non-small cell lung cancer harboring EGFR T790M. Nat Med. 2015;21:560–2.

196. Siravegna G, Bardelli A. Minimal residual disease in breast cancer: in blood veritas. Clin Cancer Res. 2014;20:2505–7.

197. Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. J Clin Oncol. 2013;31:1002–8.

198. Youssef O, Sarhadi VK, Armengol G, Piirilä P, Knuuttila A, Knuutila S. Exhaled breath condensate as a source of biomarkers for lung carcinomas. A focus on genetic and epigenetic markers-A mini-review. Genes Chromosomes Cancer. 2016;55:905–14.

199. Gessner C, Kuhn H, Toepfer K, Hammerschmidt S, Schauer J, Wirtz H. Detection of p53 gene mutations in exhaled breath condensate of non-small cell lung cancer patients. Lung Cancer. 2004;43:215–22.

200. Zhang D, Takigawa N, Ochi N, Tanimoto Y, Noujima D, Chen YY, et al. Detection of the EGFR mutation in exhaled breath condensate from a heavy smoker with squamous cell carcinoma of the lung. Lung Cancer. 2011;73:379–80.

201. Kordiak J, Szemraj J, Hamara K, Bialasiewicz P, Nowak D. Complete surgical resection of lung tumor decreases exhalation of mutated KRAS oncogene. Respir Med. 2012;106:1293–300.

202. Yang Ai SS, Hsu K, Herbert C, Cheng Z, Hunt J, Lewis CR, et al. Mitochondrial DNA mutations in exhaled breath condensate of patients with lung cancer. Respir Med. 2013;107:911–8.

203. Mozzoni P, Banda I, Goldoni M, Corradi M, Tiseo M, Acampa O, et al. Plasma and EBC microRNAs as early biomarkers of non-small-cell lung cancer. Biomarkers. 2013;18:679–86.

204. Youssef O, Sarhadi VK, Lehtimäki L, Tikkanen M, Kokkola A, Puolakkainen P, et al. Mutations by next generation sequencing in stool DNA from colorectal carcinoma patients – a literature review and our experience with this methodology. J Anal Oncol. 2016;5:24–32.

205. Dutta SK, Silverstein RJ, Fleisher AS, Vengurlekar SM, Shami S, Nair PP. Increased expression of CD44 in isolated fecal colonocytes in patients with colonic neoplasia. Gastroenterology. 1995;108:A463.

206. Smith RA, Manassaram-Baptiste D, Brooks D, Doroshenk M, Fedewa S, Saslow D, et al. Cancer screening in the United States, 2015: a review of current American cancer society guidelines and current issues in cancer screening. CA Cancer J Clin. 2015;65:30–54.

207. Youssef O, Sarhadi V, Ehsan H, Böhling T, Carpelan-Holmström M, Koskensalo S, et al. Gene mutations in stool from gastric and colorectal neoplasia patients by next-generation sequencing. World J Gastroenterol. 2017;23:8291–9.

208. Mäki-Nevala S, Knuuttila A, Knuutila S, Sarhadi VK. Concordant Results of Epidermal Growth Factor Receptor Mutation Detection by Real-Time Polymerase Chain Reaction and Ion Torrent Technology in Non-Small Cell Lung Cancer. J Clin Respir Dis Care. 2016;2:107

209. Gormally E, Vineis P, Matullo G, Veglia F, Caboux E, Le Roux E, et al. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. Cancer Res. 2006;66:6871–6.

210. Fernandez-Cuesta L, Perdomo S, Avogbe PH, Leblay N, Delhomme TM, Gaborieau V, et al. Identification of circulating tumor DNA for the early detection of small-cell lung cancer. EBioMedicine. 2016;10:117–23.

211. Olmedillas López S, García-Olmo DC, García-Arranz M, Guadalajara H, Pastor C, García-Olmo D. KRAS G12V mutation detection by droplet digital PCR in circulating cell-free DNA of colorectal cancer patients. Int J Mol Sci. 2016;17:484.

212. Le Calvez-Kelm F, Foll M, Wozniak MB, Delhomme TM, Durand G, Chopard P, et al. KRAS mutations in blood circulating cell-free DNA: a pancreatic cancer case-control. Oncotarget. 2016;7:78827–40.

213. Qin Z, Ljubimov VA, Zhou C, Tong Y, Liang J. Cell-free circulating tumor DNA in cancer. Chin J Cancer. 2016;35:36.

214. Youssef O, Knuuttila A, Piirilä P, Böhling T, Sarhadi V, Knuutila S. Presence of cancer-associated mutations in exhaled breath condensates of healthy individuals by next generation sequencing. Oncotarget. 2017;8:18166–76.

215. Armengol G, Sarhadi VK, Ghanbari R, Doghaei-Moghaddam M, Ansari R, Sotoudeh M, et al. Driver gene mutations in stools of colorectal carcinoma patients detected by targeted next-generation sequencing. J Mol Diagn. 2016;18:471–9.

216. Lamlum H, Papadopoulou A, Ilyas M, Rowan A, Gillet C, Hanby A, et al. APC mutations are sufficient for the growth of early colorectal adenomas. Proc Natl Acad Sci U S A. 2000;97:2225–8.

217. Rowan AJ, Lamlum H, Ilyas M, Wheeler J, Straub J, Papadopoulou A, et al. APC mutations in sporadic colorectal tumors: A mutational “hotspot” and interdependence of the “two hits.” Proc Natl Acad Sci U S A. 2000;97:3352–7.

218. Ghatak S, Chakraborty P, Sarkar SR, Chowdhury B, Bhaumik A, Kumar NS. Novel APC gene mutations associated with protein alteration in diffuse type gastric cancer. BMC Med Genet. 2017;18:61.

219. Fang D-C, Luo Y-H, Yang S-M, Li X-A, Ling X-L, Fang L. Mutation analysis of APC gene in gastric cancer with microsatellite instability. World J Gastroenterol. 2002;8:787–91.

220. Tahara E. Genetic pathways of two types of gastric cancer. IARC Sci Publ. 2004;327–49.

221. Tamura G, Maesawa C, Suzuki Y, Tamada H, Satoh M, Ogasawara S, et al. Mutations of the APC gene occur during early stages of gastric adenoma development. Cancer Res. 1994;54:1149–51.

222. Lyu G-Y, Yeh Y-H, Yeh Y-C, Wang Y-C. Mutation load estimation model as a predictor of the response to cancer immunotherapy. Npj Genomic Med. 2018;3:12.

223. Oberndorfer F, Müllauer L. Molecular pathology of lung cancer: current status and perspectives. Curr Opin Oncol. 2018;30:69–76.

224. Chae YK, Anker JF, Bais P, Namburi S, Giles FJ, Chuang JH. Mutations in DNA repair genes are associated with increased neo-antigen load and activated T cell infiltration in lung adenocarcinoma. Oncotarget. 2018;9:7949–60.

225. Chen CC, Chen KJ, Kong MS, Chang HJ, Huang JL. Alterations in the gut microbiotas of children with food sensitization in early life. Pediatr Allergy Immunol. 2016;27:254–62.

226. Chumpitazi BP, Hollister EB, Oezguen N, Tsai CM, McMeans AR, Luna RA, et al. Gut microbiota influences low fermentable substrate diet efficacy in children with irritable bowel syndrome. Gut Microbes. 2014;5:165–75.

227. Yang J, Martínez I, Walter J, Keshavarzian A, Rose DJ. In vitro characterization of the impact of selected dietary fibers on fecal microbiota composition and short chain fatty acid production. Anaerobe. 2013;23:74–81.

228. Lam YY, Ha CWY, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, et al. Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. PloS One. 2012;7:e34233.

229. Zhao F, Feng J, Li J, Zhao L, Liu Y, Chen H, et al. Alterations of the gut microbiota in Hashimoto's thyroiditis patients. Thyroid. 2018;28:175–86.

230. O'Callaghan A, van Sinderen D. Bifidobacteria and their role as members of the human gut microbiota. Front Microbiol. 2016;7:925.