CORSO DI DOTTORATO DI RICERCA IN
SCIENZE BIOMEDICHE INTEGRATE E BIOETICA
XXXIII CICLO A.A. 2017-2018

ANALYSIS OF TUMOR HETEROGENEITY IN BLOOD AND TISSUE SAMPLES

Coordinatore
CHIAR.MO PROF. **PAOLO POZZILLI**

Tutor
CHIAR.MO PROF. **VITO MICHELE FAZIO**
CHIAR.MA DR.SSA. **CECILIA SIMONELLI**

Ph.D. Candidate:
DR. FRANCESCO PICARDO

INDEX

CHAPTER 1: CANCER HETEROGENEITY

The only certain way
to avoid cancer
is not to be born:
to live is to incur the risk
(Robbins SL)

1.1 CANCER EPIDEMIOLOGY

Tumors in general are growing as the leading cause of death and an important barrier to increasing life expectancy in every country of the world. According to World Health Organization (WHO) estimates, in 2019 (1) cancer represents the 1st or 2nd leading cause of death before the age of 70 years in 112 of 183 countries and ranks 3rd or 4th in a further 23 countries. Cancer's rising prominence as a leading cause of death partly reflects marked declines in mortality rates of stroke and coronary heart disease, relative to cancer, in many countries (2). Overall, the burden of cancer incidence and mortality is rapidly growing world-wide; this reflects both aging and growth of the population as well as changes in the prevalence and distribution of the main risk factors for cancer, several of which are associated with socioeconomic development.

Female breast cancer represents the most commonly diagnosed cancer worldwide, in both sex (11.7% of total cases), followed by lung (11.4%), colorectal (10.0%), prostate (7.3%), and stomach (5.6%) cancers. Lung cancer is the leading cause of cancer death for both sex (18.0% of the total cancer deaths), followed by colorectal (9.4%), liver (8.3%) stomach (7.7%), and female breast (6.9%) cancers. In women, breast cancer is both the most commonly diagnosed cancer and the leading cause of cancer death, followed by colorectal and lung cancer for incidence, whereas by lung and colorectal for mortality. In men, instead, the most frequently occurring cancer and the leading cause of cancer death is represented by lung cancer, followed by prostate and colorectal cancer for incidence and liver and colorectal cancer for mortality (3).

1.2 CANCER HETEROGENEITY DEFINITION

The coexistence of different biological, morphological, phenotypic and genotypic profiles is defined as "tumor heterogeneity", and can be both inter-tumor (between tumors of the same kind) and intra-tumor (within the same tumor). Tumor heterogeneity exists at multiple levels and may be present within different tumor regions or between primary cancer and metastases (spatial heterogeneity), or during the disease progression (temporal heterogeneity). The tumor microenvironment (TME), defined as the complex ecosystem in which cancer cells interact with non-cancerous cells, represents an additional source of intra-tumor heterogeneity. The TME includes proliferating tumor cells, the

tumor stroma, surrounding blood vessels, and immune cells. In particular, the dynamic interplay between cancer and immune cells has become an issue of great interest. There is growing recognition that immunoediting, the process whereby the immune system can both counteract and promote tumor development, contributes to cancer heterogeneity and represents a potential source of biomarkers (4). Heterogeneity can be achieved through selection and expansion of tumor clones, which dynamically evolve with a branched evolution and confer growth advantage within a specific individual microenvironment. Tumor heterogeneity can be explained from an ecological perspective that focuses on interactions of organisms with their environment and among each other (5). When applying an ecological perspective to human tumors, sub clonal populations of tumor cells that differ in heritable traits are considered as distinct "species," whereas infiltrating normal cells, extracellular matrix, vessels, are considered the environment. A major implication of the ecological perspective is that co-existence of phenotypically distinct clonal populations of tumor cells should inevitably lead to the formation of a network of biological interactions, which could be either direct or mediated by the tumor microenvironment. Four key interactions are likely to exist among distinct tumor clones:

1. **Competition**, the strongest and most important biological interaction between tumor cells. The role of competition is strengthened by the limited nature of resources: while, under tissue culture conditions, tumor cells are capable of limitless exponential growth, clonal expansion within tumors is severely constrained by the limited availability of oxygen, nutrients, growth factors, and space. Limited resources intensify competitive interactions between interacting cancer sub clones) (6).
2. **Amensalism**, namely an interaction in which one interacting party is inhibited by the other without being affected itself. As tumors grow under conditions of limited resources, amensalistic interactions can provide a competitive advantage to a clone that can inhibit other clones while being resistant, like the plant "*Juglans nigra"*, whose root secrete a compound that kills or damage surrounding plants. The existence of this type of interaction has also been documented among distinct tumor clonal populations, both in cell culture and *in vivo*, although the exact mechanisms are unclear (7).
3. **Commensalism**, a positive interaction in which one interacting party benefits the other without itself being affected, like what has been shown in a human xenograft model, where an "instigator" tumor cell line augmented the proliferation and metastasis of genetically distinct "indolent" tumors that were incapable of forming macroscopic outgrowths on their own (8).

4. **Mutualism** (cooperation), a positive interaction in which both interacting parties can benefit from each other. Positive interactions between species have been widely documented in natural ecosystems. Mutualistic interactions between distinct tumor clones can play an important role in tumor evolution by maintaining the survival and proliferation of tumor cells until one of the clones achieves a "full deck" of malignant mutations required for clonal dominance and full-blown malignancy (9).

Of note, most of the biological interactions are not mutually exclusive, and the net outcome for the interacting species will depend on the net sum of the different interactions.

1.3 Cancer Heterogeneity-Related Issues

The existence of tumoral heterogeneity brings several issues in the comprehension of the natural history of cancer, which are strictly connected with the actual gold standard for tumor monitoring: the tissue biopsy. Tissue biopsy has proven extremely useful in cancer care, but shows different limitations (10):

- It is a medical procedure that involves risks (hemorrhage, perforation of adjacent structures), may cause stress, pain, and discomfort for the patient, and its performance remains challenging for specific tumor locations that are particularly troublesome to reach (11).
- Samples issues:
    - if a small region of the tumor is taken, it may not be informative about the clonal composition of the tumor as a whole;
    - if the tumor specimen is too large, most methods will only provide an average picture, usually reflecting the dominant clone: a large numbers of cells pooled will quench the "noise" and the signal from minor subpopulations will be lost.

  Furthermore, it is hard to get samples from metastasis sites: spatial heterogeneity remains unsolved.
- As an invasive procedure, it can hardly be repeatedly performed to monitor disease progression and treatment response: temporal heterogeneity remains unsolved.

1.4 Overcome Cancer Heterogeneity

1.4.1 Liquid Biopsy

The term "Liquid biopsy" (liquid biopsy) refers to the sampling and analysis of non-solid biological tissues, in particular bio fluid as blood, but also cerebrospinal fluid (12), urine (13), saliva (14). A blood sample for a liquid biopsy analysis include different circulating bio-markers, among which the

main ones are Circulating Tumor Cells (CTC), Cell-Free DNA (cfDNA), Extravesicles (EV). Although *sensu strictu* the concept of "biopsy" refers only to CTC samples, it has also been extended and accepted with reference to cfDNA and EV.

1. **Circulating Tumor Cells (CTCs)** are cells detached from the primary tumor mass to enter in the systemic circulation through direct extravasation into blood vessels or after transit through the lymphatic system. The first news of the presence of CTCs dates back to a study of 1869, (15) but the use of the same in research protocols dates back to the early 2000s. Studies performed on animal models, to investigate metastatic progression in the liver and lung, show that the least efficient steps in the survival of neoplastic cells are not the earliest ones, which involve survival in the circulation, arrest and escape from the blood vessels. neoplastic cells, but rather the later ones, which concern the failure of the growth of solitary cells and micrometastases: metastatic inefficiency (16). Contrary to the traditional view, which wants metastases as a late event of tumorigenesis, numerous studies show how the dissemination of tumor cells can also occur in early neoplastic lesions (17). These works support a model of metastasis according to which it is the CTCs that detach earlier that remain dormant in the secondary sites, eventually evolving into overt metastasis. According to this model, primary and metastatic cancers progress independently of each other, and maintain different genetic characteristics (18). This model also supports the idea of "self seeding", namely that CTCs can return to the site of the main tumor, increasing the malignancy of the previous one or causing a tumor reappearance (19). Complementary model to this is the one hypothesized by Paget with the theory of "seed and soil". This theory was formulated following the observation that different types of cancer metastasize in an organ-specific way (colon to liver and lung, breast to bone, brain, liver and lung). According to this theory, metastases occur only if the CTC (the "seed") has a specific affinity for the microenvironment ("soil") (20). This specificity can manifest itself as specificity in the surface markers of CTC (21), specificity of the CTC response tissue-specific chemotactic agents, (22) or response to tissue-specific growth conditions (20).One of the most difficult challenges in CTC research concerns their detection, being present in very low numbers, about 1 per 10^{5-7} mononuclear cells (23). The techniques used differ mainly according to the method of separation and identification of the CTC. Separation from peripheral blood can occur in different ways, mainly on an EpCAM-dependent (24) or EpCAM-independent basis (25). Epithelial Cell Adhesion Molecule (Epithelial Cell Adhesion Molecule) is an adhesion molecule present on epithelial cells and absent on other histotypes; this feature made EpCAM as one of the best

marker of CTC identification in carcinomas (26), and has been implemented as a capture antigen in CellSearch® (Menarini Silicon Biosystems, Bologna) (27), gold standard in the automated research of CTC, one and only approved device by the FDA (Food and Drug Administration) in metastatic breast (28), prostate (29) and colorectal cancer (30). CTCs are heterogeneous, and there is ample evidence that some CTCs undergo EMT (Epithelial-to-Mesenchymal Transition), causing acquisition of a mesenchymal phenotype and consequent loss of epithelial characteristics such as the expression of EpCAM, making the methods based on the detection of $EpCAM^+$-CTC useful only for certain types of CTC (31). EpCAM-independent approaches, on the other hand, exploit the differences in physico-morphological characteristics existing between CTC and blood cells. Parsortix™ (ANGLE Plc), for example, is based on the passage of the collected blood through a membrane with pores with a diameter of ≤10 μm, which allow the CTC, whose diameter is greater, to be retained inside (32). Following the separation of cells from whole blood, different methods were used to confirm whether the isolated cells are indeed CTCs. Immunological staining, i.e. with antibodies conjugated with fluorescent markers, is the most commonly used technique to confirm the presence of epithelial or tumor markers on cells. The CellSearch system involves a labeling for CD45 (leukocyte surface marker), cytokeratin (surface markers of epithelial cells), and DAPI (nuclear dye that shows the presence of a nucleus, to differentiate cell debris and apoptotic cells): cells were defined as CTC when they showed positivity for different isoforms of cytokeratin (CK8, 18, 19) and DAPI and negativity for CD45 (33). Other post-enrichment identification strategies involve the use of RT-PCR and FISH (fluorescent in situ hybridization) for the identification of mutations on target genes (34).

2. **Circulating free DNA**: the discovery of the presence of freely circulating DNA not associated with cells (cfDNA) dates back to 1948 (35). The potential applications of cfDNA in cancer diagnosis and prognosis were subsequently demonstrated when elevated cfDNA values were detected in the serum of cancer patients (36). Subsequently, some tumor-associated oncogene mutations, but also a profile of methylation consistent with that of primary cancer have been discovered in DNA from plasma or serum of patients with carcinoma. These findings confirm that a fraction of cfDNA is actually circulating tumor DNA (ctDNA) being derived from cancer cells (37). Numerous studies have shown the presence of cfDNA even in healthy individuals, reporting, however, an average concentration 10 times higher in patients with different types of neoplasia (38). ctDNA can originate following the lysis of CTC or micrometastases (39), due to necrosis process (40), or by active release from live tumor cells

(41). The modality that gives the greatest contribution to the total DNA in circulation seems to be instead the DNA release from apoptotic cells (42), as suggested by the fact that most of the cfDNA fragments have a length of about 180-200bp, typical of the apoptotic process. The concentration of cfDNA is generally low in healthy individuals, as it is rapidly removed from the circulation by phagocytosis (43), but a higher amount of ctDNA can be accumulated in bloodstream of cancer patients. ctDNA can provide the same genetic information of the cell of origin such as point mutations (EGFR, KRAS), arrangements (EML4-ALK), amplifications (HER2, MET), methylation (GALT) and aneuploidy. ctDNA analysis shows some limitations including, due to the modest quantity and high fragmentation of the DNA, the not easy execution of a rapid, efficient and reliable isolation of ctDNA, and often the measurement it is not perfectly reproducible (44, 45). The basis for successful ctDNA detection is the selection of an isolation method that ensures that sufficient ctDNA is extracted (45, 46). Molecular analysis of ctDNA can be used to verify the presence of tumor-specific genetic and epigenetic alterations, linked to the development of resistance to certain drugs or that can drive the use of personalized therapies (46, 47). Liquid biopsy can be used also in screening programs: ctDNA alterations are already present 5 to 14 months before clinical diagnosis in patients who later developed cancer (37). ctDNA analysis can provide information from single mutation to Whole-Genome Amplification (WGA). PCR-based methods (48) have been applied since the early 2000s for the detection of mutations in serum and plasma, and some assays are available as approved kits for clinical use, but still show limited sensitivity. Droplet digital PCR (ddPCR) is a highly sensitive quantitative method, extensively used to quantify ctDNA levels (49). The multiplexing capacity of these assays lies in the different binding affinity between mutant and wild type allele and, since in most cases it requires specific primers and probes for a known and defined mutation, it has limited applicability. The assays are set up to search for a small number of mutations and are often used for the analysis of known mutations.

3. **Extracellular Vesicles**: identification of extracellular vesicles (EV) dates back to 1940, with the discovery in the plasma of the presence of subcellular factors capable of promoting blood clotting Chargaff (50). The discovery that gave the decisive impetus recently concerned the identification of the tissue factor functionally active in the supernatant of monocytes stimulated with lipopolysaccharide (51). The tissue factor, initiator of coagulation, functions only if bound to the cell membrane. This meant that the stimulated cells are able to release biologically active parts of the membrane to which the tissue factor is bound. The further

discovery that EVs contain RNA and microRNA, has increased interest in these particles, identifying them as mediators of intercellular communication (52). EVs and their contents therefore represent a class of potential therapeutic targets, due to their ability of modulating the tissue regeneration and the immune response. EVs are a heterogeneous group of particles that can be mainly divided into three groups: exosomes, microvesicles and apoptotic bodies (53). This classification is made in relation to particle size, lipid composition and density (54):

- ***Exosomes*** are the smallest vesicles, with a diameter between 40-100 nm and generally homogeneous in size. They originate from the formation of multivesicular compartments at the cytoplasmic level, which are released into the extracellular environment by fusion with the plasma membrane (55).
- ***Microvesicles (MV)*** have an irregular shape with a maximum diameter of 1000nm and a density not well determined. MV are released into the extracellular space by budding from plasma membrane, as a consequence of activation mainly due to cellular stress, an increase in intracellular calcium or a decrease in cholesterol levels of the cellular plasma membrane (56).
- ***Apoptotic bodies (AB)*** are the largest vesicles, with a diameter between 50-4000 nm and greater density than the other two classes. AB show an extremely irregular shape and this makes their morphological discrimination very complex. As for the lipid composition, mass spectrometry experiments, in vesicles isolated from different cell lines, have identified the presence of different classes of lipids characterized by phosphoglycerides, sphingolipids and steroids (57). AB are released during the terminal stages of apoptosis and originate from apoptotic cells by fragmentation (58).

The particular lipid composition of each category helps to determine their stiffness and transport efficiency. Moreover, all the vesicles show the presence of residues of phosphatidylserine more or less exposed according to the type of vesicles (59). There is not yet a standard process to separate EVs from biological fluids, nor to count them or to extract nucleic acid or protein from EVs, due to the biological complexity of the medium in which they are contained. The difficulties in vesicle detection and isolation partly explain the differences in classification criteria and exposes one of the main problems that has to be solved to increase EVs use as information tool. In the majority of the studies, EVs are purified using differential centrifugation protocols, in which centrifugal force is used, at increasing accelerations, to settle the particles. A high centrifugal acceleration of 100,000 to 200,000g can cause vesicle fusion and contamination of the pellets with proteins, thus hindering

proteomics studies or co-precipitation together with soluble protein vesicles or free miRNA (60). Furthermore, the functional properties of EVs may change during isolation: centrifugation may increase phosphatidylserine exposure and strengthening the vesicles' ability to promote clotting (61). In addition to ultracentrifuge, other methods of EV separation require the use of dedicated commercial kits according to the type of analysis to be subsequently carried out on the sample.

In summary, liquid biopsy is a minimally invasive procedure, allowing the serial monitoring of the molecular changes that occur along the course of the disease and the evaluation of temporal heterogeneity. Liquid biopsy can also be used to track spatial heterogeneity, because analytes can represent the entire tumor and metastasis, providing a "snapshot" of the status of the disease. In addition, it avoids possible adverse events related to tissue biopsy. The information provided by liquid biopsy offer unique opportunities for understanding the biology of the tumor, its evolution and therapeutic efficacy and resistance to therapy (62).

1.4.2 Single Cell analysis

Single-cell analysis is an alternative to bulk sequencing to explore tumor evolution (63). In theory, sequencing individual cells removes the time bias inherent to bulk sequencing, as all genetic mutations within the sequenced cell should be detectable irrespective of when the mutations arose. Clonal identity also becomes evident, removing the need for allelic copy number correction. However, calling a "single-nucleotide variant" (SNVs) in single-cell sequencing remains challenging owing to the level of noise and missing data. Combining information from multiple cells addresses this issue (64), although at the cost of losing single-cell resolution. By contrast, copy number aberrations (CNAs) can be reliably identified in single cells (65); however, because the background CNA rate is still not well understood, drawing inferences about temporal evolutionary dynamics from these data is not straightforward. Nevertheless, single-cell sequencing offers a powerful route to learning how CNAs accrue, because sequencing individual cells means that some newly born cells can be analyzed before the effects of selection, informing the background CNA mutation rate. Single-cell sequencing of cells from a large cancer risks sequencing many cells that are 'evolutionary dead ends' and would not contribute to future disease progression. Simply sequencing large numbers of cells would abrogate this issue and moreover gives a direct means to detect and characterize negative selection on CNAs (66), which cannot be identified by bulk sequencing.

Other than the research for genomic mutations, single cell can be analyzed to explore transcriptomic, to determine what genes are being expressed in each cell. The transcriptome is often used to quantify

the gene expression instead of the proteome because of the difficulty currently associated with amplifying protein levels (67). There are three major reasons gene expression has been studied in single cell: to study gene dynamics, RNA splicing, and cell typing. Gene dynamics are usually studied to determine what changes in gene expression effect different cell characteristics. RNA splicing studies are focused on understanding the regulation of different transcript isoforms. Cell typing represents the identification of a single cell type based on genes expressed in the cell. The main goal in cell typing is to determine the identity of cells that don't have known genetic markers (67).

In addition, the proteome can be studied from single cell, to better understand the activity of cells at the single cells level. Understanding the proteome of single cell may provide the best understanding of how a cell operates, and how gene expression changes in a cell due to different environmental stimuli. Respect transcriptomic, that can be easier than proteomics, the latter has more accuracy in determining the actual kind of protein expressed, taking into account post-transcriptional regulation (68). Transcriptomic is anyway still important to study the difference between RNA levels and protein levels and therefore provide insight on which genes are post-transcriptionally regulated.

Single-cell technologies have advanced rapidly in the past several years. Currently available protocols vary in cell capture method, library preparation chemistry and throughput. Most protocols require single-cell suspension, so the first critical consideration is optimizing tumor dissociation to generate a cell suspension that is fully representative of the intact tumor in terms of cell populations, their frequencies and expression programs.

Among different devices for single cell capture, the DEPArray™ (Menarini Silicon Biosystems) have been used to identify different singe cell population from FFPE samples (69) or CTCs (70). The DEPArray is based on dielectrophoresis (DEP), an electrokinetic principle by which, creating a non-uniform electric field, it is possible to exert forces on neutral particles such as cells. The electric field is created inside a microfluidic chip (DEPArray™ cartridge) containing cells suspended in a liquid, where it forms DEP "cages" around cells trapping them. The DEPArray™ is equipped with a six-channel fluorescent microscope and a CCD camera that captures images and identifies cells by their fluorescence labelling and morphological characteristics. The instrument offers the possibility to analyze the collected images with the CellBrowser™ software, that allows for multiple parameters from fluorescence and bright field images for cell selection, and recover them in a tube for further analyses, through opening and closing the DEP cages.

The DEPArray system allow single cell isolation for different further purpose, such as Whole Genome Amplification or DNA methylation.

CHAPTER 2: EPIGENETICS

2.1 History and Definition

The term "epigenetics" derives from the union of the Greek suffix "επι" (with the meaning of "beyond", "above") and "γενετικός" (which means "relative to birth", here with reference especially to genetics), assuming the meaning of "what is beyond genetics". The concept of epigenetics, and the definition itself, have changed several times since, in the late 1930s, Conrad Hal Waddington, referring above all to its implications in embryonic development, proposed the first: "the causal interactions between genes and their products that lead the phenotype to manifest"(71). With the discovery of the implications of epigenetics in various biological processes, the previous definition has been made more complete by the one from Shikhar Sharma: "the study of heritable variations in gene expression that occurs independently of changes in the primary DNA structure."(72)

2.2 Epigenetics Mechanisms

Chromatin is the form in which nucleic acid are found in eukaryotic cells. Chromatin is made up of repeating units of nucleosomes that consist of approximately 146 base pairs of DNA wrapped around an octameric core of four histone proteins (H3, H4, H2A and H2B)(73). Protein H1, on the other hand, binds to each nucleosome by coming into contact with the DNA and proteins of the core by changing the path of the DNA itself that emerges from the nucleosome.

Three primary epigenetic codes have been well studied, and are listed below:

- DNA methylation
- Histone modifications
- Non-coding RNA (ncRNAs)

All these mechanisms work together to regulate the functioning of the genome by altering the dynamic structure of chromatin, particularly through the regulation of its accessibility to transcription factors and compactness. The interaction of these mechanisms creates an epigenetic substrate that regulates how the genome manifests itself in different cell types, in the different stages of development, and in different stages of various diseases, including cancer (74).

2.2.1 DNA Methylation

DNA methylation is the most important epigenetic mechanism that has been intensively investigated. There are different DNA methylation modifications, such as 5-methylcytosine (5mC), N6-methyladenine (6mA) and 4-methylcytosine (4mC)(75). While 6mA and 4mC are commonly found

in prokaryotic genome, 5mC is the most widely distributed methylation type in eukaryotes, and the most studied and understood DNA modification pattern overall (76). There are many conventional techniques to analyze the approximate or exact methylation contents of DNA. Bisulfite modification is the foundation for the majority of DNA methylation assays, which converts cytosine to uracil in single-stranded DNA but does not affect 5mC (77).

DNA methylation results in the stable silencing of a gene, and has a very important role in regulating gene expression and chromatin architecture, in association with changes in associated histone or non-histone proteins. DNA methylation occurs mainly through addition of a methyl group (CH_3) to a cytosine contained within CG dinucleotides series, which are not uniformly distributed in the genome, but are concentrated in small tracts of CpG-rich DNA, called "CpG islands" and in regions of widely repeated sequences (centromeric repeats, retrotraceable elements, rDNA) (78). CpG islands are preferentially located at the 5' end of a gene and occupy about 60% of the promoter sequence (79). Most CpG islands remain unmethylated during development and in differentiated tissues, 45 although some are, resulting in long-term silencing. The inactivation of one of the two X chromosomes in women and the existence of imprinted genes are classic examples of the existence of methylated CpG islands during development (80). Repetitive genomic sequences scattered throughout the genome are heavily methylated to prevent the occurrence of chromosomal instability by silencing non-coding DNA and transposable elements (81). DNA methylation leads to gene silencing by both preventing (blocking the access of transcription factors to the target site) and promoting (providing binding sites for proteins with methyl-binding domains, which can mediate gene repression by interacting with histone deacetylases) the recruitment of DNA regulatory proteins.

The DNA methylation pattern in the genome is generated and inherited through the coordinated activity of *de novo* DNA methyltransferases (DNMT3A and 3B), which act independently of replication and show equal affinity for both unmethylated and hemimethylated DNA (methylated only on a strand), and is kept by maintenance DNA methyltransferase (DNMT1), which acts during replication mainly by methylating the hemimethylated DNA (81).

2.2.2 Histone modifications

Histone proteins, which include the nucleosomal core, have a c-terminal domain with a globular structure and an unstructured N-terminal one (74). The N-terminal tail can undergo numerous post-transcriptional covalent modifications, such as methylation, acetylation, ubiquitination, SUMOylation, and phosphorylation on specific residues. These modifications regulate critical cellular processes such as transcription, replication and repair (82). The various modifications serve

to preserve the epigenetic memory in a cell in the form of a "histone code" that determines the structure and activity of the various chromatin regions (83).

Histone modifications work both by modifying the accessibility to chromatin and by recruiting or inhibiting the non-histone effector proteins that decode the message contained in the pattern of modifications.

The mechanism by which the histone code is inherited is still far from being fully understood. Unlike DNA methylation, histone modifications can lead to both activation and repression of transcription, based on the type of modification present and where it occurs: acetylation of lysine correlates with transcriptional activation, while methylation of lysine can lead to both activation (such as trimethylation of lysine 4 on histone 3 - H3K4me3) (84) and repression (trimethylation of lysine 9 or 27 on histone 3 - H3K9me3 and H3K27me3) (83).

Specific patterns of histone modifications are present in different cell types and appear to play a key role in determining cell identity, such as embryonic stem cells have divalent domains consisting of activating (H3K4me3) and inhibitory (H3K27me3) modifications, which serve to maintain cellular plasticity; instead the differentiated cells lose this ambivalence and show a more rigid chromatin structure (85).

Histone modifications are regulated by histone acetyltransferases (HAT) and histone methyltransferases (HMT) which add acetyl and methyl groups respectively, whereas deacetylases (HDAC) and histone demethylases (HDM) remove them.

The DNA and histone methylation processes interact with each other to better determine the state of greater or lesser repression of a gene; for example, several HMTs can drive DNA methylation either by directly recruiting DNMTs (86) or by stabilizing DNMTs themselves (87). In turn, DNMTs can recruit HDACs and HMTs to increase the transcriptional repressive state of chromatin (88).

2.2.3 Non coding RNA

Only 2% of the genes contained in the human genome encode proteins, while the remaining 98% are non-coding DNA, initially thought to be "junk DNA" (89). With the advancement of knowledge about the genome, this DNA, transcribed below form of non-protein coding RNA (ncRNA), has been found to be responsible for regulating and modulating the expression of numerous "protein coding" genes. Among the ncRNAs with a regulatory function, there are:

- miRNA (miRNA): small RNA molecules, of about 19-25 nt, which target and bind to a specific sequence of 8 nucleotides ("seed" region) in mRNA and inhibit its translation;

- Piwi-interacting RNA (piRNA): the most expressed ncRNA class in vertebrates. piRNA typically range from 25 to 33nt, and play a key role in spermatogenesis, protecting germ cells from the action of transposons by selectively silencing them (90). piRNAs are found in conserved clusters across the genome, although the piRNAs themselves all internal of these clusters are not. A single cluster can code for a number of individual piRNAs ranging from ten to over a thousand.
- Small nuclear RNA (snoRNA): they are usually 60-200nt large, but only small portions participate in the recognition of the target mediated by antisense interactions. In mammals, they originate from introns of pre-mRNA transcripts. SnoRNAs cause a wide range of post-transcriptional modifications of ribosomal RNA (rRNA) and transfer RNA (tRNA) (91).
- Long non coding RNA (lncRNA): the longest class of ncRNA, they are composed, on average, of at least 200 nt. These transcripts are capped and polyadenylated suggesting their potential role within the cell. Overall, there are five coding modes for lncRNAs:(92)
    - Transcription from protein coding genes: for example, exon and promoter of the XIST gene (X Inactive-Specific Transcript), originate from Lnx3; (93)
    - Gene recombination
    - As an effect of a retrotransposition that generates both a functional transcription gene and a non-functional pseudogene
    - Sequences repeated in tandem, such as the transcripts of XIST and Kcnq1ot1
    - Transposable elements are inserted within the genes

  LncRNAs are involved in chromatin remodeling, in direct control of both transcription and post-transcription.

2.2.4 Epigenetic in cancer

The development of a tumor may found its origins on genetic and epigenetic foundations. Clonal genetic alterations are common in tumors and account for genetic changes in both blood (especially chromosomal rearrangements) and solid cancers (mutations in gatekeepers genes). Epigenetic alterations, on the other hand, are ubiquitous and function as "surrogates" of genetic alterations (activation of oncogenes, silencing of tumor suppressors), mimicking their effects, as happens for example in silencing/hypermethylation of SFRP (Secreted Frizzled-Related Protein) gatekeepers in CRC (94) or VHL (Von Hippel-Lindau protein) in renal cell carcinomas (95), or BCL2 activation / hypomethylation (B-cell CLL / lymphoma 2) in B-cell leukemia (96). Epigenetic changes occur very early in the tumorigenesis process, and are found in healthy tissues before cancer arises. Epigenetic

alterations in stem progenitors play a decisive role not only in predicting the risk of cancer, but also in the progression and final heterogeneity of the tumors that arise from these cells: epigenetic alterations set the stage for genetic alterations. Furthermore, early epigenetic variations may explain most of the heterogenic properties commonly associated with tumor growth, invasive and metastatic phenotype.

The epigenetic-stem model, therefore, predicts the onset of the tumor in a three-phasic process:

- Genetic alterations in progenitor cells lead to a polyclonal population of cells able to become a neoplasm. These cells represent the main target of environmental, genetic and age-dependent exposure. Epigenetic alterations can alter the normal balance between undifferentiated progenitors and committed cells within a given anatomical compartment, both in number and in their aberrant differentiation capacity
- A mutation within the subpopulation of epigenetically altered progenitors in the early stages of a tumor can be considered the first step towards the formation of a tumor (97). These mutations are tumor specific, such as the mutations of APC / β-catenin in CRCs or chromosomal rearrangement of BCR – ABL in CML. As mentioned above, an epigenetic alteration can replace a genetic mutation with the same meaning.
- Genetic-epigenetic plasticity is the ability to modify the phenotype and transmit it stably, both on a genetic and epigenetic basis. In many cases, the basis underlying the increase in genetic plasticity are well known, as happens, for example, in the case of telomerases. These enzymes are expressed only in stem cells and are used to synthesize telomeres, the structures in the terminal 5 'region of chromosomes, which shorten with each cell division in the absence of telomerases. Telomere erosion reaches the point where the 5' end can no longer be protected by a polyadenosine cap and a dicentric chromosome is created which interrupts the anaphase, blocking cell replication. In tumors, however, telomerase synthesis is re-established and telomeres no longer shorten, resulting in perpetual cell division (94).

CHAPTER 3: β-1,4-GALACTOSYLTRANSFERASE

β1,4-galactosyltransferase (β4GALT1) is a type II membrane-bound glycoprotein involved in several functions, among which, of particular interests, is the direct interaction with EGFR. Particularly, β4GALT1 interacts directly with Epithelial Growth Factor Receptor (EGFR), inhibiting the dimerization of the receptor and the tyrosine phosphorylation and its activation in human hepatocellular carcinoma cells, suggesting an inhibitory role of β4GALT1 in EGFR signaling pathway. β4GALT1 has been demonstrated to be specifically downregulated, due to an hypermethylation of the promoter in different tumors, such as in colon cancer.

3.1 β4GALT1 STRUCTURE AND TRANSCRIPTION

The β4GALT1 gene belongs to a family of 7 genes that code for a family of proteins involved in galactosylation mechanisms. These genes encode membrane-bound glycoproteins (98). Due to the similarity between the different strutture-4-GALT protein structures, they are divided into four groups: β4GALT1 and β4GALT2 belong to the first group, β4GALT3 and β4GALT4 to the second group, to the third β4GALT5 and β4GALT6, and to the fourth group β4GALT7. The main function of these enzymes is to add galactose to an N-acetylglucosamine. However, β4GALT1 is unique in the family, as it encodes an enzyme that is involved in both glycoconjugation and lactose biosynthesis. For glycoconjugation, β4GALT1 adds galactose to an N-acetylglucosamine residue, which is both a monosaccharide and the non-reducing end of a glycoprotein chain of carbohydrates. The biosynthesis of lactose, on the other hand, is restricted to the tissues of the breast-feeding mammary gland. Two different isoform of β4GALT1 are responsible of the different functions. The two enzymatic forms are the result of an alternation of the transcription start site and post-translational changes. The transcription of the longest mRNA (4.1 Kb) begins upstream of the two ATG codons and encodes a protein of 399 amino acids (Long form). The transcription of the shortest RNA (3.9 Kb), which is also the most abundant one, starts between the two ATG codons and encodes a protein of 386 amino acids (Short form). The only difference between the two protein forms is represented by an extension of 13 amino acids present in the long form at the N-terminal end compared to the short form. The longer transcript encodes the protein involved in glycoconjugation, the shorter one encodes the protein involved in lactose synthesis. The β4GalT1 gene contains two promoter sequences with numerous putative regulatory sequences, such as the response elements to SP1, AP1, glucocorticoids and CTF / NF1. Positive regulatory regions acting in -cis are located distal and proximal to the transcription initiation sites of Long and Short transcripts, while negative regulatory regions acting in-cis repress transcription from the proximal site. In addition to the regulation of the transcription, it

has also been hypothesized that the Long form can be translated by virtue of a secondary structure stable at 5' of the transcript (99).

3.2 β4GALT1 FUNCTIONS

The diverse biological functions of surface β4GALT1 indicate that it is capable of interacting with multiple extracellular ligands or other intracellular proteins. On the cell surface, the catalytic region of β4GALT1 is exposed. Therefore surface β4GALT1 can recognize and bind to terminal GlcNAc residues, which are found in a variety of extracellular and cell surface glycoproteins and glycolipids, such as laminin1, zona glycoprotein ZP3 and E-cadherin (100).

The β4GALT1 molecule present on the cell surface contributes to the metastatic and invasive phenotype. β4GALT1 causes a rapid induction of the activity of caspase-3, one of the effector proteins of the apoptosis pathway, and stimulates the cut of PARP. The release of cytochrome C from mitochondria also increases thanks to the action of β4GALT1.

3.3 β4GALT1 INTERACTIONS WITH EGFR

EGFR is a transmembrane glycoprotein, part of the tyrosin kinase receptor, and is involved in apoptosis inhibition (trough activation of Pathway PI3K/AKT) and to cell cycle regulation (trough activation of MAP-chinasi Pathway). Tang et al. showed that β4GALT1 interact with EGFR both in vitro and in vivo (101). The interaction between β4GALT1 and EGFR occurs through the [3H] galactosylation of EGFR in the presence of UDP [3H] galactose. This association leads to a marked reduction of the autophosphorylation of the Tyr1068 residue, which inhibits the phosphorylation and activity of EGFR, as also shown, inversely, by the fact that a downregulation of β4GALT1 increases the activity of EGFR (102). Phosphorylation of the Tyr1045 residue of EGFR, which leads to ubiquitination and degradation of the receptor, however, undergoes no significant changes (103).

Tyr1068 residue is one of the major EGFR autophosphorylation sites, as it provides a direct binding site for Grb2, which, together with Sos, activates Ras through the exchange of a GDP for a GTP, and, in cascade, the activation of the effectors of the MAPkinase cascade and PKB / AKT pathways (104). Both pathways are involved in promoting cell survival, and lead to inactivation of Bad and Bax, key proteins of apoptosis, through serine phosphorylation mediated by PKB/Akt or ERK. (105).

Bad and Bax are present in the cytoplasm of healthy cells but translocate to the mitochondria to allow the release of cytochrome C following proapoptotic stimuli. The release of cytochrome C into the cytosol activates caspase 3, with the consequent irreversible triggering of cellular apoptosis. The

overexpression of β4GALT1 causes a marked increase in the levels of mitochondrial Bax and Bad and, therefore, of apoptosis.

CHAPTER 4: OBJECTIVES

In the perspective to better understanding intra tumor heterogeneity, both in solid and liquid compartment, we first aimed to analyze methylation level of β4GALT1 in primary and metastatic tissue, and successively in blood. The hypothesis of finding the same epigenetic alterations present in the primary tumor in ctDNA and in CTC is confirmed by several previous studies. In particular, β4GALT1 has shown a good frequency of methylation and a 54% sensitivity and a 91.7% specificity (106). Furthermore, β4GALT1 showed a 0% of methylation in normal mucosa, in addition to the important role in regulation and inhibition of the EGFR pathway and its possible implications for target therapy. In addition to our cohort, we have analyzed other cohorts from five GEO public data repositories.

The second part of the project unravels from the concept of tumor heterogeneity in primitive tissue, namely the evaluation of tumor heterogeneity in samples of formalin-preserved primitive tissue (FFPE). In order to evaluate the heterogeneity at the single-cell level through the use of DEPArray™ and Ampli1 (Menarini Silicon Biosystems), samples of different carcinomas have been obtained from University of Siena. The DNA extracted from tumoral tissue, resulting from FFPE samples collected in different years from 2015 to 2020, will be analyzed using the "DEPArray QC kit" to determine the quality and integrity of the DNA. This latter work has been planned in two steps: the first one had the purpose to define and identify the feasibility of retrospective studies with Low Pass analysis on Single Cell from FFPE samples, thus allowing the access to single cell data of patients who already has a clinical follow up. Sample characteristics, such as year of preparation, size of the sample, type of the tumor, preservation, have been collected and correlated to QC score. FFPE features may also be used to indicate which are the settings where the Low Pass on Single Cell may be performed at its best. The second step consisted in analyze single cell from FFPE samples with adequate QC score sorted with DEPArray, to define single cell features. This second step has been postponed due to the spread of Covid-19 pandemic.

CHAPTER 5: MATHERIALS AND METHODS

5.1 PATIENTS ENROLMENT

109 patients with metastatic CRC have been evaluated for the analysis of β4GALT1 in tissue and blood from different health facilities (Università "Campus Bio-Medico" (UCBM) – Roma, IT; IRCCS "Casa Sollievo della Sofferenza" (CSS) – San Giovanni Rotondo (FG), IT; Complejo Hospitalario Universitario de Santiago de Compostela (CHUS), ES. CHUS has provided two cohorts that have been evaluated independently) and a total of 102 FFPE samples of different tumor from University of Siena (UNISI) have been analyzed for the evaluation of the DEPArray™ FFPE QC Kit (Menarini Silicon Biosystems).

All the CRC solid tissue for β4GALT1 evaluation (tumoral, matched normal and metastatic) of all the cohorts derived from FFPE tissue. The anamnestic and pathologic data of the patients included pTNM classification, tumor grading, histologic subtype, intra colic localization, metastasis site and stage at diagnosis. All the materials used have been collected accordingly to Ethical Committee guidelines of the respective institution.

As negative control for ctDNA, has been used 22 samples of cfDNA obtained from plasma of healthy donor from CHUS (10 samples) and from a collaboration with National Institute for Infectious Disease "Lazzaro Spallanzani" (10 samples) with same procedure of patients' blood drawn.

5.2 DNA EXTRACTION PROCEDURE FOR FFPE SAMPLES

The solid tissues have been included in paraffin. From each paraffin block, one 5µm-slide was prepared and colored with hematoxylin and eosin to identify the quantity of tumor cell and highlight the tumor area:

- Tumor with >70% of tumor cell were directly used for DNA extraction
- Tumor with <70% of tumor cell have been microdissected and scraped to enrich tumor cells

Additional 10 µm sections have been cut for DNA extraction, with the following procedure:

1. Paraffin tissue or scraping-obtained tissue were added to an Eppendorf tube with 1mL of xylene
2. Pipetting to dissolve the tissue and gently vortex
3. Cover the tubes with parafilm and incubate at 65°C for 15', then gently vortex for 30''
4. Take off parafilm and centrifuge 5' at 13000 rpm.
5. Discard supernatant preserving the pellet
6. Repeat point 1 - 5
7. Discard supernatant

8. Add 1 ml of 100% EtOH
9. Vortex 30"
10. Centrifuge 5' at 13000 rpm e remove supernatant
11. Dry the pellet 15-20'
12. Add 333µl of a solution of 270µl TE-9, 33µl 10X SDS/PK, 30µl di PKa;
13. Overnight incubation at 48°C, adding 15-20µl PK if needed;
14. Prepare a new 2ml Eppendorf tube with 300µl of phenol-chloroform (Sigma-Aldrich) and add the sample
15. Vortex 20'' and centrifuge at 13000 20';
16. Prepare a new 2ml Eppendorf tube with 300µl of phenol-chloroform (Sigma-Aldrich) and add the aqueous phase of the sample
17. Vortex 20'' and centrifuge at 13000 20'
18. Prepare a new 1,5 Eppendorf tube including:
    - Aqueous phase of the sample
    - 150µl of Ammonium Acetate 7,5 M
    - 1ml of cold EtOH 100%
19. Vortex 20'' and centrifuge at 13000 60' at 4° C
20. Discard supernatant
21. Wash the pellet with 1 ml EtOH 70% and centrifuge 10' a 13000rpm 4° C
22. Discard supernatant
23. Centrifuge 2' 13000rpm to remove all residual EtOH
24. Discard supernatant and dry the pellet
25. Resuspend in 22µl of LoTE buffer and store at -20°C

5.3 DNA EXTRACTION PROCEDURE FOR PLASMA SAMPLES

The protocol used for plasma DNA extraction was the following:

1. Add to 1ml of plasma:
    - 50µl (in proportion to plasma volume) SDS al 25%
    - 30µl (in proportion to plasma volume) PK (20mg/ml)
2. Incubate at 55°C for 16h
3. Add 1mL (in proportion to plasma volume) of PC8 and gently vortexing
4. Centrifuge Maxtract high density Column (Qiagen®) for 1-2' 1500g
5. Transfer the sample solution in the column and centrifuge at 3500g room temperature for 10'

6. Transfer the supernatant in a new 2ml Eppendorf tube and add PC8 in proportion 1:1 with the supernatant and vortex
7. Transfer the solution in the previously prepared Maxtract high density Column and centrifuge 10' 3500g at room temperature
8. Transfer the supernatant in a new 2ml Eppendorf tube and add EtOH and ammonium acetate in proportion 1:10 with the supernatant
9. Incubate O.N. -20°C
10. Centrifuge 35' a 5200g a 4°C
11. Discard supernatant and dry the pellet
12. Resuspend DNA in 50µl LoTE and store at -20°C

5.4 CTC EXTRACTION PROCEDURE

The CTCs have been captured with Dynabeads® (Thermofisher) and separated from blood at CHUS, with the following procedures and then sent to our laboratory for DNA extraction (procedure was the same of FFPE tissue). The Dynabeads® protocol used was the following:

1. Add 250µl Dynabeads® at 5 mL of whole blood
2. Incubate at 30' 2-8°C gently mixing the tube
3. Insert the tube in the magnet for 2'
4. While the tube is in the magnet, carefully discard the supernatant
5. Remove the tube from the magnet and add 5mL of Buffer 1
6. Vortex 2–3"
7. Insert the tube in the magnet for 2'
8. Repeat the steps 4-7 twice
9. Resuspend the cell bond to the beads in 200µL of Buffer 3 preheated at 37°C
10. Add 4µL of Release Buffer
11. Incubate 15' at room temperature gently mixing the tube
12. Pipetting with a 100-200µL tip to the bottom of the tube for at least 5–10 times to maximize cellular release, avoiding foam formation
13. Insert the tube in the magnet for 2' and transfer the supernatant with the released cells in a preheated (37°C) tube
14. Resuspend the fraction with the beads in 200µl of Buffer 3
15. Repeat once the steps 12–14
16. The cell are now beads-free and ready to be frozen at -80 or for subsequent applications.

5.5 DEPARRAY QC SCORE ANALYSIS IN FFPE SAMPLES

The "QC score" obtained with the "DEPArray™ QC kit" (Menarini Silicon Biosystems) is a "quality control (QC)" of the DNA integrity of an FFPE sample. The QC score value range from 0 to 1, with 1 representing the most intact DNA. This score allows predicting the outcome of NGS library preparation, providing a useful tool to infer the minimum recommended number of cells to obtain a given NGS performance level. The assessment consists in a qPCR amplification of two DNA fragment of different length (long amplicon of 32bp and a short amplicon of 54bp) at serial dilution to generate a standard curve. The score is then calculated as $1/2^{\Delta Cq}$, where ΔCq = mean Cq obtained by long amplicon assay – mean Cq obtained by short amplicon assay. The qPCR reaction has been prepared according to manufacturer's instructions.

5.6 SODIUM BISULFITE TREATMENT

DNA methylation is one of the most studied epigenetic phenomena, which has, as its final effect, the non-transcription of the methylated gene. It almost exclusively affects the cytosine residues contained within the CG-rich areas (known as CpG island). They are usually collected in areas of the promoter with a gene expression regulatory function.

The incubation of DNA with Sodium Bisulfite determines a conversion of the non-methylated cytosine residues into uracil, leaving the methylated cytosine residues unchanged, allowing to have a different DNA sequence (107): a random sequence of non methylated DNA, such as "N-**C**-G-N-**C**-G-N-**C**-G-N" after bisulfite treatment, will become "N-**U**-G-N-**U**-G-N-**U**-G-N"; the same sequence "N-**C**-G-N-**C**-G-N-**C**-G-N", with methylated cytosine, would remain unmodified after bisulfite treatment.

Bisulfite treatment has been performed with EpiTect® Kit (Qiagen), and the protocol used was the following:

1. Measure DNA concentration with Nanodrop®
2. Prepare new eppendorf 0.2ml tubes with:
    - 1,5µg of DNA and up to 20µl with RNAsi Free water
    - 85 µl of Bisulfite solution mix;
    - 35 µl of Protect Buffer.

3. Insert the tube in a thermal cycler with the following reaction condition::

Step	Duration	Temperature
Denaturation	5'	95° C
Incubation	25'	60° C
Denaturation	5'	95° C
Incubation	85'	60° C
Denaturation	5'	95° C
Incubation	175'	60° C
Hold	indefinite	20° C

4. Briefly centrifuge and transfer the solution in 1,5ml tube
5. Add 310µl of BL buffer (loading buffer) containing 10µg/ml of RNA carrier and vortex
6. Add 250µl of EtOH 100% and vortex
7. Centrifuge 1' 12000rpm
8. Transfer the solution in the Epitect column
9. Centrifuge 1' 12000rpm and discard the outflow liquid
10. Add 500µl of BW buffer (washing buffer) to the column
11. Centrifuge 1' 12000rpm and discard the outflow liquid
12. Add 500µl of BD buffer (desulphonation buffer) in the column
13. Incubate 15' at room temperature
14. Centrifuge 1' 12000rpm and discard the outflow liquid
15. Add 500µl of BW buffer (washing buffer) to the column
16. Centrifuge 1' 12000rpm and discard the outflow liquid
17. Repeat once steps 15-16
18. Insert the column in a new 2 ml tube
19. Centrifuge 1' 12000rpm and discard the outflow liquid
20. Insert the column in a new 1,5 ml Eppendorf tube
21. Add 22µl of EB buffer (elution buffer) in the center of the column membrane
22. Incubate 2' at room temperature
23. Centrifuge 1' 12000 rpm
24. Store the tubes at -20°C o -80°C

5.7 METHYLATION ANALYSIS

Based on the type of starting sample, two different methodologies have been used for the determination of methylation levels of β4GALT1, which are: the qMSP for DNA obtained from solid tissue (primary tumor, normal, or metastatic) and the ddQMSP, developed by us, for DNA obtained from plasma and CTCs

5.7.1. Quantitative Methylation Specific PCR

For the identification of methylated DNA in solid tissue (primary tumor, normal, or metastatic), we used qMSP (Quantitative Methylation Specific PCR). The qMSP is a technique based on Real Time PCR, specific, however, for methylated DNA sequences and exploits Taqman technology (108). In qMSP, bisulfite-treated DNA is used as a substrate for the amplification in separate reactions of a methylated target promoter and a reference sequence (β-actin – ACTB - gene).
The reactions are conducted with respect to a predetermined threshold, based on the construction of standard curves deriving from serial dilutions of methylated DNA in vitro and subjected to treatment with bisulfite with the same protocol described above used for the samples. The resulting fluorescence is converted into numerical data related to a predetermined standard curve, and calculating the range of the value of the target promoter with respect to the reference sequence. This number is an absolute value in fluorescence units and theoretically reflects the value of a methylated sequence contained in a given sample. This number can vary over a wide range depending on the prevalence of methylated CpG in the target sequence. The reference sequence is used to normalize DNA samples and determine the quality of the bisulfite conversion.[98]

The QMSP has been set up for β4GALT1 and ACTB gene, and the following couple of primers and fluorogenic probe have been used:

	β4GALT1	**ACTB**
Forward primer (5'-3')	TAGGAAACGGGTTTCGACG	TGGTGATGGAGGAGGTTTAGTAAGT
Reverse primer (3'-5')	CGGTAAAGAAAGTGGACGG	AACCAATAAAACCTACTCCTCCCTTAA
Probe (5'-3')	CGTTAAACAACGAAATCCAACCGAA	ACCACCACCCAACACACAATAACAAACACA

The amplification reaction has been performed in triplicate in a volume of 25µL constituted by:

- 600nM of each primer
- 200nM of probe
- 0.75 units of *Platinum Taq polymerase* (Invitrogen, Carlsbad, Ca)
- 200µM of each dNTPs (dATP, dGTP, dCTP, dTTP)

- 16,6 mM of Ammonium Sulphate
- 67mM of Trizma
- 6,7mM of MgCl2
- 10mM of Mercaptoethanol
- 0,1% of DMSO

The amplification reaction has been obtained in a plate of 384 well in a *fast 7900 real-time* (Applied Biosystem, Foster City, CA). The PCR conditions were the following:

- one step at 95°C 2'
- 40 cycles at 95°C 15'' and 60° C for 1'
- hold at 98 10'

A calibration curve for the β4GALT1 gene, patient DNA samples, positive controls (CpGenomeTM Universal Methylated DNA, Serologicals Corp., Norcross, GA), and multiple negative controls were included in each plate. The relative level of methylated DNA was determined by the ratio of "Quantity Mean" β4GALT1 / "Quantity Mean" β-actin multiplied by 1000 for a simpler tabulation

5.7.2. Droplet digital quantitative methylation specific PCR

For the identification of methylated DNA in plasma and CTCs, we used the " droplet digital quantitative methylation specific PCR" (ddQMSP).

The droplet digital PCR (ddPCR™) (BioRad) system allows an extremely sensitive and precise analysis of nucleic acids, allowing the detection of very rare events that differ even for a single nucleotide. This system provides absolute quantification of target DNA molecules in the presence of a high wild type DNA background. We have developed a specific assay for the analysis of methylation in digital PCR which we have called "quantitative methylation specific droplet digital PCR"; in this assay we exploited the possibility of considering the transformation of unmethylated cytosine into thymine as if it were a mutation. The analysis platform used, the QX100, consists of two components, the droplet generator and the droplet reader. The Droplet generator has the function of dividing the sample, containing the DNA to be amplified, into 20,000 droplets of nanoliter size. The division has the function of reducing the competition between the target DNA and the wild type DNA, increasing the specificity and sensitivity of the analysis. In this way, the relative abundance of methylated DNA compared to the wild type increases. The number of target ctDNA molecules present can be calculated from the fraction of positive reactions at the end-point, using the Poisson statistic according to the equation: $\lambda = -\ln(1-p)$, where "λ" is the mean number of target DNA molecules replicated per reaction and "p" is the fraction of positive reactions at the end-point. Using λ, together with the volume of each PCR replication and the total number of replications tested, it is therefore

possible to calculate an absolute estimate of the amount of target DNA present. The DNA molecules present inside the drops thus generated are independently amplified in a standard thermal cycler with the same thermal profile used for qMSP. After amplification, the plate with the amplified drops is placed in the QX100 droplet reader, which reads each well of the plate individually. The "QuantaSoft" software provides a complete set of tools for analyzing results. Negative drops (without target DNA and / or reference) and positive drops (with methylated DNA) are counted. The fraction of positive drops in a sample determines the concentration of the target DNA expressed in "copies / µl". This system has a sensitivity of 0.01%. The analysis with the ddQMSP were carried out in collaboration with the BioClarma company of Turin

5.8 EXTERNAL DATABASES

In this thesis, we have also evaluated six different cohorts from two different databases, the "Gene Expression Omnibus" (GEO) and "The Cancer Genome Atlas" (TCGA).

5.8.1. Gene Expression Omnibus

GEO is a database for Gene expression profiling and RNA methylation profiling managed by the National Center for Biotechnology Information (NCBI). These High-throughput screening genomics data are derived from Microarray or RNA-Seq experimental data. These data need to conform to the Minimum information about a microarray experiment (MIAME) format. The GEO DataSets database stores original submitter-supplied records (Series, Samples and Platforms) as well as curated DataSets that form the basis of GEO's advanced data display and analysis features, including tools to identify differences in gene expression levels and cluster heat maps. GEO Profiles are derived from GEO DataSets. The GEO DataSets database can be searched using many different attributes including keywords, organism, DataSet type and authors.

The GEO cohorts are the following:

- Cohort 1: stage I–III CRC patients (n = 226, GEO accession number GSE14333) (109).
- Cohort 2: stage II–III CRC patients (n = 130, GEO accession number GSE37892) (110).
- Cohort 3: stage I–IV CRC patients (n = 557, GEO accession number GSE39582) (110, 111)
- Cohort 4: stage I–III patients only (n = 125(112), GEO accession number GSE41258) (113)
- Cohort 5: patients with refractory mCRC (n = 80, GEO accession number GSE5851) (114).

5.8.2. The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA) is a project, begun in 2005, to catalogue genetic mutations responsible for cancer, using genome sequencing and bioinformatics. TCGA applies high-throughput

genome analysis techniques to improve the ability to diagnose, treat, and prevent cancer through a better understanding of the genetic basis of this disease.

TCGA is supervised by the National Cancer Institute's Center for Cancer Genomics and the National Human Genome Research Institute funded by the US government.

The TCGA cohort is the following:

- Cohort 6: gene-expression and methylation data were downloaded from Illumina HiSeq RNAseq and Illumina Infinium HumanMethylation450 platforms of TCGA[115]. Stage I–IV CRC patients (n=300) were considered excluding mucinous adenocarcinomas.

CHAPTER 6: RESULTS

6.1 PATIENTS COHORTS

The cohorts have been divided in:

- CSS cohort: 22 patients, used as training set;
- UCBM cohort: 27 patients used as validation set;
- CHUS cohorts: 25 patients have been used as the training set for β4GALT1 detection as circulating biomarker; 27, instead, as validation set in the same setting.
- UNISI cohort: 102 patients with five different cancers (23 cHL, 19 mCRC, 21 mBC, 20 Lung, 19 mPC) have been used to evaluate the QC score on FFPE samples

The details of the first four cohorts are in the following table:

	Cohort 1 (UCBM)	**Cohort 2 (CSS)**	**Cohort 3T (CHUS)**	**Cohort 3V (CHUS)**
Patients	27	22	25	27
Tumor Tissue	24/27	19/22	25/25	0/27
Normal Tissue	7/27	19/22	10/25	0/27
Metastatic Tissue	27/27	13/22	1/25	0/27
Perimetastatic normal tissue	15/27	0/22	0/25	0/27
cfDNA	0/27	0/22	18/25	27/27
CTC	0/27	0/22	4/25	19/27

Having different cohorts from different structures provided us the possibility to evaluate over 201 patients for different analysis.

6.2. B4GALT1 PROMOTER IS HYPERMETHYLATED IN METASTASES OF MCRC PATIENTS

The methylation status of the *β4GALT1* promoter was analyzed in an initial training set (Group 1) of 27 CRC patients with liver metastases that underwent surgery at the UCBM. In 24/27 both tumor and metastatic lesions were available; for 3/27 patients only metastasis specimens with no tumors were retrieved; 16/27 metastases were synchronous and 11/27 were metachronous lesions. The methylation analysis was conducted in primary tumors, matched normal colorectal mucosa, and liver metastases. The median value of the *β4GALT1 /ACTB* ratio of the two plates has been used for statistical analyses. Furthermore, a validation set (Group 2) of 22 mCRC cases containing liver and lung lesions has been

used. Considering the 49 cases of the training (27 cases, fig 6.1A) and validation sets (22 cases, fig 6.1B), both the tumor and matched metastasis were available for 37/49. 19/37 CRC cases showed methylation alteration both in primary tumors and metastasis tissues. The overall concordance of *β4GALT1* methylation between primary tumor tissues and metastasis samples was 62% (23/37 patients). For 31/37 cases the data related to the timing of metastasis were available. 20/39 were synchronous metastases while 11/31 were metachronous lesions. Noteworthy, considering together synchronous and metachronous metastases, 19% of cases (7/37) have shown *β4GALT1* methylation only in the metastases and no methylation in primary tumors. These data suggest that the methylation of this gene could be detected in the circulating compartment even when it is not possible to assess the methylation status of this gene in the primary tumor.

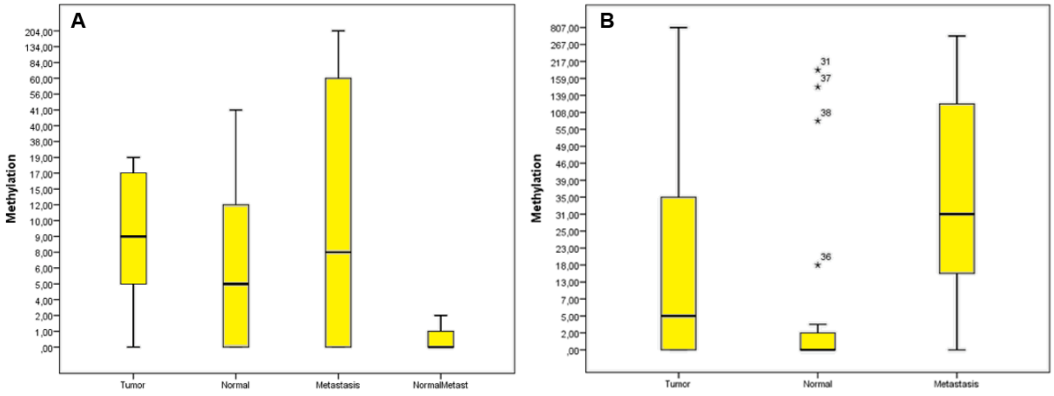

Fig. 6.1 β4GALT1 Methylation levels in UCBM (A) and CSS (B) Cohorts

We then tried to analyze if methylation level of β4GALT1 may distinguish between different sites of metastasis, but we have not find out any statistically significant difference between metastasis in liver and lung (fig 6.2). The samples come from CSS and UCBM cohorts (U di Mann-Whitney, p <0.613).

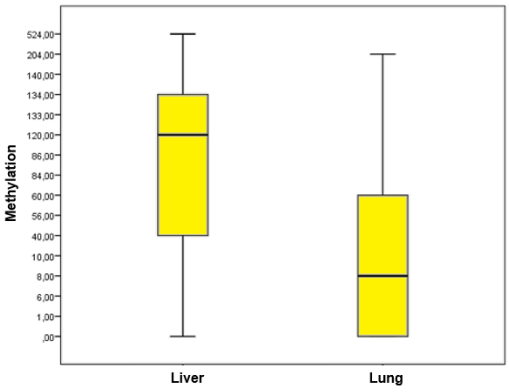

Fig. 6.2 β4GALT1 methylation level in Liver and Lung metastasis

6.3 B4GALT1 REPRESENTS A POTENTIAL "CANDIDATE BIOMARKER" IN LIQUID BIOPSY FOR MCRC

As a consequence of the results in solid tissue, we have developed a method of methylation analysis on a digital platform, the ddQMSP in collaboration with the company BioClarma of Turin. The set-up of this assay allowed us to conduct methylation analysis in the circulating compartment of patients with metastatic colon cancer (cfDNA, CTC) from CHUS.

First, we tried with the same set up used for qMSP, although to optimize the procedure, some changes were needed. We have removed β-mercaptoethanol due to its capability of inhibits the ddQMSP, such in figure 6.3A. Then, we have used ddPCR mastermix (BioRad, Pleasanton, CA, USA), to obtain amplification of the positive control (methylated DNA).

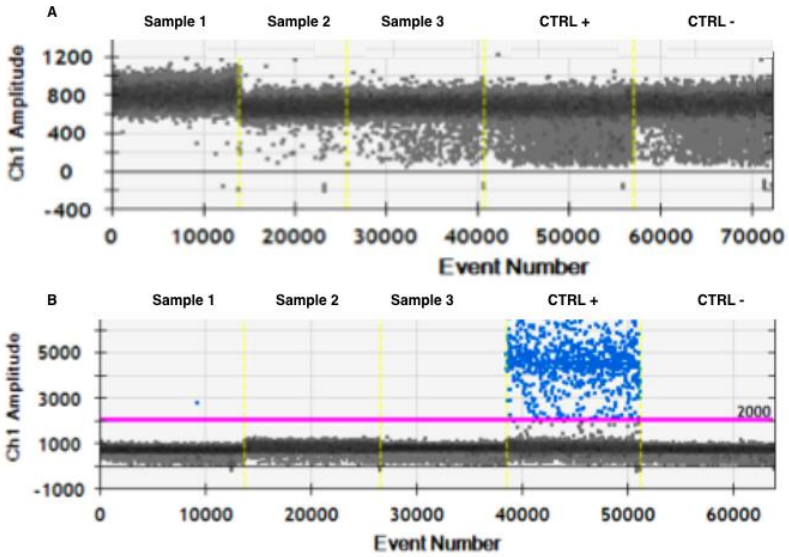

Fig 6.3 ddQMSP with (A) and without (B) β-mercaptoethanol

Different quantities of control template were also tested, and we found out that 10ng was enough to obtain an amplification of the control DNA, about 1/10 of the quantity usually used for qMSP already starting from. At quantity greater than 10ng of starting DNA, in fact, the ddQMSP tends to saturate and it is not possible to distinguish the amplified from the background drops.

As compared to the control (chemically methylated DNA), the DNA extracted from paraffin samples behaves differently, showing a better resolution of the signal at a quantity of 50 ng, this was then assumed in the subsequent experiments (fig 6.4).

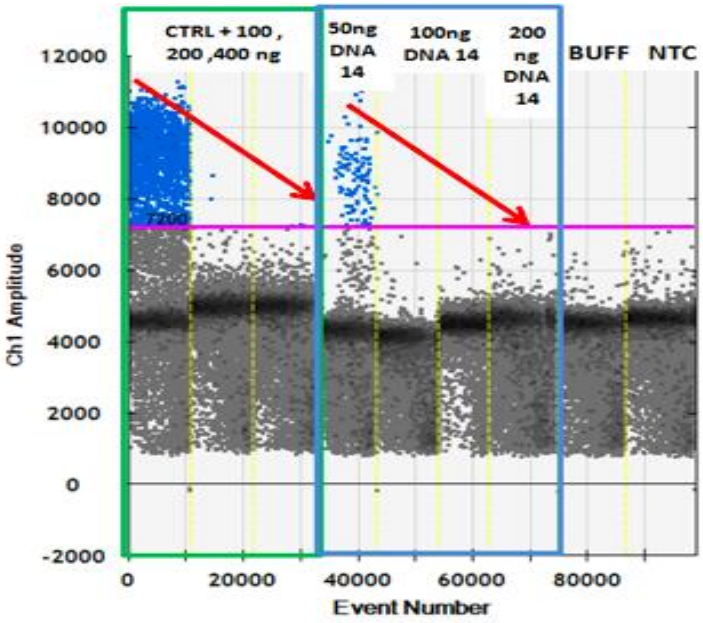

Fig 6.4 Ideal DNA quantity for samples in ddQMSP

Furthermore, to evaluate the detection limit (LoD) of ddQMSP, we carried out two standard curves (STD) with 1:10 serial dilutions of chemically methylated DNA starting from 10ng of initial template, both in water and in DNA of a patient previously analyzed and negative result. Both curves showed linearity in all four points of the curve (100% -10% -1% -0.1%). From a quantitative point of view, no differences are observed between an STD curve prepared in H_2O (fig 6.5A) or in unmethylated DNA (fig 6.5B).

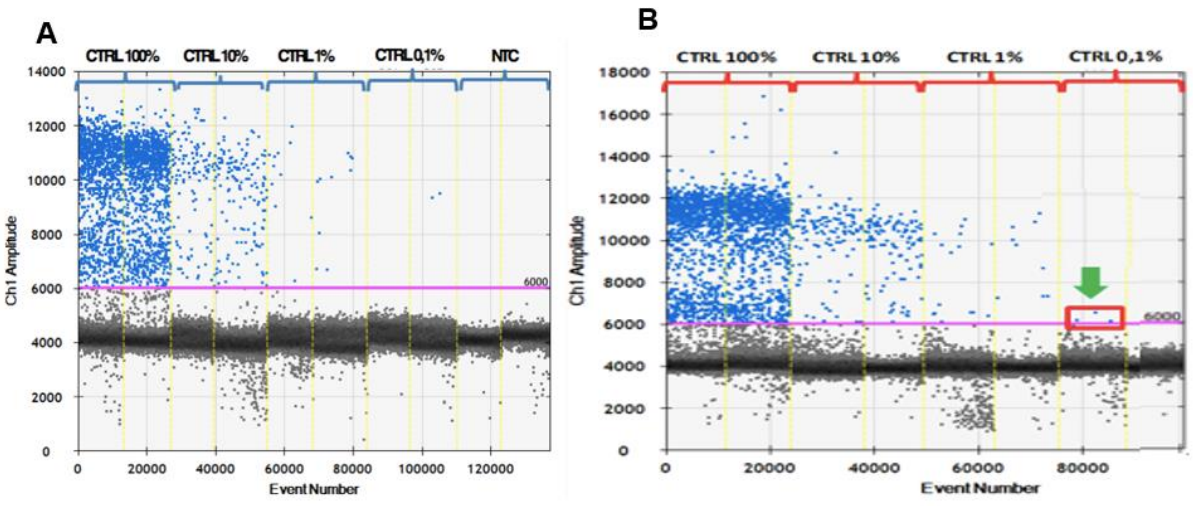

Fig 6.5 Standard Curve in H_2O and DNA sample

6.4. METHYLATION ANALYSIS OF β4GALT1 PROMOTER IN CIRCULATING COMPARTMENT OF MCRC PATIENTS

Based on the LOD results of QMSP and dd-QMSP, we assessed the discriminatory power of the *β4GALT1* dd-QMSP in 20 plasma samples of mCRC patients obtained from CHUS and 19 healthy donors (HDs) to draw a ROC curve. The AUC value was 0.750 (95% CI: 0.592–0.908, $p = 0.008$) (Fig 6.6). Based on the ROC curve, an optimal cut off value of 0.04 was assessed by maximizing the sensitivity and specificity, thereby yielding a 100% specificity and a 50% sensitivity. Methylation of *β4GALT1* showed a highly discriminative ROC curve profile, clearly distinguishing mCRC patients from healthy control subjects.

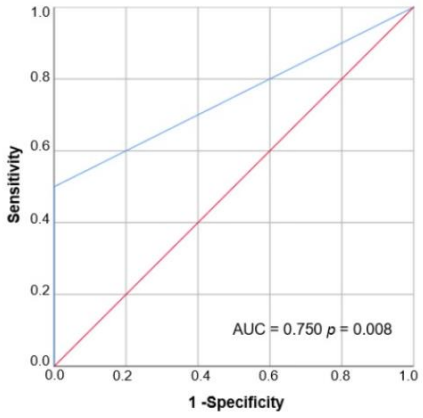

Fig 6.6 ddQMSP analyses of β4GALT1 in mCRC patient plasmas: ROC curve analysis in the training set of 20 tumor plasma samples and 19 plasma derived from HD. AUC = 0.750 (95% CI 0.592–0.908, p = 0.008)

In the training set 9/20 mCRC (45%) displayed *β4GALT1* methylation in primary tumors, among them 5/9 (56%) showed methylation in tumors and plasma as well, whereas 4/9 (44%) cases showed methylation only in the plasma compartment. In a second independent validation set of 26 plasma specimens collected from mCRC patients obtained from CHUS, *β4GALT1* promoter methylation was analyzed only in plasma. According to the 0.04 cut-off value derived from the ROC analysis 12/26 (46%) mCRC cases of the validation set showed *β4GALT1* hypermethylation in plasma.

In both training and validation set, β4GALT1 shown high methylation level in 8/23 (34,8%) CTC cases. In 10/19 (53%) where both ctDNA and CTC were available, the results were concordant, either positive (20%) or negative (80%).

The ddQMSP analysis on plasma samples from HD were always negative (fig 6.7).

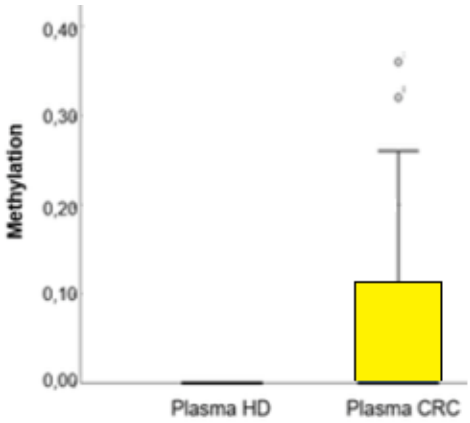

Fig 6.7 β4GALT1 methylation level in plasma from HD and CRC patients

6.5. DOWNREGULATED EXPRESSION OF β4GALT1 IS ASSOCIATED WITH POOR SURVIVAL IN CRC IN GEO AND TCGA DATASET

We have analyzed the prognostic impact of β4GALT1 gene expression using data of CRC patients with stage I–III (cohorts 1–4) from GEO datasets. DFS for cohorts 1 to 3 and CSS for cohort 4 have been used as prognostic parameter. In three out of four datasets, Kaplan–Meier curves showed a significantly worse survival duration in patients with β4GALT1 low expression than in β4GALT1 high patients (cohort 1 DFS p = 0.164; cohort 2 DFS p = 0.002; cohort 3 DFS p = 0.003; cohort 4 CSS p = 0.008) (Figure 6.8 A–D), indicating that the downregulation of β4GALT1 gene expression is related to poor prognosis for CRC. These results were confirmed also in GEO dataset cohort 5 (all patients have stage IV CRC) and TCGA cohort 6, for which PFS, OS, and DFS data were available (Figure 6.8E–G). Only cohort 5 patients (n = 80) received Cetuximab monotherapy. Patients with β4GALT1 low expression levels showed a shorter PFS duration than β4GALT1 high patients (p = 0.02; Figure 6.8E), which may suggest a possible role of β4GALT1 in influencing the Cetuximab responsiveness.

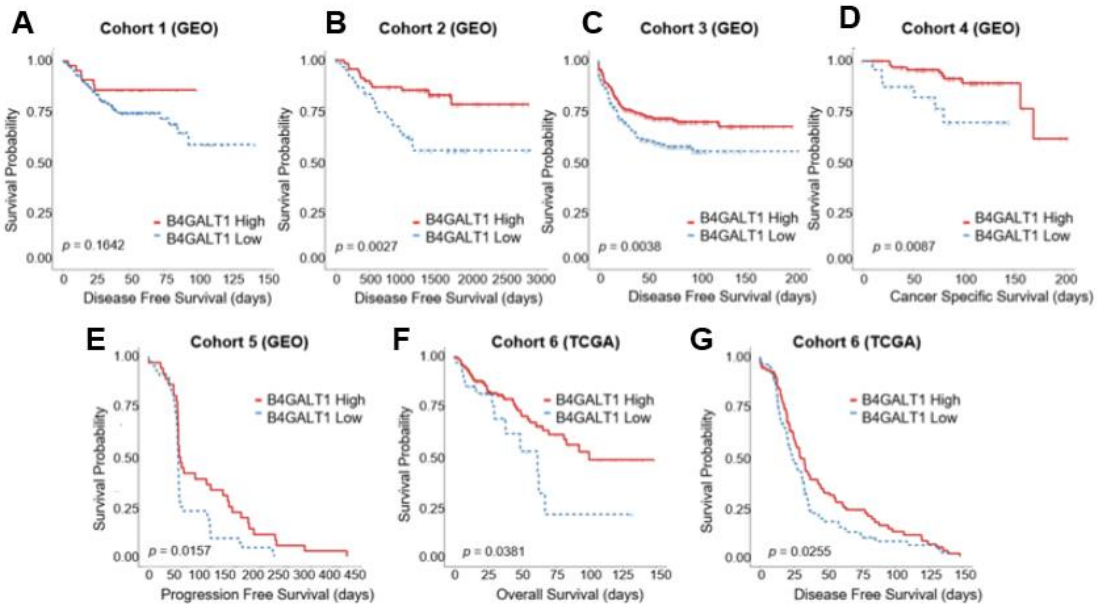

Fig 6.8 Kaplan–Meier survival curves of β4GALT1 high (red line) and low (blue line) expression groups in GEO cohorts 1-5 and TCGA cohort 6. (A) DFS in GEO Cohort 1 (n = 226), (B) DFS in GEO Cohort 2 (n = 130), (C) DFS in GEO Cohort 3 (n = 557), (D) cancer-specific survival (CSS) in GEO Cohort 4 (n = 125), (E) PFS in GEO Cohort 5 (n = 80), (F) OS in TCGA Cohort 6, (G) DFS in TCGA Cohort 6 (n = 300).

Two cohorts have been analyzed separately in details due to the specific information provided: GEO dataset cohort 5 for Cetuximab resistance and TCGA dataset cohort 6 to analyze CpG islands.

6.6. Downregulated Expression of B4GALT1 is Associated with Cetuximab Resistance in Primary Tumors of CRC in GEO Dataset

In GEO cohort 5, 110 patients had received Cetuximab monotherapy. For 80/110 cases mRNA expression data were available and 70/80 were provided KRAS mutation status information (Fig 6.9A). In those cases there was no difference in the KRAS mutation rates between the β4GALT1 high (38.9%) and low (29.5%) expression groups (p = 0.665). However we found statistical significant differences in response to Cetuximab between the two groups, where the PFS was lower in patients with downregulation of β4GALT1 (p = 0.01 (Figure 6.8E)). Interestingly, CRC patients with low expression levels of β4GALT1 showed indeed a significantly shorter PFS duration than patients with high expression levels of β4GALT1 and more so in wild-type (WT)-KRAS patients (p = 0.032 (Figure 6.9B)) than in KRAS-mutant patients (p = 0.916 (Figure 6.9C)). Furthermore, when we restricted the analyses to the group of patients with WT-KRAS the disease control rate (partial remission or stable disease) occurred only in patients with β4GALT1 high expression levels (65%) vs. β4GALT1 low group patients (30.4%) (p = 0.025).

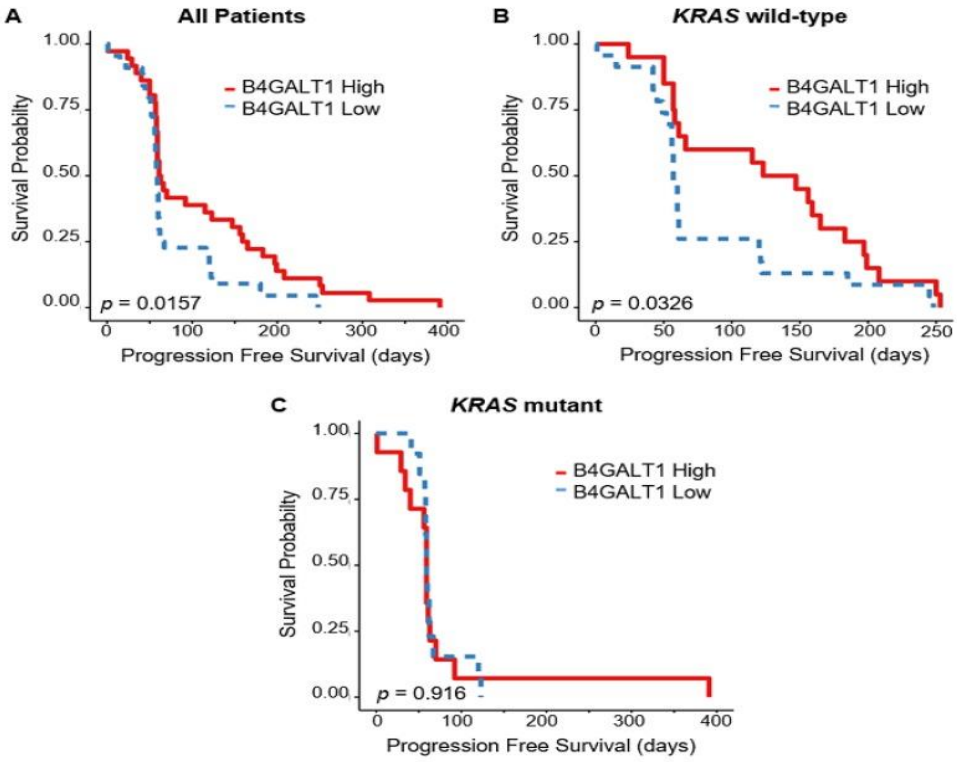

Fig 6.9 Kaplan–Meier survival curves of β4GALT1 high (red line) and low (blue line) expression groups in GEO cohort 5 (n = 70) according to KRAS mutational status. (A) PFS in all patients, (B) PFS in wild-type (WT)-KRAS patients, (C) PFS in mutant KRAS patients

6.7. HYPERMETHYLATION OF β4GALT1 IS ASSOCIATED WITH POOR SURVIVAL IN PRIMARY TUMORS OF CRC IN GEO AND TCGA DATASETS

In TCGA cohort 6, methylation data were available, thus we performed a Kaplan-Meier curve in four different CpG sites that showed a correlation between 3/4 methylated CpGs and clinical outcome. In particular, the hypermethylation of cg13834453, cg14829378, and cg14440947 is statistically significantly and is associated with a worse DFS (cg13834453 $p = 0.0172$; cg14829378 $p = 0.0184$ (Fig 6.10A and 6.10B, respectively)) and decreased OS (cg14440947 $p = 0.0018$ (Fig 6.10H)), respectively.

For some CpG (cg21316772, cg13834453 fig 6.10D and 6.10E, respectively) the lack of significance may be due to the methylation data that displayed low and, thus, little informative dynamic ranges.

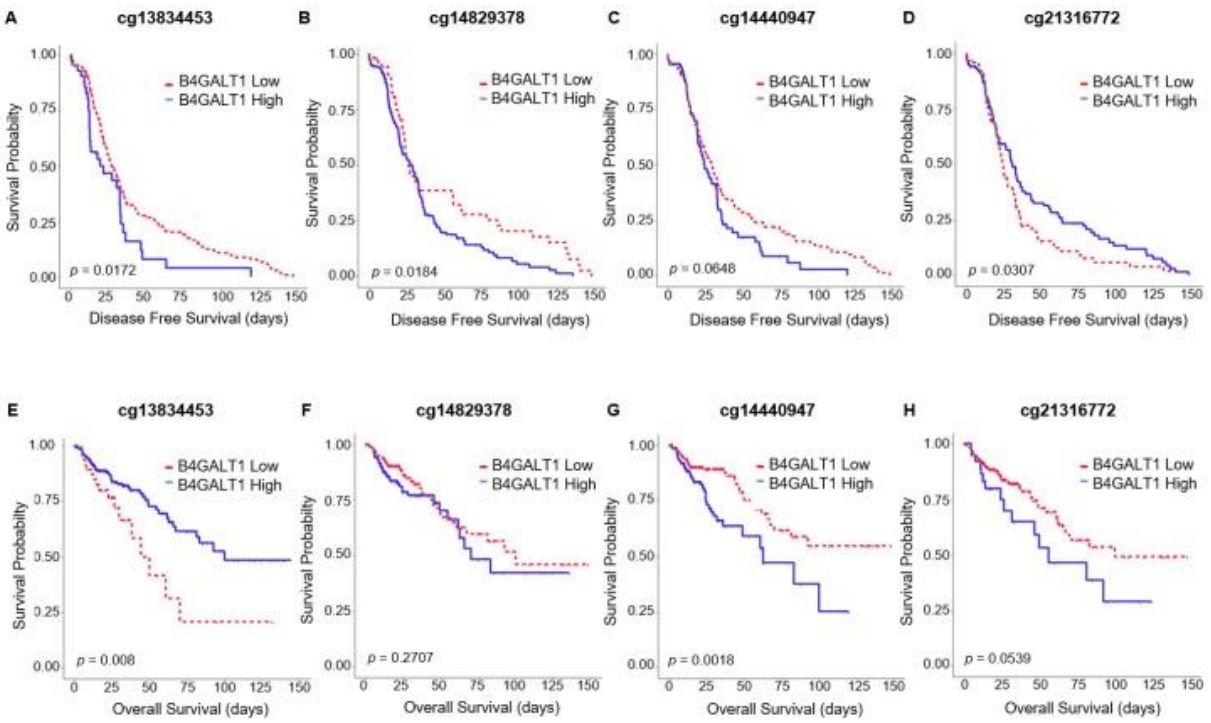

Fig 6.10 Kaplan–Meier survival curves of high (blue line) and low (red line) methylated β4GALT1 promoter CpGs in TCGA cohort 6 (n = 300). OS in (A) cg13834453, (B) cg144440947, (C) cg14829378, (D) cg21316772; DFS in (E) cg13834453, (F) cg144440947, (G) cg14829378, (H) cg21316772

6.8. QC SCORE VALUE IS NOT ASSOCIATED WITH TUMOR TYPE OR YEAR OF SAMPLE PREPARATION

To identify an adequate value for QC score, 102 FFPE Samples have been analyzed in a cohort from UNISI. We have analyzed 5 different types of tumor, organized as following: 23 classic Hodgkin Lymphoma (cHL), 19 CRCC, 21 Breast cancer, 20 Lung cancer, 19 prostate cancer. The mean QC score value was 0,4 and the median 4,1 (fig 6.11).

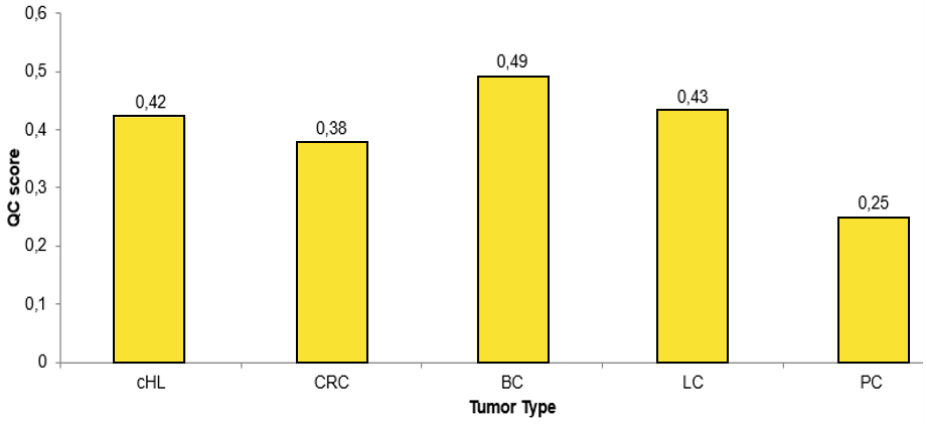

Fig 6.11 QC score according to tumor type distribution

FFPE samples have been taken from sample of different collection years, ranging from 2015 to 2019, as detailed in the table below (fig 6.12):

	Number of samples	QC Score Mean	QC Score Median
2015	18	0.37	0,39
2016	13	0,42	0,42
2017	15	0,43	0,46
2018	21	0,47	0,48
2019	24	0,42	0,44

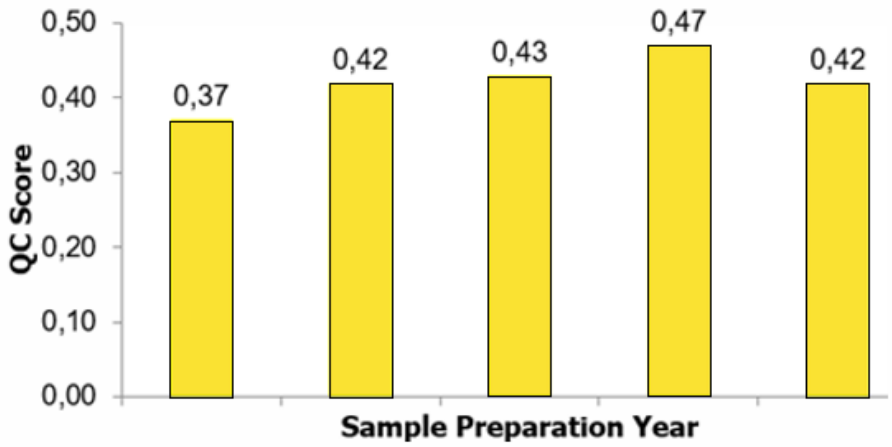

Fig 6.12 QC score according to sample preparation year

6.9. QC SCORE VALUE IS ASSOCIATED WITH PREANALYTIC PROCEDURES

To evaluate preclinical features that may influence the DNA fragmentation and QC value, we have analyzed different parameters, such as time needed for formalin fixation, storage temperature between surgery and fixation. Unfortunately, available data for those parameters were not enough for analysis. Interestingly, instead, there were a QC score difference between FFPE samples stored with low (45°C) and high (65°) paraffin melting point. This difference may suggest the importance of low temperature in maintaining DNA integrity reducing the possibility of fragmentation (fig 6.13).

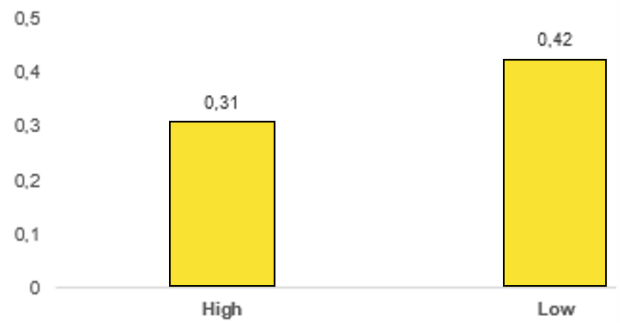

Fig 6.13 QC Score differences in high and low paraffin melting point

Noteworthy, prostate cancer samples derived from fine needle aspiration biopsy (11 samples) or from paraffin block (8 samples) have shown high QC score differences (6.14).

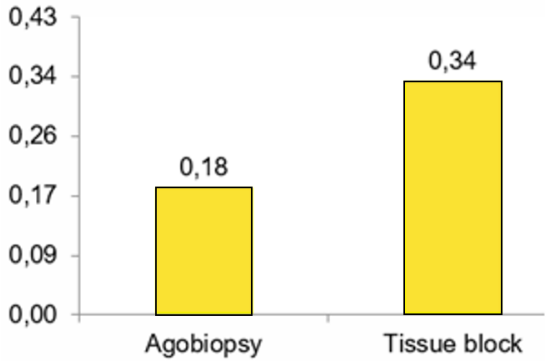

Fig 6.14 QC score differences between PC cancer samples according to material of origin

CHAPTER 7: DISCUSSION

Our study aimed to examine the possible use of β4GALT1 as a disease marker in colon cancer. Previous works in the literature have shown that the promoter of the β4GALT1 gene appears hypermethylated in 50% of CRC cases, with a specificity greater than 90% (106). Furthermore, the importance of β4GALT1 in the EGFR pathway (116) makes it a potential objective for the target therapy, in order to restore normal functions in inhibiting EGFR signal transduction, restoring the pro-apoptotic activity of the proteins Bax and Bad.

The existing literature data on the specificity and sensitivity of methylated β4GALT1 levels in primary tumor tissue constituted the rationale to analyze β4GALT1 values in metastatic tissue, both liver and lung.

The finding of β4GALT1 hypermethylation even in metastatic tissue, in the absence of significant differences with respect to the site of metastasis, provided us with the rationale for studying the levels of methylation also in the circulating compartment, both as ctDNA and within the CTC.

In order to be able to analyze small quantities of DNA, such as that extracted from plasma and CTCs, we have developed an extremely sensitive method, namely the ddQMSP. Using this technique, we found the presence of hypermethylated β4GALT1 in 55% of the plasma of patients with hypermethylated β4GALT1 also in tumor tissue. This finding paves the way for a possible use of methylated β4GALT1 as a marker of disease in patients with CRC. In addition, of fundamental importance is the fact that none of the cfDNA samples extracted from healthy subjects showed positive analysis with ddQMSP, noting the sensitivity and usefulness of the β4GALT1 assay.

In the six cohorts analyzed (1418 cases), the frequency of downregulated expression of β4GALT1 ranged from 11.6% to 80.5% of CRC patients. Among these cohorts, only five sets provided the data related to the stage. In 4/5 groups the patients with a more advanced disease showed lower expressions of the β4GALT1 gene, corroborating the hypothesis that the downregulation of the gene may occur more frequently in late phases of cancer progression.

Patients with low expressions of β4GALT1 had worse survival durations in terms of DFS and CSS than CRC cases with high expression levels. Nevertheless, in the multivariate analysis the low expression status was not significantly related to worse DFS, suggesting that it is not independent of other clinical variables. Furthermore, the Kaplan–Meier analysis conducted on the TCGA cohort revealed a correlation between methylated CpGs and clinical outcomes. Indeed, despite the fact that a large number of CpG sites should be analyzed in order to find a panel of significant CpGs that can affect the prognosis (117), we had the chance to analyze a small set of CpGs and we found that 3/4 aberrantly methylated CpGs (cg13834453, cg14829378, and cg14440947) correlate with a worse

DFS and OS. The heterogeneous cellular composition of a tumor altogether with the heterogeneous CpG methylation pattern, and the different performances of the analytical platforms available up to date to study the aberrant methylation, are all factors that make the assessment of prognostic significance of these epigenetic modifications very complex due to its multivariable dependency. So a more comprehensively methylation analysis that takes into account a larger number of CpG sites is needed to validate the prognostic impact of β4GALT1 hypermethylation in CRC.

Anti-EGFR monoclonal antibodies, such as Cetuximab and Panitumumab, used alone or in combination with chemotherapy, are still the standard treatment for mCRC, and they have shown to be more effective in wild-type RAS patients. Despite the patient selection based on the pre-treatment assessment of RAS mutational status, 40–60% are still non-responders (118). According to the most recent ASCO and ESMO guidelines (119, 120), besides mutations of RAS family members there are no recommendations for testing additional genes or proteins, suggesting the urgent need to identify additional biomarkers that can be detected in tissue or blood, to predict intrinsic or acquired anti-EGFR resistance. Several lines of evidence suggest that post-translational modifications of EGFR, such as phosphorylation of the cytoplasmic domain (121) or glycosylation can strongly influence the EGFR conformation, its activation (122, 123), and the efficiency of antibody-based therapies (124, 125), paving the way to explore a wide spectrum of enzymes able to regulate the response to antibody-based drugs. Interestingly, previous works have shown that the ectopic expression of cell surface β4GALT1 promotes apoptosis in human hepatocellular carcinoma cells inhibiting EGFR dimerization and tyrosine phosphorylation, indicating that β4GALT1 may be an inhibitor of EGFR signaling (101). In this perspective, we analyzed the predictive value of β4GALT1 in a clinical trial of CRC patients treated with anti-EGFR in monotherapy (114), and we showed that the downregulated expression of this gene is significantly associated with poor response to Cetuximab therapy, particularly in wt-KRAS cases. Among the selected datasets, only TCGA provided methylation data, but due to the lack of Cetuximab response data it was not possible to assess the relationship between aberrant β4GALT1 methylation levels and therapy response.

The use of liquid biopsy to analyze methylated β4GALT1 could therefore have numerous advantages in monitoring patients with metastatic CRC. However, further studies are needed to strengthen the data on the circulating compartment, both on patients with metastatic cancer and on patients with earlier stages of the disease, with the aim of evaluating whether β4GALT1 can also be used in screening protocols, included in of a panel of genes to increase sensitivity.

Through the collaboration with Menarini Silicon Biosystems, owner, between other devices, of the DEPArray system, we have been able to evaluate cell heterogeneity at "single-cell" level. The

DEPArray is an automated instrument whose technology is based on the ability of a non-uniform electric field to exert forces on neutral, polarizable particles, such as cells, that are suspended in a liquid. This electrokinetic principle, called dielectrophoresis (DEP), can be used to trap cells in DEP "cages" by creating an electric field. When a DEP cage is moved by a change in the electric field pattern, the trapped cell moves with it, allowing the isolation and recovery for subsequent analysis. In order to evaluate the heterogeneity at the single cell level through the use of DEPArray and Ampli1, case histories and samples of different carcinomas have been obtained from University of Siena. A Menarini Silicon Biosystems internal survey has shown that a QC score $\geq$0,3 may allow LowPass analysis on single cell (unpublished data). The DNA extracted from tumoral tissue, resulting from FFPE samples collected in different years from 2015 to 2019, have shown a mean of 0,4 and a median of 0,41, with no important differences in regard of tumor origin nor year of sample preparation, suggesting the possibility to evaluate single cell from several FFPE samples. Important note is given from the analysis of the melting point of paraffin used for FFPE preparation, due to the fact that when low melting point paraffin was used, QC score was higher. Taken together, those data highlight the importance of the preanalytic phase, with the use of low melting temperature paraffin and keeping samples in cold to prevent DNA degradation temperature-derived.

CHAPTER 8: CONCLUSIONS

Heterogeneity is an intrinsic feature of the tumor, and an issue for oncologists, pathologists, biologist and all the complex galaxy of personnel who have to deal with it to treat the patient over time.

This work confirms that liquid biopsy may overcome both spatial and temporal tumor heterogeneity. The different analytes (CTCs, ctDNA, EVs, and their characterization) available in a liquid biopsy sample provide different information that should be integrated to understand the complexity of the tumor. We have highlighted how the hypermethylation of β4GALT1 is characteristic of CRCs and the biological mechanisms that may partially explain the Cetuximab resistance of wt-KRAS tumor. The novel ddQMSP assay from ctDNA and CTCs, once validated in more robust clinical studies, may allow the possibility to postpone and reduce solid biopsy to assess β4GALT1 methylation status in blood samples.

On the other hand, single cell analysis may overcome the complexity background of a heterogeneous sample, both FFPE or Liquid Biopsy assay, gaining in specificity of evaluating features of an effective cancer cell without the background noise of non-cancer population. We have demonstrated the importance of using a QC to evaluate DNA integrity before analyzing the sample with the DEPArray system. The LowPass analysis from a single cell derived from FFPE samples with adequate QC score sorted with DEPArray, has been postponed due to the spread of Covid-19 pandemic.

Taken together, our data suggest that both approaches may provide a great opportunity to solve the puzzle of tumor treatment and drug resistance and deal with the threat of intratumor heterogeneity.

REFERENCES

1. WHO. World Health Organization (WHO). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. WHO; Accessed December 1, 2020. [Available from: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death.
2. Bray F, Laversanne M, Weiderpass E, I. S. The ever-increasing importance of cancer as a leading cause of premature death worldwide. . Cancer; In press.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021.
4. Hinohara K, Polyak K. Intratumoral Heterogeneity: More Than Just Mutations. Trends Cell Biol. 2019;29(7):569-79.
5. Merlo L, Pepper J, Reid B, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer. 2006;6(12):924-35.
6. Begon M, Harper J, Townsnd C. Ecology. 3rd edition 1998.
7. Miller BE, Miller FR, Wilburn D, Heppner GH. Dominance of a tumor subpopulation line in mixed heterogeneous mouse mammary tumors. Cancer Res. 1988;48(20):5747-53.
8. McAllister SS, Gifford AM, Greiner AL, Kelleher SP, Saelzler MP, Ince TA, et al. Systemic endocrine instigation of indolent tumor growth requires osteopontin. Cell. 2008;133(6):994-1005.
9. Lyons JG, Lobo E, Martorana AM, Myerscough MR. Clonal diversity in carcinomas: its implications for tumour progression and the contribution made to it by epithelial-mesenchymal transitions. Clin Exp Metastasis. 2008;25(6):665-77.
10. Mannelli C. Tissue vs Liquid Biopsies for Cancer Detection: Ethical Issues. J Bioeth Inq. 2019;16(4):551-7.
11. Sholl LM, Aisner DL, Allen TC, Beasley MB, Cagle PT, Capelozzi VL, et al. Liquid Biopsy in Lung Cancer: A Perspective From Members of the Pulmonary Pathology Society. Arch Pathol Lab Med. 2016;140(8):825-9.
12. Malani R, Fleisher M, Kumthekar P, Lin X, Omuro A, et al. Cerebrospinal fluid circulating tumor cells as a quantifiable measurement of leptomeningeal metastases in patients with HER2 positive cancer. J Neurooncol. 2020;148(3):599-606.
13. Printz C. Liquid biopsy test appears better at detecting bladder cancer than urine cytology. Cancer. 2021;127(5):663.
14. Rapado-González Ó, Muinelo-Romay L, Suárez-Cunqueiro MM. Letter to the editor: "Liquid biopsy based on saliva cell-free DNA as a potential biomarker for head and neck cancer". Oral Oncol. 2021;112:105016.
15. Ashworth T. A case of cancer in which cells similar to those in the tumours were seen in blood after death. Aust Med J. 1869
16. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153(3):865-73.
17. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. Cell. 2012;148(1-2):349-61.
18. Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer. 2009;9(4):302-12.
19. Janni W, Vogl FD, Wiedswang G, Synnestvedt M, Fehm T, Jückstock J, et al. Persistence of disseminated tumor cells in the bone marrow of breast cancer patients predicts increased risk for relapse--a European pooled analysis. Clin Cancer Res. 2011;17(9):2967-76.
20. Paget S. The distribution of secondary growths in cancer of the breast. 1889. Cancer Metastasis Rev. 1989;8(2):98-101.
21. Nicolson GL. Organ specificity of tumor metastasis: role of preferential adhesion, invasion and growth of malignant cells at specific secondary sites. Cancer Metastasis Rev. 1988;7(2):143-88.
22. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015;527(7578):329-35.
23. Costa C, Dávila-Ibáñez AB. Methodology for the Isolation and Analysis of CTCs. Adv Exp Med Biol. 2020;1220:45-59.
24. Went PT, Lugli A, Meier S, Bundi M, Mirlacher M, Sauter G, et al. Frequent EpCam protein expression in human carcinomas. Hum Pathol. 2004;35(1):122-8.
25. Lampignano R, Schneck H, Neumann M, Fehm T, Neubauer H. Enrichment, Isolation and Molecular Characterization of EpCAM-Negative Circulating Tumor Cells. Adv Exp Med Biol. 2017;994:181-203.
26. Balzar M, Winter MJ, de Boer CJ, Litvinov SV. The biology of the 17-1A antigen (Ep-CAM). J Mol Med (Berl). 1999;77(10):699-712.
27. Terstappen LW, Rao C, Gross S, Kotelnikov V, Racilla E, Uhr J, et al. Flow cytometry--principles and feasibility in transfusion medicine. Enumeration of epithelial derived tumor cells in peripheral blood. Vox Sang. 1998;74 Suppl 2:269-74.
28. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. 2004;351(8):781-91.
29. Attard G, Sarker D, Reid A, Molife R, Parker C, de Bono JS. Improving the outcome of patients with castration-resistant prostate cancer through rational drug development. Br J Cancer. 2006;95(7):767-74.
30. Cohen SJ, Punt CJ, Iannotti N, Saidman BH, Sabbath KD, Gabrail NY, et al. Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. J Clin Oncol. 2008;26(19):3213-21.
31. Nicolazzo C, Gradilone A, Loreni F, Raimondi C, Gazzaniga P. EpCAM. Dis Markers. 2019;2019:1718920.
32. Hvichia GE, Parveen Z, Wagner C, Janning M, Quidde J, et al. A novel microfluidic platform for size and deformability based separation and the subsequent molecular characterization of viable circulating tumor cells. Int J Cancer. 2016;138(12):2894-904.
33. Riethdorf S, Fritsche H, Müller V, Rau T, Schindlbeck C, Rack B, et al. Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system. Clin Cancer Res. 2007;13(3):920-8.
34. Krebs MG, Metcalf RL, Carter L, Brady G, Blackhall FH, Dive C. Molecular analysis of circulating tumour cells-biology and biomarkers. Nat Rev Clin Oncol. 2014;11(3):129-44.
35. Mandel P, Metais P. Nuclear Acids In Human Blood Plasma. C R Seances Soc Biol Fil. 1948;142(3-4):241-3.

36. Anker P. Quantitative aspects of plasma/serum DNA in cancer patients. Ann N Y Acad Sci. 2000;906:5-7.
37. Tong YK, Lo YM. Diagnostic developments involving cell-free (circulating) nucleic acids. Clin Chim Acta. 2006;363(1-2):187-96.
38. Silva JM, Dominguez G, Garcia JM, Gonzalez R, Villanueva MJ, Navarro F, et al. Presence of tumor DNA in plasma of breast cancer patients: clinicopathological correlations. Cancer Res. 1999;59(13):3251-6.
39. 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(9):2297-303.
40. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A. 2005;102(45):16368-73.
41. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. Clin Chim Acta. 2001;313(1-2):139-42.
42. Holdenrieder S, Dharuman Y, Standop J, Trimpop N, Herzog M, Hettwer K, et al. Novel serum nucleosomics biomarkers for the detection of colorectal cancer. Anticancer Res. 2014;34(5):2357-62.
43. Schwarzenbach H, Pantel K. Circulating DNA as biomarker in breast cancer. Breast Cancer Res. 2015;17(1):136.
44. Board RE, Williams VS, Knight L, Shaw J, Greystoke A, Ranson M, et al. Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer. Ann N Y Acad Sci. 2008;1137:98-107.
45. Benesova L, Belsanova B, Suchanek S, Kopeckova M, Minarikova P, Lipska L, et al. Mutation-based detection and monitoring of cell-free tumor DNA in peripheral blood of cancer patients. Anal Biochem. 2013;433(2):227-34.
46. Swanton C. Plasma-derived tumor DNA analysis at whole-genome resolution. Clin Chem. 2013;59(1):6-8.
47. Chan DLH, Toh GLX, Goh LL. Clinical implementation of plasma EGFR T790M testing using droplet digital PCR in TKI-resistant NSCLC patients. Exp Mol Pathol. 2020;116:104515.
48. Wu DY, Ugozzoli L, Pal BK, Wallace RB. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc Natl Acad Sci U S A. 1989;86(8):2757-60.
49. Taly V, Pekin D, Benhaim L, Kotsopoulos SK, Le Corre D, Li X, et al. Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients. Clin Chem. 2013;59(12):1722-31.
50. Chargaff E, West R. The biological significance of the thromboplastic protein of blood. J Biol Chem. 1946;166(1):189-97.
51. Satta N, Toti F, Feugeas O, Bohbot A, Dachary-Prigent J, Eschwège V, et al. Monocyte vesiculation is a possible mechanism for dissemination of membrane-associated procoagulant activities and adhesion molecules after stimulation by lipopolysaccharide. J Immunol. 1994;153(7):3245-55.
52. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654-9.
53. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol. 2013;200(4):373-83.
54. van der Pol E, Böing AN, Harrison P, Sturk A, Nieuwland R. Classification, functions, and clinical relevance of extracellular vesicles. Pharmacol Rev. 2012;64(3):676-705.
55. Denzer K, Kleijmeer MJ, Heijnen HF, Stoorvogel W, Geuze HJ. Exosome: from internal vesicle of the multivesicular body to intercellular signaling device. J Cell Sci. 2000;113 Pt 19:3365-74.
56. Boulanger CM, Amabile N, Guérin AP, Pannier B, Leroyer AS, Mallat CN, et al. In vivo shear stress determines circulating levels of endothelial microparticles in end-stage renal disease. Hypertension. 2007;49(4):902-8.
57. Weerheim AM, Kolb AM, Sturk A, Nieuwland R. Phospholipid composition of cell-derived microparticles determined by one-dimensional high-performance thin-layer chromatography. Anal Biochem. 2002;302(2):191-8.
58. Akers JC, Gonda D, Kim R, Carter BS, Chen CC. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. J Neurooncol. 2013;113(1):1-11.
59. Larson MC, Woodliff JE, Hillery CA, Kearl TJ, Zhao M. Phosphatidylethanolamine is externalized at the surface of microparticles. Biochim Biophys Acta. 2012;1821(12):1501-7.
60. György B, Szabó TG, Pásztói M, Pál Z, Misják P, Aradi B, et al. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. Cell Mol Life Sci. 2011;68(16):2667-88.
61. Berckmans RJ, Nieuwland R, Tak PP, Böing AN, Romijn FP, Kraan MC, et al. Cell-derived microparticles in synovial fluid from inflamed arthritic joints support coagulation exclusively via a factor VII-dependent mechanism. Arthritis Rheum. 2002;46(11):2857-66.
62. Lianidou E, Pantel K. Liquid biopsies. Genes Chromosomes Cancer. 2019;58(4):219-32.
63. Zhang K. Stratifying tissue heterogeneity with scalable single-cell assays. Nat Methods. 2017;14(3):238-9.
64. Roth A, McPherson A, Laks E, Biele J, Yap D, Wan A, et al. Clonal genotype and population structure inference from single-cell tumor sequencing. Nat Methods. 2016;13(7):573-6.
65. Zahn H, Steif A, Laks E, Eirew P, VanInsberghe M, Shah SP, et al. Scalable whole-genome single-cell library preparation without preamplification. Nat Methods. 2017;14(2):167-73.
66. Worrall JT, Tamura N, Mazzagatti A, Shaikh N, van Lingen T, Bakker B, et al. Non-random Mis-segregation of Human Chromosomes. Cell Rep. 2018;23(11):3366-80.
67. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. Trends Biotechnol. 2010;28(6):281-90.
68. Wu AR, Wang J, Streets AM, Huang Y. Single-Cell Transcriptional Analysis. Annu Rev Anal Chem (Palo Alto Calif). 2017;10(1):439-62.
69. Bolognesi C, Forcato C, Buson G, Fontana F, Mangano C, Doffini A, et al. Digital Sorting of Pure Cell Populations Enables Unambiguous Genetic Analysis of Heterogeneous Formalin-Fixed Paraffin-Embedded Tumors by Next Generation Sequencing. Sci Rep. 2016;6:20944.
70. Rossi T, Gallerani G, Angeli D, Cocchi C, Bandini E, Fici P, et al. Single-Cell NGS-Based Analysis of Copy Number Alterations Reveals New Insights in Circulating Tumor Cells Persistence in Early-Stage Breast Cancer. Cancers (Basel). 2020;12(9).
71. Waddington CH. The epigenotype. 1942. Int J Epidemiol. 2012;41(1):10-3.

72. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36.
73. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251-60.
74. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128(4):683-92.
75. Chen K, Zhao BS, He C. Nucleic Acid Modifications in Regulation of Gene Expression. Cell Chem Biol. 2016;23(1):74-85.
76. Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science. 2001;293(5532):1068-70.
77. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827-31.
78. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6-21.
79. Wang Y, Leung FC. An evaluation of new criteria for CpG islands in the human genome as gene markers. Bioinformatics. 2004;20(7):1170-7.
80. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465-76.
81. Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S. Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J. 2002;21(15):4183-95.
82. Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693-705.
83. Jenuwein T, Allis CD. Translating the histone code. Science. 2001;293(5532):1074-80.
84. Liang G, Lin JC, Wei V, Yoo C, Cheng JC, Nguyen CT, et al. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. Proc Natl Acad Sci U S A. 2004;101(19):7357-62.
85. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553-60.
86. Tachibana M, Matsumura Y, Fukuda M, Kimura H, Shinkai Y. G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. EMBO J. 2008;27(20):2681-90.
87. Estève PO, Chin HG, Benner J, Feehery GR, Samaranayake M, Horwitz GA, et al. Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. Proc Natl Acad Sci U S A. 2009;106(13):5076-81.
88. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem. 2003;278(6):4035-40.
89. Wright MW, Bruford EA. Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics. 2011;5(2):90-8.
90. Wang C, Lin H. Roles of piRNAs in transposon and pseudogene regulation of germline mRNAs and lncRNAs. Genome Biol. 2021;22(1):27.
91. Decatur WA, Fournier MJ. RNA-guided nucleotide modification of ribosomal and other RNA. J Biol Che 2003;278(2):695-8.
92. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136(4):629-41.
93. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature. 2002;420(6915):563-73.
94. Sharpless NE, DePinho RA. Telomeres, stem cells, senescence, and cancer. J Clin Invest. 2004;113(2):160-8.
95. Kumar A, Kumari N, Gupta V. Renal Cell Carcinoma: Molecular Aspects. Indian J Clin Biochem. 2018;33(3):246-54.
96. Hill RJ, Lou Y, Tan SL. B-cell antigen receptor signaling in chronic lymphocytic leukemia: therapeutic targets and translational opportunities. Int Rev Immunol. 2013;32(4):377-96.
97. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004;4(3):177-83.
98. Lee J, Park SH, Sundaram S, Raju TS, Shaper NL, Stanley P. A mutation causing a reduced level of expression of six beta4-galactosyltransferase genes is the basis of the Lec19 CHO glycosylation mutant. Biochemistry. 2003;42(42):12349-57.
99. Harduin-Lepers A, Shaper JH, Shaper NL. Characterization of two cis-regulatory regions in the murine beta 1,4-galactosyltransferase gene. Evidence for a negative regulatory element that controls initiation at the proximal site. J Biol Chem. 1993;268(19):14348-59.
100. Hathaway HJ. Cell surface beta1,4-galactosyltransferase function in mammary gland morphogenesis: insights from transgenic and knockout mouse models. J Mammary Gland Biol Neoplasia. 2003;8(4):421-33.
101. Tang W, Weng S, Zhang S, Wu W, Dong L, Shen X, et al. Direct interaction between surface β1,4-galactosyltransferase 1 and epidermal growth factor receptor (EGFR) inhibits EGFR activation in hepatocellular carcinoma. Biochem Biophys Res Commun. 2013;434(3):449-54.
102. Hinton DA, Evans SC, Shur BD. Altering the expression of cell surface beta 1,4-galactosyltransferase modulates cell growth. Exp Cell Res 1995;219(2):640-9.
103. Ettenberg SA, Rubinstein YR, Banerjee P, Nau MM, Keane MM, Lipkowitz S. cbl-b inhibits EGF-receptor-induced apoptosis by enhancing ubiquitination and degradation of activated receptors. Mol Cell Biol Res Commun. 1999;2(2):111-8.
104. Rojas M, Yao S, Lin YZ. Controlling epidermal growth factor (EGF)-stimulated Ras activation in intact cells by a cell-permeable peptide mimicking phosphorylated EGF receptor. J Biol Chem. 1996;271(44):27456-61.
105. Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2000;103(2):211-25.
106. Poeta ML, Massi E, Parrella P, Pellegrini P, De Robertis M, Copetti M, et al. Aberrant promoter methylation of beta-1,4 galactosyltransferase 1 as potential cancer-specific biomarker of colorectal tumors. Genes Chromosomes Cancer. 2012;51(12):1133-43.
107. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253-66.
108. Lo YM, Wong IH, Zhang J, Tein MS, Ng MH, Hjelm NM. Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. Cancer Res. 1999;59(16):3899-903.
109. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, Desai J, et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. Clin Cancer Res. 2009;15(24):7642-51.
110. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010;138(3):958-68.

111. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med. 2013;10(5):e1001453.
112. Lee KW, Lee SS, Kim SB, Sohn BH, Lee HS, Jang HJ, et al. Significant association of oncogene YAP1 with poor prognosis and cetuximab resistance in colorectal cancer patients. Clin Cancer Res. 2015;21(2):357-64.
113. Sun B, Li Y, Zhou Y, Ng TK, Zhao C, Gan Q, et al. Circulating exosomal CPNE3 as a diagnostic and prognostic biomarker for colorectal cancer. J Cell Physiol. 2019;234(2):1416-25.
114. Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, Wu S, et al. Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. J Clin Oncol. 2007;25(22):3230-7.
115. Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.
116. Li Z, Zong H, Kong X, Zhang S, Wang H, Sun Q, et al. Cell surface beta 1, 4-galactosyltransferase 1 promotes apoptosis by inhibiting epidermal growth factor receptor pathway. Mol Cell Biochem. 2006;291(1-2):69-76.
117. Jia M, Zhang Y, Jansen L, Walter V, Edelmann D, Gündert M, et al. A prognostic CpG score derived from epigenome-wide profiling of tumor tissue was independently associated with colorectal cancer survival. Clin Epigenetics. 2019;11(1):109.
118. Sorich MJ, Wiese MD, Rowland A, Kichenadasse G, McKinnon RA, Karapetis CS. Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. Ann Oncol. 2015;26(1):13-21.
119. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol. 2016;27(8):1386-422.
120. Sepulveda AR, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK, et al. Molecular Biomarkers for the Evaluation of Colorectal Cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and the American Society of Clinical Oncology. J Clin Oncol. 2017;35(13):1453-86.
121. Horita H, Law A, Hong S, Middleton K. A simple toolset to identify endogenous post-translational modifications for a target protein: a snapshot of the EGFR signaling pathway. Biosci Rep. 2017;37(4).
122. Fernandes H, Cohen S, Bishayee S. Glycosylation-induced conformational modification positively regulates receptor-receptor association: a study with an aberrant epidermal growth factor receptor (EGFRvIII/DeltaEGFR) expressed in cancer cells. J Biol Chem. 2001;276(7):5375-83.
123. Kaszuba K, Grzybek M, Orłowski A, Danne R, Róg T, Simons K, et al. N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. Proc Natl Acad Sci U S A. 2015;112(14):4334-9.
124. Korphaisarn K, Chou CK, Xia WY, Clarke CN, Katkhuda R, Davis JS, et al. Arginine methylation of EGFR: a new biomarker for predicting resistance to anti-EGFR treatment. Am J Cancer Res. 2017;7(12):2587-99.
125. Liao HW, Hsu JM, Xia W, Wang HL, Wang YN, Chang WC, et al. PRMT1-mediated methylation of the EGF receptor regulates signaling and cetuximab response. J Clin Invest. 2015;125(12):4529-43.

ACKNOWLEDGEMENTS

Difficult period makes important people more important. Therefore, I want to thank first of all the patients whose tissue have been analyzed in this study. Then, all the people who work with me to this project and without them this couldn't see lights: dr Parrella, dr Muscarella, dr Barbano, from CSS, dr Muinelo-Romay and her team from CHUS, dr Taffon and her team from UCBM, dr Amato from UNISI. I have really appreciated your help and support.

Special thanks goes to dr Poeta, my mentor and guide through this study project, and to prof Fazio, who took me as a med student and saw me leave his lab as an "MD, PhD", and to all the people who has crossed their paths with me in the LAMMB. I'll never stop to thank the organizer of the meeting in Montpellier, were I met dr Simonelli, to whom go my heartfelt thanks, who gave me the chance to finish the project and follow my dreams, even if in another role, who protects me when necessary and lets me make my mistakes, to grow better.

Last, but not least, my greatest thanks goes to who always supported me, whatever I do: my family, my girlfriend, pillar of my life, who, despite all odds, become my wife in the meanwhile, and all my friends.

Thank you all.