

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
SCIENZE CARDIO NEFRO TORACICHE**

Ciclo 32°

Settore Concorsuale: 06/D3

Settore Scientifico Disciplinare: MED/21

**ADDRESSING TUMOR HETEROGENEITY
IN ESOPHAGEAL ADENOCARCINOMA
THROUGH DIFFERENT MOLECULAR APPROACHES**

Presentata da: Federica Isidori

Coordinatore Dottorato

Prof. Gaetano Domenico Gargiulo

Supervisore

Prof. Niccolò Daddi

**Co-supervisore
Prof. ssa Elena Bonora**

Esame finale anno 2020

Abstract

Objective

Several studies have shown epidemiologic, clinical, immune-histochemical and molecular differences among esophageal adenocarcinomas (EAC). Since pathogenesis and biology of this tumor are far to be well defined, our study aimed to examine intra- and inter-tumor heterogeneity and to solve crucial controversies through different molecular approaches.

Results

Target sequencing was performed for sorted cancer subpopulations from formalin embedded material obtained from 38 EACs, not treated with neoadjuvant therapy.

35 out 38 cases carried at least one somatic mutation, not present in the corresponding sorted stromal cells. 73.7% of cases carried mutations in *TP53* and 10.5% in *CDKN2A*. Mutations in other genes occurred at lower frequency, including *HNF1A*, not previously associated with EAC. Sorting allowed us to isolate clones with different mutational loads and/or additional copy number amplifications, confirming the high intra-tumor heterogeneity of these cancers. In our cohort *TP53* gene abnormalities correlated with a better survival ($P = 0.028$); conversely, loss of SMAD4 protein expression was associated with a higher recurrence rate ($P = 0.015$).

Shifting the focus on the epigenetic characterization of EAC, miR-221 and miR-483-3p resulted upregulated from the MicroRNA Array card analysis and confirmed with further testing. The up-regulation of both miRNAs correlated with clinical outcomes, in particular with a reduced cancer-specific survival (miR483-3p $P=0.0293$; miR221 $P=0.0059$). In vitro analyses demonstrated an increase for miR-483-3p (fold-change=2.7) that appear to be inversely correlated with SMAD4 expression in FLO-1 cell-line.

Conclusions

In conclusion, selective sorting allowed to define the real mutation status and to isolate different cancer subclones.

MiRNA expression analysis revealed a significant up-regulation of miR-221 and miR-483-3p, which correlated with worst prognosis, implying that they can be considered oncogenic factors in EAC.

Therefore, cell sorting technologies, coupled with next generation sequencing, and the analysis of microRNA profiles seem to be promising strategies to guide treatment and help classify cancer prognosis.

INDEX

ABSTRACT	1
1. INTRODUCTION	6
1.1 Esophageal cancer	6
1.2 Pathogenesis of Esophageal Adenocarcinoma	8
1.3 Treatment options	9
1.4 Esophageal Adenocarcinoma: current classifications and prognostic factors	11
1.4.1 Staging of EAC	11
1.4.2 Classification according to position: the Siewert classification	12
1.4.3 Different morphological and histological spectrum of esophageal adenocarcinoma	13
1.4.4 Histological classification according to Lauren	14
1.4.5 Immuno-histologic profiles of esophageal adenocarcinoma	15
1.5 Genomic characterization of EAC	16
1.5.1 <i>TP53</i> and <i>CDKN2A</i>	19
1.5.2 TGF- β pathway – associated genes	20
1.6 Technological advantages in the study of cancer: digital sorting of pure cell populations	21
1.7 Micro-RNAs and Esophageal Adenocarcinoma	23
2. AIMS	26
3. MATERIALS AND METHODS	27
PART I	27
3.1 Genetic analysis of esophageal adenocarcinoma	27

3.1.1	Subjects	27
3.1.2	Whole-exome sequencing (WES)	28
3.1.3	Bioinformatics data analysis	29
3.1.4	SNPhylo: phylogenetic analysis of primary tumor and metastases	29
3.1.5	LOH detection from WES data on genomic DNA	30
3.1.6	Cell sorting protocol	30
3.1.7	Assessing FFPE DNA quality	32
3.1.8	OncoSeek genetic analysis	32
3.1.9	Whole-Genome Low-Pass sequencing for CNA analysis	33
3.1.10	TA-cloning for <i>SMARCA4</i> PCR product	34
3.1.11	Sanger sequencing	35
3.1.12	Droplet digital PCR (ddPCR)	35
3.1.13	In Situ Hybridization (ISH)	36
3.1.14	Immunohistochemistry (IHC)	36
3.1.15	Statistical analysis	37
Part II		38
3.2 MicroRNA expression profiles in esophageal adenocarcinoma		38
3.2.1	Subjects	38
3.2.2	RNA isolation from FFPE surgical specimens	39
3.2.3	TaqMan Human MicroRNA Array cards for microRNA expression profiling	39
3.2.4	Quantitative real-time PCR validation of microRNA profiles through single assays	40
3.2.5	Immunohistochemistry (IHC)	41
3.2.6	Cell lines	42
3.2.7	Reverse transcription PCR and real-time quantitative PCR analysis of <i>SMAD4</i>	42
3.2.8	Wound healing assay	43
3.2.9	Statistical analysis	43
4. RESULTS PART I		44
4.1 Genomic profiles of primary and metastatic esophageal adenocarcinoma identified via digital sorting of pure cell populations: results from a case report		44

4.2 High-throughput sorting of tumor cell populations reveals the composite mutational landscape of esophageal adenocarcinoma	54
4.2.1 Copy number alteration analysis	55
4.2.2 Identification of mutations in <i>TP53</i> and p53-regulated genes	57
4.2.3 Selective sorting identifies high intra-tumor heterogeneity	60
4.2.4 Correlation between <i>TP53</i> mutations and survival	61
4.2.5 <i>SMAD4</i> loss is associated with cancer recurrence	63
4.2.6 Identification of mutations in <i>HNF1A</i> , a novel mutated gene in EAC	64
4.2.7 Identification of additional mutations through Whole Exome Sequencing (WES)	65
5. RESULTS PART II	68
5.1 Identification of deregulated microRNAs (miRNAs) in esophageal adenocarcinoma patients through microfluidic cards	68
5.2 miRNA 221 and 483-3p expression analysis via single assays and correlation with clinical outcomes	69
5.2.1 miRNAs expression data from FFPE samples	69
5.2.1.1 miRNA 483-3p expression and correlation with clinical features	70
5.2.1.2 miRNA 221 expression and correlation with clinical outcomes	73
5.2.2 miRNAs expression data from fresh-frozen samples	75
5.2.3 Expression analysis of <i>SMAD4</i> : a validated target of miR-483-3p	77
5.2.4 Expression analysis of <i>PTEN</i> : a validated target of miR-221	78
5.2.5 miRNAs expression analysis in OE19 and FLO-1 EAC cell lines	79
6. DISCUSSION	81
7. CONCLUSIONS	93
8. TABLES	94
9. REFERENCES	110
ACKNOWLEDGEMENTS	123

1. Introduction

1.1 Esophageal cancer

Esophageal cancer is considered a severe malignancy in terms of prognosis and mortality rate. Despite the latest improvements in diagnostics and therapy, esophageal cancer is still the sixth most frequent cause of cancer-related death worldwide [1], with an estimated 500,000 deaths in 2018 (about 6.7 mortality rate expressed as an annual rate per 100,000 persons at risk). A worldwide map of esophageal cancer mortality is shown in Figure 1 (source <http://globocan.iarc.fr/>). Patients that suffer from this cancer have very poor prognosis, with 5-years survival rates of 15-20% [2]. This is mostly due to late clinical presentation with advance disease.

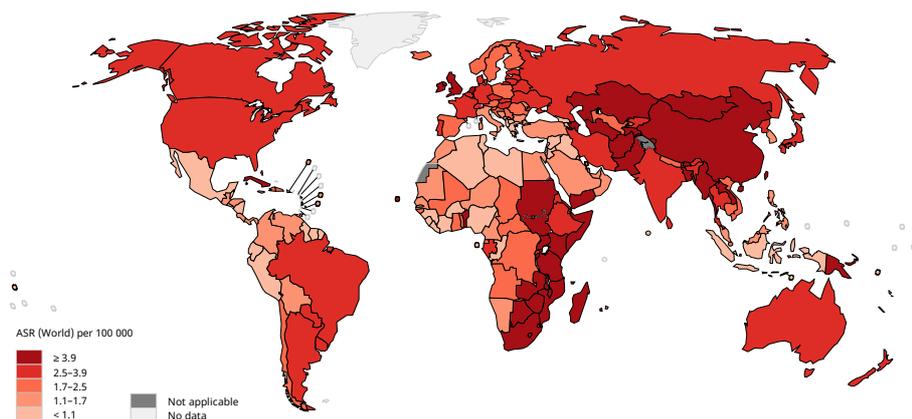


Figure 1: Esophageal cancer Age-standardized mortality Rate (ASR) per 100,000 worldwide-2018 (from GOLOCAN. Available from <http://gco.iarc.fr/>).

Esophageal cancer can be classified into two major histological subtypes: Squamous Cell carcinoma (ESCC) and Adenocarcinoma (EAC) [3]. ESCC occurs mainly in the upper and middle portions of esophagus, instead, EAC develops in the lower part, in proximity to the gastro-esophageal junction (GEJ). This classification has been recently confirmed and implemented at molecular level, where extensive genomic analysis of ESCC and EAC patients' tumor tissues clearly defined the two cancer subtypes as distinct molecular entities [4]. ESCC emerges as a disease more reminiscent of

squamous carcinomas of other organs than of EAC, which itself bears striking resemblance to chromosomally unstable (CIN) gastric cancer (Figure 2) [4].

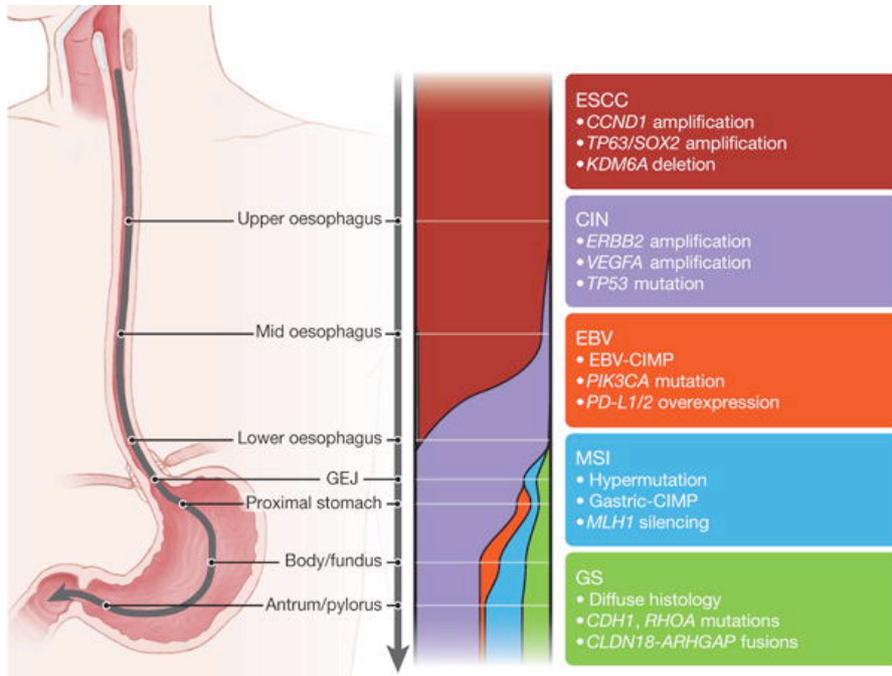


Figure 2: Schematic representation of the subtypes of esophageal carcinoma from the proximal esophagus to the distal stomach. The widths of the color bands represent the proportion of the subtypes present within anatomic regions. Key features of the different subtypes are indicated in associated text (adapted from: The Cancer Genome Atlas Research Network, 2017 [4]).

In Western countries, the incidence of esophageal adenocarcinoma has increased over the past two decades, however the cause of this alarming finding is unclear [2].

The American Joint Committee on Cancer (AJCC) TNM staging system considers EAC as a single entity (7th and 8th editions) [5;6], although different biological behaviors imply that EAC may be consistently heterogeneous. Only a deep molecular characterization of EAC heterogeneity can lead the way to solve the existing controversies on classification, prevention, early diagnostic programs and modalities of medical or surgical treatment of these tumors.

1.2 Pathogenesis of Esophageal Adenocarcinoma

A strong epidemiological association exists between esophageal adenocarcinoma and esophageal reflux disease (GERD) [7]. A complication of gastroesophageal reflux is Barrett's esophagus (BE), considered one of the most important risk factors for aggressive esophageal adenocarcinoma. Barrett's esophagus is a premalignant condition characterized by the replacement of the stratified squamous epithelium of the distal esophagus with intestinal type columnar epithelium. Chronic reflux of gastric, duodenal acid and bile is the initial stimulus needed to drive the development of a columnar phenotype [7]. The result of this process is the appearance of intestinal-metaplasia, conditions that can evolve into low-grade and high-grade dysplasia, and finally to EAC. The risk of progression from BE to EAC is 0.12-0.7% per patient per year [8;9]. The incidence of high-grade dysplasia and EAC increases up to 13.4% per year when low-grade dysplasia is present [10], and up to 25% for patients diagnosed with high-grade dysplasia [11].

Another risk factor for EAC is obesity, specifically in those individuals with predominately abdominal fat distribution. Hypertrophied adipocytes and inflammatory cells within fat deposits create an environment of low-grade inflammation that promotes tumor development through the release of adipokines and cytokines [12;13].

Cigarette smoking and alcohol consumption have been established as other two causative factors in EAC. Previous case-controls studies have found a strong association between cigarette smoking and EAC in Caucasians. There was a strong dose-response association; in particular a longer smoking cessation was associated with a decreased risk of all adenocarcinomas [14]. On the other hand, the association between alcohol consumption and EAC and/or BE is still unclear [15;16].

Case-control studies have found that a reduced risk of BE is associated with frequent intake of fruits and vegetables as well as the consumption of omega-3 fatty acids and fiber [17]. In a large prospective study in the USA on meat consumption and esophageal cancer risk, red meat intake and its cooking methods were positively associated with EAC [17].

The incidence of EAC increases with age and there is a strong male predominance, with up to eight men/one woman affected [18].

Genetic susceptibility to EAC has been studied for years. Familial clustering of Barrett's esophagus and EAC has been observed in 7% of tumor cases [19]. Genome-wide association studies (GWAS) have led to the identification of new genetic variants with a small effect that increase susceptibility to EAC development [20]. In particular, these studies identified several single nucleotide polymorphisms SNPs located in or near genes that regulate development and differentiation of the esophagus, stomach and intestine, such as *FOXP1*, *BARX1*, *FOXF1*, *CRTCI*, *GDDF1*, *ABCC5* and *CFTR* that were significantly associated with a higher risk of developing Barrett's metaplasia and esophageal adenocarcinoma [20;21;22].

In EAC prevention one practical issue is how to stratify cancer risk among patients with BE [23]. The current clinical strategy involves regular endoscopic examinations and biopsy pathology. This approach is controversial due to lack of specificity of a pathological diagnosis of dysplasia and recent data suggest that routine endoscopic surveillance in BE is not effective for early EAC detection. The debate on possible etiologic factors is still highly relevant and it is at the core of the controversies related also to prevention, early diagnosis programs, modalities of surgical treatments [24;25]. Moreover, the relationship between gastro-esophageal reflux and EAC via the sequence Barret's metaplasia, dysplasia and cancer is still highly discussed [26].

1.3 Treatment options

The treatment of patients with esophageal cancer and in particular esophageal adenocarcinoma is highly complex and requires an interdisciplinary approach. Patients with lesions limited to the mucosa can be diagnosed and treated with endoscopic resection [27]. In contrast, patients with adenocarcinoma might have a survival advantage when they undergo more extensive surgery; transthoracic esophagectomy with lymphadectomy for adenocarcinoma located in the distal esophagus, whereas transhiatal resection for tumors located at the gastroesophageal junction or gastric cardia [28;29].

However, surgical resection alone showed poor outcomes prompting considerable research on multimodality treatment options in esophageal adenocarcinoma. These options include neoadjuvant chemotherapy or chemoradiotherapy followed by surgery and adjuvant chemotherapy/radiotherapy following surgery. While the role of adjuvant therapy is questionable with current evidence not supporting its routine use [30], neoadjuvant treatment is the standard of care in resectable cancer [30]. Neoadjuvant therapy is a favored preoperative approach in the management of esophageal cancer and two different strategies can be applied: neoadjuvant chemotherapy alone or a combination of chemotherapy with radiotherapy in the neoadjuvant setting (neoadjuvant chemoradiotherapy). The most common neoadjuvant chemotherapeutic combination used in the treatment of EAC is cisplatin and 5-fluoruracil (cisplatin/5FU). Several studies have evaluated the role of cisplatin/5FU treatment followed by surgery compared with surgery alone, showing contrasting results. The Intergroup trial [31] showed no improvement in long-term survival with neoadjuvant chemotherapy, in contrast the MRC OE02 trial [32] showed a better overall survival in neoadjuvant group of patients. The reason for the difference in survival outcomes between these studies is unclear and probably depends on different tumor molecular behaviors.

More promising results are shown from neoadjuvant chemoradiotherapy, that however is considering the therapy with the higher toxicity profile [33;34;35]. The most recent neoadjuvant chemoradiotherapy approach is the Chemoradiotherapy for Esophageal Cancer Followed by Surgery Study (CROSS), consisting of weekly administration of carboplatin and paclitaxel with concurrent radiotherapy [36]. Long-term results have showed a statistically significant and clinically relevant increase in survival for both squamous cell and adenocarcinoma subtypes, with acceptable toxicity. On the basis of these results, the CROSS regimen is now a standard treatment in many countries [37].

For adenocarcinomas, several targeted therapies have been approved by the USA Food and Drug Administration (FDA) in the last few years. In 2010, based on the results of the Trastuzumab for Gastric Cancer (ToGA) study [38], the humanized anti-HER2 receptor monoclonal antibody (Trastuzumab) was approved as the first targeted therapy for use in human epidermal growth factor receptor 2 (HER2)-positive EAC (about 20% to 25% of tumors). More recently, the anti-VEGF receptor antibody Ramucirumab was approved in combination with paclitaxel in the second-line setting [39].

A novel and promising therapeutic approach for the treatment of esophageal adenocarcinoma is the immunotherapy. Immunotherapy aims to enhance the body's natural immune response by facilitating the targeting and destruction of cancer cells. Cytotoxic CD8 T-cells are able to recognize and eliminate cancer cells by inducing apoptosis or cell lysis [40]. As cancer cells mutate, they evolve to evade the anti-tumor immune response by developing immunosuppressive mechanisms. One of the suppressive immune evasion strategies employed by tumors is by controlling known natural immunosuppressive signaling pathways that are normally used as checkpoints to regulate immune cell function and prevent damage to the host during infections [41;42]. One such checkpoint signaling molecule is programmed cell death protein 1 (PD-1). PD-1 is an immune checkpoint that, under physiological conditions, suppresses the function of T cells, thus preventing autoimmunity [43]. The clinical rationale for immune checkpoint inhibition arises from the fact that tumor cells are able to overexpress PD-1, in order to decrease T cell-driven anti-tumor response [43]. Recently, FDA evaluated the effectiveness of the anti-programmed cell death protein 1 (PD-1) antibody (Pembrolizumab) in the second line setting for both esophageal adenocarcinomas and squamous cell carcinomas [44]. However, until now, despite the remarkable clinical achievements, the immunotherapeutic approaches resulted to be beneficial only for a subset of cancer patients [45;46;47].

1.4 Esophageal Adenocarcinoma: current classifications and prognostic factors

1.4.1 Staging of EAC

Esophageal cancer staging is defined by the American Joint Committee on Cancer (AJCC) TNM Staging System based on the depth of invasion of the primary tumor (T), lymph node involvement (N) and extent of metastatic disease (M) (Figure 3). The recent 7th and 8th editions AJCC TNM system harmonized EAC staging, and included all tumors in the “adenocarcinoma of the esophagus” chapter, which comprehends adenocarcinomas located in the distal thoracic esophagus, esophagogastric junction and within the first 5 cm of the stomach [5;6].

Stage	T	N	M	G
0	is (HGD)	0	0	1
IA	1	0	0	1-2
IB	1	0	0	3
	2	0	0	1-2
IIA	2	0	0	3
IIB	3	0	0	Any
	1-2	1	0	Any
IIIA	1-2	2	0	Any
	3	1	0	Any
	4a	0	0	Any
IIIB	3	2	0	Any
IIIC	4a	1-2	0	Any
	4b	Any	0	Any
	Any	N3	0	Any
IV	Any	Any	1	Any

Figure 3: Adenocarcinoma stage groupings. T=primary tumor characteristics; N= regional lymph nodes infiltration; M=distant metastases; G= histologic grade (G1 well differentiated, G2 moderately differentiated, G3 Poorly differentiated, G4 undifferentiated) (*adapted from: Rice TW et al., 2010 [5]*).

This approach is still controversial, since studies demonstrated differences of anatomy, metastatic patterns and survival depending on whether adenocarcinomas are predominantly esophageal or gastric in location, or according to immune-histological parameters [29;48].

Therefore, in recent years, different types of EAC classifications have been developed in order to resolve pending controversies and to improve the diagnosis and treatment of patients affected by adenocarcinoma of the esophagus, therefore improving the outcome of the disease.

1.4.2 Classification according to position: the Siewert classification

Siewert and colleagues proposed a classification based on anatomic-topographic features [49]. In particular, Siewert classification differentiated three tumor entities within cardia area [50]:

- EAC type I: adenocarcinoma of the distal esophagus, which usually arises from an area with specialized intestinal metaplasia of the esophagus, i.e. Barrett’s esophagus, and may infiltrate the esophago-gastric junction from above;
- EAC type II: true carcinoma of the cardia, arising from the cardia epithelium or short segments with intestinal metaplasia at the esophago-gastric junction;

– EAC type III: subcardial gastric carcinoma, which infiltrates the esophago-gastric junction and distal esophagus from below.

Siewert type II and type I carcinomas show different patterns of lymphatic dissemination, possibly implying different surgical strategies in regard to the resection area and to the level of lymphadenectomy. In particular, patients with EAC type I may achieve complete remission by radical transmediastinal esophagectomy and abdomino-thoracic esophagectomy, instead in patients with EAC type II and type III an extended total gastrectomy and transhiatal resection of the distal esophagus have proven to be the best surgical option [29].

1.4.3 Different morphological and histological spectrum of esophageal adenocarcinoma

Esophageal adenocarcinoma often arises as a stepwise progression of precursor lesions from low-grade epithelial dysplasia to high-grade dysplasia to adenocarcinoma. Therefore, dysplasia plays the major role in cancer development. There are two main types of glandular dysplasia involving the lower esophagus: adenomatous-type dysplasia and foveolar dysplasia. Based on characteristics of dysplastic lesions, EACs exhibit a wide morphological spectrum [51].

An intestinal-like morphology is characterized by the presence of pseudo-stratified columnar cells with hyperchromatic cigar-like shaped nuclei, generally in association with BE, with or without foci of adjacent adenomatous dysplasia (Figure 4A).

A gastric-foveolar-like morphology is characterized by the presence of cuboidal cells and round-oval nuclei with prominent nucleoli, generally associated to gastric intestinal columnar epithelium, not necessarily related with esophageal intestinal specialized epithelium. Dysplastic foveolar glands are positive for MUC5AC, rarely for MUC6, and negative for MUC2 and CDX2. High-grade foveolar dysplasia tends to be associated with poorly differentiated adenocarcinoma (Figure 4B).

Lastly, a cardiopyloric-like histotype, the less common type of esophageal adenocarcinoma, is characterized by large cells with clear cytoplasm, basal or centrally located nuclei, increase in nuclear/cytoplasmic ratio. Histologically, it consists of

packed pyloric-type tubules lined by cuboidal to low columnar cells with eosinophilic, ground glass cytoplasm devoid of an apical mucin cap. The glands are immunoreactive for MUC5AC and MUC6. These lesions more frequently exhibit high-grade cytoarchitectural atypia [51].

The signet-ring cell carcinoma (SRCC) represents another infrequent histotype, characterized by large vacuoles full of mucin that displace the nucleus to the cell's periphery (Figure 4C). It often occurs after the disease has advanced and consequently, shows worse prognosis than the more differentiated EACs [52].

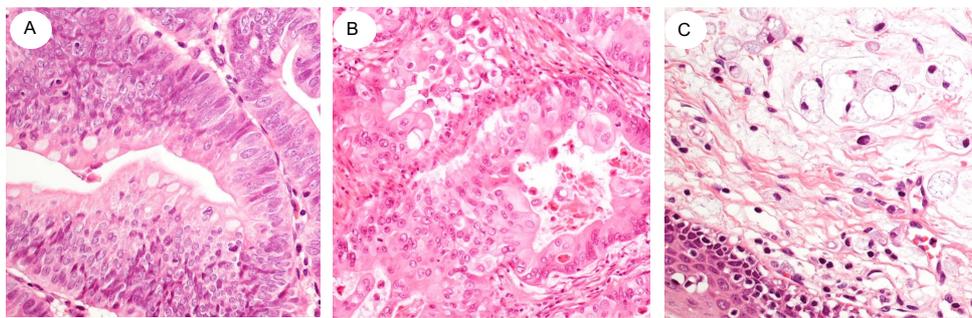


Figure 4: Morphological classification of EAC based on hematoxylin and eosin staining (20X). **A.** intestinal-like morphology, **B.** gastric-foveolar-like morphology, **C.** signet-ring cell carcinoma.

1.4.4 Histological classification according to Lauren

The Lauren classification, originally designed to classify gastric cancer, is a histopathological classification system with prognostic value expanded to esophageal adenocarcinoma [53;54].

It divides EAC into three subtypes:

- Intestinal type, which forms glands and resembles adenocarcinoma of the large intestine;
- Diffuse type, which consists of poorly cohesive cells with little or no gland formation, often containing various proportions of signet ring cells;
- Mixed type, which exhibits components of both intestinal and diffuse type carcinomas.

The majority of tumors can be classified as intestinal or mixed types, whereas the incidence of the diffuse types is lower than 15% [54]. Although the EAC diffuse type is seen with less frequency, it is associated with a significantly worse prognosis compared with intestinal type tumors (Figure 5) [53].

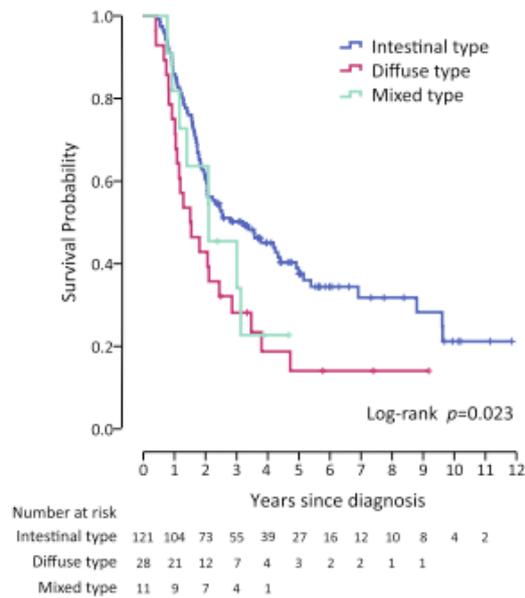


Figure 5: Overall survival according to the Lauren classification (*adapted from: Van der Kaaij RT, et al., 2017 [53]*).

Moreover, previous studies suggested a poorer response to neoadjuvant chemoradiotherapy and a higher risk of lymph node metastases in the diffuse type than in intestinal type carcinomas [55].

1.4.5 Immuno-histologic profiles of esophageal adenocarcinoma

Recent studies revealed the presence of differences in the biologic behavior of EAC, with the identification of three distinct entities on the basis of the expression pattern of tumor cytokeratin 7 (CK7) and cytokeratin 20 (CK20), the presence or absence of Barrett's intestinal metaplasia (BIM) and gastric intestinal metaplasia (GIM). These different immunoprofiles could allegedly correspond to a Barrett's esophagus-like type (CK7+/CK20-), a gastric cancer-like type (CK7-/CK20+), and a third undefined or mixed type (CK7+/CK20+) [29].

As the literature documents, different patterns of metastatic spread by means of the lymph nodes exist in EAC. It has been demonstrated that according to the presence or absence of BIM and GIM in the esophagus and cardia, adenocarcinoma show different patterns of lymphatic metastatization, which reflect different biologic and

carcinogenetic pathways: the BIM+/GIM- tumors spread to the thoracic stations, whereas abdominal lymph nodes are metastasized by BIM-/GIM- and BIM-/GIM+ tumors [48]. The BIM-/GIM- group includes patients with higher stage (III-IV) tumors and a more aggressive disease compared to the BIM+/GIM- ones. Therefore, EACs cannot be always considered a unique pathological entity and a gastroesophageal reflux disease-related tumor; other pathogenetic pathways should be taken into consideration [48].

1.5 Genomic characterization of EAC

Development of massively parallel and less costly sequencing techniques (next-generation sequencing, NGS) has improved our understanding of molecular factors associated with EAC development. The recent sequencing approaches including whole exome sequencing (WES) and whole-genome sequencing (WGS), allowed to identify frequent somatic structural rearrangements, copy number alterations and single-nucleotide mutations, demonstrating a high mutational frequency and high inter-sample heterogeneity in EAC [56].

The evolution from Barrett's esophagus to adenocarcinoma is underlined by continuous DNA damage caused by reflux and chronic inflammatory factors. In particular, previous studies have shown that the evolution from Barrett's esophagus to adenocarcinoma is dominated by loss of *TP53*, genome doubling, chromosomal instability (CIN) and a high frequency of chromothripsis events resulting in genome diversity and increase of the prevalence of focal amplifications and copy number gains/losses [57]. In fact, in consequence of *TP53* loss, whole-genome doubling occurs, which enhances tumor progression, requiring few other mutations [58]. These genomic catastrophes can occur at any stage and dramatically accelerate progression of BE to high-grade dysplasia and cancer (Figure 6).

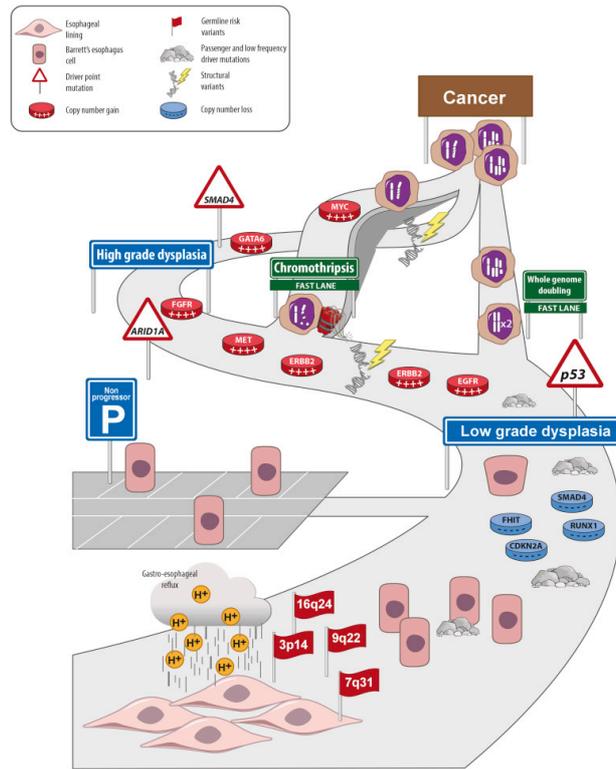


Figure 6: Paths of BE progression to EAC. Findings from next-generation sequencing studies indicate BE progression can accelerate via genome doubling, genome catastrophes, and other unknown mechanisms—even at early stages of tumor progression (*adapted from: Contino G. et al., 2017 [57]*).

The histological diversity of EAC is also mirrored by the genetic diversity of these tumors. Large-scale sequencing studies revealed three distinct mutational signatures for EAC [59]:

- enrichment for BRCA signatures with prevalent defects in the homologous recombination pathway;
- dominant T>G mutational pattern associated with a high mutational load and neoantigen burden;
- C>A/T mutational pattern with evidence of aging imprint.

Although the considerable level of genetic heterogeneity, the *TP53* gene is the most frequently mutated gene in EACs (80.5%), ahead of *CDKN2A* (30%) (TCGA, PanCancer Atlas, data updated to September 2019). However, the percentage of *CDKN2A* inactivation increased up to 76% of tumors when considering also epigenetic silencing mechanisms [60]. Therefore, the *TP53*-pathway is the most frequently mutated one in EAC.

Cell cycle regulation is affected not only by inactivation of *TP53* and *CDKN2A* but also by amplification of the 7q21 region, which harbors the *CDK6* gene that was found amplified in 35% of EACs [4; 61]. *CDK6* is a serine/threonine kinase whose activity is independent on binding to D-type cyclins (*CCND1*, *CCND2* and *CCND3*) and it phosphorylates and inhibits the pRB tumor suppressor allowing cell cycle progression during mitogen-dependent early G1-phase [62]. The deregulation of the cell cycle control pathway includes overexpression of cyclins (*CCND1* and *CCNE*) and protein kinases (*CDK6* and *CDK4*) or loss of CDK inhibitors, such as *INK4A*. Aberrant expression of these cell cycle regulators has been observed in many tumors including EAC as a result of chromosomal amplification [63;61]. The *MYC* gene, which regulates proliferation, is also amplified in approximately 30% EACs [64].

Among frequently altered genes are receptor tyrosine kinases (RTKs) of the *EGFR* family and their downstream mediators. Amplification of the *ERBB2* gene is the most prominent receptor alteration in EACs that was found in 32% of tumors [64].

In addition to *ERBB2* and *EGFR* gene amplifications, which can potentially activate the PI3K pathway, other mutations were reported in *PI3KCA*, *PI3KR1* and *PTEN* [56].

NGS analyses also revealed a dysregulation of the TGF β pathway, with its components frequently mutated. The more recurrently altered gene in this pathway is *SMAD4* (22%) (TCGA, PanCancer Atlas, data updated to September 2019).

EAC also shows loss-of-function mutations and copy number alterations (CNAs) in *ARID1A*, *ARID2*, *SMARCA4*, and *PBRM1* genes that encode components of the SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin-remodeling complex. The SWI/SNF complex is an evolutionarily conserved multi-subunit complex involved in chromatin restructuring that contribute to transcriptional activation and repression. Alterations of the SWI/SNF complex are not unique to EAC and are found in over 20% of human malignancies [60].

Except *TP53*, relatively few genes are recurrently point-mutated, demonstrating a high mutational frequency and high degree of genetic heterogeneity between EAC patients. The highly heterogeneous landscape, often difficult to correlate with clinical outcomes,

explains the problems encountered to date in finding suitable avenues for tailored therapies [56].

1.5.1 *TP53* and *CDKN2A*

TP53 is a tumour-suppressor gene located on the short arm of chromosome 17. This gene encodes the p53 protein, which acts as a transcription factor that plays a key role in cell cycle regulation, DNA synthesis inhibition, damaged DNA repair and apoptosis [65;66]. For its functions, it can be considered a driver gene in different human types of cancer.

TP53 is frequently inactivated by point mutations in many cancers [67]. These mutations are primarily missense variants that inhibit the binding of p53 protein to its target DNA regulatory regions, therefore blocking the p53-dependent transcription. *TP53* missense mutations are distributed in all coding exons of the *TP53* gene, with a strong predominance in exons 4-9, which encode the DNA-binding domain of the protein. About 30% of the mutations, found in many different cancers, fall within 6 “hotspot” residues (residues R175, G245, R248, R249, R273, and R282) [68]. In heterozygous cancer samples, where both wild-type (*wt*) and mutant alleles exist, mutant p53 can antagonize *wt* p53 in a dominant negative manner [69;70]. The generally accepted mechanism behind this dominant negative effect is the shutdown of wild-type p53 function because of heteromerization with mutant p53 [69]. However, such a heterozygous state is often transient, as *TP53* mutations are frequently followed by loss of heterozygosity (LOH) during cancer progression [71]. Nevertheless, accumulating evidence supports that a dominant negative effect can play an oncogenic activity by a gain of function mechanism [72].

For these reasons, *TP53* is intensively studied in cancer and recent analyses support the observation that *TP53* genetic status is not a prognostic but a predictive biomarker influencing survival only in the presence of effective chemotherapy in colorectal cancer [73]. In fact, *TP53* gene mutations were associated with worse survival for patients treated with adjuvant 5-fluorouracil (5FU) [74;75]. Indeed, patients with a normal *TP53* genetic status may experience notable benefits from neoadjuvant chemotherapy with

cisplatin/fluorouracil, whereas those with a mutant *TP53* status appear to be at risk for lack of response [74;75].

Given the important role played by p53 in tumor suppression and chemotherapeutic drug response, a number of compounds, such as STIMA-1, PRIMA-1, MIRA-1, RITA and others, have been identified to restore the activity of mutant p53 [76]. PRIMA-1 and its analog APR-246 are the more investigated compounds in this category. APR-246 was tested in EAC cells harboring mutant p53 and found to upregulate p53 target genes and induce apoptosis [76]. An initial phase I clinical trial has shown APR-246 to be safe in humans. Phase Ib/II study (ClinicalTrials.gov Identifier: NCT02999893) evaluating the efficacy of APR-246 in the treatment of advanced and metastatic esophageal or gastro-esophageal junction cancers is currently ongoing.

Another gene frequently mutated in EAC is *CDNK2A*, a p53 regulator which encodes two tumor suppressors with different functions: p16INK4a and p14ARF. p16INK4a is a specific inhibitor of the cyclin D/CDK4/6 complexes, whose inhibition leads to disruption of normal cell cycle and uncontrolled cell growth [77]. p14ARF interferes with the proper p53 response, because p14ARF is a critical upstream regulator of p53 that activates p53 protein by blocking its Mdm2-mediated degradation [78]. Restoring p53 activity by inhibiting the interaction between p53 and MDM2 represents an attractive approach for cancer therapy. To this end, a number of small-molecule p53-MDM2 binding inhibitors have been developed, such as Nutlin-3 [79].

However, Nutlin-3 treatment reduced viability and induced p53-mediated apoptosis only in cancer cells with wild-type p53 protein.

1.5.2 TGF- β pathway – associated genes

The transforming growth factor beta (TGF- β) pathway is implicated in regulation of cell growth, apoptosis, differentiation, development and inhibition of proliferation and inflammation in normal tissues [80;81]. However, in EAC development the TGF- β pathway can facilitate epithelial to mesenchymal transition (EMT), invasion, and metastasis [82]. TGF- β signal transducers (SMADs) are commonly lost in EACs and, among them, *SMAD4* gene is greatly affected [56]. The product of this gene forms transcription complexes with other members of the SMAD protein family and regulates TGF β -mediated transcription [83]. Interestingly, *SMAD4* is primarily mutated in EAC,

but not in high grade dysplasia (HGD), providing a genetic distinction between EAC and HGD [84]. Loss of SMAD4 expression by immunohistochemistry was found in 10% of EAC cases, and correlated with increased postoperative recurrence and poor overall survival [85]. Although mutations in *SMAD4* are quite common events in esophageal adenocarcinoma, the principal causes of reduced SMAD4 expression are represented by promoter hypermethylation, promoter deletion and protein modification [86].

1.6 Technological advantages in the study of cancer: digital sorting of pure cell populations

The development of new sequencing technologies (NGS) is arguably one of the most significant technological advances of the last 30 years in the field of genetics. The second-generation sequencing platforms have advanced rapidly to the point that several genomes can now be sequenced simultaneously in a single instrument run in few days [87]. NGS has been crucial in the identification of variants characterizing the genomic landscape of thousands of cancer genomes across many disease types. NGS technologies hold the potential to reveal the molecular underpinnings of tumor biology with the accuracy required for clinical implementation [88]. However, when the input DNA is a mixture of normal and tumor cells there is an inherent trade-off between sensitivity and specificity, further complicated by the fact that the tumor DNA derive from subpopulations with different genetic characteristics. This can dilute the signal associated with quantitative genomic features like copy number alterations (CNAs), loss of heterozygosity (LOH) and simple homozygous/heterozygous status of a variant in a tumor cell subpopulation. This masks variant alleles/copy number alterations to values close to background noise or below the detection limit and strongly impairs their accurate detection by NGS alone. Since these sequencing approaches have been unable to resolve differences in complex mixture of cells, such as heterogeneous tumors [89], innovative high-throughput cell sorting technologies were recently combined with NGS assays to isolate pure tumor cells and obtain unambiguous genetic analysis results [90]. In the last decade, different cell-sorting systems were developed to identify and recover specific individual cells of interest from complex, heterogeneous samples. In this thesis, we approached the study of pure cancer cells exploiting one of these

cell-sorting systems, named DEPArray™ (Menarini Silicon Biosystems) [90]. DEPArray™ technology is based on the ability to isolate rare single cells from low cell count samples, whereas, most traditional flow-cytometry instruments require several thousand cells to prime the system. For its characteristics, DEPArray™ system allows to obtain results in a wide range of applications, including: analysis of single circulating tumor cells (CTCs) to characterize mutational heterogeneity, examine mechanisms of disease progression, monitor drug response [91] and analysis of pure tumor cells from FFPE samples to achieve a more accurate assessment of prevalence of various mutations. In the latter case, the cell sorting system can be used to separate tumor cells from stromal or immune cells that are present in the tissue section. A slice or a punch from an FFPE sample can be disaggregated to be transformed into individual cell suspensions and stained for DNA content and cytoskeletal markers, such as cytokeratin proteins (CK) and vimentin (VIM) [92]. Cytokeratin (CK) are the largest and most diverse class of intermediate filaments which constitutes cytoskeleton components. At least twenty different CKs can be expressed. They are markers of normal epithelial differentiation, but they can be used as diagnostic tool to detect carcinoma cells. Epithelial cancer cells often express CKs 8, 18 and 19 that are used in combination with DAPI signal to discriminate tumor cells in a heterogeneous tumor tissue [92]. However, besides cytokeratin, carcinoma cells sometimes express variable levels of vimentin, depending on the epithelial-mesenchymal transition status in cancer [93]. In contrast, mesenchymal cells, such as fibroblast, chondrocytes, macrophages, and endothelial cells, express only the cytoskeletal marker vimentin, a type III intermediate filament whose expression is abundant in tumor stroma [94]. Therefore, using these cell sorting capabilities, tumor cells (CK+/VIM-) can be separated from the stromal cells (VIM+/CK-) present in a tumor tissue section (Figure 7).

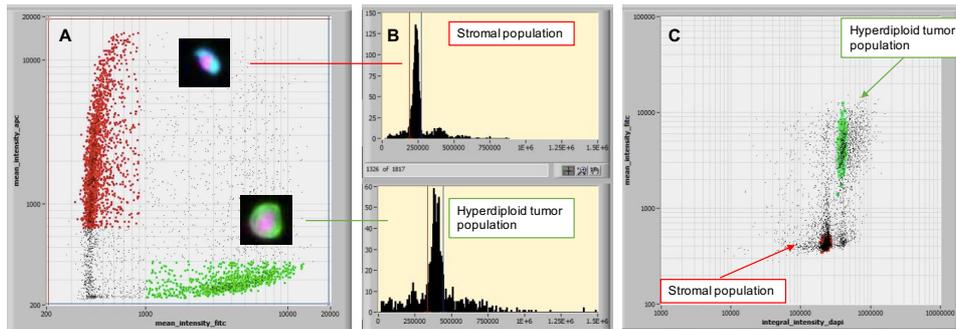


Figure 7: DEPAArray™ image-based sorting. **A.** Scatter plot of disaggregated FFPE sample showing two well defined populations, based on cytokeratin and vimentin markers signal: tumor (CK+/VIM-) and stromal (VIM+/CK-) cells. **B.** Histograms of integral intensity DAPI, which is proportional to the amount of DNA contained in the cell. Ploidy is inferred by the ratio of integral intensity DAPI of a cell population with reference to the integral intensity DAPI of stromal cells. **C.** Scatterplot of cytokeratin signal intensity measured against DAPI intensity.

In this thesis, I will outline how this cell sorting technology, combined with NGS analysis, helped us to investigate the cancer genome complexity from FFPE esophageal adenocarcinoma samples.

1.7 Micro-RNAs and Esophageal Adenocarcinoma

Progression of Barrett's esophagus toward esophageal adenocarcinoma is underlined by the somatically heritable deregulation of genes that can be a consequence of mutations, structural alterations, epigenetic events, transcriptional and post-transcriptional mechanisms [57]. Among these factors there are microRNAs (miRNAs), small single-stranded noncoding RNAs of about 18–28 ribonucleotide lengths with the ability to regulate cell processes as differentiation, proliferation or apoptosis. They mainly act by targeting 3' untranslated region of the mRNA, resulting in either translational repression, mRNA degradation, or mRNA cleavage, depending on the complementarity between the miRNA and their target mRNA [95]. MiRNAs are generated by a mechanism that involves the transcription of a long precursor (pri-miRNA) and is operated by different groups of enzymes in the nucleus or cytoplasm. The pri-miRNA, located in the nucleus, is converted in pre-miRNA through the cleavage activity of the Drosha enzyme. The produced pre-miRNA is exported to the cytoplasm by Exportin 5. Upon its arrival into cytoplasm, the pre-miRNA is processed in 18-22 nucleotide

miRNA duplexes by the cytoplasmic RNase III Dicer. Usually, one strand of this duplex is degraded, whereas the other strand accumulates as a mature miRNA. In some cases, both strands of the duplex are viable and become functional miRNA that target different mRNA molecules. When two mature microRNAs originate from opposite arms of the same pre-miRNA and are found in similar amounts, they are denoted with a -3p or -5p suffix. Only one strand (mature miRNA) is stabilized and incorporated within the heterotrimeric complex of Argonaute-2 (Ago-2)/TAR RNA-Binding Protein (TRBP)/protein kinase R-activating protein (PACT) as an RNA-Induced Silencing Complex (RISC). The single strand miRNA incorporated acts as a template to recognize complementary messenger RNA (mRNA) transcript. Once found, the Argonaute-2 protein activates and cleaves the mRNA target. This is a key process in gene silencing [96;97].

In the last decade, miRNAs have been shown to regulate several processes in normal physiology and miRNA dysregulation has been reported in many diseases, including cancer. Owing to their stable expression in serum, plasma, saliva and other body fluids, they can be potentially good diagnostic, prognostic and predictive biomarkers. Furthermore, since miRNAs act as oncogenes or tumor-suppressor genes and regulate individual biological pathways by regulating the number of different downstream molecules, they make attractive candidates as therapeutic targets. To date, several studies have revealed distinct miRNA expression profiles between tumor and normal or premalignant tissue in EAC, identifying promising miRNAs with different roles at multiple steps of tumor progression. In particular, Feber and colleagues [98] performed miRNA expression arrays on a small cohort of patients that included esophageal tissues from normal squamous epithelium, high-grade dysplasia, Barrett's esophagus and EAC. They found miRNA profiles that were distinct for each tissue type and could distinguish normal from malignant tissue. Interestingly the miRNA profile of EAC was found to be similar to BE, reflecting the tissue specificity of miRNAs with the ability to discriminate between squamous and columnar tissue. Recently, numerous studies have identified miRNAs which are aberrantly expressed in esophageal cancer: for example, miR-21, -145, -192, are upregulated, whereas miRNA-31, -203, -205 and let-7 which are often downregulated compared to normal esophageal tissue [99]. Several studies have shown that aberrant expression of specific miRNAs correlated with patient survival in esophageal cancer, presence of metastasis and response to neoadjuvant therapies [100]. Although altered miRNA expression profiles are intensively studied in

tumors of the esophagus, many analyses focused on esophageal squamous cell carcinoma or only on adenocarcinoma developed from Barrett's esophagus. To date there are no studies demonstrating a specific miRNA signature distinguishing different histological pattern of EAC (i.e. presence/absence of intestinal metaplasia or Lauren's cancer subtypes). The relatively small number of EAC patients included in these studies also makes it difficult to obtain consistent results, and prevents the detection of any robust association with clinical variables.

2. Aims

Understanding the molecular landscape of cancer can be challenging because of the genetic heterogeneity both within and between patients. Here, I reported the different molecular approaches used to examine tumor heterogeneity and solve crucial controversies on classification, staging, medical and surgical therapy of esophageal adenocarcinoma (EAC).

Starting from these considerations, the aims of this thesis were:

(i) to investigate intra- and inter-tumor heterogeneity in whole tumor samples and sorted cell populations, and correlate tumor heterogeneity/mutational profiles with clinical outcomes, i.e. recurrence and survival. To pursue this aim, we combined a high-throughput cell sorting/recovery workflow with next generation sequencing (NGS) technologies to separate and analyze the different cancer cell populations vs normal (stromal) cells (**Part I**).

(ii) to characterize the dysregulated miRNAs in EAC, compared to the non-neoplastic tissue counterparts, and describe the signatures of the dysregulated miRNAs in the different EAC histological subtypes. We correlated miRNA expression with clinical outcomes and evaluated the protein expression level of specific targets. (**Part II**).

3. Materials and Methods

Part I

3.1 Genetic analysis of esophageal adenocarcinoma

3.1.1 Subjects

The first case analyzed in this study was a 53-year-old white woman who, in January 2010 underwent subtotal esophagectomy and extended thoracic and abdominal lymphadenectomy for adenocarcinoma of the esophagus, which was classified as Barrett's type due to the presence of intestinal metaplasia in the esophageal mucosa and the absence of gastric intestinal metaplasia in the stomach. According to Lauren's classification, the cancer was intestinal type, G2 T3N0M0, stage 2B (detailed clinical features are summarized in Table 1, Row 1, EAC1). Adjuvant chemotherapy was indicated because of the stage, serosa and perineural invasion (5 cycles of cisplatin). In the autumn of 2010, she was diagnosed with sub-centimeter right lung metastases. She was treated with capecitabine and trastuzumab (HER2-positive tumor cells) as a second line of chemotherapy, and a clear reduction in the nodules was reached. Trastuzumab therapy was adopted for maintenance and achieved good control of the disease until 2014, when the progression of pulmonary metastases became evident. Because of her lack of tolerance to drug chemotherapy, in April and December of 2015, she underwent video-assisted thoracoscopic surgical (VATS) resection of two metachronous right lung metastatic clusters. She was treated with a combination of cancer therapies (Paclitaxel/Ramucirumab, Irinotecan, Capecitabine, Epirubicin) until July 2018, when she underwent ago-biopsy of a third pulmonary nodule that was confirmed as EAC metastasis. On June 2019 she was alive, again under maintenance chemotherapy, but in disease progression.

Subsequently, the genetic analysis was extended to a cohort of 37 EAC patients submitted to primary (no neoadjuvant therapy) surgical resection (clinical features are summarized in Table 1). We retrospectively utilized formalin fixed paraffin embedded samples (FFPE) of the surgical specimens collected between 2005 and 2016.

Hematoxylin and eosin (HE)-stained slides from all FFPE blocks were examined to identify the tumor areas.

The study received approval (# L3P1223) from the Ethical Committee “Comitato Etico IRST IRCCS AVR (CEIIAV)”- Italy (Reg. Sper. 109/2016 Protocol 7353/51/2016) and written informed consent was obtained from all patients before inclusion in the study.

3.1.2 Whole-exome sequencing (WES)

Whole Exome sequencing (WES) was performed on nine esophageal adenocarcinoma tissues of patients submitted to primary surgical resection and on three metachronous chest metastases developed in patient EAC1.

DNA was extracted from 3 10- μ m thick sections obtained from formalin-fixed paraffin-embedded (FFPE) blocks of tumors using the QIAMP DNA Mini Kit (Qiagen, Hilden, Germany). Dual-index paired-end libraries followed by exome enrichment were prepared according to the Nextera[®] Rapid Capture Enrichment protocol (Illumina Inc., San Diego, CA, USA) using the coding exome exon kit (Nextera). 150 ng of genomic DNA from each tissue sample was fragmented to a size of 250-300 bp with adapter sequences added by tagmentation, according to the manufacturer's instructions. Following purification, the tagmented DNA was amplified by PCR for 10 cycles with unique indices for each sample. Libraries were purified and validated for appropriate size on 2100 Bioanalyzer DNA 1000 chip (Agilent Technologies, Santa Clara, California, USA). The single DNA libraries were also quantified using Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, Waltham, Massachusetts, USA), prior to normalization and pooling. Whole exome capture of 500 ng of each indexed DNA library was performed by pooling groups of 8 samples and hybridizing to the capture probes for 16 hours at 58°C. The captured regions were then bound to streptavidin magnetic beads, washed to remove any non-specific bound products and eluted from the beads. Wash and elution procedures were repeated to further enrich the targeted regions. The enriched library is purified again prior to quantification by Qubit dsDNA HS Assay kit (ThermoFisher Scientific) and the library size was checked on 2100 Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Santa Clara, California, USA). Each pool was normalized to 1.3 pM and then sequenced on an Illumina NextSeq500 platform (Illumina) at 150 bp paired ends.

3.1.3 Bioinformatics data analysis

The bioinformatics analysis of WES data began by utilizing raw reads produced by NextSeq500 (Illumina) sequencing platform, that were checked using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and aligned using BWA (bio-bwa.sourceforge.net) to the human reference (hg19). PCR-duplicated reads were marked and removed using Picard (<http://broadinstitute.github.io/picard/>). Mapping and coverage statistics were calculated using SAMtools (<http://samtools.sourceforge.net/>). Putative somatic variants, including SNPs and small insertions/deletions (indels), were identified using GATK3 software (software.broadinstitute.org/gatk/). The raw mutation calls were filtered to exclude false calls based on base quality and allele frequency of mismatch bases. The identified mutations were further annotated with Ensembl VEP (www.ensembl.org/Tools/VEP) and their pathogenicity was evaluated using PolyPhen2 (genetics.bwh.harvard.edu/pph2), Provean (provean.jcvi.org), FATHMM (<http://fathmm.biocompute.org.uk/>) and MutationTaster (www.mutationtaster.org) for missense mutations and Human Splicing Finder (www.umd.be/HSF/) and ESEfinder v3.0 (krainer01.cshl.edu/cgi-bin/tools/ESE3) for putative splicing alterations. The identified mutations were filtered by a Minor Allele Frequency (MAF) < 0.01 and based on their functional impact using a customized program for prioritizing variants.

In parallel, the copy number analyses were performed on WES data with EXCAVATOR2 software (sourceforge.net/projects/excavator2tool), using as control whole exome data derived from a pool of normal (non-tumor) samples.

3.1.4 SNPhylo: phylogenetic analysis of primary tumor and metastases

WES data were used for a phylogenetic analysis of the primary lesion and three chest metastases of patient EAC1. After converting the data in SNP File Format, we computed the phylogenetic tree with SNPhylo, using the default parameters, according to Lee TH et al. [101]. To calculate the phylogenetic distance, the program sums each branch length between two samples. Therefore, the length of a branch represents the estimate of evolutionary distance defined as the number of elementary substitution events that occurred during the time of divergence of two samples, irrespective of the direction of time.

3.1.5 LOH detection from WES data on genomic DNA

WES data from the DNA derived from tumor samples were obtained through enrichment with Nextera capture kit (Illumina) and sequencing with Illumina NextSeq500 (Illumina), as mentioned above (paragraph 3.1.2). LoFreq software ([csb5.github.io/lofreq/](https://github.com/lofreq/lofreq)) was used to get variant calls, which were later annotated with Ensembl VEP (www.ensembl.org/Tools/VEP). Variant frequencies and annotations were used to produce a B-allele frequency plot (BAF) for investigating the presence of LOH regions. Given the lack of the matching normal sample and the subsequent inability to figure out germline heterozygous variants, only highly polymorphic sites ($GMAF > 0.1$) were used in the plot. Moreover, variant frequencies greater than 90% or lower than 10% were filtered out to exclude germline homozygous SNPs and background noise. Removal of these variants does not affect the LOH identification, since the expected tumor frequencies of LOH variants in unsorted samples are lower than 100% and greater than 0% due to contamination of normal tissue. In addition, to improve the interpretation of frequency trends along the genome, allele frequencies were smoothed using a kernel density estimation (KDE) in 20Mb overlapping bins. LOH regions were visually predicted focusing on stretches of consecutive variants with frequencies deviating from heterozygous state (~50%).

3.1.6 Cell sorting protocol

The high throughput cell sorting protocol was carried out according to the DEPArray™ (Menarini Silicon Biosystems) technology, based on the ability of a non-uniform electric field to exert forces on neutral, polarizable particles, such as cells, suspended in a liquid. Cells are introduced into a microfluidic chip and trapped in dynamically controlled dielectrophoretic (DEP) cages. Once immobilized, cells are identified using a combination of immunofluorescent staining and LED optical imaging. Selected cells are then precisely driven using DEP cage motion to a collection tube [91]. The isolation of carcinoma cells from FFPE archived samples, using the cytoskeletal markers, such as cytokeratin (CK) and vimentin (VIM) allows the recovery of the different cell populations [90]. Each section of 50- μ m was collected in a nylon biopsy bag inside a 50-ml conical tube, dewaxed by three sequential 10 min incubations with xylene and then rehydrated via decreasing ethanol washes. After washing with deionized water, the section was immersed in 10 mM sodium citrate buffer (pH 6.4) for 5 min at room

temperature and heat-treated in the same pre-warmed buffer for 1 h at 80°C. After cooling to room temperature, the section was washed with three sequential 5 min incubations with the RPMI medium (Life Technologies, Carlsbad, CA, USA). To obtain the cell suspension, the section was incubated in a solution of 10 ml 0.1% collagenase I-A (Sigma-Aldrich, St. Louis, MO, USA) and 0.1% dispase (Life Technologies) at 37°C. The dissociation process was stopped after 45 min by placing the sample tube on ice. The cell suspension was resuspended by pipetting and transferred through a 30-µm mesh nylon filter into a 15-ml conical tube. The cell suspension was washed two times with ice-cold PBATw (PBS-1%BSA-0.05%Tween20 buffer) by centrifugation at 1000 g for 5 min. The cell pellet was resuspended in 1 ml of ice-cold PBATw, and an aliquot of 5×10^5 cells was incubated with 100 µl of the primary monoclonal antibody mixture containing anti-keratin (CK) MNF116, IgG1 (DAKO, Glostrup, Denmark) (final concentration 3.2 µg/ml), anti-keratin (CK) AE1/AE3, IgG1 (Millipore-Chemicon, Burlington, Massachusetts, USA) (final concentration 10 µg/ml) and anti-vimentin (VIM) 3B4, IgG2a (DAKO) (final concentration = 3.1 µg/ml) in PBATw. After 30 min at 4°C, the cells were washed twice with ice-cold PBATw and 100 µl premixed secondary reagents Alexa Fluor® 488 Goat Anti-Mouse IgG1 (Life Technologies), a final concentration of 2.5 µg/ml for the keratin detection and Alexa Fluor® 647 Goat Anti-Mouse IgG2a (Life Technologies), a final concentration of 2.5 µg/ml for the vimentin detection, and PBATw was added to the pellet. The incubation occurred for 60 min in the dark at 4°C and was followed by two washes with ice-cold PBATw. A DNA staining solution containing 10 µM DAPI (Sigma-Aldrich) in PBATw was added, and after an incubation of 30 min at 37°C, the cells were washed twice with PBATw with a 5 min centrifugation at 1000 g, and the pellet was resuspended in the same buffer.

For cell sorting on the DEPArray™, a small amount of the labeled cell suspension was washed twice with 1 ml of SB115 buffer (Menarini Silicon Biosystems). The pellet was resuspended in the same buffer and an aliquot, corresponding to approximately 24000 cells, was loaded into the DEPArray™ A300K cartridge (Menarini Silicon Biosystems). Based on the fluorescent staining and the DNA content, a precise number of homogeneous cells in the tumor (CK+/VIM-) and stromal populations (VIM+/CK-), together with pools of mixed cells (unsorted), were recovered in different PCR tubes. The measure of DNA content is proportional to the DNA index, determined comparing the integral-intensity DAPI signal of tumor population with that of stromal fraction, used as reference.

After lysis, reagents were added to the same tubes to prepare the DEPArray™ OncoSeek libraries (Menarini Silicon Biosystems) or to perform Low Pass Whole Genome Analysis in order to detect copy number alterations (CNAs).

3.1.7 Assessing FFPE DNA quality

The DNA integrity of the cells suspension was determined using the DEPArray™ FFPE QC Kit (Menarini Silicon Biosystems), before cell sorting. DNA quality assessment is performed by qPCR-based assay yielding a QC score defined as the ratio between the quantification of 132 bp amplicon and a 54 bp amplicon. This value provides a useful tool to evaluate the sample quality and possibly select samples with QC quality optimal for subsequent NGS analysis (QC ≥ 0.20).

For each sample, the effectively amplifiable template (EAT), a measure of the amount of DNA competent for amplification, was estimated by multiplying the QC score by the ploidy (assessed by cell sorting) and by the number of cells recovered.

3.1.8 OncoSeek genetic analysis

Illumina-compatible, targeted NGS libraries from DEPArray™-sorted cell lysates were obtained using the DEPArray™ OncoSeek Panel (Menarini Silicon Biosystems), according to the manufacture's protocol. This panel allows the simultaneous detection of single nucleotide variants (SNVs), indels and copy number alterations (CNAs) from 63 clinically-relevant oncology-related genes (Table 2). Each library was diluted 1:10,000 and then quantified in triplicate by qPCR using the KAPA Library Quantification Kit (Hoffmann-La Roche, Basilea, Switzerland) following the user's instructions. All libraries were pooled and NGS was performed using MiSeq v2 (150 PE) reagents on a MiSeq instrument (Illumina) according to the manufacturer's protocol.

The FASTQ paired-end reads were trimmed using Cutadapt (cutadapt.readthedocs.io) according to the manufacturer's protocol to remove synthetic primers from overlapping amplicons. The trimmed reads were aligned to the human reference genome (hg19) using the BWA software (bio-bwa.sourceforge.net). The alignment and coverage statistics were obtained using Samtools (samtools.sourceforge.net) and BEDTools (bedtools.readthedocs.io/en/latest) packages. After filtering to discard the partial, poorly aligned and unmapped reads, variant calls were obtained using LoFreq software

(csb5.github.io/lofreq). The resulting variants were annotated using Ensembl VEP (www.ensembl.org/Tools/VEP). The copy number alteration analysis of the OncoSeek data was performed using sorted populations as tests and a set of stromal cell pools from different samples as controls. For the copy number calling, the reads mapping to target amplicons of the DEPAArray™ OncoSeek panel were counted. Then, the read counts were normalized using the following 2-step procedure: 1) between-sample normalization using the total number of aligned reads and 2) within-sample normalization using a LOWESS fitting of read counts with respect to the first component explaining >90% variation between regions in the control samples. The fold changes were computed by dividing the normalized counts in the test samples by the baseline and are represented by the median value of the normalized counts per amplicon across the control samples. The final copy number calls per gene were obtained by calculating the median fold changes of all gene-specific amplicons.

3.1.9 Whole-Genome Low-Pass sequencing for CNA analysis

Recovered cells were lysed using SB LysePrep™ Kit (Menarini Silicon Biosystems) and 46 µl of LowTE buffer (TE 0.1X) were added to the tube. The sample was then transferred into a microTUBE-50 AFA Fiber Screw-Cap for fragmentation by Covaris M220 instrument (Covaris, Woburn, Massachusetts, USA) for 3 min and 52 sec (pick power:50, duty factor:20, cycles/burst:200) to obtain a 150–200 bp fragment size. Libraries were prepared using Accel-NGS® 2S PCR-Free DNA Library kit (Swift Biosciences, Ann Arbor, MI, USA) according to the manufacturer's instructions.

20 µl of library were amplified as following: 6 µM of amplicon PCR forward primer (5'-AATGATACGGCGACCACCGAGATC-3'), 6 µM of amplicon PCR reverse primer (5'-CAAGCAGAAGACGGCATACTGA-3') and 2× KAPA HiFi HotStart Ready Mix (Hoffman-La Roche). The PCR cycling conditions were 98 °C initial denaturation for 45 sec, followed by 16 cycles and 15 cycles for ~100 cells and ~300 cells, respectively at 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 1 min, and a final extension at 72 °C for 1 min. The products were cleaned up with 0.75X Agencourt AMPure XP beads (Beckman Coulter Genomics, Chaska, USA) according to the manufacturer's protocol and eluted in 20 µl Low TE (Swift Biosciences).

Libraries were normalized and pooled to 4 nM based on qPCR quantification. Pooled samples were denatured and diluted to a final concentration of 12 pM. All samples were

multiplexed and sequenced in a single lane on the MiSeq using 2×100 bp paired-end sequencing with the MiSeq Reagent Kit V3 (Illumina).

The BWA algorithm was used to align the reads to the hg19 human reference genome. PCR duplicates and secondary alignments were filtered out using Picard MarkDuplicates (<http://broadinstitute.github.io/picard/>) and Samtools (samtools.sourceforge.net). Control-FREEC algorithm was used to obtain copy-number calls, using the mode without control sample independently for all libraries. Read counts were corrected by GC content and mappability (uniqMatch option). Main ploidy level was estimated for each library based on best fitting of profiles to underlying copy number levels.

3.1.10 TA-cloning for *SMARCA4* PCR product

40 ng of genomic DNA from the third metastasis biopsy of patient EAC1 was subjected to PCR amplification of *SMARCA4* gene (exon 32), using KAPA Taq HotStart PCR Kit (Hoffman-La Roche) in a final volume of 50 μ l, according to the manufacturer's protocol. The primer sequences of exon 32 are reported in Table 3. The 213 bp amplicon was purified with Millipore PCR clean-up plates (Millipore-Chemicon) and cloned using the pCR 2.1 TA Cloning Kit (Invitrogen- Life Technologies), according to standard procedure. The purified amplicon was combined with 2 μ l of pCR®2.1 vector (25 ng/ μ l), 1 μ l of 10X Ligation Buffer, 1 μ l of T4 DNA Ligase (5 units) and water for a final volume of 10 μ l. The ligation reaction was then incubated 16 hours at room-temperature. The ligation reaction was then gently pipetted into 50 μ l of One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen- Life Technologies) competent cells and incubated on ice for 30 min. The reaction was then incubated, without mixing or shaking, for exactly 30 seconds at 42°C (heat shock step). The transformed mixture was immediately placed again on ice for 3 min. Next, 250 μ l of pre-warmed S.O.C. medium was added to each reaction tube and the mixture was incubated at 37°C for 1 hour at 200 rpm in a shaking incubator. The content of each transformation vial was placed on an ampicillin (50 μ g/ml)-LB agar plate. After an incubation of 16 hours at 37°C, the clones were picked up and grown in ampicillin (50 μ g/ml)-LB broth at 37°C for 16 hours. The plasmid DNA of the single collected colonies was extracted with Plasmid Mini Kit (Qiagen), according to the manufacturer's instruction and directly sequenced on both strands with primers specific for pCR®2.1 vector, forward M13 and reverse

M13 (Table 3). The purified sequencing products were analyzed on 3730 DNA Analyzer (ThermoFisher Scientific, Waltham, Massachusetts, USA).

3.1.11 Sanger sequencing

The genomic DNA was extracted from different FFPE tissue sections of the same tumor tissue block with QIAamp DNA Mini Kit (Qiagen). KAPAHiFi HotStart (Hoffman-La Roche) was used for genomic DNA amplification, according to the manufacturer's protocol. The *TP53* PCR products (the specific primer sequences of *TP53* exons are reported in Table 3) and the plasmid templates (according to paragraph 3.1.9) were directly sequenced on both strands using the BigDye v1.1 kit (ThermoFisher Scientific) on 3730 DNA Analyzer (ThermoFisher Scientific). Each sequencing reaction was performed in a final volume of 10 μ l, using 2 μ l of 5X Sequencing Buffer, 0.5 μ l of Big Dye™ Terminator v1.1 Ready Reaction mix, 1 μ l of specific 10 μ M primer (forward or reverse), 1 μ l of DNA template (purified PCR products or plasmid DNA) and 5.5 μ l of water. The reaction was conducted in the 2720 Thermal Cycler (ThermoFisher Scientific) setting the following program: 1 min at 96°C; 10 sec at 96°C, 4 min at 60°C for 35 cycles. Electropherograms were visualized with Chromas version 2.0 (Chromas, Technelysium, South Brisbane, Australia) and Sequencer version 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA).

3.1.12 Droplet digital PCR (ddPCR)

The ddPCR experiments were performed for TP53_R273H (dHsaMDV2010109) and CDKN2A_R58* (dHsaMDS2512016) assays using the QX100/QX200 Droplet Digital PCR (ddPCR) system (Bio-Rad, Hercules, California, USA). The ddPCR reaction mixture consisted of 10 μ l of a 2X ddPCR Supermix for Probes no dUTP (Bio-Rad), 1 μ l of 20X primer/probe mix and 50 ng of sample DNA from unsorted tissue material in a final volume of 20 μ l. Mutant target and wild-type assays are provided together in a single tube, with FAM targeting the mutant allele and HEX targeting the wild-type allele. The entire reaction mixture was loaded into the sample well of a DG8 cartridge (Bio-Rad) together with 70 μ l of droplet generation oil (Bio-Rad) and placed in the droplet generator (Bio-Rad). After processing, the droplets generated from each sample were transferred to a 96-well plate that was heat-sealed with a foil seal and then placed on a conventional thermal cycler. Thermal cycling consisted of a 10 min activation

period at 95°C followed by 40 cycles of a two-step thermal profile of 30 sec at 94°C denaturation and 1 min at 55°C for combined annealing-extension and 1 cycle of 98°C for 10 min. The ramp rate for each step was set to 2°C/sec. Droplets were read using a QX100/QX200 Droplet Reader (Bio-Rad), and the data were analyzed using QuantaSoft software (Bio-Rad). No-template controls were performed using water in place of template in every experiment; in all cases no amplification occurred.

3.1.13 In Situ Hybridization (ISH)

In Situ Hybridization was performed using the dual-color silver/chromogenic *in situ* hybridization (SISH) test for *Her-2* gene and chromosome 17 (centromeric probe). Copy number alterations were identified by INFORM HER2 Dual ISH DNA Probe Cocktail on a Benchmark Ultra instrument (Ventana Medical Systems, Mannheim, Germany). Forty non-overlapping nuclei were counted at 600X microscope magnification.

3.1.14 Immunohistochemistry (IHC)

Immunohistochemistry was performed automatically with Benchmark XT[®] immunostainer (Ventana Medical Systems) for p53 and HER2 antigens. The immunohistochemical analysis was validated through positive controls (as an external positive control put on the slide) and negative control (by omitting the primary antibody). The immunoreactivity for the HER2 protein was scored according to American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines. [102]. The expression of p53 protein was defined as hyper-expressed if there was evidence of strong and diffuse nuclear immunoreactivity [76].

IHC for SMAD4 was performed manually on FFPE blocks with anti-SMAD4 mouse monoclonal antibody (clone B-8; 200 µg/ml, diluted 1/200; Santa Cruz Biotechnology, Dallas, Texas, USA). In brief, deparaffinized sections were immersed in 3% H₂O₂ for 10 min to abolish endogenous peroxidase activity and washed in 0.05 M TBS (pH 7.3) for 15 min. Antigen retrieval was performed in citrate buffer (pH 6.0) in a microwave oven with one cycle at 900 W for 3 min followed by a second cycle at 360 W for 13 min. Sections were then incubated with anti-SMAD4 antibody for 1 hour at room temperature. The slides were subsequently rinsed three times in TBS and incubated for 20 min with Multilink Biotinylated anti-Ig (Biogenex, Fremont, CA, USA) diluted 1:20,

followed by 3 min incubation with 0.05% 3,3'-diaminobenzidine-tetrahydrochloride in 0.02% H₂O₂. Sections were finally counterstained with Mayer's hematoxylin, dehydrated and mounted. The immunohistochemical analysis was validated using both positive controls (non-neoplastic mucosa and lymphoid cells) and negative internal controls (i.e. smooth muscle cells). The expression of the SMAD4 protein was defined by a complete loss of expression in at least 30% of cancer cells, using the same cut-off score identified for colon cancer in a previously published work [103;104].

3.1.15 Statistical analysis

Data were represented as the median and interquartile range (IQR) for continuous variables and as n (%) for categorical variables unless otherwise stated.

The Mann-Whitney test and one-way analysis of variance (ANOVA) were used to analyze continuous variables. *P*-values < 0.05 were considered significant. Data were analyzed using SPSS (version 15.0; SPSS Inc., Chicago, IL, USA).

The statistical significance of the fold-change difference among the primary tumor and metastases was determined using one-way analysis of variance (ANOVA) with a significance level of 5%.

Spearman's rank correlation was used to assess the correlation between quality of starting DNA and performance of WES analysis.

The cancer-specific survival was assessed using the Kaplan-Meier method and Log-Rank test. The comparison between the histological subtypes defined according to Lauren's classification and distribution of *TP53* mutations was performed with the chi-squared test for given probabilities (R software package; R Project for Statistical Computing, Vienna, Austria), Bonferroni's correction was applied and *P*-values < 0.025 were considered significant. Multivariate analysis applying forward stepwise approach was used to assess the joint effect upon survival of more variables, such as age, sex, cancer stage, Lauren classification, presence/absence of intestinal metaplasia and *TP53* mutational status.

Part II

3.2 MicroRNA expression profiles in esophageal adenocarcinoma

3.2.1 Subjects

We evaluated 112 esophageal adenocarcinoma patients, surgically treated between 2005 and 2017 in four European Centers: Maria Cecilia Hospital, Cotignola, Italy (47 cases); IRCCS Policlinico San Martino, Genova, Italy (8 cases); Helsinki University Central Hospital, Helsinki, Finland (10 cases) and The Academic Medical Center Hospital, Amsterdam, Netherlands (47 cases). The inclusion and the exclusion criteria consist of the presence of adenocarcinoma in the esophagogastric junction in absence of neoadjuvant treatment (chemioradiotherapy-naïve EACs). After formalin fixation, surgical resections were processed and paraffin embedded (FFPE). All cases were examined by gastrointestinal pathologist and were classified according to morphology in glandular intestinal types (intestinal-like morphology), glandular non-intestinal types (gastric-like morphology) and signet-ring. Samples were distinguished according to Lauren's and BIM/GIM classifications (detailed clinical features are summarized in Table 1).

As non-neoplastic tissues (normal controls), we selected 8 cases of FFPE healthy gastric mucosa from patients subjected to pancreatic duodenectomy for pancreatic cancer.

The number of cases was increased with a second independent cohort of 61 EAC patients from The Academic Medical Center Hospital (Amsterdam, Netherlands). Patients underwent upper gastrointestinal endoscopy during which three mucosal biopsies were taken next to each other from the tumor area of patients. Biopsies were immediately frozen and stored at -80 °C until RNA was extracted with an optimized protocol for selective isolation of microRNAs. As non-neoplastic tissues (normal controls), we selected 10 cases of healthy esophageal tissues (fresh-frozen material). Histological assessment was performed on biopsies and clinical features are summarized in Table 4. 25 out of 61 patients were excluded from the Kaplan-Meier disease-free survival analysis since they underwent neoadjuvant chemoradiotherapy

following CROSS strategy, including carboplatin, paclitaxel and 41.4Gy of radiotherapy, followed by surgical resection.

The study received approval from the Ethical Committee, as specified in paragraph 3.1.1.

3.2.2 RNA isolation from FFPE surgical specimens

From FFPE blocks of EAC and normal gastric mucosa, two 10 µm thick sections were prepared and specific tumor area was selected by hematoxylin-eosin staining. Total RNA was manually isolated from paraffin-embedded tissue samples using RecoverAll™ Total Nucleic Acid Isolation for FFPE Kit (Thermo Fisher Scientific) and treated with *DNase I* (Thermo Fisher Scientific) following manufacturer's instructions. RNA was quantified using NanoDrop spectrophotometer (Thermo Fisher Scientific) and stored at -80°C.

3.2.3 TaqMan Human MicroRNA Array cards for microRNA expression profiling

MicroRNA expression was firstly profiled in 8 FFPE EAC cases and 8 FFPE healthy gastric mucosa (controls) with the TaqMan MicroRNA Array card A2.1/B3.0 (Thermo Fisher Scientific) for a total of 754 human different miRNAs. Included on each array three TaqMan MicroRNA Assays were used as endogenous controls (U6 snRNA, RNU44 and RNU48) for data normalization.

50 ng of total RNA was reverse transcribed (RT) into cDNA using TaqMan microRNA Reverse Transcription Kit and Megaplex RT primer pools A or B (Thermo Fisher Scientific) in 7.5 µl final volume. A step of preamplification was included as follows: 2.5 µl of the RT reaction was combined with the matching Megaplex PreAmp Primer Pool and TaqMan PreAmp Master Mix (Thermo Fisher Scientific) in a final volume of 25 µl. Preamplification was run using the following cycling conditions: 10 min at 95°C; 2 min at 55°C; 2 min at 72°C; 15 sec at 95°C, 4 min at 60°C for 12 cycles; 99°C for 10 min. The preamplification product was diluted 1:4 in TE 0.1X and 9 µl of each dilution was combined with TaqMan Universal Master Mix, NoAmpErase UNG (2X) (Thermo Fisher Scientific) before loading on the matching TaqMan MicroRNA Array Card. The TaqMan Array Cards were spun, sealed and then run on a 7900 HT Real Time PCR

system (Thermo Fisher Scientific) using universal cycling conditions (10 min at 95°C; 15 sec at 95°C, 1 min at 60°C, 40 cycles).

Raw data files were analyzed using the SDS v2.2 software (Thermo Fisher Scientific), a data analysis tool that can quickly evaluate large sets of data files and allows to perform relative miRNA expression quantification (2- $\Delta\Delta$ CT method).

3.2.4 Quantitative real-time PCR validation of microRNA profiles through single assays

Real-time quantitative PCR analysis of two miRNAs, that were found to be differentially expressed in the microarray experiments on microfluidic cards, was validated using single TaqMan probes for miR-221 and miR-483-3p (Thermo Fisher Scientific). Expression levels of selected miRNAs were firstly evaluated on 112 FFPE cases and then on 61 EAC fresh-frozen tissues.

Single-strand cDNA was synthesized from 150 ng of total RNA extracted from FFPE sections using specific 5X miRNA primers (TaqMan MicroRNA Assay, Thermo Fisher Scientific) and the TaqMan MicroRNA Reverse Transcription Kit (Thermo Fisher Scientific). For fresh-frozen samples, we started from 150 ng of microRNA-enriched samples. A step of preamplification was included as follows: 2.5 μ l of the RT reaction was combined with specific primer pool, containing a 20X mix of forward and reverse primers (final concentration of each assay equals 0.2X), and TaqMan PreAmp Master Mix in a final volume of 25 μ l. Preamplification was run using the same cycling conditions specified in paragraph 3.2.3.

The preamplification product was diluted 1:8 in TE 0.1X then 2 μ l was combined with TaqMan Universal Master Mix II, NoAmpErase UNG (2X) and specific 20X TaqMan MicroRNA Assays for miR-221 and miR-483-3p (#000524, #002339). Quantitative PCR (qPCR) analysis was performed on 7500 Fast Real-Time PCR Systems (Thermo Fisher Scientific) using universal cycling conditions (10 min at 95°C; 15 sec at 95°C, 1 min at 60°C, 40 cycles). PCR reactions for each sample were run in triplicate. Control reactions included cDNA synthesized without reverse transcriptase enzyme (RNA only) and no cDNA template. RNU44 (#001094) was tested as endogenous control for data normalization. Fold change was calculated using 2- $\Delta\Delta$ CT method, comparing FFPE tumor cases versus the pool of 8 selected tissues of healthy gastric mucosa, or

comparing fresh-frozen cases versus a commercial pool of normal esophagus RNAs pooled from 5 different donors (BioChain, Newark, CA, USA).

3.2.5 Immunohistochemistry (IHC)

4 µm sections of FFPE tissue were used for IHC of SMAD4 and PTEN antigens, using the following antibodies: SMAD4 mouse monoclonal antibody (clone B-8, 200 µg/ml, diluted 1/200; Santa Cruz Biotechnology) and rabbit anti-human PTEN antibody (Clone: SP128; Abcam, Cambridge, UK).

PTEN immunohistochemistry was performed using the Ventana automated staining platform (Ventana Medical Systems), instead SMAD4 IHC was performed manually. In brief, deparaffinized sections were immersed in 3% H₂O₂ for 10 min to abolish endogenous peroxidase activity and washed in 0.05 M TBS (pH 7.3) for 15 min. Antigen retrieval was performed in citrate buffer (pH 6.0) in a microwave oven with one cycle at 900 W for 3 min followed by a second cycle at 360 W for 13 min. Sections were then incubated with specific antibody for 1 hour at room temperature. The slides were subsequently rinsed three times in TBS and incubated for a further 20 min with Multilink Biotinylated anti-Ig (Biogenex) diluted 1:20, followed by 3 min incubation with 0.05% 3-3'-diaminobenzidine-tetrahydrochloride in 0.02% H₂O₂. Sections were finally counterstained with Mayer's hematoxylin, dehydrated and mounted. The immunohistochemical analysis was validated using both positive controls (non-neoplastic mucosa and lymphoid cells) and negative internal controls (i.e. smooth muscle cells admixed with tumor).

The expression of the SMAD4 protein was defined by a complete loss of expression in at least 30% of cancer cells, using the same cut-off score identified for colon cancer in a previously published work [103;104].

Instead, a tumor tissue was considered to have PTEN protein loss if the intensity of cytoplasmic and nuclear staining was markedly decreased or entirely negative across >50% of tumor cells compared to surrounding benign glands and/or stroma, which provide internal positive controls. If the tumor showed PTEN protein expressed in >50% of sampled tumor glands, the tumor was scored as PTEN intact.

3.2.6 Cell lines

To perform functional studies of the expression profiles of both miR-483-3p and miR-221, found dysregulated in our patients' cohorts, we used 2 EAC cell lines: OE19 and FLO-1 [105].

Both cell lines were provided by Professor Kausilia K. Krishnadath of Academic Medical Center Hospital, Amsterdam, Netherlands.

The cell line OE19 was established in 1993 from an adenocarcinoma of gastric cardia/esophageal gastric junction of 72 years old male patient. The tumor was identified as pathological stage III (UICC TNM classification) and showed moderate differentiation [105;106]. The FLO-1 cell line was established from a primary distal esophageal adenocarcinoma in a 68 years old Caucasian male in 1991. The tumor was classified as pathological stage III, poorly differentiated [105;107]. The mutational landscape of these cell lines was known and representative of EAC tumors [108].

OE19 cells were cultured in RPMI 1640 (EuroClone, Milan, Italy) supplemented with 10% (v/v) fetal bovine serum, 100 U/mL penicillin, 100 µg/mL streptomycin and 2mM L-glutamine (supplements were purchased from Sigma-Aldrich). Instead, FLO-1 were grown in high-glucose Dulbecco's modified Eagle's medium (DMEM) (EuroClone) supplemented with 10% (v/v) fetal bovine serum, 100 U/mL penicillin, 100 µg/mL streptomycin and 2mM L-glutamine (Sigma-Aldrich). All cells were cultured in a 5% CO₂ incubator at 37°C.

3.2.7 Reverse transcription PCR and real-time quantitative PCR

analysis of *SMAD4*

Total RNA was extracted from FLO-1 and OE19 cells using Invitrogen™ RiboPure™ RNA Purification Kit (Thermo Fisher Scientific) and retrotranscribed with Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). TaqMan gene expression assays (Thermo Fisher Scientific) were performed following manufacturer's instructions. The assays utilized were *SMAD4* (Hs00929647_m1) and *β-actin* (Hs99999903_m1). qPCR analysis was performed on the 7500 Fast Real-Time PCR Systems (Thermo Fisher Scientific) using universal cycling conditions (10 min at 95°C; 15 sec at 95°C, 1 min at 60°C, 40 cycles). Relative gene expression was normalized to *β-actin* by comparative Ct method (2- $\Delta\Delta$ CT).

3.2.8 Wound healing assay

5×10^5 esophageal tumor cell lines (OE19 and FLO-1) were plated onto six-well plates and allowed to form a confluent monolayer. The cell monolayer was scratched in a straight line to make a 'scratch wound' with a 10- μ l tip and the cell debris was removed by washing the cells with phosphate-buffered saline (PBS). Images of the wound closure were captured at 0 and 3 days. Images were analyzed with the TScratch software [109].

3.2.9 Statistical analysis

Expression data from MicroRNA Arrays was analyzed with the statistical software Expression Suite Software v.1.0 (Life Technologies). The software included the student's t-test for sample group comparisons and built Volcano Plot comparing the size of the fold change to the statistical significance (*P*-value).

The ROC method was used to optimize cut-off values for miRNAs classification into a "high expression" and "low expression" groups.

Correlations between miRNA expression, tumor recurrence, cancer specific death, BIM/GIM and Lauren's classification were investigated using Mann-Whitney and Kruskal-Wallis tests, t-Student's test and Kaplan-Meier method, using SPSS (version 15.0; SPSS Inc., Chicago, IL, USA).

Student's t-test using Prism (GraphPad, San Diego, CA, USA) was performed to analyze the wound healing assay data. A *P*-values < 0.05 was considered statistically significant.

4. Results Part I

4.1 Genomic profiles of primary and metastatic esophageal adenocarcinoma identified via digital sorting of pure cell populations: results from a case report

In order to evaluate the importance of tumor cell selection for an unambiguous genetic analysis, we started studying the case of a woman who underwent primary radical resection for a stage 2B HER-2-positive Barrett's type esophageal adenocarcinoma (EAC) (detailed clinical features are summarized in Table 1, Row 1, EAC1). Despite HER-2 targeted therapy, her disease recurred and required repeated metastectomies.

Whole Exome Sequencing (WES) of the primary EAC and first two metachronous lung metastases (M1 and M2, resected in April and December 2015 respectively) revealed a shared heterozygous *TP53* missense mutation (chr17:7577094 G>A), that caused a single amino acid substitution (NP_00537:p.Arg282Trp, rs28934574; Figure 8A). Sanger sequencing confirmed the presence of the variant as heterozygous in the EAC primary tumor and in two metastases, but it was absent in the patient's blood DNA, indicating that is a somatic mutation (Figure 8B).

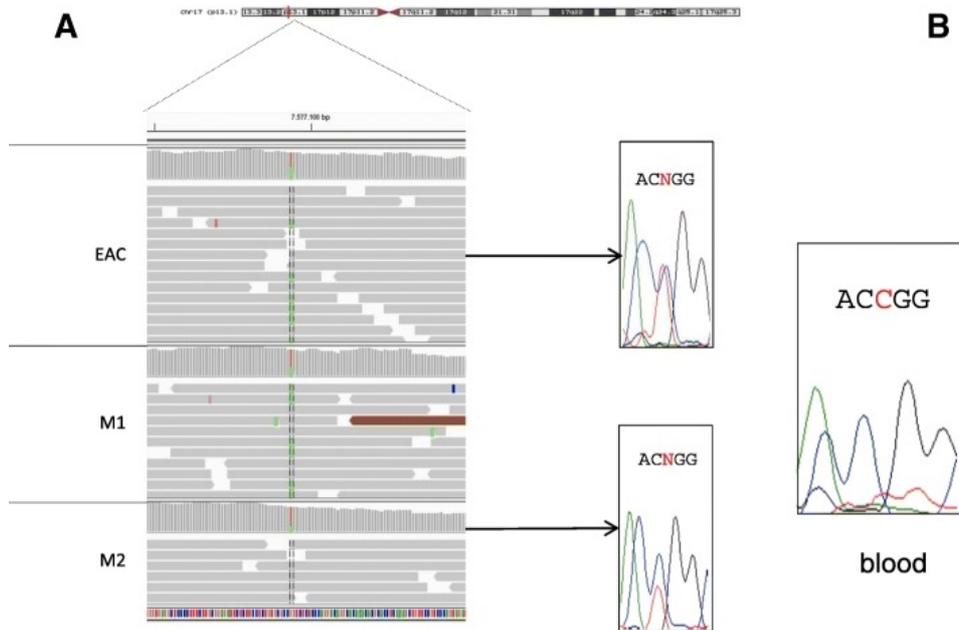


Figure 8: TP53 p.Arg282Trp mutation. A. Representation of the *TP53* mutation (Integrative Genomic Viewer, IGV). *B.* Sanger sequencing of EAC, metastasis and blood.

The *TP53* variant maps in one of the 6 “hotspot” residues frequently mutated in tumor cells and it is present in the COSMIC database (<https://cancer.sanger.ac.uk/cosmic>; COSM10704), where it has been found in different tumor samples (544 entries), including stomach and lung carcinoma. It is present in large control databases such as 1000G, ExAc and GnomAD at an extremely low allele frequency (MAF= 0.00002) and it is predicted to be associated with cancer (score -9.73) by FATHMM (<http://fathmm.biocompute.org.uk>). It is also reported in ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar>) as a pathogenic germline allele (id 27403) present in patients with Li-Fraumeni’s syndrome, characterized by the development of different tumors in the affected individuals (OMIM #15623).

Hematoxylin and eosin (HE)-stained slides from all FFPE blocks of primary tumor and metastases showed homogenous clusters of cancer cells (Figure 9A), which revealed an intense immunohistochemical staining for p53 compared to the normal counterparts, that was consistent for *TP53* missense mutations (Figure 9B).

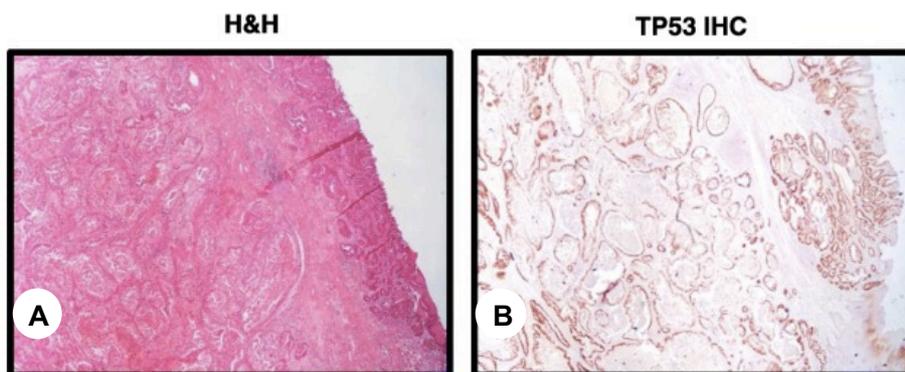


Figure 9: **A.** Histological appearance of the primary EAC (hematoxylin and eosin staining). **B.** p53-immunoreactivity in primary EAC.

In order to dissect the real zygosity status of this variant, cancer cells were separated from stromal cells, using the DEPArrayTM (Menarini Silicon Biosystems) sorting protocol. Using this technology, 9 cancer and 9 stromal populations were isolated in the EAC primary tumor and the two metastases (Table 5). All tumor populations showed a DNA index higher than 1, suggesting a hyperdiploid DNA content and confirming their tumor origin. Targeted sequencing, performed at deep coverage (mean depth ranged between 600X and 5000X) using the OncoSeek Panel (Menarini Silicon Biosystems), revealed that *TP53* was completely mutated in the EAC and metastatic clusters, whereas it was wild-type in the stromal cells (Figure 10 – row 13). This suggested that the *TP53* locus might have been involved in an early loss-of-heterozygosity (LOH) event.

The purity of the sorted samples also detected several LOH events involving the cancer-related genes on chromosome 4 *PDGFRA*, *KIT* (Figure 10 – row 1-6), and on chromosome 7 *CDK6*, *MET* in the primary EAC and in the analyzed metastases (Figure 10 – row 7-12). In these cases, LOH events were characterized by allele frequencies near 50% in stromal populations (germline heterozygous sites), that deviate to 0% or 100% in pure tumor populations.

gene	var	event	EAC primary site									metastasis 1						metastasis 2							
			stromal			tumor			gDNA			stromal			tumor			stromal			tumor				
			L2057	L2065	L2612	L2056	L2066	L2613	L1839	L2058	L2059	L2614	L2618	L2060	L2615	L2619	L2061	L2062	L2616	L2620	L2063	L2617	L2621	L2064	
1	PDGFRA	chr4:55149258-G:A	LOH	25.6	55.4	60.4	0.0	0.0	0.0	34.0	15.3	86.8	81.1	12.0	0.0	0.0	0.0	39.4	54.4	46.8	87.6	0.0	0.0	0.0	79.1
2	PDGFRA	chr4:55152040-C:T	LOH	68.3	41.4	94.6	100.0	100.0	0.0	94.8	90.1	97.8	83.0	81.5	98.2	100.0	100.0	96.4	31.5	86.8	85.0	100.0	99.9	99.9	78.7
3	KIT	chr4:55526615-C:T	LOH	63.6	67.7	63.8	100.0	100.0	100.0	75.9	83.9	53.3	49.0	60.1	100.0	100.0	100.0	80.0	70.2	46.2	67.4	100.0	100.0	97.6	67.7
4	KIT	chr4:55526615-T:G	LOH	11.2	40.4	71.2	100.0	100.0	100.0	67.5	91.3	36.4	61.0	23.2	100.0	100.0	100.0	63.2	46.7	78.5	60.0	100.0	100.0	100.0	70.2
5	KIT	chr4:55579081-G:A	LOH	65.2	70.5	59.5	0.0	0.0	0.0	36.4	66.7	36.1	39.6	86.3	1.4	0.0	0.0	86.0	60.5	36.2	51.5	0.0	0.0	6.6	33.9
6	KIT	chr4:55607973-A:G	LOH	0.0	42.3	100.0	75.9	100.0	100.0	34.0	71.8	99.6	33.1	51.9	100.0	100.0	100.0	51.6	0.0	31.5	15.1	100.0	100.0	100.0	85.7
7	CDK6	chr7:92281213-A:G	LOH	75.9	69.8	14.6	100.0	100.0	100.0	79.1	83.8	75.7	69.7	66.6	100.0	100.0	100.0	69.8	38.7	45.4	71.0	100.0	100.0	99.7	77.6
8	CDK6	chr7:92331373-A:G	LOH	38.3	82.6	58.6	0.0	0.0	0.0	38.7	17.8	34.4	54.2	35.0	0.0	0.0	0.0	45.7	0.0	39.2	35.5	0.0	0.0	0.0	11.8
9	CDK6	chr7:92333408-A:G	LOH	47.8	50.6	32.8	0.0	5.3	0.0	26.7	18.9	64.7	47.0	51.2	0.0	0.0	0.0	16.3	55.2	59.3	50.1	0.0	1.1	0.0	18.1
10	MET	chr7:116336947-C:T	LOH	80.5	34.7	10.4	100.0	100.0	100.0	79.9	64.6	62.6	39.9	49.6	97.6	99.9	100.0	63.8	84.4	63.6	25.9	100.0	100.0	100.0	77.4
11	MET	chr7:116336948-G:A	LOH	20.3	65.0	89.5	0.0	0.0	0.0	20.1	35.4	37.4	39.6	50.2	0.0	0.0	0.0	36.2	15.6	36.2	79.7	0.0	0.0	0.0	22.6
12	MET	chr7:11637094-C:T	LOH	84.6	55.6	21.7	100.0	100.0	100.0	63.9	59.0	63.9	67.1	79.4	100.0	100.0	100.0	65.7	35.7	36.9	48.4	100.0	100.0	99.3	89.4
13	TP53	chr17:7577094-G:A	SOM_HOM	0.0	0.0	0.0	100.0	99.4	90.9	49.9	55.8	0.0	0.0	0.0	96.8	100.0	99.9	51.6	0.0	0.0	0.0	99.9	100.0	98.9	48.6
14	ERBB2	chr17:37858678-A:G	amplification	78.1	53.0	38.8	99.3	99.0	99.4	97.8	98.2	68.7	46.7	62.5	88.1	98.7	98.8	97.1	73.2	53.7	57.9	98.0	97.7	98.1	97.1

Figure 10: Relevant variants identified in the sorted pure populations of tumor (red), stromal (blue) cells and unsorted fractions (violet) from primary EAC and two chest metastases. Table cells in grey highlight positions with very low coverage. Numeric values represent the alternative allele frequency.

To confirm the presence of LOH regions, we performed B-allele frequency (BAF) analysis on WES data. Despite the lower purity of data, the BAF profiles of EAC primary tumor sample were consistent with the identified LOH events, whose coordinates insisted on long regions of polymorphic markers with allelic frequencies that mostly deviate from heterozygous value (Figure 11).

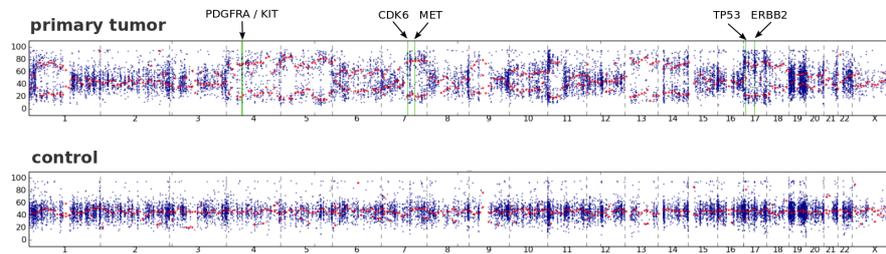


Figure 11: BAF plot obtained using WES data of the primary EAC and a control female individual (WES performed on genomic DNA derived from peripheral blood). In the tumor track, the green lines highlight the positions of genes with putative LOH events detected using the OncoSeek panel. While the control profile shows a flat signal centered around 50%, as expected for a normal germline DNA, EAC tumor profile highlights several consistent regions with abnormal allele frequency, describing putative copy-number altered regions. Given the high variability of allele frequency, due to the relative low coverage in WES, a local smoothing on 20 Mb-long regions, represented by red dots, was calculated specifically to mitigate the frequency variability and to give a sharper idea of copy-number alterations at genome-level.

In addition to *TP53* mutation and LOH events, an *ERBB2* intronic variant were also detected (chr17:37858678 A>G, rs1565923; Figure 10 – row 14), that showed a ~ 100% frequency in both tumor populations and unsorted samples, indicating a high level of

copy-gains. Conversely, *TP53* (Figure 10 – row 13) did not show such a prevalence of the mutated allele in the unsorted fraction (frequency ~ 50%), suggesting that it did not undergo the same amplification event as *ERBB2*. In concordance, DEPAarray™ OncoSeek copy-number analysis of the sorted cell populations revealed a high level of *ERBB2* amplification in all tumor subpopulations, that was completely absent in sorted stromal cells (Figure 12). Interestingly, in the sorted cell populations the *ERBB2* fold-change decreased from ≈ 70 -fold in the primary tumor to approximately ≈ 45 -fold in the two recurrent chest metastases that subsequently developed. Statistical analysis revealed that the fold-changes were significantly different between primary tumor and two metastases (one-way ANOVA, $p < 0.01$ M2/PT, $p < 0.05$ M1/PT; Figure 12).

A further confirmation of *ERBB2* amplification derived from the copy number analysis performed with EXCAVATOR2 software using WES data of unsorted material, even if the absolute value of the amplification in primary EAC and in metastases was much lower due to normal cells contamination (Figure 13, Table 6, chr17:37263657-38948823).

Despite the “diluting” effect due to stromal cells, by performing analysis of copy number alterations (CNA) on the whole primary tumor and metastases, a gain of the chromosomal region 6q21–22.33 (18 Mb) was detected. This alteration in the second chest metastasis generated a focal amplification (39 copies) spanning *RNF146* and *ECHDC1* genes (Figure 14; Table 6). *ECHDC1* copy gains are already present in COSMIC database (COSG94494; <http://cancer.sanger.ac.uk/>) in one case of esophageal cancer.

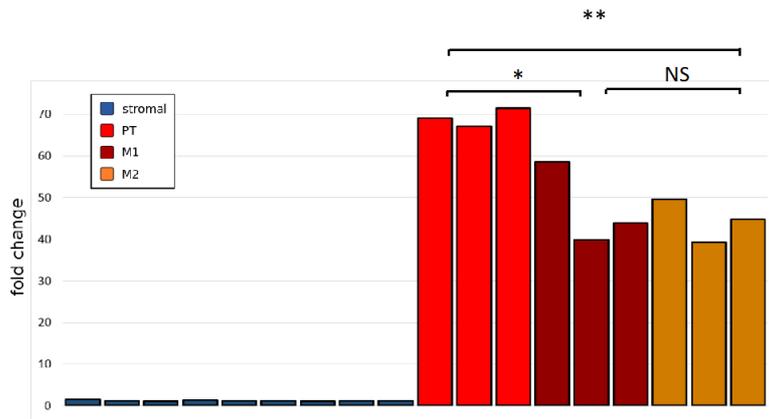


Figure 12: *ERBB2* fold-change in all sorted pure populations (stromal and tumor) of primary EAC (PT) and two metachronous metastases (M1 and M2). Histogram of CNV differences in the primary EAC and metastases. * = $p < 0.05$, ** = $p < 0.001$, NS = not significant (ANOVA test).

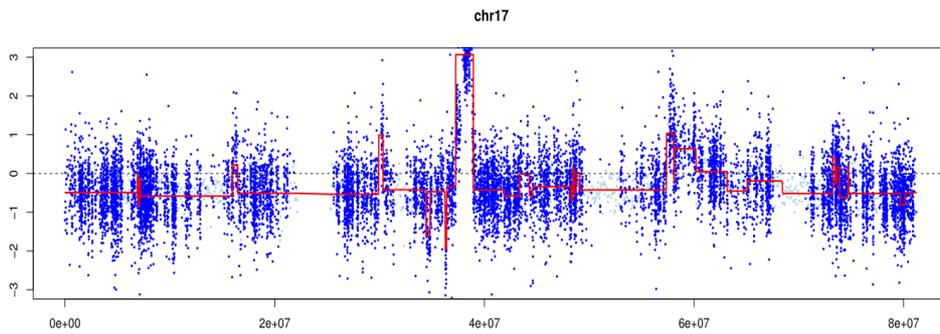


Figure 13: Her-2 Copy-number analysis using EAC WES data on unsorted material.

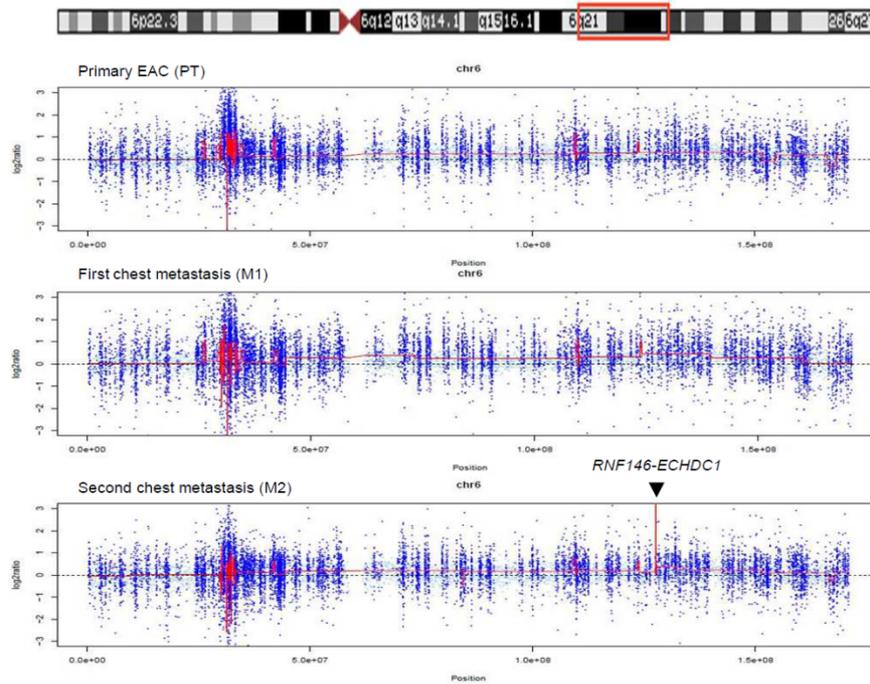


Figure 14: Plots of copy number analysis using WES data. The 18 Mb region on chromosome 6 (q21-22.33) is indicated in the red box, where CNA analysis identified a copy gain in PT (primary tumor) and M1 (first metastasis). In the second chest metastasis (M2) a focal amplification was detected in the 6q22.33 region, spanning *RNF146* and *ECHDC1* genes (black arrowhead).

Using routine diagnostic techniques, no differences were observed in HER-2 immunostaining between the primary lesion and chest metastases (Figure 15A). Silver in situ hybridization (SISH) showed clusters of *ERBB2* amplification, shared by primary and metastatic tumor sites, with a $ERBB2/CEP17 > 2$ and *ERBB2* copy number > 6 (clusters). However, differences between samples could not be appreciated with SISH, since all tissues presented the same amplification patterns (Figure 15B).

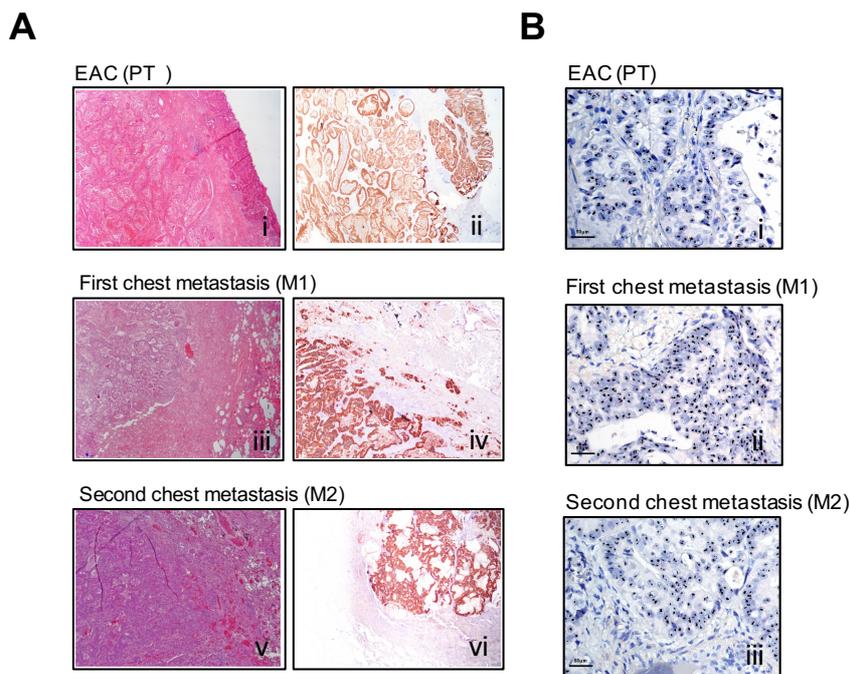


Figure 15: Diagnostic techniques to detect *ERBB2* amplification **A.** (i) Histological appearance and Her2-immunoreactivity in the primary EAC (i-ii), M1 (iii-iv) and M2 (v-vi) metastases. **B.** *Her2* cluster amplification detected by Ventana's Her2 SISH test in primary EAC (i), M1 (ii) and M2 (iii) metastases. Clusters of amplification are represented by black areas in the nuclei.

Whole exome sequencing performed on the third metastasis (M3, resected in July 2018) confirmed the presence of the *TP53* missense mutation (p.Arg282Trp) and the *ERBB2* amplification previously identified, with the same fold change as second metastasis.

In addition, we identified two neighboring variants of new onset, hitting *SMARCA4* gene on chromosome 19 (chr19:11169553 G>T and chr19:11169554 G>T). While the first variant causes a synonymous substitution (NM_001128845 p.L1511L), the second one is a nonsense mutation that introduces a stop codon (p.L1512*). This variant is not present in control databases such as GnomAD but it is already reported in COSMIC database (COSM6972998; <https://cancer.sanger.ac.uk/cosmic>) in squamous cell lung cancer. Other variants in the same gene are reported in esophageal adenocarcinoma and esophageal squamous cell cancer (from TCGA database; <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>).

These *SMARCA4* variants are found on the same allele and form a haplotype (Figure 16) present in 28 out of 194 total reads (alternative allele frequency of 15%, Figure 16),

indicating the possible presence of an advanced tumor subclone, possessing this alteration. This hypothesis was confirmed by Sanger sequencing of bacterial colonies after the TA- cloning step was carried out to separate the two alleles. Thus, we were able to clearly isolate the mutated allele showing the GG> TT substitution (Figure 17), predicted to produce a premature SMARCA4 protein termination, with a polypeptide missing part of its bromodomain.

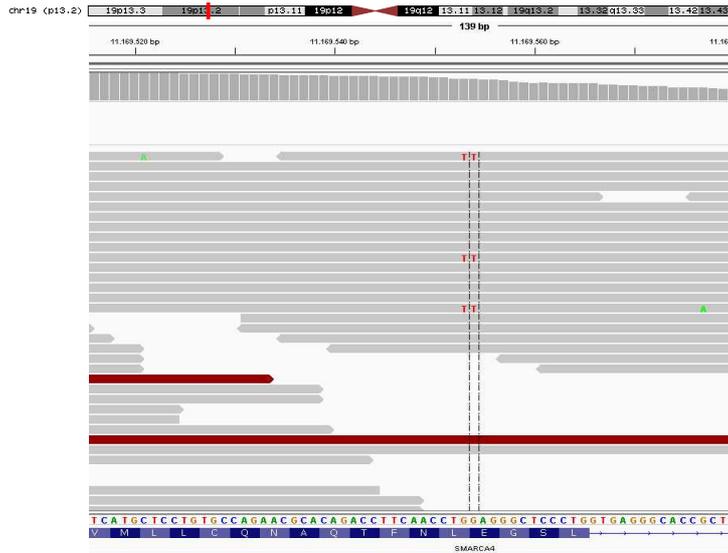


Figure 16: Representation of the *SMARCA4* mutations (Integrative Genomic Viewer, IGV), identified in the third metastasis.

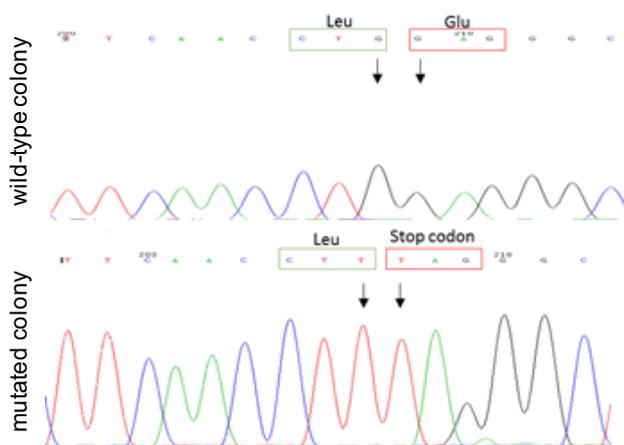


Figure 17: Sanger sequencing of individual recombinant bacterial colonies, where exon 32 of *SMARCA4* gene was cloned. In the upper panel, the electropherogram show the wild-type allele, whereas in the lower panel, the electropherogram show the mutant allele, carrying the GG> TT substitution.

Additionally, SNP-based phylogenetic analysis with SNPhylo, using the WES data, was performed and a phylogenetic tree was constructed to see how similar the three metastases and how differ from primary tumor. Results revealed that all three distant metastases descended from a common metastatic precursor, derived from the primary tumor, and two of them (M1 and M2) clustered together to finally diverge in two branches from a shared precursor clone. The first two chest metastases (M1 and M2) presented the same phylogenetic distance from primary EAC, given by the sum of each branch length between samples. Instead, the third metastasis (M3) seems to have shorter evolutionary distance from primary tumor, in terms of numbers of substitutions events (Figure 18).

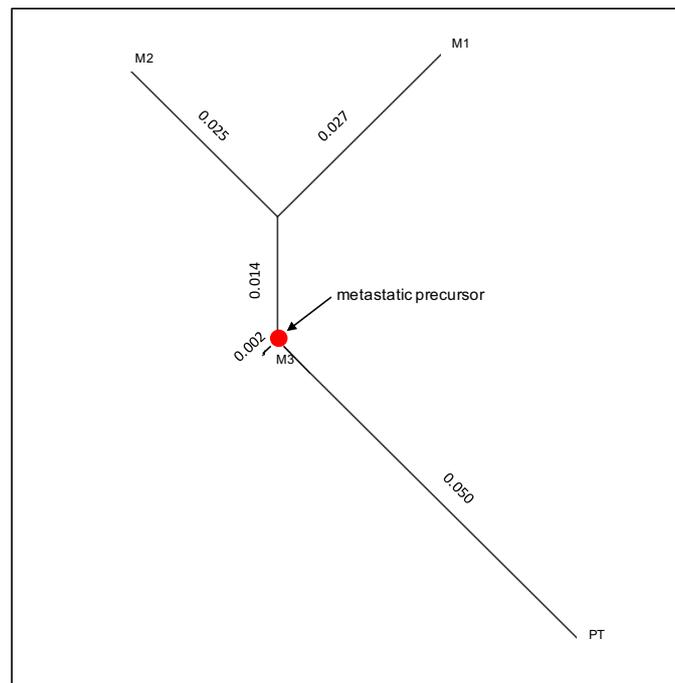


Figure 18: SNPhylo analysis results, showing the genetic distance between the four tumor samples. Numbers indicate the branch length from central node. The distance between two tumors is equal to the sum of their branch length.

4.2 High-throughput sorting of tumor cell populations reveals the composite mutational landscape of esophageal adenocarcinoma

To further improve the genetic characterization of esophageal adenocarcinoma (EAC), pure tumor cell populations were separated through high-throughput cell sorting technology from formalin embedded material (FFPE) of 38 EAC patients, who were treated with surgery alone (Table 1, patients indicated by “*”). The cells obtained from each block were sorted based on immunoreactivity to antibodies against vimentin/cytokeratin and DAPI signal, that is proportional to DNA index (DI), using DEPArray™ machine (Menarini Silicon Biosystems). Thus, we identified at least two cell populations: the stromal (vimentin positive) and the tumor population (cytokeratin positive).

Target sequencing results for the DNA extracted from the whole-tumor samples and sorted cells was performed for 63 cancer-related genes, using the OncoSeek Panel NGS approach (Menarini Silicon Biosystems). Data analysis was performed on a dedicated pipeline, comparing stromal and tumor cell population sequencing data. The same analysis was performed on unsorted cells using the whole tumor specimen. All variants identified were filtered according to presence/absence in public databases (1000 Genomes, gnomAD, COSMIC) and their pathogenic effect.

61 point mutations (missense, nonsense and frameshift) were found across 63 genes by targeted sequencing of the unsorted samples, and 9 additional somatic mutations were detected in the sorted tumor cells (Figure 19A).

The allele frequencies of gene mutations were greater in the sorted cells, where the majority of variants were found in a homozygous state (the number of reads supporting the alternative allele was >80%). Instead, the analysis of unsorted heterogeneous tumor samples revealed an abundance of low-frequency genetic variants (under 20%) that were below the limit of detection of conventional NGS analysis at lower coverage (below 4000X) (Figure 19B). Furthermore, in 5 cases (EAC13, EAC14, EAC19, EAC33 and EAC35), mutations in *HNF1A*, in *PTEN*, in *TP53* and *STK11* were missed

due to a very low percentage of alternative alleles in the analysis of the unsorted material.

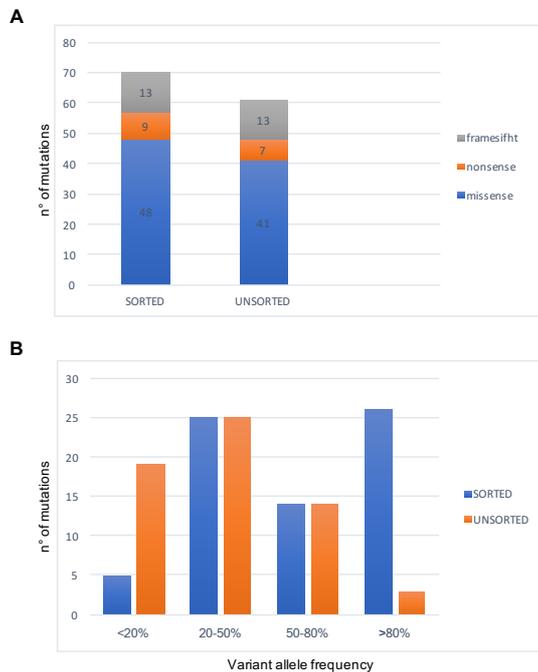


Figure 19: Point mutations detected through sorted cells sequencing and unsorted cells sequencing. **A.** Number of point mutations, distinguishing between missense (blue), nonsense (orange) and frameshift (grey), identified by targeted sequencing in sorted and unsorted cells. **B.** Number of point mutations identified in sorted (blue) and unsorted (orange) cells, clustered in 4 categories according to the number of reads supporting the alternative allele.

4.2.1 Copy number alteration analysis

Although the majority of recovered cancer populations (CK+/VIM-) showed a DI higher than 1, indicative of hyperdiploid DNA content, other cell populations positive to CK immunostaining showed a pseudodiploid DNA content (DI=1), resembling the profile of normal cells. In particular, in 13 EAC cases, we were able to isolate both hyperdiploid and pseudodiploid tumor clones. Low-pass whole-genome sequencing analysis was used to verify whether pseudodiploid cells (isolated in 9 of 13 EACs with different tumor populations) also showed an aberrant genomic profile, as expected for cancer cells. Among these 9 cases, two (EAC10 and EAC18) keratin-positive pseudodiploid cell populations through low-pass sequencing confirmed a normal copy number profile (true-diploid) resembling the corresponding stromal cells, whereas the other 7 cases showed aberrant genomic profiles (Figure 20). In these 7 cases, the sorted tumor populations showed different single-nucleotide mutational loads, and in two cases (EAC19 and EAC4) additional CNAs were also detected in hyperdiploid cells compared to those identified in the corresponding pseudodiploid populations. It is likely that these subclones might have developed during tumor progression (Figures 21A, B).

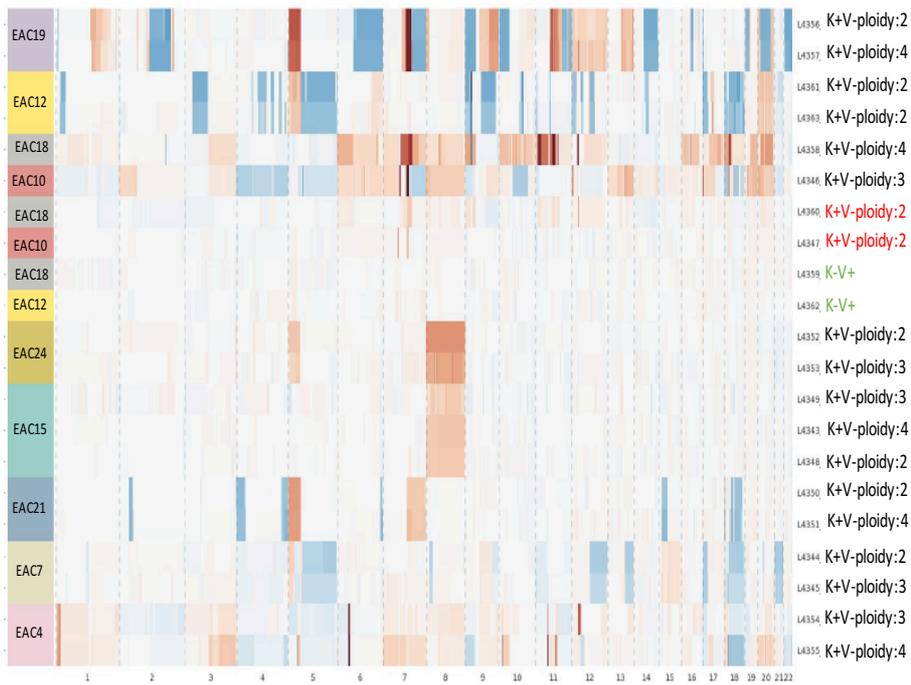


Figure 20: High-throughput image-based cell sorting and analysis of recovered cell populations. Clustering of copy number alteration (CNA) profiles inferred from low-pass whole-genome sequencing for different cell populations for 9 EACs, sorted based on antibodies against vimentin (V)/cytokeratin (CK) and based on the DAPI signal, proportional to cell ploidy. Gains and losses with respect to the estimated main ploidy are shown in red and in blue, respectively.

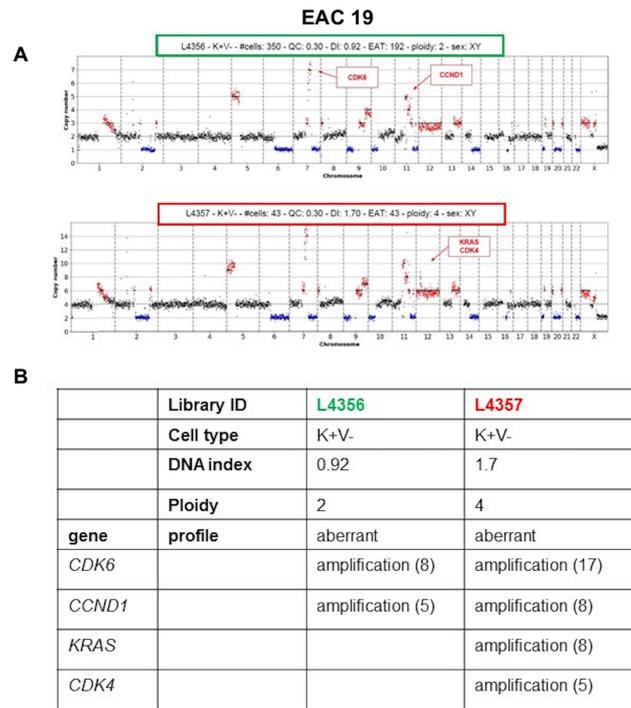


Figure 21: CNA analysis of recovered cell populations from EAC19. **A.** Low-pass whole-genome profile (chr1-22 and chrX) for 2 keratin-positive cell populations (L4356, pseudodiploid cells, green box; L4357 hyperdiploid cells, red box) sorted from sample EAC19. Ploidy values are indicated on the y-axis; on the x-axis, the alteration of different chromosomes is plotted with different colors. CNAs in the tumoral cells are indicated in red (amplification) and blue (deletion). **B.** Principal CNAs identified in pseudodiploid (L4356) and hyperdiploid cell populations (L4357) in EAC19. An approximate copy number value is indicated in brackets.

4.2.2 Identification of mutations in *TP53* and p53-regulated genes

Upon cell sorting, at least one somatic alteration across 63 cancer related genes (point mutation, small insertion/deletion or copy-number alteration), was detected in 35 out of 38 EAC analyzed (Table 7). In 5 EAC cases, only one somatic gene mutation or CNA was detected, whereas the remaining cases presented alterations in multiple genes (Table 7).

In 28/38 (73.7%) of cases a mutation in *TP53* was detected and in 4 cases we observed mutations in *CDKN2A*, a p53-regulator (Table 7). The *TP53* p.R273H hotspot mutation, found at a low percentage in unsorted tumor tissue samples of EAC6, EAC11 and EAC26, and the nonsense mutation in *CDKN2A* (p.R58*), detected in sample EAC4,

were also confirmed by Digital Droplet PCR (ddPCR), which we used as an independent technique to validate results (Figure 22).

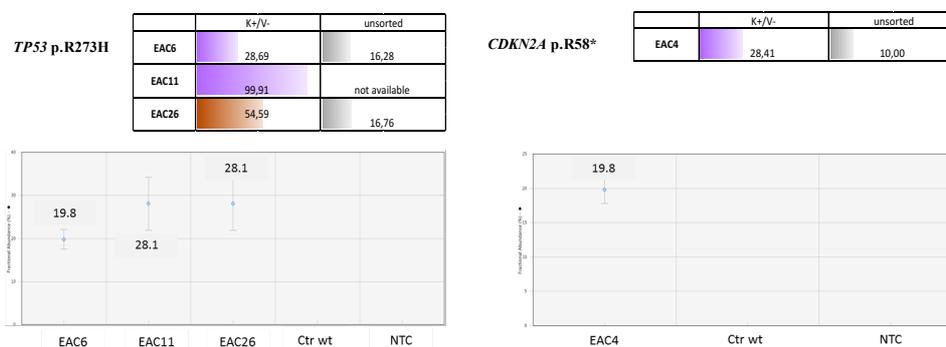


Figure 22: *TP53* and *CDKN2A* mutations identified in different patients with OncoSeek panel and validated by Droplet Digital PCR (ddPCR) specific assays. In the left panel; *TP53* hotspot mutation (p.R273H) identified by the targeted panel in the sorted pure populations of the tumor with hyperdiploid (violet) and pseudodiploid (brown) DNA content and in unsorted fractions (gray) of 3 EAC samples (EAC6, EAC11 and EAC26). Values represent the alternative allele frequency. The graph shows *TP53* mutant allele fractional abundances (%) identified with ddPCR in the DNA from unsorted material of 3 EACs, a known wild-type sample and a no-template control (NTC). In the right panel; *CDKN2A* nonsense mutation (p.R58*) identified by targeted panel in the sorted pure hyperdiploid cell population of tumor (violet) and in unsorted fractions (gray) of sample EAC4. Values represent the alternative allele frequency. The graph shows the *CDKN2A* mutant allele fractional abundance (%) identified with ddPCR in the DNA from unsorted material of EAC4, a known control wild-type sample (CTR wt) and a no-template control (NTC).

A total of 24 different *TP53* mutations, involving distinct aminoacid residues, were detected in our EAC samples. Investigation of the type of these mutations showed that the vast majority were missense mutations (16/24; 66.67%; Figure 23A), two of which are classified as functional in IARC *TP53* Database (<http://p53.iarc.fr/>) based on overall transcriptional activity (TA) (Figure 23B) [110], followed by frameshift (5/24; 20.83%) and nonsense mutations (3/24; 12.5%; Figure 23A; Table 8). The mutations were found to be quite diverse with respect to the location across the coding region of the gene; however, they occurred preferentially in the DNA-binding domain of p53 protein (aminoacids 101-300) (Figure 23C). Whereas single heterozygous *TP53* mutations were present in 35.7% (10/28) of patients with mutant *TP53*, homozygous *TP53* mutations were detected in 64.3% (18/28). Thus, a lower proportion of wild-type *TP53* allele

retention was detected because of more frequent 17p LOH events or second heterozygous TP53 mutations. In 5 cases where we isolated two tumor populations, in the hyperdiploid tumor clones we observed the complete loss of *TP53* wild-type allele and the selective retention/duplication of mutant allele compared to the respective clones with lower DNA content where the mutations were detected in heterozygous state. Thus, we confirmed the presence of multiple tumor clone at different progression stages and high frequency of LOH events at *TP53* locus during esophageal adenocarcinoma evolution.

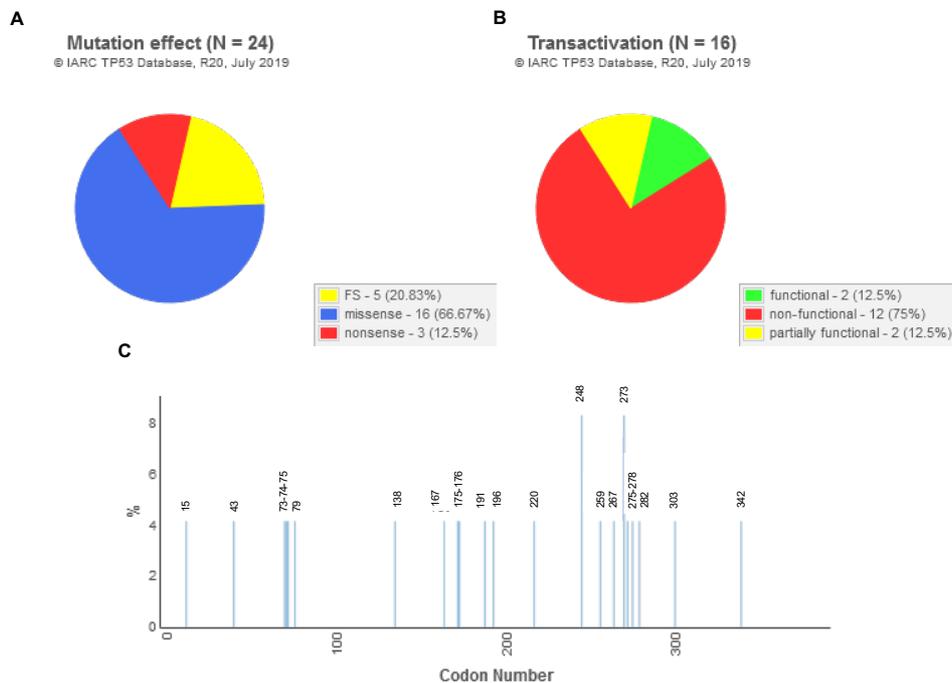


Figure 23: *TP53* mutations distribution graphs. **A.** Mutation effect. Proportion of mutations classified according to their predicted effect on protein sequence (missense, nonsense, frameshift ins/del): number of mutations of each class divided by the total number of mutations detected (% is shown). **B.** Transactivation. Proportion of missense mutations classified according to their experimentally measured transactivation activities (based on IARC TP53 Database): number of mutations of each class divided by the total number of missense mutations detected (% is shown). **C.** Codon distribution. Proportion of exonic point mutations at each codon position: number of mutations at each codon position divided by the total number of exonic mutations detected (% is shown).

In all 38 cases immunohistochemical (IHC) analysis was performed for the p53 protein, which has long been used as a surrogate method for mutations analysis in histopathological diagnostic practice [111]. A significant correlation was observed

between p53 protein overexpression by IHC and the presence of missense mutations, in line with previous data ($P = 0.0002$, Fisher's test; Tables 8 and 9; Figure 24A, B) [112;113]. In 10 EACs, that did not show p53 overexpression, 5 truncating mutations (stop-codon or frameshift changes) and three missense changes classified as nonpathogenic according to IARC *TP53* Database (<http://p53.iarc.fr/>) (Table 8), were identified.

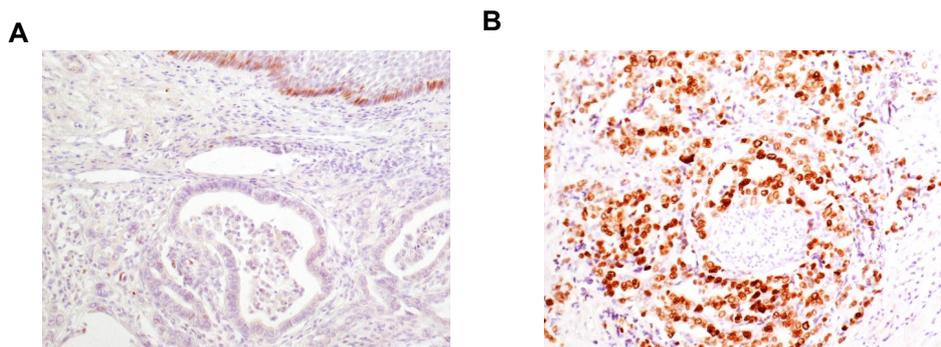


Figure 24: A. Immunohistochemistry indicating low p53 protein expression in a case with a normal p53 gene status (hematoxylin was used as a counterstain). B. Immunohistochemistry for p53 in a case of gene missense mutation and protein overexpression.

4.2.3 Selective sorting identifies high intra-tumor heterogeneity

The presence of high intra-tumor heterogeneity, observed with targeted sequencing and CNA analysis of sorted cell populations, was also supported by further validation of the mutations identified. In particular, in EAC36, the *TP53* missense mutation p.Y220C was found in 59.77% and 99.86% (Figure 25A) of the NGS reads obtained from the two subclones of the sorted tumor cells. We were able to confirm the mutation with Sanger sequencing, in two out of three sections from the same FFPE block. Analysis of the third section revealed a very low variant allele peak, almost below the detection threshold of Sanger sequencing (Figure 25A), due to the presence of different cell types within the cancer area.

Furthermore, a homozygous *TP53* mutation was identified (p.R267G) in patient EAC32 in a homozygous state in the sorted tumor cell population (Figure 25B). Further investigation with Sanger sequencing of two different tissue sections from the same tumor tissue block identified this *TP53* mutation in only the DNA isolated from the

unsorted cells of one section, confirming the high intratumor heterogeneity of these cancers (Figure 25B).

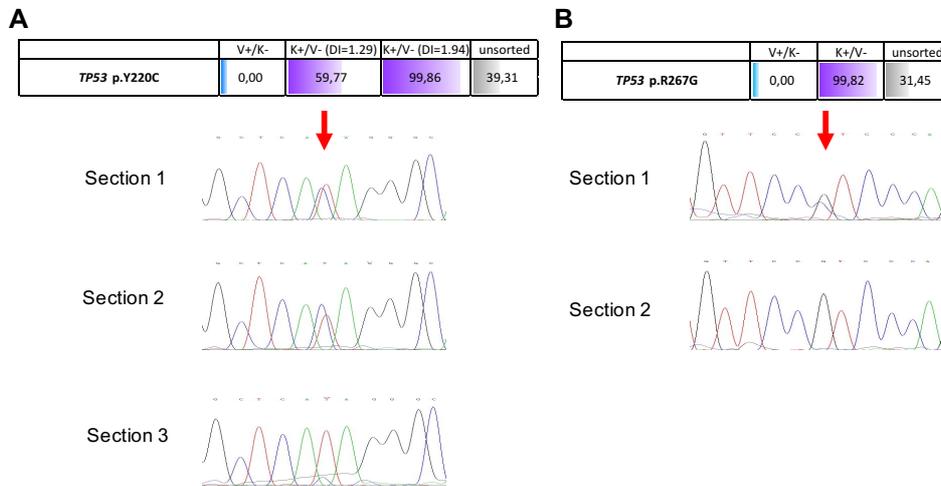


Figure 25: **A.** *TP53* p.Y220C mutation in the sorted pure populations of hyperdiploid tumor cells (violet), stromal cells (blue) and unsorted cell fractions (grey). Values represent the alternative allele frequency (upper panel). Lower panel: Sanger sequencing of DNA isolated from three different tissue sections of the same tumor tissue block, showing the presence of the mutation as a heterozygous change in only sections 1 and 2 (red arrow). **B.** *TP53* p.R267G mutation in the sorted pure populations of hyperdiploid tumor cells (violet), stromal cells (blue) and unsorted cell fractions (grey). Values represent the alternative allele frequency (upper panel). Lower panel: Sanger sequencing of DNA isolated from two different tissue sections of the same tumor tissue block, showing the presence of the mutation as a heterozygous change in only section 1.

4.2.4 Correlation between *TP53* mutations and survival

Since *TP53* is the most frequently mutated gene in EAC, we evaluated whether the presence of *TP53* mutations correlated with clinical outcomes in our cohort of cancers not treated with preoperative chemotherapy (naïve cases). Kaplan-Meier method was fitted to assess the influence of *TP53* on survival. This analysis revealed a correlation between the *TP53* mutational status and a better cancer-specific survival (Log-Rank $P = 0.028$; Figure 26), which, nevertheless, was not statistically significant when we considered the p53 protein expression according to IHC (Log-Rank $P=0.2$).

Therefore, surgery alone without neoadjuvant chemotherapy/chemoradiotherapy resulted in a significant benefit for tumor-associated survival in patients with mutant *TP53* compared with those with a wild-type *TP53*.

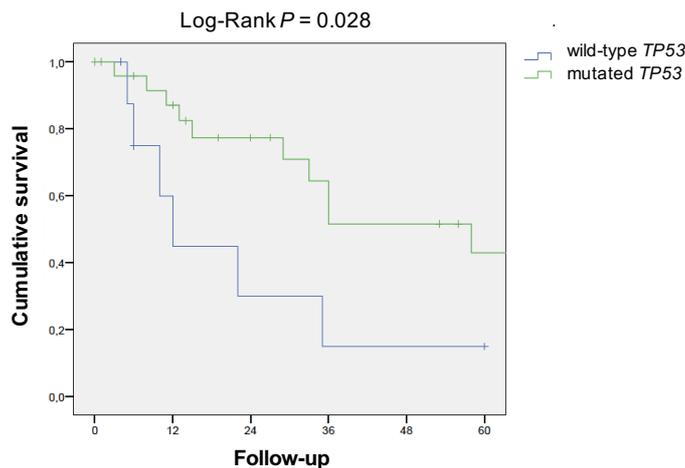


Figure 26: Kaplan-Meier plot of cancer-specific survival, as stratified for patients carrying mutated or wild-type *TP53* (Kaplan-Meier test; log-rank test $P = 0.028$).

Moreover, a difference in the distribution of *TP53* mutations was observed in the histological subtypes defined according to Lauren's classification, i.e. intestinal vs diffuse type. In our sample, 77.4% of intestinal cases, a histological type associated with better outcomes [53;55;54], had *TP53* mutations. Statistically significant differences in the frequency distribution of *TP53* mutations in intestinal type were observed (χ^2 test: $P = 0.002263$), whereas diffuse type cancer did not show this significant increase in *TP53* mutations (χ^2 test: $P = 0.7055$); however, the number of diffuse cases was small (Table 10). The value of *TP53* mutations as prognostic factor was also assessed with a multivariate analysis using forward stepwise method that allowed us to predict the effects on survival of more variables, such as age, sex, *TP53* mutational status, Lauren classification and the presence absence of intestinal metaplasia (BIM/GIM). The multivariate analysis selected the *TP53* mutations, the presence of intestinal metaplasia and the Lauren intestinal cancer subtype as the only significant prognostic factors, that were associated with a longer survival. Among them, the *TP53* mutational status was the strongest predictor of esophageal adenocarcinoma survival for our group of naïve patients. Indeed, the significant negative coefficient ($b=-1.235$, $P<0.0001$; Table 11) indicated that the hazard rate for death was greater in the *TP53* wild-type patients as opposed to the mutant *TP53* group.

4.2.5 *SMAD4* loss is associated with cancer recurrence

Another frequently mutated genes in our EAC cohort was *SMAD4*, occurring in 10.5% of our patients. *SMAD4* is an important tumor suppressor frequently altered in cancers [114]; therefore, its expression was also evaluated by IHC in 34 of the 38 sequenced EAC samples. In four FFPE tissue sections we were unable to perform and analyze IHC results due to poor quality material.

The samples mutated in *SMAD4* (EAC12, EAC17 and EAC23) showed a clear signal reduction (Figures 27A), that was also observed in a substantial number of cases with no mutations in the *SMAD4* gene (18/34, 52.9%) (Table 12).

We then investigated if there was an association between poor clinical outcomes and the loss of SMAD4 expression. For this extent, first the EAC patients were classified according to SMAD4 immunoreactivity, using a cut-off of 30% of SMAD4 loss in cancer area, based on previous works on colon cancer (103;104). Two groups of patients were identified according to this cut-off value: tumors with “low” SMAD4 protein expression (% of loss >30%) and tumors with “high” SMAD4 protein expression (% of loss < 30%). Interestingly, a statistically significant correlation was observed between the patients’ group with loss of SMAD4 expression and higher risk of developing recurrence ($P = 0.015$, Fisher’s exact test; Figure 27B). Instead, no significant correlation was found between the SMAD4 immunoreactivity and specific disease survival (Log-Rank $P=0.383$), probably due to the limited number of cases.

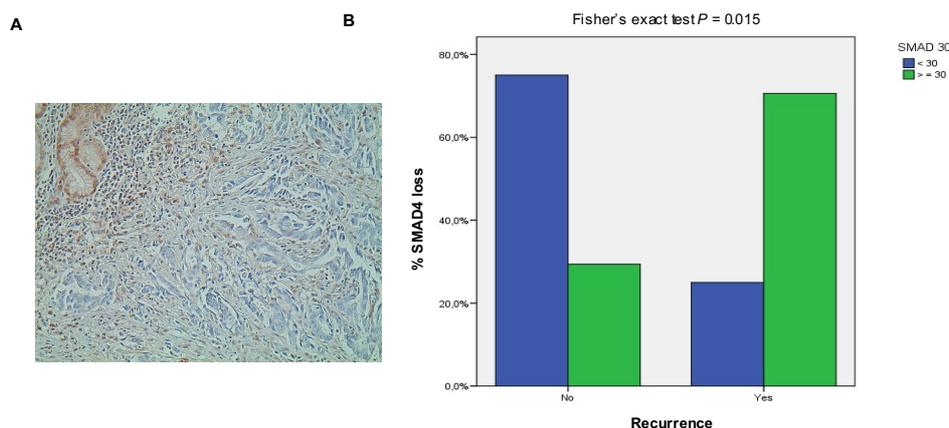


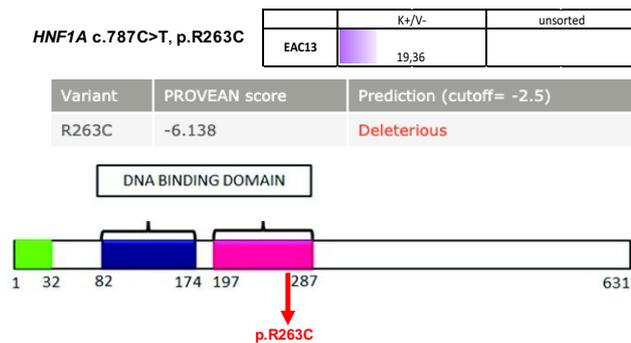
Figure 27: A. Case of SMAD4 loss in tumor cells vs normally expressed SMAD4 in non-neoplastic glands and in stromal cells (upper left corner) (hematoxylin was used as a counterstain; magnification 20X). *B.* EAC grouped according to cancer high SMAD4 (<30% loss of protein expression; green bars)

and low SMAD4 (>30% loss of protein expression; blue bars) and disease recurrence. (Fisher's exact test; $P = 0.015$).

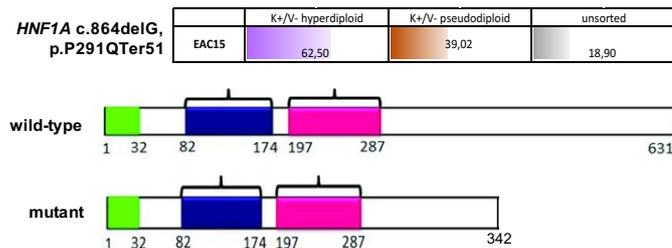
4.2.6 Identification of mutations in *HNF1A*, a novel mutated gene in EAC

In addition to *TP53*, mutations in other genes occurred at lower frequency, including a new gene not previously found mutated in EAC: *HNF1A*, which encodes for a transcription factor that acts as a tumor suppressor in pancreatic cancer [115]. Notably, two *HNF1A* mutations were identified: a missense mutation (p.R263C), occurring in a residue important for DNA binding (Figure 28A), and a deletion (c.864delG), resulting in a frameshift mutation that introduces a premature stop codon (Figure 28B). The *HNF1A* frameshift mutation, identified in the sorted tumor population, was confirmed with Sanger sequencing of DNA from unsorted tumor tissue (Figure 28C). The mutations in this gene were found in conjunction with mutations in other genes: in one case (missense change), this mutation was found in conjunction with somatic mutations of *PIK3CA*, *CDH1*, *SMARCB1*, and in the other case, the frameshift change was found with *TP53*, *EGFR*, *FLT3* and *IDH2* mutations (Table 7).

A



B



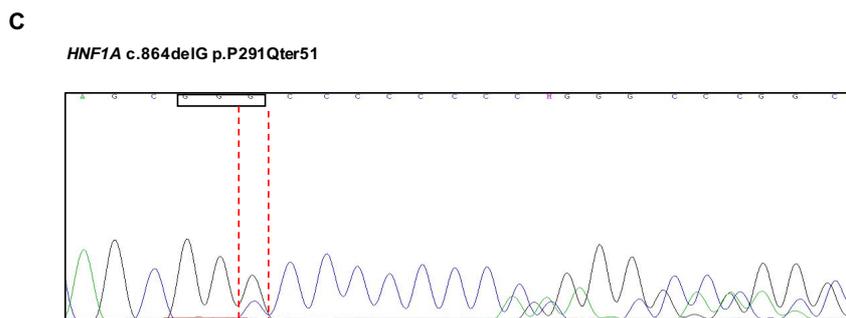


Figure 28: *HNF1A* mutations identified in the sorted populations of tumors with hyperdiploid (violet) and pseudodiploid (brown) DNA content and in unsorted fractions (gray). Values represent the alternative allele frequency. **A.** Frequency of the variant allele *HNF1A* missense mutation p.R263C in the hyperdiploid tumor cell population (violet) (upper panel). The variant pathogenicity was evaluated using PROVEAN. The protein domains are shown in the lower panel. The missense mutation (red arrow) is indicated. **B.** Frequency of the variant allele *HNF1A* deletion (c.864delG) in the different sorted tumor populations (hyperdiploid in violet and pseudodiploid in brown) (upper panel) and prediction of the stop codon inserted by the frameshift mutation into the mutant protein (lower panel). **C.** Sanger sequencing of the *HNF1A* p.P291Qter51 frameshift mutation in DNA isolated from the corresponding FFPE block of EAC15.

4.2.7 Identification of additional mutations through Whole Exome Sequencing (WES)

In order to improve our understanding of the genetic landscape of esophageal adenocarcinoma, WES analysis was performed on the DNA extracted from the whole tumor area of 8 EAC cases; 5 mutated samples at OncoSeek Panel analysis and 3 negative cases. We started from FFPE-derived DNA of low quality, mostly showing a high degree of DNA degradation (QC <0.20). Bioinformatics data analysis of mapped reads revealed an average coverage across exome ranged from 20X to 55X (Figure 29A) and a percentage of the target region covered at least 10X between 80% and 40% (Figure 29B). Using the Spearman rank correlation coefficient test we observed that the integrity of DNA starting material was directly correlated with the average coverage obtained (Figure 30), indicating that the poor quality of DNA extracted from FFPE tissues had significantly impaired quality of whole exome sequencing data.

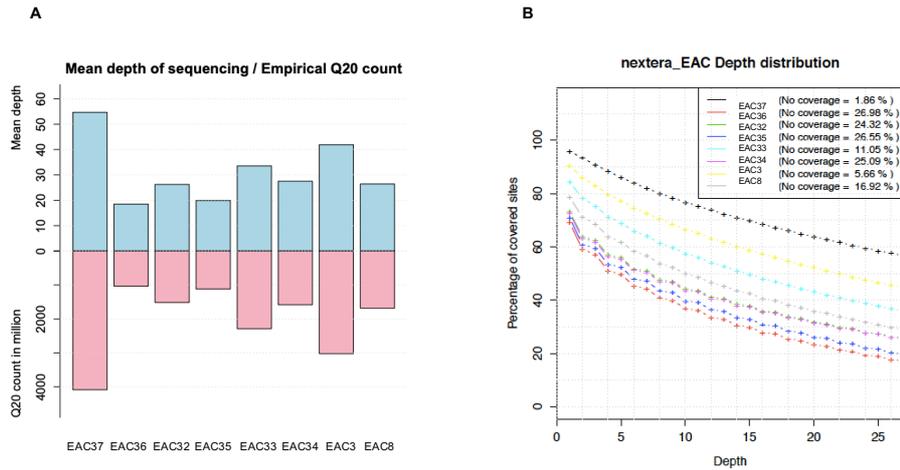


Figure 29: Quality control of WES data. **A.** In the upper panel, representation of mean coverage per sample; in the lower panel, quality of sequencing as assessed by Phred Quality Score (Q20) that represents an error rate of 1 in 100, with a corresponding call accuracy of 99%. Higher Q scores indicate a smaller probability of error. **B.** Sequencing-depth distribution on exome targets per sample.

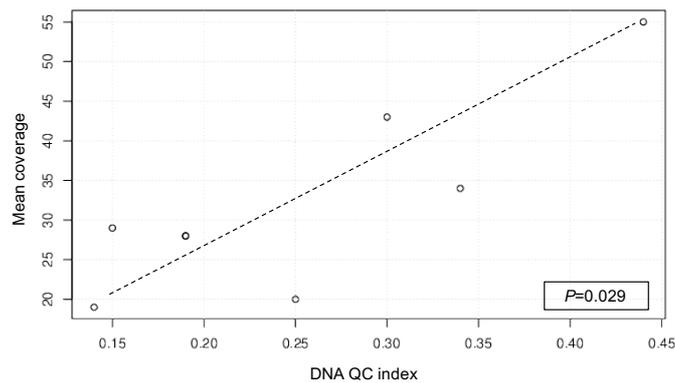


Figure 30: Associations between WES mean coverage for each sample and quality of DNA extracted from the whole tumor area of 8 EAC cases evaluated using Spearman rank correlation coefficient test.

Despite the quality of starting DNA limits the widespread application of WES in clinical practice of cancer patients, in this context we were able to confirm the presence of the mutations detected via targeted sequencing analysis of sorted cell populations, that showed a higher sequencing-depth (4000X) and a better uniformity of on-target coverage, mapping to the coding region of 63 cancer related genes. In particular, using WES mutations were identified in *CDKN2A* (p.L63Q), *APC* (p.R876*) and *TP53* genes (p.Y220C, p.R267G, p.C176F, p.A138V, p.V73RfsTer76) in heterozygous state, already found with target sequencing of pure tumor cell populations of EAC36, EAC32,

EAC35, EAC3 and EAC8, with higher alternative allele frequencies. Nevertheless, WES analysis also detected other damaging heterozygous mutations in genomic regions of oncology-relevant genes not covered by the target panel (Figure 31), such as: *ARID2*, *ATM*, *MSH6* in EAC33, *FLT3*, *ERBB2* in EAC8, and *ALK* in EAC3. These genes are candidate drug target in human cancers, therefore an early detection of such mutations is important in order to personalize a specific therapy.

<i>EAC_ID</i>	<i>Gene</i>	<i>Mutations</i>
EAC37	<i>MYLK</i> <i>TMBIM4</i>	ENST00000475616:p.I512V ENST00000556010:c.464+2T>C
EAC36	<i>TP53</i> <i>INO80D</i>	ENST00000269305:p.Y220C ENST00000424117:p.W241*
EAC32	<i>TP53</i> <i>ERBB4</i> <i>RHBDF2</i>	ENST00000269305:p.R267G ENST00000436443:p.E215K ENST00000591885:p.R320C
EAC35	<i>CDKN2A</i> <i>NOTCH1</i>	ENSP00000304494:p.L63Q ENST00000277541:p.T984M
EAC33	<i>MSH6</i> <i>ARID2</i> <i>ATM</i> <i>CDH1</i>	ENST00000540021:p.C1145F ENST00000457135:p.T148K ENST00000527805:p.R924W ENST00000422392:p.T217K
EAC3	<i>TP53</i> <i>ALK</i> <i>LRRK2</i>	ENST00000269305:p.C176F ENST00000453137:p.G35R ENSP00000298910.7:p.M2521I
EAC8	<i>TP53</i> <i>APC</i> <i>ERBB2</i> <i>FLT3</i> <i>NOTCH1</i> <i>RASAL1</i>	ENST00000269305:p.A138V ENST00000269305:p.V73RfsTer76 ENST00000512211:p.R876* ENST00000584601:p.R648Q ENST00000537084:p.E776K ENST00000277541:p.G2299S ENST00000548055:p.W53*

Figure 31: Results of WES performed in unsorted material of 8 EAC cases. Mutations already found with target sequencing are highlighted in green.

5. Results Part II

5.1 Identification of deregulated microRNAs (miRNAs) in esophageal adenocarcinoma patients through microfluidic cards

miRNA profiles of 8 EAC cases and two pools of 8 normal gastric tissues were analyzed using TaqMan MicroRNA Array card A2.1/B3 (Thermo Fisher Scientific). Different patterns of expression were observed, ranging from low to high levels compared to controls. Using Expression Suite, miRNAs with a fold change ($2^{-\Delta\Delta CT}$) greater or lower than 2 and a p-value

< 0.05 were considered to be differentially expressed.

In particular, we observed 26 miRNAs significantly up-regulated and 72 miRNAs significantly down-regulated (Figure 32; Table 13).

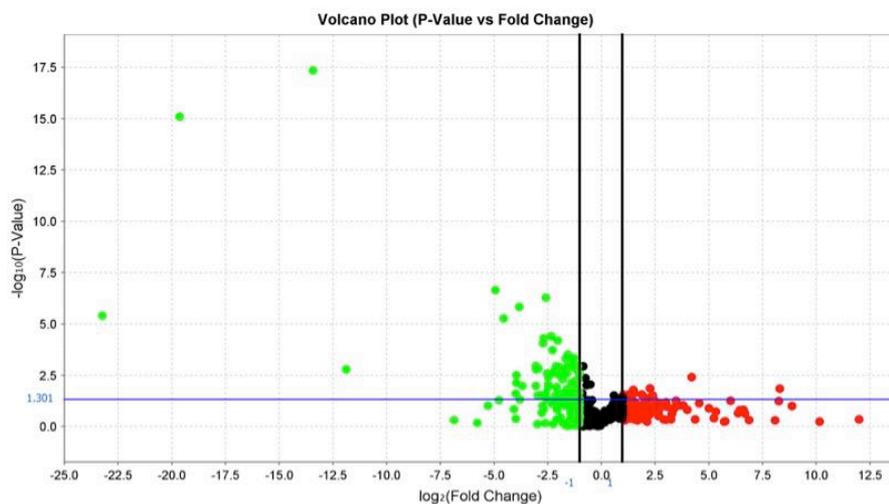


Figure 32: miRNA expression profile in EAC patients versus healthy controls. Volcano plot data from TaqMan microfluidic cards. Each dot on the plot is one miRNA. In the top left: miRNAs (green dots) with a significant negative change of expression (down-regulated). In the top right: miRNAs (red dots) with a significant positive change in expression.

Among them, two miRNAs, based on public data available (miRBase: <http://www.mirbase.org/>), resulted of particular interest: miR-221 and miR-483-3p (Figure 33).

Both miR-221 and miR-483-3p resulted significantly up-regulated in our EAC cases, showing a mean fold increase of 2.746 and 11.33, respectively.

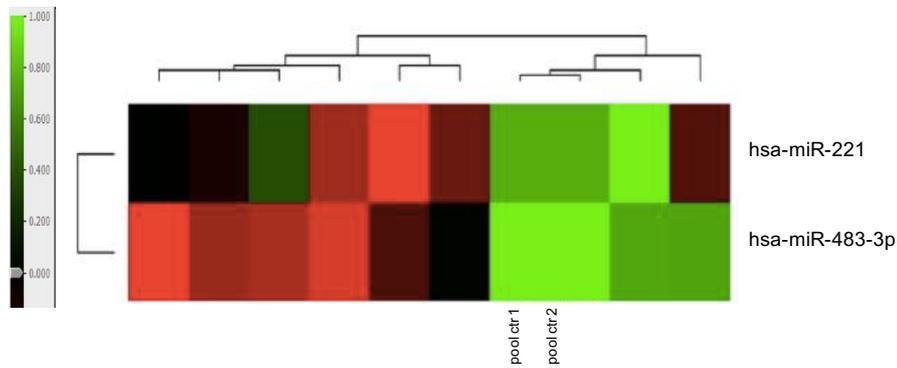


Figure 33: Heat map showing the differential expression of miR-221 and miR-483-3p in EAC cases and controls. The color key indicates the expression levels from low (green) to high (red).

5.2 miRNA 221 and 483-3p expression analysis via single assays and correlation with clinical outcomes

5.2.1 miRNAs expression data from FFPE samples

The miRNA 221 and miRNA 483-3p expression levels were validated in a cohort of 112 formalin-embedded (FFPE) surgical specimens of EAC patients, treated with surgery alone in different European Centers, as mentioned in paragraph 3.2.1, and classified according to Lauren and presence/absence of intestinal metaplasia in esophagus and in stomach. They were analyzed via single Real-time PCR assays using RNU44 as endogenous control. Fold changes were obtained comparing EAC cases versus a pool of 8 RNAs isolated from formalin-embedded healthy gastric mucosa.

The quantitative analysis showed that the expression patterns of both miR-221 and miR-483-3p were consistent with our preliminary miRNA array data. Indeed, these miRNAs were increased significantly in EAC cases compared to normal gastric tissues (miR-221 mean fold increase 2.276, Wilcoxon Signed Rank test: $P < 0.0001$; miR-483-3p mean

fold increase 5.964 $P < 0.0001$; Figure 34), suggesting that they may be novel factors associated with the development of esophageal cancer. Among them, miR-483-3p exhibited the highest up-regulation fold change (5.964) in both microarray and single assays in EAC tissues as compared to the non-cancerous tissues.

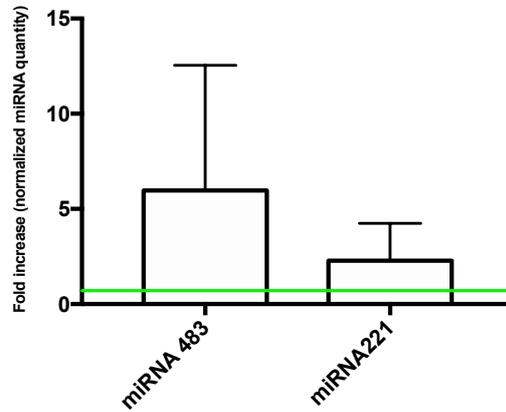


Figure 34: miR-483-3p and miR-221 expression levels in EAC cases respect to the control pool. The values are expressed as fold increase (2-DDCt) with respect to control tissues (FFPE healthy gastric tissues), corresponding to the green base line (Wilcoxon Signed Rank test, $P < 0.05$).

5.2.1.1 miRNA 483-3p expression and correlation with clinical features

In order to explore the role of miR-483-3p in esophageal adenocarcinoma, statistical analyses were performed to investigate possible correlations between miRNA expression, tumor recurrence, cancer specific death, Lauren classification and the presence/absence of intestinal metaplasia in esophagus (BIM) and in stomach (GIM). First, using ROC (Receiver Operating Characteristic) curve and Youden index, we defined the optimal cut-off value, related to survival (cut-off 3.15), that distinguished the patients into a high miR-483-3p expression group and a low miR-483-3p expression group. The Kaplan-Meier curve of the two groups showed a reduced disease-specific survival in patients with up-regulation of miR-483-3p and EAC cancer (Long-Rank $P = 0.0293$; Figure 35A), in particular in Lauren intestinal subtype (Log-Rank $P = 0.012$; Figure 35B). Moreover, an increase in expression of miR-483-3p showed a trend toward association with recurrence (Mann-Whitney test $P = 0.05$; Figure 36). Significant difference in expression was also detected for miR-483-3p among BIM/GIM cancer subtypes (overall Kruskal-Wallis analysis $P < 0.0001$; Figure 37), with a specific decrease in expression in tumors with Barrett's intestinal metaplasia (BIM+/GIM-)

compared to tumors without intestinal metaplasia (BIM-GIM-) (Mann-Whitney test $P=0.0002$; Figure 37).

The results of Kruskal-Wallis analysis also revealed that the expression of miR-483-3p was quite different among TNM stages, with advanced TNM tumors (stages II and III) showing an upregulated miR-483-3p expression (Mann-Whitney test $P=0.053$ (stages I-II); $P=0.019$ (stages I-III); Figure 38A,B).

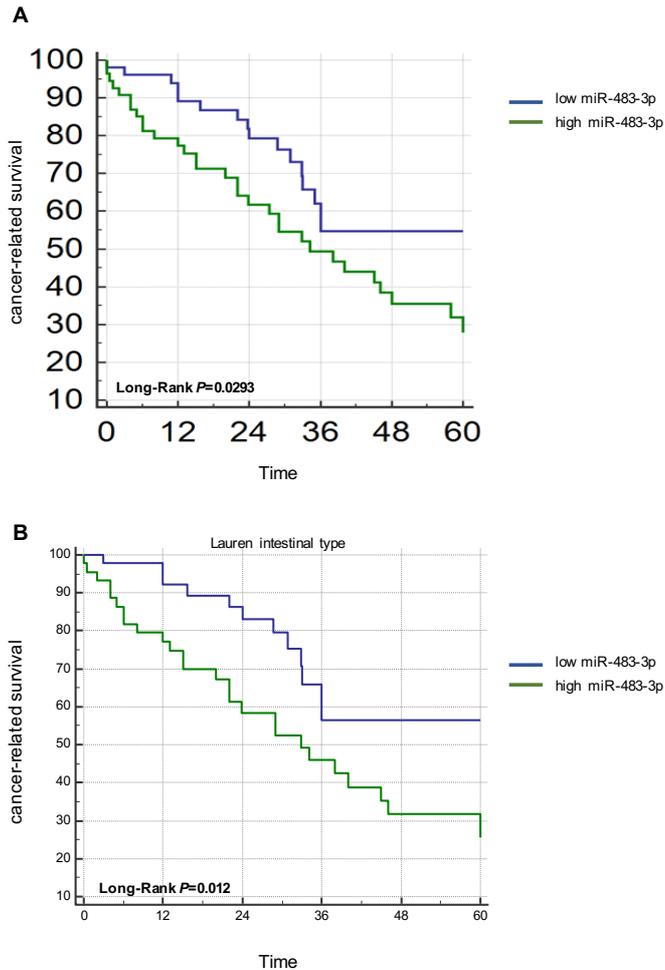


Figure 35: Kaplan-Meier curves depict cancer-related survival for groups stratified base on the expression levels of miR-483-3p, in all EACs (A) and only in Lauren intestinal subtype (B).

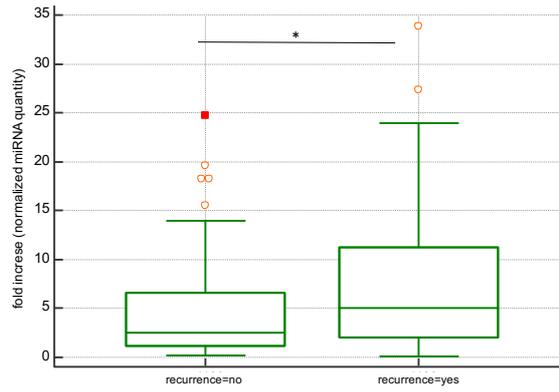


Figure 36: Values of correlations between miR-483-3p and recurrence, using Mann-Whitney test. P -value was considered weakly significant ($P=0.05$; indicated by “*”).

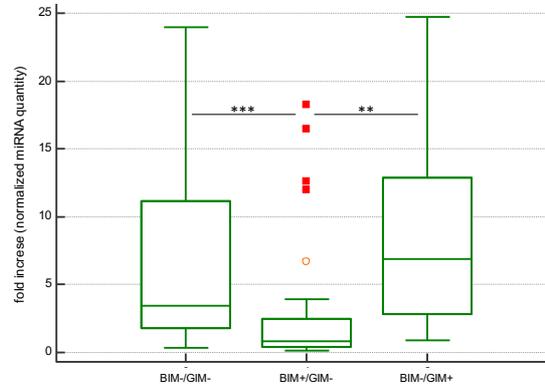


Figure 37: Values of correlations between miR-483-3p and BIM/GIM classification, using Mann-Whitney test. Significant P -values were indicated by “***” $P=0.0034$ and “**” $P=0.0002$.

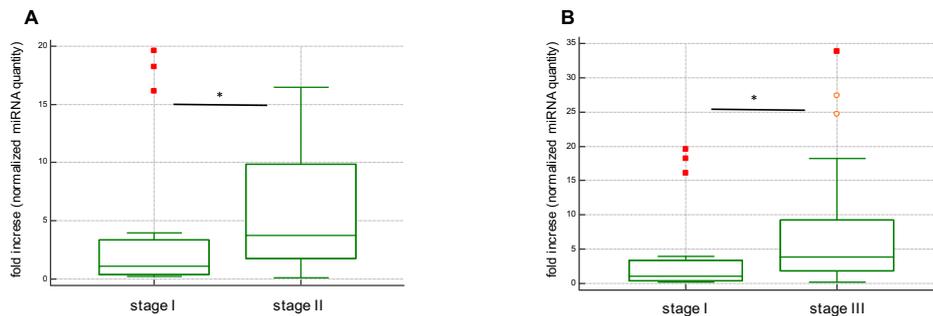


Figure 38: Values of correlations between miR-483-3p and TNM stages, using Mann-Whitney test. **A.** Differences between stage I and stage II tumors: $P=0.053$ “*”. **B.** Differences between stage I and stage III tumors: $P=0.019$ “*”.

5.2.1.2 miRNA 221 expression and correlation with clinical outcomes

To test the predictive value for prognosis, the miR-221 signature was correlated with specific clinical features. A ROC (Receiver Operating Characteristic) curve analysis was used to identify the optimal cut-off value of fold-change (1.32) that distinguished the patients into a high miRNA expression group and a low expression group. Subsequently, differences in disease-specific survival were evaluated in patients classified into these two groups and patients with high miR-221 expression have shown significantly shorter survival compared to patients with low expression (Log-Rank $P=0.0059$; Figure 39A). Kaplan-Meier analysis also revealed a reduced disease-specific survival in up-regulated cases showing Lauren intestinal type of EAC (Log-Rank $P=0.0024$; Figure 39B).

Median expression level of miR-221 was significantly higher in relapsed compare to non-relapsed patients (Mann-Whitney test $P=0.0003$; Figure 40A), and in BIM-/GIM- compare to BIM+/GIM- tumors (Mann-Whitney test $P=0.0247$; Figure 40B). Moreover, we compared miR-221 expression patterns at different stages of esophageal adenocarcinoma and we found that patients with advanced tumor stages (stage III and IV) had significantly higher median expression levels of miR-221 (Mann-Whitey test $P=0.027$ stage I vs stage III, $P=0.0146$ stage I vs stage IV; Figure 41A,B).

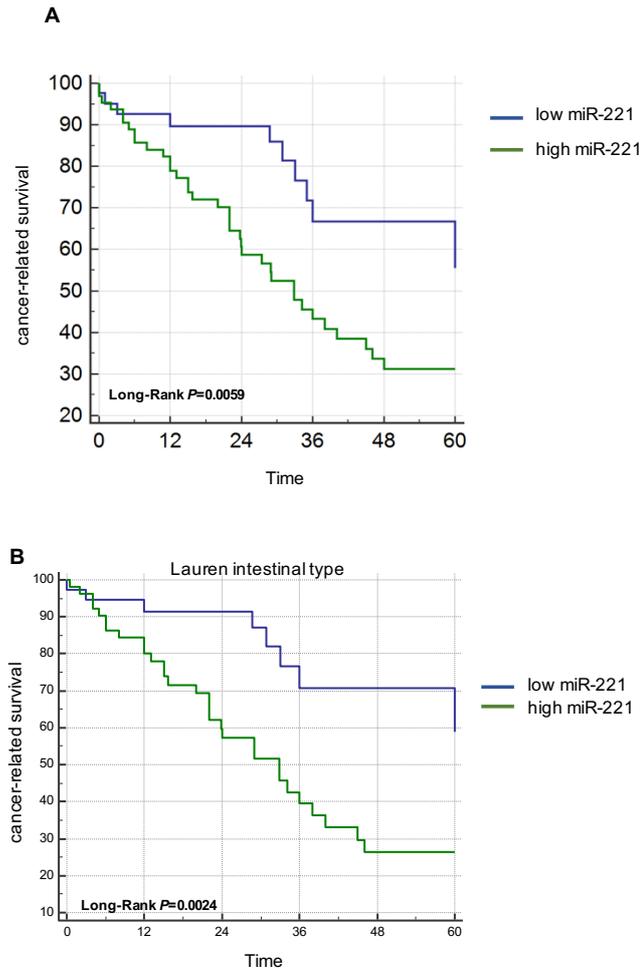


Figure 39: Kaplan-Meier curves depict cancer-related survival for groups stratified base on the expression levels of miR-221, in all EACs (A) and only in Lauren intestinal subtype (B).

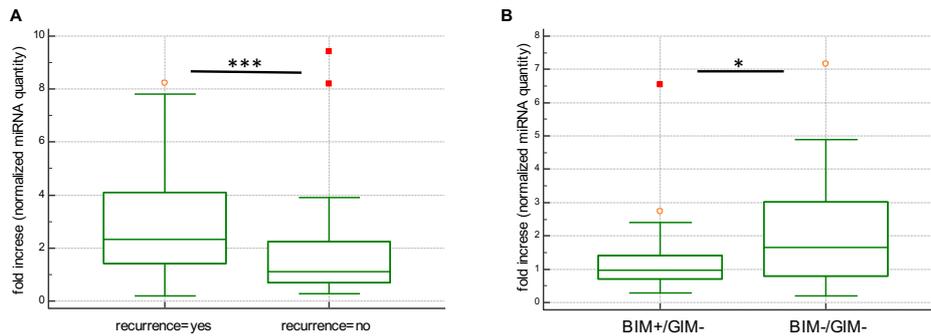


Figure 40: Values of correlations between miR-221 and recurrence (A) and BIM/GIM classification (B), using Mann-Whitney test. P-values were considered significant ($P=0.0003$, indicated by “***”; $P=0.0247$, indicated by “**”).

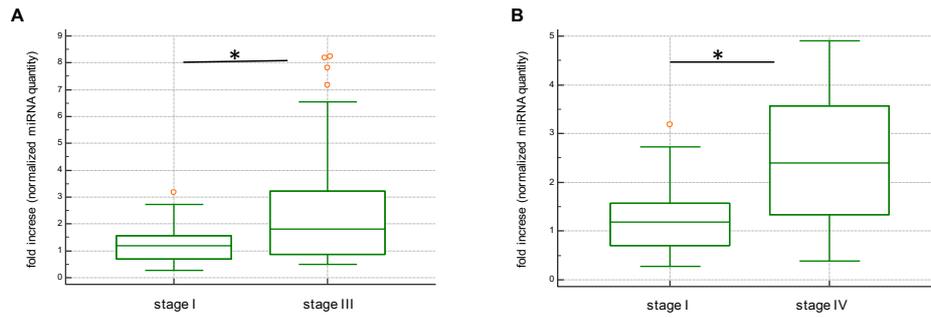


Figure 41: Values of correlations between miR-221 and TNM stages, using Mann-Whitney test. **A.** Differences between stage I and stage III tumors: $P=0.027$ “*”. **B.** Differences between stage I and stage IV tumors: $P=0.0146$ “*”.

5.2.2 miRNAs expression data from fresh-frozen samples

We subsequently evaluated the miRNAs expression patterns in an independent cohort of 61 fresh-frozen EAC biopsies and 10 normal esophageal tissue from The Academic Medical Center Hospital in Amsterdam (detailed clinical features are summarized in paragraph 3.2.1). The expression analysis was performed using single TaqMan probes of miR-221 and miR-483-3p, and RNU44 as endogenous control. The increase in expression was confirmed for both miRNAs (Figure 42A,B). Notably, in EAC cases we identified a mean fold increase in miR-483-3p and in miR-221 of 2.897 and 14.08 respectively, which were statistically different in the tumors compared to the control tissues (Mann-Whitney test: $P<0.0001$; $P=0.0048$).

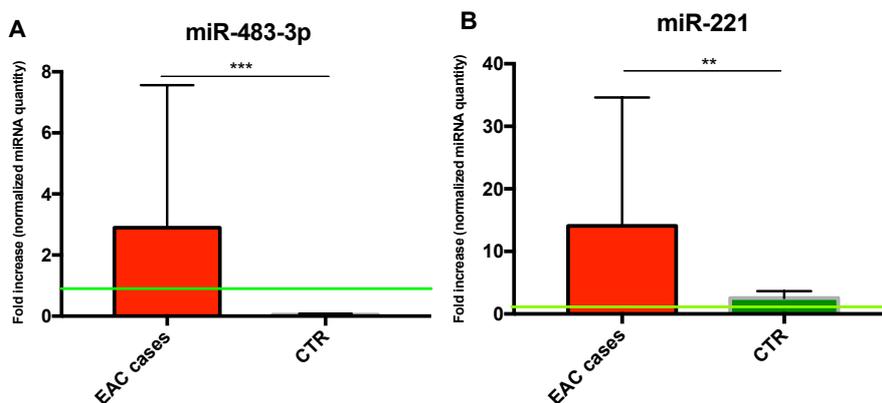


Figure 42: miR-483-3p and miR-221 expression levels in fresh-frozen tumor samples and fresh-frozen control tissues cases. The values are expressed as fold increase (2-DDCt) with respect to a commercial control RNAs pool (BioChain), corresponding to the green base line. Statistical comparisons between cases and controls was performed (Mann-Whitney test: $P<0.0001$ (***) ; $P=0.0048$ (**)).

We tested if the differences in the expression of these two miRNAs could affect the patients' cancer-related survival also in this EAC cohort, in order to confirm our previous data (paragraphs 5.2.1.1 and 5.2.1.2). Kaplan-Meier analysis was performed in 36 out of 61 EAC patients, that were not treated with chemoradiotherapy following CROSS strategy.

Using the ROC (Receiver Operating Characteristic) curve approach to evaluate Youden index and associated criterion as a cut-off (miR-483-3p: cut-off ≤ 0.76 ; miR-221: cut-off ≤ 3.88), the patients were divided into a "high expression" group and "low expression" group. In contrast with data obtained in FFPE cohort, no association between the expression levels of both miRNAs and poor cancer specific survival was detected (Figure 43A,B: Log-Rank: $P=0.116$; Log-Rank: $P=0.132$).

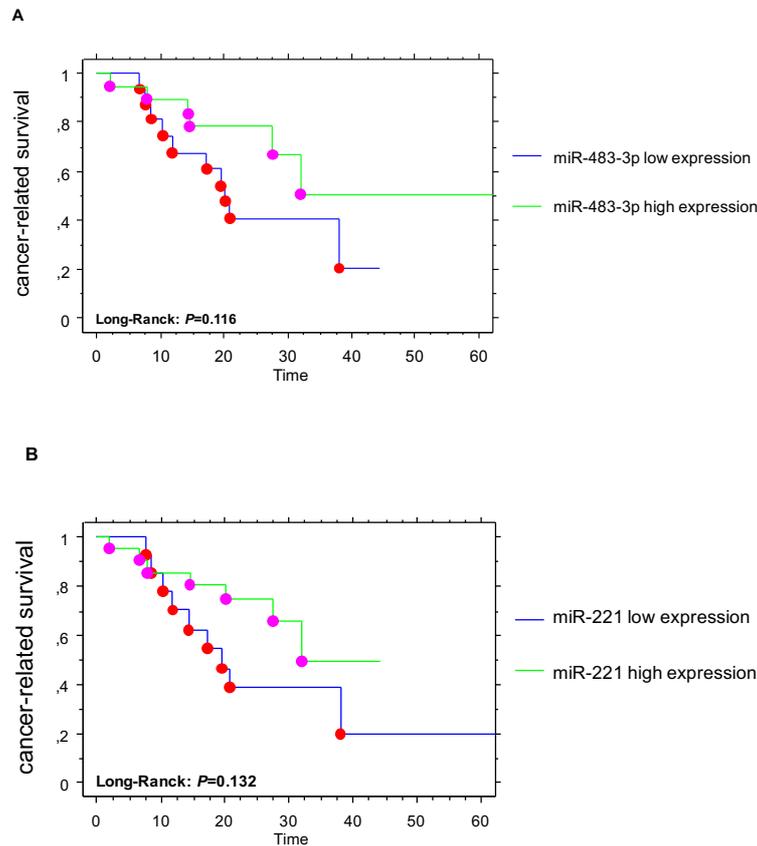


Figure 43: Kaplan-Meier curves depict cancer-related survival for groups stratified base on the expression levels of miR-483-3p (A) and miR-221 (B), as detected by Real Time PCR. We selected patients who underwent surgical resection only. P -value are based on Log-Rank test.

5.2.3 Expression analysis of *SMAD4*: a validated target of miR-483-3p

To explore the mechanisms by which miR-483-3p executes its function in esophageal adenocarcinoma, we used public data available in miRBase (<http://www.mirbase.org/>) to identify potential target genes. Among these many candidates, *SMAD4* is negatively modulated by miR-483-3p, therefore it was selected for further analysis. *SMAD4* is a tumor suppressor gene and a critical effector of *TGF β* -pathway, that regulate cell proliferation [113], and approximately in 52.9% of EAC cases it was involved in gene mutations or loss of protein expression, as mentioned above in paragraph 4.2.5.

Therefore, we focused on miR-483-3p, hypothesizing that it could directly or indirectly target *SMAD4*, leading to its reduced expression, as previously described in pancreatic cancer [116]. To determine whether a similar effect may also be presented in esophageal adenocarcinoma, the expression of SMAD4 was evaluated by immunohistochemistry in 55 formalin-embedded (FFPE) EAC surgical specimens, which were already tested for miR-483-3p expression. Loss of SMAD4 protein was defined by complete loss of expression in least 30% of cancer cells. Loss of SMAD4 immunoreactivity (IHC) was found in 28 out of 55 EAC cases (50.9%) (Table 14), however, except for few cases, we did not observe a clear relationship between miR-483-3p up-regulation and SMAD4 loss (Table 14).

Since loss of SMAD4 expression was found in several EAC cases, we examined its effects on clinical outcomes, i.e. cancer specific survival and tumor recurrence. Kaplan-Meier analysis showed that loss of SMAD4 immunoreactivity was strongly associated with poor survival (Log-Rank: $P=0.013$; Figure 44A) and recurrence (Log-Rank: $P=0.001$; Figure 44B), suggesting SMAD4 expression as potential prognostic biomarker in esophageal adenocarcinoma.

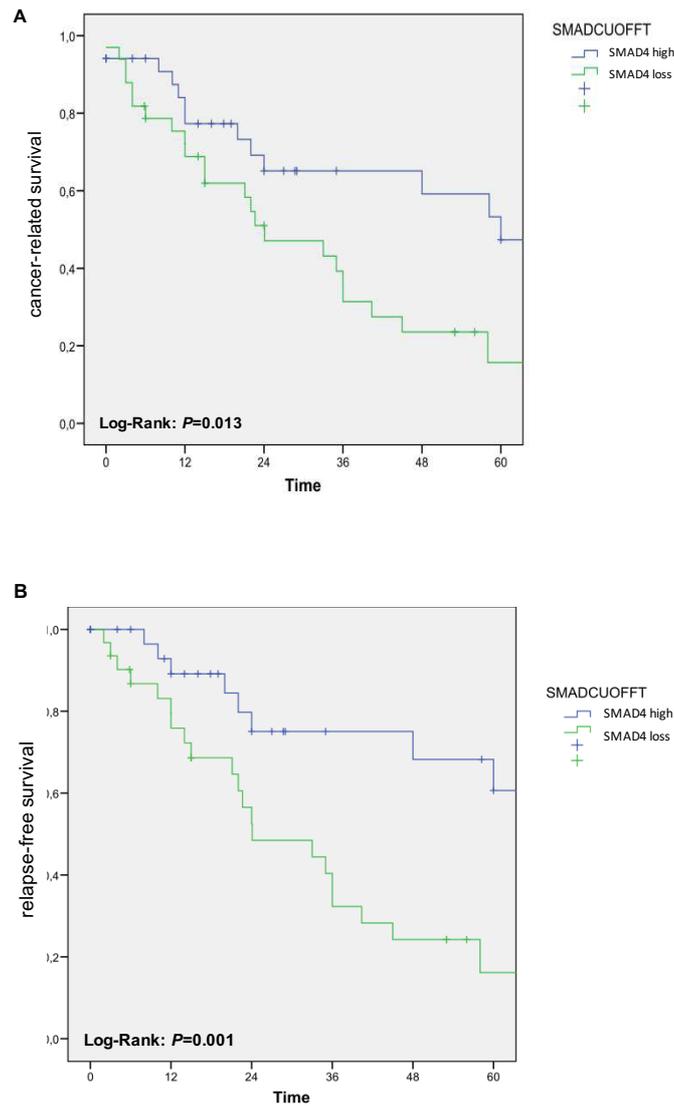


Figure 44: Kaplan-Meier curves depict cancer-related survival (A) and relapse-free survival (B) for groups stratified based on the expression levels of SMAD4 protein. *P*-value are based on Log-Rank test.

5.2.4 Expression analysis of *PTEN*: a validated target of miR-221

To determine whether also miR-221 plays a role in esophageal cancer development and progression, we focused on *PTEN* gene, one of its validated target (<http://www.mirbase.org/>). Notably, *PTEN* expression was inversely proportional to miR-221 overexpression in gastric carcinoma, according to published data [117].

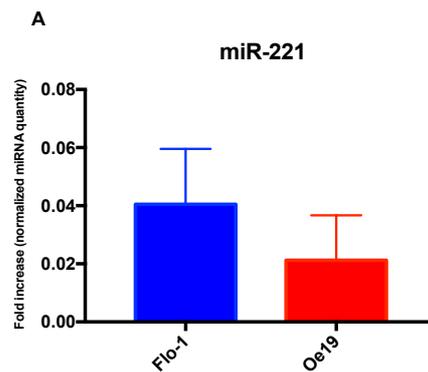
Therefore, *PTEN* protein expression was evaluated by immunohistochemistry in 47 formalin-embedded (FFPE) EAC surgical specimens, which were already tested for miR-221 expression. Loss of *PTEN* protein, defined by complete loss of expression in at least 50% of cancer cells, was detected in 40 out of 47 (about 85%) tumor sections

(Table 15). Although PTEN expression resulted downregulated in several cases, this condition was not correlated with up-regulation of miR-221 in our samples; probably other epigenetic mechanisms are involved in *PTEN* inhibition [118] (Table 15). Moreover, no significant correlation was observed between the diffuse loss of PTEN protein and survival (Long-Rank: $P=0.418$).

5.2.5 miRNAs expression analysis in OE19 and FLO-1 EAC cell lines

The expression of miR-483-3p and miR-221 was further characterized in two different esophageal adenocarcinoma cell lines, OE19 and FLO-1, via single Real-time PCR assays using RNU44 as endogenous control.

In both cell lines, there were not significant differences in miR-221 levels (Figure 45A), while miR-483-3p expression was absent in OE19 but positively increased in FLO-1 (fold-change=2.7; Figure 45B). OE-19 and FLO-1 also differed in the expression of *SMAD4* at RNA level. In particular, we performed *SMAD4* gene expression analysis with single Real Time PCR assays, using β - actin as reference gene to normalize data. Although it did not reach statistical significance, a reduction of *SMAD4* expression in FLO-1 was observed (Figure 46). Therefore, expression of miR-483-3p and *SMAD4* appear to be inversely correlated in FLO-1 cell line, suggesting the role of this miRNA in regulation of *TGF β* -pathway, as previously described [119;116].



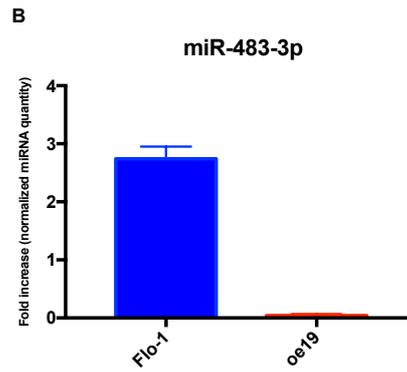


Figure 45: miR-221 (A) and miR-483-3p (B) expression levels in FLO-1 and OE19 EAC cell lines. The values are expressed as fold increase (2-DDCt) with respect to a commercial control RNAs pool (BioChain). Statistical comparisons between cell lines and controls was performed using Mann-Whitney test.

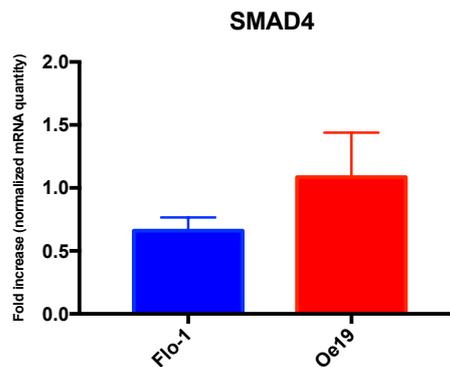


Figure 46: SMAD4 expression in FLO-1 and OE19 EAC cell lines, performed by Real Time PCR. The values are expressed as fold increase (2-DDCt) with respect to a commercial control RNAs pool (BioChain), using β -ACT as reference gene to normalize data. Statistical comparison between cell lines was performed using Mann-Whitney test ($P=0.4$).

6. Discussion

The work described in this thesis aimed to provide novel insights into the pathogenesis of esophageal adenocarcinoma (EAC), ranging from genetic profiles to protein expression patterns and miRNA signatures.

Furthermore, we sought to identify potential molecular targets and prognostic biomarkers which could lay the basis to detect novel targets for tumor therapy, predict the response to conventional treatment regimens and solve crucial controversies on EAC classification, supported by studies demonstrating differences in metastatic patterns, survival and immune-pathological parameters [29;48].

Defining the molecular landscape of cancer can be challenging because of tumor heterogeneity, therefore **in the first part** of this thesis we exploited an automatic sorting system (DEPArrayTM, Menarini Silicon Biosystems) enabling the isolation of pure tumor cell populations for an unambiguous genetic analysis with targeted next generation sequencing (NGS).

We first performed a pilot study using formalin-embedded (FFPE) tissue biopsies of the primary tumor and three metachronous chest metastases derived from a 53-year-old woman affected by an aggressive adenocarcinoma of the esophagus.

Whole exome sequencing (WES) analysis of the primary tumor and all metastases revealed a shared heterozygous *TP53* missense mutation (p.Arg282Trp), that was completely absent in patient's blood DNA, indicating that is a somatic mutation. Combining cell sorting and NGS analysis we found that the *TP53* mutant allele was completely mutated in the EAC and metastatic clusters, while it was wild-type in the separated stromal cells. This suggests that the *TP53* locus might have been involved in an early loss of heterozygosity (LOH) event in primary tumor cells, which can explain the homozygous state of the p.Arg282Trp mutation. This is in agreement with previous findings for which *TP53* mutations are considered an early genetic event in Barrett's esophagus associated with an increased risk of progression to cancer [4;57;59]. Therefore, from a technical perspective, the high-throughput sorting of the tumor cells led to the identification of somatic alterations without a "diluting" effect due to the presence of normal stromal cells. Since WES was carried out on whole tumor areas, the

presence of “normal” stromal cells within a mixed sample, such as a heterogeneous tumor tissue, prevented the identification of the real mutational status of the variant. In addition, the possibility to sort and recovery pure stromal cells provided an ideal internal control to avoid many false-positive alterations not specific to the patient’s tumor [90] and represents a valuable surrogate when matched normal tissue is unavailable, as may be the case for archival samples.

In addition to *TP53* mutation, we also detected with targeted sequencing and with WES an *ERBB2* gene amplification, that was completely absent in sorted stromal cells but shared by the primary and three metastatic tumor sites. The same amplification event was also identified by conventional diagnostic techniques (such as SISH and IHC) and had guided the use of anti-HER-2 drug Trastuzumab as the first targeted therapy approach. However, as in most Trastuzumab-responsive patients [120], the patient developed resistance after four years of treatment. The drug resistance can be due to mechanisms dependent on HER-2 receptor, such as overexpression of proteins that mask HER-2 receptor (for example, MUC1), up-regulation of target-like tyrosine kinase receptors or their ligands, and alternative splicing of *ERBB2* [120], or due to HER-2 independent mechanisms, such as altered PI3K/Akt signalling, deregulation of anti-apoptotic proteins / cell cycle regulators, TGF β signalling and epigenetic events regulating specific expression pathways, including the immune system signatures [120;121;122]. In our case, using the selective cell sorting coupled to NGS, we were able to observe that the *ERBB2* copy numbers detected were significantly lower in the metastases developed after initiation of Trastuzumab therapy, compared to the primary tumor, pointing to a selection of sub-clones HER-2 negative, therefore more resistant to treatment, although the histological appearance of cells in the tumor areas was homogeneous.

This decrease was not previously detected by standard tests such as SISH or IHC, which showed the same clusters of the *ERBB2* amplified region in the primary and metastatic tissues.

Being located on the same chromosome, *TP53* and *ERBB2* genes offer further insights for understanding the mutational events insisting in chromosome 17. The hypothesis, emerging from a combination of WES and target panel data, is that an early LOH event of the entire chromosome 17 occurred with prior/following somatic mutation in *TP53*

and large copy-gains of an *ERBB2* allele, influencing tumor development and progression. These data suggest that all metastases shared a common ancestor with *TP53* mutation and *ERBB2* amplification, and early sub-clonal evolution characterized by a partial loss of *ERBB2* amplification then occurred.

SNP-based phylogenetic analysis using WES data revealed a genomic divergence between metastases, in numbers of substitutions events during cancer progression. This is further confirmed by WES data that showed an exclusive focal amplification spanning *RNF146-ECHDC1* genes in the second chest metastasis and a *SMARCA4* nonsense mutation (p.E1512*) of new onset in the third metastasis.

RNF146 encodes for a E3 ubiquitin ligase ring finger protein 146, a critical regulator of Wnt/ β -catenin signaling, whose overexpression, reported in non-small cell lung cancer, enhances cell growth, invasion, and survival [123; 124].

ECHDC1 encodes for a proofreading enzyme involved in lipid metabolism, with an increased expression observed in resistant bladder cancer cells [125]. We hypothesize that the acquired *RNF146-ECHDC1* copy gain in the cells giving raise to the second metastasis, coupled to the loss of cells with *HER2* amplification, might contribute to resistance and progression in metastatic EAC cancer. A similar situation can be assumed in the third metastasis, where we identified a mutation hitting *SMARCA4* gene and causing a premature protein termination.

The *SMARCA4* gene encodes a catalytic subunit (BRG1) of SWI/SNF complexes, which function as regulators of gene expression by remodeling chromatin to alter nucleosome conformation, making it more accessible for transcriptional activation [127]. Mutations in *SMARCA4* resulting in complete loss of its protein (BRG1) occur frequently in non-small cell lung cancer (NSCLC) [127] and the presence of *SMARCA4*-inactivating mutations increases sensitivity to inhibitors of Aurora kinase A (AURKA), a kinase required for mitotic spindle assembly [128;129]. In addition, *SMARCA4* mutations are considered genetic biomarkers that predict enhanced sensitivity to topoisomerase II inhibitor (TopoIIi) in response to EZH2 inhibition (EZH2i). In particular, tumors with *SMARCA4* loss-of-function mutations respond to EZH2i with increased apoptosis and TopoIIi sensitivity [130].

Therefore, the genomic dissection of a primary EAC and recurrent metastases allowed to identify the tumor cell mutational status and its evolution in response to therapies.

Our preliminary study took advantage of an available target panel for cancer-related genes, for the study of specific mutations in digitally-sorted cell populations from formalin-embedded tissue biopsies.

Based on the results achieved in this pilot study, we exploited this high throughput selective sorting technology to investigate the different mutational patterns of 38 EAC patients, who were treated with surgery alone. In this line, we were able to investigate somatic genomic alterations present in different types of tumor cells without the “diluting” effect due to the presence of stromal cells, which were characterized by a normal diploid profile. In seven EAC cases different tumor cell populations were identified with hyperdiploid or pseudodiploid DNA content, that showed aberrant copy number profiles by low pass whole-genome analysis and different somatic mutation loads. This finding indicates the existence of different cell clones within the same tumor, each of which can have different tumor behavior and responses to therapy.

35 out of 38 EAC analyzed with OncoSeek Panel carried at least one alteration (point mutation, small insertion/deletion or copy number alteration), not present in the corresponding stromal cells, recovered with the same technology, confirming their somatic origin. Most of these mutations, that showed an alternative allele frequency of 50% in unsorted tumor DNA (heterozygous state), were present as almost unique alleles (homozygous state) in the sorted tumor cell populations.

Unexpectedly, mutations in *HNFI1A*, a gene not previously found mutated in EAC, were identified in two different samples. *HNFI1A* encodes the protein HNF1 α , a transcription factor predominantly expressed in the pancreas, kidney and liver [131;132;133]. Several proteins are actively involved in the interaction with HNF1 α , such as MYC (Myc protooncogene protein), HDAC1 (Histone deacetylase 1), CTNNB1 (catenin beta-1), TLE1 (Transducin-like enhancer protein1) and PDX1 (Pancreas/duodenum homeobox protein 1) [134]. Constitutive mutations in this gene cause maturity-onset diabetes of the young (MODY) [135], whereas somatic mutations were frequently observed in hepatocellular adenoma [136]. In pancreatic ductal adenocarcinoma, an aggressive cancer with poor prognosis, *HNFI1A* acts as a tumor suppressor, and loss-of-function mutations in this gene have been reported [115]. A recent study showed that the long noncoding RNA HNF1A-AS1 was markedly upregulated in human primary EACs relative to that in their corresponding normal esophageal tissues, and HNF1A-AS1

knockdown significantly inhibited cell proliferation and anchorage-independent growth in *in vitro* EAC models [137]. However, no mutations in EAC in the *HNF1A* have been reported thus far; therefore, a new gene mutated in EAC was found in this study. The mutations identified were found in conjunction with mutations in other genes and we expect that lost/mutated *HNF1A* might contribute to tumor severity/progression. Further analyses of additional cases are warranted to investigate the role of this gene in EAC to understand whether these mutations might act as a cancer driver or passenger.

Although the considerable level of genetic heterogeneity between patients, the *TP53* gene was confirmed to be the most frequently mutated gene in EACs, as previously described [4;56]. The majority of *TP53* mutations were shifted to homozygosity in the sorted tumor populations, suggesting that they are early events in tumorigenesis, as highlighted also by studies in Barrett's esophagus [57]. Interestingly, in two cases, the *TP53* mutations were present in ~100% of the sorted tumor cells, but by Sanger sequencing they were in heterozygous state only in some of the sequenced DNA samples, obtained from different extractions from serial sections of the same tumor area, reinforcing the concept of the high intra-tumor heterogeneity.

Mutations in *TP53* pave the way for many different molecular derangements that lead to diverse histopathological features of the tumors [59]. Thus, a fast and reliable detection of its variants is crucial for the accurate diagnostic decisions. Since in routine diagnostic pathology of gastric cancer [138] *TP53* mutation status is assessed based on p53 expression by IHC, we also characterized our samples for p53 expression in order to verify its ability to define the mutational status. Immunostaining for p53 correlated with the presence of missense mutations but could not help us distinguish the presence of loss-of-function mutations such as stop codon or frameshift variants from normal p53 staining, in concordance with previous data [112;113].

This discrepancy may suggest that a molecular analysis of the *TP53* mutational status, assessed with sequencing technologies, is of key importance, especially because the presence of mutations in *TP53* correlated with a better survival in our cohort of EACs. However, it is important to note that our patients were not treated with neoadjuvant cisplatin/fluorouracil chemotherapy, a treatment associated with poor response when p53 is mutated [74;75]. Our data support the idea that the use of cisplatin/fluorouracil should be selectively applied preoperatively to patients with wild-type *TP53*, whereas

patients with mutated p53 may benefit more from a timely surgical approach. *TP53* genetic status could be considered a predictive biomarker influencing survival only in the presence of effective chemotherapy, also in esophageal adenocarcinoma [73]. However, this model is based on a small sample size and further analyses in additional treatment-naïve EAC cases will help clarify this important issue in EAC treatment.

Since the complete loss of *TP53* wild-type protein in EAC provides a significant impact on prognosis and therapeutic options, its pharmacological reactivation emerged as a promising strategy for target therapy. Several small molecules that restore wild-type activity of mutant p53 have been identified such as APR-246/PRIMA-1Met, which is already under clinical trials for different cancers, including EAC [139;140]. This molecule restores *TP53* activity in presence of missense mutations and regulates several *TP53*-related pathways [139], including the proteasome machinery, a common target of *TP53* missense mutants. Upregulation of the proteasome fosters chemoresistance to proteasome inhibitors and this can be overcome by target p53 mutant proteins with APR-246/PRIMA-1Met [141]. Since p53 could be disabled either by mutation or by upstream negative regulators, including MDM2, targeting the MDM2-p53 interaction by small molecules, like Nutin-3a, could represent an additional therapeutic strategy to induced apoptosis in cancer cells with wild-type *TP53*, as previously described in primary B-acute lymphoblastic leukemia [142].

Therefore, the early identification of *TP53* mutation status in EAC is becoming increasingly instrumental not only for predicting chemotherapeutic drug response but also for selecting more efficient target therapies. In this regard, we suggest to perform the mutation analysis of selected tumor cells to determine the real zigosity of the *TP53* variants and the possible presence of a wild-type allele, which could be discriminating for some therapies, such as Nutin-3a [142].

Additional markers may help to characterize tumor progression and among them *SMAD4* expression was found to be a promising predictive factor for recurrence of esophageal adenocarcinoma. Loss of *SMAD4* protein expression was a recurrent event in EAC, linked with genetic mutations. In fact, all mutated samples showed a clear signal reduction, confirming a close correlation between the *SMAD4* genotype and the protein detection. However, *SMAD4* loss was also detected in a number of cases without *SMAD4* gene mutations (52.9%), and its expression significantly correlated

with tumor recurrence, also in absence of gene mutations, indicating that additional regulatory mechanisms might be involved, such as promoter hypermethylation, which can downregulate *SMAD4* expression [86]. These data are in line with previous studies on *SMAD4* loss in EAC, though those studies identified the immunoreactivity loss in fewer cases (10%) [85].

The incorporation of genomic differences in cancer cell sub-populations with currently available clinical variables can further stratify patients, in order to select more efficient therapeutic options. Moreover, whole exome approach in unsorted material could identify additional genomic alterations, compared to targeted sequencing, but not the true tumor cell mutational status. Therefore, a promising strategy could be to apply whole exome/genome approaches to sorted cell populations from formalin-embedded tissue samples, in order to gain a global view of all the tumor alterations.

In the second part of this thesis we aimed to investigate the dysregulated miRNAs in EAC, shifting the focus from a genetic to epigenetic characterization of these cancers.

By TaqMan Human MicroRNA Array and single Assays analysis of human EAC specimens, we identified a miRNA dysregulated expression pattern in esophageal adenocarcinoma tissues compared with esophageal/stomach normal tissues. In particular, two miRNAs implicated in tumor progression were found up-regulated: miR-483-3p and miR-221.

The hsa-miR-483-3 is a mammal-conserved microRNA that resides at the second intron of the human insulin growth factor 2 (*IGF2*) gene at the 11p15.5 chromosome region (Figure 47) [143]. The genomic localization of this microRNA is of particular interest. Indeed, the *IGF2* is an imprinted gene, expressed by the paternal allele to produce an important fetal insulin growth factor. Defects in the imprinting of the *IGF2* locus are observed in the Beckwith-Wiedemann syndrome, which increases the incidence of pediatric malignancies as nephroblastoma (Wilms' tumor), hepatoblastoma, and rhabdomyosarcoma [144]. Moreover, adult tumors are linked to genetic and epigenetic defects of this imprinted locus such as colorectal cancer, hepatocellular carcinoma and breast cancer [145;146], pointing out to *IGF2* as the main oncogene of this genomic locus. However, a transgenic mouse model for *IGF2* (without including the miR-483

gene) exhibited several features associated with the Beckwith-Wiedemann syndrome without association to any neoplasia [147]. These data suggested the *IGF2*/miR-483 locus as the oncogenic unit of the 11p15.5 chromosome region instead of the *IGF2* alone [147].

The hsa-miR-483 gene encodes for two mature miRNAs: miR-483-5p and miR-483-3p, both found deregulated in different types of cancer. MiR-483-5p is abnormally observed in the serum of cancer patients suggesting its possible utilize as circulating cancer biomarker in various malignancies [148]. On other hand, the miR-483-3p is extensively studied concerning its function on different cancer-related pathways and its expression levels is high or extremely high in many common cancer tissues, comprising 30% of cases of colon, breast and liver cancer [149].

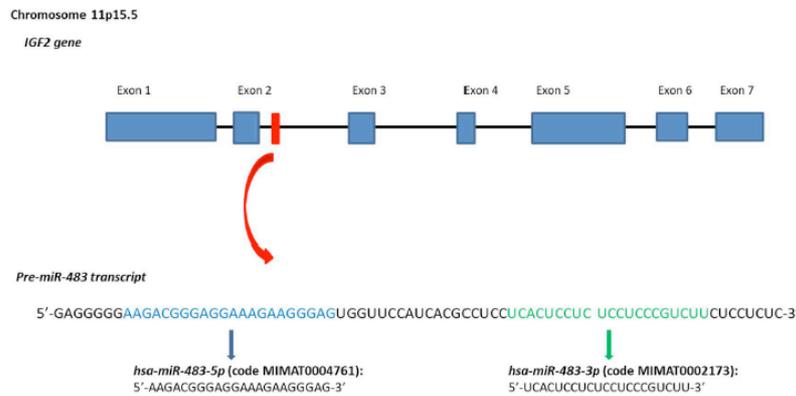


Figure 47: Stem-loop sequence of the hsa-miR-483 and mature miRNAs. In the Figure are reported genomic position, sequence of the hsa-miR-483 gene. Sequences data from miRBase database. (adapted from: Pepe F et al., 2018 [119]).

Together with our findings, it is suggested that miR-483-3p may be used a novel biomarker and therapeutic target also for esophageal adenocarcinoma. In fact, miR-483-3p up-regulation correlated in formalin embedded (FFPE) samples with reduced cancer-specific survival, in particular in Lauren intestinal subtype. Moreover, statistical analysis in FFPE EACs cohort showed an up-regulation of miR-483-3p in cancers without intestinal metaplasia in esophagus and in stomach (BIM-/GIM-) compared to Barrett's like (BIM+/GIM-) subtypes. Our data may support a biological difference of esophageal adenocarcinoma with or without intestinal metaplasia and between diffuse and intestinal subtypes. Further, an increased expression of miR-483-3p was also found in advanced cancer stages, suggesting that miR-483-3p signature might contribute to

monitoring angiogenesis and even supporting epithelial-mesenchymal transition [153]. These results were confirmed in several cancers, such as thyroid papillary carcinomas [154], breast cancer [155], hepatocellular carcinoma [156] and gastric cancer [117]. In particular, in gastric carcinoma miR-221 acts as a tumor promoting factor, essentially regulating *PTEN* expression [117]. *PTEN* is a gene located at 10q23.3 and encodes a dual-specificity phosphatase with lipid and protein phosphatase activities. *PTEN* acts as tumor suppressor gene through a variety of pathways and *PI3K* and *AKT* are its important target genes [157]. In gastric cancer, the overexpression of miR-221 significantly decreased PTEN protein which should inhibit phosphorylated *PI3K* and *AKT* resulting in an increase in proliferation, migration and invasion of cancer cells [153]. However, a similar effect may not be presented in EAC, where the up-regulation of miR-221 was not correlated with loss of PTEN protein. Probably the correlation highlighted between the expression of miR-221 and poor prognosis in our EAC samples is justified by its role in other cellular pathways.

Further *in vitro* studies are warranted to elucidate in depth the role of miR-221 and miR-483-3p in EAC.

We conducted preliminary studies in EAC cell lines, in particular to understand whether miR-483-3p, found dysregulated in FLO-1 cells, can correlate with more aggressive tumor phenotype, as observed in our EAC patients. To this aim, we used the scratch assay to test the invasive potential of esophageal cancer cells. Wound healing assay at 72 hours showed that the ability of migrated cells filling a scratch was significantly enhanced in FLO-1 cell line compared to OE19 cell line (Student's t-test $P=0.02$; Figure 49A,B), suggesting a possible correlation with upregulation of miR-483-3p. Hence, these observations prompt us to further study the functional effect of miR-483-3p via transient transfection of individual synthetic miRNA mimics or inhibitors into the different EAC cell lines, which can be a useful *in vitro* model for esophageal adenocarcinoma to assess miRNA functions and investigate possible target genes.

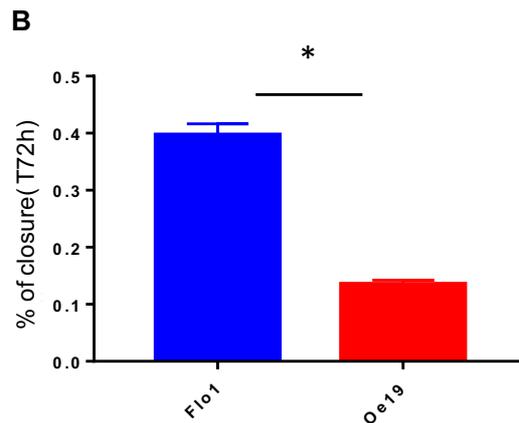
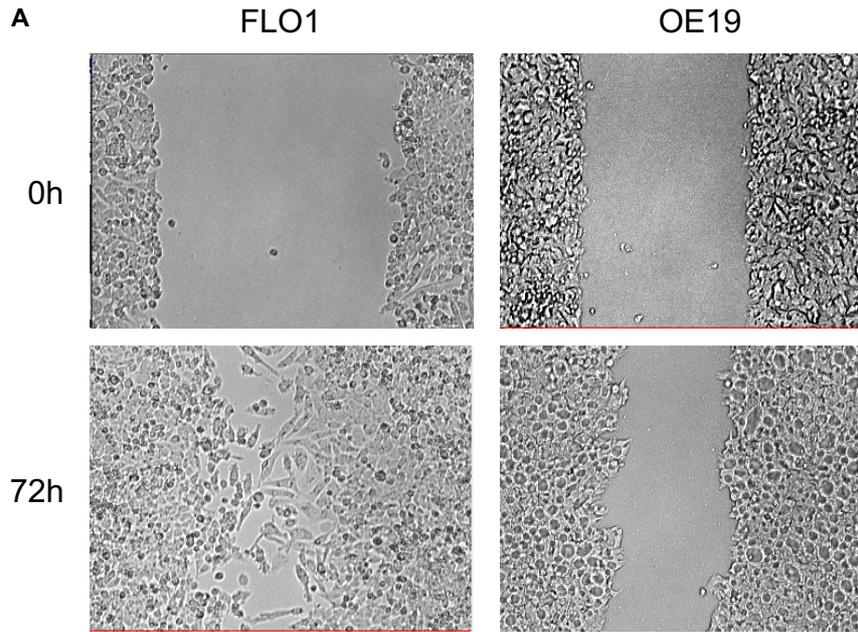


Figure 49: Representative photographs of wound healing assays showed that FLO-1 cells had an improvement in wound healing ability at 72h. The statistical results of cell migration area measured by Image J software demonstrated a significant difference in cell migration ability. * $P=0.02$

In contrast to other malignancies, such as breast and colon cancer, where the incorporation of molecular information has become part of routine practice for therapeutic stratification, in EAC the determination of disease stage and grade to guide treatment and help classify prognosis is still depending on histological assessments [158]. Even though the identification of new molecular biomarkers is still challenging in EAC, it is essential to predict patient outcome and facilitate tailored therapies based on the molecular profile.

To this aim, we used different approaches to investigate genetic and epigenetic features in EAC. In order to evaluate the genetic landscape, our studies suggest that cell sorting technologies, coupled with next generation sequencing analysis, are ideal tools to guide clinical decisions, since they can lead to a better definition of tumor mutation status and inter- and intra-tumor heterogeneity.

Concerning epigenetic factors influencing EAC development/progression, the analysis of microRNA profiles seems a promising strategy to distinguish more aggressive tumors. MiR-221 and miR-483-3p up-regulation correlated with worst prognosis in our tumors, implying that they can be considered oncogenic factors in EAC.

Further studies will be required to interpret the role of miRNAs in EAC patients.

7. Conclusions

In conclusion, our study showed the following:

(i) a combination of high-throughput sorting technology and massive parallel sequencing led to a better definition of EAC mutation status, allowing to identify mutations in *HNF1A*, a gene not previously found mutated in EAC, and to reinforce the presence of high inter- and intratumor heterogeneity in adenocarcinoma of esophagus.

(ii) we identified genetic and epigenetic biomarkers in EAC, such as *TP53*, *SMAD4* and specific microRNAs (miR-221 and miR-483-3p), correlated with tumor recurrence and survival. In particular, in our cohort of cancers not treated with neoadjuvant radio- or chemotherapy but surgery only, the presence of mutations in *TP53*, as detected more adequately in sorted cells, correlated with better cancer-specific survival (Log-Rank $P = 0.028$), and the observed loss of *SMAD4* protein (IHC) was statistically significant associated with cancer recurrence risk (P -value=0.015). Therefore, the evaluation of such markers may help facilitate tailored therapies, and predict patient outcomes in terms of survival and recurrence.

8. Tables

Table 1: Clinical and epidemiological information for EAC cases included in the study: FFPE cohort. In black: samples from Maria Cecilia Hospital, Cotignola, Italy; in red: samples from The Academic Medical Center Hospital, Amsterdam, Netherlands; in blue: samples from IRCCS Policlinico San Martino, Genova, Italy; in green: samples from Helsinki University Central Hospital, Helsinki, Finland. “*” indicates samples also subjected to genetic analysis. F=Female; M=Male. n.a.= not available

ID	Age	Sex	Cancer Specific Survival 1=dead	Follow-up (months)	Recurrence (1=yes)	Lauren	BIM/GIM	Stage (7 th ed)
EAC1 *	53	F	0	72	1	intestinal	BIM+/GIM-	II
EAC2 *	43	M	0	60	0	diffuse	BIM+/GIM-	II
EAC3 *	76	F	0	1	0	diffuse	BIM+/GIM-	III
EAC4 *	66	M	1	36	1	intestinal	BIM+/GIM-	II
EAC5 *	71	M	1	11	1	intestinal	BIM+/GIM-	III
EAC6 *	82	F	0	84	0	intestinal	BIM+/GIM-	III
EAC7 *	83	M	0	27	0	intestinal	BIM+/GIM-	III
EAC8 *	86	M	1	13	1	intestinal	BIM-/GIM+	III
EAC9 *	62	M	1	3	1	intestinal	BIM-/GIM+	III
EAC10 *	72	M	1	58	1	diffuse	BIM-/GIM+	I
EAC11 *	76	M	0	12	0	intestinal	BIM-/GIM-	III
EAC12 *	58	M	0	53	0	intestinal	BIM-/GIM-	III
EAC13 *	28	F	1	22	1	intestinal	BIM-/GIM-	IV
EAC14 *	83	F	0	0	0	intestinal	BIM-/GIM-	III
EAC15 *	60	F	0	84	0	intestinal	BIM-/GIM-	II
EAC16 *	78	M	1	36	1	intestinal	BIM-/GIM-	II
EAC17 *	59	M	0	24	1	intestinal	BIM-/GIM-	II
EAC18 *	75	M	1	8	1	intestinal	BIM-/GIM-	III
EAC19 *	44	F	1	35	1	diffuse	BIM-/GIM-	III
EAC20 *	79	M	0	0	0	diffuse	BIM-/GIM-	IV
EAC21 *	63	M	0	56	0	intestinal	BIM-/GIM-	II
EAC22 *	84	M	0	84	0	intestinal	BIM-/GIM-	II
EAC23 *	77	F	0	4	n/a*	intestinal	BIM-/GIM-	III
EAC24 *	78	F	1	10	1	intestinal	BIM-/GIM-	III
EAC25 *	80	M	0	71	0	diffuse	BIM-/GIM-	I
EAC26 *	74	M	0	19	0	intestinal	BIM-/GIM-	III
EAC27 *	68	M	0	0	0	intestinal	BIM-/GIM-	III
EAC28 *	72	M	0	6	0	intestinal	BIM-/GIM-	III
EAC29 *	67	M	0	14	0	intestinal	BIM-/GIM-	II
EAC30 *	82	M	0	14	1	intestinal	BIM-/GIM-	III
EAC31 *	66	M	1	33	1	intestinal	BIM-/GIM-	III
EAC32 *	54	M	1	15	0	intestinal	BIM-/GIM-	III
EAC33 *	87	M	0	0	0	intestinal	BIM-/GIM-	II

- Chapter 8-

EAC34 *	61	M	1	12	1	intestinal	BIM-/GIM-	III
EAC35 *	82	M	1	6	1	intestinal	BIM-/GIM-	IV
EAC36 *	65	M	1	29	1	intestinal	BIM+/GIM-	III
EAC37 *	62	M	1	5	1	intestinal	BIM+/GIM-	II
EAC38 *	54	F	0	6	0	diffuse	BIM-/GIM-	II
EAC39	75	F	1	45	1	intestinal	BIM-/GIM-	IV
EAC40	77	M	1	60	1	intestinal	BM-/GM+	IV
EAC41	75	M	0	16	0	intestinal	BM+/GM-	I
EAC42	69	M	0	24	0	intestinal	BIM-/GIM-	III
EAC43	70	F	1	48	1	diffuse	BM-/GM+	III
EAC44	66	M	1	15	1	intestinal	BM-/GM-	III
EAC45	54	F	1	12	1	intestinal	BIM-/GIM-	IV
EAC46	73	M	1	24	1	intestinal	BM+/GM+	I
EAC47	70	M	0	6	0	intestinal	BIM-/GIM-	III
EAC48	71	M	1	4	1	intestinal	BIM-/GIM-	III
EAC49	77	M	1	22	1	intestinal	BM+/GM+	III
EAC50	69	M	1	2	1	intestinal	BIM-/GIM-	III
EAC51	68	M	0	0	0	intestinal	BIM-/GIM-	III
EAC52	70	M	0	7	0	intestinal	BM+/GM-	I
EAC53	57	M	1	6	1	intestinal	BIM-/GIM-	IV
EAC54	72	M	0	7	0	intestinal	BM+/GM0	III
EAC55	80	M	0	4	0	intestinal	BM-/GM+	III
EAC56	76	M	1	n.a	1	intestinal	BIM-/GIM-	II
EAC57	69	M	0	7	0	intestinal	BM+/GM-	IV
EAC58	65	M	0	11	1	intestinal	BIM-/GIM-	IV
EAC59	80	F	0	0	n.a	diffuse	BIM-/GIM-	III
EAC60	75	M	0	35	0	intestinal	BM-/GM+	III
EAC61	82	M	1	20	1	intestinal	BIM-/GIM-	III
EAC62	58	F	0	15	0	intestinal	BM-/GM+	II
EAC63	72	M	0	53	0	intestinal	BM-/GM+	III
EAC64	72	M	0	29	0	intestinal	BIM-/GIM-	III
EAC65	63	M	0	36	0	intestinal	BM+/GM-	I
EAC66	83	M	0	18	0	intestinal	BM+/GM-	I
EAC67	67	F	0	29	0	intestinal	BM+/GM-	I
EAC68	66	M	0	31	0	intestinal	BM+/GM-	I
EAC69	63	M	0	38	0	intestinal	BM+/GM-	I
EAC70	62	M	0	58	0	intestinal	BM+/GM-	I
EAC71	52	M	0	57	0	intestinal	BM+/GM-	I
EAC72	60	F	0	41	0	intestinal	BM+/GM-	I
EAC73	66	M	0	84	0	intestinal	BM+/GM-	I
EAC74	83	F	0	6	0	intestinal	BM+/GM-	I
EAC75	72	M	1	0	n.a	diffuse	BIM-/GIM-	III
EAC76	80	F	0	67	0	mixed	BIM-/GIM-	I
EAC77	74	F	1	3	0	diffuse	BIM-/GIM-	II
EAC78	71	M	1	31	1	intestinal	BM+/GM-	III
EAC79	57	M	1	29	1	intestinal	BM+/GM-	III
EAC80	78	M	1	11	n.a	mixed	BM+/GM-	IV
EAC81	70	M	0	29	0	intestinal	BM+/GM-	III

- Chapter 8-

EAC82	54	M	0	30	1	intestinal	BM+/GM-	III
EAC83	66	M	0	57	1	intestinal	BM+/GM-	III
EAC84	68	M	0	10	0	intestinal	BM+/GM-	III
EAC85	82	M	0	24	1	intestinal	BM+/GM-	II
EAC86	46	M	0	25	0	intestinal	BM+/GM-	II
EAC87	47	M	0	28	0	intestinal	BM+/GM-	III
EAC88	60	M	0	49	0	diffuse	n.a.	II
EAC89	73	M	0	62	0	intestinal	n.a.	IV
EAC90	79	M	0	48	0	intestinal	n.a.	II
EAC91	41	M	1	33	1	intestinal	n.a.	III
EAC92	51	F	1	62	1	intestinal	n.a.	II
EAC93	48	F	0	98	0	n.a	n.a.	III
EAC94	62	F	1	46	1	intestinal	n.a.	II
EAC95	45	F	1	16	1	intestinal	n.a.	III
EAC96	61	M	1	22	1	intestinal	n.a.	II
EAC97	77	M	0	102	0	intestinal	n.a.	II
EAC98	13	M	0	84	0	mixed	n.a.	III
EAC99	78	M	1	38	1	intestinal	n.a.	III
EAC100	67	M	1	24	1	diffuse	n.a.	II
EAC101	77	M	0	56	0	intestinal	n.a.	II
EAC102	62	M	1	33	1	intestinal	n.a.	IV
EAC103	61	M	1	24	1	intestinal	n.a.	IV
EAC104	49	F	1	27	1	n.a.	n.a.	III
EAC105	62	F	1	34	1	intestinal	n.a.	III
EAC106	73	M	0	18	1	intestinal	n.a.	II
EAC107	51	M	1	0,46	1	intestinal	n.a.	III
EAC108	70	M	1	66	1	mixed	n.a.	III
EAC109	62	M	1	29	1	intestinal	n.a.	II
EAC110	69	F	0	29	1	intestinal	n.a.	III
EAC111	44	M	0	133	0	intestinal	n.a.	III
EAC112	68	F	1	40	1	intestinal	n.a.	III
EAC113	76	F	0	31	0	intestinal	n.a.	II
EAC114	77	M	0	16	0	intestinal	n.a.	II

Table 2: Genes represented in DEPAarray™ OncoSeek Panel

<i>ABL1</i>	5	<i>CDK4</i>	10	<i>ERBB4</i>	8	<i>GNAI1</i>	2	<i>KDR</i>	9	<i>MYCN</i>	9	<i>RET</i>	6
<i>AKT1</i>	2	<i>CDK6</i>	10	<i>EZH2</i>	1	<i>GNAQ</i>	2	<i>KIT</i>	24	<i>NOTCH1</i>	3	<i>SMAD4</i>	10
<i>ALK</i>	12	<i>CDKN2A</i>	2	<i>FBXW7</i>	6	<i>GNAS</i>	2	<i>KRAS</i>	11	<i>NPM1</i>	1	<i>SMARCB1</i>	4
<i>APC</i>	9	<i>CSF1R</i>	2	<i>FGFR1</i>	12	<i>HNF1A</i>	4	<i>MAP2K1</i>	5	<i>NRAS</i>	3	<i>SMO</i>	5
<i>AR</i>	10	<i>CTNNB1</i>	1	<i>FGFR2</i>	13	<i>HRAS</i>	2	<i>MET</i>	16	<i>PDGFRA</i>	13	<i>SRC</i>	1
<i>ATM</i>	19	<i>DDR2</i>	1	<i>FGFR3</i>	14	<i>IDH1</i>	1	<i>MLH1</i>	1	<i>PIK3CA</i>	19	<i>STK11</i>	5
<i>BRAF</i>	12	<i>DNMT3A</i>	1	<i>FGFR4</i>	9	<i>IDH2</i>	2	<i>MPL</i>	1	<i>PTEN</i>	14	<i>TP53</i>	21
<i>CCND1</i>	8	<i>EGFR</i>	19	<i>FLT3</i>	4	<i>JAK2</i>	2	<i>MSH6</i>	4	<i>PTPN11</i>	2	<i>TSC1</i>	1
<i>CDH1</i>	3	<i>ERBB2</i>	14	<i>FOXL2</i>	1	<i>JAK3</i>	3	<i>MYC</i>	9	<i>RBI</i>	12	<i>VHL</i>	3

Genes with contiguous, overlapping coverage
 Genes with comprehensive coding exon coverage
Genes with amplicons for CNA calling and SNV calling
Genes with only amplicons for CNA calling

Table 3: In the table, sequences of primers are reported

Primers	Sequences
SMARCA4 x32 F	5'-GTAAGACCTGCTCCTCCCG-3'
SMARCA4 x32 R	5'-ATTCAGAAGGGAAGGAGGGG-3'
TP53 x8 F	5'-CTTTTCCTATCCTGAGTAGTGGT-3'
TP53 x8 R	5'-TGTCCTGCTTGCTTACCTCG-3'
TP53 x6 F	5'-CCTCCTCAGCATCTTATCCGA-3'
TP53 x6 R	5'-TTGCAAACCAGACCTCAGG-3'
M13 F	5'-GTA AAAACGACGGCCAG-3'
M13 R	5'-CAGGAAACAGCTATGAC-3'

Table 4: Clinical and epidemiological information for EAC cases included in the study: fresh-frozen cohort.

n.a.= not available

ID	Cancer Specific Survival 1=dead	Follow-up (months)	1 (1=yes)	CROSS therapy	Lauren
T130	1	8	n.a.	no	intestinal
T131	1	21	1	yes	mixed
T133	1	28	1	no	intestinal
T136	0	30	0	no	diffuse
T137	0	33	0	no	intestinal
T138	1	12	n.a.	no	mixed
T139	1	2	n.a.	no	diffuse
T140	1	10,	1	no	diffuse
T141	1	21	n.a.	no	intestinal
T119	0	37	0	yes	intestinal
T120	0	5	0	no	diffuse
T121	1	20	n.a.	no	intestinal
T122	0	39	0	no	intestinal
T123	1	17	n.a.	no	intestinal
T124	1	8	1	yes	mixed
T125	0	36	0	yes	intestinal
T142	0	31	1	no	intestinal
T145	0	10	0	no	intestinal
T146	0	30	1	no	intestinal
T147	0	32	1	no	intestinal
T134	1	9	n.a.	no	intestinal
T150	1	8	n.a.	no	diffuse
T159	0	20	0	no	intestinal
T161	0	22	0	no	intestinal
T162	0	15	n.a.	no	intestinal
T5	1	16	1	yes	mixed
T7	1	15	1	yes	intestinal
T8	1	14	n.a.	no	diffuse
T9	1	17	1	yes	intestinal
T11	1	38	1	no	intestinal
T12	0	72	0	yes	mixed
T15	0	72	0	yes	intestinal
T16	1	32	1	no	intestinal
T17	1	40	1	yes	diffuse
T19	0	82	0	yes	intestinal
T27	0	70	0	yes	intestinal
T29	0	70	0	no	mixed
T30	0	21	0	yes	n.a.
T31	1	7	n.a.	no	intestinal
T34	1	20	1	yes	intestinal
T173	0	21	0	no	intestinal
T175	0	20	0	yes	mixed
T32	0	71	0	yes	diffuse
T43	1	19	n.a.	no	intestinal
T23	0	71	1	yes	intestinal
T24	0	5	0	yes	intestinal
T126	0	29	0	no	intestinal
T128	1	12	1	yes	intestinal
T112	0	44	0	no	intestinal
T115	1	14	1	yes	intestinal
T156	0	27	0	no	mixed
T116	0	38	0	no	intestinal
T158	0	22	0	yes	diffuse
T160	0	25	0	yes	intestinal
T163	0	24	0	no	diffuse

T166	0	21	0	yes	intestinal
T167	0	12	n.a.	no	intestinal
T169	0	22	1	no	mixed
T45	1	42	1	yes	diffuse
T155	1	15	n.a.	no	intestinal
T170	0	18	0	yes	intestinal

Table 5: Summary statistics regarding DEPArray™ recoveries and OncoSeek sequencing. DNA index is a measure of DNA content, determined comparing the integral-intensity DAPI of tumor population with that of stromal fraction, used as reference. All tumor populations show a hyperdiploid profile, compared to the expected diploid stromal profile. EAT (Effective Addressable Template) is a number estimating the usable template for DEPArray™ OncoSeek assay and can be considered a measure of sample quality. It is determined according to the following formula: $n.cells \times DNA_index \times 2 \times FFPE_QC_score$.

	type	library ID	n. cells	DNA index	FFPE QC	EAT	mean depth	uniformity
EAC primary tumor	stromal	L2057	205	1	0.21	86	656	77.9
		L2065	300	1	0.21	126	1,084	82.6
		L2612	224	1	0.21	94	1,922	83.1
	tumor	L2056	186	1.63	0.21	127	1,601	55.1
		L2066	239	1.58	0.21	159	1,027	57.4
		L2613	199	1.58	0.21	132	1,863	54.3
	unsorted	L2058	300	-	0.21	-	878	73.3
DNA-extracted	L1839	-	-	-	-	2,454	71.7	
metastasis 1	stromal	L2059	256	1	0.35	179	1,352	84.6
		L2614	300	1	0.35	210	1,472	88.6
		L2618	300	1	0.35	210	1,740	87.5
	tumor	L2060	127	1.52	0.35	135	1,191	64.7
		L2615	233	1.49	0.35	243	4,221	64.8
		L2619	213	1.46	0.35	218	4,276	66.6
	unsorted	L2061	300	-	0.35	-	1,010	79.2
metastasis 2	stromal	L2062	182	1	0.32	116	722	80.3
		L2616	284	1	0.32	182	1,774	88.1
		L2620	303	1	0.32	194	3,877	89.8
	tumor	L2063	211	1.57	0.32	212	1,245	63.0
		L2617	270	1.58	0.32	273	5,479	71.3
		L2621	300	1.59	0.32	305	4,607	67.6
	unsorted	L2064	300	-	0.32	-	1,401	74.7

Table 6: EXCAVATOR2 output for CNV on chromosome 6 and 17. EXCAVATOR2 results reports chromosome [CHROM], start [START], end [END] (positions compared to the reference genomic sequence hg19), segment length of a specific copy number [SEGMENT], copy number value (rounded to nearest integer) [CN], call probability inferred by FastCall algorithm [ProbCall], for EAC primary tumor (PT), first chest metastasis (M1) and second chest metastasis (M2).

	CHROM	START	END	SEGMENT	CN	ProbCall
EAC (PT)	6	109294558	109761795	467 kb	4	0.933
	6	123573555	123905087	331 kb	3	0.895
	17	37263657	38948823	1,7 Mb	17	0.999
M1	6	109308744	109761795	453 kb	4	0.945
	6	123573555	123905087	331 kb	4	0.780
	17	37295904	38948823	1,7 Mb	11	0.988
M2	6	108843497	109761795	918 kb	3	0.933
	6	123545236	123868517	323 kb	3	0.891
	6	127601482	127636156	34 kb	39	0.999
	17	37263657	38948823	1,7 Mb	11	0.999

Table 8: p53 immunohistochemistry and TP53 mutation status.

We report in black the missense mutation; in red stop codon (*) and frameshift (fs) mutations. wt=wild-type. TA classes and GVGD scores, according to IARC TP53 Database (<http://p53.iarc.fr/>) are based: a) for TA classes, on the overall transcriptional activity (TA) on 8 different promoters as measured in yeast assays [110]. For each mutation, the median of the 8 promoter-specific activities (expressed as percent of the wild-type protein) is calculated and missense mutations are classified as "non-functional" if the median is ≤ 20 , "partially functional" if the median is >20 and ≤ 75 , "functional" if the median is >75 and ≤ 140 , and "supertrans" if the median is >140 ; b) for the GVGD scores, classification is based on alignments obtained with Align-GVGD tool for TP53 missense variant prediction [159]. C15 is considered the best cut-off of pathogenicity.

EAC_ID	p53 immunostaining	TP53 mutation	TA class ^a	GVGD class ^b
EAC1	overexpression	p.R282W	non-functional	C65
EAC2	no overexpression	wt	-	-
EAC3	overexpression	p.C176F	partially functional	C65
EAC4	no overexpression	p.P75L	non-functional	C0
EAC5	no overexpression	p.P191Sfs*18	-	-
EAC6	no overexpression	p.R273H	non-functional	C25
EAC7	no overexpression	p.A79G / p.A74Efs*45	functional	C0
EAC8	overexpression	p.A138V / p.V73Rfs*76	partially functional	C55
EAC9	overexpression	p.R273C	non-functional	C65
EAC10	no overexpression	p.S303Afs*42	-	-
EAC11	overexpression	p.R273H	non-functional	C25
EAC12	overexpression	p.R273C	non-functional	C65
EAC13	no overexpression	wt	-	-
EAC14	no overexpression	wt	-	-
EAC15	no overexpression	p.R248Q	non-functional	C35
EAC16	overexpression	p.C275F	non-functional	C65
EAC17	overexpression	p.R175H	non-functional	C25
EAC18	no overexpression	p.R342*	-	-
EAC19	no overexpression	wt	-	-
EAC20	overexpression	p.P278S	non-functional	C65
EAC21	overexpression	p.R273C	non-functional	C65
EAC22	overexpression	p.R175H	non-functional	C25
EAC23	no overexpression	wt	-	-
EAC24	no overexpression	wt	-	-
EAC25	overexpression	p.R175H	non-functional	C25
EAC26	overexpression	p.R273H	non-functional	C25
EAC27	overexpression	p.R196*	-	-
EAC28	no overexpression	p.S15Rfs*28	-	-
EAC29	no overexpression	p.L43V	functional	C0
EAC30	overexpression	p.D259V	non-functional	C15
EAC31	overexpression	p.R248W	non-functional	C65
EAC32	overexpression	p.R267G	non-functional	C65
EAC33	no overexpression	p.Q167*	-	-
EAC34	no overexpression	wt	-	-
EAC35	no overexpression	wt	-	-
EAC36	overexpression	p.Y220C	non-functional	C65
EAC37	no overexpression	wt	-	-
EAC38	overexpression	wt	-	-

Table 9: Fisher's test output for p53 immunostaining and the presence of missense mutations.

	Missense change present	Others (no mutation / loss-of-function)	Total
p53 overexpression	17	2	19
p53 no overexpression	5	14	19
Total	22	16	38
<i>P</i> = 0.0002			

Table 10: Crosstab showing the relationship between *TP53* mutational status and Lauren's classification. The chi-squared test for given probabilities and Bonferroni's correction were applied to calculate the *P*-values (*P*) showed in table.

Lauren	<i>TP53</i>		Total	<i>P</i>
	wt	mutant		
Intestinal	7 (22.6%)	24 (77.4%)	31 (100%)	0.002
Diffuse	3 (42.9%)	4 (57.1%)	7 (100%)	0.705
Total	10 (26.3%)	28 (73.7%)	38 (100%)	

Table 11: Multivariate forward stepwise analysis of clinical and mutational parameters associated with survival. In table, we only reported the potential predictive parameters. The *TP53* mutations, the Lauren intestinal subtype and the cancers with intestinal metaplasia in esophagus (BIM+/GIM-) or in stomach (BIM-/GIM+), who showed a significant negative coefficient (b), are significantly correlated with better survival.

		b	Sig.	Exp(B)	95%CI for Exp(B)	
					Lower	Upper
TP53	mutation (yes)	-3.146	0.000	0.043	0.007	0.249
BIM/GIM						
	BIM+/GIM-	-3.522	0.001	0.030	0.004	0.241
	BIM-/GIM+	-3.451	0.001	0.032	0.004	0.262
Lauren	Intestinal	-3.035	0.005	0.048	0.006	0.405

Table 12: *SMAD4* immunoreactivity and genetic status of *SMAD4* and *TP53/CDKN2A* (TP53-pathway) genes.

EAC_ID	SMAD4 loss (%)	SMAD4 mutations	TP53-pathway mutations (1=yes)
EAC7	0	-	1
EAC6	0	-	1
EAC15	0	-	1
EAC26	0	-	1
EAC18	0	-	1
EAC14	0	-	0
EAC5	0	-	1
EAC20	0	-	1
EAC11	0	-	1
EAC33	0	-	1
EAC29	0	-	1
EAC34	10	-	0
EAC13	10	-	0
EAC3	15	-	1
EAC2	20	-	0
EAC25	20	-	1
EAC31	30	-	1
EAC9	40	-	1
EAC24	40	-	0
EAC32	50	-	1
EAC30	50	-	1
EAC23	50	p.R361C	0
EAC10	50	-	1
EAC4	60	-	1
EAC19	80	-	0

EAC21	80	-	1
EAC38	90	-	0
EAC12	90	p.S144*	1
EAC36	90	-	1
EAC22	90	-	1
EAC1	99	-	1
EAC16	100	-	1
EAC17	100	p.G176*	1
EAC35	100	-	1

Table 13: List of deregulated miRNAs determined by TaqMan MicroRNA Array cards in 8 EAC cases compared to two pools of 8 controls. Down-regulated miRNAs are highlighted in gray and up-regulated in orange.

Target name	Relative Quantity	Adjusted P-value
hsa-miR-1293-002905	316.564	0.015
hsa-miR-518b-001156	18.437	0.004
hsa-miR-483-3p-002339	11.033	0.050
hsa-miR-125b-000449	5.632	0.050
hsa-miR-708-002341	5.266	0.031
hsa-miR-604-001567	4.799	0.014
hsa-miR-146b-3p-002361	4.773	0.041
hsa-miR-214-002306	4.262	0.043
hsa-miR-1290-002863	3.693	0.028
hsa-miR-199a-3p-002304	2.885	0.024
hsa-miR-502-001109	2.815	0.018
hsa-miR-181c-000482	2.796	0.017
hsa-miR-221-000524	2.746	0.050
hsa-miR-330-000544	2.571	0.041
hsa-miR-331-000545	2.476	0.041
hsa-miR-671-3p-002322	2.398	0.028
hsa-miR-100-000437	2.334	0.050
hsa-miR-935-002178	2.213	0.041
hsa-miR-197-000497	2.091	0.050
hsa-miR-214#-002293	2.062	0.050
hsa-miR-337-5p-002156	2.026	0.028
hsa-let-7a-000377	1.983	0.038
hsa-miR-636-002088	1.981	0.050
hsa-miR-501-001047	1.960	0.047
hsa-miR-493-002364	1.573	0.039
hsa-miR-324-3p-002161	1.492	0.031
hsa-miR-484-001821	0.702	0.009
hsa-miR-30c-000419	0.620	0.010
hsa-miR-30b-000602	0.595	0.004
hsa-miR-376c-002122	0.557	0.001
hsa-miR-543-002376	0.538	0.024
hsa-miR-136#-002100	0.531	0.028
hsa-miR-186-002285	0.529	0.026
hsa-miR-148b#-002160	0.499	0.049

hsa-miR-342-5p-002147	0.492	0.003
hsa-miR-126#-000451	0.486	0.022
hsa-miR-33a#-002136	0.474	0.005
hsa-miR-26b-000407	0.469	0.001
hsa-miR-628-3p-002434	0.444	0.025
hsa-miR-10b#-002315	0.443	0.002
hsa-miR-29b-2#-002166	0.434	0.011
hsa-miR-942-002187	0.432	0.006
hsa-miR-590-3P-002677	0.424	0.040
hsa-miR-625-002431	0.422	0.011
hsa-let-7g-002282	0.412	0.000
hsa-miR-575-001617	0.402	0.025
hsa-let-7g#-002118	0.401	0.001
hsa-miR-505-002089	0.400	0.026
hsa-miR-628-5p-002433	0.385	0.001
hsa-miR-140-3p-002234	0.383	0.007
hsa-miR-146a-000468	0.347	0.041
hsa-miR-139-5p-002289	0.347	0.041
hsa-miR-411-001610	0.340	0.000
hsa-miR-199b-000500	0.339	0.001
hsa-miR-26a-1#-002443	0.332	0.003
hsa-miR-148a-000470	0.328	0.015
hsa-miR-598-001988	0.315	0.000
hsa-miR-136-000592	0.312	0.021
hsa-miR-888-002212	0.300	0.033
hsa-miR-30e-3p-000422	0.288	0.005
hsa-miR-30a-5p-000417	0.284	0.002
hsa-miR-15b#-002173	0.273	0.041
hsa-let-7a#-002307	0.258	0.003
hsa-miR-378-002243	0.257	0.038
hsa-miR-486-3p-002093	0.245	0.000
hsa-miR-200a-000502	0.244	0.026
hsa-miR-30a-3p-000416	0.235	0.001
hsa-miR-130b#-002114	0.234	0.012
hsa-miR-190b-002263	0.231	0.001
hsa-miR-192-000491	0.225	0.031
hsa-miR-148a#-002134	0.219	0.002
hsa-miR-411#-002238	0.216	0.002
hsa-miR-625#-002432	0.197	0.000
hsa-miR-326-000542	0.181	0.004
hsa-miR-485-5p-001036	0.180	0.009
hsa-miR-200a#-001011	0.178	0.013
hsa-miR-486-001278	0.173	0.047
hsa-miR-363-001271	0.173	0.003
hsa-miR-642-001592	0.167	0.000
hsa-miR-375-000564	0.167	0.025
hsa-miR-1262-002852	0.152	0.000
hsa-miR-376a#-002127	0.142	0.031
hsa-miR-144-002676	0.131	0.001
hsa-miR-520c-3p-002400	0.128	0.031
hsa-miR-30c-2#-002110	0.121	0.011
hsa-miR-135a-000460	0.121	0.002
hsa-miR-204-000508	0.120	0.001
hsa-miR-31-002279	0.078	0.011
hsa-miR-7-2#-002314	0.071	0.000

hsa-miR-1179-002776	0.065	0.003
hsa-miR-518d-5p-002389	0.064	0.008
hsa-miR-567-001534	0.063	0.026
hsa-miR-144#-002148	0.043	0.000
hsa-miR-672-002327	0.033	0.000
hsa-miR-595-001987	0.000	0.000
hsa-miR-520h-001170	0.000	0.000
hsa-miR-346-000553	0.000	0.000
hsa-miR-208b-002290	0.000	0.002

Table 14: SMAD4 immunoreactivity and miR-483-3p expression levels. Fold change was calculated using 2- $\Delta\Delta$ CT method, comparing FFPE tumor cases versus the pool of 8 control tissues. Up-regulated miRNA (cut-off value 3.15) and SMAD4 protein loss (cut-off value 30% of cancer cells) are highlighted in red; down-regulated miRNA and high SMAD4 expression are highlighted in green.

n.a.= not available

ID	% of SMAD4 LOSS	miR-483-3p fold-change
EAC7	0	2,5
EAC6	0	2,99
EAC15	0	1,51
EAC46	0	2,48
EAC42	0	53,53
EAC41	0	18,22
EAC26	0	6,54
EAC18	0	3,31
EAC47	0	18,23
EAC14	0	3,85
EAC5	0	n.a.
EAC20	0	0,72
EAC11	0	0,77
EAC33	0	3,45
EAC64	0	15,56
EAC29	0	3,74
EAC34	10	2,18
EAC43	10	6,89
EAC13	10	1,59
EAC60	10	24,7
EAC61	10	5,03
EAC59	10	7,33
EAC76	10	19,61
EAC3	15	6,69
EAC40	20	5,1
EAC2	20	n.a.
EAC25	20	2,79
EAC31	30	2,19
EAC45	40	18,05
EAC9	40	0,87

EAC24	40	37,49
EAC39	45	3,43
EAC32	50	14
EAC30	50	0,33
EAC23	50	5,61
EAC10	50	16,12
EAC4	60	0,12
EAC63	60	3,14
EAC77	60	42,25
EAC19	80	1,48
EAC75	80	15,69
EAC21	80	11,53
EAC44	90	3,55
EAC38	90	10,76
EAC12	90	1,91
EAC36	90	11,95
EAC62	90	11,78
EAC22	90	1,66
EAC48	95	12,28
EAC50	95	6,2
EAC1	99	16,45
EAC16	100	1,02
EAC17	100	1,4
EAC49	100	27,36
EAC35	100	11,75

Table 15: PTEN immunoreactivity and miR-221 expression levels. Fold change was calculated using 2- $\Delta\Delta$ CT method, comparing FFPE tumor cases versus the pool of 8 control tissues. Up-regulated miRNA (cut-off value 1.32) and PTEN protein loss (cut-off value 50% of cancer cells) are highlighted in red; down-regulated miRNA and normal PTEN expression are highlighted in green.

ID	PTEN expression	miR-221 fold-change
EAC16	LOSS	0,19
EAC40	LOSS	0,38
EAC15	LOSS	0,46
EAC30	LOSS	0,49
EAC54	LOSS	0,5
EAC33	LOSS	0,52
EAC12	LOSS	0,56
EAC29	NORMAL	0,58
EAC3	LOSS	0,61
EAC52	NORMAL	0,69
EAC31	LOSS	0,79
EAC1	NORMAL	0,82
EAC34	LOSS	0,84
EAC64	LOSS	0,88
EAC38	LOSS	0,9
EAC51	NORMAL	1

EAC60	LOSS	1,14
EAC9	LOSS	1,17
EAC28	LOSS	1,29
EAC4	LOSS	1,4
EAC32	NORMAL	1,48
EAC11	LOSS	1,58
EAC46	LOSS	1,64
EAC13	LOSS	1,69
EAC44	LOSS	2,08
EAC43	LOSS	2,09
AC59	NORMAL	2,09
EAC17	LOSS	2,12

9. Refernces

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int.J.Cancer* 2015;136:E359-86.
2. Coleman G, Xie SH, Lagergren J. The Epidemiology of Esophageal Adenocarcinoma. *Gastroenterology* 2018;154:390-405.
3. Siewert JR, Ott K. Are squamous and adenocarcinomas of the esophagus the same disease? *Semin Radiat Oncol.* 2007; 17:38-44.
4. The Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. *Nature* 2017; 541:169
5. Rice TW, Blackstone EH, Rusch VW. 7th edition of the AJCC Cancer Staging Manual: esophagus and esophagogastric junction. *Ann Surg Oncol.* 2010;17:1721-4.
6. Rice TW, Ishwaran H, Ferguson MK, et al. Cancer of the Esophagus and Esophagogastric Junction: An Eighth Edition Staging Primer. *J Thorac Oncol* 2017;12:36-42.
7. Lagergren J, Bergström R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med.* 1999;340:825-31.
8. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med.* 2011;365:1375-83.
9. Desai TK, Krishnan K, Samala N, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut.* 2012;61:970-6.
10. Duits LC, Phoa KN, Curvers WL, et al. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. *Gut.* 2015;64:700-6.
11. Kastelein F, Van Olphen S, Steyerberg EW, et al. Surveillance in patients with long-segment Barrett's oesophagus: a cost-effectiveness analysis. *Gut.* 2015;64:864-71.
12. Hoyo C, Cook MB, Kamangar F, et al. Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. *Int J Epidemiol.* 2012; 41:1706-18.
13. Singh S, Sharma AN, Murad MH, et al. Central adiposity is associated with increased risk of esophageal inflammation, metaplasia, and adenocarcinoma: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol.* 2013; 11:1399-1412.

14. Cook MB, Kamangar F, Whitman DC, et al. Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *J Natl Cancer Inst.* 2010; 102:1344-53.
15. El-Serag HB, Lagergren J. Alcohol drinking and the risk of Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology.* 2009; 136:1155-7.
16. Thrumurthy SG, Chaudry MA, Thrumurthy SSD et al. Oesophageal cancer: risks, prevention, and diagnosis. *BMJ.* 2019;366:l4373.
17. Kubo A, Block G, Quesenberry CP Jr, et al. Effects of dietary fiber, fats, and meat intakes on the risk of Barrett's esophagus. *Nutr Cancer.* 2009; 61:607-16.
18. Brown LM, Devesa SS, Chow WH. Incidence of adenocarcinoma of the esophagus among white Americans by sex, stage, and age. *J Natl Cancer Inst* 2008;100:1184-7.
19. Verbeek RE, Spittuler LF, Peute A, et al. Familial clustering of Barrett's esophagus and esophageal adenocarcinoma in a European cohort. *Clin Gastroenterol Hepatol* 2014;12:1656-63.
20. Ek WE, Levine DM, D'Amato M, et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. *J Natl Cancer Inst.* 2013; 105:1711-8.
21. Gharahkhani P, Fitzgerald RC, Vaughan TL, et al. Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. *Lancet Oncol.* 2016; 17:1363-1373.
22. Becker J, May A, Gerges C, et al. Supportive evidence for FOXP1, BARX1, and FOXF1 as genetic risk loci for the development of esophageal adenocarcinoma. *Cancer Med.* 2015; 4:1700-4.
23. Yang CS, Chen X and Tu S. Etiology and Prevention of Esophageal Cancer. *Gastrointest Tumors.* 2016; 3: 3–16.
24. Devesa SS, Blot WJ, Fraumeni JF Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer.* 1998;83:2049-53.
25. Pera M, Manterola C, Vidal O, et al. Epidemiology of esophageal adenocarcinoma. *J Surg Oncol.* 2005;92:151-9.
26. Odze R. What the gastroenterologist needs to know about the histology of Barrett's esophagus. *Curr Opin Gastroenterol* 2011; 27:389–3.
27. Peters FP, Kara MA, Rosmolen WD et al. Stepwise radical endoscopic resection is effective for complete removal of Barrett's esophagus with early neoplasia: a prospective study. *Am J Gastroenterol* 2006; 101:1449-57.

28. Hulscher JB, van Lanschot JJ. Individualised Surgical Treatment of Patients with an Adenocarcinoma of the Distal Oesophagus or Gastro-Oesophageal Junction. *Dig Surg* 2005; 22:130-4
29. Mattioli S, Ruffato A, Di Simone MP, et al. Immunopathological Patterns of the Stomach in Adenocarcinoma of the Esophagus, Cardia, and Gastric Antrum: Gastric Profiles in Siewert Type I and II Tumors. *Ann Thorac Surg* 2007;83:1814–9.
30. Purwar P, Bambarkar S, Jiwnani S, et al. Multimodality Management of Esophageal Cancer. *Indian J Surg*. 2014; 76: 494–503.
31. Kelsen DP, Ginsberg R, Pajak TF, et al. Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. *N Engl J Med*. 1998; 339:1979-84.
32. Medical Research Council Oesophageal Cancer Working Group. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial.. *Lancet*. 2002; 359:1727-33.
33. Bosset JF, Gignoux M, Triboulet JP, et al. Chemoradiotherapy followed by surgery compared with surgery alone in squamous-cell cancer of the esophagus. *N Engl J Med*. 1997; 337:161-7.
34. Apinop C, Puttisak P, Preecha N. A prospective study of combined therapy in esophageal cancer. *Hepatogastroenterology*. 1994; 4:391-3. APROC (ClinicalTrials.gov Identifier:NCT02999893).
35. Nygaard K, Hagen S, Hansen HS, et al. Pre-operative radiotherapy prolongs survival in operable esophageal carcinoma: a randomized, multicenter study of pre-operative radiotherapy and chemotherapy. The second Scandinavian trial in esophageal cancer. *World J Surg*. 1992; 16:1104-9.
36. Van Hagen P, Hulshof MC, Van Lanschot JJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012; 366:2074-84.
37. Oppedijk V, Van der Gaast A, Van Lanschot JJ, et al. Patterns of recurrence after surgery alone versus preoperative chemoradiotherapy and surgery in the CROSS trials. *J Clin Oncol*. 2014; 32:385-91.
38. Bang YJ, Van Cutsem E, Feyereislova A, et al. ToGA Trial Investigators. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet*. 2010;376:687-97.
39. Mawalla B, Yuan X, Luo X, et al. Treatment outcome of anti-angiogenesis through VEGF-pathway in the management of gastric cancer: a systematic review of phase II and III clinical trials. *BMC Res Notes*. 2018; 11: 21.

40. Barsouk A, Rawla P, Hadjinicolaou AV, et al. Targeted Therapies and Immunotherapies in the Treatment of Esophageal Cancers. *Med. Sci.* 2019, 7, 100.
41. Burnet FM. Immunological aspects of malignant disease. *Lancet.* 1967;1:1171-4.
42. Burnet FM. The concept of immunological surveillance. *Prog Exp Tumor Res.* 1970;13:1-27.
43. Sharpe AH, Pauken, KE. The diverse functions of the PD1 inhibitory pathway. *Nat. Rev. Immunol.* 2017;18:153.
44. Doi T, Bennouna J, Shen L, et al. KEYNOTE-181: Phase 3, open-label study of second-line pembrolizumab vs single-agent chemotherapy in patients with advanced/metastatic esophageal adenocarcinoma. *J Clin Oncol.* 2017;34:suppl 15:abstr TPS4140.
45. Robert C, Schachter J, Long GV, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med.* 2015;372:2521-32.
46. Dimberu PM and Leonhardt RM. Cancer Immunotherapy Takes a Multi-Faceted Approach to Kick the Immune System into Gear. *Yale J Biol Med.* 2011; 84: 371–80.
47. Fang L, Lonsdorf AS, Hwang ST. Immunotherapy for advanced melanoma. *J Invest Dermatol.* 2008;128:2596–605.
48. Ruffato A, Mattioli S, Perrone O, et al. Esophagogastric Metaplasia Relates to Nodal Metastases in Adenocarcinoma of Esophagus and Cardia. *Ann Thorac Surg.* 2013;95:1147–53
49. Siewert JR, Stein HJ. Adenocarcinoma of the gastroesophageal junction. Classification, pathology and extent of resection. *Dis Esoph.* 1996;9:173–82.
50. Siewert JR, Stein HJ. Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg.* 1998;85:1457-9.
51. Serra S, Chetty R. Non-adenomatous forms of gastro-oesophageal epithelial dysplasia: an under-recognised entity? *J Clin Pathol* 2014;67:898–902.
52. Piessen G, Messenger M, Leteurtre E, et al. Signet ring cell histology is an independent predictor of poor prognosis in gastric adenocarcinoma regardless of tumoral clinical presentation. *Ann Surg.* 2009;250:878–87.
53. Van der Kaaij RT, Snaebjornsson P, Francine EM, et al. The prognostic and potentially predictive value of the Lauren classification in oesophageal adenocarcinoma. *European Journal of Cancer* 2017;76:27-35.
54. Polkowski W, Van Sandick JW, Offerhaus GJA, et al. Prognostic Value of Laurén Classification and c-erbB-2 Oncogene Overexpression in Adenocarcinoma of the Esophagus and Gastroesophageal Junction. *Annals of Surgical Oncology* 1999;6:290-7.

55. Al-Batran S, Hofheinz RD, Pauligk C, et al. Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial. *Lancet Oncol.* 2016;17:1697-1708. 138.
56. Frankell AM, Jammula S, Li X, et al. The landscape of selection in 551 esophageal adenocarcinomas defines genomic biomarkers for the clinic. *Nat Genet.* 2019;51:506-16.
57. Contino G, Vaughan TL, Whiteman D, et al. The evolving genomic landscape of Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterology* 2017; 135:657-73.
58. Stachler MD, Taylor-Weiner A, Peng S, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet.* 2015; 47(9):1047-55.
59. Secrier M, Li X, De Silva N, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nature Genetics* 2016; 48: 1131–41.
60. Gokulana RC, Garcia-Buitragob MT, Zaikaa AI. From genetics to signaling pathways: molecular pathogenesis of esophageal adenocarcinoma. *BBA - Reviews on Cancer* 2019:1872;37–48.
61. Ismail A, Bandla S, Reveiller M, et al. Early G1 cyclin-dependent kinases as prognostic markers and potential therapeutic targets in esophageal adenocarcinoma. *Clin. Cancer Res.* 2011;17:4513-22
62. Malumbres M, Sotillo R, Santamaría D, et al. Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell.* 2004; 118:493-504.
63. Malumbres M, Carnero A. Cell cycle deregulation: a common motif in cancer. *Prog Cell Cycle Res.* 2003;5:5-18.
64. Ross-Innes CS, Becq J, Warren A, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma, *Nat. Genet.* 2015;47:1038–1046.
65. Soussi T, et al. “The p53 pathway and human cancer”. *Br J Surg.* 2005;92:1331–2.
66. Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* 1991;5:6304–11.
67. Olivier M, Hollstein M, Hainaut P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harb Perspect Biol.* 2010; 2: a001008.

68. Cho Y, Gorina S, Jeffrey PD, Pavletich NP. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*. 1994 Jul 15; 265:346-55.
69. Sigal A, Rotter V. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. *Cancer Res*. 2000; 60:6788-93.
70. Milner J, Medcalf EA, Cook AC. Tumor suppressor p53: analysis of wild-type and mutant p53 complexes. *Mol Cell Biol*. 1991 Jan; 11:12-9
71. Miller C, Mohandas T, Wolf D, Prokocimer M, Rotter V, Koeffler HP. Human p53 gene localized to short arm of chromosome 17. *Nature*. 1986; 319:783-4.
72. Oren M, Rotter V. Cold Spring Harb. Mutant p53 gain-of-function in cancer. *Perspect Biol*. 2010 Feb; 2:a001107.
73. Iacopetta B. TP53 mutation in colorectal cancer. *Hum. Mutat*. 2003;2:271–276.
74. Kandioler D, Mittlböck M, Kappel S, et al. TP53 Mutational Status and Prediction of Benefit from Adjuvant 5-Fluorouracil in Stage III Colon Cancer Patients. *EBioMedicine*. 2015; 2: 825–830.
75. Kandioler D, Schoppmann SF, Zwrtek R, et al. The biomarker TP53 divides patients with neoadjuvantly treated esophageal cancer into 2 subgroups with markedly different outcomes. A p53 Research Group study. *J Thorac Cardiovasc Surg*. 2014;148:2280-6.
76. Parrales A, Iwakuma T. Targeting Oncogenic Mutant p53 for Cancer Therapy. *Front Oncol* 2015;5:288.
77. Bian YS, Osterheld MC, Fontollet C, et al. p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus, *Gastroenterology* 2002:1113–21.
78. Huang Y, Peters CJ, Fitzgerald RC, et al. Progressive silencing of p14ARF in oesophageal adenocarcinoma. *J. Cell. Mol. Med*. 2009:13:398–409.
79. Vassilev LT, Vu BT, Graves B, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 2004;303:844-8.
80. Weiss A, Attisano L. The TGFbeta Superfamily Signaling Pathway. *WIREs Dev. Biol*. 2012;2:47–63.
81. Moses HL, Serra R. Regulation of differentiation by TGF-beta. *Curr. Opin. Genet. Dev*. 1996;6: 581–6.
82. Rees JR, Onwuegbusi BA, Save VE, et al. In vivo and in vitro evidence for transforming growth factor-beta1-mediated epithelial to mesenchymal transition in esophageal adenocarcinoma. *Cancer Res*. 2006;66:9583-90.

83. Liu F, Pouponnot C, and Joan Massague. Dual role of the Smad4/DPC4 tumor suppressor in TGF β -inducible transcriptional complexes. *Genes Dev.* 1997; 11: 3157–67.
84. Weaver MJ, Ross-Innes CS, Shannon N, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat. Genet.* 2014;46:837–843.
85. Singhi AD, Foxwell TJ, Nason K, et al. Smad4 Loss in Esophageal Adenocarcinoma Is Associated With an Increased Propensity for Disease Recurrence and Poor Survival. *Am J Surg Pathol.* 2015;39: 487–95.
86. Onwuegbusi BA, Aitchison A, Chin S-F, et al. Impaired transforming growth factor β signalling in Barrett's carcinogenesis due to frequent SMAD4 inactivation. *Gut* 2006;55:764-74.
87. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem.* 2009;55:641-58.
88. Meldrum C, Doyle MA, Tothill RW.. Next-Generation Sequencing for Cancer Diagnostics: a practical perspective. *Clin Biochem Rev.* 2011; 32: 177–95.
89. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012; 366:883-92.
90. Bolognesi C, Forcato C, Buson G, 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.
91. Di Trapani M, Manaresi N, Medoro G. DEPArray™System: An Automatic Image-Based Sorter for Isolation of Pure Circulating Tumor Cells. *Cytometry Part B* 2018; A93A: 1260–6.
92. Lee JW, Shin JY, Seo JS. Identification of novel mutations in FFPE lung adenocarcinomas using DEPArray sorting technology and next-generation sequencing. *J Appl Genet.* 2018;59:269-77.
93. Polioudaki, H. Agelaki S, Chiotaki R, et al. Variable expression levels of keratin and vimentin reveal differential EMT status of circulating tumor cells and correlation with clinical characteristics and outcome of patients with metastatic breast cancer. *BMC Cancer* 2015;15: 399.
94. Ngan CY, Yamamoto H, Seshimo I, et al. Quantitative evaluation of vimentin expression in tumour stroma of colorectal cancer. *Br J Cancer.* 2007; 96: 986–92.
95. Huntzinger E, Izaurralde E. Gene silencing by miRNAs: contributions of translational repression and mRNA decay. *Nat.Rev.Genet.* 2011;12:99-100.
96. MacFarlane LA, Murphy PR. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics.* 2010; 11: 537–561.
97. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004; 116:281-97.

98. Feber A, Xi L, Luketich JD, et al. MicroRNA expression profiles of esophageal cancer. *J. Thorac. Cardiovasc. Surg.* 2008;135:255-60.
99. Gu J, Wang Y, Wu X. MicroRNA in the pathogenesis and prognosis of esophageal cancer. *Curr Pharm Des.* 2013;19:1292-300.
100. Gao S, Zhao ZY, Zhang ZY, et al. Prognostic Value of MicroRNAs in Esophageal Carcinoma: A Meta-Analysis. *Clin Transl Gastroenterol.* 2018; 9: 203.
101. Lee TH, Guo H, Wang X, et al. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 2014;15:162.
102. Dietel M, Ellis IO, Höfler H, et al. Comparison of automated silver enhanced in situ hybridisation (SISH) and fluorescence ISH (FISH) for the validation of HER2 gene status in breast carcinoma according to the guidelines of the American Society of Clinical Oncology and the College of American Pathologists. *Virchows Arch.* 2007;451:19-25.
103. Yan P, Klingbiel D, Saridaki Z, et al. Reduced Expression of SMAD4 Is Associated with Poor Survival in Colon Cancer. *Clin Cancer Res.* 2016; 22:3037-47.
104. Bosman FT, Yan P, Tejpar S, et al. Tissue biomarker development in a multicentre trial context: a feasibility study on the PETACC3 stage II and III colon cancer adjuvant treatment trial. *Clin Cancer Res.* 2009;15:5528-33.
105. Boonstra JJ, Van Marion R, Beer DG, et al. Verification and unmasking of widely used human esophageal adenocarcinoma cell lines. *J Natl Cancer Inst.* 2010;102:271-4.
106. Rockett JC, Larkin K, Darnton SJ, et al. Five newly established oesophageal carcinoma cell lines: phenotypic and immunological characterization. *Br J Cancer.* 1997;75:258-63.
107. Hughes SJ, Nambu Y, Soldes OS, et al. Fas/APO-1 (CD95) is not translocated to the cell membrane in esophageal adenocarcinoma. *Cancer Res.* 1997;57:5571-8.
108. Contino G, Eldridge MD, Secrier M, et al. Whole-genome sequencing of nine esophageal adenocarcinoma cell lines. *F1000Res.* 2016;5:1336.
109. Gebäck T, Schulz MM, Koumoutsakos P, et al. TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays. *Biotechniques.* 2009;46:265-74.
110. Kato S, Han SY, Liu W, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A.* 2003;100:8424-9.
111. Temam S, Flahault A, Périé S, et al. p53 gene status as a predictor of tumor response to induction chemotherapy of patients with locoregionally advanced squamous cell carcinomas of the head and neck. *J Clin Oncol.* 2000;18:385-94.

112. Eicheler W, Zips D, Dörfler A, et al. Splicing mutations in TP53 in human squamous cell carcinoma lines influence immunohistochemical detection. *J Histochem Cytochem.* 2002;50:197-204.
113. Tolbert DM, Noffsinger AE, Miller MA, et al. p53 immunoreactivity and single-strand conformational polymorphism analysis often fail to predict p53 mutational status. *Mod Pathol.* 1999;12:54-60.
114. Connolly EC, Freimuth J, Akhurst RJ. Complexities of TGF- β targeted cancer therapy. *Int J Biol Sci.* 2012; 8:964-78.
115. Lou Z, Li Y, Wang H, et al. Hepatocyte nuclear factor 1A (HNF1A) as a possible tumor suppressor in pancreatic cancer. *PLoS One.* 2015;10:e0121082.
116. Hao J, Zhang S, Zhou Y, et al. MicroRNA 483-3p suppresses the expression of DPC4/Smad4 in pancreatic cancer. *FEBS Letters* 2011; 585:207–13.
117. Shi J, Zhang Y, Jin N, et al. MicroRNA-221-3p Plays an Oncogenic Role in Gastric Carcinoma by Inhibiting PTEN Expression. *Oncol. Res.* 2017;25:523-36.
118. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. *Nat Rev Mol Cell Biol.* 2012; 13:283–96.
119. Pepe F, Visone R, Veronese A. The Glucose-Regulated MiR-483-3p Influences Key Signaling Pathways in Cancer. *Cancers (Basel).* 2018;10. doi: 10.3390/cancers10060181.
120. Rexer BN, Arteaga CL. Intrinsic and acquired resistance to HER2-targeted therapies in HER2 gene-amplified breast cancer: mechanisms and clinical implications. *Crit. Rev. Oncol.* 2012;17:1-16.
121. Ebbing EA, Steins A, Fessler, et al. Esophageal adenocarcinoma cells and xenograft tumors exposed to Erb-b2 receptor tyrosine kinase 2 and 3 inhibitors activate transforming growth factor beta signaling, which induces epithelial to mesenchymal transition. *Gastroenterology* 2017;153:63-76.
122. Fumagalli D, Venet D, Ignatiadis M, et al. RNA sequencing to predict response to neoadjuvant anti-HER2 therapy: a secondary analysis of the NeoALTTO randomized clinical trial. *JAMA Oncol.* 2016;3:227-34.
123. Zhu X, Xing R, Tan R, et al. The RNF146 E3 ubiquitin ligase is required for the control of Wnt signaling and body pattern formation in *Xenopus*. *Mech Dev.* 2017; 147:28-36.
124. Gao Y, Song C, Hui L, et al. Overexpression of RNF146 in non-small cell lung cancer enhances proliferation and invasion of tumors through the Wnt/ β -catenin signaling pathway. *PLoS One.* 2014; 9:e85377.

125. Asai S, Miura N, Sawada Y, et al. Silencing of ECHDC1 inhibits growth of gemcitabine-resistant bladder cancer cells. *Oncol Lett.* 2018; 15:522-7.
126. Jelinic P, Mueller JJ, Olvera N, et al. Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. *Nature Genet.* 2014;46: 424-6.
127. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014; 511:543-50.
128. Tagal V, Wei S, Zhang W, et al. SMARCA4-inactivating mutations increase sensitivity to Aurora kinase A inhibitor VX-680 in non-small cell lung cancers. *Nat Commun.* 2017; 8:14098.
129. Barr AR, Gergely F. Aurora-A: the maker and breaker of spindle poles. *J Cell Sci.* 2007; 120:2987-96.
130. Fillmore CM, Xu C, Desai PT, et al. EZH2 inhibition sensitizes BRG1 and EGFR mutant lung tumors to TopoII inhibitors. *Nature* 2015; 520: 239–42.
131. Bjørkhaug L, Bratland A, Njølstad PR, et al. Functional dissection of the HNF-1alpha transcription factor: a study on nuclear localization and transcriptional activation. *DNA Cell Biol.* 2005;24:661-9.
132. Odom DT, Zizlsperger N, Gordon DB, et al. Control of Pancreas and Liver Gene Expression by HNF Transcription Factors. *Science* 2004; 303: 1378–81.
133. Tronche F, Yaniv M. HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays.* 1992;14:579-87.
134. Sneha P, Kumar DT, Lijo J, et al. Probing the Protein–Protein Interaction Network of Proteins Causing Maturity Onset Diabetes of the Young. *Adv Protein Chem Struct Biol.* 2018;10:167-202
135. Ellard S, Lango AH, De Franco E, et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. *Diabetologia* 2013;56:1958-63.
136. Jeannot E, Mellottee L, Bioulac-Sage P, et al. Spectrum of HNF1A somatic mutations in hepatocellular adenoma differs from that in patients with MODY3 and suggests genotoxic damage. *Diabetes.* 2010;59:1836-44.
137. Yang X, Song JH, Cheng Y, et al. Long non-coding RNA HNF1A-AS1 regulates proliferation and migration in oesophageal adenocarcinoma cells. *Gut.* 2014;63:881-90.
138. Ando K, Oki E, Saeki H, et al. Discrimination of p53 immunohistochemistry-positive tumors by its staining pattern in gastric cancer. *Cancer Med.* 2015; 4: 75–83.
139. Bykov VJ, Zhang Q, Zhang M, et al. Targeting of Mutant p53 and the Cellular Redox Balance by APR-246 as a Strategy for Efficient Cancer Therapy. *Front Oncol.* 2016; 6:21.
140. APROC (ClinicalTrials.gov Identifier:NCT02999893).

141. Lisek K, Walerych D, Del Sal G. Mutant p53–Nrf2 axis regulates the proteasome machinery in cancer. *Mol Cell Oncol.* 2017;4: e1217967.
142. Trino S, Iacobucci I, Erriquez D, et al. Targeting the p53-MDM2 interaction by the small-molecule MDM2 antagonist Nutlin-3a: a new challenged target therapy in adult Philadelphia positive acute lymphoblastic leukemia patients. *Oncotarget.* 2016; 7: 12951–61.
143. Fu H, Tie Y, Xu C, et al. Identification of human fetal liver miRNAs by a novel method. *FEBS Lett.* 2005; 579:3849-54.
144. Lapunzina P. Risk of tumorigenesis in overgrowth syndromes: a comprehensive review. *Am J Med Genet C Semin Med Genet.* 2005;137C:53-71.
145. Livingstone C. IGF2 and cancer. *Endocr Relat Cancer.* 2013; 20:R321-39.
146. Rainier S, Johnson LA, Dobry CJ, et al. Relaxation of imprinted genes in human cancer. *Nature* 1993; 362:747-9.
147. Sun FL, Dean WL, Kelsey G, et al. Transactivation of Igf2 in a mouse model of Beckwith-Wiedemann syndrome. *Nature.* 1997; 389:809-15.
148. Perge P, Butz H, Pezzani R, et al. Evaluation and diagnostic potential of circulating extracellular vesicle-associated microRNAs in adrenocortical tumors. *Sci Rep.* 2017; 7:5474.
149. Veronese A, Lupini L, Consiglio J, et al. Oncogenic role of miR-483-3p at the IGF2/483 locus. *Cancer Res.* 2010;70:3140–9.
150. Tabasi SA, Erson AE. MIR222 (microRNA 222). *Atlas Genet Cytogenet Oncol Haematol.* 2009;13:566-9.
151. Le Sage C, Nagel R, Egan DA, et al. Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation. *EMBO J.* 2007;26:3699–708.
152. Felli N, Fontana L, Pelosi E, et al. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci USA.* 2005;102(50):18081–6.
153. Garofalo M, Quintavalle C, Romano G, et al. miR221/222 in Cancer: Their Role in Tumor Progression and Response to Therapy. *Curr Mol Med.* 2012; 12: 27–33.
154. Visone R, Russo L, Pallante P, et al. MicroRNAs (miR)-221 and miR-222, both overexpressed in human thyroid papillary carcinomas, regulate p27Kip1 protein levels and cell cycle. *Endocr Relat Cancer.* 2007;14:791–8.
155. Miller TE, Ghoshal K, Ramaswamy B, et al. MicroRNA-221/222 Confers Tamoxifen Resistance in Breast Cancer by Targeting p27Kip1. *J Biol Chem.* 2008;283:29897–903.
156. Fu X, Wang Q, Chen J, et al. Clinical significance of miR-221 and its inverse correlation with p27Kip1 in hepatocellular carcinoma. *Mol Biol Rep.* 2011; 38:3029-35.

157. Bai ZG, Ye YJ, Shen DH, et al. PTEN expression and suppression of proliferation are associated with Cdx2 overexpression in gastric cancer cells. *Int J Oncol.* 2013;42:1682–91.
158. Weaver JM, Ross-Innes CS, Fitzgerald RC. The ‘-omics’ revolution and oesophageal adenocarcinoma. *Nat Rev Gastroenterol Hepatol* 2014;11:19–27.
159. Fortuno C, James PA, Young EL, et al. Improved, ACMG-compliant, in silico prediction of pathogenicity for missense substitutions encoded by TP53 variants. *Hum Mutat.* 2018;39:1061-9.

URL

- **Global Cancer Observatory**: <http://globocan.iarc.fr/>
- **FastQC**: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- **Burrows-Wheeler Aligner (BWA)**: bio-bwa.sourceforge.net
- **Picard**: <http://broadinstitute.github.io/picard/>
- **SAMtools**: <http://samtools.sourceforge.net/>
- **Genome Analysis Toolkit (GATK)**: software.broadinstitute.org/gatk/
- **Variant Effect Predictor (VEP)**: www.ensembl.org/Tools/VEP
- **PolyPhen-2**: genetics.bwh.harvard.edu/pph2
- **PROVEAN**: provean.jcvi.org
- **FATHMM**: <http://fathmm.biocompute.org.uk/>
- **MutationTaster**: www.mutationtaster.org
- **Human Splicing Finder (HSF)**: www.umd.be/HSF/
- **ESE Finder**: krainer01.cshl.edu/cgi-bin/tools/ESE3
- **EXCAVATOR2**: sourceforge.net/projects/excavator2tool
- **LoFreq**: csb5.github.io/lofreq
- **Cutadapt**: cutadapt.readthedocs.io
- **bedtools**: bedtools.readthedocs.io/en/latest
- **COSMIC**: <https://cancer.sanger.ac.uk/cosmic>
- **ClinVar**: <https://www.ncbi.nlm.nih.gov/clinvar>
- **The Cancer Genome Atlas Program (TCGA)**: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
- **IARC TP53 Database**: <http://p53.iarc.fr/>
- **miRBase**: <http://www.mirbase.org>

ACKNOWLEDGEMENTS

The study described in this thesis is part of a larger research project, called “*The mutation spectra of the whole-exome sequences and microRNA profiles of adenocarcinoma of the esophagus and cardia (ADECC)*” and developed by EACSGE group (Esophageal Adenocarcinoma Study Group Europe), whose principal investigator is Prof. Sandro Mattioli. The study received approval (# L3P1223) from the Department of Medical and Surgical Sciences (DIMEC), University of Bologna, and from the Ethical Committee “*Comitato Etico IRST IRCCS AVR (CEIIAV)*”- Italy (Reg. Sper. 109/2016 Protocol 7353/51/2016).

The project was supported by Prof. Mattioli’s personal research funds, partially supplemented by Maria Cecilia Hospital/GVM to the University of Bologna, and was partly funded by the AIRC IG1769 grant of Prof. Seri.

The PhD fellowship was supported by the University of Bologna /Regione Emilia-Romagna (Italy) for the “ONCOPENTA” project.

I thank Prof. Sandro Mattioli for giving me the opportunity to join his EACSGE research group and to do this PhD. He has coordinated the activities of all collaborators and has provided most of the analyzed cases.

I want to thank Prof. Marco Seri for having me in the Medical Genetic’s Lab and for believing in me.

I thank Prof. Niccolò Daddi, member of EACSGE research group, for finding the time to evaluate my thesis.

I want to thank Prof. Elena Bonora for supervising the experiments and giving me scientific support. Thank you for valuing my job, allowing me to work independently.

I thank all the collaborators of EACSGE research group, and in particular Prof. Kausilia K. Krishnadath for finding the time to review my works by showing always available to discuss the scientific data. She collected part of the analyzed samples and clinical data.

I am also very thankful to Prof. Roberto Fiocca, Prof. Antonietta D’Errico and their team who have conducted immunohistochemical experiments and helped me to interpret IHC results.

Also, I thank Dr. Marialuisa Lugaresi, as well part of EACSGE group, for the support you gave me for statistical analyses.

I thank all my collaborators from the Medical Genetics Lab, in particular Dr. Tommaso Pippucci for introducing me to the interesting world of bioinformatics.

Thanks to all my colleagues from the “Bonora group” for encouraging me to hold on.

Isotta, I will never thank you enough for the support you gave me in this project and for the great time we had in the lab.

Special thanks also to Chiara and Francesca for teaching me many “lab tricks” since before the start of my PhD. You made my daily lab-life easier, giving me precious advices. Thanks for the great time together in and outside the lab.

Thanks to all my closest friends to haven't left me alone during these three long years, always supporting me. Especially, thanks to Eugenia, Ilaria and Francesca, I know our friendship will last forever.

Grazie a David, per avermi supportata sin dai tempi dell'Università fino alla decisione di intraprendere questo Dottorato. Grazie per aver sempre creduto in me prima ancora che io stessa lo facessi. Sei stato la forza di cui avevo bisogno per completare questo viaggio.

Grazie in fine alla mia famiglia. In questi anni ho imparato che per risolvere qualsiasi difficoltà bastano gli sguardi un po' severi di mio padre, la voce tenera di mia madre, l'abbraccio sincero di mia sorella e i baci dolci di mia nipote.