

Exploring overexpression of PEA3 transcription factors as new possible mechanisms of resistance to EGFR targeted therapy in metastatic colorectal cancer.

Raquel Almeida Ramos

M
2023



Raquel Almeida Ramos. Exploring overexpression of PEA3 transcription factors as new possible mechanisms of resistance to EGFR targeted therapy in metastatic colorectal cancer.



Exploring overexpression of PEA3 transcription factors as new possible mechanisms of resistance to EGFR targeted therapy in metastatic colorectal cancer.

Raquel Almeida Ramos



RAQUEL ALMEIDA RAMOS

**EXPLORING OVEREXPRESSION OF PEA3 TRANSCRIPTION
FACTORS AS NEW POSSIBLE MECHANISMS OF RESISTANCE TO
EGFR TARGETED THERAPY IN METASTATIC COLORECTAL
CANCER**

Dissertação de Candidatura ao grau de Mestre em
Oncologia – Especialização em Oncologia Laboratorial
– submetida ao Instituto de Ciências Biomédicas Abel
Salazar da Universidade do Porto.

Orientador:

Doutora Paula Cristina Martins dos Santos Paulo
Investigadora Auxiliar no Grupo de Oncogenética do
Centro de Investigação do Instituto Português de
Oncologia do Porto.

Co-orientador:

Professor Doutor Manuel Rodrigues Teixeira
Professor Catedrático Convidado
Departamento de Patologia e Imunologia Molecular
Instituto de Ciência Biomédicas Abel Salazar da
Universidade do Porto.
Diretor do Serviço de Genética Laboratorial e
Coordenador do Grupo de Oncogenética do Centro de
Investigação do Instituto Português de Oncologia do
Porto.

ACKNOWLEDGMENTS

Em primeiro lugar quero agradecer ao Professor Manuel Teixeira por me ter dado a oportunidade de realizar a minha Tese de Mestrado no Grupo de Investigação de Oncogenética e por todo o tempo e atenção dispensados neste projeto.

Um dos agradecimentos mais importantes é à minha orientadora, Doutora Paula Paulo, por me ter confiado este projeto e por todo o esforço e tempo que lhe dedicou. Obrigada por acreditar em mim, por toda a paciência que teve e por todo o conhecimento que partilhou comigo, certamente acabo este ano mais rica a nível científico. Agradeço também todas as palavras de apoio, encorajamento e tranquilização quando tudo parecia correr menos bem, foi um privilégio trabalhar consigo.

À Doutora Andreia Brandão e ao Doutor Bernardo Orr pela ajuda ao longo deste ano e por todos os conselhos que me deram que, com toda a certeza, vou levar sempre comigo. Um obrigada adicional à Doutora Andreia pela grande ajuda e paciência na análise dos dados.

A toda a equipa do Serviço de Genética que desde o início me acolheu da melhor forma.

À equipa de oncologistas, Doutor Pedro Cruz, Doutora Olga Meneses, Doutora Deolinda Pereira e Doutora Dânia Santos pelo contributo essencial neste projeto.

À minha família e amigos, um obrigada não chega por tudo o que fizeram e fazem por mim.

Aos meus pais, a quem dedico este projeto, por estarem sempre ao meu lado, por me encorajarem a ultrapassar todos os obstáculos e por fazerem de mim uma pessoa cada vez melhor. Obrigada por me darem os melhores conselhos e fazerem com que tudo isto seja possível. Mais uma vez, a vossa menina conseguiu e espero que o vosso orgulho em mim continue a crescer!

À minha avó Alcina, obrigada por todo o carinho e apoio que me deste e continuas a dar. Às minhas quatro estrelinhas, avó Anunciação, avô Francisco, avô José e Alzira que, apesar de não estarem cá fisicamente, sei que estão orgulhosos do meu percurso e me guiam sempre pelo melhor caminho.

Às minhas melhores amigas, Dora, Filipa, Lécia e Leonor, por estarem sempre aqui para mim desde a licenciatura. Obrigada por todos os vossos conselhos, por ouvirem os meus áudios intermináveis sobre os altos e baixos da minha tese e, acima de tudo, por

todas as gargalhadas e todos os momentos que passamos juntas. São as melhores amigas que alguma vez poderia pedir!

Às minhas “Meninas da Genética”, Cátia, Elsa e Inês, por se terem revelado uma surpresa tão boa. Obrigada por me ouvirem e por todos os momentos que passamos este ano. Tenho a certeza de que este tempo foi só uma pequena amostra dos muitos anos de amizade que vamos ter! Ao Miguel, que passou de uma simples pessoa que me ajudou nas extrações infundáveis das minhas amostras, a um amigo que espero levar por muitos anos. À Ângela, que apesar de só nos termos conhecido no início do Mestrado tornou-se uma grande amiga sempre disponível para me ouvir e ajudar!

INDEX

ACKNOWLEDGMENTS	III
INDEX	VII
FIGURE INDEX	XIII
TABLE INDEX.....	XXI
ABBREVIATIONS	XXV
ABSTRACT	XXXI
RESUMO	XXXV
INTRODUCTION	39
1. Large bowel.....	41
1.1. Anatomy and physiology of the large bowel.....	41
2. Colorectal cancer	42
2.1. Epidemiology	42
2.2. Etiology	43
2.3. Screening and Diagnosis.....	43
2.3.1. Screening	43
2.3.1.1. Non-invasive screening tests	44
2.3.1.2. Invasive screening tests	44
2.3.1.3. Other screening tests under development	45
2.3.2. Diagnosis.....	46
2.4. Pathophysiology of colorectal cancer	48
3. Treatment	51
4. Prognostic biomarkers	52
4.1. TNM.....	52
4.2. Molecular Biomarkers	53
4.3. Histological Biomarkers.....	54
5. Epidermal Growth Factor Receptor	55
5.1. Gene, protein, and receptor.....	55
5.2. Main signalling pathways and oncogenic functions	55

5.2.1. The ETS family of transcription factors.....	57
5.2.1.1. ETS subfamilies.....	57
5.2.1.2. ETS physiologic and oncogenic functions.....	58
5.2.1.3. ETS in colorectal cancer	58
5.2.1.4. ETS as therapeutic targets.....	59
5.2.2. Gastrin-releasing Peptide Receptor.....	59
5.2.2.1. GRPR and cancer.....	59
5.2.2.2. GRPR and EGFR association	60
5.2.2.3. GRPR as a therapeutic target.....	60
5.3. EGFR inhibitors and mechanisms of resistance in colorectal cancer.....	61
AIMS	65
MATERIALS AND METHODS	69
1. Formalin-Fixed Paraffin-Embedded tissue samples.....	71
1.1. Patients' cohort validation and characterization	71
2. RNA extraction and cDNA synthesis	73
3. Quantitative Real-time Polymerase Chain Reaction (qRT-PCR).....	73
3.1. Data analysis	75
3.2. Quality control.....	76
4. Statistical analyses	77
4.1. Statistical Package for the Social Sciences (SPSS).....	77
4.2. Survival R package.....	77
RESULTS	79
1. Expression profile of PEA3 transcription factors and <i>EGFR</i> in <i>wild-type</i> tumours from mCRC patients highlights different “molecular subgroups”.....	81
2. Overexpression of PEA3 transcription factors or <i>EGFR</i> associates with OS and/or PFS.....	84
2.1. <i>EGFR</i>	84
2.2. <i>ETV5</i>	85
2.3. <i>ETV1</i>	86

2.4. <i>ETV4</i>	87
3. Prognostic impact of the major “molecular subgroups”	88
3.1. Overall Survival	88
3.2. Progression Free Survival	91
3.3. Predictive value of PEA3 and <i>EGFR</i> expression levels	95
4. Influence of clinicopathological characteristics in the association between PEA3 and <i>EGFR</i> “molecular subgroups” with OS and PFS	96
5. Definition of molecular predictors of response to anti-EGFR therapy considering tumour laterality	100
6. <i>GRPR</i> expression and its association with response to anti-EGFR therapy ...	101
DISCUSSION	103
1. The expression levels of <i>EGFR</i> and PEA3 transcription factors are associated with anti-EGFR therapy response	105
2. The prognostic value of each PEA3 transcription factor and of <i>EGFR</i> may be modulated by co-expression of one or more PEA3 members	107
3. Proposal of a “molecular profile” with potential clinical utility	107
4. <i>GRPR</i>	109
CONCLUSION	111
FUTURE PERSPECTIVES	115
REFERENCES	119
APPENDIX	131
Appendix 1	133
Appendix 2	134
Appendix 3	136
Appendix 4	141

FIGURE INDEX

Figure 1. Anatomy of the large bowel. (A) Representation of the anatomic sections that constitute the large bowel: cecum, ascending colon, transverse colon, descending colon, sigmoid, and rectum. (B) Representation of the different tissue layers that form the colon wall [adapted from (Shelton, B. K., 2002) (1)].....	41
Figure 2. Estimates of colorectal cancer incidence and deaths worldwide, in both sexes, in 2020 [adapted from GLOBOCAN 2020, available at https://gco.iarc.fr/].	42
Figure 3. Estimates of colorectal cancer incidence and deaths in Portugal, in both sexes, in 2020 [adapted from GLOBOCAN 2020, available at https://gco.iarc.fr/].	43
Figure 4. Representation of CRC stages, considering TNM classification, according AJCC [adapted from (Shek et al., 2021) (22)]	47
Figure 5. Systematization of colorectal cancer types (25).	48
Figure 6. Carcinogenic pathways in colorectal cancer. (A) Chromosomal Instability (CIN) pathway is characterized by inactivating mutations in tumour suppressor genes (<i>APC</i> , <i>TP53</i>) and activation of oncogenes (<i>KRAS</i>). (B) Microsatellite Instability (MSI) pathway is mainly characterized by inactivation of the DNA mismatch repair genes, such <i>MLH1</i> , leading to alterations in the microsatellite sequences. (C) Cytosine-phosphoguanine (CpG) island methylator phenotype (CIMP) pathway is associated with MSI and is characterized by methylation of CpG island of several tumour suppressor genes, being also associated with mutations in the <i>BRAF</i> oncogene [adapted from (Mundade et al., 2014) (33)].	51
Figure 7. EGFR structure and activation. (A) EGFR is composed by four extracellular domains - I, II, III, and IV - followed by the transmembrane domain (TM), and the intracellular tyrosine kinase domain (TK) and carboxy tail. (B) EGF ligand binds to EGFR extracellular domain (C) and leads to alteration in EGFR conformation allowing the homo or heterodimerization with other family receptors inducing TK autophosphorylation and activation of signalling pathways [adapted from (Sigismund et al., 2018) (51)].	55
Figure 8. Signalling pathways triggered by EGFR activation. The main intracellular effectors, leading to cell proliferation, inhibition of apoptosis, and angiogenesis, are shown (52).....	56
Figure 9. Clustering of ETS subfamilies according to the similarity of DNA binding domain sequence. The yellow, the orange and the red dots represent the most common rearrangements in cancer. The PEA3 subfamily is highlighted in green, and the ERG	

subfamily in blue. ETS members with known tumour suppressive activity are highlighted in pink [adapted from (Nicholas et al.,2019) (59)]..... 57

Figure 10. Schematic representation of the intracellular signalling inhibition using the monoclonal antibodies cetuximab or panitumumab. The anti-EGFR antibodies prevent ligand-induced activation of EGFR downstream signalling pathways, impairing tumour aggressiveness [adapted from (Leto & Trusolino, 2014) (93)]..... 62

Figure 11. Mechanisms of resistance to anti-EGFR therapy. (A) Activating mutations in oncogenes such as *RAS*, *RAF*, or *PI3K* can lead to signalling pathway activation despite EGFR inhibition. **(B)** Activation and amplification of other EGFR family tyrosine kinase receptors, such as HER2 and HER3, that activate the same oncogenic pathways as EGFR. **(C)** Mutations in EGFR ectodomain inhibits mAbs binding, decreasing EGFR inactivation and increasing downstream pathways activation. **(D)** Alterations in angiogenesis increase the secretion of VEGF anti-apoptotic factors and EGFR ubiquitination and degradation, which cause modifications in the expression of the receptor (93)..... 63

Figure 12. Principle of TaqMan probes. The proximity of the quencher dye to the fluorophore represses the fluorescence emitted by the fluorophore. After the annealing of the primers, the Taq DNA Polymerase initiates the extension phase adding nucleotides to synthesize the complementary DNA strand. When it reaches the annealed probe, its exonuclease activity cleaves the probe allowing the emission of fluorescence by the fluorophore. The fluorescence is measured, and the initial copy number is inversely proportional to the cycle when fluorescence is higher than the background, crossing the threshold in the amplification plot [adapted from (Butler, 2012) (99)] 75

Figure 13. Diagram representing the samples available for analysis in each step of the study, from the initial patient cohort (N=487) until the final patient cohort (N=111). The red boxes represent the samples that are not analysable due to insufficient RNA or lack of clinical information. The yellow boxes represent the samples that can be analysed in the future (N=285), and the green box represents the final patients' cohort that followed statistical analyses in the present study (N=111). 76

Figure 14. Percentage of samples with overexpression of *ETV1*, *ETV4*, *ETV5*, or *EGFR* considering the different expression cut-offs. 82

Figure 15. “Molecular subgroups” considering the Optimal-OS cut-off. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, *ETV5*, or *EGFR* - and samples with overexpression of

two or more genes (*right*). The “Others” group includes both negative samples for all the genes under study (N=8) and gene combinations found in less than four cases (N=2)... 83

Figure 16. “Molecular subgroups” considering the Optimal-PFS cut-off. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, or *EGFR* - and samples with overexpression of two or more genes (*right*). The “Others” group includes both negative samples for all the genes under study (N=3) and gene combinations found in less than four cases (N=5)..... 84

Figure 17. Kaplan-Meier curves of OS and PFS stratifying patients for *EGFR* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *EGFR* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown... 85

Figure 18. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV5* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV5* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown... 86

Figure 19. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV1* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV1* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown... 87

Figure 20. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV4* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV4* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown... 88

Figure 21. Kaplan-Meier curve of OS stratifying patients for PEA3 “molecular subgroups”. Survival probability over 125 months of patients with “*ETV1* Only”, “*ETV4* Only” and (A) “*ETV1+ETV4*”, or (B) “*ETV4+ETV5*” tumours. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown..... 89

Figure 22. Kaplan-Meier curve of OS comparing patients with extreme ETS “molecular subgroups”. Survival probability over 125 months of patients with “*ETV4* Only” tumours and tumours with overexpression of *ETV1* or *ETV5* and low expression of *ETV4* or *EGFR*. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown. 90

Figure 23. Kaplan Kaplan-Meier curve of OS comparing the different *ETV1/EGFR* “molecular subgroups”. Survival probability is shown over 125 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown. 90

Figure 24. Kaplan-Meier curve of PFS stratifying patients for *ETV1/ETV4* “molecular subgroups”. Progression-free probability is shown over 75 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown. 92

Figure 25. Kaplan-Meier curve of PFS stratifying patients for *PEA3* “molecular subgroups” with or without *ETV5*. Progression-free probability is shown over 75 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown. 92

Figure 26. Kaplan-Meier curve of PFS comparing patients with extreme ETS “molecular subgroups”. Progression-free probability over 75 months of patients with “*ETV4* Only” tumours and tumours with overexpression of *ETV1* or *ETV5* and low expression of *ETV4* or *EGFR*. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown. 93

Figure 27. Kaplan-Meier curve of PFS comparing the different *ETV1/EGFR* “molecular subgroups”. Progression-free probability is shown over 100 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown. 94

Figure 28. Multivariate Cox proportional hazards regression considering expression levels of *ETV1*, *ETV4*, *ETV5*, and *EGFR* for OS and PFS. For each outcome, the *Hazard Ratio (HR)*, the confidence interval (CI), and the *p*-value are shown for each gene. 96

Figure 29. Kaplan-Meier curves of OS stratifying patients’ tumours for both “molecular markers” and tumour laterality. Survival probability over 150 months for

patients with tumours from the right or left side of the colon and **(A)** high/low *EGFR* expression, **(B)** high/low *ETV1* expression, **(C)** high/low *ETV4* expression, and **(D)** high/low *ETV5* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown..... 99

Figure 30. Kaplan-Meier curves of PFS stratifying patients’ tumours for both “molecular markers” and tumour laterality. Progression-free probability over 100 months for patients with tumours from the right or left side of the colon and **(A)** high/low *EGFR* expression, **(B)** high/low *ETV1* expression, **(C)** high/low *ETV4* expression, and **(D)** high/low *ETV5* expression. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.....100

Figure 31. Kaplan-Meier curves of OS and PFS stratifying patients’ tumours for GRPR expression. Survival probability over 120 months (*left*), and progression-free probability over 120 months (*right*) for patients with “GRPR Only”, “GRPR+ETV5” and “Negative” tumours. The *p*-value of Log-rank and Generalized Wilcoxon statistical tests and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.....102

TABLE INDEX

Table 1. Sensitivity and recommended screening interval of each screening test used in colorectal cancer [adapted from (Simon K, 2016; Shaukat and Levin, 2022) (8, 15)].	45
Table 2. TNM classification for CRC, according to the AJCC and UICC [adapted from (Centelles, 2012) (21)].	47
Table 3. Staging groups of CRC according to the AJCC and UICC [adapted from (Centelles, 2012) (21)].	47
Table 4. Amsterdam II Criteria and Bethesda Guidelines used for the diagnosis of Lynch Syndrome [adapted from (Sinicrope, 2018) (31)].	50
Table 5. Clinicopathological characteristics of the mCRC patients involved in this study with available clinical data at IPO Porto, treated with anti-EGFR therapy, and with molecular information.	72
Table 6. Sequences of the primers and TaqMan probes designed to assess the expression levels of <i>ETV1</i> , <i>ETV4</i> , and <i>GRPR</i> . The amplicon size is also represented.	74
Table 7. Summary of the 5-years survival rate (%) and median survival rate (months) associated with the most important “molecular subgroups” identified.....	91
Table 8. Summary of the 5-years progression-free survival rate (%) and median progression rate (months) associated with the most important “molecular subgroups” identified.....	95
Table 9. Multivariate Cox proportional hazards regression analysis for OS considering each “molecular marker” and the clinicopathological features.	97
Table 10. Multivariate Cox proportional hazards regression analysis for PFS considering each “molecular marker” and the clinicopathological features.	98
Table 11. Proposed “molecular markers” of response to anti-EGFR therapy in mCRC patients considering tumour laterality.	101

ABBREVIATIONS

5-FU	5-Fluorouracil
ACTB	Bete-actin
AFAP	Attenuated Familial Adenomatous Polyposis
AJCC	American Joint Committee on Cancer
APC	Adenomatous Polyposis Coli
ASOs	Antisense oligonucleotides
BLPs	Bombesin-like peptides
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRS-3	Bombesin receptor subtype 3
BsMo	Bispecific molecules
cDNA	Complementary DNA
CIMP	Cytosine-phosphoguanine Island methylator phenotype
CIN	Chromosomal Instability
CpG	Cytosine-phosphoguanine
CRC	Colorectal Cancer
CT	Computed Tomography
DNA	Deoxyribonucleic acid
EGF	Epidermal growth factor
EGFR	Epidermal Growth Factor Receptor
ERG	v-ets avian erythroblastosis virus E26 oncogene homolog
ETS	E-26 transformation-specific
ETV1	ETS variant 1
ETV4	ETS variant 4

ETV5	ETS variant 5
FAP	Familial Adenomatous Polyposis
FDA	Food and Drug Administration
FFPE	Formalin-Fixed Paraffin-Embedded
FIT	Fecal Immunochemical Test
gFOBT	Guaiac Fecal Occult Blood Test
GPCR	G-protein coupled receptors
GRP	Gastrin Releasing Peptide
GRPR	Gastrin Releasing Peptide Receptor
HER	Human Epidermal Receptor
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HR	Hazard Ratio
iPARP	Poly (ADP-ribose) Polymerase Inhibitors
KRAS	Kristen rat sarcoma viral oncogene homolog
mAbs	Monoclonal antibodies
MAPK	Mitogenic-activated protein kinase
mCRC	Metastatic Colorectal Cancer
MLH1	Human mutL homolog 1
MMR	Mismatch repair
MRI	Magnetic Resonance Imaging
MSH2	MutS homolog 2
MSH6	MutS homolog 6
MSI	Microsatellite Instability

mt-sDNA	Multitarget stool DNA
NMB-R	Neuromedin B-receptor
NRAS	Neuroblastoma RAS viral oncogene homolog
OS	Overall Survival
PET	Positron Emission Tomography
PFS	Progression Free Survival
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PMS2	PMS1 homolog 2
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
SMIs	Small molecule inhibitors
TKIs	Tyrosine Kinase Inhibitors
TMPRSS2	Transmembrane Protease Serine 2
TNM	Tumour-Node-Metastasis
TP53	Tumor protein p53
UICC	International Union Against Cancer
VEGF	Vascular endothelial growth factor
VEGFR	Vascular Endothelial Growth Factor Receptor
Wnt	Wingless-Type MMTV Integration Site Family

ABSTRACT

Colorectal cancer is one of the most common cancers and the second cause of cancer deaths worldwide. In metastatic colorectal cancer (mCRC), the Epidermal Growth Factor Receptor (EGFR) inhibitors – cetuximab and panitumumab – are the most common therapy for tumours without activating mutations in *KRAS*, *NRAS*, or *BRAF* oncogenes – named *wild-type*. However, about 80% of the patients develop resistance to this therapy. Recently, members of the PEA3 subfamily of ETS transcriptions factors (which includes ETV1, ETV4 and ETV5) have been associated with CRC progression. Additionally, the Gastrin-Releasing Peptide Receptor (GRPR), known to transactivate EGFR, was also described to be up-regulated in a subset of the CRCs.

To understand if overexpression of PEA3 transcription factors, *GRPR* and *EGFR* can underly therapeutic resistance to anti-EGFR therapy, we analysed their expression profile in 111 “*wild-type*” carcinomas from mCRC patients treated with anti-EGFR therapy and followed at IPO Porto. We assessed the influence of gene expression levels in Overall Survival (OS) and in Progression Free Survival (PFS) using Kaplan-Meier curves considering the optimal *cut-off* for each outcome. Multivariate Cox proportional hazards regression analyses were used to investigate the independence of the “molecular markers” from each other and from the clinicopathological features.

Our results suggest that overexpression of *ETV5* is associated with worse OS and PFS ($p=0.015$ and $p=0.048$, respectively), while overexpression of *ETV1* is associated with worse PFS ($p=0.018$). Conversely, overexpression of *EGFR* or *ETV4* is associated with better OS and PFS, respectively ($p=0.04$ and $p=0.019$, respectively). In addition, *ETV4* seems to antagonise the worse prognosis of *ETV1* in OS and PFS ($p=0.012$ and $p=0.001$, respectively), while overexpression of *ETV1* seems to reverse the good therapy response associated with *EGFR* overexpression ($p=0.018$). In multivariate analyses, the prognostic value of some of the “molecular markers” showed to depend on tumour laterality. Thus, we observed that, overexpression of *ETV5* or *ETV1* is associated with worse OS and PFS, respectively ($p=0.00033$ and $p=0.00029$, respectively), independently of tumour laterality. Additionally, in *right* colon tumours, high *EGFR* expression is associated with better PFS ($p=0.00071$), whereas high *ETV4* expression is associated with better OS ($p<0.0001$). Finally, *GRPR* overexpression revealed to be a possible biomarker of good response to anti-EGFR therapy when *ETV5* is not co-expressed.

We conclude that the PEA3 transcription factors, *EGFR*, and *GRPR* are possible biomarkers for mCRC, demonstrating prognostic value in the response to anti-EGFR therapy and potentially contributing to optimize patients’ stratification and treatment management, opening opportunities for the development of new targeted therapies.

RESUMO

O cancro colorretal (CRC) é um dos cancros mais comuns e a segunda causa de morte por cancro em todo o mundo. Os inibidores do *Epidermal Growth Factor Recetor* (EGFR) – cetuximab e panitumumab – são a terapia mais comumente usada no tratamento de carcinomas metastáticos sem mutações ativantes nos oncogenes *KRAS*, *NRAS* ou *BRAF* – designados "*wild-type*". No entanto, cerca de 80% dos doentes desenvolvem resistência a esta terapêutica. Recentemente, os membros da subfamília PEA3 de fatores de transcrição ETS (que inclui o ETV1, ETV4 e ETV5) foram associados à progressão do cancro colorretal. Adicionalmente, o *Gastrin-Releasing Peptide Receptor* (GRPR), envolvido na transactivação do EGFR, foi também descrito como sobre-expresso num subgrupo de carcinomas colorretais.

De forma a avaliar se a sobre-expressão dos fatores de transcrição PEA3, bem como de *GRPR* e *EGFR* estão relacionados com o desenvolvimento de resistência à terapia dirigida, analisamos o perfil de expressão genética em 111 carcinomas metastáticos "*wild-type*" tratados com terapia anti-EGFR e seguidos no IPO Porto. A influência dos níveis de expressão na Sobrevivência Global (SG) e na Sobrevivência Livre de Progressão (SLP) foi analisada através de curvas de Kaplan-Meier considerando o *optimal cut-off* de expressão para cada alvo. Para avaliar a independência entre "marcadores moleculares" e entre estes e as características clínico patológicas dos doentes foram realizadas análises multivariadas usando o modelo de riscos proporcionais de Cox.

Os resultados obtidos sugerem que a sobre-expressão de *ETV5* está associada a uma pior SG e SLP ($p=0.015$ and $p=0.048$, respetivamente), enquanto a sobre-expressão de *ETV1* se associa a pior SLP ($p=0.018$). Por outro lado, a sobre-expressão de *EGFR* ou *ETV4* associa-se a melhor SG e SLP, respetivamente ($p=0.04$ and $p=0.019$, respetivamente). Além disso, o *ETV4* parece reverter o pior prognóstico dado pelo *ETV1* na SG e SLP ($p=0.012$ and $p=0.001$, respetivamente), enquanto a sobre-expressão de *ETV1* parece anular a melhor resposta terapêutica associada à sobre-expressão de *EGFR* ($p=0.018$). Na análise multivariada, o valor prognóstico de alguns dos "marcadores moleculares" mostrou ser dependente da lateralidade do tumor. Assim, observámos que a sobre-expressão de *ETV5* e *ETV1* está associada a pior SG e SLP, respetivamente ($p=0.00033$ and $p=0.00029$, respetivamente), de forma independente da lateralidade. Além disso, em tumores do cólon direito, a expressão elevada de *EGFR* está associada a uma melhor SLP ($p=0.00071$), enquanto a expressão elevada de *ETV4* associa-se a uma melhor SG ($p<0.0001$). Finalmente, a sobre-expressão de *GRPR* revelou ser um possível

biomarcador de boa resposta à terapêutica anti-EGFR na ausência de co-expressão de *ETV5*.

Concluimos que a subfamília de fatores de transcrição PEA3, o *EGFR* e o *GRPR* são possíveis biomarcadores na doença metastática, apresentando valor prognóstico na resposta à terapia anti-EGFR, com potencial utilidade na otimização da estratificação dos doentes e na gestão terapêutica, bem como no desenvolvimento de novas abordagens terapêuticas.

INTRODUCTION

1. Large bowel

1.1. Anatomy and physiology of the large bowel

The large bowel is a component of the digestive tract and averages approximately 1,5m in length (1, 2). Looking at its functions, the large intestine is responsible for absorbing water, electrolytes, and vitamins from indigestible food, being also involved in the formation of feces (2, 3).

The large bowel is mainly composed by four parts: cecum, ascending colon, transverse colon, and descending colon (Figure 1A). More specifically, the cecum is the blind portion of the ascending colon, a retroperitoneal organ that is continuous to the transverse colon. This last one is the most mobile and the longest part of the large intestine, and is continuous to the descending colon, which connects to the rectum by an S-shaped loop called sigmoid (3).

The colon is composed by several tissue layers (Figure 1B). The mucosal layer is the one that is more susceptible to injuries, once is in direct contact with all the substances. It is also in this layer that the crypts of Lieberkuhn are located, and where the colon cell development occurs. Consequently, any damage in these crypts can lead to abnormal cell proliferation and inflammation, resulting in an adenoma (1).

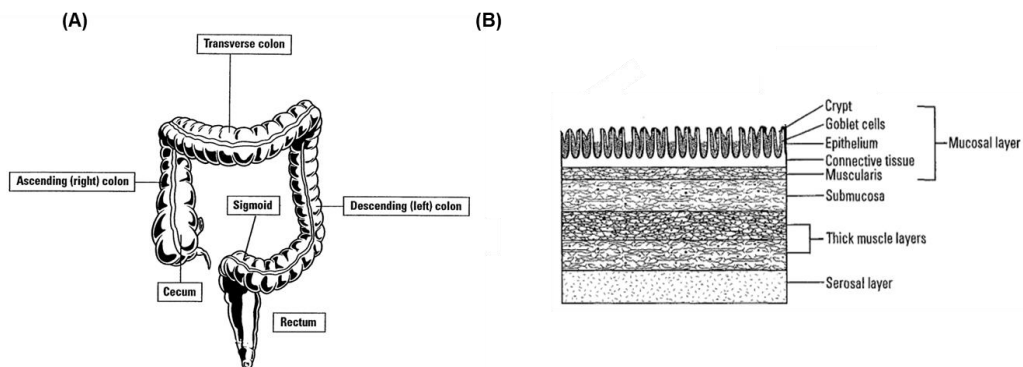


Figure 1. Anatomy of the large bowel. (A) Representation of the anatomic sections that constitute the large bowel: cecum, ascending colon, transverse colon, descending colon, sigmoid, and rectum. **(B)** Representation of the different tissue layers that form the colon wall [adapted from (Shelton, B. K., 2002) (1)].

2. Colorectal cancer

2.1. Epidemiology

According to GLOBOCAN statistics for 2020, colorectal cancer (CRC) was the third most common cancer worldwide, with over 1.9 million new diagnosed cases, and the second cause of cancer death, with over 935,000 deaths, representing 10% of the global cancer incidence and 9,4% of all cancer deaths (Figure 2). Additionally, CRC is the third most common cancer in males and the second most frequently diagnosed cancer among females (4). Projections for 2040 point to an increase in both the number of new CRC cases and associated deaths to about 3 and 1.6 million, respectively. Moreover, CRC is the most common cancer in developed countries, namely Australia and Europe, a fact that is not expected to be changed in 2040 (5, 6).

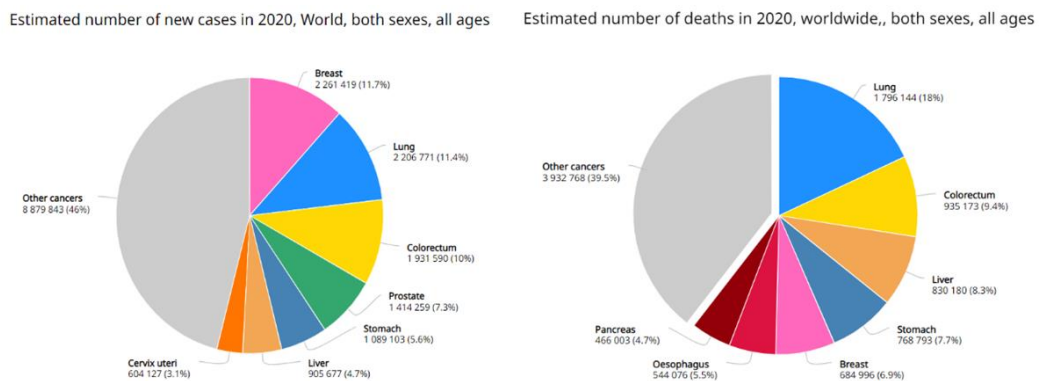
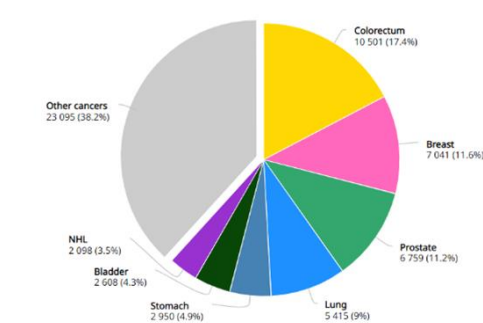


Figure 2. Estimates of colorectal cancer incidence and deaths worldwide, in both sexes, in 2020 [adapted from GLOBOCAN 2020, available at <https://gco.iarc.fr/>].

In Portugal, estimates for 2020 point to ≈10,500 new CRC cases, 64.2 hundreds in males and 40.8 hundreds in females. In the same period, 25.6 hundreds deaths in males, and 17.6 hundreds in females, were accounted for CRC (7). According to GLOBOCAN (<https://gco.iarc.fr>), in 2020, CRC was the most common cancer in Portugal, and the second cause of cancer deaths, in both sexes, after lung cancer (Figure 3).

Estimated number of new cases in 2020, Portugal, both sexes, all ages



Estimated number of deaths in 2020, Portugal, both sexes, all ages

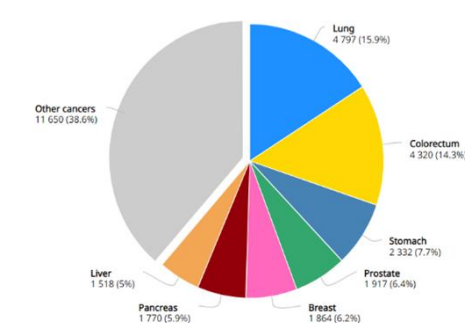


Figure 3. Estimates of colorectal cancer incidence and deaths in Portugal, in both sexes, in 2020 [adapted from GLOBOCAN 2020, available at <https://gco.iarc.fr/>].

2.2. Etiology

As mentioned before, CRC is one of the top three most common cancer worldwide and despite everyone can develop this disease, many factors contribute to increase that risk. Some of these factors, like diet, alcohol, sedentary lifestyle, and tobacco, are modifiable. In contrast, others are classified as nonmodifiable, such as age, ethnicity, and personal or familial history of other diseases, as Lynch syndrome, type two diabetes, familial history of colorectal polyps or colorectal cancer, and personal history of inflammatory bowel disease, such as Crohn's disease (8, 9). In fact, people over the age of fifty have higher risk to develop the disease, and both the mortality and the incidence rates are higher in non-Hispanic people (9, 10). Genetic factors also increase the risk to develop CRC, which will be discussed in Chapter 2.4.

2.3. Screening and Diagnosis

Colorectal cancer is, usually, a silent disease. Symptoms are more common in advanced disease and may include blood in stools, changes in bowel habits, abdominal pain, fatigue, anemia, and weight loss (11). All these symptoms have an important predictive value for CRC detection and should be complemented with a clinical evaluation using screening tests. Moreover, the implementation of these technics in the population is important in order to remove precancerous polyps, detect early stages of CRC, and reduce the mortality risk (12, 13).

2.3.1. Screening

The screening tests may include non-invasive tests, as the Fecal Immunochemical Test (FIT), the guaiac Fecal Occult Blood Test (gFOBT), the multitarget stool DNA (mt-

sDNA), and the colon capsule, or invasive tests, which include Colonoscopy, Sigmoidoscopy, and Computed Tomography colonography (CT colonography) (8, 13). Each approach has different sensitivities and recommended screening intervals, depending on the principle of the technique used (Table 1).

2.3.1.1. Non-invasive screening tests

FIT and gFOBT are two non-invasive tests used to detect occult blood in the stools, targeting different analytes (13). Whereas gFOBT is not specific for human stool and it detects pseudo peroxidase activity of heme, FIT only reacts to human hemoglobin using human globin-specific antibodies, not being affected by dietary intake of hemoglobin (13). Both tests can be performed at home and do not require extensive bowel preparation. However, there is a high risk of both false-negative and false-positive results, considering that precancerous lesions and early carcinomas bleed less frequently than advanced lesions, and the test interference with some compounds like aspirin, vitamin C, red meat, and raw vegetables (8, 13). The multitarget stool DNA (mt-sDNA) is a non-invasive test that combines molecular assays to detect methylation profiles and tumour mutations in DNA with an immunochemical assay for human hemoglobin detection in stool samples. Despite mt-sDNA has higher sensitivity than FIT, once it combines two different types of analyses it is also more costly and technically complex (14). Colon capsule is a disposable camera, pill-sized, that is swallowed and activated in the colon. This technic takes pictures of all colon mucosa without being necessary bowel insufflation, radiation or sedation. Despite the practicability of this device, bowel preparation is required (15).

Non-invasive tests do not allow the removal of hypothetical lesions at the moment of the analyses, and thus, a positive result requires subsequent colonoscopy for the removal of lesions (8).

2.3.1.2. Invasive screening tests

Colonoscopy is the most recommended test in colorectal cancer screening and is also the first option to confirm a positive result from a non-invasive test due to its high sensitivity and specificity (8, 15). Moreover, this exam allows the examination of the full colon (proximal and distal) and the removal of precancerous and cancerous lesions during the exam (8). Nonetheless, colonoscopy is an invasive technique that requires sedation (in most cases) and a previous unpleasant bowel preparation where patients need to adopt a liquid diet to ensure tests' accuracy. Apart from the risks of bleeding upon polypectomy, there is also the risk of bowel perforation (15). Flexible sigmoidoscopy is also an invasive

test that, as in colonoscopy, allows the removal of the polyps using a flexible endoscope (13). In opposite to colonoscopy, sigmoidoscopy provides only visualization of the distal colon, for which sedation is not required and bowel preparation does not involve diet restrictions (13). One of the most disadvantages of this test is the fact that it cannot detect lesions in the proximal colon, leading to a decrease in sensitivity when compared to traditional colonoscopy (16). CT colonography is a radiologic exam that provides two and three-dimensional images of the colon, allowing the identification of lesions in the full colon. Despite CT colonography being preferred by patients, this is less sensitive than colonoscopy and, for better visualization, colon inflation with air is required. Moreover, if lesions are detected a colonoscopy needs to be performed. The main advantages of the CT colonography include the no need of sedation and the null risk of bowel perforation (13, 17).

Table 1. Sensitivity and recommended screening interval of each screening test used in colorectal cancer [adapted from (Simon K, 2016; Shaukat and Levin, 2022) (8, 15)].

Test	Sensitivity	Screening interval
FIT	60% - 85%	Annually
gFOBT	33% - 75%	Annually
mt-sDNA	92%	3 years
Colon Capsule	88%	5 years
Colonoscopy	>95%	5 years beginnign at 50 years old
Sigmoidoscopy	>95%	5 years combined with gFOBT
CT colonography	>90%	5 years

2.3.1.3. Other screening tests under development

Despite the variability of the screening tests already in use and described above, new tests are being developed, namely blood- and imaging-based tests. The blood-based tests are more attractive for patients once are less invasive and can be performed at the same time as the other routine blood tests. New blood techniques are focused in the detection of cell-free DNA (cfDNA), circulating tumour DNA (ctDNA), or methylated ctDNA, despite the low levels of these DNAs in circulation. In addition, at the imaging level, some techniques under development focus in the colon capsule, with the aim of improving its technology in order to make it easier to administer to the patients and enable clinical results to be analysed remotely. Other two screening tests are the CT capsule, an X-ray imaging

capsule that does not need bowel preparation and emits low-dose X-ray, and the MRI colonography, which allows the detection of colon polyps, extracolonic findings, and cancer metastasis without using ionizing radiation (15).

2.3.2. Diagnosis

As mentioned before, colonoscopy is the most common screening test for CRC (13). Nevertheless, 15 to 30% of the patients already have metastatic disease at diagnosis, and 20-50% of the patients that are diagnosed with a localized tumour will develop metastasis, mainly in the liver, lung, peritoneum, and lymph nodes (18).

In order to select the best therapy, the metastatic status of the disease needs to be confirmed by radiologic imaging. Additionally, histologic analysis of the primary tumour and/or metastasis by an experienced pathologist and biomarker analysis are fundamental for proper disease staging and management. The biomarker analyses may include the mismatch repair status, the mutations status for *KRAS*, *NRAS*, and *BRAF* oncogenes, HER2 amplification, Dihydropyrimidine Dehydrogenase (DPD) deficiency, and genomic fusions (18).

Tumour Staging

Tumour staging has a huge importance in both diagnosis and therapeutic decision. The most recommended staging system is the Tumour-Node-Metastasis (TNM) system of the American Joint Committee on Cancer (AJCC) and of the International Union Against Cancer (UICC) (19, 20). In this system, “T” defines the extent of invasion of the tumour into the wall of the intestine, “N” encodes the level of invasion of the lymph nodes by cancer cells, and “M” represents the presence or not of distant metastasis, whereby the most common organs for metastasis are the lungs and liver (Table 2). This tumour classification ends into a general stage definition – stages 0, I (the least advanced), II, III, and IV (metastatic stage) (Table 3) (Figure 4) (21).

Table 2. TNM classification for CRC, according to the AJCC and UICC [adapted from (Centelles, 2012) (21)].

T (Local tumour invasion)	
Tx	No available information about local tumour invasion
Tis	Tumour <i>in situ</i> . Restricted to mucosa
T1	Tumour invades submucosa
T2	Tumour invades muscularis propria
T3	Tumours grows through the muscularis propria and invade the peripheral layers of the colon or rectum
T4	Tumour directly invades serosa and other tissues/ organs

N (Regional lymph nodes)	
Nx	No available information about lymph node involvement
N0	No lymph node involvement
N1	Tumour cells in 1 to 3 regional lymph nodes
N2	Tumour cells in 4 or more regional lymph nodes

M (Distant metastasis)	
Mx	No available information about distant metastasis
M0	No distant metastasis
M1	Distant metastasis

Table 3. Staging groups of CRC according to the AJCC and UICC [adapted from (Centelles, 2012) (21)].

Stage 0	Tis, N0, M0
Stage I	T1-T2, N0, M0
Stage II	T3, N0, M0 T4, N0, M0
Stage III	Any T, N1, M0 Any T, N2, M0
Stage IV	Any T, Any N, M1

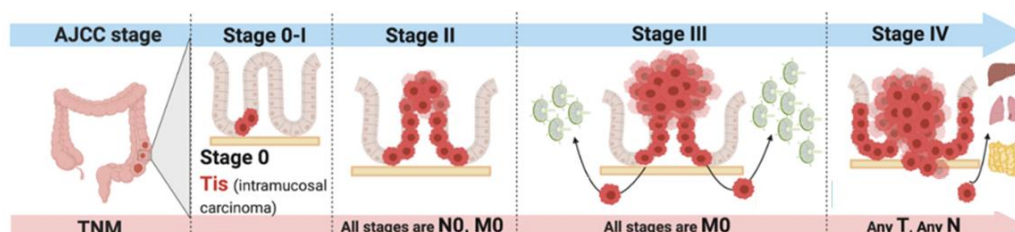


Figure 4. Representation of CRC stages, considering TNM classification, according AJCC [adapted from (Shek et al., 2021) (22)]

Beyond tumour staging, it is also important to define the grade of the cancer, which corresponds to the histological similarity between tumour and normal colorectal tissue. The scale used ranges from G1, where cancer looks like normal tissue (low grade), to G4 where the cancer looks very abnormal (high grade). High grade (G3 and G4) cancers tend to grow faster than low grade (G1 and G2) cancers (21).

To support tumour staging, several modalities of radiologic imaging technics are used namely, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). CT is the most commonly used for diagnosis and staging since it allows to access the location and extent of the primary tumour, the presence of metastasis, and the state of local and distant lymph nodes. MRI is largely used to detect hepatic metastasis but also to perform a primary tumour evaluation, conversely to PET, that is mostly used in the detection of occult disease and is also capable to detect alterations in glucose, being helpful to evaluate changes in tumour metabolism (23).

2.4. Pathophysiology of colorectal cancer

As previously mentioned, many environmental and genetic factors cause alterations in colon epithelial cells. In fact, progressive accumulation of genetic and epigenetic alterations in oncogenes and tumour suppressor genes leads to cancer development, and ultimately, to increased tumour aggressiveness (12). Although colorectal carcinomas are extremely heterogeneous, they can be classified as sporadic or inherited, depending on the nature of the identified mutation, and the underlying genetic or epigenetic alterations (Figure 5) (9, 24).

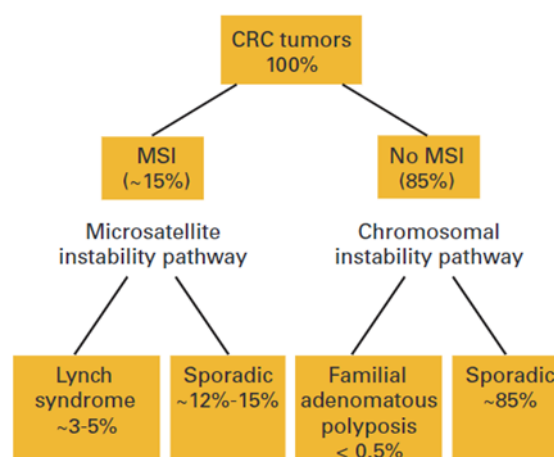


Figure 5. Systematization of colorectal cancer types (25).

Although 10-30% of the colorectal carcinomas are considered to have an hereditary component, only 5% of all colorectal cancer cases have a known genetic cause, and are

included in two main groups: the polyposis syndromes; and the nonpolyposis syndrome – Hereditary non-polyposis colorectal cancer (HNPCC) – also known as Lynch Syndrome (9).

Among the polyposis syndromes, the Familial Adenomatous Polyposis (FAP) is the most frequent inherited syndrome, comprising approximately 1% of the CRC cases. FAP is characterized by the formation of multiple polyps throughout the colon since childhood, which have a high probability of becoming malignant (26). FAP can also exhibit extraintestinal manifestations, notably osteomas, desmoid tumours, dental abnormalities, and other carcinomas, as of thyroid, liver, and central nervous system (27). This is an inherited autosomal dominant syndrome associated with a germline mutation in the *APC* (adenomatous polyposis coli) tumour suppressor gene, which is localized on the long arm of chromosome 5 in band q21 (5q21), divided into 15 exons, and displays an important role in cell adhesion and transcriptional activation. *APC* is involved in the Wnt signalling pathway, regulating β -catenin accumulation in the cytoplasm. In classical FAP, the most aggressive form of the disease, characterized by several hundred polyps, the *APC* mutations occur between exon 5 and the 5' portion of exon 15. In the attenuated FAP (AFAP), manifested with fewer adenomas and in more advanced age, the mutations are concentrated in the extreme 5' portion of the gene, exon 9, and the 3' portion of exon 15 (27, 28). Germline *APC* mutations are found in $\approx 10\%$ of the AFAP patients, and $\approx 7\%$ have a mutation in *MUTYH*.

Other polyposis syndromes include *MUTYH*-associated polyposis (MAP), Peutz–Jeghers syndrome (PJS), Juvenile polyposis syndrome (JPS), PTEN hamartoma tumour syndrome (PHTS), and Serrated polyposis syndrome (SPS), each with specific clinical manifestations and molecular profiles, which overall account for $<1\%$ of all inherited CRCs (26).

HNPCC/Lynch Syndrome is an inherited autosomal dominant syndrome and is the most common cause of hereditary CRC, accounting for 3 to 5% of all CRC cases (29, 30). This syndrome is characterized by germline mutations in DNA mismatch repair genes – *MLH1*, *MSH2*, *MSH6*, and *PMS2* – leading to microsatellite instability (MSI). Besides colorectal carcinomas, these patients also have a high risk to develop other cancer types, namely gynecological (endometrial and ovarian) and digestive cancers (gastric, pancreas, and small bowel) (30). In order to facilitate the diagnosis of Lynch Syndrome the clinical CRC diagnosis is supported by patient's and family history of cancer, the age of cancer diagnosis, and the types of the tumours, using the Amsterdam II criteria and the Bethesda guidelines (Table 4) (31).

Table 4. Amsterdam II Criteria and Bethesda Guidelines used for the diagnosis of Lynch Syndrome [adapted from (Sinicrope, 2018) (31)].

Amsterdam II Criteria	Bethesda Guidelines
≥ 3 relatives with Lynch Syndrome associated cancer*, where one of whom is a first degree relative of the other two	Diagnosis of CRC or endometrial cancer in patients with < 50 years of age
≥ 2 successive generations are affected	Diagnosis of synchronous or metachronous CRC, or other Lynch Syndrome associated cancer at any age
≥ 1 Lynch Syndrome associated cancer cases diagnosed before 50 years	CRC with high MSI in a patient diagnosed before 60 years of age
	≥ 1 first degree relatives diagnosed with CRC or Lynch Syndrome associated cancer* younger than 50 years of age
	≥ 2 first- or second-degree relatives diagnosed with CRC or Lynch Syndrome associated cancer* at any age

*Lynch Syndrome associated cancer includes colorectal, endometrial, stomach, ovary, pancreas, ureter, renal pelvis, biliary tract, brain, small bowel, and sebaceous glands, as well as keratoacanthomas.

Regarding the molecular classification, CRCs can be divided into three different molecular categories: Chromosomal Instability (CIN), Microsatellite Instability (MSI), and cytosine-phosphoguanine (CpG) island methylator phenotype (CIMP) (32).

CIN is characterized by the presence of chromosomal abnormalities, which lead to the loss of tumour suppressor genes and the gain of oncogenes (32). About 80-85% of the CRCs arise via CIN and 70-75% of both adenomas (benign lesion) and sporadic CRCs have a somatic mutation in the *APC* tumour suppressor gene (24, 32). Other mutations such as in *TP53*, *PIK3CA*, *TGF-β*, and constitutively activation of the *KRAS/BRAF* pathway, contribute to the progression of the adenoma to a malignant tumour, which has the capability of invasion (32).

The MSI phenotype, found in approximately 15% of the sporadic colorectal carcinomas, is characterized by insertions or deletions in the short nucleotide repeat sequences – the microsatellites (24, 32). Hypermethylation of the DNA mismatch repair (MMR) gene *MLH1* is the most common cause of MSI in sporadic cancer. MSI is associated

with the CIMP carcinogenesis pathway, where the promoter CpG island of several tumour suppressor genes is methylated and, therefore, silenced (32) (Figure 6).

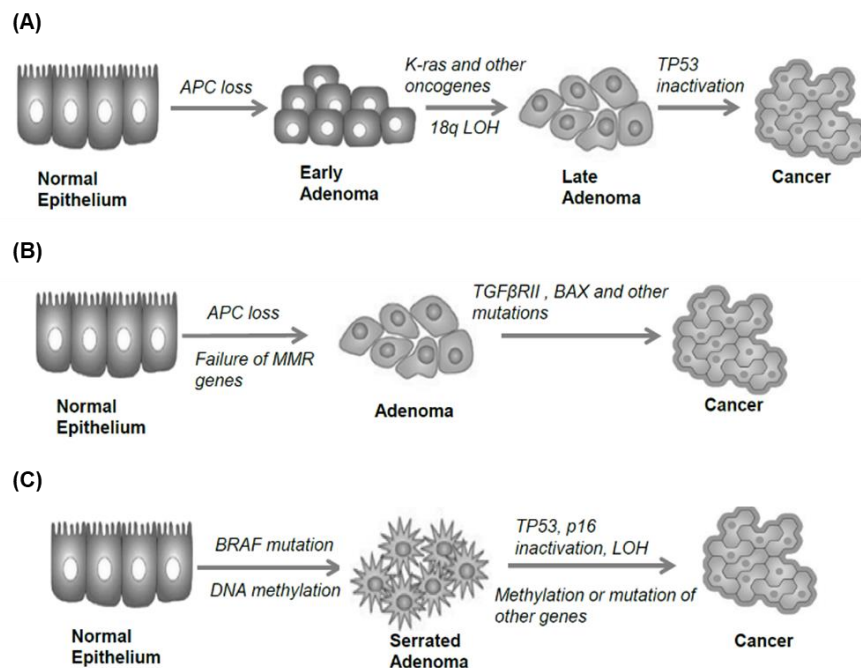


Figure 6. Carcinogenic pathways in colorectal cancer. (A) Chromosomal Instability (CIN) pathway is characterized by inactivating mutations in tumour suppressor genes (*APC*, *TP53*) and activation of oncogenes (*KRAS*). **(B)** Microsatellite Instability (MSI) pathway is mainly characterized by inactivation of the DNA mismatch repair genes, such *MLH1*, leading to alterations in the microsatellite sequences. **(C)** Cytosine-phosphoguanine (CpG) island methylator phenotype (CIMP) pathway is associated with MSI and is characterized by methylation of CpG island of several tumour suppressor genes, being also associated with mutations in the *BRAF* oncogene [adapted from (Mundade et al., 2014) (33)].

3. Treatment

The different treatment approaches for CRC include surgery, chemotherapy, radiotherapy, and targeted therapies, which choice will depend on the stage of the disease and the location of the carcinoma (right or left colon, or rectum). In patients with an early stage (stage 0 or I), surgery can be used exclusively with curative intent, while for patients with stage II or III, the use of adjuvant therapy is recommended, also with a curative intent. For stage IV disease (metastatic) the approach is palliative in the majority of the cases and may include chemotherapy or targeted therapies along with surgery (11, 34). The standard adjuvant therapy for stage III carcinomas consists in the use of 5-Fluorouracil (5-FU) combined with oxaliplatin and leucovorin – named FOLFOX (34). For metastatic patients, the chemotherapy used is usually the FOLFOX or a combination of 5-FU, leucovorin, and

irinotecan – FOLFIRI. The main groups of target therapies include monoclonal antibodies (mAbs) against EGFR, monoclonal antibodies against VEGFR, tyrosine kinase inhibitors, and immune checkpoint inhibitors (12).

Among the monoclonal antibodies, the anti-EGFR – cetuximab and panitumumab – are the most commonly used. However, they are only effective in patients without activating mutations in *KRAS*, *NRAS*, and *BRAF* oncogenes. This mechanism of action will be further discussed (Chapter 5.4). Despite having the same target, these monoclonal antibodies differ in some characteristics. Cetuximab is a mouse/human chimeric mAb and is an IgG1 isotype, while panitumumab is a human mAb and is an IgG2 isotype. Moreover, cetuximab is used in first-line of treatment in combination with FOLFOX, while panitumumab appears in the first and second line in combination with FOLFOX and FOLFIRI, and with FOLFIRI alone, respectively (35).

Among the monoclonal antibodies, bevacizumab (anti-VEGFR) is the mostly used, while regorafenib is the mostly used multi-kinase inhibitor. Tumours with VEGF overexpression comprise about 40-60% of all colorectal cancers, and anti-VEGFR therapy cannot be combined with monoclonal antibodies anti-EGFR (12, 18, 36).

In addition, immunotherapy has also become an alternative treatment approach to CRC, specifically for carcinomas that have high MSI, which associates with high mutational burden, and, consequently, higher tumour immunogenicity. Colorectal carcinomas with these characteristics are mainly metastatic, however, only 4% of mCRC have this profile (37, 38). Currently, the immunotherapies already approved by the Food and Drug Administration (FDA) for mCRC treatment are the immune checkpoint inhibitors pembrolizumab and nivolumab, targeting PD-1, and the CTLA-4 inhibitor ipilimumab, used in combination with nivolumab (37).

4. Prognostic biomarkers

Biomarkers are an important tool in the detection of CRC, analysis of disease progression and recurrence, being also important for predicting therapeutic response. Thus, the TNM staging system, molecular, and histological biomarkers have an important impact on predicting patient prognosis and should be used for treatment decision-making (39).

4.1. TNM

As seen in Chapter 2.3, TNM classification is the most recommended system for CRC tumour staging, where the extent of tumour invasion (T), the state of lymph nodes (N),

and the presence of distant metastasis (M) are analysed and combined for tumour staging. Tumours are classified from stage I to IV, where the 5-years survival declines from 90% to 10%. This is an anatomic-based staging with a huge impact on patient prognosis, allowing treatment management for each patient in order to reduce over or undertreatment (19, 40).

4.2. Molecular Biomarkers

Despite the importance of TNM staging, this classification must be complemented with the molecular information of each tumour, in order to understand the risks and benefits of the therapy applied and, consequently, improve the efficacy of the cancer treatment. Regarding all the treatments described in Chapter 3, the use of 5-FU as part of FOLFOX or FOLFIRI improves overall survival and progression-free survival in more advanced stages of CRC, which is not seen on earlier tumour stages (Stage II) (39). Mutations in the *KRAS* gene were also associated with a worse overall survival and a high probability of recurrence and hepatic metastasis, being related with a poor response to anti-EGFR target therapy (39). In fact, patients with and without mutations in *KRAS* oncogene, both treated with monoclonal antibodies, have shown a median overall survival of 4.5 and 9.5 months, respectively, and a median progression-free survival of 1.8 and 3.7 months, respectively, demonstrating the relevance of mutations in *KRAS* in anti-EGFR therapy efficacy (41). In another study, combination therapy of cetuximab and FOLFIRI in CRCs without *KRAS* mutations has shown to increase the patients' survival by $\approx 16\%$. Concomitantly, the same combined therapy has shown to improve the median progression-free time by $\approx 11\%$ (39, 42). Furthermore, mutations in *BRAF* oncogene, specifically the variant V600E, also affect downstream signalling pathways, being associated with $\approx 50\%$ decreased survival when comparing to non-mutated *BRAF* tumours (39). Some clinical studies have already shown that, in mCRC, combination therapy with anti-EGFR and target therapy for *BRAF*, namely vemurafenib, improved therapeutic efficacy (43). Other markers, such as *VEGF* overexpression and mutations in *PTEN* or *PIK3CA*, are also associated with poorer prognosis and increased metastatic risk, with the last two being associated with cetuximab and panitumumab resistance (36, 42).

Regarding the different carcinogenesis pathways in CRC, CIN and MSI have also different prognosis, with CIN being associated with worse prognosis, and MSI-high with better survival and response to 5-FU. The prognostic value of the CIMP tumours is not yet clear (44).

Considering the most common hereditary cause of CRC – Lynch Syndrome – patients with non-metastatic tumours and deficient mismatch repair genes are associated

with better prognosis due to the eligibility to immunotherapy, which is also applied in the case of metastatic disease (31).

4.3. Histological Biomarkers

Apart from TNM staging and molecular biomarkers, tumour histopathology is also an important feature that should be considered in the moment to assess patient's prognosis (45). Thus, among the tumour stages defined by the TNM system, defining the tumour grade is also important to understand the similarity between the tumour and normal tissue, once the more similar it is (low grade) the slower is tumour progression and the lower is the risk of metastasis, resulting in better patient prognosis. In contrast, a high grade tumour is linked with worse prognosis, due to its higher capability of growing and spreading (21).

Laterality is another important histopathological biomarker that allows the differentiation between right-side colon tumours (ascending colon) and left-side colon tumours (descending colon). In fact, it has already been proved that these two portions of the colon have different prognoses, molecular characteristics, and clinical responses to therapy (46). *Right* colon tumours have more predisposition for deficiencies in MMR genes and mutations in *KRAS* and *BRAF* oncogenes, being the more common origin of CRCs. Otherwise, *left* colon tumours are more associated with CIN, and mutations in *TP53* and *NRAS* genes. Thus, a study showed that laterality is an independent prognostic factor for CRC in all stages of the disease, where patients with tumours from the left colon have a better prognosis and 19% less risk of death compared to tumours originated in the right colon (47). In addition, the National Comprehensive Cancer Network (NCCN) guidelines only recommend anti-EGFR therapy for *left* colon tumours, in which, as shown by Brulé et al. team, treatment with cetuximab is associated with slower tumour progression (median progression time of 5.4 months for *left* colon tumours *versus* 1.9 months for *right* colon tumours) (48, 49).

Additionally, information about metastatic site and the number of metastasis can also be informative for patient prognosis. A study has concluded that patients with lung and liver metastasis have better survival than patients with brain metastasis, with the first group having median overall survival of 16-20 months and 5-years survival rates of 44-90%, while patients with brain metastasis are associated with a median survival of 5 months. Regarding the number of affected organs, patients with a single metastatic site have better survival than those with three or four metastatic sites (median survival of 16, 5, and 2 months, respectively) (50).

5. Epidermal Growth Factor Receptor

5.1. Gene, protein, and receptor

The *EGFR* gene, located in the short arm of chromosome 7 and composed by 28 exons, encodes the Epidermal Growth Factor Receptor (EGFR), a transmembrane receptor which belongs to the ErbB family of receptor tyrosine kinases (RTKs) (51, 52). Apart from EGFR, also known as ErbB1 or HER1, this family includes ErbB2 (HER2), ErbB3 (HER3), and ErbB4 (HER4) which can form homo or heterodimers, leading to the activation of different signalling pathways (51, 53).

EGFR, like the other receptors of the HER family, is composed by an extracellular ligand binding domain with four domains, a hydrophobic transmembrane domain, an intracellular protein tyrosine kinase, and a C-terminal domain, which allows the activation of different signalling pathways (Figure 7) (53, 54).

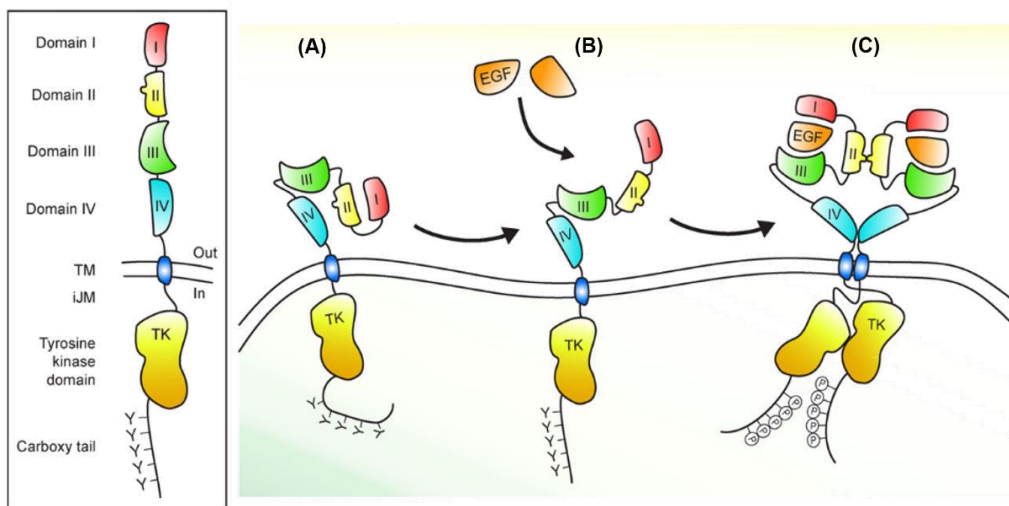


Figure 7. EGFR structure and activation. (A) EGFR is composed by four extracellular domains - I, II, III, and IV - followed by the transmembrane domain (TM), and the intracellular tyrosine kinase domain (TK) and carboxy tail. (B) EGF ligand binds to EGFR extracellular domain (C) and leads to alteration in EGFR conformation allowing the homo or heterodimerization with other family receptors inducing TK autophosphorylation and activation of signalling pathways [adapted from (Sigismund et al., 2018) (51)].

5.2. Main signalling pathways and oncogenic functions

As mentioned before, EGFR activation depends on the extracellular binding of the ligand, such as the epidermal growth factor (EGF), the transforming growth factor- α (TGF- α), the amphiregulin, the heparin-binding epidermal growth factor, or the betacellulin

(54), which trigger several downstream signalling cascades, namely, RAS/RAF/MAPK, PI3K/AKT, and JAK/STAT (Figure 8) (51). Besides the importance of these signalling pathways in normal cell survival, they also have a crucial influence on cancer, leading to cancer cell proliferation, apoptosis inhibition, cell invasion capacity, stimulating tumour neovascularization, and metastasis (51, 52).

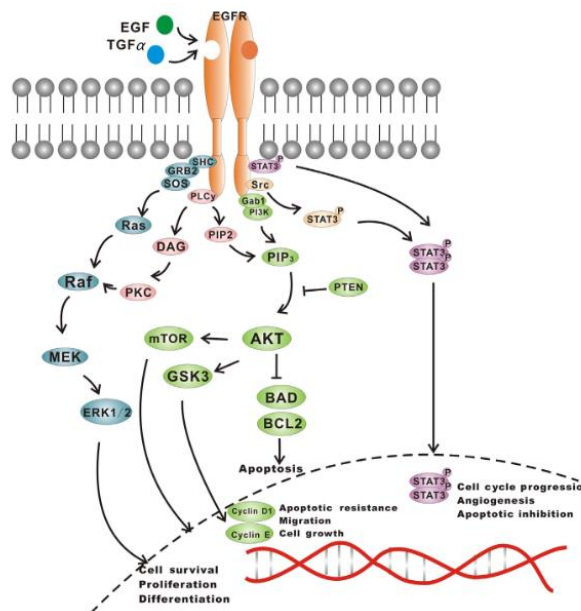


Figure 8. Signalling pathways triggered by EGFR activation. The main intracellular effectors, leading to cell proliferation, inhibition of apoptosis, and angiogenesis, are shown (52).

In cancer cells, EGFR is commonly altered by point mutations, gene amplification, and/or protein overexpression, leading to an increase in the activation of the downstream pathways. Approximately 90% of the mutations occur in exons 19-21 (51, 52), thus affecting the intracellular tyrosine kinase domain.

EGFR overexpression is observed in numerous cancers, namely, bladder, breast, head and neck, kidney, non-small cell lung, and prostate cancer. Specifically in colorectal cancer, nearly 80% of the carcinomas show upregulation of the *EGFR* gene, which associates with increased metastatic risk (54). Apart from the ligand-mediated activation of EGFR-dependent signalling pathways, EGFR has also several “non-canonical” functions mediated by receptor endocytosis, receptor ubiquitination, translocation to mitochondria or ligand-independent receptor activation (51).

Recently, the Cancer Genetics Group of IPO-Porto has unveiled that overexpression of ETV1, a member of the ETS family of transcription factors overexpressed in ~10% of the prostate carcinomas, is sufficient to increase the expression and activation of EGFR and downstream STAT effectors (Azeredo et al, *in preparation*). Additionally, in the same cancer

model, the G-protein coupled receptor GRPR (Gastrin Releasing Peptide Receptor), known to be involved in EGFR transactivation (55, 56), is upregulated by ETV1 overexpression (57, 58). Interestingly, the three members of the PEA3-subfamily of ETS transcription factors, which comprises ETV1, ETV4 and ETV5 (59, 60), have been associated with CRC progression, being the focus of the present study.

5.2.1. The ETS family of transcription factors

5.2.1.1. ETS subfamilies

The ETS is a family of transcription factors encoded by 28 genes that act as activators or repressors of the transcription. The members of this family are characterized by a conserved ETS DNA binding domain composed by 85 amino acids. This domain binds to DNA as a monomer recognizing the GGAA/T sequence in the regulatory regions of target genes. These proteins can be divided into 11 subfamilies, which include the ERG-subfamily composed by ERG, FEV, and FLI1, and the PEA3-subfamily, which comprises ETV1, ETV4, and ETV5 (59, 60).

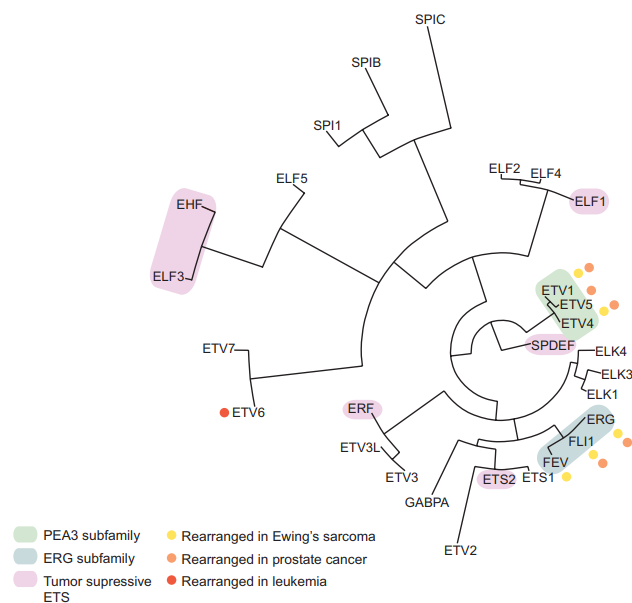


Figure 9. Clustering of ETS subfamilies according to the similarity of DNA binding domain sequence. The yellow, the orange and the red dots represent the most common rearrangements in cancer. The PEA3 subfamily is highlighted in green, and the ERG subfamily in blue. ETS members with known tumour suppressive activity are highlighted in pink [adapted from (Nicholas et al.,2019) (59)].

5.2.1.2. ETS physiologic and oncogenic functions

ETS transcription factors are involved in various physiological processes, including cell cycle control, proliferation, differentiation, angiogenesis, tissue remodelling, and apoptosis, being also related with development of embryonic epithelial tissues, as the lung, kidney, salivary gland, and mammary gland (59, 61, 62). In cancer, ETS members are frequently involved in gene fusions, leading to the expression of chimeric proteins and/or upregulation of ETS proteins. Thus, ETS overexpression is associated with carcinogenesis processes such as tumour initiation, progression, and metastasis (59, 60).

The oncogenic role of ETS members was first described in Ewing Sarcomas, upon the identification of EWS-FLI1 rearrangement in 85-90% of the tumours (60, 63). Later, the EWS-ERG and EWS-FEV rearrangements were identified in the remaining 10-15% of the Ewing Sarcomas. During this period (1983-1997), genomic rearrangements involving other ETS members were identified in leukaemia, albeit rarer (60). In 2005, the identification of ETS rearrangements in 50-60% of the prostate carcinomas (PrCa) opened a new avenue in cancer research (64). Specifically, 50% of the prostate carcinomas show the *TMPRSS2-ERG* translocation involving *TMPRSS2* – the most common 5' partner and an androgen-regulated gene – and *ERG*, leading to overexpression of a truncated form of the ERG transcription factor. Additionally, genes encoding the PEA3 subfamily members, *ETV1*, *ETV4*, and *ETV5*, are also rearranged in PrCa, however in lower frequencies, with *ETV1* rearrangements representing 8-10% of the PrCa and *ETV4/ETV5* rearrangements being found in 2-5% of all prostate tumours. While the implications of *ERG* fusions in PrCa is not yet well established, rearrangements/overexpression of PEA3 members has been consistently associated with worse prognosis (65, 66).

5.2.1.3. ETS in colorectal cancer

In 2019, Cheng et al. described *ETV5* overexpression as a potential biomarker of poor prognosis in CRC, with these tumours having a more advanced TNM stage and a larger size, and being associated with cell proliferation and angiogenesis (67). In the same year, Park et al. described a panel of genes, including *ETV5*, to be associated with resistance to cetuximab in CRC cell lines, and showed that *ETV5* knockdown increased the sensitivity to this mAb (68). More recently, Shen et al. validated the previous observations of Cheng et al. and, additionally, suggested that both *ETV1* and *ETV4* also had a role in CRC, with overexpression of *ETV1* showing a positive correlation with worse prognosis, and overexpression of *ETV4* being associated with better outcome. An association between overexpression of *ETV1* and *ETV5* with immune infiltration levels and metastasis was also

described (69). Interestingly, in 2021, Fonseca et al. described overexpression of *ETV4* in various adenocarcinoma samples, which was associated with increased cell proliferation and migration *in vitro*, suggesting that *ETV4* plays an important role in the primary events underlying colorectal carcinogenesis (70). Additionally, Leng et al. showed that *ETV4* is capable of regulating EGFR transcription, promoting tumour growth (71). All these findings reinforce the association between members of the PEA3 family and EGFR, an important target in mCRC cancer.

5.2.1.4. ETS as therapeutic targets

Considering the oncogenic activity of ETS transcription factors, specifically relevant in Ewing sarcomas, leukaemias, and prostate cancer, several compounds and small molecules have been described to inhibit ETS transcription factors (72), however, since ETS transcription factors are crucial for tissue maintenance and development, the major challenge concerns tissue specificity (59, 73), and, thus, the potential clinical utility of these compounds is still under investigation.

5.2.2. Gastrin-releasing Peptide Receptor

The Gastrin-releasing Peptide (GRP) is a neuropeptide that belongs to the bombesin-like peptides (BLPs) family. BLPs activate several signalling pathways by binding to G-protein coupled receptors (GPCR) (74). GRP is a regulatory peptide responsible for many physiological functions in the brain, gastrointestinal tract, and vascular and endocrine system, being responsible for the regulation of the smooth musculature contraction, the release of gastrointestinal and pancreatic hormones, insulin secretion, and control of cell proliferation. GRP receptor (GRPR) is the most common mediator of GRP signalling in human tissues, and frequently found overexpressed in several carcinomas, including, colon, breast, endometrial, pancreas, brain, lung, and prostate (74, 75).

5.2.2.1. GRPR and cancer

In 1999, Saurin et al. detected higher GRPR mRNA levels in colon carcinoma samples, reporting an association with poor/moderate tumour differentiation and lymphatic vessel invasion (76). These observations were corroborated by Carrol et al., who described overexpression of GRPR or GRP in 84% of the colon carcinomas (77) and by Chave et al., who observed increased expression of both GRPR and GRP in colorectal carcinomas comparing with normal mucosa (78). Despite GRPR/GRP is not normally expressed in normal adult colonic epithelium, they are present in an initial phase of the epithelial cell development (79). Thus, it was concluded that when GRPR is overexpressed in colorectal

cancer, its impact on tumour growth is lower. In accordance with this, results from Glover et al. suggested that GRPR/GRP overexpression may have initial impact in tumour development, maintaining the cancer cells in a better differentiation stage (79, 80). Concordantly, Rivera et al. validated these results by demonstrating that carcinomas with GRPR or GRPR/GRP overexpression are associated with better patients' survival and decreased risk of metastasis (81).

In prostate and ovarian cancer, GRPR was associated with cancer progression and poor prognosis, leading to the development of androgen independence in prostate carcinomas, and conferring angiogenic properties to the tumours in ovarian cancer (82, 83). In prostate cancer, Paulo et al. described that GRPR can be regulated by ETS transcription factors, where ERG and ETV1 are capable to bind to the GRPR promoter, activating its transcription and overexpression (57).

5.2.2.2. GRPR and EGFR association

In 1994, Liebow et al. validated previous studies that described GRP as being responsible for the upregulation of EGF, leading to EGFR amplification. Thus, the study suggested an association between GRP/GRPR phosphorylation and increased EGFR signalling (84). In 2003, a study in head and neck squamous cell carcinomas (HNSCC), showed that GRP/GRPR is responsible for EGFR phosphorylation, leading to the activation of signalling pathways important for tumour development. This EGFR transactivation by GRPR, leading to downstream activation of the MAPK signalling pathway and cell proliferation, was also later described in non-small cell lung cancer (NSCLC) (55, 56).

5.2.2.3. GRPR as a therapeutic target

Considering the membrane localization of GRPR and its overexpression in multiple carcinomas, several approaches have been developed to inhibit this receptor, including, receptor agonists and antagonists, monoclonal antibodies (mAbs), vaccines, antisense oligonucleotides (ASOs), and bispecific molecules (BsMo) (74, 85).

GRPR agonists and antagonists can be classified according to their mode of action. Whereas GRPR agonists induce receptor internalization, GRPR antagonists block cell signalling without being internalized (75). GRPR agonists were the first-generation of bombesin inhibitors, showing good performance against cancer cells. However, aggressive side effects in the gastrointestinal were reported, leading to the development of the GRPR antagonists (second-generation) (86). Although the processes involved in tumour inhibition

by the second-generation of bombesin inhibitors are not well established, some of them seem to involve reduction of EGFR and inhibition of neovascularization (75, 87, 88).

Considering the GRPR/EGFR relation previously described, the potential efficacy of a combination therapy using an EGFR inhibitor and a GRPR antagonist has been suggested (56, 89). In colorectal cancer, a recent study described a new GRPR-targeting peptide that showed a high uptake by tumour cells, allowing early tumour detection, imaging guidance for GRPR-targeted therapy, and monitoring of cancer progression (90).

5.3. EGFR inhibitors and mechanisms of resistance in colorectal cancer

There are two strategies for EGFR blocking – anti-EGFR monoclonal antibodies (mAbs) and small-molecule tyrosine kinase inhibitors (TKIs) (54). TKIs interfere with ATP-binding in the tyrosine kinase receptor inhibiting the autophosphorylation of the receptor. Besides its use in mCRC some of these compounds, namely gefitinib and erlotinib, have been also approved by the FDA for the treatment of advanced non-small cell lung cancer and pancreatic cancer (42).

The mAbs, cetuximab and panitumumab were already approved by FDA for routine clinical use for the treatment of mCRC and squamous cell carcinomas of the head and neck. They bind to the extracellular domain of EGFR blocking ligand binding, and, therefore, inhibiting downstream pathways activation (Figure 10) (54). However, this therapy is only effective in a restricted group of mCRC patients – named “*wild-type*” – whose tumours do not have activating mutations in *KRAS*, *NRAS*, or *BRAF* oncogenes (91). The reason for the restriction in patient selection for this anti-EGFR therapy is related to the mutational status in these genes, which can lead to constitutive activation of the downstream signalling pathways independently of EGFR activation, and thus lead to drug resistance (92).

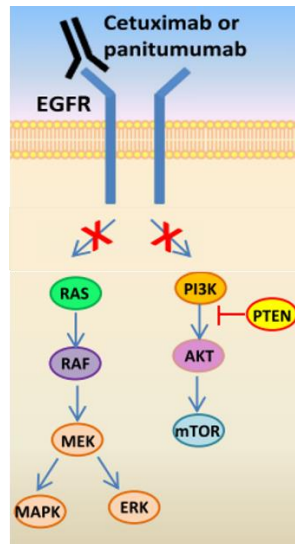


Figure 10. Schematic representation of the intracellular signalling inhibition using the monoclonal antibodies cetuximab or panitumumab. The anti-EGFR antibodies prevent ligand-induced activation of EGFR downstream signalling pathways, impairing tumour aggressiveness [adapted from (Leto & Trusolino, 2014) (93)].

Nevertheless, despite the initial efficacy of anti-EGFR therapy, approximately 80% of the *wild-type* patients develop resistance to monoclonal antibodies, by not yet completely understood underlying mechanisms (94, 95). Some studies suggest that this resistance could be acquired after treatment, while others defend that it can be innate, resulting from the expansion of a pre-existing cell clone (95). Acquired resistance can arise from selective pressure induced by monoclonal antibodies, from amplification of other ErbB family tyrosine-kinase receptors, as HER2, which will lead to activation of the same downstream pathways as EGFR (96), from alterations in the EGFR ectodomain, and from changes in processes such as angiogenesis, EGFR ubiquitination, and a higher liberation of cytokines and anti-apoptotic growth factors (93, 95, 97, 98). Recently, characteristics of the tumour microenvironment, namely, activated cancer-associated fibroblasts (CAFs) and an epithelial-mesenchymal transition (EMT) phenotype have been associated with poor response to target therapy (95). Overexpression of the EGFR receptor is found in 25-82% of all CRCs and has been suggested as a possible mechanism of resistance to target therapy (36). Some of these mechanisms of resistance are illustrated in Figure 11.

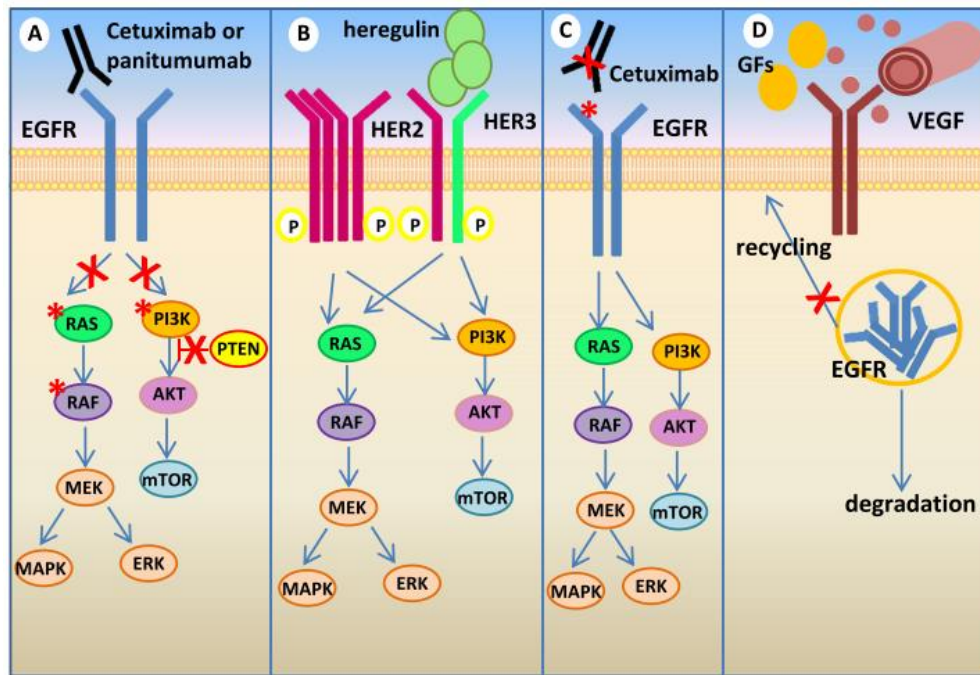


Figure 11. Mechanisms of resistance to anti-EGFR therapy. (A) Activating mutations in oncogenes such as *RAS*, *RAF*, or *PI3K* can lead to signalling pathway activation despite EGFR inhibition. (B) Activation and amplification of other EGFR family tyrosine kinase receptors, such as HER2 and HER3, that activate the same oncogenic pathways as EGFR. (C) Mutations in EGFR ectodomain inhibits mAbs binding, decreasing EGFR inactivation and increasing downstream pathways activation. (D) Alterations in angiogenesis increase the secretion of VEGF anti-apoptotic factors and EGFR ubiquitination and degradation, which cause modifications in the expression of the receptor (93).

AIMS

The main goal of this project is to understand if, in metastatic colorectal cancer, overexpression of PEA3 transcription factors, *GRPR* and/or *EGFR* have prognostic value for patients eligible for anti-EGFR therapy. In order to achieve this main goal, some specific points will be addressed, namely:

- The frequency of *KRAS/NRAS/BRAF wild-type* metastatic colorectal carcinomas with overexpression of *ETV1*, *ETV4*, *ETV5*, *GRPR*, and/or *EGFR*;
- The correlation between the expression profile of *ETV1*, *ETV4*, *ETV5*, *GRPR*, and *EGFR* and clinical response to anti-EGFR therapy, namely, their association with overall survival and progression-free survival;
- The association between that expression profile and clinicopathological characteristics of the patients.

MATERIALS AND METHODS

1. Formalin-Fixed Paraffin-Embedded tissue samples

1.1. Patients' cohort validation and characterization

For this project, we gathered 487 FFPE (Formalin-Fixed Paraffin-Embedded) cancer tissue sections from *wild-type* metastatic colorectal carcinomas (negative for mutations in *KRAS*, *NRAS*, and *BRAF* - clearly stated in the medical report) characterized at the Department of Genetics of IPO Porto between 2013 and 2018.

All the patients' data were kept confidential, in accordance with the General Data Protection Regulation. This project was approved by the institutional review board (Comissão de Ética para a Saúde – CES 51/023) of the Portuguese Oncology Institute of Porto (IPO Porto).

Table 5. Clinicopathological characteristics of the mCRC patients involved in this study with available clinical data at IPO Porto, treated with anti-EGFR therapy, and with molecular information.

Characteristics	N=111
Gender	
Male	73
Female	38
Age at diagnosis (years)	
≤50	18
51-69	64
≥70	29
Period of diagnosis (years)	
2006-2012	28
2013-2015	45
2016-2018	38
Tumour grade	
I	5
II	57
III	15
IV	2
Unknown	32
Surgery	
Yes	102
No	9
Metastasis site	
Hepatic	76
No Hepatic	34
No metastasis	1
Number of metastasis	
0	1
1	59
2	36
3	13
4 or more	2
Laterality	
Right	26
Left	85
Treatment	
Cetuximab	2
Cetuximab + FOLFIRI	83
Cetuximab + FOLFOX	3
Cetuximab + Irinotecan	12
Cetuximab + LV5FU	1
Panitumumab	6
Panitumumab + FOLFIRI	2
Panitumumab + FOLFOX	2

2. RNA extraction and cDNA synthesis

FFPE tissue sections were deparaffinized and hydrated, and RNA was extracted using the High Pure FFPE RNA Isolation Kit (Roche, Basel, Switzerland), following the manufacturer's instructions. For deparaffinization, tissues were incubated in 100% DL-Limonene for 2 minutes, twice, and finally 2 minutes in 100% Ethanol.

For RNA extraction, tissue samples were incubated overnight with 100 µL of Tumour Lysis Buffer (TLB), 16 µL of SDS, and 120 µL of Proteinase K, to digest tissue proteins and allow access to nuclear acids. On the next day, 325 µL of Reagent Preparation Buffer (RPB) and 325 µL of 100% Ethanol were added, in order to precipitate the nucleic acids in the solution. Then, the sample was transferred to a silica column for RNA isolation and purification and a centrifugation for 30 seconds at 6.000 xg was performed. To digest single and double-stranded DNA, 100 µL of diluted DNase I (1:10) was added to the column and incubated for 15 minutes. After that, the column was washed with 500 µL of Wash Buffer I, followed by two washes with 500 µL of Wash Buffer II. Between washes, the centrifugations were performed for 30 seconds at 6.000 xg and the final centrifugation was for 2 minutes at 16.000 xg to dry the column. Finally, the RNA was eluted with 30 µL of nuclease-free water and concentration was measured using the NanoDrop™ Spectrophotometer (ThermoFisher Scientific™, Waltham, MA, USA).

RNA was converted into cDNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher Scientific™, Waltham, MA, USA), according to the manufacturer's instructions.

Briefly, 100 ng of RNA were converted into cDNA using 1 µL of random hexamer primer (0.2 µg/µL) in a 20 µL buffered reaction containing 1X Reaction Buffer, 20 U of RiboLock RNase Inhibitor, 1 mM of dNTP Mix, and 200 U of RevertAid H Minus M-MuLV Reverse Transcriptase. First, the samples were incubated with the random hexamers in a thermocycler at 65°C for 5 minutes, to denature the samples and provide the primer binding. After an incubation on ice for 2 minutes, the remaining reagents were added and the cDNA synthesis proceeded with an incubation of 5 minutes at 25°C, followed by 60 minutes at 42°C, and, finally, by enzyme denaturation at 70°C for 5 minutes.

3. Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

The qRT-PCR allows accurate quantification of a specific target region in a sample by measuring the fluorescence released during each cycle of amplification, using a fluorescent dye or a probe. This technic uses the cycle threshold (C_t), that is inversely

proportional to the initial amount of the target, and corresponds to the cycle where the fluorescence level is higher than the background, defined by the baseline. In the exponential phase, duplication of the target sequences occurs until the reaction reaches the plateau phase, where product amplification stops.

The expression levels of *ETV1*, *ETV4*, *ETV5*, *GRPR*, and *EGFR* were accessed by qRT-PCR using Taqman primer/probe assays for each gene. *ACTB* was used as housekeeping gene to normalize variations between samples, such as pipetting errors, and correct for differences in the target expression levels. *ACTB* was previously selected as the best housekeeping gene for our samples, in comparison with *GAPDH* and *GUSB* (Figure A-1, Appendix 1). For *ETV1*, *ETV4*, and *GRPR*, primer/probe assays were designed using Primer Blast (58) (Table 6), for *EGFR*, *ETV5*, and *ACTB*, we used pre-developed assays from Integrated DNA Technologies (IDT, Coralville, Iowa, USA), namely, the assays Hs.PT.58.39601461 (113 bp), Hs.PT.58.879535 (94 bp), and Hs.PT.39a.22214847 (110 bp), respectively.

Table 6. Sequences of the primers and TaqMan probes designed to assess the expression levels of *ETV1*, *ETV4*, and *GRPR*. The amplicon size is also represented.

	Primer Forward	Primer Reverse	Probe	Amplicon size
ETV1	5' - TTT TAT GAC CAG CAA GTG CCT TAC - 3'	5' - TAG CAT TTG TTG GTT TCT CGT TACA - 3'	5' - TCT CCC ACG CTG ACT ATT GGT GAC CA - 3'	79 bp
ETV4	5' - GGA ATG GAG TTC AAG CTC ATT GA - 3'	5' - TAA TTC ATG GCT GGC CGG TTCT - 3'	5' - CCA GAG CCT GGC GAC CTC CTC AG - 3'	80 bp
GRPR	5' - TCC AGA GTG CTT ACA ATC TTC C - 3'	5' - CAC AAA CAC CAG CAC TGT CTT G - 3'	5' - AGT CGC TTC CGG GAT TCA ATC TGC T - 3'	136 bp

The TaqMan probes have a FAM fluorophore attached at the 5'-end and a TAMRA quencher dye at the 3'-end. The principle of TaqMan probes is explained in the Figure 12.

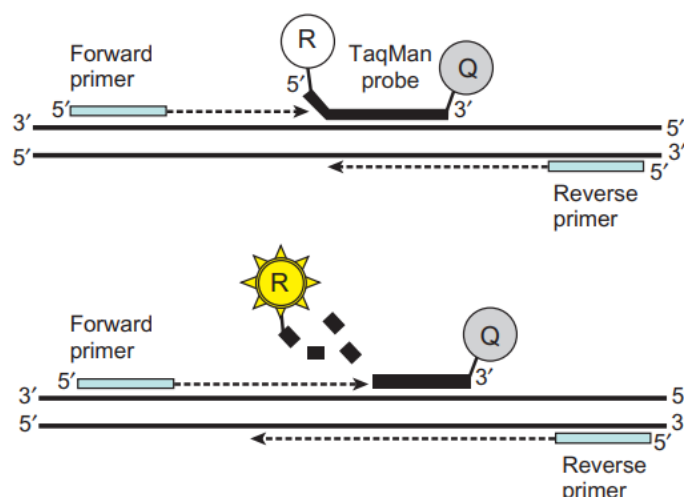


Figure 12. Principle of TaqMan probes. The proximity of the quencher dye to the fluorophore represses the fluorescence emitted by the fluorophore. After the annealing of the primers, the Taq DNA Polymerase initiates the extension phase adding nucleotides to synthesize the complementary DNA strand. When it reaches the annealed probe, its exonuclease activity cleaves the probe allowing the emission of fluorescence by the fluorophore. The fluorescence is measured, and the initial copy number is inversely proportional to the cycle when fluorescence is higher than the background, crossing the threshold in the amplification plot [adapted from (Butler, 2012) (99)] .

The qPCR reactions were optimized to run in 384-well plates (Bio-Rad, Hercules, CA, USA). The samples, blanks, and standards were performed in quadruplicate, in a final reaction volume of 10 μ L, which contains 1 μ L of cDNA and 9 μ L of the reaction mix composed by 5 μ L Xpert Fast Probe, 2X master mix (Grisp, Porto, Portugal), 375 nM of each primer and 250 nM of the TaqMan probe. Reactions took place in the LightCycler® 480 II (Roche, Basel, Switzerland) under the following conditions: 2 minutes at 50°C, followed by a pre-incubation of 10 minutes at 95°C, and 45 cycles of amplification, each including a denaturation step of 15 seconds at 90°C and an annealing and amplification step of 1 minute at 60°C. Finally, the reaction ends with an extension cycle of 5 minutes at 72°C.

In order to evaluate the amplification efficiency of each target, a standard curve with the dilutions 1:10; 1:100; 1:500; 1:1000 was constructed, using cDNA from the metastatic prostate cancer cell line PC3, known to be a positive control for the expression of all the targets under study.

3.1. Data analysis

The expression levels of each target gene (*ETV1*, *ETV4*, *ETV5*, *GRPR*, and *EGFR*) was normalized to the housekeeping gene (*ACTB*), using the comparative threshold cycle

(C_t) method. To obtain the C_t , the threshold was manually adjusted for each target, considering the background fluorescence of the “blank” reactions, using the Fit Points method. C_t values were exported to an Excel spreadsheet and gene expression data was obtained upon normalization for *ACTB* using the formula $2^{-\Delta C_t}$, where the number 2 corresponds to exponential duplication of the target regions in each cycle of the PCR reaction and ΔC_t is the difference between the C_t values of the target gene and *ACTB*.

For a better understanding of the criteria used to obtain the final sample cohort subjected to statistical analyses (N=111) the follow diagram was constructed (Figure 13).

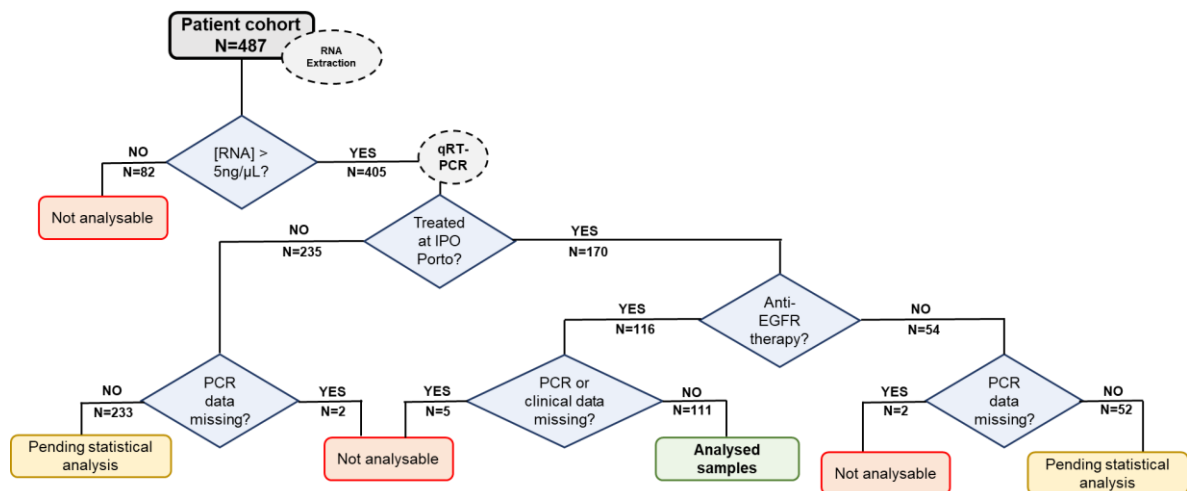


Figure 13. Diagram representing the samples available for analysis in each step of the study, from the initial patient cohort (N=487) until the final patient cohort (N=111). The red boxes represent the samples that are not analysable due to insufficient RNA or lack of clinical information. The yellow boxes represent the samples that can be analysed in the future (N=285), and the green box represents the final patients' cohort that followed statistical analyses in the present study (N=111).

3.2. Quality control

Since both the FFPE procedure and the time in storage can affect RNA integrity, an amplification result was considered reliable when at least two of the four replicates carried out for each sample had C_t deviations lower than 1. Additionally, we tested qPCR results' credibility in all analysed samples (N=396) by plotting the time in storage (the days passed between the date of tumour collection and the date of RNA extraction), with the standard deviation obtained for *ACTB* levels (C_t) for each analysable sample (Figure A-2, Appendix 1).

4. Statistical analyses

4.1. Statistical Package for the Social Sciences (SPSS)

Considering the expression values of each gene, a first division of the samples was performed considering two different *cut-offs* – Quartiles and Outlier/Extreme values – using IBM SPSS Statistics 28.0.1.0. Using “quartiles” *cut-off* - from now on called “Q3” *cut-off* -, samples were grouped in three expression ranges: “<Q1” comprise samples with the lowest expression levels (below 1st quartile); “Q1-Q3” includes samples with medium expression levels (Interquartile Range - IQR); and “>Q3” is composed by samples with the highest expression levels (above 3rd quartile). Using the “Outliers” *cut-off*, samples were grouped in the three expression ranges: the “Outlier” group, that comprises samples with an expression value higher than Q3 by $1.5 \times \text{IQR}$; the “non-outlier” group, that contains samples with expression levels ranging from Q1 to that below “outliers”; and the “negative” group, which includes samples with expression levels <Q1.

Kaplan-Meier curves with Log-rank and Generalized Wilcoxon test were used to analyse the impact of the molecular profile in OS and PFS. A *p*-value lower than 0.05, corresponding to a confidence interval $\geq 95\%$, was considered statistically significant.

4.2. Survival R package

Using the statistical software R (version 4.2.1), we defined a third *cut-off* – the optimal *cut-off* – which splits the samples in low and high expression groups, for each gene. The optimal *cut-off* value corresponds to the *cut-off* that better differentiates sample groups considering the endpoint value (either Overall Survival – OS, or Progression Free Survival – PFS), using the Log-rank statistical test. OS was calculated considering the time difference between the onset of metastasis and last follow-up or date of death. PFS was considered the time difference between the first anti-EGFR treatment and disease progression (clinical, biochemical, and/or imagological) or death. Thus, two “optimal *cut-offs*” were defined, one for OS (Optimal-OS) and one for PFS (Optimal-PFS), and samples were grouped in “High” and “Low” expression groups.

Multivariate Cox proportional hazards regression analysis was performed to investigate the independence of the genes between them and the influence of the clinical features, namely, age, surgery, site of metastasis, and tumour origin (laterality), in patient survival and tumour progression considering the molecular subgroups. For these, samples with low gene expression, no surgery performed, hepatic metastasis, and tumours from the right side of the colon, were taken as reference groups. For age, the risk was calculated considering the increase in age with one-year increments. For *Hazar Ratio* interpretation,

an $HR=1$ indicates no differences between the groups, and an $HR<1$ or $HR>1$ reveals a lower or a higher risk of death/progression than the reference group, respectively.

Kaplan-Meier curves with Log-rank were used to analyse the impact of the molecular profile in OS and PFS. All 111 patients had the clinical information necessary for analysis of OS. Three cases had no information regarding date of progression, and, thus, 108 cases were analysed for PFS.

For all statistical analysis, a p -value lower than 0.05, was considered statistically significant.

RESULTS

As there are no guidelines for which expression *cut-off* to use, for the purpose of this thesis, the results obtained with the Optimal *cut-offs* identified with the *Survival R package* will be presented and discussed in detail. Comparison with “Q3” and “Outliers” *cut-offs* performed in SPSS will be mentioned if already performed and shown in the *Appendix* chapter.

Additionally, since the sample distribution considering the “Outliers” *cut-off* is similar when comparing the 111 samples analysed in detail in this study and the full series of 396 samples with analysable expression data, we consider the 111 samples to be an adequate representation of the “molecular subgroups” found in the full patient series (Figure A-4, *Appendix 2*).

Analyses related with *GRPR* expression, and considering that only 7 samples showed detectable expression levels, the “Outliers” *cut-off* was used. For simplification, *GRPR* data will be shown and discussed in separate subchapters.

1. Expression profile of PEA3 transcription factors and *EGFR* in *wild-type* tumours from mCRC patients highlights different “molecular subgroups”

To understand the influence of the different expression *cut-offs* in the distribution of samples according to their molecular profile, we compared the frequency of cases with overexpression of each gene under study considering the different *cut-offs*, namely, “Q3”, “Outliers”, Optimal-OS and Optimal-PFS (Figure 14). We can see that, comparing with the “Q3” and the “Outliers” *cut-offs*, the Optimal *cut-offs* lead to a greater number of samples to be considered with high expression of *ETV1* or *ETV4*, while for *ETV5*, a more restricted number of samples are considered to have high expression using the Optimal *cut-offs*. For *EGFR*, the Optimal-OS *cut-off* results in the same separation of samples as the “Outliers” *cut-off*, while the Optimal-PFS *cut-off* leads to an increase in the number of samples to be considered with high expression.

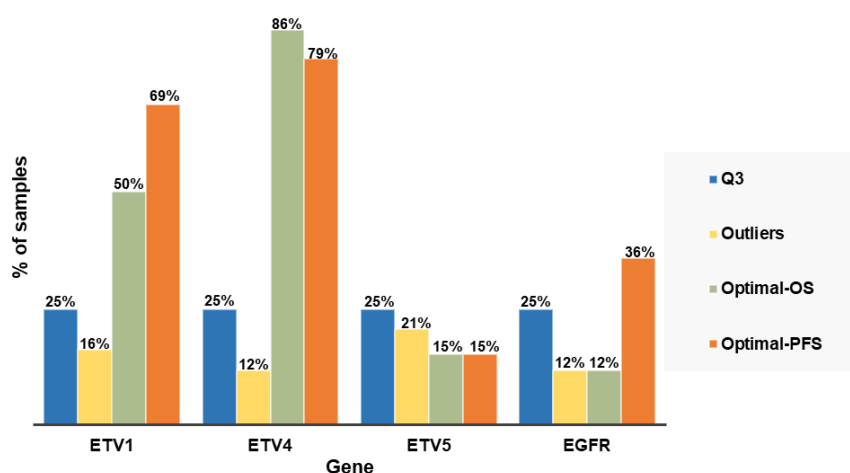


Figure 14. Percentage of samples with overexpression of *ETV1*, *ETV4*, *ETV5*, or *EGFR* considering the different expression *cut-offs*.

Then, we investigated the possible “molecular subgroups” considering the most prevalent combinations of gene overexpression, using the Optimal *cut-offs*. The molecular subgroups obtained with the “Q3” and the “Outliers” *cut-offs* are represented in *Appendix 2* (Figures A-5 and A-6).

Considering OS, and starting by the isolated expression of each gene (Figure 15), the majority of the samples have *ETV4* as the single overexpressing gene (33%), 4% have single overexpression of *ETV1*, 1% of *ETV5* and 1% of *EGFR*. Looking into samples with concomitant expression of two or more genes, the most represented “molecular subgroups” overexpress *ETV1* and *ETV4*, 30% without any other gene overexpressed, 8% of the samples with concomitant overexpression of *ETV5* (PEA3 group) and 6% with concomitant overexpression of *EGFR*. The 4th and 5th more common “molecular subgroups” include *ETV4* overexpressed with *ETV5* (4%) or with *EGFR* (4%).

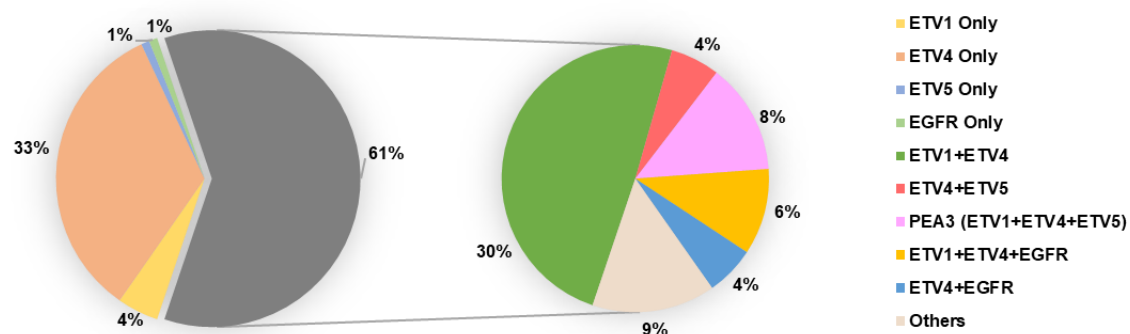


Figure 15. “Molecular subgroups” considering the Optimal-OS *cut-off*. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, *ETV5*, or *EGFR* - and samples with overexpression of two or more genes (*right*). The “Others” group includes both negative samples for all the genes under study (N=8) and gene combinations found in less than four cases (N=2).

Considering the “molecular subgroups” for PFS, as in OS, most of the samples have overexpression of *ETV4* only (16%). Overexpression of *ETV1* alone is found in 6% of the samples, followed by high expression of *EGFR* only (3%), and none of the samples overexpresses *ETV5* only (Figure 16). Additionally, the “molecular subgroup” with overexpression of both *ETV1* and *ETV4* remains the most common (30%), followed by the “molecular subgroups” with simultaneous overexpression of *ETV4* and *EGFR*, with 14% having also overexpression of *ETV1*, and 7% having no other overexpressing gene. The PEA3 “molecular subgroup” also represents 7% of the samples. The 5th and 6th more common “molecular subgroups” include *ETV1* overexpressed with *EGFR* (6%) or with *ETV4*, *ETV5* and *EGFR* (4%).

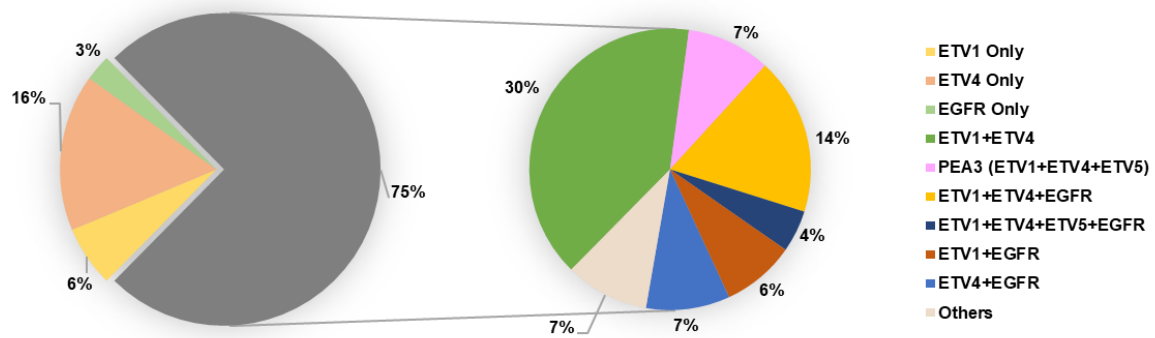


Figure 16. “Molecular subgroups” considering the Optimal-PFS cut-off. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, or *EGFR* - and samples with overexpression of two or more genes (*right*). The “Others” group includes both negative samples for all the genes under study (N=3) and gene combinations found in less than four cases (N=5).

2. Overexpression of PEA3 transcription factors or *EGFR* associates with OS and/or PFS

Having established the most prevalent “molecular subgroups”, the next aim was to correlate them with patients’ medical records. We started by analysing the influence of each gene independently of the expression of the others.

2.1. *EGFR*

Starting by investigating the impact of *EGFR* in OS, high expression of *EGFR* is associated with better prognosis ($p=0.040$, Figure 17), and a decreased risk of death by 49% ($HR=0.51$; CI: 0.26-0.98, $p=0.043$), comparing with patients with low gene expression.

Comparatively, looking to PFS, high *EGFR* expression is not associated with patient outcome ($p=0.47$, Figure 17), where low *EGFR* expression only increase in 17% the risk of progression ($HR=1.17$; CI: 0.76-1.80, $p=0.468$).

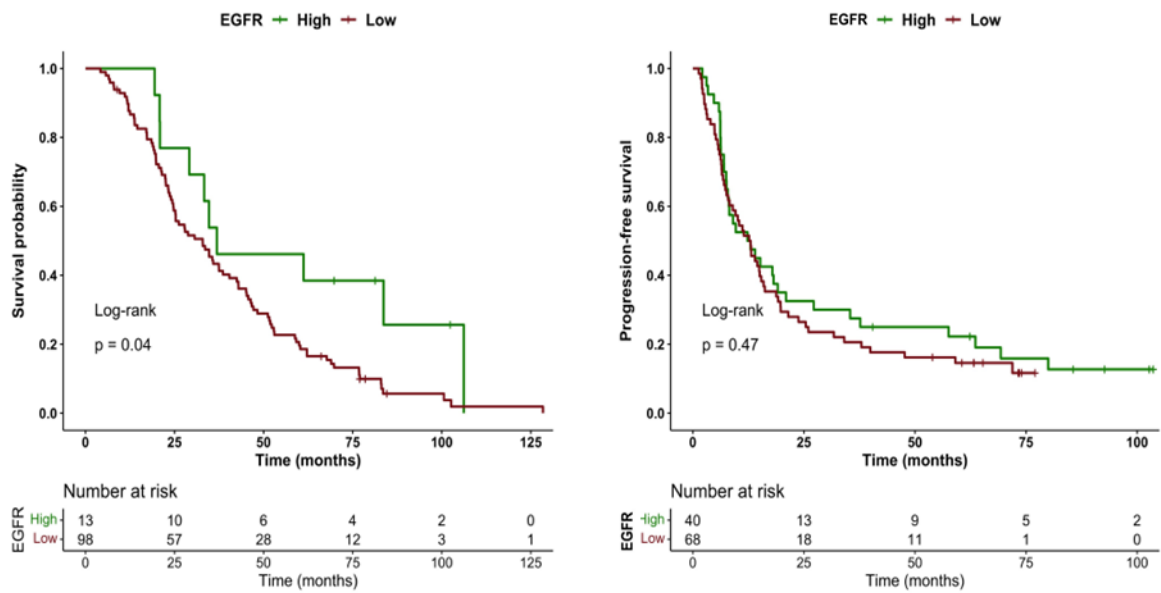


Figure 17. Kaplan-Meier curves of OS and PFS stratifying patients for *EGFR* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *EGFR* expression. The p -value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.

2.2. *ETV5*

Regarding high and low *ETV5* overexpression, Kaplan-Meier curves show an association between patients with high *ETV5* gene expression and worse OS ($p=0.015$, Figure 18), with an estimated 1.91-fold increased risk of death ($HR=1.91$; CI: 1.13-3.25, $p=0.016$).

Similarly, high *ETV5* expression shows a tendency for an association with worse PFS ($p=0.048$, Figure 18), with low expressing tumours having a decreased risk of progression by 42% ($HR=0.58$; CI: 0.34-1.00, $p=0.051$).

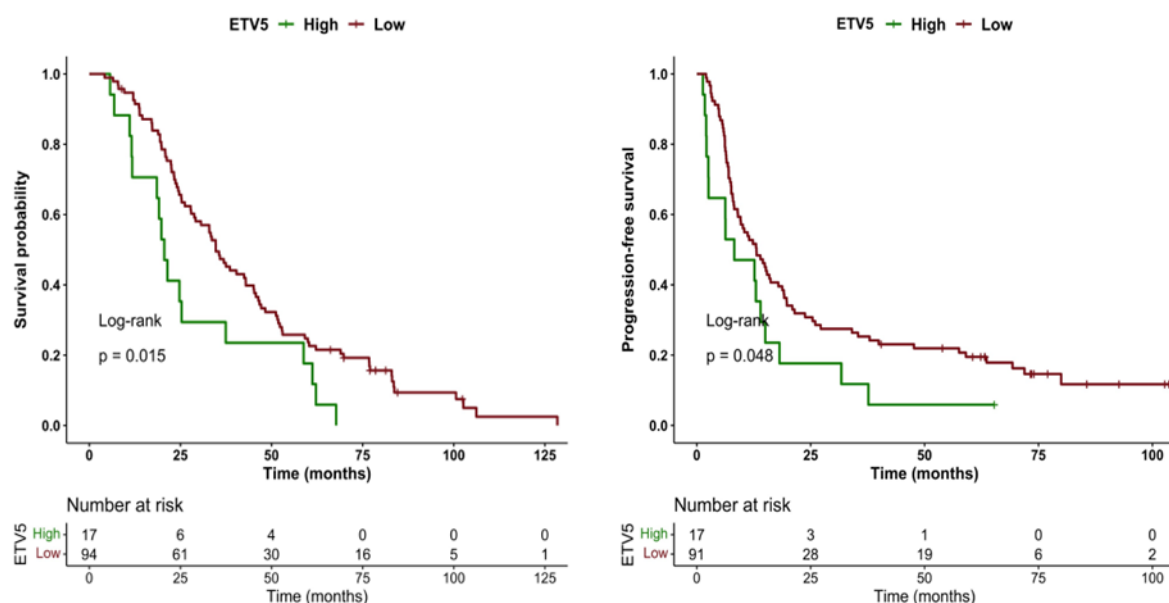


Figure 18. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV5* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV5* expression. The p -value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.

2.3. *ETV1*

Regarding *ETV1*, in OS, a high expression of *ETV1* seems to be associated with worse outcome, even though the analysis does not reach statistical significance ($p=0.089$, Figure 19). This tendency is complemented by the *hazard ratio* which shows that *ETV1* overexpression associates with a 42% increased risk of death ($HR=1.42$; CI: 0.95-2.12, $p=0.091$).

Investigating the influence of *ETV1* expression levels in PFS (Figure 19) there is a clear association between high expression of *ETV1* and worse PFS, compared with low *ETV1* expression ($p=0.018$), which is confirmed by a 1.75-fold decreased risk of tumour progression in patients with low *ETV1* expression ($HR=0.57$; CI: 0.36-0.91, $p=0.019$).

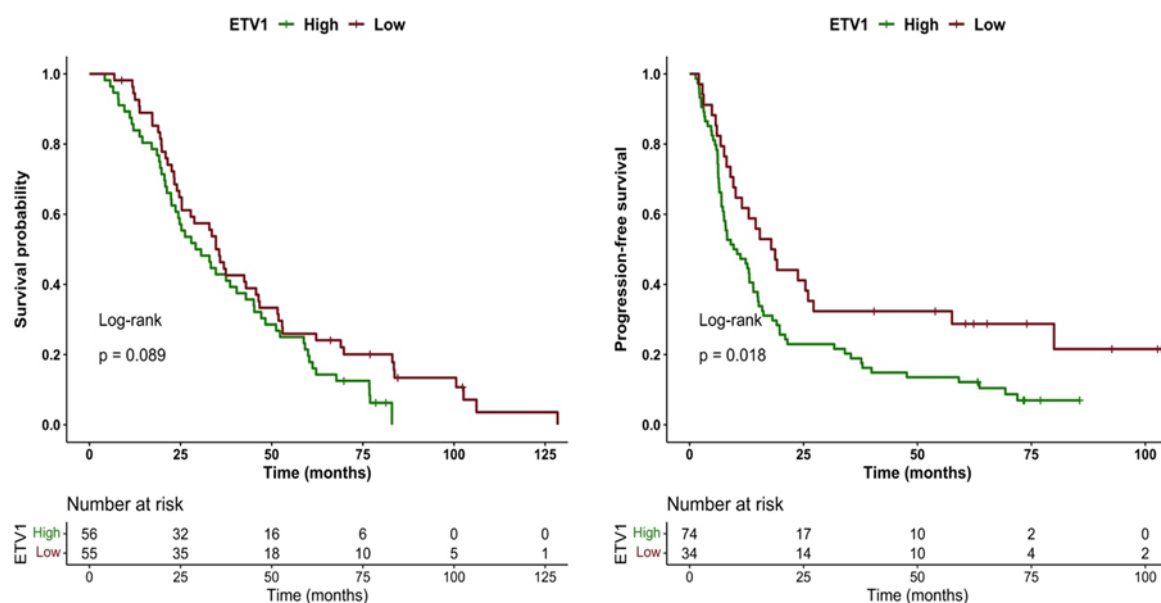


Figure 19. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV1* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV1* expression. The p -value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.

2.4. *ETV4*

Focusing the attention on *ETV4*, high *ETV4* expression shows a tendency for an association with better OS (Figure 20), despite the lack of significance in Log-rank statistical test ($p=0.12$). This observation is supported by a reduced risk of death by 35% for patients with high *ETV4* expressing tumours, compared with those with low tumour gene expression ($HR=0.65$; CI: 0.37-1.13, $p=0.126$).

Regarding the influence of *ETV4* in disease progression, high expression of *ETV4* seems to associate with better PFS (Figure 20), when compared with patients with low gene expression ($p=0.019$). This association is supported by a 1.8-fold increase in the risk of tumour progression in patients with low *ETV4* expression when compared to those with high gene expression ($HR=1.80$; CI: 1.10-2.97, $p=0.020$).

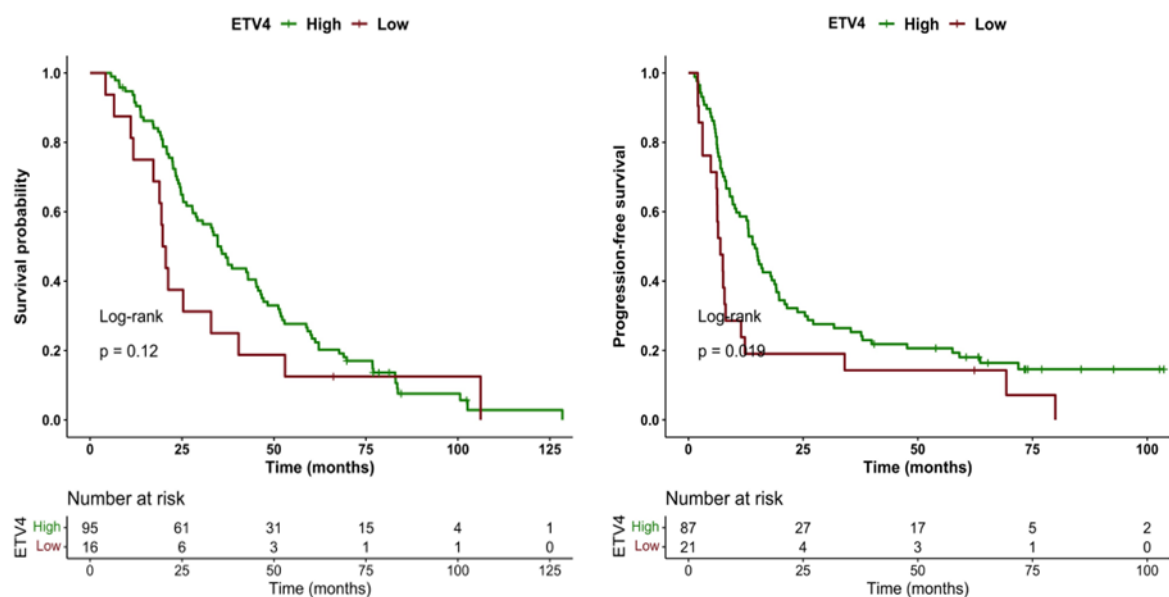


Figure 20. Kaplan-Meier curves of OS and PFS stratifying patients for *ETV4* expression. Survival probability over 125 months (*left*), and progression-free probability over 100 months (*right*) for patients with high and low *ETV4* expression. The p -value of Log-rank statistical test and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.

3. Prognostic impact of the major “molecular subgroups”

In order to understand the impact that the co-expressed genes have in patients’ prognosis we also investigated the influence of the major “molecular subgroups” identified, considering the number of patients available in each group for the analysis.

3.1. Overall Survival

First, we investigate the influence of the ETS groups in patients’ overall survival, and observed that tumours with *ETV4* overexpression alone (*ETV4* Only) are associated with better OS (Figure 21) and a lower risk of death by 73% ($HR=0.27$; CI: 0.09-0.76, $p=0.014$) when compared with tumours “*ETV1* Only”. On the other hand, compared with patients that express only *ETV4*, the simultaneous expression of *ETV4* and *ETV1* associates with better prognosis while the simultaneous expression of *ETV4* and *ETV5* associates with worse prognosis ($p=0.0074$, Figure 21B).

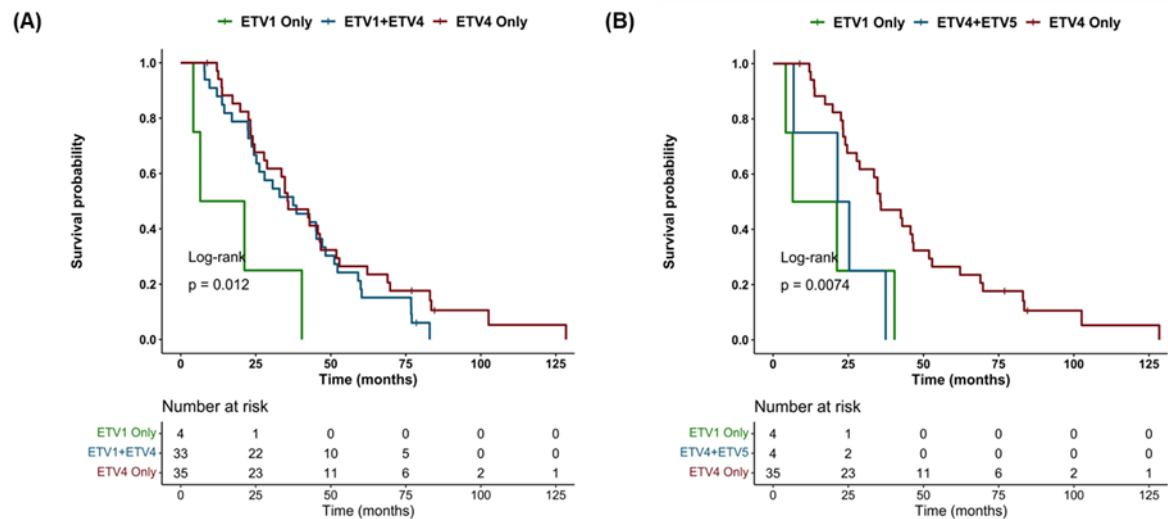


Figure 21. Kaplan-Meier curve of OS stratifying patients for PEA3 “molecular subgroups”. Survival probability over 125 months of patients with “*ETV1* Only”, “*ETV4* Only” and **(A)** “*ETV1+ETV4*”, or **(B)** “*ETV4+ETV5*” tumours. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

Moreover, considering the previous associations between *EGFR* or *ETV4* overexpression with better outcome, and between *ETV1* or *ETV5* overexpression with worse outcome, we evaluated the impact of extreme “molecular subgroups” in OS, comparing tumours with high expression of either *ETV1* or *ETV5* and low expression of either *ETV4* or *EGFR* with tumours “*ETV4* Only” (Figure 22). Highly significant differences were observed between these groups, with both Kaplan-Maier survival curves ($p < 0.0001$) and hazard ratio showing that “*ETV4* Only” tumours associate with a 5.26-fold lower risk of death ($HR = 0.19$; CI: 0.08-0.47, $p < 0.001$) comparing to the extreme subgroup (*ETV1/ETV5* High + *ETV4/EGFR* Low).

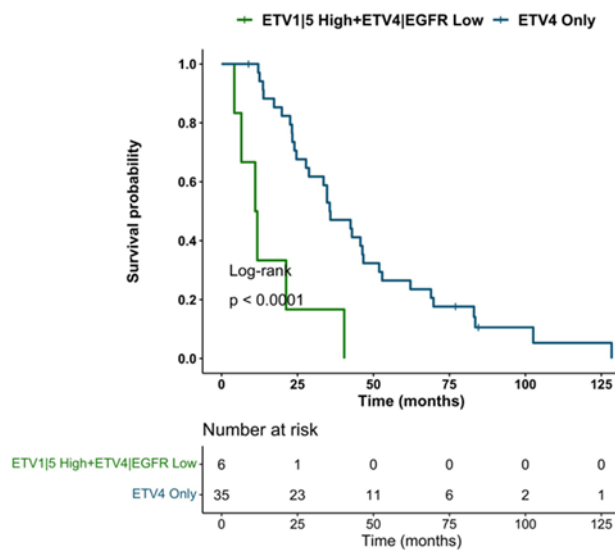


Figure 22. Kaplan-Meier curve of OS comparing patients with extreme ETS “molecular subgroups”. Survival probability over 125 months of patients with “ETV4 Only” tumours and tumours with overexpression of *ETV1* or *ETV5* and low expression of *ETV4* or *EGFR*. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

Next, we aimed to investigate whether different “molecular subgroups” with or without *EGFR* overexpression could associate with OS. For that purpose, and considering the most prevalent *EGFR* overexpressing groups, we compared the four different molecular subgroups combining *ETV1* and *EGFR* expression (Figure 23). Despite the lack of statistically significant differences between the four groups ($p=0.057$), tumours with overexpression of *EGFR* alone (*ETV1* Low + *EGFR* High) show a tendency for better OS, when comparing with the other “molecular subgroups”.

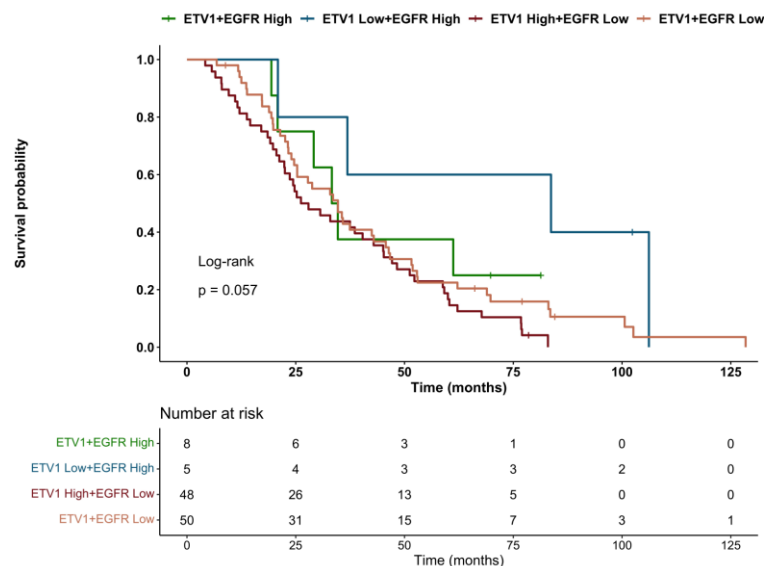


Figure 23. Kaplan Kaplan-Meier curve of OS comparing the different *ETV1/EGFR* “molecular subgroups”. Survival probability is shown over 125 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

Table 7 summarises the impact of the most significant “molecular subgroups” in OS, reinforcing the associations between *ETV4* or *EGFR* overexpression and a better OS, with median survival times of ≈ 36 months in both subgroups, and 5-years survival rates of 26% and 46%, respectively. Conversely, the associations with worse OS obtained for *ETV1* or *ETV5* overexpression are reflected in the median survival times – ≈ 14 months for *ETV1* and ≈ 21 months for *ETV5* – and in the 5-years survival rates of 0% and 17%, respectively. The positive influence in OS of concomitant *ETV4* overexpression in tumours with *ETV1* overexpression is also supported by the association with the higher median survival rates (≈ 37 months) and 5-years survival rates (18%) in comparison with tumours “*ETV1* Only”.

Table 7. Summary of the 5-years survival rate (%) and median survival rate (months) associated with the most important “molecular subgroups” identified.

	5-years survival (%)	Median survival rate (months)
<i>ETV1</i> Only	0%	13.88
<i>ETV4</i> Only	26%	35.71
<i>ETV1</i> + <i>ETV4</i>	18%	37.47
<i>ETV5</i> High	17%	20.57
<i>EGFR</i> High	46%	36.87

3.2. Progression Free Survival

Analysing the influence of the most relevant “molecular subgroups” in PFS, we observed, that, as for OS, the isolated expression of *ETV1* (*ETV1* Only) is associated with worse PFS comparing with patients with either “*ETV4* Only” tumours or tumours with concomitant overexpression of *ETV1* and *ETV4* ($p=0.001$, Figure 24). Specifically, “*ETV4* Only” tumours are associated with a 75% lower risk of disease progression ($HR=0.25$; CI: 0.08-0.77, $p=0.016$) compared with tumours “*ETV1* Only”, while the concomitant expression of *ETV1* and *ETV4* associates with 80% lower risk ($HR=0.20$; CI: 0.07-0.59, $p=0.004$) comparing with the same “molecular subgroup”.

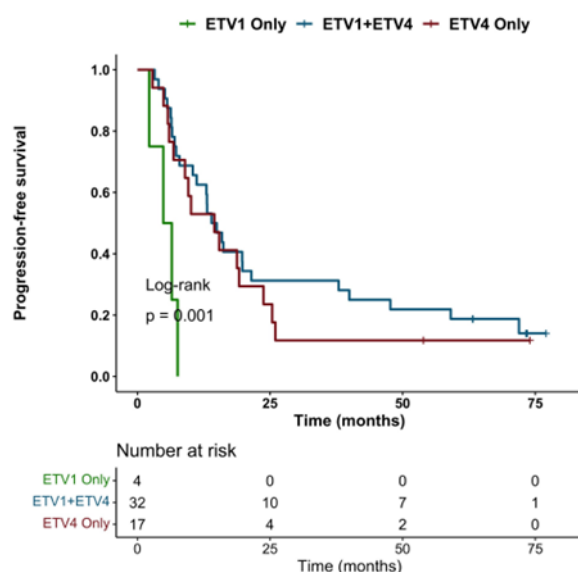


Figure 24. Kaplan-Meier curve of PFS stratifying patients for *ETV1/ETV4* “molecular subgroups”. Progression-free probability is shown over 75 months. The p -value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

Due to the low number of samples with concomitant overexpression of *ETV4* and *ETV5* using the Optimal-PFS *cut-off*, the influence in PFS of *ETV5* overexpression in a *ETV4*-positive background was assessed by comparing PEA3 tumours (with *ETV1+ETV4+ETV5* overexpression) with “*ETV1+ETV4*” tumours. Again, the negative association between *ETV5* overexpression and worse patients’ outcome was validated by the association with worse PFS of PEA3 overexpressing tumours ($p=0.02$, Figure 25) and by the 2.56-fold increased risk of disease progression ($HR=2.56$; CI: 1.05-6.24, $p=0.039$).

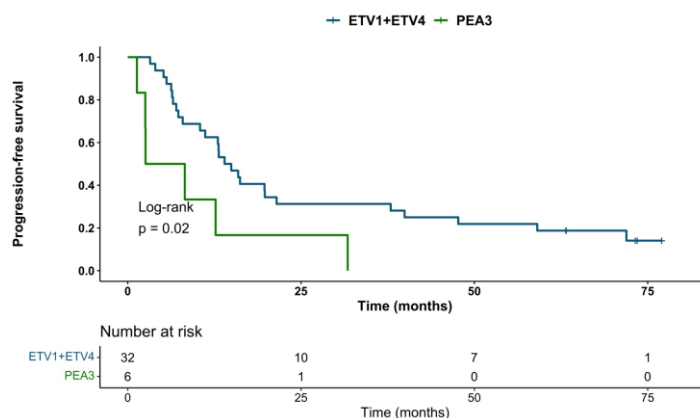


Figure 25. Kaplan-Meier curve of PFS stratifying patients for PEA3 “molecular subgroups” with or without *ETV5*. Progression-free probability is shown over 75 months. The p -value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

When evaluating the impact of *extreme* “molecular subgroups” in PFS, highly significant differences were, again, observed with both Kaplan-Maier survival curves ($p=0.0038$, Figure 26) and *hazard ratio* analysis, in which “*ETV4* Only” tumours associate with a 76% lower risk of death ($HR=0.24$; CI: 0.08-0.74, $p=0.013$) comparing to the *extreme* subgroup (*ETV1/ETV5* High + *ETV4/EGFR* Low).

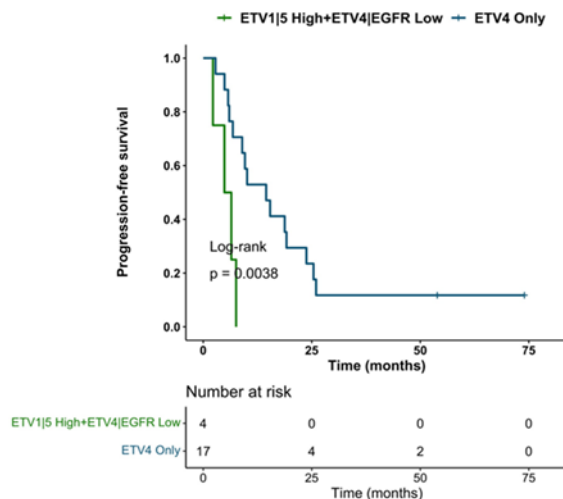


Figure 26. Kaplan-Meier curve of PFS comparing patients with *extreme* ETS “molecular subgroups”. Progression-free probability over 75 months of patients with “*ETV4* Only” tumours and tumours with overexpression of *ETV1* or *ETV5* and low expression of *ETV4* or *EGFR*. The p -value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

Analysing the influence of *EGFR* in PFS, we observed that overexpression of *EGFR* in a low *ETV1* background associates with better PFS ($p=0.018$) in comparison with the survival curves obtained for patients with *ETV1*-overexpressing tumours (with or without co-overexpression of *EGFR*) or with tumours with low expression of both genes (Figure 27). In fact, comparing with “*ETV1+EGFR* High” tumours, tumours with high *EGFR* expression and low *ETV1* expression are associated with a 73% lower risk of tumour progression ($HR=0.27$; CI: 0.12-0.64, $p=0.003$).

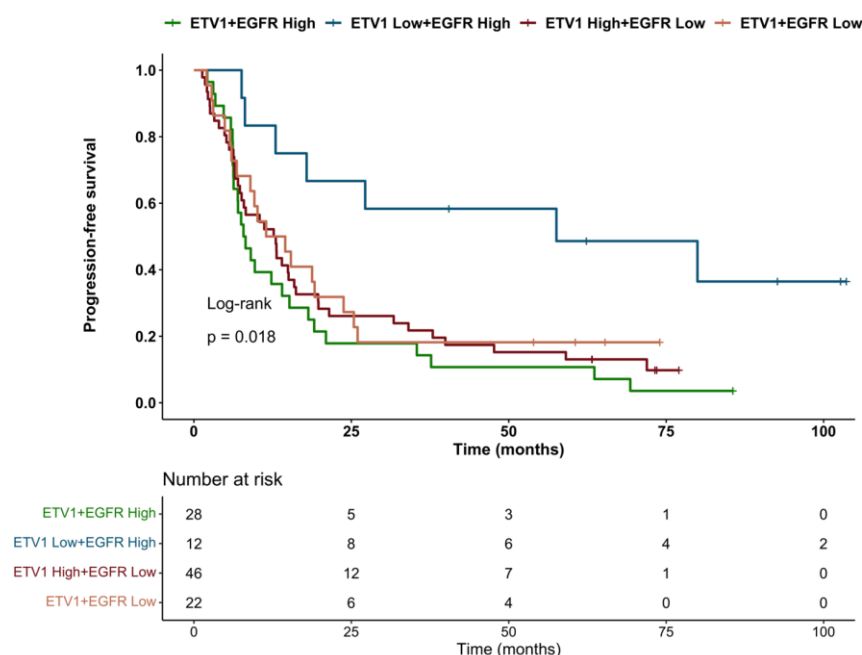


Figure 27. Kaplan-Meier curve of PFS comparing the different *ETV1/EGFR* “molecular subgroups”. Progression-free probability is shown over 100 months. The *p*-value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

Table 8 summarizes the impact of the main “molecular subgroups” in PFS. Concordant with the observation described above, among PEA3 “molecular subgroups”, the association between *ETV4* overexpressing tumours and lower PFS is supported by the ≈14.5 months of median progression rate estimated for patients with either “*ETV4* Only” or “*ETV1+ETV4*” tumours, comparing with the ≈6-8 months estimated for patients with “*ETV1* Only” or “*ETV5* High” tumours. Additionally, the low anti-EGFR therapeutic efficacy of these “molecular subgroups” is also supported by the fact that none of the patients harbouring those tumours fulfilled 5-years after first anti-EGFR therapy without evidencing tumour progression, in comparison with the 10-20% 5-year PFS rates observed in patients harbouring “*ETV4* Only”, “*ETV1+ETV4*” or “*ETV1+EGFR*” tumours. On the other hand, these “molecular subgroups” are associated with faster disease progression when comparing with the 48% of cases with tumours “*EGFR* High+*ETV1* Low” (*EGFR-ETV1*) who survived 5 years without disease progression (median progression of ≈60 months).

Table 8. Summary of the 5-years progression-free survival rate (%) and median progression rate (months) associated with the most important “molecular subgroups” identified.

	5-years progression-free survival (%)	Median progression rate (months)
<i>ETV1</i> Only	0%	5.67
<i>ETV4</i> Only	11%	14.50
<i>ETV1</i> + <i>ETV4</i>	18%	14.45
<i>ETV5</i> High	0%	8.23
<i>EGFR</i> + <i>ETV1</i>	10%	18.4
<i>EGFR</i> - <i>ETV1</i>	48%	59.24

3.3. Predictive value of PEA3 and *EGFR* expression levels

To clarify if the influence of the expression levels of each gene has independent predictive value for OS or PFS or is affected by the expression levels of the other genes, we performed a Multivariate Cox analysis for both outcomes under study (Figure 28). In OS, despite the association with increased risk, *ETV1* overexpression showed no independent predictive value ($p=0.102$). On the other hand, *ETV4*, *ETV5*, and *EGFR* showed to have independent predictive value for OS (Figure 28A), with *ETV4* and *EGFR* overexpression being independent predictors of better OS ($p=0.019$ and $p=0.020$, respectively) and *ETV5* of worse OS ($p=0.043$). On the other hand, *EGFR* expression is not an independent predictor of PFS ($p=0.186$), but the expression levels of each of the three PEA3 transcription factors has independent predictive value for PFS (Figure 28B), with *ETV1* or *ETV5* overexpression being independent predictors of worse PFS ($p=0.037$ and $p=0.037$, respectively), and *ETV4* overexpression being an independent predictor of better PFS ($p=0.004$).

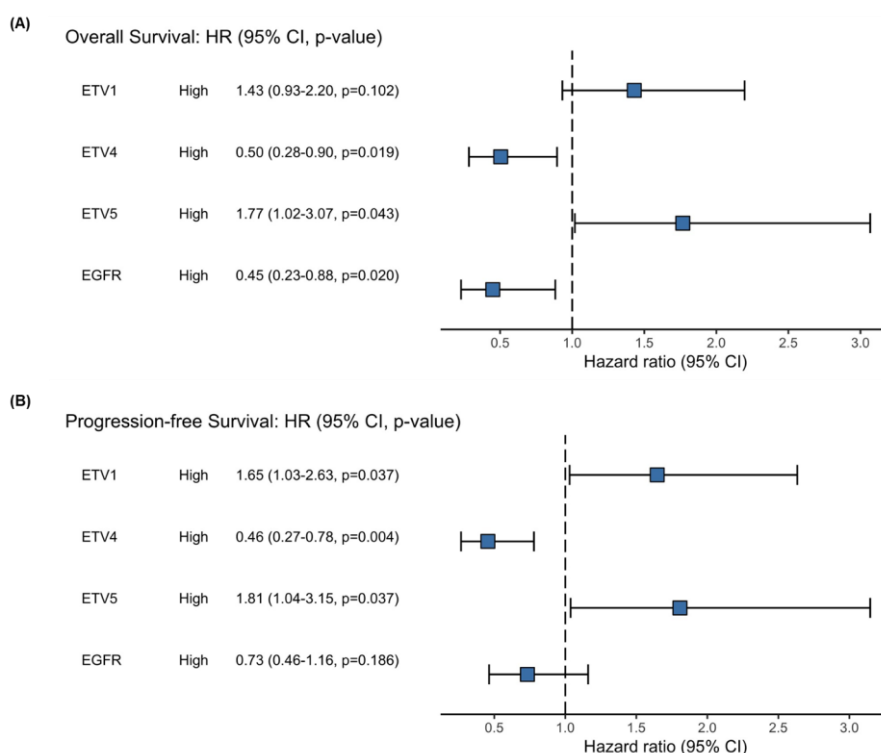


Figure 28. Multivariate Cox proportional hazards regression considering expression levels of *ETV1*, *ETV4*, *ETV5*, and *EGFR* for OS and PFS. For each outcome, the *Hazar Ratio (HR)*, the confidence interval (CI), and the *p*-value are shown for each gene.

4. Influence of clinicopathological characteristics in the association between PEA3 and *EGFR* “molecular subgroups” with OS and PFS

Considering the associations between the main “molecular subgroups” in OS and PFS, we wanted to clarify whether these have prognostic value independently of the clinicopathological features (covariates), namely surgery, age, metastatic site, and tumour laterality. For this purpose, we used Multivariate Cox proportional hazards regression analyses for both OS and PFS.

We started by performing univariate analysis in order to understand the impact of each clinical feature in OS and PFS, as well as to validate the results obtained for the associations between *ETV1*, *ETV4*, *ETV5* and *EGFR* expression levels and patients’ outcome (Table A-1, *Appendix 4*). Overexpression of *ETV5* and *EGFR* remain associated with worse patients’ OS ($p=0.016$ and $p=0.043$, respectively), while *ETV1* and *ETV4* remain associated with worse and better patients’ PFS ($p=0.019$ and $p=0.020$, respectively). Regarding clinicopathological features, age ($p=0.022$) and tumour laterality ($p<0.001$) are associated with OS, where the risk of worse OS increases 3% with each year of age ($HR=1.03$; CI: 1.00-1.05), and tumours originated from the left colon associate with better

OS ($HR=0.42$; CI: 0.27-0.67) (Table A-1.1, *Appendix 4*). Regarding PFS, only laterality shows a significant association with patients' outcome ($p=0.002$), with the tumours originating in left colon being associated with a decreased risk of PFS by 52% ($HR=0.48$; CI: 0.30-0.76) (Table A-1.2, *Appendix 4*).

In multivariate analyses, *ETV5* is the single “molecular marker” with an independent predictive value, being significantly associated with OS ($p=0.022$). In addition, laterality also shown to have independent influence in OS, being associated with patients' survival considering each “molecular marker” ($p=0.006$, $p=0.004$, $p=0.006$, $p=0.007$, for *ETV1*, *ETV4*, *ETV5*, and *EGFR*, respectively) (Table 9). On the other hand, *ETV1* and *ETV4* show significant associations with PFS, being independent from the other clinical features ($p=0.029$ and $p=0.025$, respectively). As in OS, tumour laterality showed association with PFS, independently from the other characteristics, considering each “molecular marker” ($p=0.007$, $p=0.008$, $p=0.015$, $p=0.009$, for *ETV1*, *ETV4*, *ETV5*, and *EGFR*, respectively) (Table 10).

Table 9. Multivariate Cox proportional hazards regression analysis for OS considering each “molecular marker” and the clinicopathological features.

				OS			
		N	HR (Multivariable)			N	HR (Multivariable)
ETV1				ETV4			
	Low	55	-		Low	16	-
	High	56	1.29 ($p=0.248$)		High	95	0.61 ($p=0.104$)
Surgery				Surgery			
	No	9	-		No	9	-
	Yes	97	0.71 ($p=0.404$)		Yes	97	0.64 ($p=0.278$)
Age				Age			
	Mean (SD)	61.7	1.02 ($p=0.088$)		Mean (SD)	61.7	1.03 ($p=0.047$)
Metastasis Site				Metastasis Site			
	Hepatic	76	-		Hepatic	76	-
	No Hepatic	34	1.18 ($p=0.503$)		No Hepatic	34	1.21 ($p=0.444$)
Laterality				Laterality			
	Right	26	-		Right	26	-
	Left	85	0.49 ($p=0.006$)		Left	85	0.47 ($p=0.004$)
ETV5				EGFR			
	Low	94	-		Low	98	-
	High	17	1.96 ($p=0.022$)		High	13	0.55 ($p=0.089$)
Surgery				Surgery			
	No	9	-		No	9	-
	Yes	97	0.68 ($p=0.333$)		Yes	97	0.66 ($p=0.303$)
Age				Age			
	Mean (SD)	61.7	1.02 ($p=0.086$)		Mean (SD)	61.7	1.03 ($p=0.043$)
Metastasis Site				Metastasis Site			
	Hepatic	76	-		Hepatic	76	-
	No Hepatic	34	1.16 ($p=0.556$)		No Hepatic	34	1.15 ($p=0.569$)
Laterality				Laterality			
	Right	26	-		Right	26	-
	Left	85	0.49 ($p=0.006$)		Left	85	0.49 ($p=0.007$)

Table 10. Multivariate Cox proportional hazards regression analysis for PFS considering each “molecular marker” and the clinicopathological features.

PFS							
		N	HR (Multivariable)		N	HR (Multivariable)	
ETV1				ETV4			
	Low	34	-		Low	23	-
	High	77	1.73 (<i>p</i> = 0.029)		High	88	0.54 (<i>p</i> = 0.025)
Surgery				Surgery			
	No	9	-		No	9	-
	Yes	97	0.77 (<i>p</i> = 0.537)		Yes	97	0.78 (<i>p</i> = 0.544)
Age				Age			
	Mean (SD)	61.7	1.02 (<i>p</i> = 0.144)		Mean (SD)	61.7	1.02 (<i>p</i> = 0.046)
Metastasis Site				Metastasis Site			
	Hepatic	76	-		Hepatic	76	-
	No Hepatic	34	1.14 (<i>p</i> = 0.592)		No Hepatic	34	1.14 (<i>p</i> = 0.598)
Laterality				Laterality			
	Right	26	-		Right	26	-
	Left	85	0.51 (<i>p</i> = 0.007)		Left	85	0.51 (<i>p</i> = 0.008)
ETV5				EGFR			
	Low	94	-		Low	71	-
	High	17	1.73 (<i>p</i> = 0.074)		High	40	0.80 (<i>p</i> = 0.344)
Surgery				Surgery			
	No	9	-		No	9	-
	Yes	97	0.79 (<i>p</i> = 0.576)		Yes	97	0.82 (<i>p</i> = 0.641)
Age				Age			
	Mean (SD)	61.7	1.02 (<i>p</i> = 0.069)		Mean (SD)	61.7	1.02 (<i>p</i> = 0.071)
Metastasis Site				Metastasis Site			
	Hepatic	76	-		Hepatic	76	-
	No Hepatic	34	1.08 (<i>p</i> = 0.768)		No Hepatic	34	1.14 (<i>p</i> = 0.607)
Laterality				Laterality			
	Right	26	-		Right	26	-
	Left	85	0.54 (<i>p</i> = 0.015)		Left	85	0.52 (<i>p</i> = 0.009)

Considering the identified influence of tumour laterality and each “molecular marker” in the patients’ prognostic, we questioned whether subgroup analysis considering both the “molecular marker” (High/Low *ETV1*, *ETV4*, *ETV5* or *EGFR*) and tumour laterality (*right/left*) would show significant associations with patients’ outcome. We thus queried the impact in OS and PFS of the different molecular and clinical subgroups using Kaplan-Maier survival curves (Figure 29 and Figure 30, respectively).

Overall, both in OS and PFS, tumours from the right colon seem to be associated with worse patient outcome than tumours from the left colon. Analysing the impact of specific groups in OS (Figure 29), *EGFR* overexpression seems to be associated with better OS among *left* colon tumours (Figure 29A). The influence of *EGFR* expression among tumours from the right colon cannot be interpreted due to the low number of samples in the “*Right & EGFR* High” group (N=2). On the other hand, *ETV1* expression doesn’t seem to be associated with OS among tumours from the right or the left colon, with *right* colon tumours showing a tendency for worse OS than *left* colon tumours, independently of *ETV1* expression (Figure 29B). Among *ETV4* subgroups, high expression of *ETV4* (Figure 29C) seems to be associated with better OS among tumours from the right side of the colon, while OS of patients with *left* colon tumours show similar survival curves (Figure 29C). Conversely, *ETV5* overexpression seems to be associated with worse OS among patients

with tumours from either *left* or *right* colon (Figure 29D). Log-rank statistical tests for each two-group comparisons are being performed to assess the significance of the associations.

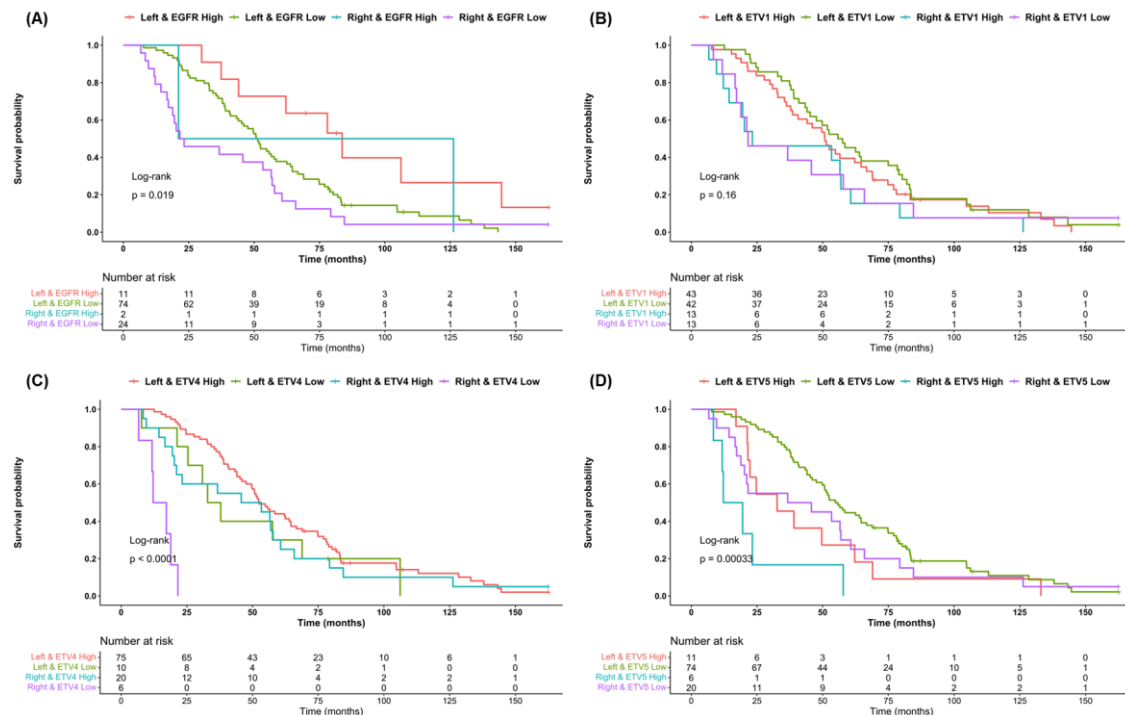


Figure 29. Kaplan-Meier curves of OS stratifying patients' tumours for both "molecular markers" and tumour laterality. Survival probability over 150 months for patients with tumours from the right or left side of the colon and **(A)** high/low *EGFR* expression, **(B)** high/low *ETV1* expression, **(C)** high/low *ETV4* expression, and **(D)** high/low *ETV5* expression. The p -value of Log-rank statistical test and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

Analysing the influence of gene expression and laterality in PFS (Figure 30), we observed that tumours with *EGFR* overexpression seem to be associated with better PFS when considering tumours from the right side of the colon. Among patients with *left* colon tumours, the PFS curves from "high" and "low" *EGFR* groups show high overlap (Figure 30A). Conversely to the impact in OS, in PFS, overexpression of *ETV1* seems to be associated with faster tumour progression in patients with tumours from either the right or the left colon (Figure 30B). Regarding *ETV4*, and considering that PFS curves show high overlap, in lack of statistical analysis (two-group Log-rank testes pending) we cannot infer possible associations, although *extreme* groups (with or without both risk factors – *right* colon tumours and *ETV4* Low) show completely distinct PFS curves (Figure 30C). Analysing *ETV5* subgroups, *ETV5* overexpression seems to be associated with worse PFS in patients having tumours from the left colon only, as PFS curves from patients with *right* colon

tumours show high overlap (Figure 30D). Again, Log-rank statistical tests for each two-group comparisons are being performed to assess the significance of the associations.

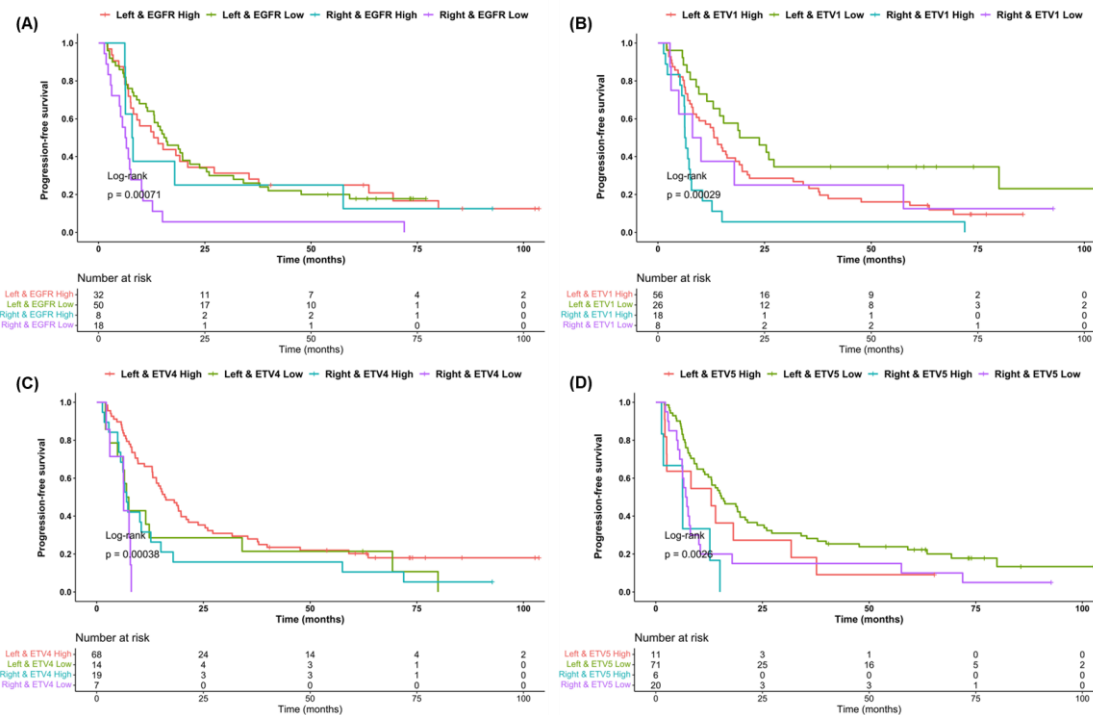


Figure 30. Kaplan-Meier curves of PFS stratifying patients' tumours for both "molecular markers" and tumour laterality. Progression-free probability over 100 months for patients with tumours from the right or left side of the colon and **(A)** high/low *EGFR* expression, **(B)** high/low *ETV1* expression, **(C)** high/low *ETV4* expression, and **(D)** high/low *ETV5* expression. The p -value of Log-rank statistical test and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

5. Definition of molecular predictors of response to anti-EGFR therapy considering tumour laterality

In light of the observed associations between the expression of the studied genes and OS or PFS, we defined a molecular profile with potential clinical utility for patient management according to tumour laterality (Table 11). Specifically, assessment of the expression levels of *ETV4* and *ETV5* could be useful to predict OS associated with *right* colon tumours, while, for *left* colon tumours, the expression profile of *EGFR* and *ETV5* may be useful. On the other hand, the expression profile of *ETV1*, and *EGFR* can be useful to predict response to anti-EGFR therapy in *right* colon tumours, while the expression profile of *ETV1*, *ETV4*, and *ETV5* can be useful to predict therapeutic response of *left* colon tumours.

Table 11. Proposed “molecular markers” of response to anti-EGFR therapy in mCRC patients considering tumour laterality.

	<i>Right colon</i>	<i>Left colon</i>
OS	<i>ETV4, ETV5</i>	<i>EGFR, ETV5</i>
PFS	<i>ETV1, EGFR</i>	<i>ETV1, ETV4, ETV5</i>

6. *GRPR* expression and its association with response to anti-EGFR therapy

Of the 7 cases with detectable *GRPR* expression, 3 samples (43%) have *GRPR* as the single gene with detectable expression, while 4 samples (57%) have expression of *GRPR* along with expression of *ETV5*, considering the “Outliers” *cut-off*.

Despite the low number of samples with *GRPR* expression, we compared the impact of the two groups in OS and PFS, along with a group of 8 patients with “Negative” tumours (without expression of any of the genes under study). Using the Log-rank test, statistically significant associations were found for both OS ($p=0.009$) and PFS ($p=0.015$) with the “*GRPR+ETV5*” group showing worse OS and PFS when compared with both the “*GRPR* only” group and the “Negative” group (Figures 31). Furthermore, having the “Negative” group as reference, “*GRPR+ETV5*” group is associated with a 4.72-fold risk of death ($HR=4.72$; CI: 1.12-19.76) and a 3.11-fold risk of tumour progression ($HR=3.11$; CI: 0.80-12.0). On the other hand, the isolated expression of *GRPR* associates with decreased probabilities of risk of death and tumour progression by 66% ($HR=0.34$; CI: 0.071-1.63) and 85% ($HR=0.15$; CI: 0.017-1.29), respectively.

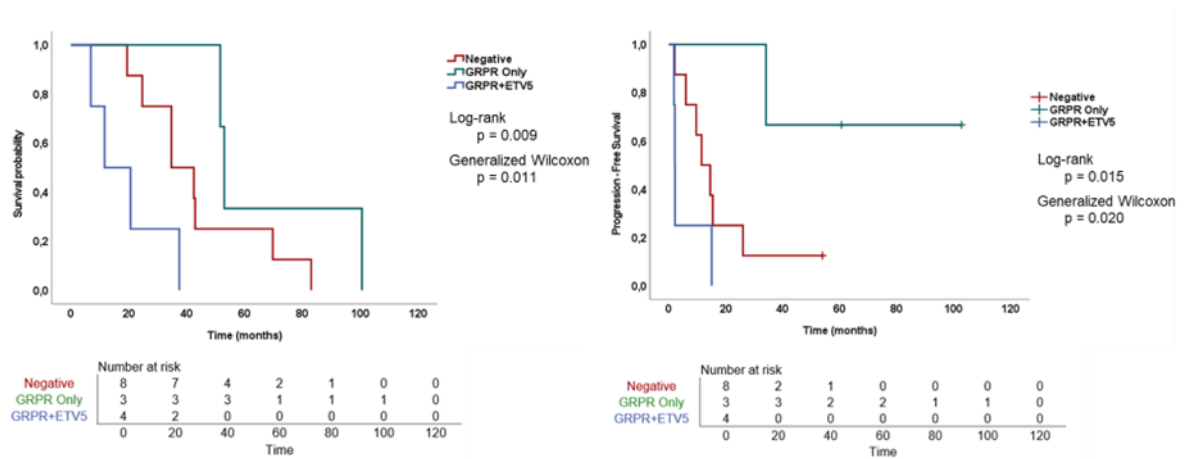


Figure 31. Kaplan-Meier curves of OS and PFS stratifying patients' tumours for GRPR expression. Survival probability over 120 months (*left*), and progression-free probability over 120 months (*right*) for patients with "GRPR Only", "GRPR+ETV5" and "Negative" tumours. The p -value of Log-rank and Generalized Wilcoxon statistical tests and a life table representing the number of patients alive (*left*) and without tumour progression (*right*) for each molecular subgroup, over time, are shown.

DISCUSSION

As described before, colorectal cancer is the third most common cancer worldwide and is responsible for a significant number of deaths, being the second most mortal cancer (4). Despite improvements in screening approaches, 15-30% of the patients are already diagnosed with metastatic colorectal cancer (mCRC) (18). Targeted therapies with monoclonal antibodies against EGFR are the most commonly used treatment approaches in mCRC (12, 34). Nevertheless, anti-EGFR therapies are only eligible for tumours without activating mutations in *KRAS/NRAS/BRAF* – named *wild-type* – and approximately 80% of the patients will develop therapy resistance. The underlying mechanisms of the “acquired” resistance are not yet completely understood, being this issue the main focus of the present study (91, 95).

Previous studies have highlighted the possible involvement of the PEA3 subfamily of transcription factors in EGFR signalling (71, 100). Moreover, GRPR has been described to interact with EGFR (84) and transactivate EGFR signalling (74).

Thus, we aimed to understand if overexpression of PEA3 transcription factors and/or *GRPR* was also observed in *wild-type* mCRC tumours and whether expression levels were associated with response to anti-EGFR target therapy.

1. The expression levels of *EGFR* and PEA3 transcription factors are associated with anti-EGFR therapy response

The PEA3 subfamily of transcription factors – including *ETV1*, *ETV4*, and *ETV5* – has been linked to cancer development and progression. Our findings indicate that overexpression of these transcription factors is associated with mCRC patient outcomes (OS and PFS), with high levels of *ETV1* and *ETV5* being associated with poorer, and *ETV4* overexpression with better, OS and PFS. An association with better OS was also observed for High *EGFR* expressing tumours, which seems to be modulated by PEA3 co-expression.

In this work, and in lack of an optimal method to define expression *cut-off* values to separate *high* and *low* expression samples, we have used the *Survival R package* to identify the optimal expression *cut-off* for each marker under study. However, we were interested in understand how the “optimal” *cut-off* could affect sample distribution in comparison with a *cut-off* based in the “Q3” or “Outlier” expression. Overall, we observed that sample stratification is different depending on the *cut-off* used – Q3, Outliers, Optimal-OS, or Optimal PFS – and, consequently, the “molecular subgroups” are also affected (Figure 14, Figure A-3, Appendix 2). Nevertheless, the three main molecular subgroups “*ETV1* Only”, “*ETV4* Only”, and “*ETV1+ETV4*” are maintained with all *cut-offs* (Figure A-5 and A-6,

Appendix 2). In fact, *ETV5* is the single gene where the proportion of samples with *High* expression is lower by applying the Optimal *cut-offs*, in comparison with “Q3” or “Outlier” *cut-offs* (Figure A-4, Appendix 2). Since the Optimal *cut-offs* are defined considering the effect of the marker in the outcome (either OS or PFS), this may suggest the lower impact of “intermediate” *ETV5* expression levels in OS and PFS, or the influence of the co-expression of other genes in the same samples. Another relevant difference between the expression *cut-offs* is the impact of the “molecular markers” in OS and PFS, since, using the Optimal *cut-offs*, statistically significant associations were found, while with “Q3” and “Outliers” *cut-offs* no associations were found with patients’ outcome (Figures-A7-A14, Appendix 3). Thus, the Optimal *cut-offs* approach allows to identify clinically different molecular subgroups that otherwise would not be established.

Regarding the possible influence of PEA3 subfamily and *EGFR* expression levels in patients’ prognosis, our results suggest an association between *ETV1* or *ETV5* overexpression and worse OS and PFS, while a high expression of *ETV4* is associated with better patients’ outcomes. Our findings validate a study from Shen et al. that evaluated the expression levels of PEA3 in CRC tumours and their influence on patients’ prognosis. They concluded that all these transcription factors are overexpressed in tumour tissues, and that *ETV1* and *ETV5* were associated with worse OS, whereas *ETV4* was associated with better patient survival (69). However, we further show a specific involvement of this transcription factors in response to anti-EGFR therapy. According to the literature, *ETV4* is involved in EGFR regulation, binding to its promoter, and leading to tumour growth and a more aggressive tumour behaviour (71, 101). On the other hand, *in vitro* studies by the Cancer Genetics Group of IPO-Porto have shown that *ETV1* overexpression, in a benign model of prostate cells, is sufficient to drive acquisition of oncogenic characteristics, namely, increased proliferation, invasion and anchorage independent growth (102), leading to increased activation of EGFR signalling (Azeredo et al., *in preparation*). In spite of the apparent overlapping activity of *ETV1* and *ETV4* in positively regulating EGFR, the expression levels of these transcription factors have shown to be inversely associated with response to anti-EGFR therapy. Considering the multiple phosphorylation sites of EGFR (51), it is possible that *ETV4*-mediated EGFR oncogenic signalling pathways are more affected by anti-EGFR inhibition than those mediated by *ETV1*, eventually as a result of differentially activated phosphorylation sites that may become affected by conformational changes of the receptor upon extracellular antibody binding. However, further studies are needed to clarify the EGFR-associated signalling pathways activated by PEA3 expression.

Our results also showed an association between high expression of *EGFR* and better OS. However, no statistically significant differences were observed in PFS, suggesting that transcriptomic *EGFR* expression levels may not be a good predictor of therapy response. In fact, this hypothesis is confirmed by multivariate analyses, which have showed that the prognostic value of *EGFR* expression, although independent from PEA3 expression in predicting OS, is not independent from other clinicopathological variables.

2. The prognostic value of each PEA3 transcription factor and of *EGFR* may be modulated by co-expression of one or more PEA3 members

To the best of our knowledge, no previous study has evaluated the impact of the combined overexpression of PEA3 transcription factors on patient prognosis. Our results suggest that the profile of different ETS transcription factors modulates the response to anti-*EGFR* therapy. Specifically, overexpression of *ETV4* seems to abrogate the worse OS and PFS associated with *ETV1* overexpression, leading to better prognosis. This effect is particularly evident when comparing the 5-year survival rate and the median survival associated with carcinomas “*ETV1* Only”, “*ETV4* Only” and “*ETV1+ETV4*” (Table 7). Conversely, *ETV5* overexpression abrogates the association of *ETV4* overexpression with high PFS, with PEA3-positive tumours having poorer therapy response than “*ETV1+ETV4*” tumours. Our *ETV5* findings validate Park et al. study where a link between *ETV5* expression and cetuximab resistance in colorectal cancer cell lines was established, concluding that *ETV5* silencing improves the sensitivity to mAb (68). Thus, this conclusion strengthens the association between *ETV5* overexpression and the worse therapy response suggested in our work. However, curiously, a group of tumours with combined expression of *ETV5* and *EGFR* (*ETV5+EGFR*) was not found, although a few samples have shown co-expression when considering the “Q3” cut-off (Figure A-5, Appendix 2). This may be related to the insufficient number of samples available in the *ETV5* “molecular subgroup”, and/or result from the individual gene adjustment used for the identification of the Optimal cut-off, leading to the exclusion of this “molecular subgroup” from the *ETV5 high* group.

3. Proposal of a “molecular profile” with potential clinical utility

Considering the association between the gene expression under study and OS or PFS, we propose a “molecular profile” with potential clinical utility considering the tumour laterality, the only clinical feature with an independent prognostic value. Despite some of the “molecular markers” showed to depend on tumour laterality, both *ETV5* and *ETV1* showed to have independent prognostic value for OS and PFS, respectively, being associated with worse prognosis among tumours from either the right or the left colon. High

ETV4 expression also showed an association with better OS and PFS in *right* and *left* colon tumours, respectively, while *EGFR* overexpression seems to associate with better OS and PFS in *left* and *right* colon tumours, respectively.

In fact, tumour laterality has shown to have independent prognostic value in OS and PFS, revealing to be important in clinical evaluation. These results are in accordance with the literature, where tumours from the right side of the colon have been associated with poorer prognosis and clinical response, when compared with tumours from the left colon (46). Additionally, our results support the NCCN guidelines for mCRC treatment, where anti-EGFR target therapy is recommended for *wild-type* patients with a primary tumour location in left colon (48). However, considering the gene expression evaluation, we also identified a potential utility of anti-EGFR therapy in *right* colon tumours depending on the “molecular subgroup”. Specifically, our results suggest that *right* colon tumours with *EGFR* overexpression are associated with good therapy response. This association strengthens the idea that tumour laterality and expression profile should be analysed together for a better clinical evaluation. Therefore, our results suggest that *ETV5* and *ETV1* have independent prognostic value in OS and PFS, respectively, since *ETV5* overexpression is associated with worse patients’ survival, and *ETV1* overexpression is associated with faster tumour progression, in tumours from either the left or the right colon.

In addition, although the statistical analysis of the “molecular subgroups” initially defined was not yet performed considering tumour laterality, the Kaplan-Meier curves suggest important associations with OS and PFS. Considering the *ETV5* independent prognostic value in patients’ OS, its expression profile can be useful to predict patients’ OS, independently of the clinicopathological characteristics of the tumours. Nevertheless, since the proportion of samples with high *ETV5* expression is much lower than those with *ETV4* or *EGFR* overexpression, the expression profiles of these markers, which seem to be associated with better OS in the *right* and the *left* colon, respectively, can add increased power to a panel of genes designed to predict the response to anti-EGFR therapy. Regarding PFS, our results suggest that *ETV1* expression levels should be analysed along with *EGFR* expression profile in tumours from the right colon, considering the worse prognosis associated with *ETV1* overexpression, which seems to abrogate the good therapy response conferred by high *EGFR* expression. Despite the independent action of *ETV1* in PFS, in *left* colon tumours the expression of *ETV5* can also be informative to predict therapy response. Concomitantly, since *ETV4* overexpression seems to abrogate the worse therapy response associated with *ETV1* overexpression, the expression profile of *ETV4* in *left* colon tumours can add increased predictive value to anti-EGFR therapy.

Considering the other clinicopathological features, despite age was found to be associated with OS and PFS, the risk only increases by 2/3% per year. Therefore, its influence on the final prognosis has no significant increased value. Regarding the metastatic site, our patients' cohort shows the liver as the most common site for metastasis, which is in accordance with the literature (103). Statistical analysis showed a 10-20% increase in the risk of mortality and tumour progression associated with metastasis beyond the liver (Table 9 and 10), even though statistical significance was not reached. Nevertheless, this tendency is in accordance with Luo et al. findings, where the influence of metastasis in different organ sites was analysed, and the liver and the brain were associated with better and worse survival, respectively (50).

4. GRPR

Overall, GRPR expression was observed in only 6.3 % of the mCRC cases analysed. We exclude, however, any technical bias resulting from the degradation of the FFPE samples in this low frequency of cases with GRPR expression since the primers designed for qRT-PCR have an amplicon size similar to the other genes.

Despite the low number of samples with *GRPR* expression, our results suggest an association between *GRPR* expression and better OS and PFS, that can, however, be lost when *ETV5* is co-overexpressed. In fact, our results suggest that simultaneous overexpression of *ETV5* abrogates the association of *GRPR* with better OS and PFS. The better OS and PFS associated with the single expression of *GRPR* comparing with the "Negative" and "*GRPR+ETV5*" tumours is in agreement with the studies from Rivera et al. and Glover et al. where *GRPR* overexpression was associated with better survival and lower tumour proliferation, which was linked to a more differentiated stage of tumour cells (79, 81). In light of this observations, it is possible that the majority of the GRPR-overexpressing tumours are underrepresented among metastatic CRCs.

On the other hand, the observed association between *GRPR* expression and good therapy response may be related with a previously reported mechanism, where GRPR transactivates EGFR (74) leading to higher receptor activity, and, eventually, better response to anti-EGFR target therapy. Curiously, the *GRPR+EGFR* "molecular subgroup" was not identified in our study, which may be due to the small number of samples with *GRPR* amplification and/or because EGFR transactivation by GRPR probably does not translate into increased expression of EGFR at transcriptomic level, but rather in receptor activation. Nevertheless, our results do support the possible utility of characterizing mCRC

tumours for *GRPR* expression in combination with PEA3 expression, as a marker of good response to anti-EGFR therapy.

CONCLUSION

Overall, we conclude that PEA3 transcription factors, *EGFR* and *GRPR* are possible biomarkers in mCRC, having predictive and/or prognostic value in response to anti-EGFR therapy. Specifically, we conclude that in mCRC patients treated with anti-EGFR therapy:

- *ETV1* and *ETV5* overexpression are independent predictors of therapy response, with *ETV5* having independent prognostic value for overall survival and *ETV1* for progression-free survival;
- *ETV4* overexpression associates with a better response to targeted therapy;
- *ETV1* overexpression abrogates the good therapeutic response associated with *EGFR* overexpression;
- The prognostic value of PEA3/*EGFR* expression in patient survival and therapy response is affected by tumour laterality;
- *GRPR* overexpression may identify a subgroup of patients that respond to anti-EGFR therapy.
- The definition of PEA3/*EGFR* molecular subtypes may be useful to guide patient stratification and the identification of new targeted therapies aiming to improve patients' survival.

FUTURE PERSPECTIVES

The relevance of the results obtained in this project demands the development of complementary analyses to increase the robustness of the findings and potentially improve patients' outcome. Some approaches are:

- Analyse the *PEA3/EGFR/GRPR* expression profiles and their association with patient outcome in the 52 patients not eligible for anti-EGFR therapy, in order to understand whether the “molecular profile” could be useful to stratify tumours with mutations in *KRAS/NRAS*;
- Extend the analysis already performed to all samples with ETS profile (N=396), in order to improve the statistical power of the study. To achieve this, external institutions will be invited to participate and provide clinical data;
- Identify targetable downstream effectors of *ETV1* and/or *ETV5* signalling by RNA sequencing of the whole transcriptome, opening a new avenue of potential targeted therapies for these molecular subgroups of mCRCs;
- Establish colorectal cancer cell models to test new therapeutic approaches targeting PEA3-associated pathways.

REFERENCES

1. Shelton BK. Introduction to colorectal cancer. *Semin Oncol Nurs*. 2002;18(2 Suppl 2):2-12.
2. Azzouz LL, Sharma S. Physiology, Large Intestine. *StatPearls*: StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.; 2023.
3. Kahai P, Mandiga P, Wehrle CJ, Lobo S. Anatomy, Abdomen and Pelvis: Large Intestine. *StatPearls*: StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.; 2023.
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209-49.
5. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology*. 2021;14(10):101174.
6. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023;72(2):338-44.
7. Dyba T, Randi G, Bray F, Martos C, Giusti F, Nicholson N, et al. The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. *European Journal of Cancer*. 2021;157:308-47.
8. Simon K. Colorectal cancer development and advances in screening. *Clinical Interventions in Aging*. 2016;Volume 11:967-76.
9. Mármol I, Sánchez-De-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi M. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *International Journal of Molecular Sciences*. 2017;18(1):197.
10. Li J, Ma X, Chakravarti D, Shalapour S, Depinho RA. Genetic and biological hallmarks of colorectal cancer. *Genes & Development*. 2021;35(11-12):787-820.
11. Labianca R, Nordlinger B, Beretta GD, Mosconi S, Mandalà M, Cervantes A, et al. Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2013;24(suppl 6):vi64-vi72.
12. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG, et al. Colorectal cancer. *Nature Reviews Disease Primers*. 2015;1(1):15065.

13. Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for Colorectal Cancer Screening. *Gastroenterology*. 2020;158(2):418-32.
14. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *New England Journal of Medicine*. 2014;370(14):1287-97.
15. Shaukat A, Levin TR. Current and future colorectal cancer screening strategies. *Nature Reviews Gastroenterology & Hepatology*. 2022;19(8):521-31.
16. Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ*. 2014;348(apr09 1):g2467-g.
17. de Haan MC, Pickhardt PJ, Stoker J. CT colonography: accuracy, acceptance, safety and position in organised population screening. *Gut*. 2015;64(2):342-50.
18. Cervantes A, Adam R, Roselló S, Arnold D, Normanno N, Taïeb J, et al. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals of Oncology*. 2023;34(1):10-32.
19. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Annals of Surgical Oncology*. 2010;17(6):1471-4.
20. Chapuis PH, Chan C, Dent OF. Clinicopathological staging of colorectal cancer: Evolution and consensus-an Australian perspective. *Journal of Gastroenterology and Hepatology*. 2011;26:58-64.
21. Centelles JJ. General Aspects of Colorectal Cancer. *ISRN Oncology*. 2012;2012:1-19.
22. Shek D, Akhuba L, Carlino MS, Nagrial A, Moujaber T, Read SA, et al. Immune-Checkpoint Inhibitors for Metastatic Colorectal Cancer: A Systematic Review of Clinical Outcomes. *Cancers*. 2021;13(17):4345.
23. Nasser Y, Langenfeld SJ. Imaging for Colorectal Cancer. *Surg Clin North Am*. 2017;97(3):503-13.

24. Legolvan MP, Taliano RJ, Resnick MB. Application of molecular techniques in the diagnosis, prognosis and management of patients with colorectal cancer: a practical approach. *Hum Pathol*. 2012;43(8):1157-68.
25. De La Chapelle A, Hampel H. Clinical Relevance of Microsatellite Instability in Colorectal Cancer. *Journal of Clinical Oncology*. 2010;28(20):3380-7.
26. Ma H, Brosens LAA, Offerhaus GJA, Giardiello FM, de Leng WWJ, Montgomery EA. Pathology and genetics of hereditary colorectal cancer. *Pathology*. 2018;50(1):49-59.
27. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. *Orphanet J Rare Dis*. 2009;4:22.
28. Hegde MR, Roa BB. Detecting mutations in the APC gene in familial adenomatous polyposis (FAP). *Curr Protoc Hum Genet*. 2006;Chapter 10:Unit 10.8.
29. Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *Journal of Clinical Oncology*. 2017;35(10):1086-95.
30. Perrod G, Rahmi G, Cellier C. Colorectal cancer screening in Lynch syndrome: Indication, techniques and future perspectives. *Digestive Endoscopy*. 2021;33(4):520-8.
31. Sinicrope FA. Lynch Syndrome-Associated Colorectal Cancer. *N Engl J Med*. 2018;379(8):764-73.
32. Ogino S, Goel A. Molecular Classification and Correlates in Colorectal Cancer. *The Journal of Molecular Diagnostics*. 2008;10(1):13-27.
33. Mundade R, Imperiale TF, Prabhu L, Loehrer PJ, Lu T. Genetic pathways, prevention, and treatment of sporadic colorectal cancer. *Oncoscience*. 2014;1(6):400-6.
34. Carethers JM. Review: Systemic treatment of advanced colorectal cancer: Tailoring therapy to the tumor. *Therapeutic Advances in Gastroenterology*. 2008;1(1):33-42.
35. García-Foncillas J, Sunakawa Y, Aderka D, Wainberg Z, Ronga P, Witzler P, et al. Distinguishing Features of Cetuximab and Panitumumab in Colorectal Cancer and Other Solid Tumors. *Front Oncol*. 2019;9:849.

36. Janani B, Vijayakumar M, Priya K, Kim JH, Prabakaran DS, Shahid M, et al. EGFR-Based Targeted Therapy for Colorectal Cancer—Promises and Challenges. *Vaccines*. 2022;10(4):499.
37. Johdi NA, Sukor NF. Colorectal Cancer Immunotherapy: Options and Strategies. *Front Immunol*. 2020;11:1624.
38. Weng J, Li S, Zhu Z, Liu Q, Zhang R, Yang Y, et al. Exploring immunotherapy in colorectal cancer. *Journal of Hematology & Oncology*. 2022;15(1).
39. Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in Colorectal Cancer: Current Research and Future Prospects. *International Journal of Molecular Sciences*. 2020;21(15):5311.
40. Wolpin BM, Meyerhardt JA, Mamon HJ, Mayer RJ. Adjuvant Treatment of Colorectal Cancer. *CA: A Cancer Journal for Clinicians*. 2007;57(3):168-85.
41. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. *K-ras* Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *New England Journal of Medicine*. 2008;359(17):1757-65.
42. Khelwatty SA, Essapen S, Seddon AM, Modjtahedi H. Prognostic significance and targeting of HER family in colorectal cancer. *Front Biosci (Landmark Ed)*. 2013;18(2):394-421.
43. Grothey A, Fakhri M, Tabernero J. Management of BRAF-mutant metastatic colorectal cancer: a review of treatment options and evidence-based guidelines. *Annals of Oncology*. 2021;32(8):959-67.
44. Luo X-J, Zhao Q, Liu J, Zheng J-B, Qiu M-Z, Ju H-Q, et al. Novel Genetic and Epigenetic Biomarkers of Prognostic and Predictive Significance in Stage II/III Colorectal Cancer. *Molecular Therapy*. 2021;29(2):587-96.
45. Bender U, Rho YS, Barrera I, Aghajanyan S, Acoba J, Kavan P. Adjuvant Therapy for Stages II and III Colon Cancer: Risk Stratification, Treatment Duration, and Future Directions. *Current Oncology*. 2019;26(11):43-52.
46. Baek SK. Laterality: Right-Sided and Left-Sided Colon Cancer. *Annals of Coloproctology*. 2017;33(6):205-6.

47. Petrelli F, Tomasello G, Borgonovo K, Ghidini M, Turati L, Dallera P, et al. Prognostic Survival Associated With Left-Sided vs Right-Sided Colon Cancer. *JAMA Oncology*. 2017;3(2):211.
48. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2021;19(3):329-59.
49. Brulé SY, Jonker DJ, Karapetis CS, O'Callaghan CJ, Moore MJ, Wong R, et al. Location of colon cancer (right-sided versus left-sided) as a prognostic factor and a predictor of benefit from cetuximab in NCIC CO.17. *Eur J Cancer*. 2015;51(11):1405-14.
50. Luo D, Liu Q, Yu W, Ma Y, Zhu J, Lian P, et al. Prognostic value of distant metastasis sites and surgery in stage IV colorectal cancer: a population-based study. *Int J Colorectal Dis*. 2018;33(9):1241-9.
51. Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. *Molecular Oncology*. 2018;12(1):3-20.
52. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. *Acta Pharm Sin B*. 2015;5(5):390-401.
53. Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, et al. Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*. 2006;366(1):2-16.
54. Chan DLH, Segelov E, Wong RS, Smith A, Herbertson RA, Li BT, et al. Epidermal growth factor receptor (EGFR) inhibitors for metastatic colorectal cancer. *Cochrane Database of Systematic Reviews*. 2017.
55. Lui VW, Thomas SM, Zhang Q, Wentzel AL, Siegfried JM, Li JY, et al. Mitogenic effects of gastrin-releasing peptide in head and neck squamous cancer cells are mediated by activation of the epidermal growth factor receptor. *Oncogene*. 2003;22(40):6183-93.
56. Thomas SM, Grandis JR, Wentzel AL, Gooding WE, Lui VW, Siegfried JM. Gastrin-releasing peptide receptor mediates activation of the epidermal growth factor receptor in lung cancer cells. *Neoplasia*. 2005;7(4):426-31.
57. Paulo P, Ribeiro FR, Santos J, Mesquita D, Almeida M, Barros-Silva JD, et al. Molecular Subtyping of Primary Prostate Cancer Reveals Specific and Shared Target Genes of Different ETS Rearrangements. *Neoplasia*. 2012;14(7):600-IN15.

58. Santos J, Mesquita D, Barros-Silva JD, Jerónimo C, Henrique R, Morais A, et al. Uncovering potential downstream targets of oncogenic GRPR overexpression in prostate carcinomas harboring ETS rearrangements. *Oncoscience*. 2015;2(5):497-507.
59. Nicholas TR, Strittmatter BG, Hollenhorst PC. Oncogenic ETS Factors in Prostate Cancer. *Adv Exp Med Biol*. 2019;1210:409-36.
60. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer*. 2017;17(6):337-51.
61. Chotteau-Lelievre A, Montesano R, Soriano J, Soulie P, Desbiens X, de Launoit Y. PEA3 transcription factors are expressed in tissues undergoing branching morphogenesis and promote formation of duct-like structures by mammary epithelial cells in vitro. *Dev Biol*. 2003;259(2):241-57.
62. Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. *Gene*. 2003;303:11-34.
63. Riggi N, Stamenkovic I. The Biology of Ewing sarcoma. *Cancer Lett*. 2007;254(1):1-10.
64. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644-8.
65. Attard G, Clark J, Ambrosine L, Mills IG, Fisher G, Flohr P, et al. Heterogeneity and clinical significance of ETV1 translocations in human prostate cancer. *British Journal of Cancer*. 2008;99(2):314-20.
66. Segalés L, Juanpere N, Lorenzo M, Albero-González R, Fumadó L, Cecchini L, et al. Strong cytoplasmic ETV1 expression has a negative impact on prostate cancer outcome. *Virchows Arch*. 2019;475(4):457-66.
67. Cheng X, Jin Z, Ji X, Shen X, Feng H, Morgenlander W, et al. ETS variant 5 promotes colorectal cancer angiogenesis by targeting platelet-derived growth factor BB. *International Journal of Cancer*. 2019;145(1):179-91.
68. Park SM, Hwang CY, Cho SH, Lee D, Gong JR, Lee S, et al. Systems analysis identifies potential target genes to overcome cetuximab resistance in colorectal cancer cells. *The FEBS Journal*. 2019;286(7):1305-18.

69. Shen X, Zhou C, Feng H, Li J, Xia T, Cheng X, et al. ETV1 Positively Correlated With Immune Infiltration and Poor Clinical Prognosis in Colorectal Cancer. *Front Immunol.* 2022;13:939806.
70. Fonseca AS, Ramão A, Bürger MC, De Souza JES, Zanette DL, De Molfetta GA, et al. ETV4 plays a role on the primary events during the adenoma-adenocarcinoma progression in colorectal cancer. *BMC Cancer.* 2021;21(1).
71. Leng Y, Chen Z, Ding H, Zhao X, Qin L, Pan Y. Overexpression of microRNA-29b inhibits epithelial-mesenchymal transition and angiogenesis of colorectal cancer through the ETV4/ERK/EGFR axis. *Cancer Cell International.* 2021;21(1).
72. Hsing M, Wang Y, Rennie PS, Cox ME, Cherkasov A. ETS transcription factors as emerging drug targets in cancer. *Med Res Rev.* 2020;40(1):413-30.
73. Konstantinopoulos PA, Papavassiliou AG. Seeing the future of cancer-associated transcription factor drug targets. *Jama.* 2011;305(22):2349-50.
74. Hohla F, Schally AV. Targeting gastrin releasing peptide receptors: New options for the therapy and diagnosis of cancer. *Cell Cycle.* 2010;9(9):1738-41.
75. Mansi R, Fleischmann A, Mäcke HR, Reubi JC. Targeting GRPR in urological cancers--from basic research to clinical application. *Nat Rev Urol.* 2013;10(4):235-44.
76. Saurin JC, Rouault JP, Abello J, Berger F, Remy L, Chayvialle JA. High gastrin releasing peptide receptor mRNA level is related to tumour dedifferentiation and lymphatic vessel invasion in human colon cancer. *Eur J Cancer.* 1999;35(1):125-32.
77. Carroll RE, Matkowskyj KA, Chakrabarti S, McDonald TJ, Benya RV. Aberrant expression of gastrin-releasing peptide and its receptor by well-differentiated colon cancers in humans. *Am J Physiol.* 1999;276(3):G655-65.
78. Chave HS, Gough AC, Palmer K, Preston SR, Primrose JN. Bombesin family receptor and ligand gene expression in human colorectal cancer and normal mucosa. *British Journal of Cancer.* 2000;82(1):124-30.
79. Glover SC, Tretiakova MS, Carroll RE, Benya RV. Increased frequency of gastrin-releasing peptide receptor gene mutations during colon-adenocarcinoma progression. *Mol Carcinog.* 2003;37(1):5-15.

80. Jensen JA, Carroll RE, Benya RV. The case for gastrin-releasing peptide acting as a morphogen when it and its receptor are aberrantly expressed in cancer. *Peptides*. 2001;22(4):689-99.
81. Rivera CA, Ahlberg NC, Taglia L, Kumar M, Blunier A, Benya RV. Expression of GRP and its receptor is associated with improved survival in patients with colon cancer. *Clin Exp Metastasis*. 2009;26(7):663-71.
82. Fleischmann A, Waser B, Reubi JC. Overexpression of gastrin-releasing peptide receptors in tumor-associated blood vessels of human ovarian neoplasms. *Cell Oncol*. 2007;29(5):421-33.
83. Nagasaki S, Nakamura Y, Maekawa T, Akahira J, Miki Y, Suzuki T, et al. Immunohistochemical analysis of gastrin-releasing peptide receptor (GRPR) and possible regulation by estrogen receptor β cx in human prostate carcinoma. *Neoplasma*. 2012;59(02):224-32.
84. Liebow C, Crean DH, Lee MT, Kamer AR, Mang TS, Schally AV. Synergistic effects of bombesin and epidermal growth factor on cancers. *Proceedings of the National Academy of Sciences*. 1994;91(9):3804-8.
85. Cornelio DB, Roesler R, Schwartzmann G. Gastrin-releasing peptide receptor as a molecular target in experimental anticancer therapy. *Ann Oncol*. 2007;18(9):1457-66.
86. Baratto L, Jadvar H, Iagaru A. Prostate Cancer Theranostics Targeting Gastrin-Releasing Peptide Receptors. *Mol Imaging Biol*. 2018;20(4):501-9.
87. Bajo AM, Schally AV, Krupa M, Hebert F, Groot K, Szepeshazi K. Bombesin antagonists inhibit growth of MDA-MB-435 estrogen-independent breast cancers and decrease the expression of the ErbB-2/HER-2 oncoprotein and c-jun and c-fos oncogenes. *Proceedings of the National Academy of Sciences*. 2002;99(6):3836-41.
88. Schally AV, Comaru-Schally AM, Nagy A, Kovacs M, Szepeshazi K, Plonowski A, et al. Hypothalamic hormones and cancer. *Front Neuroendocrinol*. 2001;22(4):248-91.
89. Zhang Q, Bhola NE, Lui VWY, Siwak DR, Thomas SM, Gubish CT, et al. Antitumor mechanisms of combined gastrin-releasing peptide receptor and epidermal growth factor receptor targeting in head and neck cancer. *Molecular Cancer Therapeutics*. 2007;6(4):1414-24.

90. Liu P, Tu Y, Tao J, Liu Z, Wang F, Ma Y, et al. GRPR-targeted SPECT imaging using a novel bombesin-based peptide for colorectal cancer detection. *Biomater Sci*. 2020;8(23):6764-72.
91. Martinelli E, Ciardiello D, Martini G, Troiani T, Cardone C, Vitiello PP, et al. Implementing anti-epidermal growth factor receptor (EGFR) therapy in metastatic colorectal cancer: challenges and future perspectives. *Annals of Oncology*. 2020;31(1):30-40.
92. Zhao B, Wang L, Qiu H, Zhang M, Sun L, Peng P, et al. Mechanisms of resistance to anti-EGFR therapy in colorectal cancer. *Oncotarget*. 2017;8(3):3980-4000.
93. Leto SM, Trusolino L. Primary and acquired resistance to EGFR-targeted therapies in colorectal cancer: impact on future treatment strategies. *Journal of Molecular Medicine*. 2014;92(7):709-22.
94. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature*. 2012;486(7404):532-6.
95. Zhou J, Ji Q, Li Q. Resistance to anti-EGFR therapies in metastatic colorectal cancer: underlying mechanisms and reversal strategies. *Journal of Experimental & Clinical Cancer Research*. 2021;40(1).
96. Birkman EM, Avoranta T, Ålgars A, Korkeila E, Lintunen M, Lahtinen L, et al. EGFR gene copy number decreases during anti-EGFR antibody therapy in colorectal cancer. *Hum Pathol*. 2018;82:163-71.
97. Lu Y, Li X, Liang K, Luwor R, Siddik ZH, Mills GB, et al. Epidermal growth factor receptor (EGFR) ubiquitination as a mechanism of acquired resistance escaping treatment by the anti-EGFR monoclonal antibody cetuximab. *Cancer Res*. 2007;67(17):8240-7.
98. Bertotti A, Papp E, Jones S, Adleff V, Anagnostou V, Lupo B, et al. The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature*. 2015;526(7572):263-7.
99. Butler JM. Chapter 3 - DNA Quantitation. In: Butler JM, editor. *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego: Academic Press; 2012. p. 49-67.
100. Seth A, Watson DK. ETS transcription factors and their emerging roles in human cancer. *Eur J Cancer*. 2005;41(16):2462-78.

101. Yao D, Bao Z, Qian X, Yang Y, Mao Z. ETV4 transcriptionally activates HES1 and promotes Stat3 phosphorylation to promote malignant behaviors of colon adenocarcinoma. *Cell Biol Int*. 2021;45(10):2129-39.
102. Mesquita D, Barros-Silva JD, Santos J, Skotheim RI, Lothe RA, Paulo P, et al. Specific and redundant activities of ETV1 and ETV4 in prostate cancer aggressiveness revealed by co-overexpression cellular contexts. *Oncotarget*. 2015;6(7):5217-36.
103. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet*. 2014;383(9927):1490-502.

APPENDIX

Appendix 1

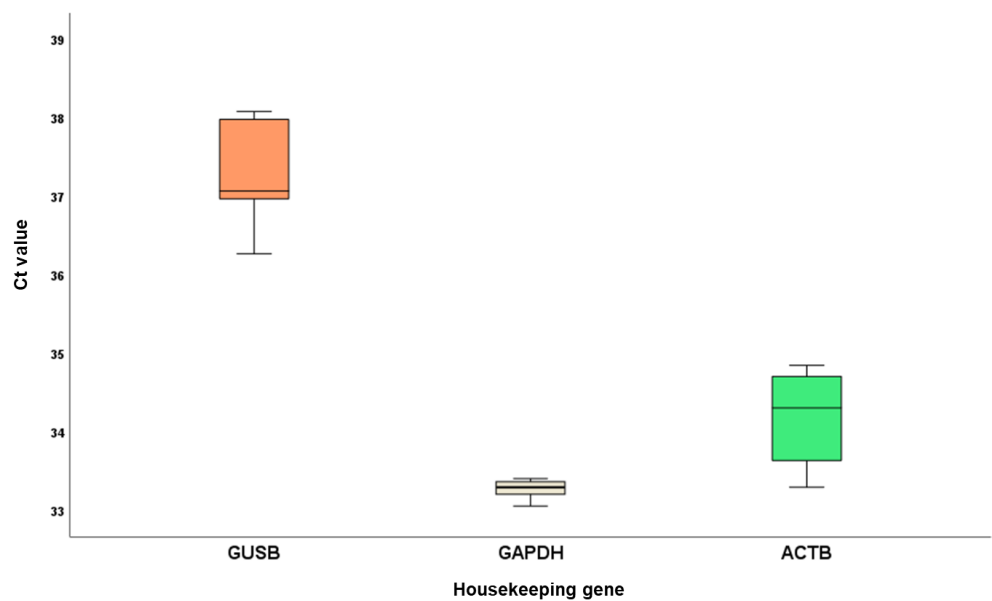


Figure A-1. Evaluation of the most adequate housekeeping gene. Ct value distribution of 9 samples from 2013 for three housekeeping genes – *GUSB*, *GAPDH*, and *ACTB*.

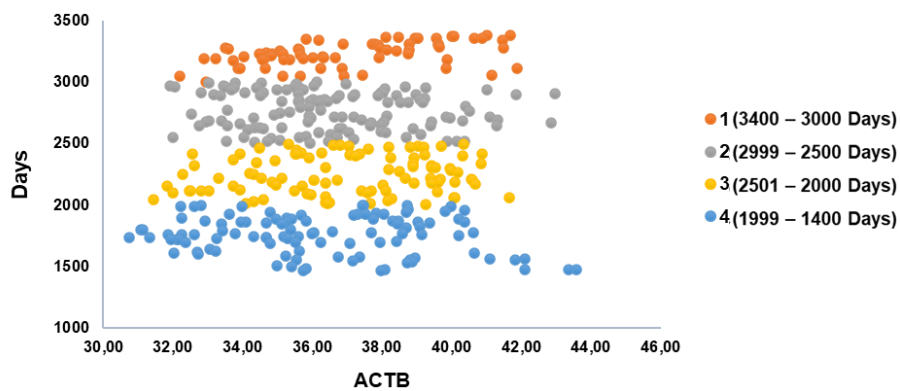


Figure A-2. Quality evaluation of the samples under study. Distribution of *ACTB* expression considering the age of the 396 total samples.

Appendix 2

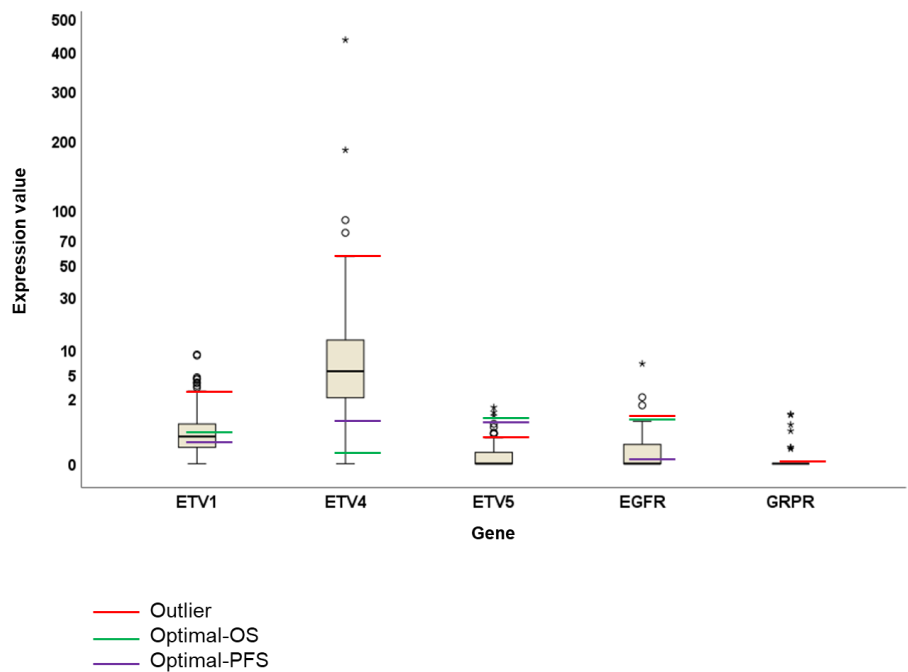


Figure A-3. Different expression *cut-offs* for each gene in the 111 analysed samples. Representation of the threshold from which the samples have high gene expression considering the different *cut-offs* defined - “Q3”, “Outliers”, Optimal-OS, and Optimal-PFS.

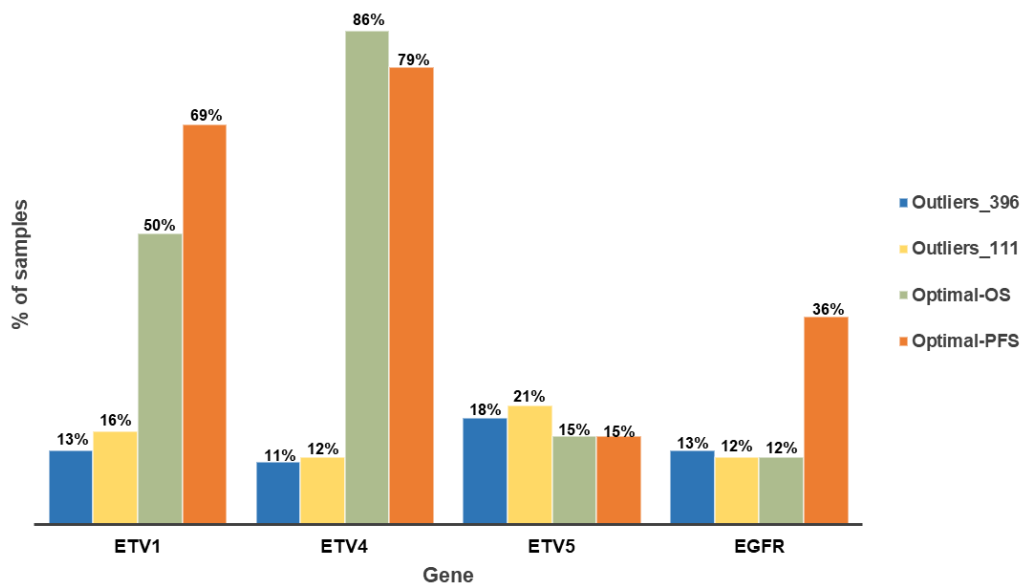


Figure A-4. Frequency of the 396 and 111 samples with overexpression of *ETV1*, *ETV4*, *ETV5*, or *EGFR* considering the “Outliers” *cut-off* for both groups, and the Optimal *cut-offs* for the 111 samples.

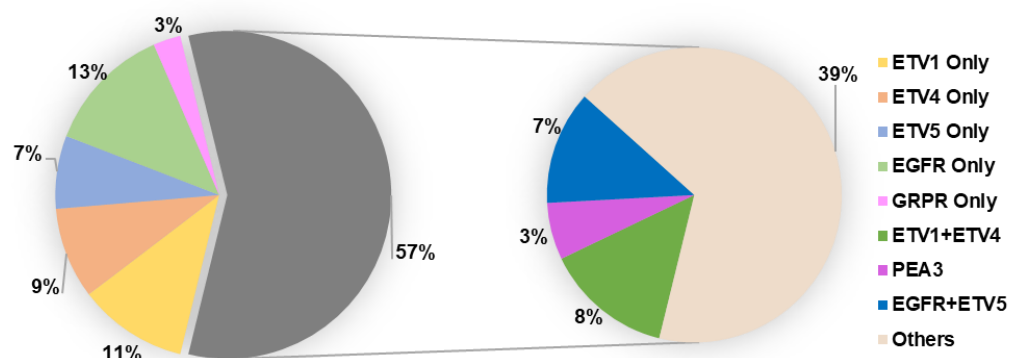


Figure A-5. “Molecular subgroups” considering the “Q3” cut-off. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, *ETV5*, *EGFR*, or *GRPR* and samples with overexpression of two or more genes (*right*). The “Others” group includes negative samples for all the genes under study (<Q1) (N=8), samples with expression <Q3 (N=23) and gene combinations found in less than four cases (N=12).

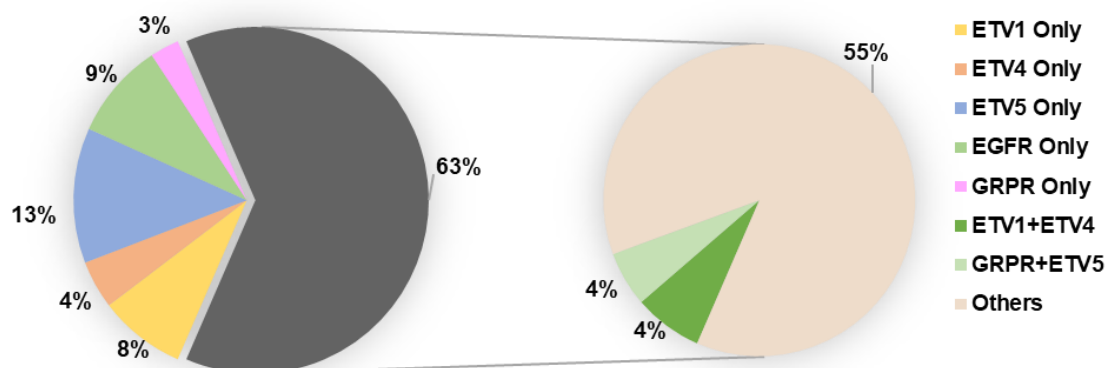


Figure A-6. “Molecular subgroups” considering the “Outliers” cut-off. Distribution of samples according to the “molecular subgroup” highlighting samples with overexpression of a single gene (*left*) - *ETV1*, *ETV4*, *ETV5*, *EGFR*, or *GRPR* and samples with overexpression of two or more genes (*right*). The “Others” group includes negative samples for all the genes under study (<Q1) (N=8), samples with expression <Q3 (N=46) and gene combinations found in less than four cases (N=7).

Appendix 3

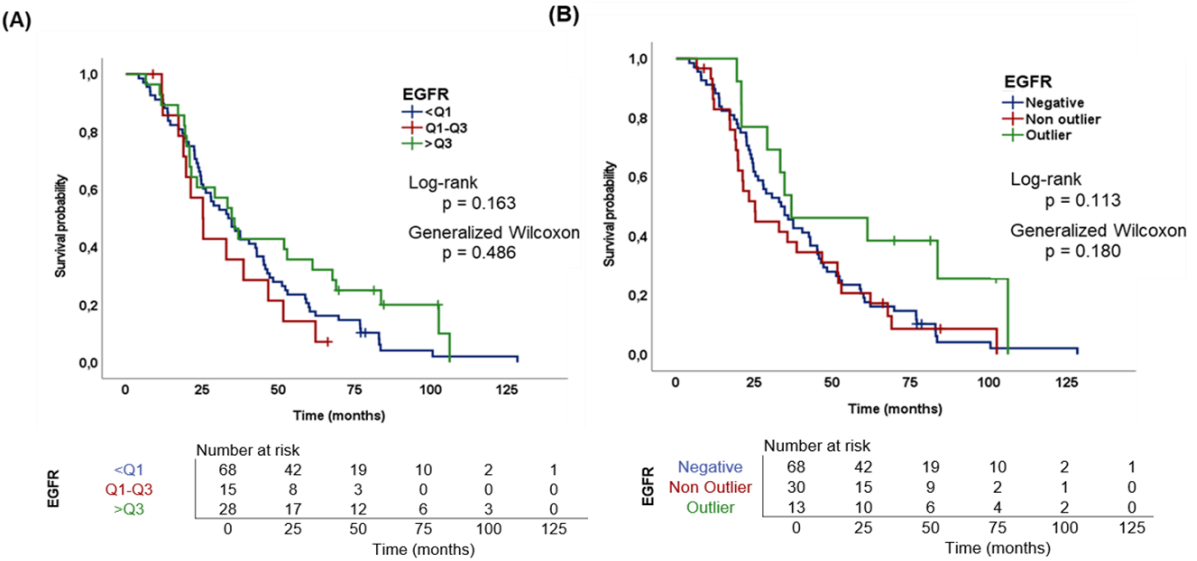


Figure A-7. Kaplan-Meier curves of OS stratifying patients for *EGFR* expression. Survival probability over 125 months for patients with high and low *EGFR* expression using (A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.163$ and $p=0.113$, respectively) and the Generalized Wilcoxon ($p=0.486$ and $p=0.180$, respectively) statistical tests and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

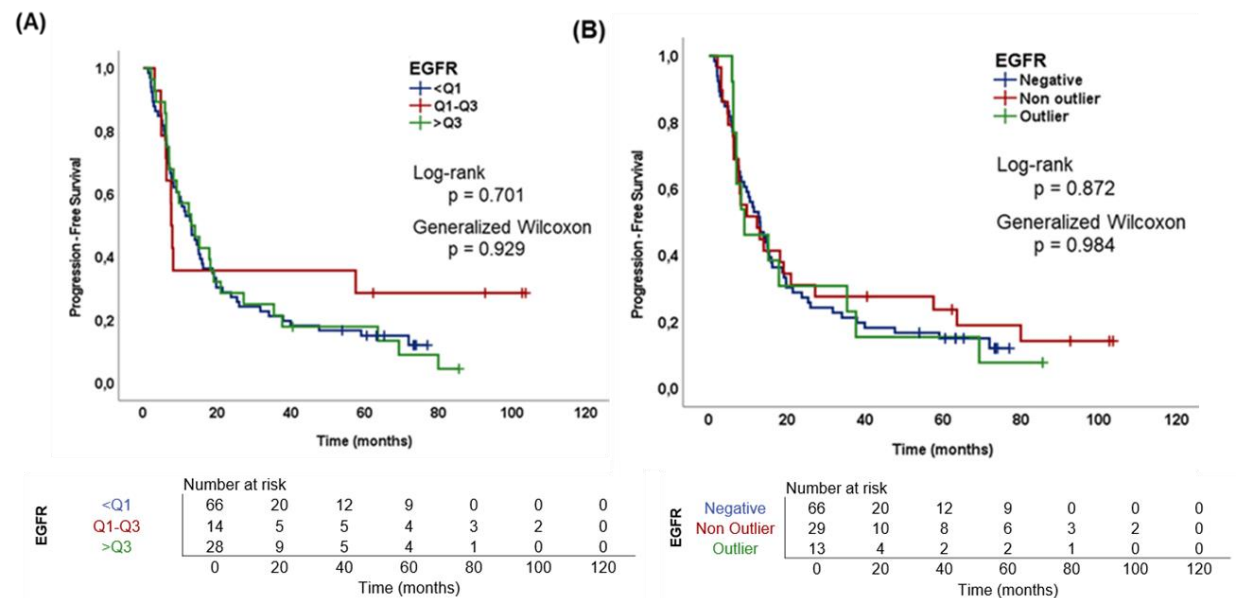


Figure A-8. Kaplan-Meier curves of PFS stratifying patients for *EGFR* expression. Progression-free probability over 120 months for patients with high and low *EGFR* expression using (A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.701$ and $p=0.872$, respectively) and the Generalized Wilcoxon ($p=0.929$ and $p=0.984$, respectively) statistical

tests and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

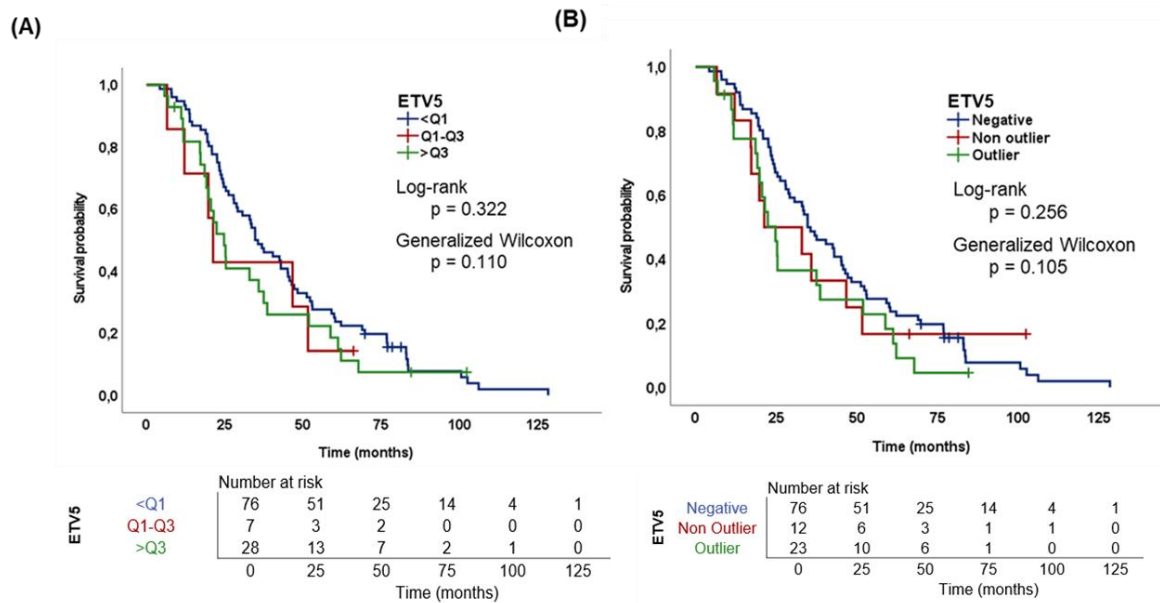


Figure A-9. Kaplan-Meier curves of OS stratifying patients for *ETV5* expression. Survival probability over 125 months for patients with high and low *ETV5* expression using **(A)** the “Q3” cut-off and **(B)** the “Outliers” cut-off. For “Q3” and “Outliers” the *p*-value of Log-rank ($p=0.322$ and $p=0.256$, respectively) and the Generalized Wilcoxon ($p=0.110$ and $p=0.105$, respectively) statistical tests and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

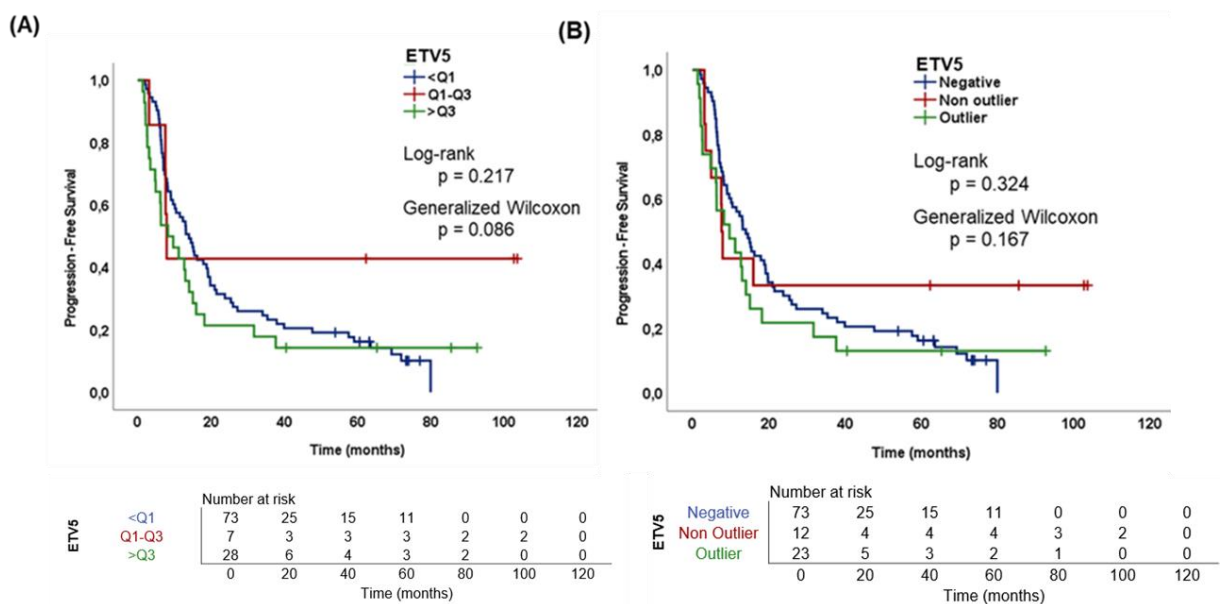


Figure A-10. Kaplan-Meier curves of PFS stratifying patients for *ETV5* expression. Progression-free probability over 120 months for patients with high and low *ETV5* expression using **(A)** the “Q3” *cut-off* and **(B)** the “Outliers” *cut-off*. For “Q3” and “Outliers” the *p*-value of Log-rank

($p=0.217$ and $p=0.324$, respectively) and the Generalized Wilcoxon ($p=0.086$ and $p=0.167$, respectively) statistical tests and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

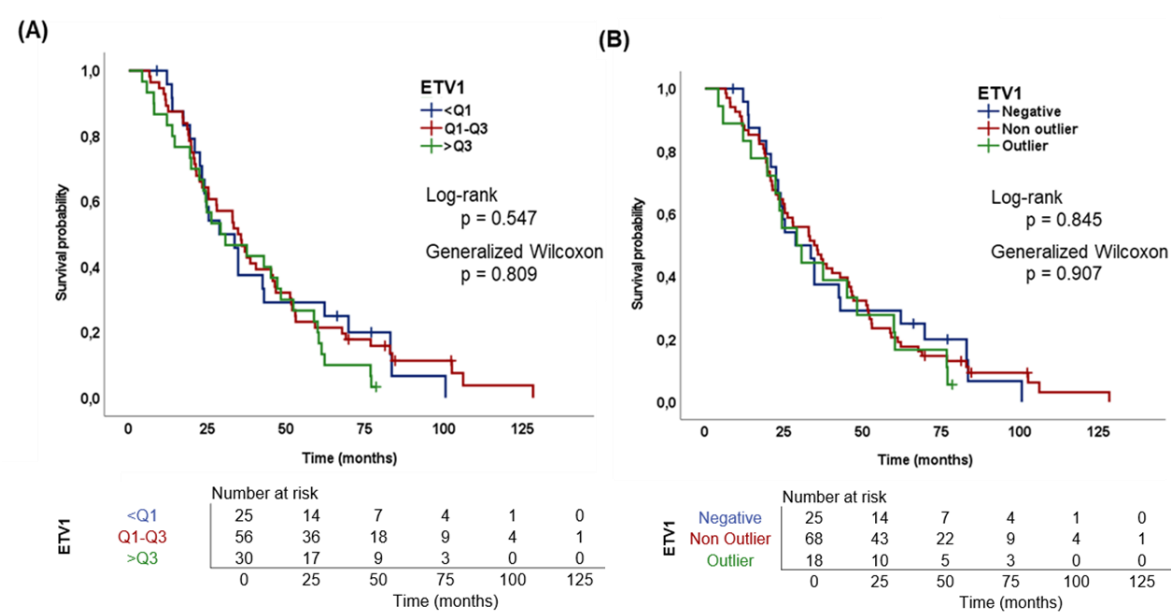


Figure A-11. Kaplan-Meier curves of OS stratifying patients for *ETV1* expression. Survival probability over 125 months for patients with high and low *ETV1* expression using (A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.547$ and $p=0.845$, respectively) and the Generalized Wilcoxon ($p=0.809$ and $p=0.907$, respectively) statistical tests and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

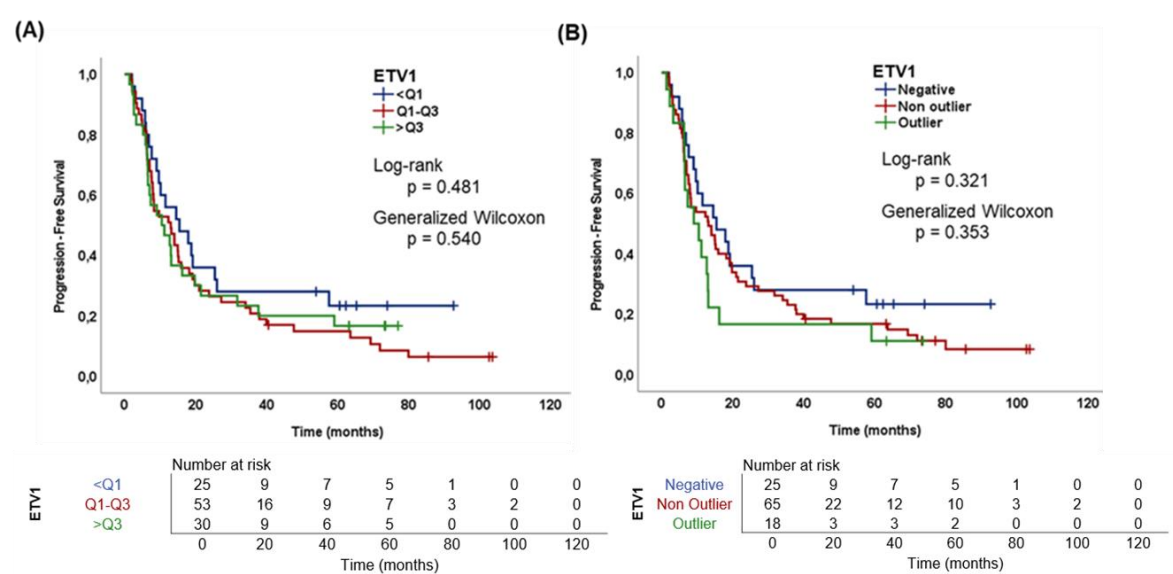


Figure A-12. Kaplan-Meier curves of PFS stratifying patients for *ETV1* expression. Progression-free probability over 120 months for patients with high and low *ETV1* expression using

(A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.481$ and $p=0.321$, respectively) and the Generalized Wilcoxon ($p=0.540$ and $p=0.353$, respectively) statistical tests and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

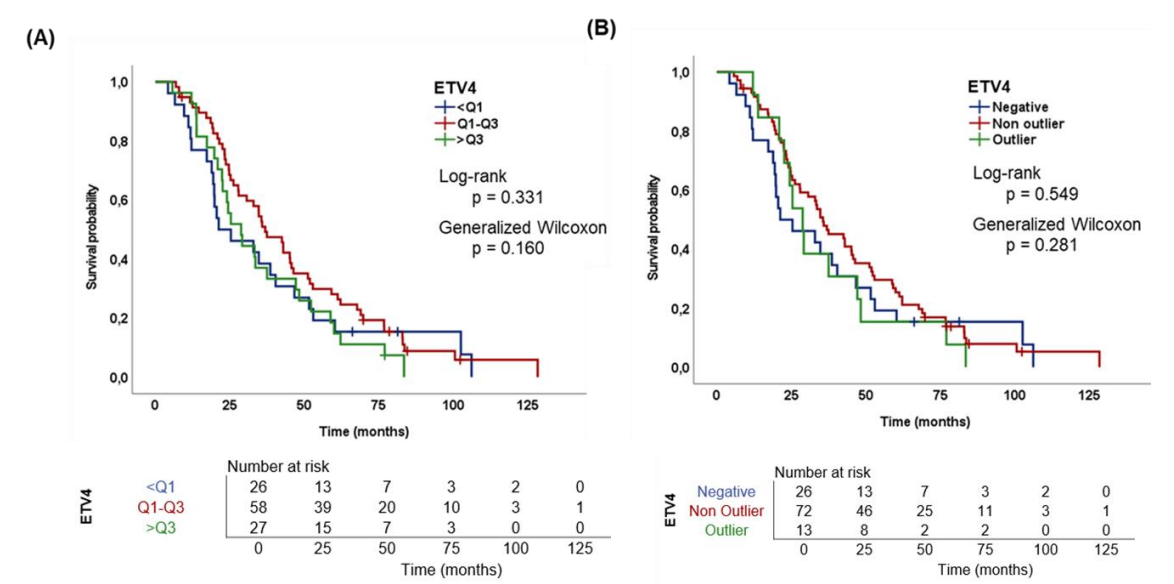


Figure A-13. Kaplan-Meier curves of OS stratifying patients for *ETV4* expression. Survival probability over 125 months for patients with high and low *ETV4* expression using (A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.331$ and $p=0.549$, respectively) and the Generalized Wilcoxon ($p=0.160$ and $p=0.281$, respectively) statistical tests and a life table representing the number of patients alive for each molecular subgroup, over time, are shown.

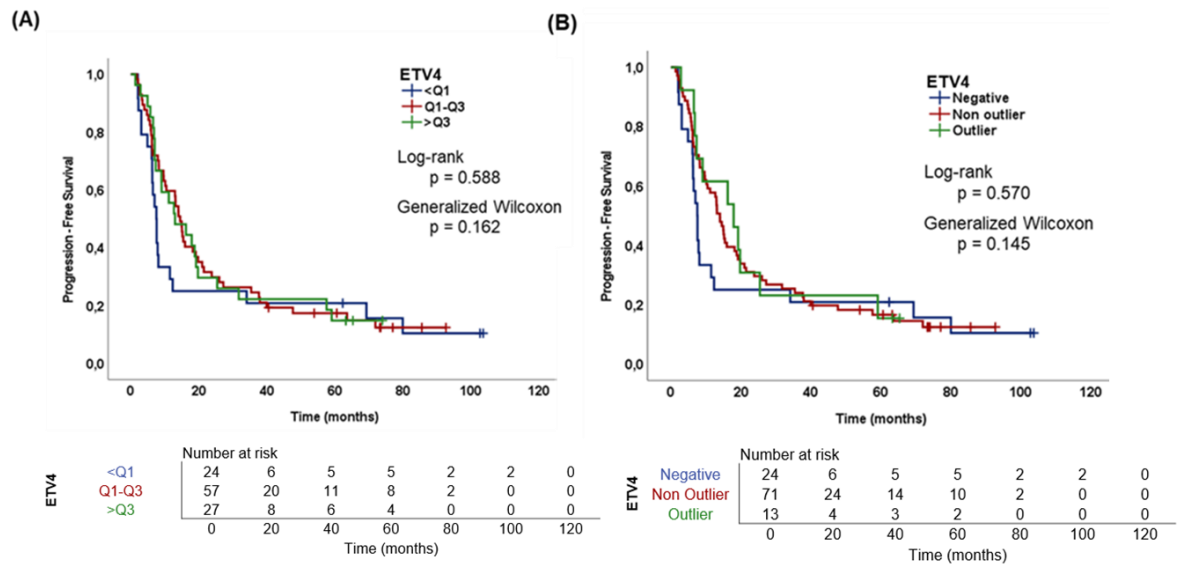


Figure A-14. Kaplan-Meier curves of PFS stratifying patients for *ETV4* expression. Progression-free probability over 120 months for patients with high and low *ETV4* expression using (A) the “Q3” cut-off and (B) the “Outliers” cut-off. For “Q3” and “Outliers” the p -value of Log-rank ($p=0.588$ and $p=0.570$, respectively) and the Generalized Wilcoxon ($p=0.162$ and $p=0.145$, respectively) statistical tests and a life table representing the number of patients without tumour progression for each molecular subgroup, over time, are shown.

Appendix 4

Table A-1.1. Univariate analysis of OS for the molecular profile and the clinicopathological features surgery, age, metastasis site, and laterality.

OS			
		N	HR (Univariable)
ETV1	Low	55	-
	High	56	1.42 ($p=0.091$)
ETV4	Low	16	-
	High	95	0.65 ($p=0.126$)
ETV5	Low	94	-
	High	17	1.91 ($p=0.016$)
EGFR	Low	98	-
	High	13	0.51 ($p=0.043$)
Surgery	No	9	-
	Yes	97	0.63 ($p=0.189$)
Age			
	Mean (SD)	61.7	1.03 ($p=0.022$)
Metastasis Site			
	Hepatic	76	-
	No Hepatic	34	1.25 ($p=0.313$)
Laterality			
	Right	26	-
	Left	85	0.42 ($p<0.001$)

Table A-1.2. Univariate analysis of PFS for the molecular profile and the clinicopathological features surgery, age, metastasis site, and laterality.

PFS			
		N	HR (Univariable)
ETV1	Low	34	-
	High	77	1.74 ($p=0.019$)
ETV4	Low	23	-
	High	88	0.55 ($p=0.020$)
ETV5	Low	94	-
	High	17	1.72 ($p=0.051$)
EGFR	Low	71	-
	High	40	0.85 ($p=0.468$)
Surgery	No	9	-
	Yes	97	0.83 ($p=0.613$)
Age			
	Mean (SD)	61.7	1.02 ($p=0.059$)
Metastasis Site			
	Hepatic	76	-
	No Hepatic	34	1.12 ($p=0.625$)
Laterality			
	Right	26	-
	Left	85	0.48 ($p=0.002$)