

---

REZUMATUL TEZEI DE DOCTORAT

# Studiul unor markeri imunohistochimici și moleculari în cancerul colo-rectal

---

Doctorand **Constantin-Ioan Busuioc**

---

Conducător de doctorat Prof.dr. **Ioana Berindan-Neagoe**

---



**UMF**  
UNIVERSITATEA DE  
MEDICINĂ ȘI FARMACIE  
IULIU HAȚIEGANU  
CLUJ-NAPOCA

# CUPRINS

<b>INTRODUCERE</b>	13
<b>STADIUL ACTUAL AL CUNOAȘTERII</b>	
<b>1. Adenocarcinomul colo-rectal</b>	17
1.1 Incidență	17
1.2 Factori de risc	17
1.3. Patogeneză	18
1.3.1 Modificări genetice	19
1.3.2 Clasificarea genomică și transcriptomică a cancerului colo-rectal	20
1.4. Clasificarea patologică a cancerului colo-rectal	22
1.5 Factori de prognostic morfologici și conceptul de muguri tumorali.	25
1.5.1 Mugurii tumorali	25
<b>2. Tranziția epitelial-mezenchimală (EMT)</b>	27
2.1 EMT: date generale	27
2.2 EMT și rezistența la terapie	27
2.3 Mecanisme moleculare în EMT	28
2.3.1 Factori de transcripție implicați în EMT	29
2.3.2 miARN în EMT	30
2.4 Rolul EMT în cancerul colorectal	31
2.5 Markerii imunohistochimici ai EMT	32
<b>3. Calea de semnalizare MAPK și rolul ei în cancerul colo-rectal</b>	35
3.1 Calea MAPK și EMT	37
<b>4. Implicarea genelor MMP3 și TESC în cancerul colo-rectal</b>	39
4.1 Gena MMP3 în cancerul colo-rectal	39
4.2 Gena TESC în cancerul colo-rectal	39
<b>CONTRIBUȚIA PERSONALĂ</b>	
<b>1. Ipoteza de lucru/obiective</b>	43
<b>2. Metodologie generală</b>	45
<b>3. Studiul 1. Semnătura genelor implicate în Tranziția Epitelial-Mezenchimală și legătura cu prognosticul în COAD</b>	47
3.1 Introducere	47
3.2 Materiale și metode	48

3.3 Rezultate	51
3.4 Discuții	61
3.5 Concluzii	64
<b>4. Studiul 2. Analiza diferentiaa a genelor exprimate, a valorilor prognostice si a căilor moleculare in adenocarcinomul de colon: o abordare bioinformatică cu accent pe MMP3 și TESC</b>	65
4.1 Introducere	65
4.2 Materiale și metode	67
4.3 Rezultate	71
4. Discuții	80
5. Concluzii	83
<b>5. Studiul 3. Tranzitia epitelial-mezenchimala in cancerul colorectal: studiu imunohistochimic</b>	85
5.1 Introducere	85
5.2 Materiale și metdodă	87
5.2 Rezultate	89
5.4 Discuții	100
5.5 Concluzii	102
<b>6. Discuții generale</b>	103
<b>7. Concluzii generale</b>	105
<b>8. Originalitatea și contribuțiile inovative ale tezei</b>	107
<b>REFERINȚE</b>	109
<b>ANEXE</b>	129

**Cuvinte cheie: cancer colo-rectal, tranziție epitelial-mezenchimală, analiza genelor, bioinformatică, qRT-PCR, prognostic, biomarkeri, Vimentină, E-caderină, OCT3/4**

# INTRODUCERE

Cancerul colo-rectal este o importantă problemă de sănătate publică, de talie globală, ce prezintă o incidență în creștere în rândul tinerilor și în rândul populațiilor țărilor în curs de dezvoltare, din cauza modificării stilului de viață în urma occidentalizării acestor țări.

Managementul clinic al cancerului colo-rectal s-a îmbunătățit considerabil în ultimul timp. Cu toate acestea, el se confruntă cu limitări majore în ce privește tratamentul, mai ales în stadiile avansate, care sunt din păcate majoritare în momentul diagnosticului. Astfel, chimio și radio-terapia se confruntă cu probleme ce țin de toxicitate și rezistență la terapie, în timp ce tratamentul chirurgical cu tentă curativă este limitat la cazuri incipiente.

Incidența în creștere și problemele ce țin de rezistența la terapie ne obligă să încercăm să identificăm noi actori moleculari implicați în patogeneză, cu roluri potențiale de biomarkeri sau ținte terapeutice, cu scopul de a ne lărgi arsenalul terapeutic.

În această lucrare ne-am concentrat atenția în mare parte pe un mecanism biologic fundamental, implicat în procese fiziologice ca vindecarea tisulară sau dezvoltarea embriologică, dar și în progresia cancerului, proces denumit tranziția epitelial-mezenchimală. Acest mecanism complex induce celulelor tumorale proprietăți migratorii și invazive, făcând posibile invazia tisulară și metastazarea, fiind practic responsabil de majoritatea complicațiilor nefavorabile și a deceselor de cauză canceroasă. Fiind un proces central și indispensabil pentru metastazare, este tentant să credem că înțelegerea sa mai bună va putea scoate în evidență vulnerabilități ce pot duce la inhibarea acestui proces și va putea priva cancerul de trăsăturile sale cele mai distructive, oferind posibilități terapeutice suplimentare. Cancerul de colon este un candidat remarcabil pentru acest tip de studiu, întrucât dispune de un corespondent morfologic al tranziției epitelial-mezenchimale sub forma mugurilor tumorali. În plus, este un tip de cancer foarte bine studiat atât la nivel genomic cât și transcripomic, ceea ce îl face un foarte bun candidat pentru studii bioinformatică ce utilizează numeroase seturi de date disponibile în baze de date publice și care au fost obținute în urma numeroaselor analize genomice, transcriptomice și post-tranlaționale care au luat amploare în ultimul timp.

În prima parte a tezei, intitulată "Stadiul actual al cunoașterii", vom expune cadrul teoretic al problematicii cancerului colo-rectal și al tranziției epitelial mezenchimale, oferind ultimele date demografice și cele mai recente clasificări și trăsături moleculare. În partea a doua a tezei, intitulată "Contribuții personale", vom expune pe larg 3 studii care folosesc analize bioinformatică complexe, teste de biologie moleculară, evaluarea unor parametri clinico-histologici și expresia imunohistochimică a unor proteine implicate în tranziția epitelial-mezenchimală în adenocarcinomul colonic.

În primul studiu a fost identificată o semnătură genetică distinctă a EMT constând din două gene supraexprimate (*NOX4* și *IGF2BP3*), precum și cinci gene subexprimate (*BMP5*, *DACT3*, *EEF1A2*, *GCNT2* și *SFRP1*) care au demonstrat legătură cu prognosticul în cancerul colo-rectal. O etapă de ulterioară de validare prin qRT-PCR a fost efectuată într-o cohortă de pacienți cu cancer colo-rectal care cuprinde 29 de țesuturi tumorale și 29 de țesuturi normale adiacente, aprobând nivelul de expresie pentru *BMP5*, precum și pentru

două dintre miARN-urile care vizează genele cheie legate de EMT, evidențiind în același timp supraexpresia miR-27a-5p și miR-146a-5p.

În al doilea studiu, folosind baza de date UALCAN, au fost identificate primele 25 de gene supra și subexprimate din COAD. Apoi, un plotter Kaplan-Meier a fost folosit pentru aceste gene pentru analiza supraviețuirii, dezvăluind corelații cu rata globală de supraviețuire doar pentru *MMP3* și *TESC*, care sunt ținte relevante ce ar trebui validate suplimentar ca biomarkeri pentru diagnosticarea precoce și ținte terapeutice.

În al treilea studiu au fost evaluate piese de rezecție chirurgicală de la 30 de pacienți cu cancer de colon și au fost efectuate teste imunohistochimice pentru a investiga expresia markerilor legați de EMT și OCT 3/4 în celulele tumorale comparativ cu țesut normal. Pozitivitatea Vimentinei, OCT3/4 și pierderea E-caderinei în tumori au fost asociate semnificativ cu gradul tumorii, mugurii tumorali, frontul tumoral invaziv și metastazele limfoganglionare. Acești anticorpi ar putea servi ca un panou de biomarkeri care ar putea ajuta la prognosticul pacienților, cu potențialul suplimentar de a fi ținte terapeutice.

## CONTRIBUTII PERSONALE

### Ipoteza de lucru

Așa cum am arătat în Capitolul 1, COAD reprezintă o problemă de importanță ridicată, cu potențial de a deveni mai și mai acută în viitor. În evoluția tumorilor epiteliale maligne, pe măsură ce acestea dobândesc pas cu pas trăsături care duc la invazie și metastazare, momentul în care celulele trec prin EMT este probabil cel mai important din cauza efectului determinant pe care EMT îl are asupra diseminării cancerului, pe de o parte. Pe de altă parte, așa cum am arătat în Capitolul 2, EMT se asociază și cu rezistența la terapie, contribuind și mai mult la arsenalul tumorilor maligne. În aceste condiții, nu putem să nu ne întrebăm dacă nu cumva acest mecanism complex ar putea fi exploatat și transformat în viitor într-o vulnerabilitate a cancerului.

Am plecat de la premiza că EMT este un program de o importanță centrală în evoluția cancerului și că este un program biologic extrem de complex, ce cuprinde numeroase componente moleculare localizate la multiple niveluri, atât intracelulare cât și extracelulare. Am dorit în primul rând să studiem îndeaproape acest program pentru a înțelege mai bine genele codificante și non-codificante asociate, în vederea identificării unor noi biomarkeri sau ținte terapeutice eficiente. Aceste gene sau transcripții lor pot avea deci valoare diagnostică, prognostică sau terapeutică. Prin urmare, în primul studiu am dorit să evaluăm semnătura genelor implicat în TEM în COAD și cum se corelează expresia acestor gene la prognosticul pacienților. Un obiectiv suplimentar a fost validarea semnăturii genelor prin qRT-PCR în țesuturi tumorale și normale.

Al doilea obiectiv principal sa identificam pe baza studierii celor mai exprimate sau subexprimate 25 de gene în COAD, gene care se corelează cu prognosticul. După ce au fost identificate două gene (*TESC* și *MMP3*), ne-am propus ca obiectiv suplimentar analize bioinformatică care să integreze date disponibile de la nivel genomic, transcriptomic și proteomic și realizarea corelații cu parametrii clinici disponibili pentru a înțelege biologia acestor două gene.

Apoi am dorit să studiem EMT din punct de vedere imunohistochimic, pentru a putea vizualiza în COAD expresia diferențială a doi markeri IHC consacrați în EMT (VIM și E-CAD) și cum se corelează cu indicatorul morfologic al TEM în COAD (mugurii tumorali) și restul parametrilor clinico-patologici relevanți, care se

asociază cu prognostic negativ. Un al treilea marker, implicat în biologia celulelor stem (OCT3/4) a fost studiat, din cauza că a fost demonstrat în mod consistent că este de asemenea implicat în EMT.

## Metodologie generală

Cele trei studii s-au desfășurat cu respectarea strictă a normelor deontologice și etica cercetării științifice, după ce a fost obținut acordul Comisie de Etică a Universității de Medicină și Farmacie "Iuliu Hațieganu" din Cluj-Napoca.

Pentru analizele bioinformatică, în primul studiu, am folosit inițial date de expresie din baza de date a The Cancer Omics Atlas (TCOA), o resursă integrativă pentru informațiile genomice, transcriptomice și proteomice din cancer, care permite utilizatorului să efectueze diferite tipuri de analiză pe aceste informații. Aici am interogat baza de date TCOA pentru a ne afișa o listă cu primele 20 de gene mutate în COAD, atât supraexprimate cât și subexprimate, în comparație cu țesuturi non-tumorale. Apoi am comparat nivelurile de expresie în funcție de stadiu cu ajutorul instrumentului bioinformatic GEPIA, instrument care a fost ulterior folosit și pentru a corela datele de supraviețuire legate de genele EMT în COAD. Pentru analiza rețelei de gene implicate în EMT și reprezentarea interacțiunilor mRNA-miARN, am utilizat instrumentele String 11.0 și respectiv miRNE. Suplimentar, au fost analizate datele de supraviețuire pentru miARN-urile relevante ce vizează genele cheie în EMT, cu ajutorul StarBase. Apoi, am efectuat o analiză a modelelor mutaționale a genelor EMT în COAD în 169 de studii de cancer bazate pe trei seturi de date, colosind cBioPortal. Pentru corelațiile între genele EMT și COAD precum și între gena BMP5 și miARN-urile asociate ei au fost utilizate instrumentele CANCERTOOL și respectiv StarBase. Pentru validarea semnăturii genetice descoperite de studiul bioinformatic, au fost analizate prin qRT-PCR 29 de probe de țesut colonic normal și tumoral.

În al doilea studiu au fost utilizate exclusiv instrumente bioinformatică care au analizat date din TCGA, ce conțin o cohortă de 286 de tumori COAD și 41 de țesuturi normale adiacente. Cu ajutorul STARBASE am identificat primele 25 de gene supra și subexprimate, acestea fiind analizate ulterior individual, pentru corelații cu supraviețuirea. Cele două gene asociate semnificativ cu supraviețuirea (*MMP3* și *TESC*) au fost apoi evaluate pentru analiza expresiei proteice pe baza datelor disponibile de la CPTAC (Clinical Proteomic Tumor Analysis Consortium). Căile de semnalizare în care sunt aceste gene implicate au fost evaluate funcțional cu programul software IPA (ingenuity pathway analysis) pentru a evidenția funcțiile biologice în care sunt implicate.

Al treilea studiu a cuprins 30 de cazuri de COAD care au fost evaluate histologic pentru parametrii care se asociază cu prognosticul nefavorabil (Stadiul T, Stadiul N, invazia limfovaculară, tipul invaziei limfovaculară, invazia perineurală, tipul frontului tumoral, gradul tumoral, mugurii tumorali, localizarea tumorilor și tipul histologic). Apoi au fost efectuate teste IHC pentru VIM, E-CAD și OCT3/4 care au fost corelate cu toți acești parametri. Datele au fost prelucrate statistic folosind SPSS versiunea 23.0, rulând pe un sistem de operare Windows 10. Pentru statistica descriptivă s-au calculat mediile și abaterile standard. Pentru variabilele cantitative s-au calculat medianele și quartilele, iar pentru cele calitative, s-au calculat frecvența și procentul.

Comparativ cu datele cantitative și în funcție de normalitatea datelor, testul t Student (testul T al eșantioanelor independente) (pentru două grupuri cu date normal distribuite) și Mann-Whitney U (pentru datele care nu au o distribuție normală) au fost folosite, respectiv. Datele cantitative au fost testate pentru

normalitate (testul Shapiro-Wilk) și omogenitatea variantelor folosind testul Levene. Testul exact al lui Fisher a fost folosit pentru date. Probabilitatea de eroare mai mică de 5% ( $p < 0,05$ ) a fost considerată pragul de semnificație statistică în toate cele 3 studii.

## **Studiul 1. Semnătura genelor implicate în Tranziția Epitelial-Mezenchimală și legătura cu prognosticul în COAD**

Adenocarcinomul de colon (COAD) rămâne o cauză importantă a mortalității cauzate de cancer la nivel mondial. Tranziția epitelial-mezenchimală (EMT) este un mecanism cheie, promovând nu numai fenotipul invaziv sau metastatic, ci și rezistența la terapie. Folosind abordări bioinformatiche, am studiat alterarea genelor legate de EMT și implicația acestora asupra prognosticului COAD bazat pe seturi de date publice. Pentru mecanismele EMT, au fost identificate două gene supraexprimate (NOX4 și IGF2BP3), precum și cinci gene subreglate (BMP5, DACT3, EEF1A2, GCNT2 și SFRP1) care au fost strâns legate de prognosticul în COAD. O etapă de validare a qRT-PCR a fost efectuată într-o cohortă de pacienți COAD care cuprinde 29 de țesuturi tumorale și 29 de țesuturi normale adiacente, aprobând nivelul de expresie pentru BMP5, precum și pentru două dintre miARN-urile care vizează genele cheie legate de EMT, dezvoltând supra-reglarea miR-27a-5p și a miR-146a-5p. Semnătura EMT poate fi utilizată pentru a dezvolta un panou de biomarkeri pentru predicția recurenței la pacienții cu COAD, care poate contribui la îmbunătățirea stratificării riscului pentru pacienți.

## **Studiul 2. Analiza diferențială a genelor exprimate, a valorilor prognostice și a căilor moleculare în adenocarcinomul de colon: o abordare bioinformatică cu accent pe MMP3 și TESC**

În ciuda progresului mare în diagnosticul și tratamentul precoce, adenocarcinomul de colon (COAD) încă pune probleme importante pentru managementul clinic. Prin urmare, identificarea de noi biomarkeri sau ținte terapeutice pentru această boală este importantă. Folosind UALCAN, au fost identificate primele 25 de gene reglate în sus și în jos din COAD. Apoi, un plotter Kaplan-Meier a fost folosit pentru aceste gene pentru analiza supraviețuirii, dezvoltând corelații cu rata globală de supraviețuire numai pentru *MMP3* (Matrix Metalloproteinază 3) și *TESC* (Tescalcin). În ciuda acestui fapt, nivelurile de expresie a mRNA nu au fost corelate cu stadiile tumorii sau cu statusul metastatic nodal. *MMP3* și *TESC* sunt ținte relevante în COAD care ar trebui validate suplimentar ca biomarkeri pentru diagnosticarea precoce și prevenirea. Analiza bioinformatică a căilor de semnalizare a dezvoltat o rețea principală legată de modificarea post-translațională, degradarea proteinelor și sinteza proteinelor, unde *MMP3* a fost în centrul rețelei. O altă rețea importantă a fost legată de reglarea ciclului celular, *TESC* fiind o componentă a acesteia. De asemenea, nu ar trebui să subestimăm mecanismele complexe de reglare mediate de interacțiunea multiplelor alte molecule de reglare, subliniind interconectarea cu moleculele legate de invazia și migrarea implicate în COAD, care ar putea servi drept bază pentru dezvoltarea de noi biomarkeri și ținte terapeutice.

## Studiul 3. Tranziția epitelial-mezenchimală în cancerul colorectal: studiu imunohistochimic

Deși managementul clinic al cancerului colorectal a fost îmbunătățit semnificativ, acesta se confruntă cu o incidență în creștere în rândul tinerilor și în rândul celor din țările în curs de dezvoltare. Mai mult, diagnosticul apare mai ales în stadii avansate, când resursele terapeutice sunt limitate. Prin urmare, avem nevoie de noi biomarkeri pentru diagnosticare și ținte terapeutice. Evenimentul cheie care duce la invazie și metastază este tranziția epitelial-mezenchimală (EMT), care poate fi studiată cu ajutorul unor markeri IHC. Am urmărit corelarea expresiei markerilor legați de EMT (Vimentină și E-cadherină) și a unui marker de celule stem (OCT 3/4) cu parametrii clinicopatologici ai tumorilor. În acest studiu au fost evaluate probe de rezecție chirurgicală de la 30 de pacienți cu cancer de colon netratați neoadjuvant, internați în perioada 2018-2021. Au fost efectuate teste imunohistochimice pentru a investiga expresia markerilor legați de EMT și OCT 3/4 în celulele tumorale. Ca rezultate, Vimentina, pozitivitatea OCT3/4 și pierderea E-cadherinei au fost asociate semnificativ cu gradul tumorii, înmugurirea tumorii, frontul tumoral invaziv și metastaza ganglionilor limfatici. În concluzie, Vimentina, E-cadherina și OCT 3/4 ar putea servi ca un panou de biomarkeri care pot ajuta la prognosticul pacienților, cu potențialul adăugat de a fi ținte terapeutice.

### Concluzii generale

a) Pe lângă rolul adaptativ pe care EMT îl joacă în biologia cancerului și implicarea sa în invazie, metastazare și rezistența la terapie, am demonstrat că semnătura genelor implicate în EMT se corelează cu prognosticul în COAD.

b) Semnătura genelor implicate în EMT prezintă interacțiuni complexe cu miARN, care ar putea avea importanță clinică.

c) Semnătura genelor implicate în EMT susține progresia tumorii prin interacțiuni cu căile majore de semnalizare, ceea ce sugerează că pot reprezenta ținte terapeutice pe lângă rolul lor ca markeri de prognostic.

d) Dintre cele mai frecvent mutate gene supraregulate și subregulate în COAD, două gene cheie (*TESC* și *MMP3*) s-au corelat cu supraviețuirea.

e) Pe lângă potențialul lor de a fi utilizate ca biomarkeri, studiile noastre au descris interacțiunile și procesele biologice în care sunt implicate aceste două gene, deschizând ușa pentru investigarea mecanismelor reglatoare implicate în COAD:

e) Din cauza puținelor informații disponibile în legătură cu *MMP3* și *TESC* și având în vedere potențialul lor imens ca biomarker și respectiv țintă terapeutică, este necesară studierea lor pe cohorte mari de pacienți pentru validarea lor și corelarea cu trăsăturile clinico-patologice ale cancerului, atât în COAD cât și în alte tipuri de cancer.

f) Expresia IHC crescută a Vimentinei în COAD s-a corelat în studiul nostru cu gradul tumoral, categoria T, categoria N, numărul ganglionilor limfatici metastatici, tipul de invazie vasculară, frontul tumoral invaziv și mugurii tumorali.

g) Pierderea expresiei IHC pentru E-cadherină s-a corelat semnificativ statistic cu gradul tumorii, mugurii tumorali, categoriile N mai mari și frontul tumoral invaziv, aspect invers față de Vimentină, susținând datele din literatură cu privire la imunohistochimia EMT.

h) OCT3/4, un marker de celule stem a prezentat un profil de expresie IHC similar cu al Vimentinei (expresie corelată cu mugurii tumorali, frontul tumoral de tip invaziv, gradul tumoral și numărul de ganglioni limfatici pozitivi), sugerând că poate fi utilizat ca un marker al EMT.

i) expresia IHC a celor 3 markeri ar putea fi integrată în practica diagnostică curentă pentru evaluarea prognosticului în COAD.



---

PhD THESIS SUMMARY

# Study of immunohistochemical and molecular markers in colo-rectal cancer

---

PhD student **Constantin-Ioan Busuioc**

---

PhD thesis advisor Prof.dr. **Ioana Berindan-Neagoe**

---



**UMF**  
UNIVERSITATEA DE  
MEDICINĂ ȘI FARMACIE  
IULIU HAȚIEGANU  
CLUJ-NAPOCA

# TABLE OF CONTENTS

<b>INTRODUCTION</b>	13
<b>CURRENT STATE OF KNOWLEDGE</b>	
<b>1. Colorectal adenocarcinoma</b>	17
1.1 Incidence	17
1.2 Risk factors	17
1.3. Pathogenesis	18
1.3.1 Genetic alterations	19
1.3.2 Genomic and transcriptomic classification of colo-rectal cancer	20
1.4. Pathological classification of colo-rectal cancer	22
1.5 Morphological prognostic factors and the concept of tumor buds	25
1.5.1 Tumor budding	25
<b>2. Epithelial-mesenchymal transition (EMT)</b>	27
2.1 EMT: general data	27
2.2 EMT and therapy resistance	27
2.3 Molecular mechanisms of EMT	28
2.3.1 Transcription factors involved in EMT	29
2.3.2 miARN and EMT	30
2.4 The role of EMT in colo-rectal cancer	31
2.5 Immunohistochemical markers of EMT	32
<b>3. The MAPK signaling pathway and its role in colo-rectal cancer</b>	35
3.1 The MAPK pathway and EMT	37
<b>4. MMP3 și TESC gene involvement in colo-rectal cancer</b>	39
4.1 The MMP3 gene in colo-rectal cancer	39
4.2 The TESC gene in colo-rectal cancer	39
<b>PERSONAL CONTRIBUTIONS</b>	
<b>1. Work hypothesis/objectives</b>	43
<b>2. General methodology</b>	45
<b>3. Study 1. The signature of genes involved in Epithelial-Mesenchymal Transition and the link with prognosis in COAD</b>	47
3.1 Introduction	47
3.2 Materials and methods	48
3.3 Results	51

3.4 Discussion	61
3.5 Conclusions	64
<b>4. Study 2. Differential analysis of expressed genes, prognostic values and molecular pathways in colon adenocarcinoma: a bioinformatics approach with emphasis on MMP3 and TESC</b>	65
4.1 Introduction	65
4.2 Materials și methods	67
4.3 Results	71
4. Discussion	80
5. Conclusions	83
<b>5. Study 3. Epithelial-mesenchymal transition in colorectal cancer: an immunohistochemical study</b>	85
5.1 Introduction	85
5.2 Materials și methods	87
5.2 Results	89
5.4 Discussion	100
5.5 Conclusion	102
<b>6. General discution</b>	103
<b>7. General conclusions</b>	105
<b>8. Originality and innovative contributions of the thesis</b>	107
<b>REFERENCES</b>	109
<b>APPENDIX</b>	129

**Keywords: colorectal cancer, epithelial-mesenchymal transition, gene analysis, bioinformatics, qRT-PCR, prognosis, biomarkers, Vimentin, E-cadherin, OCT3/4**

# INTRODUCTION

Colo-rectal cancer is an important global public health problem, with an increasing incidence among young people and among the populations of developing countries, due to lifestyle changes following the westernization of these countries.

The clinical management of colo-rectal cancer has improved considerably in recent times. However, it still faces major limitations in terms of treatment, especially in the advanced stages, which are unfortunately the majority at the time of diagnosis. Furthermore, chemotherapy and radiation therapy face problems related to toxicity and resistance to therapy, while curative surgical treatment is limited to early cases.

The increasing incidence and problems related to therapy resistance force us to try to identify new molecular players involved in the pathogenesis of colo-rectal cancer, with potential roles as biomarkers or therapeutic targets, with the aim of expanding our therapeutic arsenal.

In this work, we focused our attention mostly on a fundamental biological mechanism, involved in physiological processes such as tissue healing or embryological development, but also in cancer progression, a process that is called the epithelial-mesenchymal transition. This complex mechanism induces migratory and invasive properties in tumor cells, making tissue invasion and metastasis possible, being practically responsible for the majority of adverse complications and deaths from cancer. Being a central and indispensable process for metastasis, it is tempting to believe that its better understanding will be able to highlight vulnerabilities that can lead to its inhibition and thus we might be able to deprive cancer of its most destructive features, offering additional therapeutic possibilities. Colon cancer is a remarkable candidate for this type of study, as it has a morphological correspondent of the epithelial-mesenchymal transition in the form of tumor buds. In addition, it is a very well-studied type of cancer both at the genomic and transcriptomic level, which makes it a very good candidate for bioinformatics studies that use numerous data sets available in public databases that were obtained from numerous genomic, transcriptomic and post-translational analyses that have gained momentum recently.

In the first part of the thesis, entitled "Current state of knowledge", we will present the theoretical framework of the issue of colorectal cancer and epithelial-mesenchymal transition, providing the latest demographic data and the most recent classifications and molecular features. In the second part of the thesis, entitled "Personal Contributions", we will present in detail 3 studies that use complex bioinformatic analyses, molecular biology tests, the evaluation of some clinical-histological parameters and the immunohistochemical expression of three proteins involved in the epithelial-mesenchymal transition in adenocarcinoma colonic.

The first study identified a distinct genetic signature of EMT consisting of two overexpressed genes (NOX4 and IGF2BP3) as well as five underexpressed genes (BMP5, DACT3, EEF1A2, GCNT2 and SFRP1) that demonstrated association with prognosis in colorectal cancer . A further qRT-PCR validation step was performed in a cohort of colorectal cancer patients comprising 29 tumor tissues and 29 adjacent normal tissues, confirming the expression level for BMP5 as well as two of the miRNAs - targeting key EMT-related genes, while highlighting the overexpression of miR-27a-5p and miR-146a-5p.

In the second study, using the UALCAN database, the top 25 over- and under-expressed genes in COAD were identified. Then, a Kaplan-Meier plot was used for these genes for survival analysis, revealing correlations with overall survival rate only for MMP3 and TESC, which are relevant targets that should be further validated as biomarkers for early diagnosis and therapeutic targets.

In the third study, surgical resection specimens from 30 colon cancer patients were evaluated and immunohistochemical assays were performed in order to investigate the expression of EMT-related markers and OCT 3/4 in tumor cells relative to normal tissue. Vimentin positivity, OCT3/4, and loss of E-cadherin in tumors were significantly associated with tumor grade, tumor budding, invasive tumor front, and lymph node metastasis. These antibodies could serve as a panel of biomarkers that could aid in the prognosis of patients, with the additional potential of being therapeutic targets.

## **PERSONAL CONTRIBUTIONS**

### **Work hypothesis**

As we showed in Chapter 1, COAD is a problem of high importance, with the potential to become more and more acute in the future. In the evolution of malignant epithelial tumors, as they acquire traits in a step-by-step fashion, that ultimately lead to invasion and metastasis, the moment when cells undergo EMT is probably the most important because of the determining effect that EMT has on cancer dissemination. On the other hand, as we showed in Chapter 2, EMT is also associated with therapy resistance, further contributing to the arsenal of malignant tumors. Under these conditions, we cannot help but wonder if this complex mechanism could be exploited and turned into a cancer vulnerability in the future.

We started from the premise that EMT is a program of central importance in the evolution of cancer and that it is an extremely complex biological program, which includes numerous molecular components located at multiple levels, both intracellular and extracellular. We primarily wanted to closely study this program to better understand the associated coding and non-coding genes, with a view to identifying new biomarkers or effective therapeutic targets. These genes or their transcripts can therefore have diagnostic, prognostic or therapeutic value. Therefore, in the first study we wanted to evaluate the gene signature involved in TEM in COAD and how the expression of these genes correlates with patients' prognosis. An additional objective was to validate the gene signature by qRT-PCR in tumor and normal tissues.

The second main objective was to identify based on the study of the most expressed or underexpressed 25 genes in COAD, genes that correlate with prognosis. After two genes (TESC and MMP3) were identified, we proposed as an additional objective bioinformatic analyzes that would integrate available data from the genomic, transcriptomic and proteomic levels and to make the correlation with the available clinical parameters to understand the biology of these two genes.

Then we wanted to study EMT from an immunohistochemical point of view, to be able to visualize in COAD the differential expression of two established IHC markers in EMT (VIM and E-CAD) and how it correlates with the morphological indicator of TEM in COAD (tumor buds) and the rest of the relevant clinical-pathological parameters, which are associated with a negative prognosis. A third marker involved in stem cell biology (OCT3/4) was studied because it has been consistently shown to be involved in EMT as well.

## General methodology

The three studies were carried out in strict compliance with the deontological norms and the ethics of scientific research, after the approval of the Ethics Committee of the "Iuliu Hațieganu" University of Medicine and Pharmacy in Cluj-Napoca was obtained.

For bioinformatic analyses, in the first study, we initially used expression data from The Cancer Omics Atlas (TCOA) database, an integrative resource for cancer genomic, transcriptomic, and proteomic information that allows the user to perform various types of analysis on this information. Here we queried the TCOA database to show us a list of the top 20 genes mutated in COAD, both overexpressed and underexpressed, compared to non-tumor tissues. We then compared expression levels by stage using the bioinformatics tool GEPIA, a tool that was later also used to correlate survival data related to EMT genes in COAD. For the analysis of the network of genes involved in EMT and the representation of mRNA-miRNA interactions, we used String 11.0 and miRNE tools, respectively. Additionally, survival data for relevant miRNAs targeting key genes in EMT were analyzed using StarBase. Next, we performed an analysis of mutational patterns of EMT genes in COAD in 169 cancer studies based on three datasets, colossal cBioPortal. For the correlations between the EMT and COAD genes as well as between the BMP5 gene and its associated miRNAs, the CANCERTOOL and StarBase tools were used, respectively. To validate the genetic signature discovered by the bioinformatics study, 29 samples of normal and tumoral colonic tissue were analyzed by qRT-PCR.

The second study exclusively used bioinformatics tools that analyzed data from the TCGA, containing a cohort of 286 COAD tumors and 41 adjacent normal tissues. With the help of STARBASE we identified the first 25 over- and under-expressed genes, which were subsequently analyzed individually for correlations with survival. The two genes significantly associated with survival (MMP3 and TESC) were then evaluated for protein expression analysis based on data available from CPTAC (Clinical Proteomic Tumor Analysis Consortium). The signaling pathways in which these genes are involved were functionally evaluated with the software program IPA (ingenuity pathway analysis) to highlight the biological functions in which they are involved.

The third study included 30 cases of COAD that were histologically evaluated for parameters associated with poor prognosis (Stage T, Stage N, lympho-vascular invasion, type of lympho-vascular invasion, perineural invasion, type of tumor front, tumor grade, tumor buds, location of tumors and histological type). Then IHC tests were performed for VIM, E-CAD and OCT3/4 which were correlated with all these parameters. Data were statistically processed using SPSS version 23.0, running on a Windows 10 operating system. Means and standard deviations were calculated for descriptive statistics. For quantitative variables, medians and quartiles were calculated, and for qualitative variables, frequency and percentage were calculated.

Compared with quantitative data and according to data normality, Student's t-test (independent samples T-test) (for two groups with normally distributed data) and Mann-Whitney U (for non-normally distributed data) were used, respectively. Quantitative data were tested for normality (Shapiro-Wilk test) and homogeneity of variances using Levene's test. Fisher's exact test was used for the data. Probability of error less than 5% ( $p < 0.05$ ) was considered the threshold for statistical significance in all 3 studies.

## **Study 1. Epithelial-Mesenchymal Transition Gene Signature Related to Prognosis in COAD**

Colon adenocarcinoma (COAD) remains an important cause of cancer-related mortality worldwide. Epithelial-mesenchymal transition (EMT) is a key mechanism, promoting not only the invasive or metastatic phenotype but also resistance to therapy. Using bioinformatics approaches, we studied the alteration on EMT related genes and its implication on COAD prognostic based on public datasets. For the EMT mechanisms, two overexpressed genes were identified (NOX4 and IGF2BP3), as well as five downregulated genes (BMP5, DACT3, EEF1A2, GCNT2 and SFRP1) that were related to prognosis in COAD. A qRT-PCR validation step was conducted in a COAD patient cohort comprising of 29 tumor tissues and 29 normal adjacent tissues, endorsing the expression level for BMP5, as well as for two of the miRNAs targeting key EMT related genes, revealing upregulation of miR-27a-5p and miR-146a-5p. The EMT signature can be used to develop a panel of biomarkers for recurrence prediction in COAD patients, which may contribute to the improvement of risk stratification for the patients.

## **Study 2. Analysis of Differentially Expressed Genes, MMP3 and TESC, and Their Potential Value in Molecular Pathways in Colon Adenocarcinoma: A Bioinformatics Approach**

Despite the great progress in its early diagnosis and treatment, colon adenocarcinoma (COAD) is still poses important issues to clinical management. Therefore, the identification of novel biomarkers or therapeutic targets for this disease is important. Using UALCAN, the top 25 upregulated and downregulated genes in COAD were identified. Then, a Kaplan–Meier plotter was employed for these genes for survival analysis, revealing the correlation with overall survival rate only for MMP3 (Matrix Metalloproteinase 3) and TESC (Tescalcin). Despite this, the mRNA expression levels were not correlated with the tumor stages or nodal metastatic status. MMP3 and TESC are relevant targets in COAD that should be additionally validated as biomarkers for early diagnosis and prevention. Ingenuity Pathway Analysis revealed the top relevant network linked to Post-Translational Modification, Protein Degradation, and Protein Synthesis, where MMP3 was at the core of the network. Another important network was related to cell cycle regulation, TESC being a component of this. We should also not underestimate the complex regulatory mechanisms mediated by the interplay of the multiple other regulatory molecules, emphasizing the interconnection with molecules related to invasion and migration involved in COAD, that might serve as the basis for the development of new biomarkers and therapeutic targets.

## **Study 3. Epithelial-mesenchymal transition in colorectal cancer: an immunohistochemical study**

Although the clinical management of colorectal cancer has improved significantly, it faces an increasing incidence among young people and among those in developing countries. Moreover, the diagnosis occurs mostly in advanced stages, when therapeutic resources are limited. Therefore, we need new biomarkers for

diagnosis and therapeutic targets. The key event leading to invasion and metastasis is epithelial-mesenchymal transition (EMT), which can be studied using IHC markers. We followed the correlation of the expression of EMT-related markers (Vimentin and E-cadherin) and a stem cell marker (OCT 3/4) with the clinicopathological parameters of the tumors. In this study, surgical resection specimens from 30 neoadjuvant untreated colon cancer patients hospitalized between 2018 and 2021 were evaluated. Immunohistochemical assays were performed to investigate the expression of EMT-related markers and OCT 3/4 in tumor cells. As results, Vimentin, OCT3/4 positivity, and E-cadherin loss were significantly associated with tumor grade, tumor budding, invasive tumor front, and lymph node metastasis. In conclusion, Vimentin, E-cadherin, and OCT 3/4 could serve as a panel of biomarkers that can aid in patient prognosis, with the added potential of being therapeutic targets.

## General conclusions

- a) In addition to the adaptive role that EMT plays in cancer biology and its involvement in invasion, metastasis and resistance to therapy, we demonstrated that the signature of genes involved in EMT correlates with prognosis in COAD.
- b) The seeding of genes involved in EMT shows complex interactions with miRNAs, which could be of clinical importance.
- c) The signature of genes involved in EMT support tumor progression through interactions with major signaling pathways, suggesting that they may represent therapeutic targets in addition to their role as prognostic markers.
- d) Among the most frequently mutated upregulated and downregulated genes in COAD, two key genes (TESC and MMP3) correlated with survival.
- e) In addition to their potential to be used as biomarkers, our studies have described the interactions and biological processes in which these two genes are involved, opening the door to investigate the regulatory mechanisms involved in COAD:
- e) Due to the little information available regarding MMP3 and TESC and considering their huge potential as a biomarker and therapeutic target respectively, it is necessary to study them in large cohorts of patients for their validation and correlation with clinicopathological features of cancer, both in COAD as well as other cancers.
- f) Increased IHC expression of Vimentin in COAD correlated in our study with tumor grade, T category, category, N, number of metastatic lymph nodes, type of vascular invasion, invasive tumor front and tumor buds.
- g) Loss of IHC expression for E-cadherin was statistically significantly correlated with tumor grade, tumor budding, larger N categories and invasive tumor front, inverse aspect to Vimentin, supporting literature data on EMT immunohistochemistry.
- h) OCT3/4, a stem cell marker showed an IHC expression profile similar to that of Vimentin (expression correlated with tumor buds, invasive tumor front, tumor grade and number of positive lymph nodes), suggesting that it can be used as a marker of EMT.
- i) IHC expression of the 3 markers could be integrated into the current diagnostic practice for the assessment of prognosis in COAD.