**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

**Centro de Biotecnologia**

**Programa de Pós-Graduação em Biologia Celular e Molecular – PPGBCM**

**TESE DE DOUTORADO**

# ESTUDO DE VARIANTES TRANSCRICIONAIS NO CÂNCER COLORRETAL E NA METÁSTASE HEPÁTICA

**Joice de Faria Poloni**

**Porto Alegre, novembro de 2018.**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

**Centro de Biotecnologia**

**Programa de Pós-Graduação em Biologia Celular e Molecular – PPGBCM**

# Estudo de variantes transcricionais no câncer colorretal e na metástase hepática

**Joice de Faria Poloni**

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular (PPGBCM) da UFRGS como requisito parcial para obtenção do grau de Doutor em Biologia Celular e Molecular.

Orientador: Prof. Dr. Diego Bonatto

Porto Alegre, novembro de 2018.

## INSTITUIÇÕES E FONTES FINANCIADORAS


### Agências financiadoras

Conselho Nacional de Desenvolvimento Científico e Tecnológico CAPES (Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior).



### Instituição de origem

Laboratório de Biologia Molecular e Computacional

Centro de Biotecnologia da UFRGS

Universidade Federal do Rio Grande do Sul

## AGRADECIMENTOS

Começo os meus agradecimentos pelos meus pais e meu irmão, que sempre me incentivaram a continuar meus estudos e a fazer o que me faz feliz. Obrigada por seu amor incondicional, por sempre entenderem os meus momentos de ausência e por comemorarem as minhas conquistas.

Ao meu marido Bruno pelos momentos de apoio, pela compreensão, e por sempre me segurar quando eu estava prestes a cair. Obrigada por sempre insistir em mim, e por me ensinar todos os dias o que é amar verdadeiramente.

À minha sogra Heloísa por todos os conselhos e amparo nos momentos difíceis.

Ao Ivan por todo o apoio que me deste, e por me ajudar a conquistar um dos meus sonhos mais valiosos.

Ao meu orientador Diego Bonatto por me aceitar como aluna e acreditar no meu potencial. Obrigada pelo aprendizado acadêmico e por tantas outras lições em todos esses anos.

Aos meus amigos de tantos anos Edna, Tahila, Nicolas, Manuel e Lê, por fazerem parte da minha vida, sendo sempre tão carinhosos e atenciosos.

Aos meus amigos de laboratório, vocês são como uma família pra mim e jamais conseguirei expressar todo o sentimento que tenho por vocês Baldi, Beatriz, Chapola, Gabino, Kendi, Larissa e Raquel. Obrigada por dividirem comigo tantas vitórias e aflições da vida acadêmica e pessoal nesses anos.

Aos professores Márcio Dorn e Hugo Verli pelo suporte em etapas cruciais do trabalho, e ao professor Guido Lenz pelo compartilhamento de seus conhecimentos.

Ao Luís Fernando por sempre fazer o possível para me ajudar em tantos momentos complicados.

Por fim, agradeço a todos que colaboraram no decorrer da minha trajetória acadêmica, sempre me incentivando a buscar mais conhecimento e a ser a melhor profissional que eu posso ser.

## ESTRUTURA DA TESE

Esta tese é organizada em introdução geral, justificativa, objetivos (gerais e específicos), dois capítulos redigidos em forma de artigo, sendo o primeiro capítulo uma revisão publicada sobre o tema e o segundo capítulo um artigo de dados a ser submetido, discussão geral, conclusões e perspectivas.

A introdução consiste em uma apresentação do câncer colorretal e os principais avanços no entendimento desta doença. Adicionalmente, a introdução é seguida de uma abordagem sobre alguns aspectos gerais da metástase e o ambiente hepático que as células disseminadas encontram quando migram em direção ao nicho pré-metastático. Por fim, um tópico final disserta sobre o mecanismo de *splicing* alternativo e o impacto de sua desregulação no câncer.

Após a introdução, a justificativa sintetiza as razões da escolha do tema desta tese, seguida dos objetivos propostos.

O capítulo 1 consiste de uma revisão sobre o impacto do *splicing* alternativo na metástase, e aponta os principais genes descritos na literatura por apresentarem uma alteração do mecanismo e regulação do *splicing*. Além disso, são descritas algumas isoformas transcricionais que estão diretamente relacionadas à vantagem que algumas células tumorais apresentam, resultando na aquisição do potencial metastático.

O capítulo 2 aborda as análises do uso diferencial de isoformas no câncer colorretal entre as condições normal, tumor primário e metástase hepática. Também é realizada uma predição de consequências e eventos de *splicing* associados à expressão destas isoformas. Adicionalmente, é feita uma análise da presença de mutações somáticas nas amostras de tumor primário e metástase com o objetivo de identificar possíveis alterações genéticas que possam estar associadas às isoformas transcricionais.

Os capítulos são seguidos de uma discussão geral dos resultados obtidos, interligando as isoformas descritas na revisão (Capítulo 1) com os resultados obtidos no Capítulo 2, conclusão do tema e perspectivas.

**SUMÁRIO**

## LISTA DE ABREVIATURAS

| | |
|---|---|
| ADAMTS2 | *ADAM Metallopeptidase With Thrombospondin Type 1 Motif 2* |
| ADAR | *Adenosine Deaminase, RNA Specific* |
| AKT | *AKT Serine/Threonine Kinase* |
| ANGPT1 | *Angiopoietin I* |
| APC | *Adenomatosis Polyposis Coli Tumor Suppressor* |
| BRAF | *B-Raf Proto-Oncogene, Serine/Threonine Kinase* |
| CALD1 | *Caldesmon 1* |
| CCHNP | Síndrome de câncer colorretal hereditário não-poliposo |
| CCL14 | *C-C Motif Chemokine Ligand 14* |
| CCND1 | *Cyclin D1* |
| CCR | Câncer colorretal |
| CD44 | *Extracellular Matrix Receptor III* |
| CDK14 | *Cyclin Dependent Kinase 14* |
| CDK8 | *Cyclin dependent kinase-8* |
| CDKN2C | *Cyclin Dependent Kinase Inhibitor 2C* |
| CEA | *Carcinoembryonic Antigen* |
| CLEC12A | *C-Type Lectin Domain Family 12 Member A* |
| CLK1 | *CDC Like Kinase 1* |
| c-Myc | *MYC Proto-Oncogene, BHLH Transcription Factor* |
| COL18A1 | *Collagen Type XVIII Alpha 1 Chain* |
| COL4A2 | *Collagen Type IV Alpha 2 Chain* |
| COL11A1 | *Collagen Type XI Alpha 1 Chain* |
| COL12A1 | *Collagen Type XII Alpha 1 Chain* |
| COX-2 | *Cyclooxygenase-2* |
| CP | *Ceruloplasmin* |
| CPEB2 | *Cytoplasmic Polyadenylation Element Binding Protein 2* |
| CTCA | *Cancer Treatment Centers of America* |
| CTTN | *Cortactin* |
| CXCL12 | *C-X-C Motif Chemokine Ligand 12* |
| DES | *Desmin* |
| DUSP10 | *Dual Specificity Phosphatase 10* |

| | |
|---|---|
| EGF | *Epidermal Growth Factor* |
| EPCAM | *Epithelial Cell Adhesion Molecule* |
| GSK-3β | *Glycogen synthase kinase-3β* |
| FAP | *Fibroblast Activation Protein Alpha* |
| hnRNP | *Heterogeneous nuclear ribonucleoproteins* |
| hnRNP K | *Heterogeneous Nuclear Ribonucleoprotein K* |
| hnRNP I | *Heterogeneous Nuclear Ribonucleoprotein I* |
| ICAM-1 | *Intercellular Adhesion Molecule 1* |
| ITGA5 | *Integrin Subunit Alpha 5* |
| ITGAE | *Integrin Subunit Alpha E* |
| ITGB1BP1 | *Integrin Subunit Beta 1 Binding Protein 1* |
| KRAS | *Kirsten Rat Sarcoma Viral Oncogene Homolog* |
| LAMA1 | *Laminin Subunit Alpha 1* |
| MAPK | *Mitogen-activated protein kinase* |
| MGMT | *O-6-Methylguanine-DNA Methyltransferase* |
| MIF | *Macrophage Migration Inhibitory Factor* |
| MLH1 | *MutL Homolog 1* |
| MMPs | *Matrix Metallopeptidase* |
| MSH2 | *MutS Homolog 2* |
| MSH6 | *MutS Homolog 6* |
| MSR1 | *Macrophage Scavenger Receptor 1* |
| MST1L | *Macrophage Stimulating 1 Like* |
| MUTYH | *MutY DNA Glycosylase* |
| NF-κB | *Nuclear Factor Kappa B Subunit 1* |
| NK | *Natural killer* |
| NO | *Nitric oxide* |
| PAF | Polipose adenomatosa familiar |
| PAM | Polipose associadas ao gene MUTYH |
| PDK1 | *Pyruvate Dehydrogenase Kinase 1* |
| PI3K | *Phosphoinositide-3 Kinase (PI3K)* |
| PIK3CA | *Phosphoinositide-3 Kinase Catalytic Subunit Alpha* |
| PH_3 | *pleckstrin homology domain* |

| | |
|---|---|
| PMS2 | *PMS1 Homolog 2, Mismatch Repair System Component* |
| PTEN | *Phosphatase and Tensin Homolog* |
| RB1 | *RB Transcriptional Corepressor 1* |
| ROS | *Reactive oxygen species* |
| SEER | *Surveillance, Epidemiology, and End Results Program* |
| SERPINA1 | *Serpin Family A Member 1* |
| SERPIND1 | *Serpin Family D Member 1* |
| SFRP1 | *Secreted Frizzled Related Protein 1* |
| SJF | Síndrome juvenil familiar |
| SLUG | *Snail Family Transcriptional Repressor* |
| SMAD2 | *SMAD Family Member 2* |
| SMAD4 | *SMAD Family Member 4* |
| SNAIL | *Snail Family Transcriptional Repressor* |
| SPJ | Síndrome de Peutz-Jeghers |
| SR | *Serine And Arginine Rich* |
| SRCR | *Scavenger receptor cysteine-rich* |
| SRPK | *Serine/Arginine-Rich Protein-Specific Kinase* |
| SRSF1 | *Serine And Arginine Rich Splicing Factor 1* |
| SRSF6 | *Serine And Arginine Rich Splicing Factor 6* |
| SRSF9 | *Serine And Arginine Rich Splicing Factor 9* |
| Src | *SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase* |
| ST | Síndrome de Turcot |
| ST6GALNAC6 | *ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 6* |
| STK11 | *Serine/Threonine Kinase 1* |
| STPH | Síndrome do tumor PTEN-hamartoma |
| TCF | *Transcription Factor* |
| TGF-β | *Transforming Growth Factor Beta 1* |
| TNC | *Tenascin C* |
| TP53 | *Tumor Protein P53* |
| TPM1 | *Tropomyosin 1* |
| TRAIL | *TNF-related apoptosis-inducing ligand* |
| TWIST | *Twist Family BHLH Transcription Factor* |

| | |
|---|---|
| VCAM-1 | *Vascular Cell Adhesion Molecule 1* |
| VEGF | *Vascular Endothelial Growth Factor* |
| WNT | *Wingless-Type* |
| WNT2B | *Wnt Family Member 2B* |
| ZEB | *Zinc Finger E-Box Binding Homeobox* |

## RESUMO

O desenvolvimento do câncer colorectal (CCR) é frequentemente derivado de lesões não-cancerosas chamadas de pólipos que podem evoluir para um estado maligno quando não ressectados. Grandes avanços têm sido feitos na tentativa de compreender as diferenças adquiridas pelas células malignas, como a identificação de alterações genéticas e epigenéticas, modificações do perfil transcricional e a influência do microambiente. Com essas observações, foi possível definir as características presentes em uma condição maligna, as quais são definidas como as marcas do câncer. Entretanto, apesar dos avanços no conhecimento propiciarem melhoras no diagnóstico e no tratamento, os pacientes com CCR metastática apresentam uma menor chance de cura pelos tratamentos atuais e, por consequência, uma baixa taxa de sobrevida. Para as células adquirirem a capacidade de disseminar a partir do tumor primário e metastatizar, é necessário a aquisição de um potencial iniciador da metástase, que confere a habilidade de sobreviver em condições extremas, migrar e colonizar outros tecidos. No caso do CCR, a metástase hepática é mais frequente em relação aos outros tecidos devido às características anatômicas do fígado. Dentre todos os mecanismos estudados na caracterização da doença, o estudo do *splicing* alternativo tem ganhado espaço nas pesquisas oncológicas, na qual distintas isoformas provenientes de um mesmo gene podem apresentar diferentes atividades, contribuindo na obtenção das características celulares necessárias para o desenvolvimento tumoral e para as células malignas disseminarem e colonizarem outros tecidos. Nesta tese de doutorado, foram analisadas as isoformas de mRNA diferencialmente usadas entre os estágios do cólon normal, lesão primária e metástase hepática. Os principais processos alterados pela desregulação do mecanismo de *splicing* foram adesão celular, migração, sobrevivência e proliferação, morte celular, e resposta imune. Esses resultados demostram a importância de aprofundar o estudo dos mecanismos de *splicing* e as consequências associadas à troca das isoformas entre condições normais e patológicas, assim como das descobertas de potenciais biomarcadores e posterior desenvolvimento de novas terapias.

## ABSTRACT

Colorectal cancer (CRC) development is usually derived from noncancerous lesions called polyps that can progress to a malignant state when not resected. Major advances have been made in the attempt to understand the differences acquired by malignant cells, such as the identification of genetic and epigenetic alterations, modifications of the transcriptional networks, and the influence of the microenvironment. Thus, it was possible to characterize the main molecular mechanisms present in the malignant condition, which are defined as the hallmarks of cancer. Although the advances in knowledge provide improvements in the diagnosis and treatment of cancer, patients with metastasis have lower survival rates by standard treatments. Disseminated and metastatic cells from primary tumor carry a metastasis-initiating potential that confers the ability to survive under extreme conditions, to migrate and to colonize other tissues. In the case of CRC, liver metastasis is more frequent than to other tissues due to anatomical characteristics. Among all mechanisms considered in the characterization of cancer, the study of alternative splicing has gained space in oncological analyzes, where distinct isoforms from the same gene can present different activities. Thus, the abnormal splicing mechanism may generate aberrant isoforms that may contribute to tumor development. In this thesis, isoforms that are differentially used between normal epithelium, primary lesion, and hepatic metastasis were analyzed. It was observed that the deregulation of splicing mechanisms majorly impacted in the bioprocesses of cell adhesion, migration, survival and proliferation, cell death, and immune response. These results show the importance of deepening the study of splicing mechanisms and the consequences associated with the isoforms exchange between normal and pathological conditions, becoming potential biomarkers and new therapeutic targets.

# INTRODUÇÃO

## 1. Aspectos epidemiológicos do câncer colorretal

O câncer colorretal (CCR) é o nome dado para todos os tumores que acometem regiões do cólon e/ou reto e são frequentemente agrupados por apresentarem características em comum. No mundo inteiro, o CCR é o terceiro tipo de câncer mais comum, a segunda principal causa de morte por câncer e a primeira causa de morte por cânceres do trato gastrointestinal [1]. Adicionalmente, a prevalência é maior em mulheres e em indivíduos entre 65 e 74 anos de idade [1].

No Brasil, de acordo com as estimativas do Instituto Nacional do Câncer de 2018, o CCR é o terceiro mais frequente em homens e o segundo entre as mulheres [2]. A incidência de CCR no Brasil varia conforme a região estudada, sendo mais prevalente nos estados de Goiás, Mato Grosso do Sul, Paraná, Rio Grande do Sul, Rio de Janeiro e São Paulo, com incidência de 17,89 a 27,74 casos por 100.000 habitantes nos homens e de 18,18 a 23,57 casos por 100.000 habitantes nas mulheres [2]. Além disso, são estimados para 2018 e 2019, 17.380 casos novos em homens e 18.980 em mulheres [2].

Dentre os principais fatores de risco estão a obesidade, o sedentarismo, as doenças inflamatórias intestinais, os fatores genéticos, presença de pólipos e adenomas, os maus hábitos nutricionais (como a alta ingestão de lipídeos e proteínas), o tabagismo e a idade avançada [1].

Os principais sintomas do CCR incluem: dor abdominal, perda de peso, náusea, vômito, mal-estar, anorexia, distensão abdominal, e alteração de hábitos e movimentos intestinais [1]. O CCR em regiões distais, pode causar sangramento retal, que, quando misturado com as fezes, pode não ser percebido, e por consequência, levar à anemia como sintoma secundário [1].

O estágio de detecção do CCR é fundamental para determinar o prognóstico, sobrevivência e tratamento do paciente. A identificação do CCR em estágios iniciais é crucial para o sucesso do tratamento. Contudo, dados dos Estados Unidos e Europa mostram que mais de 30% dos pacientes já apresentavam metástase quando diagnosticados [3]. A principal modalidade de tratamento para a doença é a ressecção cirúrgica, que pode ou não estar associada com a quimioterapia, ou radioterapia, dependendo do estadiamento do tumor [4]. Infelizmente, mesmo quando diagnosticado nos estágios iniciais, a recorrência do tumor ainda é alta (cerca de 40 a 50%) [5].

Atualmente, o melhor exame diagnóstico para o CCR é a colonoscopia completa com biópsia do tecido para exames histológicos, que permite também a excisão endoscópica dos pólipos de forma preventiva [6]. Após o diagnóstico do CCR, as taxas de sobrevida em 5 anos é de 65%, enquanto que em 10 anos a sobrevida cai para 58% [1]. Contudo, essa informação é relativa, e não considera determinados fatores como a idade, o tipo histológico, o estágio da doença, ou ainda as modificações genéticas relacionadas com a agressividade do tumor.

## 2. História do CCR

A primeira descrição do CCR na literatura foi em maio de 1927 por Lockhart-Mummery e Dukes, em um trabalho intitulado "*The precancerous changes in the rectum and colon*" [7]. Neste estudo, os autores relacionaram o tecido adenomatoso com o CCR, estabelecendo o conceito de que o CCR origina-se a partir de uma lesão pré-existente [7]. Durante a década de 1930, foi desenvolvido o sistema de estadiamento para o CCR e foi demonstrado que pacientes diagnosticados e submetidos a procedimentos cirúrgicos em estágios iniciais apresentavam uma melhor sobrevida [7].

Posteriormente, Morson em 1968 propôs o termo sequência adenoma-carcinoma (ou pólipo-câncer), que descreve que a base morfológica do surgimento do CCR ocorre a partir de pólipos adenomatosos (**Figura 1**) [8]. Em 1993, o Estudo Nacional de Pólipos (*National Polyp Study*) sustentou a ideia da progressão do adenoma colorretal para adenocarcinoma e mostrou que a colonoscopia com polipectomia reduz a incidência de câncer colorretal, demostrando efetivamente a importância da remoção dos pólipos adenomatosos durante a colonoscopia [9]. No entanto, apesar de 95% dos casos de CCR serem derivados de adenomas, já foi observado a existência de lesões "de novo" como uma forma atípica do carcinoma que não apresenta o adenoma como precursor [10,11]. Nesse caso, o carcinoma se desenvolve diretamente do epitélio normal a partir de lesões pequenas e não polipoides, sem passar pelo estágio de adenoma [10].

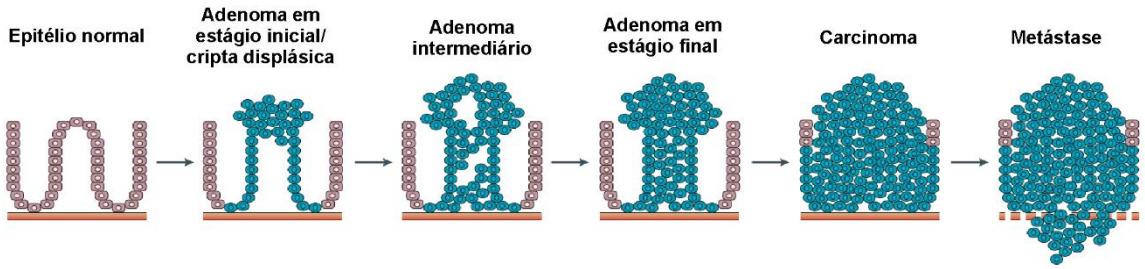



**Figura 1**. Sequencia adenoma-câncer.[1]

Em 1988, Vogelstein demonstrou que alterações genéticas no gene *Ras* eram frequentes em 58% dos adenomas maiores que 1 cm e em 47% dos carcinomas [12]. Essas descobertas levaram Fearon e Vogelstein a propor um modelo de carcinogênese que propõe que mutações em oncogenes e genes supressores de tumor seguem uma determinada ordem, culminando no CCR [13]. No entanto, os autores ressaltam que, embora exista uma ordem de alterações genéticas, a acumulação das mutações é mais consistente com as observações clínicas e histopatológicas [13].

Durante o desenvolvimento do CCR, diferentes mutações genéticas e epigenéticas em oncogenes e genes supressores tumorais conferem às células provenientes do trato gastrointestinal a capacidade de hiperproliferarem e de auto-renovarem-se, dando origem subsequentemente ao adenoma benigno [6]. Quando não diagnosticado e tratado, o adenoma benigno pode progredir para um carcinoma e posteriormente para a metástase [6].

É estimado que em torno de 70% de todos os CCR são esporádicos, ou seja, ocorrem devido à presença de mutações somáticas [6]. Além disso, 10 a 30% dos casos de CCR não são decorrentes de doença mendeliana, mas sim uma predisposição genética ao desenvolvimento do carcinoma [6]. Mutações em determinados genes (serão discutidos adiante) são causadores de síndromes de CCR hereditárias e correspondem a 5–7% dos casos [6]. Dentre as síndromes hereditárias de CCR podemos citar: polipose adenomatosa familiar, síndrome de câncer colorretal hereditário não-poliposo (também conhecida como síndrome de Lynch), síndrome de Peutz-Jeghers, síndrome de Turcot, polipose associadas ao gene *MUTYH*, síndrome do tumor PTEN-hamartoma, e síndrome juvenil familiar.

[1] Imagem modificada de Davies RJ, et al. Nat Rev Cancer. 2005; 5(3):199–209.

## 3. Mecanismos moleculares envolvidos no CCR

Quando consideramos o CCR ao nível molecular, as causas que dão origem ao seu desenvolvimento e progressão são muito heterogêneas. Estudos já comprovaram a presença de mutações somáticas em mais de 100 genes, com uma média de 80 mutações em um único caso de CCR [14]. A perda da integridade genômica contribui muito para a acumulação de mutações, que como visto anteriormente, é a principal causa da formação do CCR. Algumas mutações são capazes de conduzir o desenvolvimento tumoral, e são chamadas mutações *drivers* [15]. Contudo, ao longo das divisões neoplásicas, centenas de mutações *passengers* ocorrem devido ao funcionamento anormal da célula [15]. Muitos genes são descritos por apresentarem alta frequência de mutações no CCR. Dentre estes genes, os mais representativos para o desenvolvimento do CCR estão descritos na **Quadro 1** [5,14,16].

**Quadro 1**. Descrição dos principais genes cujas mutações estão fortemente associadas com CCR. PAF, polipose adenomatosa familiar; CCHNP, síndrome de câncer colorretal hereditário não-poliposo; SPJ, síndrome de Peutz-Jeghers; ST, síndrome de Turcot; PAM, polipose associadas ao gene *MUTYH*; STPH, síndrome do tumor PTEN-hamartoma; SJF, síndrome juvenil familiar.

| Genes | Relação com CCR | Síndrome | Mecanismos | Ref. |
|---|---|---|---|---|
| ***APC*** | Mutações inativantes estão relacionadas com aumento da atividade da via WNT e com a perda da estabilidade dos microtúbulos durante a separação das cromátides-irmãs | PAF, ST | Instabilidade cromossômica | [6,16] |
| ***TP53*** | Inativação deste gene está associada com perda da regulação do ciclo e morte celular | - | Instabilidade cromossômica | [16,17] |
| ***KRAS*** | Oncogene relacionado com o crescimento celular | - | Instabilidade cromossômica | [14,16] |
| ***BRAF*** | Oncogene relacionado com o crescimento celular | - | Fenótipo metilador de ilhas CpG, instabilidade de microssatélites | [16,18] |
| ***PIK3CA*** | Mutações neste gene podem levar à inibição da apoptose | - | Instabilidade cromossômica | [14,19] |
| ***MLH1*** | Atua no reparo de bases mal pareadas | CCHNP | Instabilidade de microssatélite | [6,19] |
| ***MSH2*** | Atua no reparo de bases mal pareadas | CCHNP | Instabilidade de microssatélite | [6] |

| | | | | |
|---|---|---|---|---|
| ***MSH6*** | Atua no reparo de bases mal pareadas | CCHNP | Instabilidade de microssatélite | [6,19] |
| ***PMS2*** | Atua no reparo de bases mal pareadas | CCHNP | Instabilidade de microssatélite | [6] |
| ***EPCAM*** | Mutações em 3' levam ao silenciamento epigenético de *MSH2* | CCHNP | - | [14] |
| ***TGF-β*** | É um supressor tumoral | - | Instabilidade de microssatélite | [14] |
| ***PTEN*** | É um supressor tumoral que regula negativamente a via PI3K/AKT | STPH | - | [6,20] |
| ***MUTYH*** | Pertence ao grupo de genes atuantes no reparo por excisão de bases, e mutações germinativas neste gene estão relacionadas com mutações somáticas em APC | PAM | Instabilidade cromossômica | [6,14] |
| ***STK11*** | Supressor tumoral | SPJ | - | [21] |

A progressão do CCR é seguida por uma série de alterações mostradas na **Figura 2**, e existem três principais mecanismos que, quando alterados de forma individual ou combinados, são comumente relacionados com a transformação maligna das células epiteliais do intestino: instabilidade cromossômica, fenótipo metilador de ilhas CpG (*CpG island methylator phenotype*) e instabilidade de microssatélite (**Figura 2**) [16].

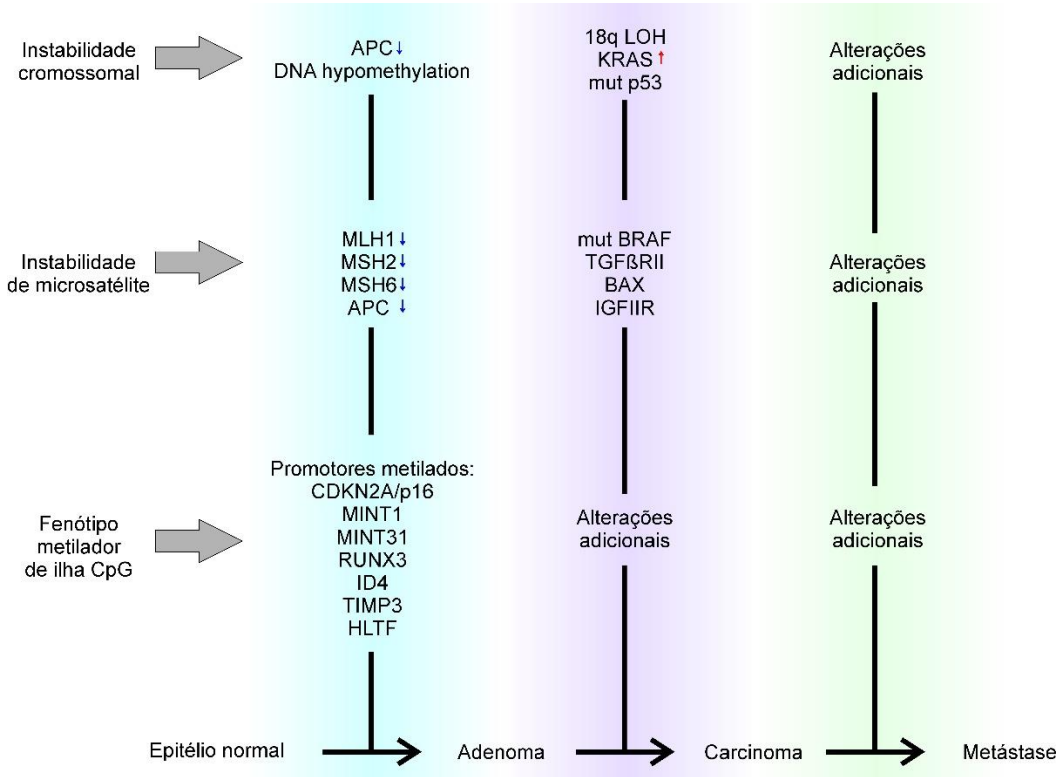


**Figura 2.** Sequência das alterações genéticas, epigenéticas e moleculares que ocorrem durante a progressão do CCR.

A principal causa da instabilidade cromossômica está associada aos defeitos durante o reparo a danos no DNA, anormalidades durante a segregação cromossômica e no funcionamento dos telômeros, além de mutações em oncogenes e genes supressores de tumor [16]. Adicionalmente, a instabilidade cromossômica está fortemente associada com 65 a 70% dos casos esporádicos de CCR, sendo principalmente relacionada com aneuploidia e perda de heterozigosidade [16]. Um dos principais genes envolvidos na instabilidade cromossômica é a inativação do gene *APC*, um bloqueador da via de sinalização de WNT [19].

Microssatélites são sequências mono- ou dinucleotídicas repetidas ao longo do genoma, e que devido a suas características, são muito vulneráveis a erros durante o processo de replicação [14]. A instabilidade de microssatélite é observada pela presença de mutações em genes atuantes na reparação de bases mal pareadas, como, por exemplo, os genes *MSH2*, *MSH6*, *MLH1* e *PMS2* [14,16]. A deficiência no reparo das sequências microssatélite é especialmente importante quando a sequência encontra-se dentro de genes codificantes, podendo alterar o produto gênico [14]. A instabilidade de microssatélite é observada em 15% dos tumores esporádicos e em >95% de casos de síndrome de câncer colorretal hereditário não-poliposo [19]. Essa síndrome é considerada uma doença autossômica dominante, e sugere-se que a síndrome não seja a única causa responsável pelo CCR, mas que o risco aumente consideravelmente quando uma mutação somática ocorre no alelo normal remanescente [14].

A metilação do DNA é um mecanismo de regulação da expressão gênica, promovendo o silenciamento do gene quando ocorre em regiões promotoras contendo ilhas CpG na sua extensão. Em aproximadamente 15% dos casos de CCR é observada o silenciamento da expressão gênica a partir da hipermetilação dos promotores, sendo tal fenômeno conhecido como fenótipo metilador de ilhas CpG [14]. Frequentemente, a hipermetilação de promotores está associada com a instabilidade de microssatélites, principalmente devido ao silenciamento de genes envolvidos no reparo de DNA, como *MGMT* e *MLH1*, mas também com mutações em *BRAF* [14,16].

### 3.1. Principais vias de sinalização relacionadas com o CCR

Durante o desenvolvimento do CCR, as vias de sinalização que regulam a homeostase das células epiteliais do intestino tornam-se desreguladas, independentemente dos sinais fornecidos pelo nicho das criptas [22]. Algumas vias de sinalização fortemente associadas ao CCR são WNT/APC/β-catenina, PI3K/AKT, TGF-β/Smad, NF-κB e Ras [6].

As alterações relacionadas com a via WNT estão frequentemente relacionadas à falha na degradação de β-catenina e por isso, ao aumento da expressão de genes alvos de TCF [19]. Muitos destes genes alvos têm importante papel na proliferação, diferenciação, migração e adesão como, por exemplo, c-Myc e CCND1, que atuam na progressão do ciclo celular durante a proliferação [19,23]. Frequentemente esta falha ocorre por mutações inativantes em *APC*, um componente do complexo que leva à degradação de β-catenina, proporcionando uma maior estabilidade de β-catenina e está associado à formação de adenomas [24]. Mutações no gene codificador de β-catenina, *CTNNB1*, que conferem maior estabilidade à proteína resultante e impedem a sua degradação estão relacionados com a hiperplasia intestinal [25]. Adicionalmente, a atividade de β-catenina pode estar indiretamente relacionada com mutações em outros genes. Um exemplo são amplificações no gene *CDK8*, cuja atividade elevada estimula β-catenina e Notch1, aumentando a transcrição e a diferenciação celular [19]. Muitos casos de CCR esporádicos apresentam mutações em elementos da via WNT, sendo que 85% das mutações ocorrem em *APC* e 10% em *CTNNB1* [26].

KRAS é um subtipo de proteína Ras e a principal atuante no CCR [27]. *KRAS* é responsável pela codificação de uma GTPase relacionada com a transdução e propagação de sinais extracelulares através da ativação da via da MAPK [19]. Mutações em *KRAS* levam à permanente ativação da via da MAPK, contribuindo para a resistência à apoptose e conferindo vantagens no crescimento celular [19]. *KRAS* mutado está altamente relacionado com mutações em PI3K. A ativação de PI3K promove ativação de AKT1 e AKT2, que contribuem para a transição epitelial-mesenquimal e aumento do crescimento tumoral [16]. O aumento da sinalização por PI3K resulta em maior atividade de COX-2, levando à inibição da apoptose [28]. COX-2 é responsável pela síntese de prostaglandina E2, que contribui para a formação de um microambiente inflamatório no cólon, capaz de induzir a expansão das células-tronco tumorais [29,30].

A sinalização mediada por NF-κB também está envolvida com o desenvolvimento tumoral e relacionada com processos inflamatórios, resposta imune, crescimento celular e sobrevivência, angiogênese e prevenção da apoptose [30,31]. Além disso, a ativação

constitutiva de NF-κB é observada em 40% das amostras tumorais humanas, promovendo a proliferação celular e resistência à quimioterápicos [32].

As mutações em *SMAD2* e *SMAD4* são comumente observadas em casos de CCR que apresentam alta instabilidade em microssatélites [19]. Esses genes fazem parte da via de sinalização mediado por TGF-β, que, uma vez ativados, promovem a expressão de muitos genes relacionados com a transição epitelial-mesenquimal, como *SNAIL*, *SLUG*, *TWIST* e *ZEB* [6,19]. Nas células disseminadas, é frequentemente observada uma maior sinalização de TGF-β, e isso está fortemente associado ao sucesso das células metastáticas em colonizar outro órgão [22].

## 4. Progressão do CCR

A heterogeneidade do CCR é dependente de vários fatores, como a ordem de aparecimento e acúmulo das mutações, os polimorfismos, a composição policlonal do tumor e a influência que o microambiente exerce sobre as células tumorais [5]. Além disso, as manifestações fenotípicas destas variações podem determinar a agressividade do tumor, e alteram-se de acordo com os hábitos individuais e influências externas [5]. Todos esses fatores são cruciais na progressão tumoral, que culmina com o desprendimento das células malignas e a sua disseminação para outros órgãos através dos vasos linfáticos e/ou sanguíneos [33]. Esse processo é chamado de metástase e é responsável por 90% das mortes associadas ao CCR em aproximadamente 5 anos (**Figura 3**) [34].

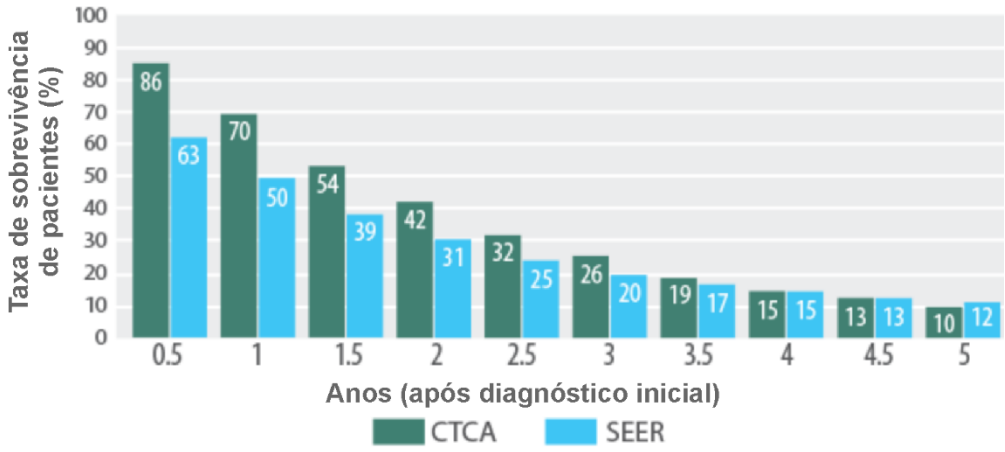



**Figura 3**. Taxa de sobrevivência de pacientes diagnosticados com CCR metastático acompanhados ao decorrer de 5 anos. Dados provenientes do *Cancer Treatment Centers of*

*America* (CTCA) e *Surveillance, Epidemiology, and End Results Program* (SEER), ambos localizados nos Estados Unidos.[2]

A metástase pode ser dividida em cinco etapas: desprendimento das células do tumor primário e invasão local através da matriz extracelular, intravasão, sobrevivência na circulação, extravasão e, por fim, a colonização de um órgão distante [35]. Contudo, a metástase é um processo extremamente ineficiente e a transformação oncogênica não é suficiente para que as células malignas tenham competência para metastatizar [35]. Para isso, as células tumorais precisam adquirir habilidades invasivas e de sobrevivência a partir de alterações em diversos processos biológicos, apresentando assim um "potencial formador de metástase", discutidos no **Capítulo 1**.

O cirurgião Stephen Paget propôs em 1889 uma hipótese que ele chamou de "a teoria semente e do solo" ("*seed and soil theory*") que sugere que as células tumorais (as "sementes") encontrem em determinados órgãos (o "solo") um ambiente mais propício à colonização [33]. Assim, determinados órgãos podem oferecer um ambiente mais favorável para o crescimento das células malignas disseminadas [33]. O microambiente tumoral é composto de diferentes tipos celulares (como células do estroma, efetoras do sistema imune, plaquetas, fibroblastos, células endoteliais) que, juntamente com o tumor secretam diferentes moléculas, como citocinas, proteases, fatores de crescimento e hormônios e fatores humorais, favorecendo a progressão do tumor neste contexto celular [34]. Muitos destes fatores secretados são circulantes na corrente sanguínea e podem atuar em locais diferentes do tumor primário, influenciando o microambiente de um órgão distante e o tornando mais permissivo ao crescimento das células circulantes tumorais, chamado nicho pré-metastático [36,37].

Dentre os principais órgãos alvos como sítio único de metástase do CCR está o fígado que abrange cerca de 50% dos casos e o pulmão, que representa 10 a 15% dos casos [38]. As células tumorais podem se disseminar com menor frequência para o peritônio, e em raros casos, para o cérebro, os ossos, as glândulas suprarrenais e o baço [38]. Quando a disseminação ocorre para um órgão único e a metástase é ressectável, existe um maior potencial curativo. Em alguns casos onde isso não é possível, pode-se usar a quimioterapia para a redução da massa tumoral, tornando-a ressectável após o tratamento [38,39].

---

[2] Imagem modificada de https://www.cancercenter.com/colorectal-cancer/statistics/tab/colon-cancer-survival-statistics/2017

## 5. Metá stase hepática

O fígado é o principal órgão para metástase de muitos tipos de tumores, incluindo o CCR (**Figura 4**) [40]. As principais razões para isso são as suas características morfofisiológicas: possui uma importante função na desintoxicação do sangue e também na biossíntese de proteínas e, por isso, recebe e processa cerca de 30% do volume sanguíneo perfusionado por minuto [40]. Para isso, recebe cerca de 20% de sangue oxigenado pela artéria hepática e quase 80% de sangue desoxigenado e rico em nutrientes por meio da veia porta [40]. Devido à grande quantidade de sangue que passa através do fígado, esse órgão é rico em uma rede de vasos sinusoides [40]. Esses vasos apresentam um fluxo sanguíneo mais lento e são recobertos pelo endotélio fenestrado sinusoidal [40,41]. Além disso, a composição celular do fígado compreende diferentes tipos celulares: células hepáticas estreladas, células endoteliais sinusoidais, células de Kupffer, fibroblastos e células do sistema imune [37].

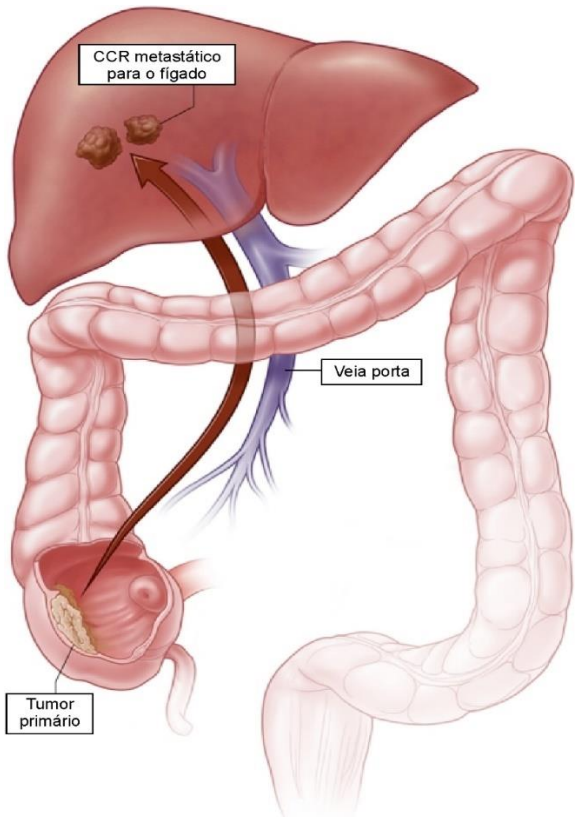


**Figura 4**. Progressão do tumor primário para metástase hepática através da veia porta.[3]

A metástase hepática pode ser dividida em diferentes fases: (*i*) pré-metastática, com estabelecimento de nicho no fígado em resposta a fatores secretados pelo tumor; (*ii*) fase microvascular, onde as células tumorais ficam aderidas e infiltram através dos vasos sinusóides; (*iii*) fase extravascular, quando ocorre a extravasão das células tumorais infiltrantes para o espaço de Disse e ativação da resposta local; (*iv*) fase angiogênica, com formação de micrometástase e vascularização; e (*v*) fase de crescimento, com expansão da metástase [37].

As células tumorais circulantes do CCR disseminam-se por meio da circulação mesentérica, sendo o fígado o primeiro órgão que encontram. Devido à fenestração dos vasos hepáticos associado a uma membrana basal pouco organizada, o fígado se torna permissivo para a extravasão das células metastáticas [40]. Porém, além da alta acessibilidade, o estabelecimento da metástase é um processo onde a seletividade das células tumorais pelo tecido alvo é também sustentada pela presença de fatores que contribuam para a adesão, extravasão e colonização [42].

As células da lesão primária secretam exossomas contendo MIF, que é fortemente correlacionado com a formação do nicho pré-metastático e o desenvolvimento da metástase hepática [37,43]. Esses exossomas estimulam a produção de TGF-β pelas células de Kupffer, que levam a produção de fibronectina pelas células estreladas e recrutamento de macrófagos derivados da medula óssea (**Figura 5**) [37,43]. Além disso, os exossomas que chegam ao fígado se fusionam preferencialmente com as células de Kupffer através das integrinas αvβ5, α6β1 e α6β4, levando a superexpressão de fatores pró-inflamatórios e contribuindo para a formação do nicho pré-metastático [37]. Além dos exossomos, algumas linhagens de CCR secretam CEA, que também leva à ativação das células de Kupffer [40]. A interação de CEA ao seu receptor nas células de Kupffer induz a secreção das citocinas pró-inflamatórias IL-1β, TNF-α e IL-6 [40].

As células tumorais circulantes que param nos vasos podem ser destruídas pelo estresse mecânico, fagocitadas pelas células de Kupffer ou sofrer citólise pela liberação de granzimas e perforinas das células NK, além de outras moléculas liberadas pelas células endoteliais

[3] Imagem modificada de https://www.hopkinsmedicine.org/liver_tumor_center/conditions/cancerous_liver_tumors/colorectal_liver_metastases.html2018

sinusoidais, como NO e ROS [37]. Um dos meios das células metastáticas escaparem dos efeitos citotóxicos e protegerem-se do estresse hemodinâmico é a formação de agregados com outras células tumorais e com células sanguíneas, como as plaquetas [35,37]. Além de proteger a célula metastática, esses agregados formados com plaquetas contribuem para a adesão das células tumorais ao tecido endotelial dos vasos sanguíneos devido à expressão de fatores de adesão, que também cooperam para a migração extravascular [44]. Uma das medidas usadas para evadir da morte celular mediada por células NK é a resistência à morte induzida por TRAIL através da ativação de Src (**Figura 5**).

Embora a resposta inflamatória inicial possa levar muitas células tumorais à morte, a inflamação pode aumentar a expressão de moléculas de adesão nas células endoteliais sinusóides, contribuindo para a migração transendotelial das células tumorais para o espaço de Disse (**Figura 5**) [37]. Dentre as moléculas adesivas expressas nas células endoteliais em resposta ao processo inflamatório estão E-selectina, VCAM-1 e ICAM-1 [37]. Adicionalmente, já foi mostrado o importante papel das integrinas αvβ5, αvβ6, α6β1 e α6β4 na adesão entre as células tumorais circulantes com as células endoteliais sinusoidais, onde as integrinas α6β1 e α6β4 também contribuem para a extravasão [42].

As células tumorais que sobrevivem ao ambiente hostil e conseguem chegar ao espaço de Disse, precisam adaptarem-se para garantir a sua sobrevivência. A colonização dessas células é limitada pelos sinais provenientes do microambiente, à pressão do sistema imune e também à falta de vascularização apropriada para garantir nutrientes (**Figura 5**) [40]. Consequentemente, as células metastáticas modificam o ambiente ao qual estão inseridas, adquirindo estímulo mitogênico das citocinas e fatores de crescimento que estão sendo produzidos [40]. Neste sentido, ocorre um equilíbrio entre o crescimento e a morte desta população de células [40].

As células estreladas hepáticas ativadas e as células de Kupffer liberam diferentes fatores em resposta às injúrias e o processo inflamatório estabelecido, iniciando uma resposta de reparo ao tecido hepático [37]. Essa resposta leva ao aumento de colágeno e fatores de crescimento e pró-angiogênicos, como EGF, VEGF e ANGPT1, além da elevar a produção de MMPs [37,40]. Esses fatores contribuem para a remodelagem da matriz extracelular e para a angiogênese, cooperando ainda mais para a invasão das células metastáticas e com o processo de colonização e expansão da metástase [37,40].

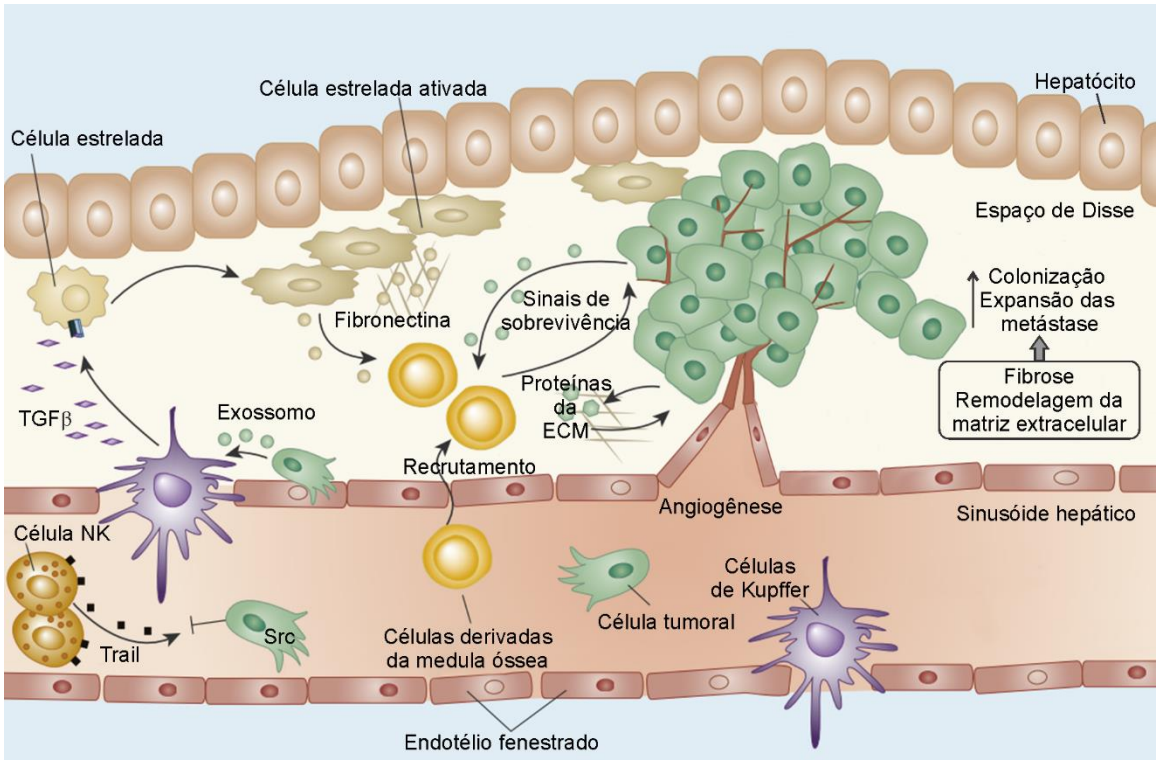



**Figura 5**. Representação da metástase hepática. As células do tumor primário secretam exossomos que contribuem para a formação do nicho pré-metastático. Os exossomos podem conter MIF e estimulam uma resposta pró-inflamatória, levando a produção de TGF-β pelas células de Kupffer, produção de fibronectina pelas células estreladas e ao recrutamento de células derivadas da medula óssea. Quando as células metastáticas chegam ao vaso sinusóide, elas podem ser destruídas pela presença de células NK ou evadir deste mecanismo através da ativação de Src. Contudo, a inflamação local também estimula a expressão de moléculas de adesão nas células endoteliais, facilitando a migração transendotelial das células metastáticas e a posterior entrada no espaço de Disse. Devido ao processo inflamatório local e às lesões teciduais, as células de Kupffer iniciam um processo de reparo, aumentando a produção de fatores de crescimento e de remodelagem da matriz extracelular. Por consequência, esse processo contribui para o estabelecimento e expansão da metástase.[4]

## 6. *Splicing* alternativo e sua importância no câncer

[4] Modificado de Obenauf AC & Massagué J. Trends in Cancer. 2015; 1(1):76–91.

A progressão do adenoma inicial até a metástase é o resultado de uma sequência de alterações genéticas e epigenéticas a qual uma célula maligna sofre, acumulando mutações em diferentes genes que lhe irão conferir resistência à apoptose, imortalidade, habilidade de resistir ao estresse ambiental e evadir dos diferentes mecanismos de defesa. Além das mutações, também estão fortemente associados à neoplasia as fusões gênicas, translocações cromossômicas, alterações no números de cópias, modificações epigenéticas e o mecanismo de *splicing* alternativo [15,45].

O *splicing* é um processo que ocorre no pré-mRNA e controla a excisão de sequências de introns e ligação dos éxons, compondo o mRNA maduro. Contudo, algumas sequências de introns e éxons podem ser alternativamente incluídas (ou excluídas) e diferencialmente combinadas, formando um mRNA maduro que pode apresentar diferentes atuações no contexto celular. Esse processo é chamado de *splicing* alternativo e será melhor discutido no **Capítulo 1**.

Em humanos, cerca de 95% dos genes codificantes apresentam variantes transcricionais e sofrem *splicing* alternativo, proporcionando uma grande diversidade transcritômica e proteômica pela geração de múltiplas isoformas de mRNA a partir de um único gene [46,47]. O controle do *splicing* alternativo pode variar de acordo com o tipo celular, tecido, estágio de desenvolvimento e o estímulo ambiental. Nesse sentido, o *splicing* alternativo foi observado na regulação de processos biológicos distintos, como diferenciação, proliferação, adesão celular, apoptose e manutenção do estado pluripotente [46]. Contudo, muitas doenças apresentam uma desregulação deste mecanismo, e por consequência, muitos processos biológicos também são afetados. Durante a tumorigênese, a anormalidade dos eventos de *splicing* alternativo podem afetar muitos genes importantes na regulação de processos biológicos cruciais, contribuindo para a progressão e agressividade do tumor, como proliferação, migração, adesão, morte celular, entre outros (**Figura 6**). Muitos estudos têm investigado a presença de isoformas unicamente expressas no tumor em busca de biomarcadores e potenciais alvos para tratamentos terapêuticos [48].

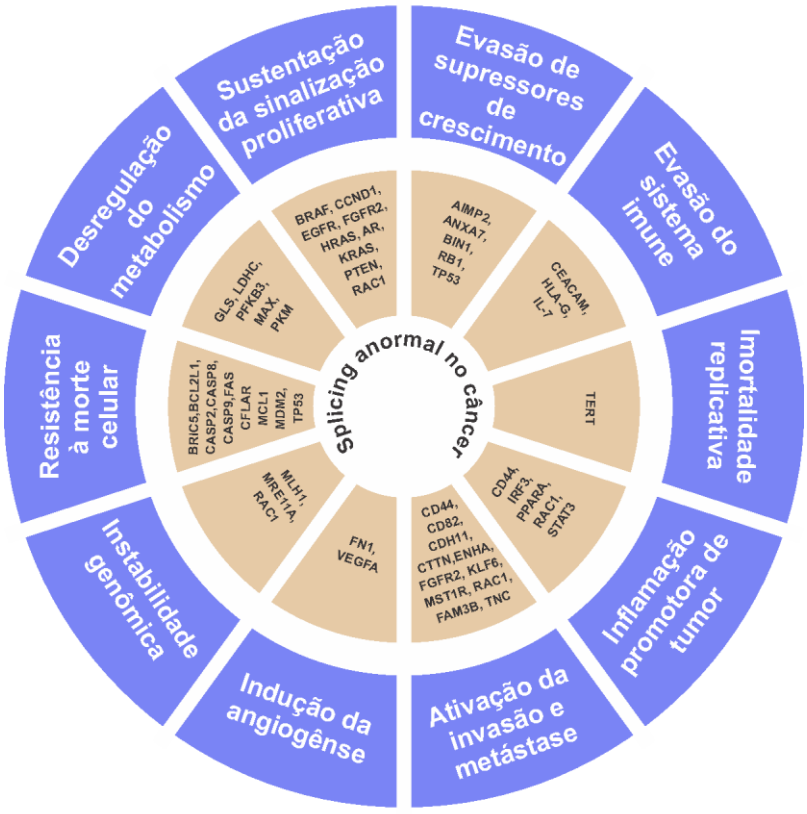



**Figura 6**. Exemplos de alguns genes que durante o câncer são submetidos a um *splicing* alternativo aberrante e os processos aos quais eles estão envolvidos.[5]

Em câncer, fatores de *splicing* podem apresentar propriedades oncogênicas, uma vez que a integridade genômica e epigenética é crucial para o correto funcionamento do mecanismo de *splicing* [49,50].

A primeira evidência que correlacionou diretamente a regulação do *splicing* com o câncer foi a presença de mutações em fatores de *splicing* observados em diferentes tumores (**Figura 7A**) [51]. Além disso, a atividade anormal dos fatores de *splicing* pode também ser em virtude de alterações na sua regulação transcricional, ou ainda, em vias de sinalização que ativam esses fatores (**Figura 7A**) [51]. A anormalidade da atividade dos fatores de *splicing* é uma das principais marcas relacionadas com o desenvolvimento tumoral e a superexpressão dos fatores de *splicing* é comumente observada em tumores, incluindo o CCR (**Figura 7B**) [52]. Por exemplo, o fator de *splicing* SRSF1 é um alvo transcricional de c-Myc, e o aumento da sua expressão está correlacionado com maior atividade de c-Myc em tumores [52]. A

[5] Modificado de Sveen A, et al. Oncogene. 2015; 35(19):2413–2427.

superexpressão de SRSF1 aumenta a agressividade do tumor e confere resistência à quimioterápicos em células de adenocarcinoma de pulmão [53]. Também foi observada a superexpressão de SRSF6 no CCR, que atua regulando positivamente a proliferação e sobrevivência das células neoplásicas [54]. Outros fatores de *splicing*, como hnRNP K, hnRNP I e SRSF9, também já apresentaram superexpressão no CCR ou em linhagens de câncer de cólon [51,55]. Além disso, os fatores de *splicing* SR podem ser regulados por fosforilação mediada por cinases, como SRPK e AKT, comumente alteradas em diferentes tumores [52].

Em tumores, os fatores de *splicing* hnRNPs e SR podem atuar regulando os eventos de *splicing* alternativo de forma a aumentar a expressão de uma isoforma oncogênica e reduzir a expressão de isoformas supressoras de tumor [51]. Um exemplo de caso é o *skipping* do exon 4 do pré-mRNA de CCND1 promovido por SRSF1, resultando em uma isoforma que contribui para a progressão tumoral [51].

No CCR, já foram descritos algumas mutações *drivers* em fatores de splicing, como PCBP1, RBM10 e SON, além de ocupar a quarta posição como o tipo de câncer que mais apresenta superexpressão de fatores de splicing quando comparados ao tecido normal (**Figura 7B**) [56].

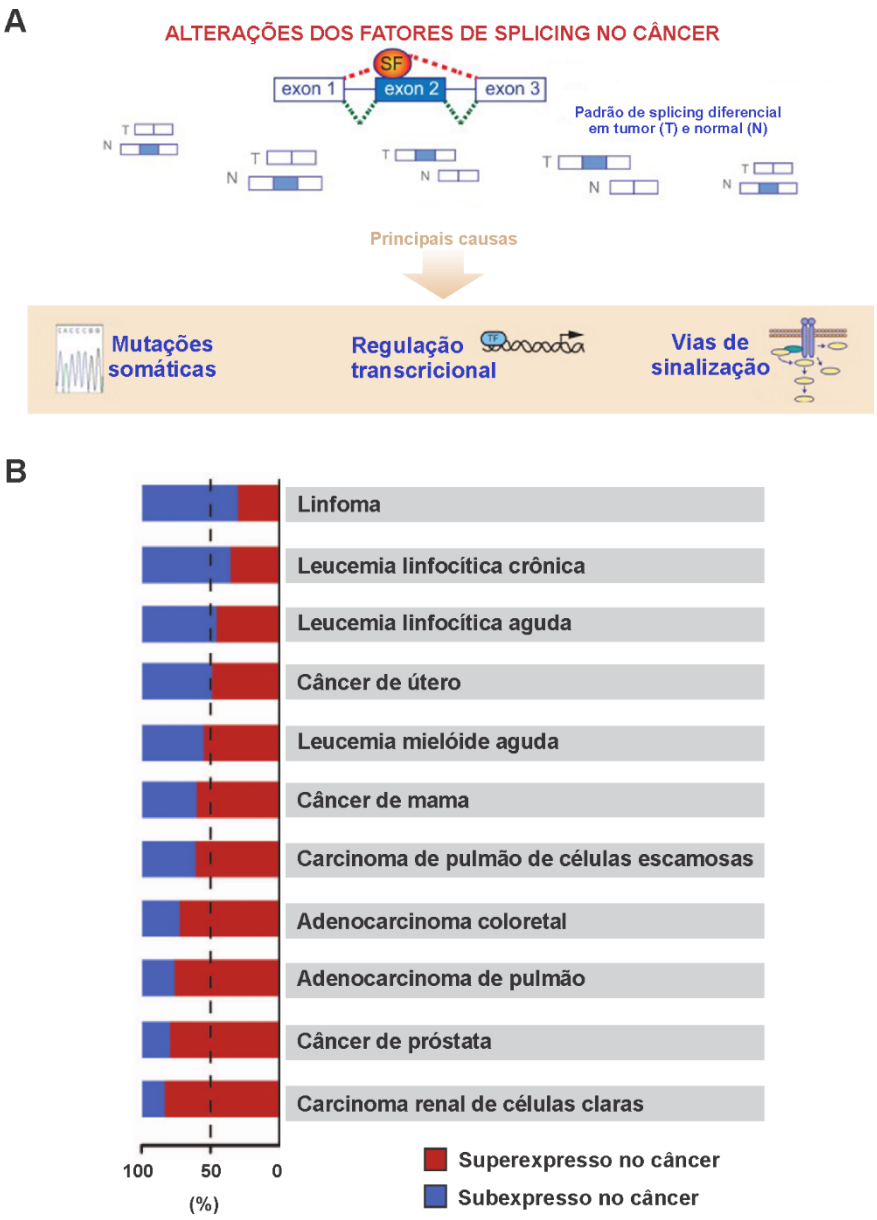



**Figure 7**. Alterações de fatores de *splicing* no câncer. **A)** Representação das principais causas de anormalidade na atividade dos fatores de *splicing* encontradas no tumor e os processos afetados. **B)** Porcentagem de fatores de *splicing* diferencialmente expressos em diferentes tipos de câncer em relação à sua contraparte normal.[6,7]

[6] Modificado de Sveen A, et al. Oncogene. 2015; 35(19):2413–2427.
[7] Modificado de Anczuków O & Krainer AR. RNA. 2016; 22(9):1285–1301.

# JUSTIFICATIVA

A instabilidade genômica e a falha nos mecanismos de reparo no DNA são as principais causas associadas ao desenvolvimento da malignidade em CCR. Contudo, muitas outras alterações ocorrem ao longo da progressão do tumor primário, como acúmulo de mutações e desregulação da expressão gênica que, por consequência, podem desencadear na aquisição de um potencial formador de metástase. Apesar de o processo metastático ser ineficiente, algumas células adquirem a capacidade de sobreviver e proliferar em condições extremas, tornando difícil o tratamento e impactando diretamente na sobrevida desses pacientes.

Muitos estudos têm mostrado a importância do *splicing* alternativo na promoção e manutenção de células malignas, e muitos alvos já são discutidos por serem cruciais para o sucesso da tumorigênese. Atualmente, é estimado que cada gene alvo de *splicing* alternativo (95% dos genes) possuem em média 6,3 isoformas, e deste valor, 3,9 delas podem codificar para proteínas [57]. Devido a isso, não é surpreendente que diferentes isoformas sejam expressas de acordo com seu contexto biológico, e que uma regulação anormal do mecanismo de *splicing* seja relacionada com a formação de processos patológicos. Por isso, as perturbações que ocorrem na regulação do *splicing* alternativo já são consideradas *hallmarks* do câncer [58,59].

Comparativamente, existem muitas alterações observadas entre o transcritoma tumoral e normal, e estas são consequências tanto de mudanças no contexto genômico quanto da regulação transcricional. Além disso, diferentes isoformas geradas a partir do mesmo gene podem apresentar funções completamente distintas, podendo impactar diretamente na classificação molecular do tumor e no tipo de tratamento ao qual determinado paciente deveria ser submetido.

Embora seja evidente a importância de se estudar o *splicing* alternativo durante a progressão tumoral, ainda são poucos os trabalhos que consideram avaliar a expressão dos genes ao nível de isoformas transcricionais. Portanto, essa pesquisa tem o potencial de ampliar o conhecimento sobre as isoformas transcricionais que podem ser alteradas durante a progressão do CCR e as principais consequências associadas às mudanças transcritômicas.

# OBJETIVOS

## Objetivo Geral

Avaliar o perfil das isoformas transcricionais associadas com o tumor primário e com a metástase hepática do câncer colorretal.

## Objetivos Específicos

- Identificar as isoformas transcricionais mais relacionadas com cada condição;
- Verificar a troca no uso das isoformas de RNA derivadas do mesmo gene entre as condições;
- Avaliar as possíveis consequências funcionais, como presença de domínios, potencial codificante, presença de peptídeo sinal, e decaimento de RNA, que podem estar relacionadas com a troca das isoformas;
- Analisar e comparar os principais processos e vias biológicas afetadas entre as condições;
- Avaliar se existe diferença no tipo de *splicing* entre as condições estudadas;
- Analisar a presença de mutações nos pacientes analisados;
- Verificar se existe relação entre a presença de mutações e a troca de isoformas.

# CAPÍTULO I – ARTIGO DE REVISÃO

## INFLUENCE OF TRANSCRIPTIONAL VARIANTS ON METASTASIS

Artigo publicado na revista RNA Biology

RNA BIOLOGY
https://doi.org/10.1080/15476286.2018.1493328

Taylor & Francis
Taylor & Francis Group

REVIEW



# Influence of transcriptional variants on metastasis

Joice De Faria Poloni and Diego Bonatto

Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

**ABSTRACT**

Cancer metastasis is defined as the dissemination of malignant cells from the primary tumor site, leading to colonization of distant organs and the establishment of a secondary tumor. Metastasis is frequently associated with chemoresistance and is the major cause of cancer-related mortality. Metastatic cells need to acquire the ability to resist to stresses provided by different environments, such as reactive oxygen species, shear stress, hemodynamic forces, stromal composition, and immune responses, to colonize other tissues. Hence, only a small population of cells has a metastasis-initiating potential. Several studies have revealed the misregulation of transcriptional variants during cancer progression, and many splice events can be used to distinguish between normal and tumoral tissue. These variants, which are abnormally expressed in malignant cells, contribute to an adaptive response of tumor cells and the success of the metastatic cascade, promoting an anomalous cell cycle, cellular adhesion, resistance to death, cell survival, migration and invasion. Understanding the different aspects of splicing regulation and the influence of transcriptional variants that control metastatic cells is critical for the development of therapeutic strategies. In this review, we describe how transcriptional variants contribute to metastatic competence and discuss how targeting specific isoforms may be a promising therapeutic strategy.

**ARTICLE HISTORY**
Received 16 February 2018
Accepted 12 June 2018

**KEYWORDS**
Alternative splicing; transcriptional isoforms; cancer-specific isoforms; cancer; metastasis; metastasis-initiating potential

## 1. Introduction

In recent years, intensive studies have contributed to advances in early cancer detection and in the treatment of surgically resected primary tumors and adjuvant therapy [1,2]. These efforts increased survival rates for the majority of patients diagnosed during the initial stages of cancer [1,2]. However, metastatic cancer treatment faces resistance to chemotherapeutic compounds, making it necessary to understand the underlying mechanisms of the metastatic cascade [1,2]. In this sense, metastasis is responsible for 90% of cancer-related mortality and morbidity in patients [2].

Metastasis is a multistage process in which malignant cells disseminate from the primary site and grow on distant organs, leading to new tumors [3]. Nonetheless, malignant cell transformation is not sufficient for the acquisition of metastatic competence [3]. Metastasis mainly occurs due to genome instability, the loss of effective DNA repair, checkpoint control failure and gene expression regulation [4]. Hence, genome comparisons between primary tumor and distant metastasis show distinct mutations in metastatic cells compared with the primary tumors [4].

Recently, several studies have established the importance of alternative transcription and alternative splicing in promoting the initiation and maintenance of cancer cells [5–10]. The use of alternative promoters and/or alternative termination sites can result in multiples pre-mRNA that undergoes splicing, generating a huge diversity of transcripts from a single gene [6]. During splicing, premature mRNA (pre-mRNA) is processed and exons are assembled together to compose a mature mRNA, while the remaining introns are subsequently degraded [11]. Additionally, some sequences can be alternatively included (or excluded) and differentially combined to give rise to different mature mRNA isoforms via a process named alternative splicing [11]. This event may have different consequences in the final product: (i) the alternative splicing does not alter the reading frame, and the resulting in-frame protein may have (or not) a different function from canonical protein, or (ii) alternative splicing causes a frameshift, which leads to the generation of a premature termination codon (PTC) that is directed to nonsense-mediated decay (NMD), a different protein arrangement or a truncated protein production [12,13]. It is estimated that reading frame alteration occurs approximately in 50% of all alternative splicing events [14].

In this sense, alternative splicing provides proteomic diversity through the production of multiple mRNA transcripts from protein-coding genes [11]. More than 95% of human genes are differentially spliced depending on the tissue/cell type or signal transduction in response to developmental stages and environmental stimuli [11,15,16]. Currently, the role of alternative splicing has been associated with different diseases, and disturbances in alternative splicing regulation are considered a hallmark of cancer [10,15–17].

This review is focused on describing the influence of alternative splicing during the metastatic cascade, focusing on current knowledge about alternative transcripts that may be intrinsically related to metastatic competence and how targeting specific isoforms that contribute to metastatic progression may be a promising therapeutic strategy.

**CONTACT** Joice de Faria Poloni joicefpoloni@gmail.com Centro de Biotecnologia da UFRGS - Sala 107 Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul – UFRGS, Avenida Bento Gonçalves 9500 - Prédio 43421, Porto Alegre, Rio Grande do Sul 15005, BRAZIL

## 2. General view of splicing

Alternative RNA isoforms are regulated by genetic and epigenetic mechanisms and can be formed by means of alternative promoters or by undergoing alternative splicing [6]. In this section, we will briefly summarize these mechanisms, focusing on alternative splicing. For more information, please refer to [6,11,18–20].

The alternative transcription mechanism is defined by the use of alternative promoters and transcriptional termination sites, resulting in multiple pre-mRNA [6]. Its regulation involves the presence of *cis*-regulatory elements, DNA methylation, histone modifications and chromatin remodeling [6]. Transcription and RNA processing are highly interdependent events and occur simultaneously [6,11]. During transcription, RNA polymerase II (RNA Pol II) elongation produces pre-mRNA, which undergoes cotranscriptional splicing, resulting in mature mRNA, during which pre-mRNA processing introns are removed and exons are joined together [6,19]. However, some exons and introns can be alternatively spliced, generating different isoforms with distinct activities, and such transcriptome variability is observed at the protein level [10]. This process is catalyzed by the spliceosome, a large complex of RNA composed of five small nuclear ribonucleoproteins (snRNPs; U1, U2, U4, U5 and U6) and several associated proteins cofactors, resulting in a highly dynamic conformation [5,21].

The spliceosome recognizes short consensus sequences, called splice sites, in nascent transcripts located at the intron-exon junctions, which are 5′ splice sites, 3'splice sites, and a branch point region near the 3'splice site [5,21]. These splice sites can be classified as 'strong' when they are similar to consensus sequences and, thus, more efficiently recognized [21]. In general, cognate splicing factors tend to more efficiently recognize 'strong' splice sites, which leads to constitutive splicing [21]. However, neighboring 'strong' and 'weak' splice sites compete with themselves for affinity with the splicing factors, and different positions of these splice sites can cause different alternative splicing outcomes [21]. Thus, additional splicing regulation is performed by *trans*-acting factors and *cis*-regulatory sequences [21]. *Cis*-regulatory sequences include exonic and intronic splicing silencers (ESSs and ISSs) and enhancers (ESEs and ISEs) (Figure 1(a)) [21]. It is common for alternative exons to have weaker splice sites and ESE, indicating an evolutionary advantage of the use of alternative exons [22].

*Trans*-acting factors bind to *cis*-regulatory sequences to either promote or prevent spliceosome assemblement at an adjacent splice site [5,11,21]. Most of these factors are serine- and arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoprotein (hnRNP), and numerous tissue-specific factors and additional proteins have been described to stimulate or inhibit splicing (Figure 1(a)) [5,11,21]. SR proteins and hnRNPs are described to act as proto-oncogenes and tumor suppressors. For example, the RNA-binding protein hnRNP I is a pre-mRNA splicing regulator in alternative splicing and is related to biological processes, such as cell structure and motility, cell cycle, immunity and others [23]. Its overexpression is observed in colorectal, brain, ovarian and breast cancers, and it has been related to tumor aggressiveness, especially in glioma and ovarian tumors [23]. By contrast, hnRNP I inhibition decreases cell proliferation, migration and invasion in colorectal cancer cells [23].

Another example is the SR splicing factor SRSF1, which is related to malignant transformation and is overexpressed in lung, breast and colon cancers [24]. This protein leads to activation of the mTOR pathway, inducing the missplicing of MNK2 and RPS6KB1 and the alternative splicing of BCL2L11 and BIN1 to produce isoforms without pro-apoptotic activity [24]. In addition, overexpression of SRSF3, SRSF6 and SRSF10 is also related to tumorigenesis [24]. Although SR

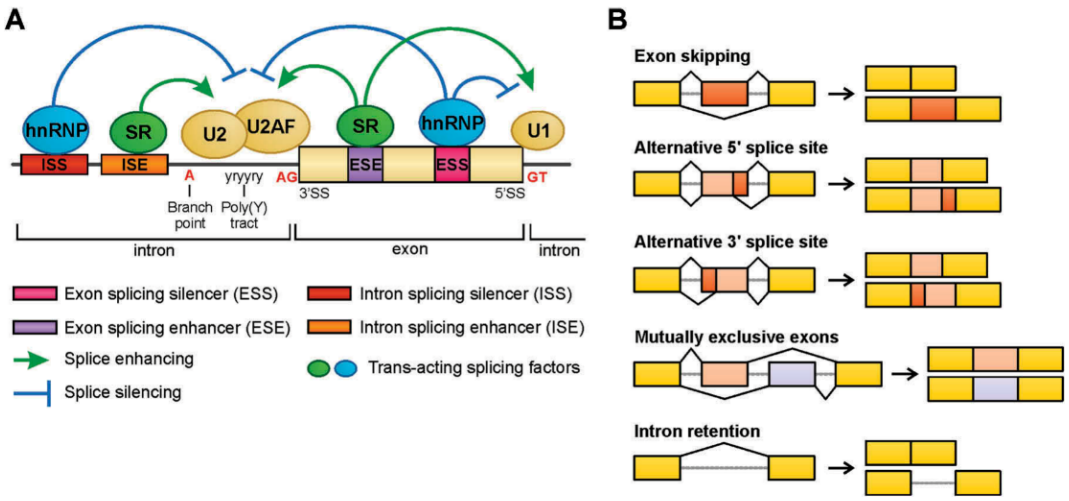



**Figure 1.** (a) Sequences located at the intron-exon junctions, termed 5′ and 3′ splice sites, together with branch point positioned near the 3′ splice site, are responsible for defining the intron and exon limits. These sequences are recognized by the spliceosome machinery and associated factors. Additional regulation of the site splice choice can be mediated by *cis*-regulatory sequences, including exonic and intronic splicing silencers (ESSs and ISSs) and enhancers (ESEs and ISEs), and *trans*-acting factors, such as serine- and arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoprotein (hnRNP). These splicing factors are able to inhibit or promote the recognition of nearby splice sites affecting the activity of the spliceosome components, such as the snRNPs U1 and U2. (b) Representation of the five main alternative splicing events.

proteins generally act as oncogenes, hnRNPs can exert both roles, suppressing and promoting tumorigenesis [24].

Alternative splicing regulation is typically influenced by three different aspects of co-transcriptional RNA processing: (i) CTD (RNA Pol II carboxy-terminal domain)-mediated recruitment, (ii) kinetic competition and (iii) chromatin structure [21,25]. In CTD-mediated recruitment, splicing factors are recruited to transcription sites by associating splicing factors with the CTD [21,25]. In kinetic competition, the RNA Pol II elongation rate influences the output timing of splice sites and regulatory sequences from the nascent pre-mRNA [21,25]. In this sense, a fast RNA Pol II elongation rate limits kinetic competition and favors the inclusion of exons flanked by strong splice sites, while weaker splice sites are favored by a slow elongation rate [25]. The coupling of transcription and splicing processes is suggested to be associated with the effects of histone modification and nucleosome positioning [25]. In fact, histone post-translational modifications are related to active transcription, such as histone H3K36me3, which is enriched in exonic compared with intronic regions and is implicated in the alternative splice site choice [21,25]. In human kidney cancer, mutations in *SETD2*, a histone methyltransferase that specifically catalyzes H3K36me3, promote increased chromatin accessibility and lead to widespread changes in RNA processing and aberrant splicing [26]. In this tumor, chromatin accessibility is pronounced in the proximal region of misspliced exons, where the deficiency of H3K36me3 is related to decreased nucleosome occupancy of these exons [26]. In fact, evidence provides a key role for nucleosomes in splice site recognition and exon definition [25]. Moreover, the roles of nucleosome are enriched in exons, preferentially exons flanked by weaker splice sites [25]. Thus, nucleosome positioning contributes not only to exon definition but also to splicing regulation [25].

Currently, five main types of alternative splicing are recognizable:(i) exon skipping; (ii) alternative 5′ splice site; (iii) alternative 3′ splice site; (iv) mutually exclusive exons; and (v) intronic retention (Figure 1(b)) [11]. In vertebrates and invertebrates, the most prevalent alternative splicing type is exon skipping, which represents approximately 30% of all alternative splicing events [11]. However, mutations in splice sites and regulatory sequences may result in aberrant intron retention with pathological consequences [13]. Actually, intron retention is enriched in tumor-suppressor genes when somatic single-nucleotide variants (SNVs) are present [27]. Consequently, these transcripts undergo proteins truncation or NMD by the generation of the PTC [27].

A subset of retained introns has been investigated recently, which was called exitrons because they have features of both introns and protein-coding exons [28]. Exitrons are a class of alternatively spliced introns defined as a protein-coding intron sequence within protein-coding exons, whose regulation is made by internal splice site motifs in the exon [12,28,29]. They have different characteristics from others introns, such as the absence of stop codons, high GC content, a preferential size that corresponds to multiples of three, and prevalence of synonymous substitutions, also frequently observed in exons [29]. Unlikely of retained conventional introns transcripts that are inclined to stay in the nucleus, retained exitrons transcripts are transported to the cytoplasm and translated, being enriched in disordered regions of the codified proteins [12,28,29]. Marquez et al. [28] investigated the role of exitrons in ERBB2-positive breast cancer and identified 29 cases with differential exitron splicing between breast cancer and normal tissue, including the suggested tumor suppressors CIZ1, ZNF238 and EXT1.

Alternative splicing precision, efficiency and plasticity generate the vast transcriptomic and proteomic diversity required for eukaryotic cell biology. However, disturbances in this regulation can lead to incorrect decisions and, consequently, disease development.

## 3. Metastasis and implications of alternative splicing for the metastatic cascade

Cancer metastasis is described as a dissemination of malignant cells from a primary tumor site to different organs through the blood and lymphatic circulation. In 1889, Stephen Paget described metastasis not as a random process, but rather, he proposed that cells of certain tumors have a preferential affinity to specific organs [2,30]. Paget hypothesized the 'seed-and-soil' theory of metastatic outgrowth, where the establishment and growth of malignant cell ('seeds') are dependent on particular organs ('soil') to survive and proliferate [2,30].

Successful metastasis consists of achieving the five basic steps of the invasion-metastasis cascade: local invasion, intravasation, circulation survival, extravasation and colonization (Figure 2). In epithelium-derived cancers, the cells are maintained together by cell-to-cell adherence and by adherence to the extracellular matrix (ECM) [2].

Once normal cells have lost their adhesion to other cells or the basement membrane, the detached cells undergo anoikis, a mechanism of apoptosis induced by cell detachment [2]. However, metastatic cells are able to overcome this phenomenon by altering the signaling pathways responsible for anoikis regulation, including NF-κB, small GTPases and its effectors, some receptor tyrosine kinases and cytoplasmic kinases, and epithelial-to-mesenchymal transition (EMT) factors [31].

Overcoming anoikis is followed by the invasion of surrounding tissues of the primary tumor, which is initiated by metastatic cells migrating through the basement membrane and ECM [32]. In this sense, malignant cells need to acquire invasive features that are frequently associated with the suppression of epithelial properties and the acquisition of mesenchymal markers to disrupt cell-cell and cell-ECM adhesion [33]. This plastic program, called EMT, is reversible and may provide the ability to the tumor cell to travel to metastatic sites and establish the growth of a secondary tumor [33]. During EMT, epithelial cells acquire a mesenchymal cell phenotype that consists of a (i) loss of adhesion structures, (ii) increased cell death resistance, (iii) altered polarity, (iv) cytoskeleton reorganization, and (v) enhanced migratory capacity [33,34]. However, the reversion of disseminated tumor cells into an epithelial phenotype is required for seeding in distant organs and the formation of a secondary tumor through a process called mesenchymal-epithelial transition (MET) [35]. Some studies using breast and pancreatic cancer cell lines have shown that EMT is not indispensable for

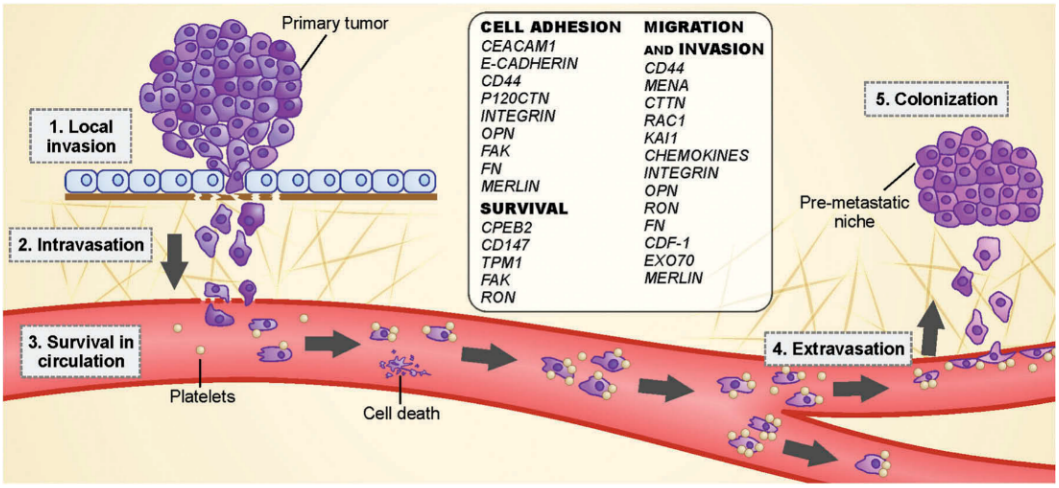



**Figure 2.** Cell spreading from the primary tumor involves the acquisition of capabilities to disrupt the basement membrane, invade the surrounding tissue (stroma) and enter the blood circulation. After that, the spreading cells in the circulation must survive in the hostile environment until they can reach distant organs and establish a niche to grow and form a new lesion (secondary tumor).

metastasis establishment but contributes largely to its aggressiveness by increasing resistance to anticancer therapy [36].

Invasive cells undergo actin cytoskeleton reorganization and the formation of membrane protrusions enriched in filamentous actin (F-actin) that, along with adhesive interactions, are crucial for invasion and migration through the stroma [2,36]. Malignant cells exit the tumor as single cells or as clusters and enter the vasculature, spreading throughout the whole body (Figure 2) [36]. Additionally, platelets and granulocytes are recruited to form niches that contribute to metastatic progression by forming a shield around tumor cells, promoting adhesion to endothelial cells and increasing endothelial permeability [37].

These tumor cells halt their circulation in capillaries at distant sites and start to extravasate into the target organ parenchyma, where these cells adopt strategies to survive and initiate colonization (Figure 2) [3,36].

Although metastasis itself is believed to be an inefficient process, the dissemination of cells from a one-centimeter-sized primary tumor corresponds to an infiltration rate of one million cells/day in the circulatory system [32]. However, due to the limited compatibility of distal sites ('seed-and-soil' theory), only < 0.1% of tumor cells reach an appropriate site and are able to survive in the new 'soil' and avoid immune surveillance [32]. This low metastatic competence can be explained by the death of tumor cells due to their inability to resist the many types of stresses, such as reactive oxygen species, shear stress, hemodynamic forces, stromal composition, detachment, and immune response [38,39]. For these reasons, only a small population of tumor cells are able to exert an adaptive response to the environmental stress, thus carrying a metastasis-initiating potential [38].

Many splice events can be used to distinguish normal and tumor tissue and contribute to an adaptive response of tumor cells and the success of the metastatic cascade. These contributions frequently alter normal biological processes, such as cell cycle control, cytoskeleton organization, migration, cell-cell adhesion and insulin signaling [40]. Splicing regulatory proteins have been the main investigated targets because the abnormal activity of these effectors could be the key related to pathological splicing alterations. Many articles have provided excellent reviews of master changes affecting the splicing program [24,41,42]. Moreover, cellular features acquired by tumor cells drive a splice program shift that could be exerted even beyond the sole aberrant function of splicing factors. In the next section, we will address the most relevant genes related to metastasis with reported transcriptional variants that contribute to tumor cell adaptation. A brief summary is represented in Table 1.

### 3.1. E-cadherin

E-cadherin is an adhesion regulator that is also known as an important tumor suppressor, which, along with catenins, maintains cell-cell adhesion and cell-matrix interactions [30]. During EMT, cadherin type switching occurs, in which epithelial E-cadherin is downregulated and the mesenchymal marker N-cadherin is upregulated [43]. Nonetheless, a study has demonstrated that E-cadherin downregulation can result from missplicing [44]. Sharma et al. [44] showed that E-cadherin exon 11 is skipped, resulting in a frameshift and PTC that will direct the transcript to NMD. The splicing factor SRSF2 has been suggested to be responsible for the aberrant splicing of E-cadherin, since in cancer cells with skipped exon 11, SRSF2 is upregulated, and its silencing increases wild-type E-cadherin RNA levels [44].

In gastric cancer, the lack of the final 83 base pairs of exon 8 (1054*del*83) results in a nonfunctional transcript [45]. Although it does not result from abnormal activity of

**Table 1.** Summary of the main mentioned isoforms and representative examples of therapeutic approaches to targeting them. M.C., metastasis contribution (↑ high contribution; ↓ low contribution).

| Gene | Isoforms | Splicing | M.C. | Consequence | Therapeutic potential | Ref. |
|---|---|---|---|---|---|---|
| E-cadherin | E-cadherin exon 11 skipping | exon 11 skipping | ↑ | NMD | HDAC inhibitor MS-275 | [44,173] |
| | E-cadherin exon 8 1054*del*83 | exon 8 partial deletion of the final 83 base pairs | ↑ | NMD | HDAC inhibitor TSA | [45] |
| CEACAM1 | CEACAM1-L | exon 7 inclusion | ↑ | Lower survival rate; Accelerated metastasis progression | - | [48,49] |
| | CEACAM1-S | exon 7 skipping | ↓ | Normal phenotype | - | [48,50] |
| CD44 | CD44s | exon 6–15 (v1-v10) skipping | ↑ | Associated to mesenchymal phenotype | - | [54,57] |
| | CD44v3–10 | Standard exons + variable exons 3-10 | ↑ | Chemotaxis induced by OPN | Anti-CD44v antibodies (non-specific to v3-10)<br>CD44v-xCT system (non-specific to v3-10)<br>Soluble CD44v3–v10–Fc fusion protein | [61,174,175] |
| | CD44v6 | Standard exons + variable exons 6 | ↑ | Induce migration and expression of mesenchymal markers; Related to drug resistance | CD44v-xCT system (non-specific to v6)<br>Anti-CD44v6 antibodies<br>Package CD44v6-shRNA/ nanoparticles<br>Competitive peptide targeting aa in exon 6 | [174–177] |
| Mena | Mena$^{11a}$ | inclusion of exon 11 | ↑ | Epithelial phenotype | - | [68] |
| | MenaΔv6 | exclusion of exon 6 | ↑ | Increase invasiveness and drug resistance in the absence of Mena$^{11a}$ | - | [71] |
| | Mena$^{INV}$ | inclusion of exon 'INV' | ↑ | Discohesive cell morphology; Cell motility | - | [69] |
| Rac1 | Rac1b | Exon 3b inclusion | ↑ | Increase cell motility; Cell transformation | Sanguinarine; Ibuprofen | [79,83,178,179] |
| CTTN | CTTN-A | Exon 11 inclusion | ↑ | Cell migration | - | [23] |
| CPEB2 | CPEB2B | Exon 4 inclusion | ↑ | Anoikis resistance | - | [89] |
| KAI1 | KAI1-SP | Exon 7 exclusion | ↑ | Increase cell motility and cell-ECM adhesion | - | [90] |
| CD147 | CD147/Bsg-2 | - | ↑ | Promote cell proliferation and invasiveness | - | [94] |
| | CD147/Bsg-3 | - | ↓ | Inhibit proliferation and invasion | - | [94] |
| RON | RONΔ155 | Exon 5, 6 and 11 deletions | ↑ | Cell transformation and tumor growth | - | [97,98] |
| | RONΔ160 | In-frame deletion of exon 5 and 6 | ↑ | Cell transformation and tumor growth | - | [97,98] |
| | RONΔ165 | In-frame deletion exon 11 | ↑ | Cell motility and invasion | - | [97,98] |
| | SF-RON | Alternative promoter | ↑ | Increase cell proliferation and invasiveness | ASLAN002 | [95,96,180] |
| IL-7 | IL-7δ5 | Exon 5 deletion | ↑ | Promote cellular proliferation and EMT | - | [107] |
| CXCR3 | CXCR3-A | Formed by exon 1 and 3 | ↑ | Increase proliferation, invasion and migration | AMG487 | [109,110] |
| | CXCR3-B | Formed by exon 2 and 3 | ↓ | Inhibits proliferation and migration | - | [110] |
| FAK | FAKΔ26 | Exon 26 deletion | ↑ | Inhibit apoptosis | - | [119] |
| OPN | OPN-a | Full-length | - | Promote cell migration and growth | - | [124] |
| | OPN-b | Exon 5 deletion | ↑ | Increase cell migration; Pro-survival | - | [125] |
| | OPN-c | Exon 4 deletion | ↑ | Decreased cell migration; Increase detachment | - | [122,125] |
| FN | EDA-FN | Exon EDA inclusion | ↑ | Related to angiogenesis, migration and proliferation | Vaccine targeting EDA-FN antigen | [129,131,181] |
| | EDB-FN | Exon EDB inclusion | ↑ | Tumor growth and angiogenesis | peptide CTVRTSADC (ZD2); monoclonal antibody BC-1 | [129,182,183] |
| | MSF | Intron 12 retention followed by cleavage events | ↑ | Highly chemotactic | - | [135] |
| TP53 | Δ133p53β | Initiation of transcription in intron 4 and alternative intron 9 splicing | ↑ | Promote cell invasion, chemoresistance and expression of pluripotency factors | - | [139,140] |

(*Continued*)

**Table 1.** (Continued).

| Gene | Isoforms | Splicing | M.C. | Consequence | Therapeutic potential | Ref. |
|---|---|---|---|---|---|---|
| Prrx1 | Prrx1a | - | ↑ | Contribute to MET | - | [35,146] |
| | Prrx1b | - | ↑ | Promote invasion and EMT | - | [35,146] |
| EGFR | EGFRvIII | Exon 2–7 deletion | ↑ | Contribute to malignant transformation | peptide vaccine rindopepimut, CAR-T targeting EGFRvIII | [150,151,169] |
| | de4 EGFR | Exon 4 deletion | ↑ | Contribute to malignant transformation, increase invasiveness | - | [147,150,151] |
| Id1 | Id1a | - | - | Increase proliferation | - | [152] |
| | Id1b | - | ↑ | Contribute to stem cell-like phenotype and cell quiescence | - | [152] |
| XAF1 | XAF1F | Lacks exon 5 and contain an exon 3-ext | ↑ | Related to cell invasion | - | [155] |
| KLF6 | KLF6-SV1 | - | ↑ | Associated to mesenchymal phenotype | polyethylenimine/si-SV1 intratumoral injection | [159,160,184] |
| Exo70 | Exo70-M | Lacks part of exon 8 relative to Exo70-E | ↑ | Promote invasion | - | [161] |
| Merlin | Merlin$^{\Delta 2-4}$ | Exons 2–4 deletion | ↑ | Promote migration and invasion | - | [163] |
| CD99 | CD99wt | Full-length | ↓ | Associated to less migrating phenotype and higher adhesion | - | [164] |
| | CD99sh | Deletion of a fragment and insertion of 18 bp at the boundary of exons 8 and 9, introducing an in-frame stop codon | ↑ | Stimulate cell motility | - | [164] |

SRSF2, this phenomenon is related to decreased acetylation of histones H3 and H4K16Ac on the *E-cadherin* gene surrounding the alternative exon 8 [45]. H3K36me3 surrounding the exon 8 region is correlated with abnormal splicing of this same region, which indicates that epigenetic modifications patterns can be critical in gastric cancer [45].

### 3.2. CEACAM1

Carcinoembryonic antigen-related cell adhesion molecule-1 (CEACAM1) is a transmembrane glycoprotein involved in cell-cell adhesion, and its abnormal expression is associated with a variety of human malignancies [46,47]. CEACAM1 alternative splicing generates 12 isoforms, resulting in three C2-like domains and forming isoforms that differ in the length of the extracellular region [48]. Furthermore, CEACAM1 splicing is able to encode two major cytoplasmic domains that differ by the inclusion (termed long (-L) tail) or exclusion (termed short (-S) tail) of exon 7, respectively, forming CEACAM1-L and CEACAM1-S [47]. Dery et al. [47] observed that stress-induced cytoplasmic accumulation of hnRNP A1 leads to the formation of CEACAM1-L in breast cancer and, consequently, leads to an inappropriate deregulation of CEACAM1-S expression. The prevalence of CEACAM1-L isoform expression is associated with accelerated metastasis progression and a lower survival rate in patients with colon tumors [49]. By contrast, transfection of CEACAM1-S has been described to mediate the reversion of mammary tumor cells to a normal phenotype, as noted by acini structures with the central lumen formation via the apoptosis mechanism [50].

### 3.3. CD44

The cell surface glycoprotein CD44 also plays an important role in cell adhesion and EMT, in addition to being involved in proliferation, migration and invasion during metastasis, in an extracellular matrix ligand-dependent manner [51,52]. CD44 pre-mRNA has 20 exons and shows several isoforms encoded by standard exons 1–5 (c1-c5) and 16–20 (c6-c10), whereas exons 6–15 (v1-v10) are considered variant exons that are regulated by alternative splicing [52]. In this sense, the inclusion and exclusion of a variant exon represent tissue and environment-specific-dependent factors, for which a single or combined exons can be selected to compose the final transcript [52]. Through these variable combinations, CD44 can interact with different cognate receptors and cooperate in the activation of several signaling pathways that contribute to tumor progression and dissemination [53]

In mesenchymal cells, variant exons are skipped, and only standard exons are included (termed standard isoform or CD44s) [52]. Nonetheless, proliferating and epithelial cells predominantly show variant-exon-encoded sequences (termed the variant isoform or CD44v) [52]. Both CD44s and CD44v may be associated with cellular malignancy depending on their partners, type and tumor stage. For example, CD44v is enriched in normal breast tissue and in lower-grade breast tumors; however, a switch in the expression of CD44v rather than CD44s accelerates the EMT process and contributes to cancer progression [34,54]. Furthermore,

CD44s has been reported to be an integral component of invadopodia, which contribute to extracellular matrix degradation and the dissemination of metastatic cells from the primary site to distant organs [55]. In pancreatic cancer, CD44s is related to the EMT phenotype, collaborating to participate in invasiveness and chemoresistance *in vivo* [56].

In addition, the switch from CD44v to CD44s has been correlated to the mesenchymal phenotype and is prevented by the splicing factors ESRP1 and ESRP2 [41,54,57]. ESRP1 and ESRP2 are involved in alternative splicing events of epithelial cell phenotypes, and their targets are frequently involved in cell-cell adhesion, cell motility, actin cytoskeletal organization and EMT [41,58]. These splicing factors mediate the variable exon inclusion and, consequently, the generation of CD44v transcripts [41,54,57]. Nonetheless, ESRP1 and ESRP2 downregulation by TGF-β regulates CD44 alternative splicing and also leads the phosphorylation of the T179 residue of SMAD3, mediating an interaction of this latter gene product with PCBP1, an RNA-binding protein [59]. This SMAD3-PCBP1 complex interacts with the variable exon region of CD44 pre-mRNA, inhibiting spliceosome assembly and favoring the expression of the mesenchymal isoform CD44s [59]. Finally, another splicing factor, hnRNP M, competes with ESRP1 to bind to the *cis*-regulatory element in the CD44 pre-mRNA, leading to the generation of CD44s by skipping the variable exons and, thus, promoting EMT and contributing to breast cancer metastasis [60].

Although CD44s is more frequently associated with cellular malignancy, CD44v has also been described in tumorigenesis. A study revealed that a subpopulation of breast cancer stem cells with increased expression of ESRP1 and CD44v (mainly CD44v3-10) had enhanced metastatic capacity to invade lung tissue [61]. Furthermore, in the lung microenvironment, OPN interacts with CD44v through the V3-V7 exon region, but not CD44s, promoting cancer cell invasion in an ESRP1-dependent manner and lung colonization [61]. In fact, CD44-dependent chemotaxis is induced by OPN, while CD44-dependent cell aggregation can be stimulated by hyaluronan, a glycosaminoglycan that belongs to the extracellular membrane [51,62].

CD44v has been correlated to malignancy in colorectal cancer stem cells, where CD44v6 expression and its cooperation with c-MET drive a highly tumorigenic population with metastatic potential and have been correlated with a poor prognosis [63]. This expression is regulated by HGF, OPN and SDF-1, which induce the expression of EMT-related genes [63]. Furthermore, CD44v6 acts as a coreceptor and activates c-MET and VEGFR-2 in response to HGF and VEGF-A, inducing endothelial cell migration, sprouting and tubule formation [64]. Some SR proteins (SRSF1, SRSF6 and SRSF9) also regulate CD44 splicing, inhibiting formation of the CD44v6 isoform, while SRSF2 reduces the expression of V6 but increases the expression of isoforms containing exons V6-10 and V6, V8-V10 [65]. CD44v6 has been suggested to be important for drug resistance since its knockdown prevents colony formation in response to several antitumoral drugs [66]. In addition, CD44v6 knockdown reduces the expression of mesenchymal markers, such as vimentin, Snail1/2 and Twist [66]. Another study showed that hnRNP LL is involved in the regulation of CD44 splicing in colorectal cancer since the knockdown of this splicing factor leads to increased expression of CD44v6 [67]. In colon cancer cell lines and clinical samples, the downregulation of hnRNP LL is associated with EMT and contributes to the increased cell invasion ability in these samples [67].

### 3.4. Mena

Human enabled homolog (Mena), a relative of enable/vasodilator-stimulated phosphoprotein (ENA/VASP), is an actin regulatory protein that influences cell adhesion, protrusion, motility and invasion [68]. The upregulation of this protein has been linked to cell malignancy in breast, pancreas, colon and cervical cancers [68]. In agreement with this statement, Mena deficiency is related to reduced invasion, intravasation and metastatic dissemination [68]. The variant isoforms of Mena that are produced during tumorigenesis are Mena$^{11a}$, MenaΔv6 and Mena$^{INV}$ [68–71]. The first one is an epithelial-associated isoform that has 21 additional amino acids in the EVH2 domain, which is encoded by exon 11a and is described in highly cohesive and well differentiated tumor phenotypes, whereas is downregulated in invasive cells and ones that have undergone EMT [68,69]. In fact, the epithelial regulatory splicing factors ESRP1 and ESRP2 regulate exon 11a inclusion, which contributes to epithelial phenotype and the formation of tumors with a differentiated morphology [68]. Furthermore, Mena$^{11a}$ contributes to survival and resistance to phosphatidylinositol-3-kinase (PI3K) inhibition therapies [72]. MenaΔv6 lacks exon 6 and is related to increased invasiveness and drug resistance of breast tumors in the absence of Mena$^{11a}$ [71]. In pancreatic cancer, TGF-β1 differentially regulates MenaΔv6 upregulation along with the downregulation of Mena$^{11a}$, promoting EMT [70].

Mena$^{INV}$ has an additional 19 amino acids codified by the 'INV' exon. Its expression is not enriched in highly proliferative cells and reveals an inverse correlation of E-cadherin expression in breast cancer [69]. In primary tumor from breast cancer, this isoform contributes to a discohesive morphology and the loss of epithelial cell-cell contact, as well as increased cell motility [73]. The expression of Mena$^{INV}$ is responsible for sensitizing tumor cells to EGF, eliciting tumor cells to a more invasive phenotype in low EGF microenvironments [68,69]. Additionally, a subpopulation of highly motile cancer cells can respond to the chemotactic stimuli provided by macrophages through EGF and promotes CSF1 production by tumor cells [74,75]. This motile tumor cells show increased expression of both Mena$^{classic}$ and Mena$^{INV}$ and decreased expression of the Mena$^{11a}$ isoform [75,76]. Mena-expressing tumor cells and macrophages mutually exert chemotaxis to each other so tumor cells can co-migrate with macrophages towards blood vessels and, consequently, intravasate at sites enriched with perivascular macrophages [74,75]. These intravasation sites are called tumor microenvironment of metastasis, and the number of these regions are positively correlated with the Mena$^{INV}$ abundance [75,76]. Furthermore, macrophages also can facilitate cancer cell intravasation by physically interacting with perivascular macrophages and tumor cells, increasing the formation of invadopodia by inducing RhoA activity and leading to macrophage-dependent transendothelial migration [76]. In this case, Mena$^{INV}$ was observed to stimulate the maturation of invadopodia by promoting the phosphorylation state of CTTN at

the tyrosine 421 residue and inhibit normal dephosphorylation mediated by phosphatase PTP1B [77].

### 3.5. Rac1

Rac1 belongs to the Rho family of small GTPase and acts by cycling between an inactive GDP-bound state to an active GTP-bound state in response to different stimuli [78]. Rac1 is crucial in the regulation of the cell cycle, cytoskeletal dynamics and formation of filopodia, lamellipodia and stress fibers [79]. Rac1 mRNA is alternatively spliced by the inclusion of exon 3b, generating Rac1b, which shows the insertion of 19 amino acids at the end of the switch II region, leading to a predominantly GTP-bound state, preventing the inactive state [78]. This mRNA isoform is overexpressed in colorectal, lung, pancreatic, thyroid and breast cancers [79,80].

In breast cancer, hnRNP A1 prevents Rac1b expression, whereas the activation of SRSF1 (through phosphorylation mediated by SR kinases (SRPK1)) in colorectal cancer is related to increased expression of Rac1b [81,82]. A study has shown that upon EGF treatment, hnRNP A1 is ubiquitylated and retained in the cytoplasm, preventing its repressive splicing activity [83]. Furthermore, EGF signaling stimulates the action of SR kinases, which in turn leads to the activation of SR proteins that mediate exon 3b inclusion, generating Rac1b and promoting cell migration [83]. Conversely, Rac1b promotes growth transformation, causing a loss of density and anchorage-dependent growth regulation in the NIH3T3 cell line in comparison to Rac1 [79].

However, Rac1b lacks some Rac1 functions and is incapable of forming lamellipodia or interacting with PAK1 and Rho-GDI [78]. Rac1b also induces NF-κB activation, which stimulates proliferation and apoptosis resistance in neoplastic thyroid cells [78,84].

### 3.6. CTTN

Cortactin (CTTN) is a cytoplasmic F-actin binding protein and acts as a Src substrate, participating in the assembly and cross-linkage of F-actin, both of which are controlled Rac1 activity [85,86]. Branched actin offers structural support for the plasma membrane and, in response to a specific stimulus, provides lamellipodial protrusions by branching of existing filaments, which requires the activity of CTTN to promote Arp2/3 complex activation leading to actin nucleation [85]. CTTN has three transcriptional isoforms: CTTN-A, CTTN-B and CTTN-C [23]. CTTN-A is the dominant isoform and has a region that consists of six motifs containing tandem amino acids repeats, where the 3$^{rd}$ to 6$^{th}$ units are codified by exons 8, 9, 10 and 11, respectively [87]. These six units represent the full functional domain responsible for filamentous actin-binding, cross-linking and cell migration [23,87]. CTTN-B and CTTN-C have five and four CTTN repeats, respectively, and bind to F-actin less robustly [87]. In this sense, cells overexpressing CTTN-B and CTTN-C exhibit reduced migration in comparison to cells overexpressing CTTN-A due to the absence of the full F-actin binding domain [23]. In colorectal cancer, hnRNP I promotes the inclusion of exon 11, leading to the overexpression of CTTN-A in comparison to normal tissue and increasing cell migration and invasion [23]. However, when hnRNP I was knocked down, the total mRNA of CTTN remains the same, although CTTN-A levels decline, indicating the preference of hnRNP I in the generation of CTTN-A in colorectal cancer [23]. This isoform is also overexpressed in B cells of patients with chronic lymphocytic leukemia, which suggests that it is related to disease progression [88].

### 3.7. CPEB2

In triple-negative breast cancer (TNBC), a subpopulation with anoikis resistance has been observed to exhibit altered pre-mRNA splicing of cytoplasmic polyadenylation element binding 2 (CPEB2), a translation regulator [89]. CPEB2 is associated with TNBC resistance by regulating cellular stress response [89]. CPEB2 is localized to RNA stress granules, where it binds to specific transcripts, maintaining them in an inactive state [89]. During the stress response, dissociation of the CPEB2 target transcript occurs, and the transcript is translated [89]. However, altered alternative splicing of CPEB2 leads to CPEB2B isoform expression in a subpopulation of TNBC cells [89]. This alternative isoform includes exon 4, is sufficient to promote anoikis resistance and is associated with higher metastasis and mortality rates [89].

### 3.8. KAI1

The transmembrane glycoprotein Kangai1 (KAI1, also known as CD82) has been described as a metastasis suppressor, and its loss or downregulation is correlated with a poor prognosis in prostate, lung, gastric, breast, bladder, colon, skin, cervix, endometrium and ovary tumors [90,91]. The expression of KAI1 reduces cell motility and invasiveness through the cross-talk with integrin signaling, adherens and tight junctions components [92]. A splice variant of KAI1 (KAI1-SP) has been detected in gastric cancer and associated with a poor prognosis and increased invasive potential *in vivo* [90,91]. KAI1-SP lacks exon 7 encoding 28 amino acids from the distal part of the second extracellular loop (ECL2) to the proximal part of the fourth transmembrane region [90,91]. Colon carcinoma cell lines transfected with KAI1-SP display a greater binding capacity to ECM components, such as type IV collagen and fibronectin, compared to those transfected with wild-type KAI1 [91]. In this study, KAI1-SP transfected cells showed increased motility mediated by cell-ECM adhesion components (especially fibronectin) and a weaker interaction with integrin α3β1, the major receptor for laminin [91]. Successful cell migration depends on an intermediate level of cell adhesiveness to the underlying matrix, but stronger cellular attachment may be related to cell immobilization [90]. In this sense, fibronectin and its receptors have been reported to act not only in attachment and cellular adhesion but also to provide signals that regulate cellular locomotion [91]. Therefore, *in vitro* ovarian cancer cells expressing KAI1-SP display decreased cell adhesiveness to vitronectin and type I collagen and increased cell motility mediated by αvβ3 integrin and vitronectin [90]. Finally, KAI1-SP fails to promote E-cadherin-mediated cell-to-cell contacts, which may prevent cell spreading [90].

### 3.9. CD147

CD147, called basigin, is a highly enriched glycosylated transmembrane protein found on the plasma membrane of various human cancer cells and is related to the increased tumorigenicity and invasive potential of such cells [93]. The most known activity is related to the expression, secretion, and activity regulation of matrix metalloproteinases (MMPs) [94]. A study has shown that CD147 expression increases cancer cell viability by modulating intercellular contacts through fibronectin assembly and preventing anoikis by downregulating Bim via the MAP kinase pathway [93]. CD147 has four isoforms, termed CD147/Bsg-1 to 4. Among them, the most predominant is CD147/Bsg-2, while CD147/Bsg-1 is retina-isoform specific and CD147/Bsg-3 and −4 were first identified in human endometrial cells and cervical carcinoma cell lines [94]. In hepatocellular carcinoma, the expression levels of CD147/Bsg-2 and −3 were higher compared to adjacent normal tissue [94]. However, these isoforms have contrasting effects, while CD147/Bsg-2 overexpression promotes cell proliferation and invasiveness, CD147/Bsg-3 is responsible for inhibiting the aforementioned processes, possibly by hetero-oligomerization of CD147/Bsg-3 with CD147/Bsg-2, thus preventing the invasive role of the latter [94].

### 3.10. RON

The *recepteur d'origine nantais* (RON) proto-oncogene encodes a transmembrane tyrosine kinase receptor and is activated by interacting with MSP, leading to RON kinase activity upregulation [95]. RON downstream signaling involves the regulation of many intracellular processes, such as EMT stimulation, cell proliferation, survival, migration, invasion, chemoresistance and angiogenesis [95]. In normal and malignant cells, two transcripts encoded by different *RON* gene promoters are observed. These transcripts are the short-form RON (SF-RON) and the full-length RON, although different isoforms can be generated through post-transcriptional modifications (mostly alternative mRNA splicing as well as protein truncation) [95].

SF-RON is generated by an alternative transcription start site controlled by a second promoter located between introns 8 and 10 [95]. This isoform is constitutively active with ligand-independent activity [96]. In ovarian cancer, SF-RON expression is related to increased tumor growth and spreading to the abdominal cavity [96]. Additionally, the RONΔ165 variant is formed by an in-frame exon 11 and results in the accumulation of the single-chain pro-RONΔ165 [97]. Furthermore, RONΔ165 is constitutively phosphorylated by oligomerization due to the deletion of exon 11 and is involved in mobile and invasive phenotypes [97,98].

RONΔ170 is a defective variant, the alternative splicing of which generates the loss of exon 19, which leads to the absence of kinase activity, inhibiting the oncogenic activity mediated by another oncogenic variant, RONΔ160 [97,98]. RONΔ160 results from an in-frame deletion of 109 amino acids encoded by exons 5 and 6 [95,97]. RONΔ160 and RONΔ155 are associated with tumor growth and metastasis, and RONΔ160 leads to the accumulation of β-catenin and the increased gene expression of β-catenin targets [97,98]. RONΔ155 carries deletions of exons 5, 6 and 11, and due to the loss of exon 11, the translated protein is retained in the cytoplasm [95,97,98].

Moreover, RONΔ110 is a constitutively active proteolytic truncated isoform that is present on the cell surface of cancer cells overexpressing RON, but its role in tumorigenicity it is not completely understood [95,97,98]. Among the splicing factors, SRSF1 is responsible for inhibiting the inclusion of exon 11, while hnRNP A2/B1 and hnRNP H are also described to mediate glioma cell transformation through exclusion of exon 11 from the proto-oncogene *RON* [99–101].

### 3.11. TPM1

Tropomyosin-1 (TPM1) belongs to a family of actin-binding proteins that provide stability to actin filaments in non-muscle cell cytoskeleton and the contractility of skeletal and smooth muscle cells, as well as affecting integrin expression and activity [102–104]. Previous studies have shown that TPM1 is downregulated in bladder, breast, colorectal, esophageal, oral cavity and oropharynx, ovarian and renal cancers [102,105]. Apoptosis evasion and cell invasiveness may be associated with disrupted cytoskeleton structures and components. Hence, TPM1 seems to have a crucial role as a potential tumor suppressor by mediating cytoskeletal reorganization and preventing metastasis [102,105,106]. In this sense, the restoration of TPM1 expression in breast cancer cells promotes cytoskeletal reorganization and recovery of cell-ECM interactions, reestablishing cell adhesion and promoting anoikis of detached cells [102].

The TPM1 isoforms are classified according to the number of amino acids in length as high molecular weight (HMW) or low molecular weight (LMW) and result from alternative promoter usage [103]. The use of exon 1a and 2a (or 2b) generates HMW, whereas exon 1b is used for the LMW isoform [103]. Furthermore, alternative splicing of exon 6 may occur via mutually exclusive splicing of 6a (or 6b) and exon 9, and 14 isoforms have been identified to date [103,106]. Among them, TPM1δ and TPM1γ are HMW expressed in normal human breast epithelial cells (but downmodulated in human breast cancer), the product of which is incorporated into stress fibers [106]. In fact, the downregulation of HMW isoforms is related to increased anoikis resistance, and tumorigenesis and malignant cells often undergo to a switch from HMV to LMW [103,104]. By contrast, the HMW isoform TPM1λ is upregulated in malignant breast cell lines and shows an inverse correlation with stress fiber formation [106].

Other tropomyosin isoforms, such as the LMW TM5NM1 and TM4, have been related to metastasis progression in melanoma and breast cancer, respectively [103]. This information indicates that their regulation is crucial for the fate of malignant cells that can potentially undergo anoikis when specific isoforms are expressed or possess increased invasiveness through isoform switching.

### 3.12. IL-7

Interleukin 7 (IL-7) is a cytokine related to tumor development and progression and, together with its receptor IL-7R, has been

identified in several cancers, such as colorectal, renal, breast, lung and central nervous system cancers [107]. An IL-7 variant lacking exon 5 (IL-7δ5) has been related to promoting cellular proliferation and cell cycle progression in breast cancer via PI3K/Akt pathway activation [107]. Furthermore, IL-7δ5 is able to induce EMT through the PI3K/Akt pathway, in addition to exogenous IL-7δ5, that promotes downregulation of E-cadherin and upregulation of N-cadherin expression [107].

### 3.13. Chemokines

Chemokines are chemotactic cytokines that are subdivided into four main classes (CC, CXC, C and $CX_3C$) according to the position of the first two cysteine residues [108]. Considering the tumor microenvironment, chemokines can be secreted by tumor cells or stromal and immune cells [108]. Specific chemokines drive immune cell homing to the tumor microenvironment, influencing immune-cell infiltration and affecting tumor progression [108]. Furthermore, chemokines can target tumor and stromal cells when these cells express chemokine receptors and are able to respond to different chemokine gradients [108]. In this sense, chemokines affect cell proliferation, invasiveness, cancer stem-like cell properties and metastasis [108]. Several chemokine receptors have been described to undergo aberrant splicing, which may contribute to tumor progression and metastasis.

CXCR3 is a receptor for CXC chemokines and is activated by the ligands CXCL4/PF4, CXCL9/MIG, CXCL10/IP10 and CXCL11/IP9/I-TAC [109]. The activity observed for this receptor is related to its two major splice isoforms: CXCR3-A, formed by exon 1 and exon 3 and considered the classical isoform, and CXCR3-B, which is assembled by exon 2 and exon 3 [110]. Interestingly, these isoforms have opposing functions, CXCR3-A is a pro-proliferative isoform that is related to invasion and migration promotion in several tumors, while CXCR3-B has no chemotactic functions and inhibits migration [109,110]. CXCR3-A, activated by its ligand CXCL9, has been shown to change the levels of ERK1/2 phosphorylation, a member of MAPK signaling, leading to the overexpression of MMP2 and MMP9 and, consequently, contributing to the invasion and metastasis of $CD133^+$ liver cancer cells [111]. CXCR3-A is expressed at higher levels than CXCR3-B in malignant cells and supports cell proliferation and migration, invasiveness and metastasis, whereas CXCR3-B reduces the metastatic potential and colonizing capacity [109,110]. However, in bulk populations, CXCR3-B inhibits cell proliferation and shows a supportive role for cells with cancer stem-like properties [110]. In this sense, CXCR3-B may contribute to resistance in therapies that target highly proliferative cells after those stem cells have achieved a low proliferation rate, contrasting with general populations of malignant cells [110].

SDF-1, also known as CXCL12, is a chemokine member of the CXC class and the ligand of the chemokine receptors CXCR4 and CXCR7 [112]. Serum SDF-1 concentration is positively correlated with the occurrence of distant metastasis in patients with gastric cancer [113]. This chemokine has six alternative spliced isoforms, comprising SDF-1α, SDF-1β, SDF-1γ, SDF-1δ, SDF-1ε and SDF-1θ [112]. In tumors, there is an increase in the expression of the SDF-1α G801A polymorphism associated with lymph node metastasis in patients with colorectal cancer, whereas SDF-1β is involved in angiogenesis and, together with SDF-1γ, is related to tumor size in patients with metastasis [112].

### 3.14. Integrin

Integrins are the major cell surface receptors and are responsible for mediating cell adhesion to the ECM [114]. They are crucial for the maintenance of homeostatic balance in the cell, but are also involved in inflammation, leukocyte trafficking, tumor metastasis and wound repair [114]. Integrins can transmit an 'outside-in' signal through the plasma membrane, as well as an 'inside-out' signal, functioning as an important intermediate a bidirectional communication between cells and their microenvironment [114]. In tumors, integrins regulate different processes through the tumor cell-microenvironment interface, influencing the cancer response and progression [114].

The α6 integrin subunit (CD49f) can heterodimerize with either β1 or β4, forming α6β1 or α6β4 integrins and then acting as the main laminin receptor [115,116]. The α6 subunit undergoes alternative splicing, forming two cytoplasmic isoforms: α6A, which is enriched in epithelial populations, and α6B, which is enriched in mesenchymal populations, where the latter has an important role in slow-cycling stem cell functions and is directly involved in the stem cell phenotype; it is related to a poor outcome in TNBC patients [116]. In this sense, VEGF signaling can affect the alternative splicing of the α6 subunit, promoting the repression of α6A and favoring the expression of α6B via induction of the oncogene GLI1 and transcriptional inhibition of the splicing factor ESRP1 [116]. However, isoform α6A knockdown reduces tumor growth, although it does not affect α6B levels [115]. In fact, α6A has been described to have a positive proliferative effect in colorectal cancer, which is involved in the Wnt/β-catenin pathway, where the knockdown of α6A leads to enhanced phosphorylation of β-catenin by GSK3β [115].

Integrin β1 alternative splicing leads to the formation of four isoforms, called β1A, β1B, β1C and β1D [117]. The expression levels of isoforms β1B, β1C and β1D are downregulated in several tumors, whereas β1A is related to focal adhesion formation, cell spreading and motility [117]. A study has shown that β1A expression is related to metastasis in pancreatic cancer, while Fyn inhibition and hnRNP E1 overexpression promote the splicing of β1C, which reduces the migration and invasiveness of tumor cells [117].

### 3.15. FAK

Focal adhesion kinase (FAK) mediates cell communication with the extracellular environment, triggering intracellular signaling in response to ECM remodeling, nutrient availability and growth factors [118]. Furthermore, its action is involved in cell adhesion, motility, survival, vasculogenesis and proliferation [118]. During tumorigenesis, FAK overexpression is frequently observed in breast, colon, cervical, ovary and

prostate cancers, and it is associated with metastasis [118,119]. When detached cells undergo anoikis, FAK is proteolyzed by caspase-3 and −7, and anoikis evasion is recurrently associated with increased FAK levels or mutations [119]. FAK shows four traditional alternatively spliced exons: 13, 14, 16 and 31 [119,120]. However, in human breast cancers, an aberrant alternative isoform that lacks exon 26, a region that contains the caspase-3/caspase-7-like cleavage motif, was observed [119,120]. Thus, although this isoform has the same kinase activity as wild-type FAK, the missing segment causes aberrant FAK to become resistant to caspase-mediated proteolysis, promoting tumor cell survival [119].

### 3.16. OPN

Osteopontin (OPN) is a glycophosphoprotein that is secreted by several cells and has been broadly overexpressed in plasma and tumor tissues, as well as described to promote tumor progression and metastasis [121]. OPN interacts with different integrins and CD44, activating signal transduction pathways involved in cell proliferation, adhesion, migration and invasion [121].

The primary transcript of OPN can be alternatively spliced in the upstream region of the central integrin binding domain and the C-terminal CD44 binding domain to form (i) the full-length OPN-a, which consists of 7 exons; (ii) the isoform OPN-b, which lacks exon 5; (iii) and OPN-c, which is generated by the lack of exon 4 [121].

A study has shown that OPN-a and OPN-b both act through integrin signaling, while OPN-c has an integrin-independent function [122]. Furthermore, OPN-b augments cell migration *in vitro* compared with OPN-c [122]. In the same study, stable cells transfected with OPN-b showed increased cell adhesion, proliferation and invasion, while OPN-c transfected cells exhibited decreased cell invasion and migration but enhanced cell detachment [122]. OPN-c is present in 75% of breast cancers, and it is directly related to anchorage-independent growth and metastasis [123]. In non-small cell lung cancer, OPN-a is highly expressed compared to OPN-b and OPN-c [124]. However, both OPN-b and OPN-c are overexpressed in gastric cancer and showed anti-apoptotic activity when induced by $H_2O_2$, but not OPN-a [125]. This pro-survival effect is promoted by OPN-b, possibly through the interaction with CD44v6 and CD44v7, which triggers downstream anti-apoptotic signaling [125].

In addition, OPN-a overexpression has been related to the promotion of proliferation, migration, adhesion and invasion in A549 cells through the interaction with αvβ3 integrin [124]. This interaction allows lung cancer cells to invade the bone matrix during metastasis, and once OPN-a is silenced, significant inhibition of bone matrix attachment occurs [124]. In thyroid cancer cell lines, increased levels of OPN-a have been related to cell growth and enhancement of cell migration and motility, possibly through the induction of MMP2 and MMP9 activity, leading to extracellular matrix degradation [126]. In papillary thyroid carcinoma, OPN-a is correlated with extrathyroid extension, vascular invasion and $BRAF^{V600E}$ mutation, which are significantly related to a poor prognosis [126].

Finally, in breast cancers, all tumor-derived OPN transcriptional isoforms increase the tumor expression of the immunosuppressive cytokines TGF-β1 and MCP-1 [127]. Additionally, these tumor-derived OPN transcripts lead to the inhibition of pro-inflammatory cytokines TNF-α and increase anti-inflammatory IL-10 production by monocytes [127]. This mechanism may be related to alternative immune cell activation, which is associated with tumor-associated macrophages (TAMs) and possibly contributes to immunological evasion mediated by OPN [127].

### 3.17. FN

Fibronectin (FN) is a glycoprotein present in the plasma and ECM that has cell adhesiveness and migration activities [128]. Fibronectin is formed by the combination of three repeating domains, called Types I, II and III [128]. The Type III domain is alternatively spliced by the inclusion of extra domain-A (EDA) or extra domain-B (EDB) and a variable type III connecting segment (IIICS), which can be excluded (or included) completely or partially [128,129]. FN can also be found either as a plasmatic FN (pFN) or as a cellular FN (cFN), and only the latter comprises variable proportions of EDA or EDB [128].

In normal tissues, these isoforms are not detected by antibodies, but they were highly representative in ECM components of aggressive tumors and other diseases that exhibit considerable tissue remodeling, serving as markers for tumor angiogenesis [128,130]. EDA-FN has been related to the promotion of neolymphangiogenesis in colorectal carcinoma cells through integrin α9 activation, in addition to the stimulation of angiogenesis, cell adhesion, migration and proliferation in malignant tumors [131]. EDB-FN can be regulated by ECM stiffness, which mediates splicing through PI3K/Akt signaling activation, promoting the regulation of SR protein phosphorylation, such as SRp40 [132]. The high level of proteolytic degradation and remodeling of the stroma surrounding the tumors alters the stiffness, enabling the invasion of migrating cells [133]. This remodeling leads to the release of activated growth factors stored in the ECM, influencing cell signaling [133]. Finally, in urinary bladder cancer, IIICS shows superexpression by 21.5-fold in invasive cells compared with noninvasive cells, as well as increased *de novo O*-linked glycosylation [129,134].

Another alternatively spliced FN isoform known as migration stimulating factor (MSF) is overexpressed by TAMs [133,135]. The MSF isoform consists of intron 12 retention followed by a two-step series of intra-intronic cleavage events, which results in a 70-kDa truncated protein [133,136]. Furthermore, MSF is highly chemotactic for monocytes and tumor cells, stimulating cell migration and possibly contributing to metastasis, which it is not observed for full-length FN [135]. In established tumors, M2-like macrophages become TAMs and show poor cytotoxicity against tumor cells, acting in the production and secretion of chemotactic molecules that stimulate tumor cell migration to distant sites and metastatic niche preparation [135]. MSF has been reported to be upregulated in several tumors, and in addition to TAMs, MSF is also expressed by neoplastic cells and vascular endothelial cells [133,136].

### 3.18. Tp53

The tumor suppressor TP53 is well known for its multiple roles in cell cycle arrest, genomic stability, DNA repair, apoptosis, autophagy, differentiation, stemness, migration, drug resistance, metabolism and angiogenesis [137]. The relationship of TP53 with cancer is frequently associated with mutations that are located in the TP53 DNA binding domain (DBD) or its conformational mutants [137]. *TP53* encodes twelve isoforms through different mechanisms, which involve the usage of the alternative promoter and splicing: TAp53α, TAp53β, TAp53γ, Δ40p53α, Δ40p53β, Δ40p53γ, Δ133p53α, Δ133p53β, Δ133p53γ, Δ160p53α, Δ160p53β and Δ160p53γ [137,138]. TAp53 and Δ40p53 are long TP53 isoforms that contain the transactivation domain, which it is not present in either of the short isoforms Δ133p53 or Δ160p53 [137]. Furthermore, β and γ have an additional domain with an unknown function and do not contain a canonical C-terminal oligomerization domain [137,138]. The isoform TAp53α is the canonical isoform expressed and found in normal tissues, but Δ133p53β has been shown to increase the risk of recurrence by eight times and the risk of death by three times in breast cancer [139]. High levels of Δ133p53 positively regulate the cancer stem cell potential in the MCF-7 cell line, stimulating mammosphere formation and the expression of pluripotency and stemness regulators, such as SOX2, OCT3/4, NANOG and CD24/CD44 [140]. Concomitantly, the metastatic capacity, invasiveness and chemoresistance are associated with increased Δ133p53 expression in different cancer cell lines, which is not associated with *TP53* mutations; the depletion of Δ133p53 in wild-type *TP53* eliminates the scattering and invasive activities, as well cancer stem cell formation [139,140].

### 3.19. CSF-1

Colony stimulating factor 1 (CSF-1) is a hematopoietic growth factor found in the bloodstream and has been related to the survival, differentiation and proliferation of mononuclear phagocytic lineages, as well as to macrophage spreading and motility [141,142]. *CSF-1* is composed of at least 10 exons, and its mRNA undergoes alternative splicing, which it involved in the alternative use of a 3′ non-coding region represented by exon 9 and 10 [142]. Exon 6 is also alternatively spliced, during which a large portion of the 5′ end is removed in one of the alternative transcripts [142]. In addition, exon 6 has a variable length, depending on which of the two spliced acceptor sites is used, forming different transcripts with different roles [142,143]. After alternative splicing and differential post-translational proteolytic processing, CSF-1 generates three mature homodimeric proteins: a cell surface-associated glycoprotein (csCSF-1), a secreted extracellular matrix-anchored proteoglycan (spCSF-1) and a secreted soluble glycoprotein (sgCSF-1) [141,142]. spCSF-1 and sgCSF-1 are secreted isoforms and regulate cells nearby or at a distance, while cell surface isoforms act on local regulation [141,142].

CSF-1 has an important role in the chemotaxis and activation of macrophages, which are important for innate immunity and for tissue morphogenesis and/or remodeling [141]. CSF-1 is related to cancer, and its activity has been associated with the recruitment and re-education of macrophages, which are also known as TAMs [133,144]. In fact, blood circulating monocytes are recruited locally and differentiate into TAMs in response to signals from the tumor microenvironment, such as chemokines and growth factors secreted by stromal and tumor cells [133,144]. TAMs are regarded to have an M2-like phenotype and are suggested to have pro-tumor effects due to their higher levels of anti-inflammatory cytokines, scavenging receptors, angiogenic factors and proteases, promoting an immunosuppressive microenvironment and contributing to tumor progression [144,145]. This ability results from to secretion of molecules that directly affect cell growth, motility, migration, angiogenesis, intravasation, invasion, inflammatory response, ECM remodeling and proliferation [135,145].

Different CSF-1 isoforms have been described to be associated with TAM activation [145]. Elevated levels of CSF-1 have been detected in breast, ovarian, endometrial, cervical, and colorectal cancers, leukemia, lymphoid malignancies and myelodysplastic syndromes [145]. Among the isoforms, sgCSF-1 is related to poor survival outcomes in patients with colorectal and breast cancer [145]. One study detected more infiltrating TAMs in the microenvironment containing csCSF-1 and sgCSF-1 isoforms [145]. Both isoforms were able to promote macrophage re-education, but they induced different spectra of M2 phenotype-related genes [145]. In this sense, csCSF-1 is more related to the stimulation of TAM proliferation in the tumor microenvironment, whereas sgCSF-1 is more prone to the recruitment and induction of TAM differentiation [145].

### 3.20. Prrx1

Alternative splicing of the paired-related homeobox transcription factor (Prrx1) results in two isoforms, termed Prrx1a and Prrx1b, which differ from each other at the C-terminus but are identical in the 199 N-terminal amino acids [35,146]. Both isoforms maintain the homeobox domain, but only Prrx1a has a conserved OAR domain with potentially important functions in transactivation [35,146].

In pancreatic carcinogenesis, Prrx1 isoform switching regulates epithelial and mesenchymal phenotypes in primary and metastatic cancer: Prrx1b is upregulated in primary tumor cells, where it promotes EMT, tumor invasion and differentiation [35]. In addition, Prrx1b facilitates tumor cell invasion by upregulating HGF through signaling mediated by its receptor (c-MET), driving metastasis by promoting the invasion of tumor cells and EMT [35].

However, higher levels of Prrx1a expression have been found in metastatic cells, contributing to MET, which is crucial to the colonization of distant organs after extravasation, as well as an increased capacity for anchorage-independent growth and self-renewal [35]. Furthermore, knockdown of both isoforms in the KPC2 cell line decreased the invasiveness and self-renewal capacity, and *in vivo* transplantation generated a smaller primary tumor volume and a reduced number of circulating tumor cells [35].

### 3.21. EGFR

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein and a member of the ErbB receptor tyrosine kinase family of growth factor receptors [147]. It is described as providing a substantial contribution to cell survival and proliferation motility [147]. Its overexpression indicates a poor prognosis in breast, lung, liver, prostate and ovarian cancers [147].

EGFR is composed of 30 exons that can be spliced to form EGFRv1, EGFRv2, EGFRv3, EGFRv4, EGFRvIII and de4 EGFR [148–150]. The EGFRvIII isoform results in a gain-of-function mutation that leads to genomic deletion of exons 2–7, and its expression is restricted to tumors, such as lung carcinoma, glioma, glioblastoma and head and neck squamous cell carcinoma [150]. de4 EGFR forms a truncated EGF variant due to a deletion of exon 4, and it has been detected in glioma, prostate and ovarian cancers, and like EGFRvIII, de4 EGFR is also found exclusively in tumoral tissues [150,151]. Both de4 EGFR and EGFRvIII have minimal EGF-binding activity and ligand-independent autophosphorylation, exhibiting similarities in their biological activities and extracellular structures [151]. Some de4 EGFR tyrosine residues are constitutively phosphorylated even without ligand activation, which leads to downstream signaling pathway activation [147,150]. In this sense, de4 EGFR promotes the upregulation of FAK, Src, ERK, Akt and MMP9, whereas E-cadherin is downregulated in ovarian cancer cells [147]. Furthermore, de4 EGFR contributes to the increase in cell proliferation and malignant transformation, stimulating metastasis-promoting activities and higher cisplatin tolerance [147].

### 3.22. Id1b

The inhibitor of differentiation (Id1) belongs to the bHLH family consisting of three other members, Id2 to Id4 [152]. Its role in tumorigenesis, angiogenesis and metastasis has been described, and it is associated with a poor prognosis in patients with glioblastoma and breast cancer [152]. Id1 has an important role in maintaining self-renewal and pluripotency in embryonic stem cells, and its role extends to cancer stem cells [152]. The alternative splicing of Id1 forms two isoforms: the classical isoform, known as Id1a, containing 155 amino acids; and the alternative isoform, referred as Id1b, containing 149 amino acids [152]. The difference between the two isoforms is found in the C terminal region, which alters their capacity to interact with specific transcription factors [152]. Id1a is related to increased cell proliferation, whereas overexpression of Id1b reduces proliferation, promotes self-renewal, and leads to a stem cell-like phenotype, increasing the expression of cancer stem cell markers, such as ALDH1A1, Notch-1, Sca-1, Tert, Sox-2 and Oct-4 [152]. Thus, Id1b may contribute to cell quiescence, a strategy adopted by many primary tumor or metastatic cells [152,153]. These quiescent cells are maintained in a non-proliferative state for long periods, and the awakening of these dormant cells leads to tumor progression and metastatic relapse [153,154].

### 3.23. XAF1

XIAP-associated factor 1 (XAF1) is a nuclear protein that when overexpressed leads to apoptosis and tumor growth inhibition [155]. XAF1 splice variants have been detected in several cancer cell lines and are regulated by aberrant epigenetic modifications [155]. The aberrant variants frequently contain a PTC, which leads to NMD [155]. However, suppression of the NMD pathway may occur due to cellular stresses influenced by the tumor microenvironment, stabilizing the NMD-target mRNA [155]. In this sense, the upregulated XAF1 variant was found to lack exon 5 and contain an extended region from exon 3, termed XAF1F [155]. XAF1F harbors a PTC and was found in highly metastatic cell lines compared to parental cells and in peripheral blood containing circulating tumor cells from patients with gastric cancer, indicating a suppression of the NMD pathway and, consequently, an accumulation of XAF1F [155]. Furthermore, although the function of XAF1F has not yet been reported, its mRNA expression levels in circulating tumor cells is correlated with venous invasion, lymph node metastasis, and the depth of tumor invasion [155].

### 3.24. FAM3B

Family with sequence similarity 3 member B (FAM3B) is a member of the FAM3 family of cytokine-like genes and has been described to have at least seven alternatively spliced isoforms [156]. Among them, FAM3B-258 is a non-secretory variant with 258 amino acids that is highly expressed in colon adenocarcinomas. It is related to poor cancer cell differentiation and has been shown to promote invasion and migration *in vitro* and metastasis in nude mice [156]. FAM3B-258 promotion of invasion is regulated by Slug-dependent activation, which inhibits the expression of JAM and E-cadherin, leading to the breakdown of cell-cell interactions by downregulating cellular adhesion proteins [156].

### 3.25. KLF6

The transcription factor Krüppel-like factor 6 (KLF6) encodes a zinc finger protein that regulates the cell cycle, differentiation, and apoptosis [157]. KLF6 acts as a tumor suppressor and is inactivated in many cancers [157]. KLF6 alternative splicing generates three cytoplasmic isoforms: KLF6-SV1, KLF6-SV2 and KLF6-SV3 [158]. In prostate cancer, a germline single nucleotide polymorphism located in an intronic region of KLF6 is responsible for disruption of their own splicing regulation by creating a novel functional SRp40 DNA binding site and removing two overlapping SR protein sites [158]. This process leads to increased expression of three alternatively spliced isoforms from *KLF6* and an augmented risk of prostate cancer [158].

KLF6-SV1 and KLF6-SV2 are described to antagonize the activity of wild-type KLF6 during p21 upregulation and to suppress cell proliferation [158,159]. Furthermore, KLF6-SV1 promotes metastasis, and its upregulation is considered to be a poor prognostic factor for patients with prostate, lung, ovarian and breast cancers [159,160]. Its expression is

detected in both healthy and tumor tissues, but it is frequently upregulated in the latter [159].

In breast cancer, KLF6-SV1 expression in primary tumors is correlated with an increased metastatic potential and EMT-like phenotype [159]. Furthermore, overexpression of KLF6-SV1 leads to increased colony formation in tumorigenic and non-tumorigenic cell lines, promoting cell survival, migration and invasion [159]. KLF6-SV1 knockdown decreases the metastatic capacity via diminished invasion and reduced expression of N-cadherin with concomitant E-cadherin overexpression, reverting the mesenchymal-like phenotype [159].

### 3.26. Exo70

Exocyst is an octameric protein complex composed of Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70 and Exo84 [161]. This complex is crucial for exocytosis and is related to plasma membrane expansion during migration [161,162]. Furthermore, Exo70 is crucial for cell invasion by mediating MMP secretion at focal degrading sites [162]. However, during the transition of the EMT state, the epithelial Exo70 isoform (Exo70-E) is switched to a mesenchymal Exo70 isoform (Exo70-M) by the splicing factor ESRP1 [161]. Exo70-E differs from Exo70-M by an insertion of a portion of exon 8 generated by ESRP1-mediating alternative splicing [161]. This switch does not affect exocyst complex function, but it is implicated to have different functions according to the assembled Exo70 isoform [161].

When Exo70 knockdown cell lines are transfected with Exo70-M, these cells recover their invasive ability, while Exo70-E transfection fails to do the same [161]. These observations were explained by the high invadopodial activity levels in cells expressing Exo70-M, whereas Exo70-E cells had lower invadopodial activity [161]. Additionally, Exo70-M interacts with the Arp2/3 complex and stimulates actin branching, which is crucial for membrane protrusion formation during migration and invasion [161].

### 3.27. Merlin

Moesin-ezrin-radixin-like protein (Merlin) is a tumor suppressor that participates in the regulation of cell proliferation, motility and survival [163]. Merlin is codified by the *Nf2* gene, which is heterozygous in mice and is related to the development of highly metastatic malignant tumors [163]. Merlin is important for the prevention of tumorigenesis by promoting adherens junctions stabilization and controlling cell-cell contact mediated by cadherin and activation of Rac [163]. Splicing variants of Merlin lacking exon 2 ($Merlin^{\Delta 2}$), exon 3 ($Merlin^{\Delta 3}$), or both of them ($Merlin^{\Delta 2/3}$) are related to their inactivity as a tumor suppressor [163]. In addition, another alternative splicing variant lacking exons 2, 3 and 4 ($Merlin^{\Delta 2-4}$) was identified in hepatocellular carcinoma and in portal vein tumor thrombus [163]. $Merlin^{\Delta 2-4}$ overexpression promotes migration and invasion by upregulating Twist1, and it impairs the ability of wild-type Merlin to bind to β-catenin and Ezrin/Radixin/Moesin proteins [163]. This isoform also leads to stemness activity by increasing spheroid formation and inducing the expression of stem-related genes, including Sox2, OCT4, Klf4, c-MYC and Nanog [163].

### 3.28. CD99

CD99 is a transmembrane glycoprotein that is involved in the cell adhesion, migration, apoptosis, differentiation, activation and proliferation of lymphocytes, monocyte migration, transport of several transmembrane proteins, and intercellular adhesion between lymphocytes and endothelial cells [164,165].

*CD99* produces two distinct transcripts: the full-length CD99 isoform (CD99wt) and a truncated shorter isoform (CD99sh) resulting from alternative splicing [164,165]. CD99sh carries a deletion of the intracytoplasmic fragment and an 18-bp insertion at the boundary of exons 8 and 9, which introduces an in-frame stop codon [164]. Studies in breast cancer cell lines have revealed that CD99sh stimulates MMP9 activity and has an FN-augmented binding potential, in addition to increasing cell motility and invasiveness [165]. The higher invasion capacity can result in diminished cell-cell adhesion in cells expressing CD99sh and an augmented interaction between FN-CD99sh [165]. Furthermore, CD99sh upregulates the AP-1 signaling pathway, which mediates cell motility and MMP9 expression [165]. The higher activity of proteolytic enzymes (such as MMP) allows ECM degradation, contributing to cell movement and invasion [165].

In osteosarcoma and prostate cancer cells, CD99 isoforms have opposing effects: CD99wt-expressing cells show a less migrating phenotype, with organized actin filaments and greater collagen type I adhesion, laminin and fibronectin and sensitivity to anoikis [164]; CD99sh activates c-Src kinase activity, which can act as a mediator of CD99 functions in stimulating cell motility [164].

## 4. Therapy

The differences observed between the tumor and normal transcriptome are vast as a consequence of several changes in genomic and transcriptomic regulation. In summary, the major genomic and transcriptomic changes are as follows: (i) the presence of mutations in transcription factor binding and splice sites; (ii) mutations that generates aberrant proteins; (iii) different DNA arrangements and histone modifications; (iv) changes in chromatin accessibility; and (v) altered expression of splicing factors and transcription factors [6]. In the past, drugs that act on specific cellular targets were used without regard for the aberrant isoforms generated during oncogenesis, mostly due to missing information regarding the isoform variants in a gene [6,166]. The differential expression of several isoforms encoded by the same gene might be a contributing factor to the observed variation in patient responses during drug treatment [6,166].

Studies have shown that the analysis of isoform-level expression offers an effective discrimination of cancer and non-cancer cells, providing a better identification of different subgroups within a cancer-type and allowing a robust signature to classify patients in molecular subgroups [6,167]. Furthermore, during

tumor progression, alternative splicing occurs in a disease-specific manner and contributes to the disease outcome [168]. The observation that some isoforms are detected only in tumor cells suggests that cancer-specific isoforms may be good diagnostic markers and therapeutic targets.

Validation of the isoform-specific expression and role in tumor cells highlights a therapeutic potential and an opportunity to develop drugs that target specific isoforms. However, it is important to maintain deep large-scale studies that provide information concerning whether an isoform is really cancer-specific or has a pro- or anti-oncogenic role, before developing a therapeutic design to ensure the correct drug-to-isoform (or drug-to-isoform product) assignment.

Some studies addressing cancer-specific isoforms as targets have shown promising results and are listed in Table 1. A successful example is EGFRvIII, which is the target of several immunotherapies for patients with glioblastoma, such as peptide vaccines, dendritic cell vaccination therapy, monoclonal antibodies and modified T cells [169]. Among them, the peptide vaccine rindopepimut was approved by the US FDA based on phase I and II clinical trials showing a significantly higher survival rate in patients with glioblastoma expressing EGFRvIII [169]. Advances have also been realized with cell therapy using the chimeric antigen receptor T (CAR-T) as an antitumor treatment, and CAR-T targeting the EGFRvIII isoform efficiently eliminates glioma cells [169].

The modulation of alternative splicing may reduce the severity of disease and prevent metastasis and its incidence. In this sense, different approaches to perform this manipulation have been studied, including targeting protein isoforms, isoform expression, alternative splicing through *trans*-acting elements, oligonucleotides specific to specific *cis*-acting elements in the transcript and inhibition of the specific isoform by RNAi [168]. One example is the use of compounds that bind to SF3B (a component of the U2 snRNP), leading to disruption of the early stages of spliceosome assembly [5]. These compounds are E7107, spliceostatin A and sudemycins, which promote unspliced (or incompletely spliced) pre-mRNA accumulation in the nucleus and show anti-proliferative activity *in vitro* [5]. E7107 was used in phase 1 clinical trials, but its use was suspended due to its high toxicity in patients [5,170]. Moreover, the spliceosome modulation in MYC-driven cancers has been suggested as an alternative to therapeutic intervention. MYC is a transcription factor that is frequently amplified in human cancers. It is highly related to cell transformation, and studies have revealed that the role for MYC as an oncogene is dependent on spliceosome activity [171,172].

## 5. Conclusive remarks and future directions

Although great advances have been achieved, understanding all of the mechanisms that contribute to tumor formation and progression is a great challenge. The presence or absence of specific isoforms may affect the metastatic potential of malignant cells, and variations in isoform expression from different genes are observed in cancer compared to normal tissues. Currently, the growth of large-scale studies has allowed us to integrate information and fill knowledge gaps in all areas of molecular biology – including information regarding splicing mechanisms and isoform variety. Certainly, further investigations are still necessary to understand the links between the splicing mechanisms, isoform diversity and the context of the disease. Finally, comprehension of the functional characteristics of abnormal splicing and their relationship with genomic instability is crucial for the development of novel therapeutic approaches.

## Acknowledgments

This work was supported by CNPq under Grant 302969/2016–0 and CAPES.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico [302969/2016–0];

## ORCID

Joice De Faria Poloni http://orcid.org/0000-0002-3896-6404
Diego Bonatto http://orcid.org/0000-0001-8679-2448

## References

1. Valastyan S, Weinberg RA. Tumor Metastasis : molecular Insights and Evolving Paradigms. Cell. 2011;147(2):275–292.
2. Guan X. Cancer metastases: challenges and opportunities. Acta Pharm Sin B. 2015;5(5):402–418.
3. Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9(4):274–284.
4. Steeg PS. Targeting metastasis. Nat Rev Cancer. 2016;16(4):201–218.
5. Lee SC-W, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat Med. 2016;22(9):976–986.
6. Pal S, Gupta R, Davuluri RV. Alternative transcription and alternative splicing in cancer. Pharmacol Ther. 2012;136(3):283–294.
7. Corkery DP, Holly AC, Lahsaee S, et al. Connecting the speckles: splicing kinases and their role in tumorigenesis and treatment response. Nucleus. 2015;6(4):279–288.
8. Silipo M, Gautrey H, Tyson-Capper A. Deregulation of splicing factors and breast cancer development. J Mol Cell Biol. 2015;7(5):388–401.
9. Lu Z, Huang Q, Park JW, et al. Transcriptome-wide Landscape of Pre-mRNA Alternative Splicing Associated with Metastatic Colonization. Mol Cancer Res. 2015;13(2):305–318.
10. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene. 2013;33(46):5311–5318.
11. Wang Y, Liu J, Huang BO, et al. Mechanism of alternative splicing and its regulation. Biomed reports. 2015;3(2):152–158.
12. Sibley CR, Blazquez L, Ule J. Lessons from non-canonical splicing. Nat Rev Genet. 2016;17(7):407–421.
13. Jacob AG, Smith CWJ. Intron retention as a component of regulated gene expression programs. Hum Genet. 2017;136(9):1043–1057.
14. Kovacs E, Tompa P, Liliom K, et al. Dual coding in alternative reading frames correlates with intrinsic protein disorder. Proc Natl Acad Sci USA. 2010;107(12):5429–5434.

15. Sveen A, Kilpinen S, Ruusulehto A, et al. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene. 2015;35(19):2413–2427.
16. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016;17(1):19–32.
17. Paronetto MP, Passacantilli I, Sette C. Alternative splicing and cell survival: from tissue homeostasis to disease. Cell Death Differ. 2016;23(12):1919–1929.
18. Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol. 2014;15(2):108–121.
19. Merkhofer EC, Hu P, Johnson TL. Introduction to Cotranscriptional RNA Splicing. Methods Mol Biol. 2014;1126:83–96.
20. Papasaikas P, Valcárcel J. The Spliceosome: the Ultimate RNA Chaperone and Sculptor. Trends Biochem Sci. 2016;41(1):33–45.
21. Kornblihtt AR, Schor IE, Alló M, et al. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol. 2013;14(3):153–165.
22. Kelemen O, Convertini P, Zhang Z, et al. Function of alternative splicing. Gene. 2013;514(1):1–30.
23. Wang Z, Liu D, Yin B, et al. High expression of PTBP1 promote invasion of colorectal cancer by alternative splicing of cortactin. Oncotarget. 2017;8(22):36185–36202.
24. Dvinge H, Kim E, Abdel-Wahab O, et al. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016;16 (7):413–430.
25. Iannone C, Valcárcel J. Chromatin's thread to alternative splicing regulation. Chromosoma. 2013;122(6):465–474.
26. Simon JM, Hacker KE, Singh D, et al. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. 2014;24(2):241–250.
27. Jung H, Lee D, Lee J, et al. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat Genet. 2015;47(11):1242–1248.
28. Marquez Y, Höpfler M, Ayatollahi Z, et al. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. Genome Res. 2015;25(7):995–1007.
29. Staiger D, Simpson GG. Enter exitrons. Genome Biol. 2015;16 (1):136.
30. Jiang WG, Sanders AJ, Katoh M, et al. Tissue invasion and metastasis: molecular, biological and clinical perspectives. Semin Cancer Biol. 2015;35:S244–75.
31. Buchheit CL, Weigel KJ, Schafer ZT. Cancer cell survival during detachment from the ECM: multiple barriers to tumour progression. Nat Rev Cancer. 2014;14(9):632–641.
32. Van Zijl F, Krupitza G, Mikulits W. Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res - Rev Mutat Res. 2011;728(1–2):23–34.
33. Chaffer CL, San Juan BP, Lim E, et al. EMT, cell plasticity and metastasis. Cancer Metastasis Rev. 2016;35(4):645–654.
34. Warzecha CC, Carstens RP. Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). Semin Cancer Biol. 2012;22(5–6):417–427.
35. Takano S, Reichert M, Bakir B, et al. Prrx1 isoform switching regulates pancreatic cancer invasion and metastatic colonization. Genes Dev. 2016;30(2):233–247.
36. Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298–306.
37. Labelle M, Begum S, Hynes RO. Platelets guide the formation of early metastatic niches. Proc Natl Acad Sci. 2014;111(30):E3053–61.
38. Senft D, Ronai ZA. Adaptive Stress Responses During Tumor Metastasis and Dormancy. Trends in Cancer. 2016;2(8):429–442.
39. Lorusso G, Rüegg C. New insights into the mechanisms of organ-specific breast cancer metastasis. Semin Cancer Biol. 2012;22 (3):226–233.
40. Abou Faycal C, Gazzeri S, Eymin B. RNA splicing, cell signaling, and response to therapies. Curr Opin Oncol. 2016;28(1):58–64.
41. Anczuków O, Krainer AR. Splicing-factor alterations in cancers. RNA. 2016;22(9):1285–1301.
42. Luz FAC, Brígido PC, Moraes AS, et al. Aberrant Splicing in Cancer: mediators of Malignant Progression through an Imperfect Splice Program Shift. Oncology. 2016.
43. Seidel B, Braeg S, Adler G, et al. E- and N-cadherin differ with respect to their associated p120ctn isoforms and their ability to suppress invasive growth in pancreatic cancer cells. Oncogene. 2004;23(32):5532–5542.
44. Sharma S, Liao W, Zhou X, et al. Exon 11 skipping of E-cadherin RNA downregulates its expression in head and neck cancer cells. Mol Cancer Ther. 2011;10(9):1751–1759.
45. Li X-W, Shi B-Y, Yang Q-L, et al. Epigenetic regulation of CDH1 exon 8 alternative splicing in gastric cancer. BMC Cancer. 2015;15(1):954.
46. Yang CC, He P, Liu Y, et al. Down-regulation of CEACAM1 in breast cancer. Acta Biochim Biophys Sin (Shanghai). 2015;47 (10):788–794.
47. Dery KJ, Gaur S, Gencheva M, et al. Mechanistic control of carcinoembryonic antigen-related cell adhesion molecule-1 (CEACAM1) splice isoforms by the heterogeneous nuclear ribonuclear proteins hnRNP L, hnRNP A1, and hnRNP M. J Biol Chem. 2011;286(18):16039–16051.
48. Dankner M, Gray-Owen SD, Huang YH, et al. CEACAM1 as a multi-purpose target for cancer immunotherapy. Oncoimmunology. 2017;6(7):1–16.
49. Ieda J, Yokoyama S, Tamura K, et al. Re-expression of CEACAM1 long cytoplasmic domain isoform is associated with invasion and migration of colorectal cancer. Int J Cancer. 2011;129(6):1351–1361.
50. Kirshner J, Chen C-J, Liu P, et al. CEACAM1-4S, a cell-cell adhesion molecule, mediates apoptosis and reverts mammary carcinoma cells to a normal morphogenic phenotype in a 3D culture. Proc Natl Acad Sci U S A. 2003;100(2):521–526.
51. Senbanjo LT, Chellaiah MA. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front cell Dev Biol. 2017;5:18.
52. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol. 2003;4(1):33–45.
53. Birzele F, Voss E, Nopora A, et al. CD44 Isoform status predicts response to treatment with Anti-CD44 antibody in cancer patients. Clin Cancer Res. 2015;21(12):2753–2762.
54. Brown RL, Reinke LM, Damerow MS, et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. J Clin Invest. 2011;121(3):1064–1074.
55. Zhao P, Xu Y, Wei Y, et al. The CD44s splice isoform is a central mediator for invadopodia activity. J Cell Sci. 2016;129(7):1355–1365.
56. Zhao S, Chen C, Chang K, et al. CD44 expression level and isoform contributes to pancreatic cancer cell plasticity, invasiveness, and response to therapy. Clin Cancer Res. 2016;22(22):5592–5604.
57. Prochazka L, Tesarik R, Turanek J. Regulation of alternative splicing of CD44 in cancer. Cell Signal. 2014;26(10):2234–2239.
58. Warzecha CC, Shen S, Xing Y, et al. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol. 2009;6(5):546.
59. Tripathi V, Sixt KM, Gao S, et al. Direct Regulation of Alternative Splicing by SMAD3 through PCBP1 Is Essential to the Tumor-Promoting Role of TGF-β. Mol Cell. 2016;64(5):1010.
60. Xu Y, Gao XD, Lee JH, et al. Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. Genes Dev. 2014;28(11):1191–1203.
61. Hu J, Li G, Zhang P, et al. A CD44v+ subpopulation subpopulation of breast cancer stem-like cells with enhanced lung metastasis capacity. Cell Death Dis. 2017;8(3):e2679.
62. Weber GF, Ashkar S, Glimcher MJ, et al. Receptor-Ligand Interaction Between CD44 and Osteopontin (Eta-1). Science (80-). 1996;271(5248):509–512.
63. Todaro M, Gaggianesi M, Catalano V, et al. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell. 2014;14(3):342–356.
64. Tremmel M, Matzke A, Albrecht I, et al. A CD44v6 peptide reveals a role of CD44 in VEGFR-2 signaling and angiogenesis. Blood. 2009;114(25):5236–5244.

65. Loh TJ, Moon H, Jang HN, et al. SR proteins regulate V 6 exon splicing of CD44 pre-mRNA. BMB Rep. 2016;49(11):612–616.
66. Hernandez JR, Kim JJ, Verdone JE, et al. Alternative CD44 splicing identifies epithelial prostate cancer cells from the mesenchymal counterparts. Med Oncol. 2015;32(5):159.
67. Sakuma K, Sasaki E, Kimura K, et al. HNRNPLL, a newly identified colorectal cancer metastasis suppressor, modulates alternative splicing of CD44 during epithelial-mesenchymal transition. Gut. 2018;67(6):1103–1111.
68. Balsamo M, Mondal C, Carmona G, et al. The alternatively-included 11a sequence modifies the effects of Mena on actin cytoskeletal organization and cell behavior. Sci Rep. 2016;6:35298.
69. Oudin MJ, Hughes SK, Rohani N, et al. Characterization of the expression of the pro-metastatic MenaINV isoform during breast tumor progression. Clin Exp Metastasis. 2016;33(3):249–261.
70. Melchionnaa R, Iapiccaa P, Di MF, et al. The pattern of hMENA isoforms is regulated by TGF-β1 in pancreatic cancer and may predict patient outcome. Oncoimmunology. 2016;5(12):e1221556.
71. Bonomi S, Gallo S, Catillo M, et al. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. Int J Cell Biol. 2013;2013:962038.
72. Trono P, Di Modugno F, Nisticò P. hMENA $^{11a}$, a hMENA isoform sending survival signals. Mol Cell Oncol. 2016;3(2):e1083648.
73. Roussos ET, Goswami S, Balsamo M, et al. Mena invasive (MenaINV) and Mena11a isoforms play distinct roles in breast cancer cell cohesion and association with TMEM. Clin Exp Metastasis. 2011;28(6):515–527.
74. Roussos ET, Condeelis JS, Patsialou A. Chemotaxis in cancer. Nat Rev Cancer. 2011;11(8):573–587.
75. Oktay MH, Jones JG. TMEM: A novel breast cancer dissemination marker for the assessment of metastatic risk. Biomark Med. 2015;9(2):81–84.
76. Pignatelli J, Goswami S, Jones JG, et al. Invasive breast carcinoma cells from patients exhibit MenaINV-and macrophage-dependent transendothelial migration. Sci Signal. 2014;7(353):ra112.
77. Weidmann MD, Surve CR, Eddy RJ, et al. MenaINV dysregulates cortactin phosphorylation to promote invadopodium maturation. Sci Rep. 2016;6(1):36142.
78. Fiegen D, Haeusler L-C, Blumenstein L, et al. Alternative Splicing of Rac1 Generates Rac1b, a Self-activating GTPase. J Biol Chem. 2004;279(6):4743–4749.
79. Singh A, Karnoub AE, Palmby TR, et al. Rac1b, a tumor associated, constitutively active Rac1 splice variant, promotes cellular transformation. Oncogene. 2004;23(58):9369–9380.
80. Faria M, Capinha L, Simões-Pereira J, et al. Extending the Impact of RAC1b Overexpression to Follicular Thyroid Carcinomas. Int J Endocrinol. 2016;2016:1–7.
81. Gonçalves V, Henriques A, Pereira J, et al. Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. RNA. 2014;20(4):474–482.
82. Pelisch F, Khauv D, Risso G, et al. Involvement of hnRNP A1 in the matrix metalloprotease-3-dependent regulation of Rac1 pre-mRNA splicing. J Cell Biochem. 2012;113(7):2319–2329.
83. Wang F, Fu X, Chen P, et al. SPSB1-mediated HnRNP A1 ubiquitylation regulates alternative splicing and cell migration in EGF signaling. Cell Res. 2017;27(4):540–558.
84. Faria M, Matos P, Pereira T, et al. RAC1b overexpression stimulates proliferation and NF-kB-mediated anti-apoptotic signaling in thyroid cancer cells. PLoS One. 2017;12(2):e0172689.
85. Weaver AM. Cortactin in tumor invasiveness. Cancer Lett. 2008;265(2):157–166.
86. Head JA, Jiang D, Li M, et al. Cortactin Tyrosine Phosphorylation Requires Rac1 Activity and Association with the Cortical Actin Cytoskeleton. Mol Biol Cel. 2003;14(8):3216–3229.
87. Katsube T, Togashi S, Hashimoto N, et al. Filamentous actin binding ability of cortactin isoforms is responsible for their cell-cell junctional localization in epithelial cells. Arch Biochem Biophys. 2004;427(1):79–90.
88. Gattazzo C, Martini V, Frezzato F, et al. Cortactin, another player in the Lyn signaling pathway, is over-expressed and alternatively spliced in leukemic cells from patients with B-cell chronic lymphocytic leukemia. Haematologica. 2014;99(6):1069–1077.
89. Johnson RM, Vu NT, Griffin BP, et al. The alternative splicing of cytoplasmic polyadenylation element binding protein 2 drives anoikis resistance and the metastasis of triple negative breast cancer. J Biol Chem. 2015;290(42):25717–25727.
90. Upheber S, Karle A, Miller J, et al. Alternative splicing of KAI1 abrogates its tumor-suppressive effects on integrin αvβ3-mediated ovarian cancer biology. Cell Signal. 2015;27(3):652–662.
91. Lee JH, Seo Y, Park SR, et al. Expression of a Splice Variant of KAI1, a Tumor Metastasis Suppressor Gene, Influences Tumor Invasion and Progression Expression of a Splice Variant of KAI1, a Tumor Metastasis Suppressor Gene, Influences Tumor Invasion and Progression. Group. 2003;63(21):7247–7255.
92. Miranti CK. Controlling cell surface dynamics and signaling: how CD82/KAI1 suppresses metastasis. Cell Signal. 2009;21(2):196–211.
93. Yang JM, O'Neill P, Jin W, et al. Extracellular matrix metalloproteinase inducer (CD147) confers resistance of breast cancer cells to anoikis through inhibition of Bim. J Biol Chem. 2006;281(14):9719–9727.
94. Liao C-G, Kong L-M, Song F, et al. Characterization of basigin isoforms and the inhibitory function of basigin-3 in human hepatocellular carcinoma proliferation and invasion. Mol Cell Biol. 2011;31(13):2591–2604.
95. Yao H-P, Zhou Y-Q, Zhang R, et al. MSP–RON signalling in cancer: pathogenesis and therapeutic potential. Nat Rev Cancer. 2013;13(7):466–481.
96. Moxley KM, Wang L, Welm AL, et al. Short-form Ron is a novel determinant of ovarian cancer initiation and progression. Genes Cancer. 2016;7(5–6):169–181.
97. Lu Y, Yao HP, Wang MH. Multiple variants of the RON receptor tyrosine kinase: biochemical properties, tumorigenic activities, and potential drug targets. Cancer Lett. 2007;257(2):157–164.
98. Mayer S, Hirschfeld M, Jaeger M, et al. RON alternative splicing regulation in primary ovarian cancer. Oncol Rep. 2015;34(1):423–430.
99. LeFave CV, Squatrito M, Vorlova S, et al. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. EMBO J. 2011;30(19):4084–4097.
100. Golan-Gerstl R, Cohen M, Shilo A, et al. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. Cancer Res. 2011;71(13):4464–4472.
101. Pagliarini V, Naro C, Sette C. Splicing Regulation: A Molecular Device to Enhance Cancer Cell Adaptation. Biomed Res Int. 2015;2015:543067.
102. Bharadwaj S, Thanawala R, Bon G, et al. Resensitization of breast cancer cells to anoikis by tropomyosin-1: role of Rho kinase-dependent cytoskeleton and adhesion. Oncogene. 2005;24(56):8291–8303.
103. Gunning P, Neill GO, Hardeman E. Tropomyosin-Based Regulation of the Actin Cytoskeleton in Time and Space. Physiol Rev. 2008;88(1):1–35.
104. Desouza M, Gunning PW, Stehn JR. The actin cytoskeleton as a sensor and mediator of apoptosis. Bioarchitecture. 2012;2(3):75–87.
105. Pan H, Gu L, Liu B, et al. Tropomyosin-1 acts as a potential tumor suppressor in human oral squamous cell carcinoma. PLoS One. 2017;12(2):e0168900.
106. Dube S, Yalamanchili S, Lachant J, et al. Expression of Tropomyosin 1 Gene Isoforms in Human Breast Cancer Cell Lines. Int J Breast Cancer. 2015;2015:859427.
107. Yang J, Zeng Z, Peng Y, et al. IL-7 splicing variant IL-7δ5 induces EMT and metastasis of human breast cancer cell lines MCF-7 and BT-20 through activation of PI3K/Akt pathway. Histochem Cell Biol. 2014;142(4):401–410.
108. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat Rev Immunol. 2017;17(9):559–572.
109. Wu Q, Dhir R, Wells A. Altered CXCR3 isoform expression regulates prostate cancer cell migration and invasion. Mol Cancer. 2012;11:3.

110. Li Y, Reader JC, Ma X, et al. Divergent roles of CXCR3 isoforms in promoting cancer stem-like cell survival and metastasis. Breast Cancer Res Treat. 2015;149(2):403–415.
111. Ding Q, Xia Y, Ding S, et al. An alternatively spliced variant of CXCR3 mediates the metastasis of CD133 + liver cancer cells induced by CXCL9. Oncotarget. 2016;7(12):14405–14414.
112. Allami RH, Graf C, Martchenko K, et al. Analysis of the expression of SDF-1 splicing variants in human colorectal cancer and normal mucosa tissues. Oncol Lett. 2016;11(3):1873–1878.
113. Woo IS, Hong SH, Byun JH, et al. Circulating stromal cell derived factor-1α (SDF-1α) is predictive of distant metastasis in gastric carcinoma. Cancer Invest. 2008;26(3):256–261.
114. Seguin L, Desgrosellier JS, Weis SM, et al. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. Trends Cell Biol. 2015;25(4):234–240.
115. Groulx J-F, Giroux V, Beauséjour M, et al. Integrin α6A splice variant regulates proliferation and the Wnt/β-catenin pathway in human colorectal cancer cells. Carcinogenesis. 2014;35(6):1217–1227.
116. Goel HL, Gritsko T, Pursell B, et al. Regulated splicing of the α6 integrin cytoplasmic domain determines the fate of breast cancer stem cells. CellReports. 2014;7(3):747–761.
117. Jiang P, Li Z, Tian F, et al. Fyn/heterogeneous nuclear ribonucleoprotein E1 signaling regulates pancreatic cancer metastasis by affecting the alternative splicing of integrin β1. Int J Oncol. 2017;51(1):169–183.
118. Yoon H, Dehart JP, Murphy JM, et al. Understanding the Roles of FAK in Cancer: inhibitors, Genetic Models, and New Insights. J Histochem Cytochem. 2015;63(2):114–128.
119. Yao L, Li K, Peng W, et al. An aberrant spliced transcript of focal adhesion kinase is exclusively expressed in human breast cancer. J Transl Med. 2014;12(1):136.
120. Despeaux M, Chicanne G, Rouer E, et al. Focal adhesion kinase splice variants maintain primitive acute myeloid leukemia cells through altered Wnt signaling. Stem Cells. 2012;30(8):1597–1610.
121. Hao C, Cui Y, Owen S, et al. Human osteopontin: potential clinical applications in cancer (Review). Int J Mol Med. 2017;39 (6):1327–1337.
122. Lin J, Myers AL, Wang Z, et al. Osteopontin (OPN/SPP1) isoforms collectively enhance tumor cell invasion and dissemination in esophageal adenocarcinoma. Oncotarget. 2015;6(26):22239–22257.
123. Shen H, Weber GF. The osteopontin-c splice junction is important for anchorage-independent growth. Mol Carcinog. 2014;53 (6):480–487.
124. Hao C, Cui Y, Hu M, et al. OPN-a Splicing Variant Expression in Non-small Cell Lung Cancer and its Effects on the Bone Metastatic Abilities of Lung Cancer Cells In Vitro. Anticancer Res. 2017;37(5):2245–2254.
125. Tang X, Li J, Yu B, et al. Osteopontin splice variants differentially exert clinicopathological features and biological functions in gastric cancer. Int J Biol Sci. 2013;9(1):55–66.
126. Bueno Ferreira L, Tavares C, Pestana A, et al. Osteopontin-a splice variant is overexpressed in papillary thyroid carcinoma and modulates invasive behavior. Oncotarget. 2016;7(32):52003–52016.
127. Sun J, Feng A, Chen S, et al. Osteopontin splice variants expressed by breast tumors regulate monocyte activation via MCP-1 and TGF-β1. Cell Mol Immunol. 2013;10(2):176–182.
128. White ES, Muro AF. Fibronectin splice variants: understanding their multiple roles in health and disease using engineered mouse models. IUBMB Life. 2011;63(7):538–546.
129. White E, Baralle F, Muro A. New insights into form and function of fibronectin splice variants. J Pathol. 2008;216(1):1–14.
130. Frey K, Fiechter M, Schwager K, et al. Different patterns of fibronectin and tenascin-C splice variants expression in primary and metastatic melanoma lesions. Exp Dermatol. 2011;20(8):685–688.
131. Ou JJ, Wu F, Liang HJ. Colorectal tumor derived fibronectin alternatively spliced EDA domain exserts lymphangiogenic effect on human lymphatic endothelial cells. Cancer Biol Ther. 2010;9 (3):186–191.
132. Bordeleau F, Califano JP, Abril YLN, et al. Tissue stiffness regulates serine/arginine-rich protein-mediated splicing of the extra domain B-fibronectin isoform in tumors. Proc Natl Acad Sci. 2015;112(27):8314–8319.
133. Liguori M, Solinas G, Germano G, et al. Tumor-associated macrophages as incessant builders and destroyers of the cancer stroma. Cancers (Basel). 2011;3(4):3740–3761.
134. Richter P, Junker K, Franz M, et al. IIICS de novo glycosylated fibronectin as a marker for invasiveness in urothelial carcinoma of the urinary bladder (UBC). J Cancer Res Clin Oncol. 2008;134 (10):1059–1065.
135. Solinas G, Schiarea S, Liguori M, et al. Tumor-Conditioned Macrophages Secrete Migration-Stimulating Factor: A New Marker for M2-Polarization, Influencing Tumor Cell Motility. J Immunol. 2010;185(1):642–652.
136. Perrier S, Woolston A-M, Purdie CA, Kazmi S, Preece PE, Davey KJ, Schor SL, Schor AM. Migration Stimulating Factor (MSF). A Novel Biomarker of Breast Cancer Progression. Transl Med. 2012; S1:003.
137. Olivos DJ, Mayo LD. Emerging Non-Canonical Functions and Regulation by p53: p53 and Stemness. Int J Mol Sci. 2016;17 (12):E1982.
138. Pflaum J, Schlosser S, Müller M. p53 Family and Cellular Stress Responses in Cancer. Front Oncol. 2014;4:285.
139. Gadea G, Arsic N, Fernandes K, et al. TP53 drives invasion through expression of its Δ133p53β variant. Elife. 2016;5:e14734.
140. Arsic N, Gadea G, Lagerqvist EL, et al. The p53 isoform Δ133p53β promotes cancer stem cell potential. Stem Cell Reports. 2015;4 (4):531–540.
141. Pixley FJ, Stanley ER. CSF-1 regulation of the wandering macrophage : complexity in action. Trends Cell Biol. 2004;14(11):628–638.
142. Stanley ER, Berg KL, Einstein DB, et al. Biology and Action of Colony – stimulating Factor-1. Mol Reprod Dev. 1997;46(1):4–10.
143. Ladner MB, Martin GA, Noble JA, et al. Human CSF-1: gene structure and alternative splicing of mRNA precursors. EMBO J. 1987;6(9):2693–2698.
144. Yang L, Zhang Y. Tumor-associated macrophages : from basic research to clinical application. J Hematol Oncol. 2017;10(1):58.
145. Liao J, Feng W, Wang R, et al. Diverse in vivo effects of soluble and membrane-bound M-CSF on tumor-associated macrophages in lymphoma xenograft model. Oncotarget. 2016;7(2):1354–1366.
146. Norris RA, Kern MJ. The Identification of Prx1 Transcription Regulatory Domains Provides a Mechanism for Unequal Compensation by the Prx1 and Prx2 Loci. J Biol Chem. 2001;276(29):26829–26837.
147. Zhang P, Zhang P, Zhou M, et al. 4 decletion variant of epidermal growth factor receptor enhances invasiveness and cisplatin resistance in epithelial ovarian cancer. Carcinogenesis. 2013;34 (11):2639–2646.
148. Greenall SA, Johns TG. EGFRvIII: the promiscuous mutation. Cell Death Discov. 2016;2:16049.
149. Guillaudeau A, Durand K, Bessette B, et al. Egfr soluble isoforms and their transcripts are expressed in meningiomas. PLoS One. 2012;7(5):e37204.
150. Wang H, Zhou M, Shi B, et al. Identification of an Exon 4-Deletion Variant of Epidermal Growth Factor Receptor with Increased Metastasis-Promoting Capacity. Neoplasia. 2011;13(5):461–471.
151. Wang H, Shi B, Zhang Q, et al. Growth and metastasis suppression of glioma xenografts expressing exon 4-deletion variant of epidermal growth factor receptor by monoclonal antibody CH12-mediated receptor degradation. FASEB J. 2012;26(1):73–80.
152. Manrique I, Nguewa P, Bleau AM, et al. The inhibitor of differentiation isoform Id1b, generated by alternative splicing, maintains cell quiescence and confers self-renewal and cancer stem cell-like properties. Cancer Lett. 2015;356(2 Pt B):899–909.
153. Giancotti FG. Mechanisms Governing Metastatic Dormancy and Reactivation. Cell. 2013;155(4):750–764.
154. Chen W, Dong J, Haiech J, et al. Cancer stem cell quiescence and plasticity as major challenges in cancer therapy. Stem Cells Int. 2016;2016:1740936.

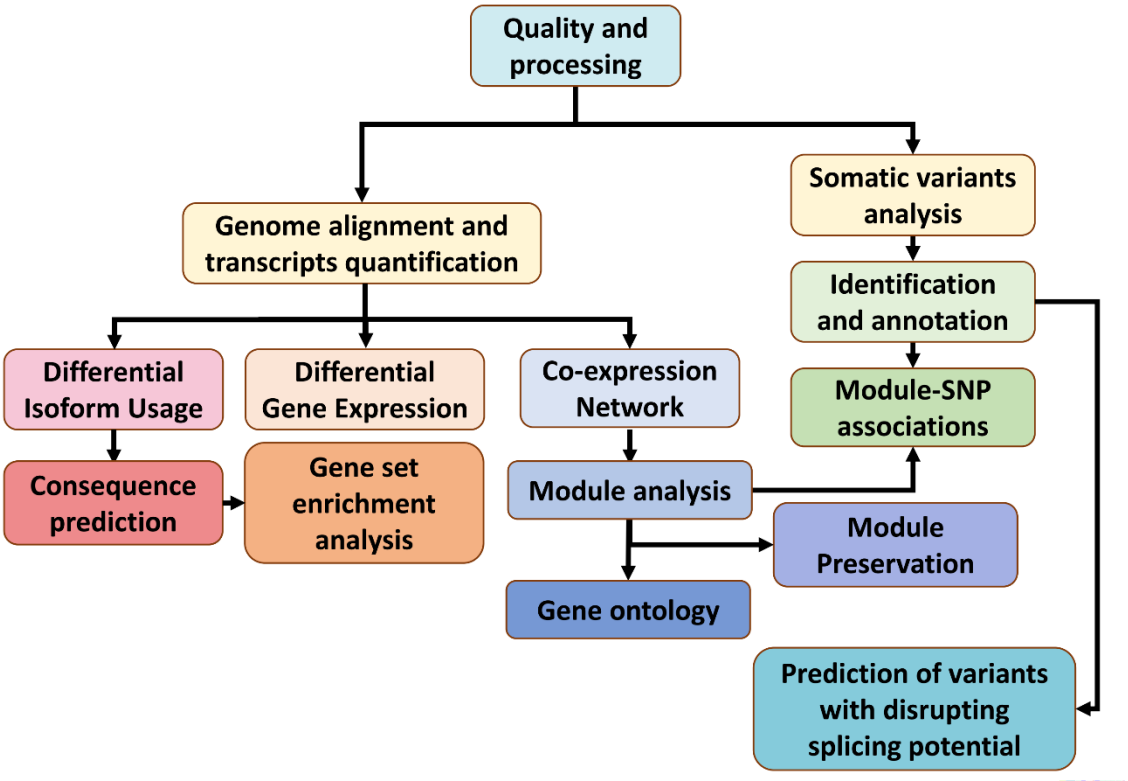
155. Hatakeyama K, Yamakawa Y, Fukuda Y, et al. A novel splice variant of XIAP-associated factor 1 (XAF1) is expressed in peripheral blood containing gastric cancer-derived circulating tumor cells. Gastric Cancer. 2015;18(4):751–761.
156. Li Z, Mou H, Wang T, et al. A non-secretory form of FAM3B promotes invasion and metastasis of human colon cancer cells by upregulating Slug expression. Cancer Lett. 2013;328(2):278–284.
157. Kong L, Yao L, Lu N, et al. Interaction of KLF6 and Sp1 regulates basigin-2 expression mediated proliferation, invasion and metastasis in hepatocellular carcinoma. Oncotarget. 2016;7(19):27975–27987.
158. Narla G, Difeo A, Yao S, et al. Targeted inhibition of the KLF6 splice variant, KLF6 SV1, suppresses prostate cancer cell growth and spread. Cancer Res. 2005;65(13):5761–5768.
159. Hatami R, Sieuwerts AM, Izadmehr S, et al. KLF6-SV1 Drives Breast Cancer Metastasis and Is Associated with Poor Survival. Sci Transl Med. 2013;5(169):169ra12.
160. DiFeo A, Martignetti JA, Narla G. The role of KLF6 and its splice variants in cancer therapy. Drug Resist Updat. 2009;12(1–2):1–7.
161. Lu H, Liu J, Liu S, et al. Exo70 isoform switching upon epithelial-mesenchymal transition mediates cancer cell invasion. Dev Cell. 2013;27(5):560–573.
162. Xiao L, Zheng K, Lv X, et al. Exo70 is an independent prognostic factor in colon cancer. Sci Rep. 2017;7(1):5039.
163. Luo Z, Cheng S, Shi J, et al. A splicing variant of Merlin promotes metastasis in hepatocellular carcinoma. Nat Commun. 2015;6:8457.
164. Scotlandi K, Zuntini M, Manara MC, et al. CD99 isoforms dictate opposite functions in tumour malignancy and metastases by activating or repressing c-Src kinase activity. Oncogene. 2007;26(46):6604–6618.
165. Byun HJ, Hong IK, Kim E, et al. A splice variant of CD99 increases motility and MMP-9 expression of human breast cancer cells through the AKT-, ERK-, and JNK-dependent AP-1 activation signaling pathways. J Biol Chem. 2006;281(46):34833–34847.
166. Da Silva MR, Moreira GA, Gonçalves Da Silva RA, et al. Splicing regulators and their roles in cancer biology and therapy. Biomed Res Int. 2015;2015:150514.
167. Zhang ZF, Pal S, Bi Y, et al. Isoform level expression profiles provide better cancer signatures than gene level expression profiles. Genome Med. 2013;5(4):33.
168. Le K-Q, Prabhakar BS, Hong W-J, et al. Alternative splicing as a biomarker and potential target for drug discovery. Acta Pharmacol Sin. 2015;36(10):1212–1218.
169. Ren P, Li M, Li T, et al. Anti-EGFRvIII Chimeric Antigen Receptor-Modified T Cells for Adoptive Cell Therapy of Glioblastoma. Curr Pharm Des. 2017;23(14):2113–2116.
170. Hong DS, Kurzrock R, Naing A, et al. A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. Invest New Drugs. 2014;32(3):436–444.
171. Koh CM, Bezzi M, Low DHP, et al. MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. Nature. 2015;523(7558):96–100.
172. Hsu TY, Simon LM, Neill NJ, et al. The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature. 2015;525(7569):384–388.
173. Jordaan G, Liao W, Sharma S. E-cadherin gene re-expression in chronic lymphocytic leukemia cells by HDAC inhibitors. BMC Cancer. 2013;13(1):88.
174. Yan Y, Zuo X, Wei D. Concise Review: emerging Role of CD44 in Cancer Stem Cells: A Promising Biomarker and Therapeutic Target. Stem Cells Transl Med. 2015;4(9):1033–1043.
175. Zöller M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? Nat Rev Cancer. 2011;11(4):254–267.
176. Misra S, Hascall VC, De Giovanni C, et al. Delivery of CD44 shRNA/nanoparticles within cancer cells. Perturbation of hyaluronan/CD44v6 interactions and reduction in adenoma growth in Apc Min/+mice. J Biol Chem. 2009;284(18):12432–12446.
177. Orian-Rousseau V, Ponta H. Perspectives of CD44 targeting therapies. Arch Toxicol. 2015;89(1):3–14.
178. Ying L, Li G, Wei -S-S, et al. Sanguinarine inhibits Rac1b-rendered cell survival enhancement by promoting apoptosis and blocking proliferation. Nat Publ Gr. 2014;36(10):229–240.
179. Matos P, Kotelevets L, Jordan P, et al. Ibuprofen Inhibits Colitis-Induced Overexpression of Tumor Related Rac1b. Neoplasia. 2013;15(1):102–111.
180. Bieniasz M, Radhakrishnan P, Faham N, et al. Pre-clinical efficacy of Ron kinase inhibitors alone and in combination with PI3K inhibitors for treatment of sfRon- expressing breast cancer patient-derived xenografts. Clin Cancer Res. 2015;21(24):5588–5600.
181. Femel J, Huijbers EJM, Saupe F, et al. Therapeutic vaccination against fibronectin ED-A attenuates progression of metastatic breast cancer. Oncotarget. 2014;5(23):12418–12427.
182. Han Z, Zhou Z, Shi X, et al. EDB fibronectin specific peptide for prostate cancer targeting. Bioconjug Chem. 2015;26(5):830–838.
183. Menrad A, Menssen HD. ED-B fibronectin as a target for antibody-based cancer treatments. Expert Opin Ther Targets. 2005;9(3):491–500.
184. Chen H, Chen L, Sun L, et al. A small interfering RNA targeting the KLF6 splice variant, KLF6-SV1, as gene therapy for gastric cancer. Gastric Cancer. 2011;14(4):339–352.

# CAPÍTULO II – ARTIGO DE DADOS

## IMPACT OF TRANSCRIPTS ISOFORMS SWITCHING DURING COLORECTAL CANCER AND IN THE HEPATIC METASTASIS

## PREFÁCIO

Este Capítulo segue uma metodologia de análise cujo pipeline está representado abaixo:



# Impact of transcripts isoforms switching during colorectal cancer and in the hepatic metastasis

Joice de Faria Poloni[1]*, Márcio Dorn[2] and Diego Bonatto[1]

[1] Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS – Brasil.

[2] Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.

* To whom correspondence should be addressed:
Joice de Faria Poloni
ORCID 0000-0002-3896-6404
Centro de Biotecnologia da UFRGS - Sala 107
Departamento de Biologia Molecular e Biotecnologia
Universidade Federal do Rio Grande do Sul - UFRGS
Avenida Bento Gonçalves 9500 - Prédio 43421
Postal Code 15005
Porto Alegre – Rio Grande do Sul
BRAZIL
91509-900
Phone: (+55 51) 3308-7765
Fax: (+55 51) 3308-7309

**Abstract**

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of death by cancer in the world. From all patients with CRC, approximately 50% will develop liver metastasis, which represents the major contributor to the mortality by CRC. During the primary tumor development, the accumulation of mutations and deregulation of gene expression may contribute to increasing metastatic potential of the malignant cells. Nonetheless, as alternative splicing emerges as one of the most important regulators of the gene expression, it is not surprising that the aberrant splicing is associated with tumor development and progression, with aberrant isoforms usage during malignant transformation. In this study, it was identified 888 isoforms switches between primary tumor and normal epithelium and 336 switches between metastasis and primary tumor. Furthermore, the main biological processes related to the isoforms switches were migration, cell adhesion, survival and proliferation, cell death, and immune response. These results support the importance of the understanding of the splicing mechanisms and the consequences associated with the isoforms switches between normal and malignant condition, suggesting potential targets for the discovery of biomarkers and the development of new therapies.

**Keywords**: colorectal cancer, hepatic metastasis, alternative splicing, transcriptional isoforms switching, isoforms usage, bioinformatics

**Abbreviations:** A3, alternative 3' splice site; A5, alternative 5' splice site; APAF1, apoptotic peptidase activating factor 1; APC, adenomatosis polyposis coli tumor suppressor; ATSS, alternative transcription start site; ATTS, alternative transcription termination site; BCAR3, breast cancer anti-estrogen resistance 3; BRI3BP, BRI3 binding protein; CRC, colorectal cancer; CTC, circulating tumor cells; CYP4F11, cytochrome P450 family 4 subfamily f member 11; DES, desmin; DUOX2, dual oxidase 2; DUOXA2, dual oxidase maturation factor 2; EGF, epidermal growth factor; EMT, epithelial-to-mesenchymal transition; ES, exon skipping; GATD1, glutamine amidotransferase like class 1 domain containing 1; GEO, gene expression omnibus; GS, gene significance; H6PD, hexose-6-phosphate dehydrogenase/glucose 1-dehydrogenase; IF, isoform fraction; IκB, inhibitor of nuclear factor kappa b kinase subunit beta; IL1β, interleukin 1 beta; IR, intron retention; kME, module membership; KRAS, KRAS proto-oncogene; KRT17, keratin 17; KRT80, keratin 80; LPAL2,

lipoprotein(a) like 2, pseudogene; MBNL1, muscleblind like splicing regulator 1; ME, module eigengenes; MEE, mutually exclusive exon; MES, multiple exon skipping; MET, metastasis; MIGA1, mitoguardin 1; MM, module membership; MRE11, MRE11 homolog, double strand break repair nuclease; MST1L, macrophage stimulating 1 like; MYO10, myosin x; NE, normal epithelium; NECTIN1, nectin cell adhesion molecule 1; NF-κB; nuclear factor kappa b subunit 1; NKD1, naked cuticle homolog1; NMD, nonsense-mediated decay; PCP, planar cell polarity; PDGF, platelet derived growth factor; PHAX, phosphorylated adaptor for rna export; PKD1, polycystin 1, transient receptor potential channel interacting; PT, primary tumor; PTC, premature stop-codon; PTPRC, protein tyrosine phosphatase, receptor type C; PTPRO, protein tyrosine phosphatase, receptor type O; RAB37, RAB37, member RAS oncogene family; RAD51, RAD51 recombinase; RBM39, RNA binding motif protein 39; ROS, reactive oxygen species; S100A2, S100 calcium binding protein a2; SERF2, small EDRK-rich factor 2; SNV, single nucleotide variants; TANGO2, transport and golgi organization 2 homolog; TGF-β, transforming growth factor beta 1; TLCD2, TLC domain containing 2; TNFAIP8, TNF alpha induced protein 8; TNFAIP8L1, TNF alpha induced protein 8 like 1; TNF-α, tumor necrosis factor; TOM, topological overlap matrix; TPM1, tropomyosin 1; TRAF1, TNF receptor associated factor 1; VEGF, vascular endothelial growth factor A; MYO10, myosin X; WNT5B, Wnt family member 5B; ZNF468, zinc finger protein 468.

## 1. Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death, and the third most prevalent type of cancer worldwide [1]. The reasons underlying CRC development and progress are considered heterogeneous, but genomic instability and impairments in DNA repair mechanisms are strongly associated with CRC malignancy [2,3]. However, other alterations occur during the progression of the primary tumor, such as the accumulation of mutations and gene expression deregulation, consequently triggering a metastatic potential [2,3]. Although metastasis is an inefficient process, some cells acquire the ability to survive and proliferate in extreme conditions, making treatment difficult and directly affecting patient survival. Numerous studies have already established the importance of alternative splicing in the initiation and maintenance of malignant cells, and various targets were already discussed as essential for successful tumorigenesis [4–6].

Currently, it is estimated that 95% of human genes have on average 6.3 alternatively spliced isoforms, where 3.9 are protein-coding [7]. Because of this variability, it is not surprising that RNA isoforms of the same gene are differentially expressed according to the biological context there are found, including diseases. Thus, disturbances in the regulation of alternative splicing during tumor progression are considered hallmarks of cancer by conferring selective tumoral advantages, such as angiogenesis, cell motility, proliferation and resistance to cell death [4,6,8]. In this sense, changes in the relative abundance of the transcript isoforms may contribute to an adaptive response during the establishment of the malignant lesion, directly impacting on the choice of the best therapeutic approach [4].

The goal of this work is to analyze isoforms switches in CRC, considering normal, primary tumor and liver metastasis to better understand the impact of alternative splicing and transcript isoforms in CRC progression together with a system biology approach. By subjecting isoforms expression profile of different conditions to a correlation network analysis, it is possible to predict regulators and malignancy drivers, since module structure may reflect on the regulatory mechanisms common to their components [9]. In addition, we conducted a bioinformatics analysis of the somatic mutational profile of CRC and its impact on isoforms switching across the studied conditions.

We report that several isoforms were mainly related to immune response, cellular adhesion, and migration. Our analyses show that transcript isoforms switches provide a novel

understanding of the biology of CRC progression, which could potentially lead to the new biomarker discovery and development of molecular targeted therapy.

## 2. Materials and Methods

### 2.1. RNA-seq data acquisition and processing

To perform the bioinformatics analysis, we collected an RNA-seq dataset at the NCBI Gene Expression Omnibus (GEO) repository under the accession number GSE50760. This dataset is composed of samples from matched normal epithelium (NE, collected > 5 cm from the tumor border), primary tumor (PT) and synchronous liver metastasis (MET). All patients, except for one, presented microsatellite stable status. The RNA extraction and sequencing protocols are described elsewhere [10]. The raw data was submitted to a quality analysis using FastQC application (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), followed by the trimming of low-quality bases (maximum of 10 bases per read), poly-N sequences and adaptor sequences using the Trimmomatic 0.35 software [11]. The resulting data were mapped against the reference genome of *Homo sapiens* (Ensembl version GRCh38.87) using the software STAR v2.5.2b in combination with RSEM v1.2.31 to achieve the transcript abundance quantification [12,13].

### 2.2. Identification of differential isoforms and gene expression analysis

The transcript quantification resulted from RSEM was used as input in the IsoformSwitchAnalyzeR R package to evaluate the isoforms switches between different conditions [14,15]. Data was filtered to remove non-expressed isoforms, lowly expressed genes, and genes with only one annotated isoform. As gene expression is represented by the multivariate expression of its isoforms, changes in the proportion of isoform expression from a corresponding gene may be related to significant phenotypic consequences [16]. In this sense, the differential isoform usage evaluates the relative expression of a specific isoform in relation to its corresponding gene expression across the compared conditions. Therefore, isoform usage is represented as a fraction of the total gene expression (IF). The switching of the isoforms expressed by the corresponding gene across the tested conditions is quantified by $dIF = meanIF_2$

– meanIF$_1$ [14,16]. Significant changes in the isoform usage was considered when the |dIF| > 0.1 and FDR < 0.05. To estimate differential gene expression, the tximport and DESeq2 R packages were employed [15,17]. Differential gene expression was determined by considering FDR < 0.05 and |log2FC| > 1.

### 2.3. Prediction of functional consequences related to the isoform switching

To evaluate the possible functional consequences related to isoforms switching, the premature stop-codon (PTC) using the 50 nucleotides rule was annotated, which establishes that to the recognition of a stop codon by nonsense-mediated decay (NMD) apparatus must be located at least 50 nucleotides upstream of the last exon-exon junction [18]. In order to perform the intron retention analysis, the retained introns were annotated by the pairwise comparison of the isoforms to its hypothetical pre-RNA, based on the exon information of the corresponding gene [19].

By using isoforms annotation, the nucleotide and coding sequence was obtained and exported to further analysis. To identify the protein domains related to each isoform with coding potential, the Pfam database 31.0 was employed [20,21]. To predict the potential isoforms with secretory capacity, the SignalP 4.1 tool was used [22]. Finally, the coding potential of the isoforms was analyzed by the Coding Potential Assessment Tool (CPAT) using the nucleotide sequences as input, which also increased the accuracy of ORF predictions [23]. Annotated ORF identification was used to predict the isoform sensitivity to NMD. The results were integrated into the isoform switches analysis by IsoformSwitchAnalyzeR.

### 2.4. Weighted isoforms co-expression network analysis of NE, PT, and MET conditions

The isoform-level quantification was normalized using the EBSeq R package, followed by filtering out isoforms whose corresponding gene are liver-specific using the TiGER database [24]. Additionally, the HTSFilter R package was used to remove low and constant levels of expression across the conditions [25]. Afterward, the results were ranked and the top 14000 varying isoforms were selected to proceed with the network analysis.

Co-expression analysis was conducted using the WGCNA R package and an individual network was created for each condition [26,27]. This analysis was carried out initially by the construction of a pairwise correlation matrix considering a soft-threshold parameter chosen based on the scale-free topology criterion ($r^2 > 0.8$), which was transformed to an adjacency matrix that represents the connection strengths between pairs of nodes [28,29]. Next, a topological overlap matrix (TOM) was calculated on the adjacency matrix for each isoform pair and a hierarchical clustering based on the dissimilarity TOM measure allowed the module detection by branches of the resulting clustering tree [30]. The following setting were chosen for the generation of individual networks: signed network type; β according to power estimate; deepSplit = 2; mergecutHeight = 0.25, minmodulesize = 30.

Module eigengenes (ME), which summarizes the first principal component of the module isoforms expression profile, was used for module-trait relationships analysis [27]. The hubs isoforms were selected according to their module membership (kME) value, defined by the correlation between module ME and the isoform expression.

Furthermore, individual networks were submitted to a module preservation analysis to assess whether the module properties from a test network are preserved when compared to a reference network [31]. The assessment of density and connectivity-based measures between the test and reference networks are carried out by permutation tests (n=500) and the results are summarized by *Zsummary* (*Z-score* values). Module preservation represented according to $Z$ statistics demonstrate that values ranging $2 < Z < 10$ have moderate preservation, $Z < 2$ have weak or no preservation and $Z > 10$ have high preservation.

### 2.5. Somatic variants and module-trait relationship

We hypothesized that somatic variants acquired during tumor progression may be related to the disturbance of the alternative splicing regulation, which may result in isoform switching. Therefore, aiming for a better understanding of molecular drivers and regulators of CRC progression, we performed a somatic variants analysis. For such purpose, we followed the best practices recommendations from GATK 3.7 software for RNA-seq analysis [32–34]. Firstly, a STAR 2-pass alignment method was performed [13,35]. Picard [http://broadinstitute.github.io/picard] and GATK tools were applied to data processing recalibrate base quality and optimize local realignment, while Mutect2 was employed to single

nucleotide variants (SNVs) and indels calling [36]. Variants annotation were performed by Variant Effect Predictor (VEP) and SNPNexus [37–39]. Variants were chosen based on the following criteria: variant allele frequency ≥ 0.2, nonsynonymous mutations, found in at least two different patients. Read alignments were reviewed using the Integrative Genomics Viewer version 2.3.98 and only variants affecting at least 10% of reads were taken into consideration for further analysis [40,41]. In order to verify if the somatic variants found may be related with abnormal splicing and isoforms switches, it was employed the software MutPred Splice [42].

To investigate whether network modules were associated with the selected variants, a binary matrix of somatic variants from each sample was created. The relationship of the interesting traits and network modules were estimated by calculating the Pearson's correlations of the binary matrix and ME module. To identify the intramodular importance of isoforms in relation to the somatic variants, the following measurements were selected: the "gene significance" (GS) by the correlation of isoforms expression profile and the traits, and the module membership (MM) by the correlation of the ME value and the isoforms expression profile. Only results with p-values < 0.05 were considered.

### 2.6. Gene ontology and gene-set enrichment analysis

The corresponding genes from isoforms switches with functional consequences were submitted to a permutation-based pre-ranked Gene Set Enrichment Analysis (GSEA) to evaluate the biological processes at the gene-set level [43]. To perform this analysis, the default settings were employed and the predefined GO gene sets of the Molecular Signature Database 6.2 were applied considering a nominal p-value < 0.05 [44].

Furthermore, an enrichment analysis of each module was conducted to identify the main biological processes in the co-expression networks by means of Fisher's exact test, using the topGO R package [45]. For the purpose of analyzing the enrichment at isoform-level, ontology terms of isoforms from the biomaRt 2.37.6 R package was applied as a customized annotation [46]. The analyses were made using the default settings and the p-value was adjusted by FDR < 0.05. Additional R packages and web server database were adopted to provide biological information concerning the results obtained, such as biomaRt 2.37.6, AnnotationDbi 1.42.1, GO.db 3.6.0, org.Hs.eg.db 3.6.0, GeneCards 3.0, APPRIS and UniProt [46–52].

## 3. Results and Discussion

### 3.1. Identification of changes in isoforms usage and splicing events across the different conditions

To evaluate the regulation of isoforms usage during the CRC and CRC liver metastasis, we analyzed matched samples from normal epithelium (NE), primary tumor (PT) and synchronous liver metastasis (MET), totalizing 54 samples. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) were employed to explore the distributions of the data before analysis and outliers samples were removed. The mapping of the sequenced fragments to the human reference genome resulted in > 85% of uniquely mapping reads for all samples.

To understand how alternative splicing and isoform usage could impact in CRC progression, we wondered which isoforms could be differentially regulated in PT relative to NE; MET relative to NE; and MET relative to PT. According to the results obtained with the isoform switch analysis, 1988 unique isoforms alternatively spliced from 1306 genes between the three comparisons were identified (**Figure 1A and B**). These isoforms were distributed across the comparisons, where we found 888 isoforms switches corresponding to 606 genes in PT relative to NE, and 1708 isoforms switches corresponding to 1125 genes were observed in MET relative to NE (**Figure 1A and B**). In addition, 336 isoforms switches corresponding to 220 genes were identified in MET when compared to PT (**Figure 1A and B**). Volcano plots (**Figure 1C**) illustrates the distribution of the significant under and overrepresented isoforms usage between the comparisons, as well as the relations of isoform usage with the gene expression in **Figure 1D**.

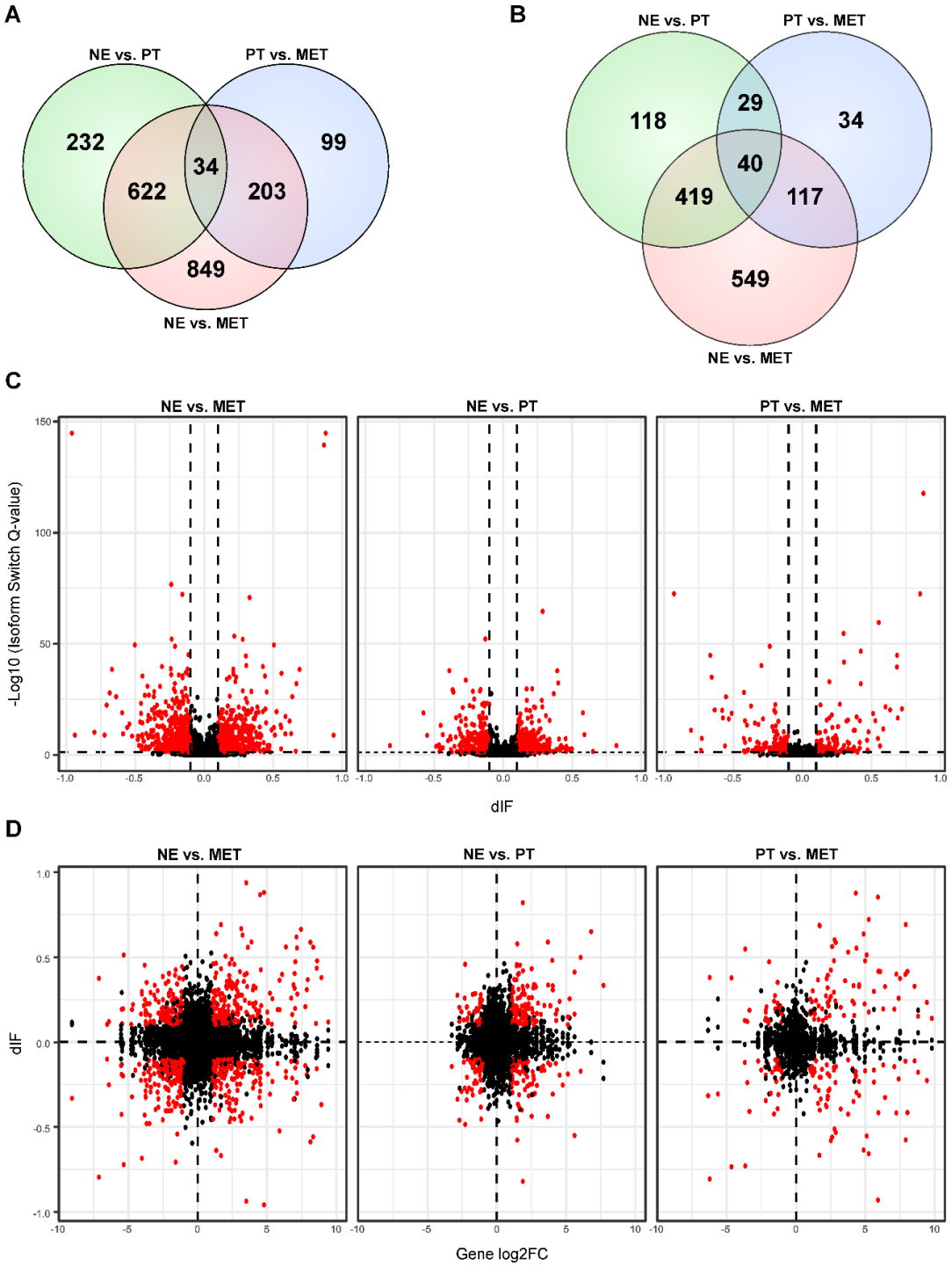



**Figure 1**. **A)** Overlap of isoform switches between the tested conditions. **B)** Overlap of the corresponding genes to the switched isoforms in tested conditions. **C)** Volcano plot of the differential isoforms usage. **D)** Volcano plot of the differential gene expression (x-axis) and its relation with the differential isoform usage (y-axis).

The differential usage of isoforms indicates that there is a distinct splicing regulation between normal and malignant conditions and even a difference regulation between PT and MET. Another study that used array-based technologies considering normal tissue, primary lesion and hepatic metastasis had similar results, where alternative splicing events are more abundant in the transition from normal colon mucosa to the primary tumor, and few changes were observed in the metastatic process [5]. These results may indicate that the core molecular mechanisms present in the primary tumor and metastases are similar, favoring the idea that few molecular changes are necessary to the acquisition of metastatic-initiating potential.

Subsequently, the next step was characterizing alternative splicing events to understand if there is a prevalent mechanism regulating isoforms usage during the malignant process. Thus, we evaluate the following events: mutually exclusive exon (MEE), alternative transcription termination site (ATTS), multiple exon skipping (MES), alternative 3' splice site (A3), exon skipping (ES), alternative transcription start site (ATSS), alternative 5' splice site (A5), and intron retention (IR). The results revealed that PT and MET have fewer isoforms which undergo to ATTS when compared to NE (**Figure 2**). This analysis considered as canonical the most downstream transcription termination site. Thus, there are a significantly small fraction of isoforms in the normal epithelium that are using exons in more upstream regions when compared to malignant tissues. In mammalian, almost 70% of genes have multiple polyadenylation sites, and it is suggested that both ATTS and ATSS are the main contributors to transcript diversity [53]. Changes in the termination site selection may generate modifications in polyadenylation and in the 3' UTR size, which were already linked to gastric cancer [54].

In addition, IR is enriched in MET when compared to NE (**Figure 2**). These results are in agreement with the outcome described in another study that observed an increase of intronic retention in several cancer types when compared to normal tissue [55]. New discoveries have consolidated the understanding that IR is not a mis-splicing consequence but actively regulated in mature mRNA [56]. The main functions related to IR involve the regulation of protein diversity and alternative splicing, mRNA localization, translation efficiency, and gene expression control [56,57]. It was also proposed that IR is a common cause of inactivation of tumor suppressor genes in cancer [56,57]. An isoform switch found in our analysis related to intron retention corresponds to the *NKD1* gene. NKD1 is a passive antagonist of Wnt signaling, reducing the aberrant activity of this pathway [58]. In CRC, the abnormal activity of the Wnt

pathway is one of the most important drivers and is frequently hyperactivated in more than 90% of CRC cases [58,59]. We identified a switch of NKD1-201 (ENST00000268459) to NKD1-203 (ENST00000566396) in both malignant conditions (PT and MET) relative to NE. NKD1-203 is a non-coding isoform with retained intron that may not exert the negative feedback of the Wnt pathway, consequently contributing to tumorigenesis.

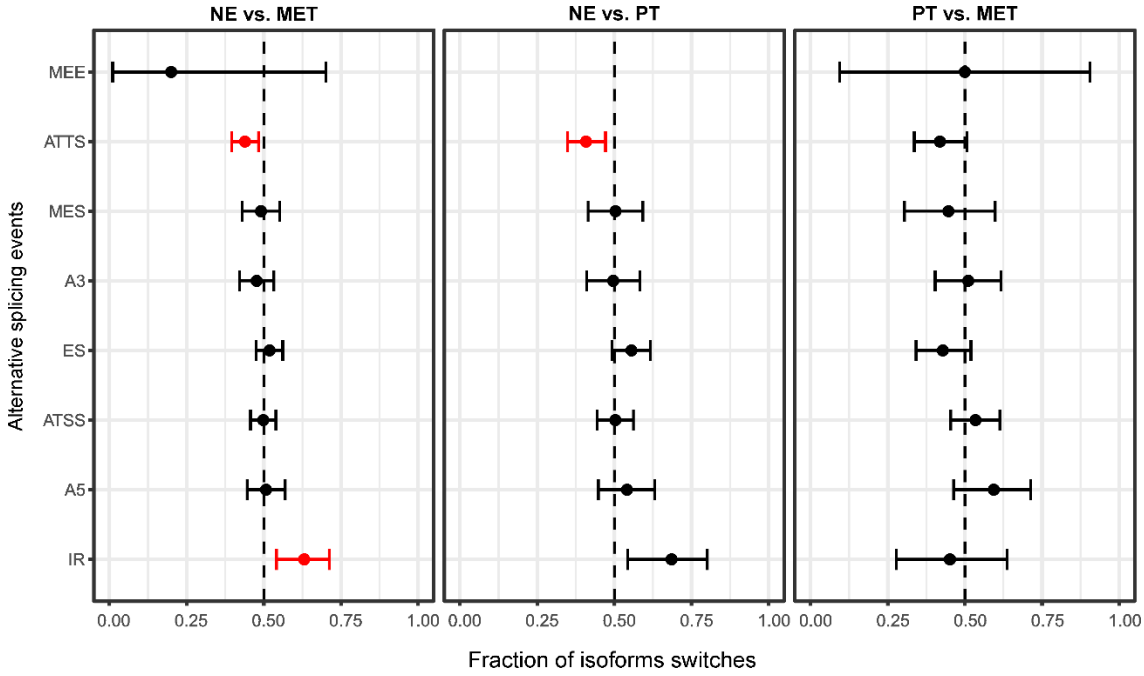



**Figure 2**. Schematic representation of alternative splicing event enrichment from isoform switches. The dashed line represents no enrichment, the right side indicates that the event is enriched in the tested condition (MET, PT, and MET, respectively), and left side indicates that the event is enriched in the reference condition (NE, NE, and PT, respectively). Red color represents significant results. MEE, mutually exclusive exon; ATTS, alternative transcription termination site; MES, multiple exon skipping; A3, alternative 3' splice site; ES, exon skipping; ATSS, alternative transcription start site; A5, alternative 5' splice site; IR, intron retention.

### 3.2. Predicted functional consequences resulting from isoforms switches and gene-set enrichment analysis of the corresponding genes

To analyze the possible overall impact related to the isoforms switches, we evaluated whether the isoform changes between conditions may be associated with potential functional consequences, such as the gain or a loss of protein domains, which in turns, affect the cellular

signaling. Hence, we analyzed if the isoform switch results in gain or loss of the following consequences: non-coding transcripts, signal peptide, intron retention, protein domain, and NMD sensitivity. From the 1988 switching isoforms identified across comparisons, 999 were associated with a gain or loss of some functional consequences, corresponding 548 different genes. The enrichment analysis of the consequences which showed significant differences between the conditions are represented in **Figure 3**.

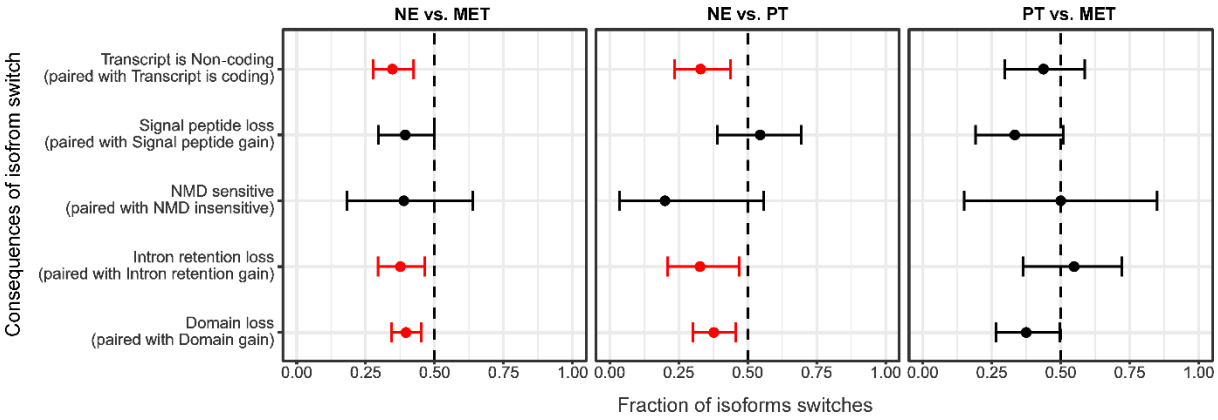



**Figure 3**. Enrichment/depletion of functional consequences in isoform switch according to the gain/loss of a consequence. The dashed line represents no enrichment, the right side indicates that the event is enriched in the tested condition (MET, PT, and MET, respectively) and the left side indicates that the event is enriched in the reference condition (NE, NE, and PT, respectively). Red color represents significant results.

The results showed that the normal tissue was more enriched with non-coding transcripts, intron retention loss, and domain loss when compared to both PT and MET. Otherwise, this result represents that the malignant process is marked by changes in isoforms expression, and these changes result in the modification of features that may have functional consequences, such as increased coding transcriptions, retained intron, and domain gain. Non-coding transcripts are described to possess critical roles in gene expression regulation and the maintenance of intestinal mucosa integrity [60]. The intestinal epithelium is in constant self-renewal, showing a high rate of apoptosis and proliferation to keep balanced cell numbers and fast repairs itself after acute injury [60]. In this sense, non-coding RNA contributes to the proper mucosa integrity through post-transcriptional regulation of the gene expression [60].

To obtain a global view of how isoforms with functional consequence changed across the comparisons, a hierarchical clustering was constructed considering the dIF values represented

by isoforms and log2FC of the corresponding genes (**Figure 4**). To aggregate information, we include known annotation about transcript biotype, which demonstrates that 81% of isoforms switches with functional consequences are protein-coding. The biological processes related to these isoforms and genes were manually curated, and the most represented processes were: (*i*) immune and inflammatory processes; (*ii*) cell adhesion; (*iii*) migration; and (*iv*) extracellular matrix organization. Carcinogenesis-related pathways showed in **Figure 4** includes isoforms involved in signaling pathways commonly altered during colorectal malignant transformation, such as Wnt, TNF-α, and TGF-β signaling pathway [59,61–63].

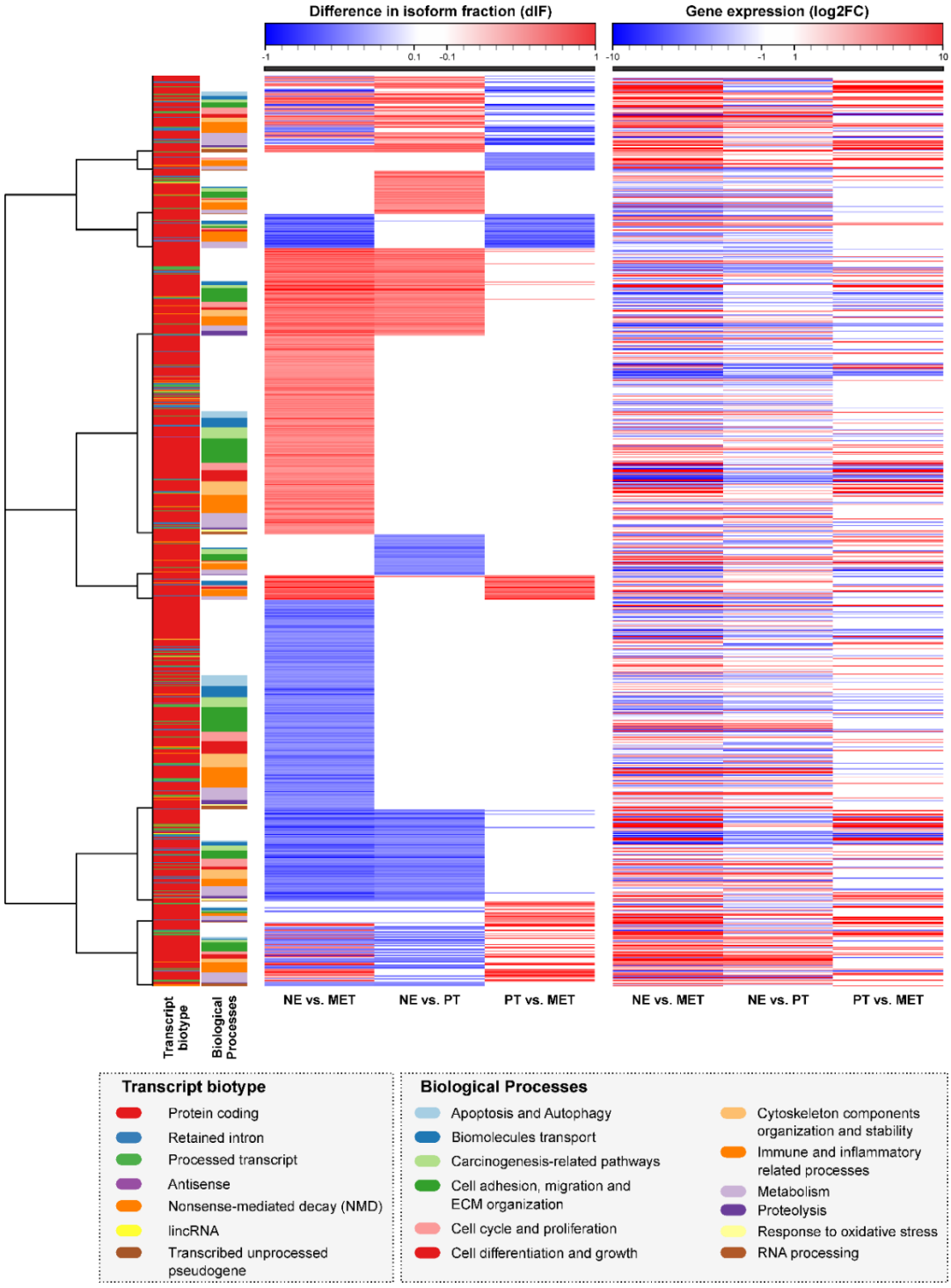



**Figure 4.** Heatmap illustrating differentially used isoforms (characterized by dIF) and differentially expressed genes (characterized by log2FC). Left bands represent transcript biotype and biological processes related to the isoform.

Moreover, we performed a GSEA considering all genes whose isoforms switches had a functional consequence (**Figure 5**). According to the results, angiogenesis and blood vessel

morphogenesis were enriched in PT relative to NE. Tumor progression is directly related to vascular support, once that the tumor growth depends on oxygen and nutrients supply [64]. In this sense, the releasing of angiogenesis and wound factors promotes the formation of disorganized and leaky new vessels, providing oxygen and nutrients, but also allowing the escape of tumor cells to the blood circulation, which can lead to metastasis [64,65].

Furthermore, considering MET relative to NE, it was observed the response to wounding bioprocess (**Figure 5**). Several mechanisms involved in wound healing overlaps with signaling pathways activated in malignancy, thus, the tumor is recognized as a non-healing wound [65]. During the wounding processes, platelets are activated, changing their shape, degranulating and secreting several factors that recruit others platelets and immune cells in response to injury with exposures of extracellular matrix proteins [65]. Interestingly, exocytosis, platelet degranulation, immune response activation, inflammatory response, and hemostasis, all involved in wound response, also were enriched in MET relative to NE (**Figure 5).** Platelet degranulation is a marker of platelet activation, where exocytosis of platelets secretory granules to the external milieu occurs [66]. These granules contain more than 300 active molecules, such as VEGF, PDGF, EGF, and TGF-β, which contribute to hemostasis, coagulation, wounding, immune processes, cancer progression, and cancer metastasis [65,66]. Platelet activation also has a permissive and protective role, contributing to tumor progression and metastasis through several mechanisms [65,66]. Additionally, circulating tumor cells (CTCs) enter in the blood circulation and recruit platelets inducing platelet aggregation, contributing to immune escape and increase adhesion to endothelial cells, facilitating vessels permeability and extravasation in a distant organ [4,65].

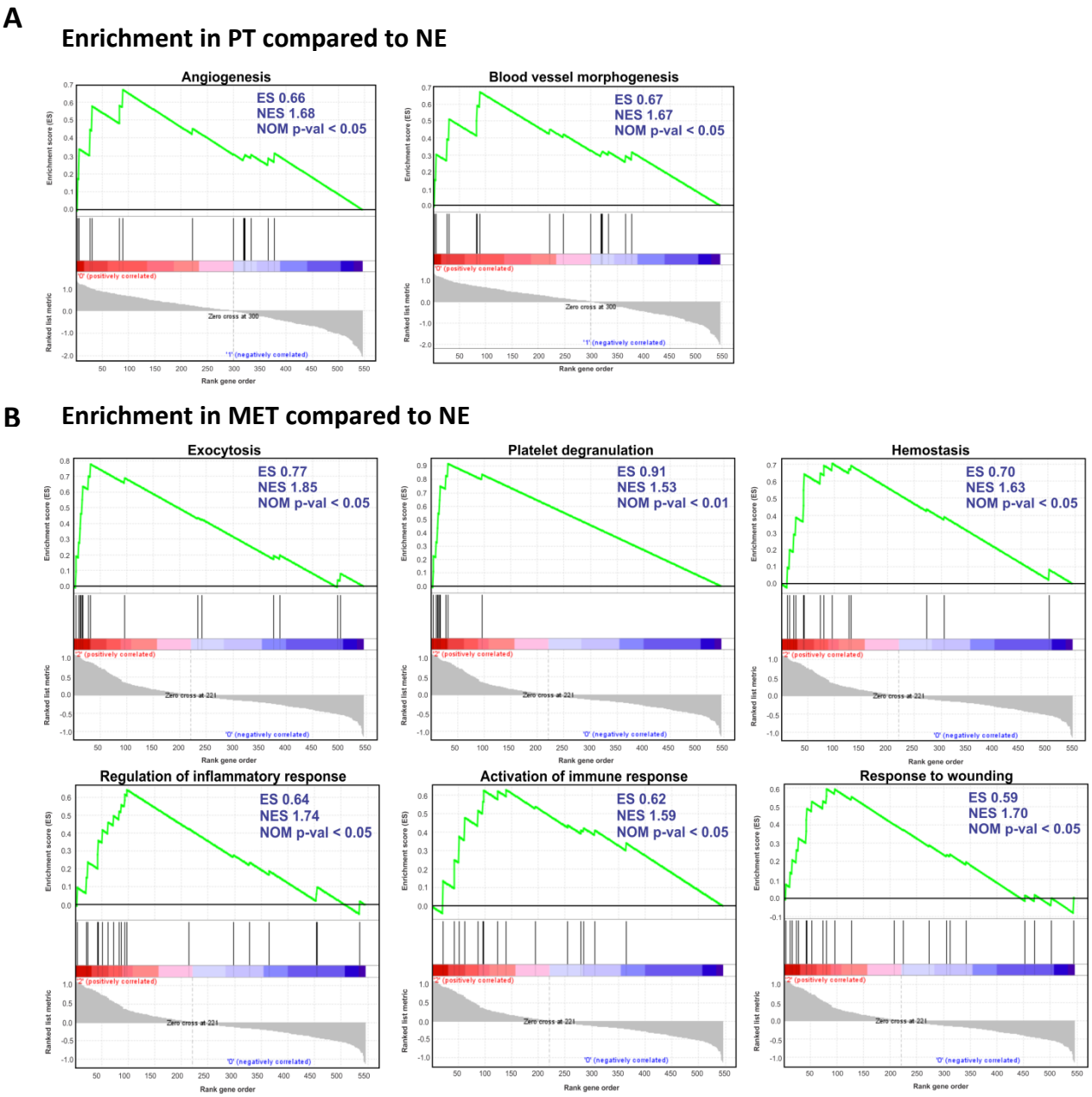



**Figure 5. A)** GSEA results from comparisons between PT relative to NE; **B)** GSEA results from comparisons between MET relative to NE.

Activation of immune response and inflammatory responses enriched in MET relative to PT and observed in **Figure 5**, are processes closely related to cancer progression and metastasis, where there is a direct association between the high cancer incidence and tissues with chronic inflammation. Potentially immunogenic tumor cells are also vulnerable to recognition and destruction by the immune system in several steps of metastatic cascade [67]. However, malignant cells with metastasis-initiating potential are able to avoid the immune surveillance and recruits immune cells populations that contribute to tumor cell spreading due to their

phenotypic plasticity, such as M2 macrophages [67,68]. Immune cells (mostly myeloid cells) and primary tumor-derived factors contribute to the formation of the pre-metastatic niche in a distant organ [67,68]. Nevertheless, CTC extravasation within the liver faces resistance by encountering a host innate immune response [69,70]. The interaction of tumor cells with the new environment triggers an inflammatory response which is damaging to a part of the tumor population, while to other population seem to favor their establishment and growth by enhancing cell adhesion and transendothelial migration into the Disse space [69,70].

### 3.3. Co-expression network denotes multiple modules preserved between PT and MET networks

To gain further insights at isoform level and differentially used isoforms across the conditions, a co-expression network for each condition was constructed, resulting in 13583 common isoforms between the three conditions after the filtering step. Hierarchical clustering identified 33, 37 and 24 modules for NE, PT and MET, respectively (**Figure 6A, B and C**).

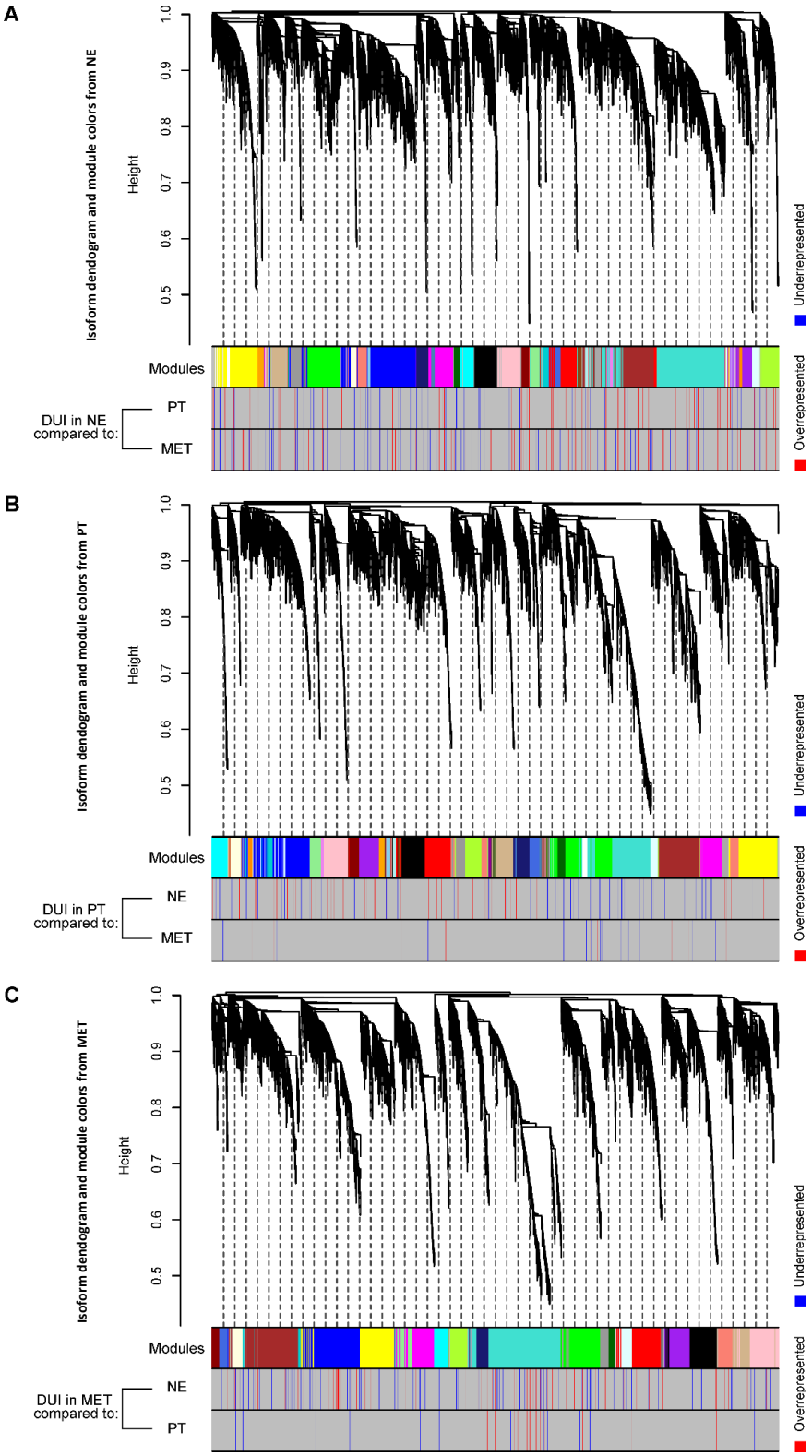



**Figure 6.** Dendrogram from the weighted gene co-expression analysis, denoting modules identified that are illustrated with different colors. A) NE network, B) PT network, and C) MET network. Bottom color bands represent under and overrepresented isoforms. DUI, differentially used isoforms.

To establish the reproducibility of these modules in each condition, a preservation statistical analysis was performed. In this sense, to investigate which PT modules are conserved in NE and MET, a series of permutation tests for each reference-test module pair were calculated (**Table 1**). The same evaluation was made to explore which MET modules are conserved in PT and NE. Considering the results from the preservation analyses, several modules that may represent significant biological processes specific to each condition tested (PT or MET) were identified. From 37 PT modules, 6 were weakly or moderately preserved (*Zsummary* < 5) relative to both MET and NE, indicating that these modules are more specific from PT condition (**Table 1**). Moreover, 11 modules were highly preserved (*Zsummary* > 10) among PT and MET and weakly or moderately preserved (*Zsummary* < 10) with NE, suggesting that these modules may be more related with the malignant processes (**Table 1**). When MET modules were tested relative to NE and PT modules, 6 modules showed a weak or moderate preservation (*Zsummary* <= 5) and may represent more specific processes from metastasis, while 8 modules were highly preserved (*Zsummary* > 10) only with PT modules (**Table 1**). In MET preservation analysis it was not found any module that showed high preservation exclusively with NE.

**Table 1.** Module preservation results. Red values indicate that the module was highly preserved between both malignant conditions (PT and MET), while is lowly preserved with NE. Blue values indicate that the module has low preservation between the comparisons.

| PT network preservation | | | | MET network preservation | | | |
|---|---|---|---|---|---|---|---|
| | | Z summary | | | | Z summary | |
| Modules | Size | Compared to MET | Compared to NE | Modules | Size | Compared to PT | Compared to NE |
| **Black** | 613 | 25.18 | 23.35 | **Black** | 760 | 19.20 | 5.75 |
| **Blue** | 1147 | 11.86 | 3.73 | **Blue** | 1271 | 25.91 | 8.58 |
| **Brown** | 1102 | 4.46 | 2.40 | **Brown** | 1267 | 20.06 | 10.44 |
| **Cyan** | 376 | 6.13 | 1.13 | **Cyan** | 346 | 4.58 | 1.98 |
| **Darkgreen** | 201 | 8.20 | 5.09 | **Darkgreen** | 159 | 1.82 | 0.04 |
| **Darkgrey** | 164 | 4.39 | 2.94 | **Darkred** | 182 | 4.52 | 0.34 |
| **Darkmagenta** | 65 | 3.07 | -0.98 | **Darkturquoise** | 102 | 10.82 | 4.05 |
| **Darkolivegreen** | 68 | 19.78 | 6.90 | **Green** | 862 | 12.07 | 6.72 |
| **Darkorange** | 120 | 11.66 | 3.45 | **Greenyellow** | 410 | 2.31 | 0.66 |
| **Darkred** | 236 | 12.72 | 8.61 | **Grey60** | 265 | 10.02 | 5.85 |
| **Darkturquoise** | 200 | 7.74 | 5.73 | **Lightcyan** | 287 | 10.32 | 4.21 |
| **Green** | 913 | 10.45 | 21.58 | **Lightgreen** | 260 | 10.51 | 5.37 |
| **Greenyellow** | 455 | 13.89 | 2.29 | **Lightyellow** | 256 | 12.70 | 2.97 |
| **Grey60** | 277 | 9.91 | 5.29 | **Magenta** | 671 | 5.00 | 1.26 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Lightcyan** | 322 | 8.78 | 12.14 | **Midnightblue** | 318 | 8.01 | 3.72 |
| **Lightgreen** | 272 | 10.36 | 1.18 | **Pink** | 715 | 5.91 | 5.38 |
| **Lightyellow** | 268 | 8.19 | -0.12 | **Purple** | 562 | 31.24 | 26.25 |
| **Magenta** | 500 | 30.46 | 7.91 | **Red** | 804 | 7.83 | 5.28 |
| **Midnightblue** | 340 | 11.67 | 17.52 | **Royalblue** | 193 | 2.88 | 0.23 |
| **Orange** | 140 | 7.45 | 3.77 | **Salmon** | 357 | 5.73 | 1.18 |
| **Paleturquoise** | 84 | 4.47 | 4.42 | **Tan** | 360 | 13.97 | 9.55 |
| **Pink** | 562 | 18.26 | 2.13 | **Turquoise** | 1887 | 6.27 | 0.37 |
| **Purple** | 479 | 15.49 | 14.78 | **Yellow** | 974 | 27.10 | 39.56 |
| **Saddlebrown** | 87 | 5.99 | 0.50 | | | | |
| **Salmon** | 397 | 12.07 | 3.66 | | | | |
| **Sienna3** | 48 | 6.01 | 0.62 | | | | |
| **Skyblue** | 110 | 7.96 | 1.79 | | | | |
| **Steelblue** | 86 | 5.37 | 2.74 | | | | |
| **Tan** | 454 | 9.06 | 0.83 | | | | |
| **Turquoise** | 1177 | 10.97 | 2.37 | | | | |
| **Violet** | 69 | 19.53 | 7.24 | | | | |
| **White** | 120 | 12.13 | 6.02 | | | | |
| **Yellow** | 964 | 24.86 | 14.52 | | | | |
| **Yellowgreen** | 43 | 8.58 | 2.13 | | | | |

### 3.4. GO enrichment analysis of the modules

To better understand the significant biological processes related to the modules found in each condition, we performed a GO enrichment analysis. The main GO results for NE, PT, and MET are summarized in **Table 2, Table 3 and Table 4**, respectively. The GO represented in this table were chosen based on statistical significance and the number of hubs isoforms associated with the biological processes.

**Table 2.** GO enrichment for NE network modules.

| Modules | GO ID | Term | Annot. | Signif. | FDR | Hubs |
|---|---|---|---|---|---|---|
| **Black** | GO:0006955 | immune response | 1178 | 78 | 0.007 | 43 |
| | GO:0007267 | cell-cell signaling | 597 | 30 | 0.015 | 11 |
| | GO:0050870 | positive regulation of T cell activation | 106 | 15 | 0.015 | 7 |
| | GO:0006935 | chemotaxis | 277 | 28 | 0.023 | 16 |
| | GO:0007049 | cell cycle | 867 | 56 | 0.024 | 31 |
| **Brown** | GO:0009988 | cell-cell recognition | 29 | 7 | 0.027 | 5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | GO:0008630 | intrinsic apoptotic signaling pathway in response to DNA damage | 65 | 9 | 0.032 | 7 |
| **Cyan** | GO:0019883 | antigen processing and presentation of endogenous antigen | 11 | 5 | 0.000 | 5 |
| | GO:0006952 | defense response | 831 | 45 | 0.000 | 25 |
| | GO:0006954 | inflammatory response | 367 | 21 | 0.004 | 13 |
| | GO:0016064 | immunoglobulin mediated immune response | 129 | 13 | 0.009 | 8 |
| **Darkgreen** | GO:0030449 | regulation of complement activation | 79 | 18 | 0.000 | 15 |
| | GO:0038096 | Fc-gamma receptor signaling pathway involved in phagocytosis | 109 | 18 | 0.000 | 15 |
| **Darkgrey** | GO:0060337 | type I interferon signaling pathway | 58 | 7 | 0.000 | 6 |
| | GO:0006955 | immune response | 1178 | 33 | 0.000 | 26 |
| | GO:0006952 | defense response | 831 | 25 | 0.000 | 21 |
| **Greenyellow** | GO:0055114 | oxidation-reduction process | 624 | 25 | 0.014 | 12 |
| **Ligthcyan** | GO:0006914 | autophagy | 279 | 6 | 0.018 | 2 |
| | GO:0090398 | cellular senescence | 22 | 3 | 0.018 | 0 |
| **Magenta** | GO:0007586 | digestion | 60 | 13 | 0.000 | 9 |
| | GO:0007154 | cell communication | 2741 | 119 | 0.030 | 61 |
| | GO:0097503 | sialylation | 15 | 8 | 0.000 | 8 |
| **Midnigthblue** | GO:0045199 | maintenance of epithelial cell apical/basal polarity | 5 | 2 | 0.017 | 0 |
| **Orange** | GO:0030154 | cell differentiation | 1504 | 14 | 0.048 | 7 |
| | GO:0035148 | tube formation | 52 | 3 | 0.048 | 3 |
| **Paleturquoise** | GO:0045087 | innate immune response | 493 | 11 | 0.013 | 9 |
| | GO:0007155 | cell adhesion | 736 | 16 | 0.045 | 13 |
| **Pink** | GO:0043312 | neutrophil degranulation | 366 | 38 | 0.002 | 18 |
| | GO:0008154 | actin polymerization or depolymerization | 140 | 14 | 0.025 | 8 |
| **Red** | GO:0006936 | muscle contraction | 185 | 28 | 0.005 | 17 |
| | GO:0007155 | cell adhesion | 736 | 61 | 0.005 | 27 |
| **Royalblue** | GO:0048167 | regulation of synaptic plasticity | 42 | 4 | 0.028 | 4 |
| **Steelblue** | GO:0006915 | apoptotic process | 987 | 12 | 0.049 | 7 |
| **Tan** | GO:0051301 | cell division | 288 | 46 | 0.000 | 41 |
| | GO:0006281 | DNA repair | 214 | 20 | 0.044 | 13 |
| **White** | GO:0030198 | extracellular matrix organization | 231 | 13 | 0.001 | 8 |
| | GO:0030212 | hyaluronan metabolic process | 21 | 4 | 0.013 | 2 |
| | GO:2000095 | regulation of Wnt signaling pathway, planar cell polarity pathway | 6 | 2 | 0.015 | 1 |
| | GO:2000647 | negative regulation of stem cell proliferation | 6 | 2 | 0.015 | 2 |
| **Yellow** | GO:0070098 | chemokine-mediated signaling pathway | 41 | 11 | 0.002 | 5 |
| | GO:0030198 | extracellular matrix organization | 231 | 44 | 0.005 | 25 |

According to the results observed in NE, we highlight digestion, muscle contraction, regulation of synaptic plasticity, cell division, DNA repair, extracellular matrix organization, and cell differentiation. All these processes are expected and crucial for the tissue homeostasis (**Table 2**). In addition, different processes related to immune system response were observed in several modules, such as Darkgreen, Cyan, Black, Darkgrey, Pink, and Paleturquoise. Among the overrepresented processes across these modules are defense response, immune response, complement activation, inflammatory response (**Table 2**). In the GO analysis of NE modules some processes that were already related with cancer development in literature, such as regulation of Wnt signaling pathway/planar cell polarity pathway (PCP), type I interferon signaling pathway, and hyaluronan metabolic process were detected (**Table 2**).

The intestinal epithelium is in constant exposition to alimentary antigens and potential pathogens in the luminal environment, therefore, immune cells surveillance against harmful antigens and tolerance to keep luminal bacteria is crucial for the individual health [71,72]. However, exacerbated immune response leads to a chronic inflammatory environment, which is thought to result in tumor initiation [73,74]. In this sense, the processes detected in NE network may indicate that an inflammatory condition is established. For example, hyaluronan (hyaluronan metabolic process found in White module, **Table 2**) is a component of the extracellular matrix that has been directly associated with intestinal inflammation and wound healing due this molecule be organized in structures that allow leucocyte adhesion, playing an important role in the initiation of intestinal inflammation [75]. Sialylation process (found in Magenta module, **Table 2**) act as an immunological modulator and affects directly the immune response, however, abnormal sialylation alters the immune recognition of cancer cells, contributing to the metastasis and therapeutic resistance [76–78]. Furthermore, type I interferon (found in Darkgrey module, **Table 2**) are immunomodulators cytokines that orchestrate the innate immune system in response to injuries and pathogens that may penetrate the intestinal epithelium [79].

The Wnt pathway/PCP (found in White module, **Table 2**) is a non-canonical Wnt pathway responsible to control the orientation of epithelial cells during development and the upregulation of components of this pathway has a role in cell proliferation, migration and cell death resistance [80].

In the PT GO enrichment analysis, it was observed that the modules conserved between PT and MET, according to the preservation analysis, were related to the following processes:

inflammatory response, adaptive immune response, leukocyte migration, chemotaxis, cell cycle regulation, blood vessel development, cell differentiation, and extracellular matrix organization (**Table 3**). All these processes are proposed to be crucial to the malignant transformation and are broadly observed in several types of cancer [81–84].

**Table 3.** GO enrichment for PT network modules.

| Modules | GO ID | Term | Annot. | Signif. | FDR | Hubs |
|---|---|---|---|---|---|---|
| **Blue** | GO:0006096 | glycolytic process | 70 | 26 | 0.004 | 21 |
| **Brown** | GO:0002682 | regulation of immune system process | 768 | 39 | 0.004 | 24 |
| | GO:0030162 | regulation of proteolysis | 473 | 25 | 0.007 | 14 |
| **Cyan** | GO:0006958 | complement activation, classical pathway | 105 | 21 | 0.000 | 15 |
| | GO:0050900 | leukocyte migration | 302 | 27 | 0.000 | 20 |
| **Darkgrey** | GO:0001525 | angiogenesis | 285 | 8 | 0.047 | 7 |
| | GO:0030198 | extracellular matrix organization | 231 | 7 | 0.047 | 6 |
| **Darkmagenta** | GO:0006958 | complement activation, classical pathway | 105 | 4 | 0.006 | 4 |
| | GO:0038096 | Fc-gamma receptor signaling pathway involved in phagocytosis | 109 | 4 | 0.006 | 3 |
| **Darkolivegreen** | GO:0045087 | innate immune response | 493 | 17 | 0.000 | 16 |
| | GO:0006954 | inflammatory response | 367 | 13 | 0.001 | 12 |
| **Darkorange** | GO:0033033 | negative regulation of myeloid cell apoptotic process | 6 | 2 | 0.027 | 1 |
| | GO:0007219 | Notch signaling pathway | 76 | 3 | 0.048 | 3 |
| **Grey60** | GO:0000398 | mRNA splicing, via spliceosome | 235 | 22 | 0.002 | 18 |
| | GO:0008283 | cell proliferation | 945 | 33 | 0.005 | 24 |
| **Lightcyan** | GO:0042407 | cristae formation | 29 | 9 | 0.000 | 6 |
| | GO:0006700 | C21-steroid hormone biosynthetic process | 12 | 4 | 0.008 | 4 |
| | GO:0006783 | heme biosynthetic process | 17 | 4 | 0.016 | 3 |
| | GO:0012501 | programmed cell death | 1028 | 32 | 0.037 | 22 |
| **Ligthgreen** | GO:0006958 | complement activation, classical pathway | 105 | 14 | 0.000 | 11 |
| | GO:0006955 | immune response | 1178 | 47 | 0.004 | 25 |
| | GO:0050900 | leukocyte migration | 302 | 19 | 0.005 | 12 |
| **Ligthyellow** | GO:0097190 | apoptotic signaling pathway | 335 | 12 | 0.038 | 7 |
| **Magenta** | GO:0030198 | extracellular matrix organization | 231 | 63 | 0.000 | 53 |
| | GO:0007155 | cell adhesion | 736 | 84 | 0.000 | 57 |
| | GO:0001568 | blood vessel development | 339 | 60 | 0.000 | 41 |
| **Pink** | GO:0002456 | T cell mediated immunity | 38 | 7 | 0.019 | 2 |
| | GO:0030855 | epithelial cell differentiation | 301 | 23 | 0.021 | 14 |
| | GO:0006935 | chemotaxis | 277 | 15 | 0.046 | 9 |
| | GO:0010564 | regulation of cell cycle process | 356 | 20 | 0.046 | 7 |
| **Purple** | GO:0006281 | DNA repair | 214 | 29 | 0.001 | 16 |
| **Royalblue** | GO:1903265 | positive regulation of TNF-mediated signaling pathway | 8 | 3 | 0.004 | 3 |
| | GO:0050718 | positive regulation of IL1β secretion | 11 | 3 | 0.004 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | GO:0043123 | positive regulation of I-κB kinase/NF-κB signaling | 126 | 7 | 0.006 | 5 |
| **Steelblue** | GO:0038158 | granulocyte colony-stimulating factor signaling pathway | 5 | 2 | 0.006 | 1 |
| | GO:0045454 | cell redox homeostasis | 55 | 4 | 0.006 | 4 |
| | GO:0016241 | regulation of macroautophagy | 84 | 3 | 0.047 | 2 |
| **Tan** | GO:0060337 | type I interferon signaling pathway | 58 | 13 | 0.000 | 10 |
| | GO:0042659 | regulation of cell fate specification | 6 | 3 | 0.002 | 1 |
| | GO:2000515 | negative regulation of CD4-positive, α-β T cell activation | 5 | 2 | 0.015 | 1 |
| **Turquoise** | GO:0030154 | cell differentiation | 1504 | 195 | 0.001 | 132 |
| | GO:0034329 | cell junction assembly | 152 | 29 | 0.009 | 18 |
| | GO:0070527 | platelet degranulation | 98 | 19 | 0.005 | 13 |
| **Violet** | GO:0002250 | adaptive immune response | 237 | 29 | 0.000 | 29 |
| | GO:0050900 | leukocyte migration | 302 | 17 | 0.000 | 17 |
| **Yellow** | GO:0007155 | cell adhesion | 736 | 54 | 0.041 | 26 |
| | GO:0043535 | regulation of blood vessel endothelial cell migration | 42 | 5 | 0.041 | 2 |
| | GO:0038061 | NIK/NF-κB signaling | 92 | 16 | 0.016 | 4 |
| | GO:0001708 | cell fate specification | 17 | 4 | 0.043 | 4 |

PT modules that were not conserved in NE or MET are represented by Fc-gamma receptor signaling pathway involved in phagocytosis, regulation of proteolysis, positive regulation of IL1β secretion, positive regulation of IκB kinase/NF-κB signaling, and positive regulation of TNF-mediated signaling pathway (**Table 3**). IL1β signaling (found in the Royalblue module, **Table 3**) mediates cancer inflammation and is related with the promotion of invasiveness and tumor growth in CRC, and to epithelial-to-mesenchymal transition (EMT) [85]. EMT is frequently observed in the promotion of metastasis in epithelium-derived carcinoma, and confers the loss of cell-cell adhesion, increase death resistance and migratory capacity [4].

Furthermore, the signaling pathways involved with the TNF and NF-κB pathways (found in Royalblue module, **Table 3**) are strongly related to the cancer growth and metastasis [86,87]. TNF-α is pro-inflammatory cytokine secreted by macrophages and part of the inflammation-related carcinogenesis process, playing roles in angiogenesis, proliferation, invasion, and metastasis [86,87]. NF-κB is activated by TNF-α and is the main cell survival promoter, eliciting anti-apoptotic signals, promoting cell proliferation and angiogenesis, which cooperates with cell invasion and metastasis promotion [86–88].

The highly preserved modules between PT and MET were enriched in cell differentiation, leukocyte migration, cell adhesion, extracellular matrix organization, Notch signaling pathway and inflammatory response processes (**Table 1 and 3**). Notch signaling pathway (found in Darkorange module, **Table 3**) is frequently activated in tumorigenesis and has been described

to contribute to metastasis [89]. In addition, Notch is related to the regulation of cell proliferation, cell fate specification, cell differentiation and stem cell maintenance [89].

In the MET network, it was observed that the modules preserved with PT condition were enriched in cellular response to interferon-gamma, inflammatory response, leukocyte chemotaxis, canonical Wnt signaling pathway, DNA repair, positive regulation of memory T cell differentiation, positive regulation of cell migration and platelet degranulation.

**Table 4.** GO enrichment for MET network modules.

| Modules | GO ID | Term | Annot. | Signif. | FDR | Hubs |
|---|---|---|---|---|---|---|
| **Black** | GO:0043382 | positive regulation of memory T cell differentiation | 5 | 3 | 0.009 | 2 |
| | GO:0044336 | canonical Wnt signaling pathway involved in negative regulation of apoptotic process | 5 | 3 | 0.009 | 1 |
| | GO:0044340 | canonical Wnt signaling pathway involved in regulation of cell proliferation | 5 | 3 | 0.009 | 1 |
| | GO:1901796 | regulation of signal transduction by p53 class mediator | 94 | 12 | 0.009 | 8 |
| | GO:0000077 | DNA damage checkpoint | 73 | 14 | 0.015 | 6 |
| **Blue** | GO:0030198 | extracellular matrix organization | 231 | 86 | 0.000 | 68 |
| | GO:0007155 | cell adhesion | 736 | 156 | 0.000 | 99 |
| | GO:0007010 | cytoskeleton organization | 620 | 121 | 0.000 | 70 |
| | GO:0001558 | regulation of cell growth | 190 | 38 | 0.010 | 30 |
| | GO:0001568 | blood vessel development | 339 | 97 | 0.014 | 66 |
| **Brown** | GO:0006325 | chromatin organization | 291 | 47 | 0.027 | 16 |
| | GO:0000086 | G2/M transition of mitotic cell cycle | 150 | 25 | 0.032 | 6 |
| | GO:0097503 | sialylation | 15 | 5 | 0.032 | 4 |
| **Cyan** | GO:0007267 | cell-cell signaling | 597 | 22 | 0.012 | 15 |
| **Darkgreen** | GO:0002223 | stimulatory C-type lectin receptor signaling pathway | 84 | 6 | 0.004 | 1 |
| | GO:0033209 | TNF-mediated signaling pathway | 118 | 7 | 0.004 | 3 |
| | GO:0042981 | regulation of apoptotic process | 782 | 14 | 0.012 | 6 |
| | GO:0090136 | epithelial cell-cell adhesion | 13 | 2 | 0.019 | 2 |
| **Green** | GO:0071346 | cellular response to interferon-gamma | 101 | 32 | 0.000 | 21 |
| | GO:0006954 | inflammatory response | 367 | 75 | 0.003 | 58 |
| | GO:0050852 | T cell receptor signaling pathway | 106 | 16 | 0.003 | 10 |
| | GO:0030595 | leukocyte chemotaxis | 113 | 21 | 0.005 | 15 |
| **Greenyellow** | GO:0002504 | antigen processing and presentation of peptide or polysaccharide antigen via MHC class II | 89 | 8 | 0.000 | 7 |
| | GO:0006955 | immune response | 1178 | 51 | 0.000 | 29 |
| | GO:0042127 | regulation of cell proliferation | 751 | 35 | 0.008 | 17 |
| **Grey60** | GO:0002224 | toll-like receptor signaling pathway | 74 | 7 | 0.001 | 5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | GO:0006897 | endocytosis | 445 | 13 | 0.030 | 8 |
| **Ligthcyan** | GO:0001525 | angiogenesis | 285 | 21 | 0.003 | 17 |
| | GO:0030335 | positive regulation of cell migration | 240 | 16 | 0.003 | 14 |
| | GO:0030198 | extracellular matrix organization | 231 | 24 | 0.006 | 21 |
| | GO:0002576 | platelet degranulation | 98 | 8 | 0.010 | 7 |
| **Ligthyellow** | GO:0006281 | DNA repair | 214 | 14 | 0.017 | 7 |
| **Pink** | GO:0001837 | epithelial to mesenchymal transition | 72 | 7 | 0.009 | 5 |
| | GO:0008380 | RNA splicing | 283 | 26 | 0.042 | 5 |
| | GO:0006281 | DNA repair | 214 | 19 | 0.042 | 7 |
| | GO:0007049 | cell cycle | 867 | 42 | 0.050 | 21 |
| **Purple** | GO:0006260 | DNA replication | 134 | 17 | 0.008 | 6 |
| | GO:0000398 | mRNA splicing, via spliceosome | 235 | 18 | 0.014 | 8 |
| | GO:0007155 | cell adhesion | 736 | 22 | 0.022 | 18 |
| | GO:0030593 | neutrophil chemotaxis | 51 | 4 | 0.028 | 3 |
| **Red** | GO:0022617 | extracellular matrix disassembly | 81 | 15 | 0.001 | 11 |
| | GO:0050777 | negative regulation of immune response | 49 | 14 | 0.001 | 10 |
| | GO:0050900 | leukocyte migration | 302 | 38 | 0.001 | 24 |
| | GO:0044319 | wound healing, spreading of cells | 32 | 8 | 0.003 | 5 |
| | GO:0034612 | response to TNF | 187 | 13 | 0.006 | 10 |
| **Royalblue** | GO:0007275 | multicellular organism development | 1910 | 25 | 0.008 | 17 |
| | GO:0030031 | cell projection assembly | 203 | 7 | 0.029 | 6 |
| | GO:0072593 | reactive oxygen species metabolic process | 132 | 4 | 0.037 | 2 |
| **Salmon** | GO:0051726 | regulation of cell cycle | 565 | 23 | 0.033 | 7 |
| | GO:0045785 | positive regulation of cell adhesion | 207 | 10 | 0.039 | 5 |
| **Tan** | GO:0006281 | DNA repair | 214 | 25 | 0.001 | 11 |
| | GO:0008283 | cell proliferation | 945 | 40 | 0.003 | 22 |
| | GO:0044770 | cell cycle phase transition | 312 | 30 | 0.018 | 17 |
| | GO:0006915 | apoptotic process | 987 | 38 | 0.044 | 20 |

Among the modules exclusive to MET, it was observed neutrophil chemotaxis, stimulatory C-type lectin receptor signaling pathway, regulation of apoptotic process, TNF-mediated signaling pathway, and epithelial cell-cell adhesion (**Table 4**). Neutrophils are described to cooperate with all steps of metastatic cascade, and their chemotaxis (found in Cyan module, **Table 4**) may occur in response to different primary tumor-derived factors, contributing to metastatic niche formation [90]. Neutrophils also collaborate to the tumoral cells extravasation by increasing their adherence to the vascular endothelium, such as liver sinusoidal endothelium (LSEC) [90]. The increasing adhesion of cancer cells to LSEC under inflammatory conditions contribute to the transendothelial migration of malignant cells to the space of Disse [70]. In this sense, C-type lectin receptors are members of the pattern recognition receptors that mediate cell

adhesion and are involved in the activation of inflammatory responses through the innate system [91,92]. These observations are in agreement to the processes stimulatory C-type lectin receptor signaling pathway and epithelial cell-cell adhesion (both found in Darkgreen module, **Table 4**), that may be related with the extravasation of the cancer cells and the metastatic outgrowth.

### 3.5. Module relevance to malignant transformation and its relation with somatic variants

To further understand the disturbances related to the isoform switch in the malignant transformation, somatic variants present in PT relative to NE and in MET compared to PT were identified. Among the genes that presented somatic variants, only *DES*, *MST1L* and *KRAS* showed isoform switch. However, according to Mutpred Splice results, the isoform switch of these genes was not directly related with the presence of somatic variant in the splice sites and splicing regulatory sequences (data not shown). Although it is known that splicing events may be influenced by genetic mutations, our work indicates that the isoforms switches are not directly affected by the presence of those mutations, corroborating with a previous study that found similar results [93]. Among the mentioned genes with isoform switch and somatic mutation is *KRAS*, one of the most frequent mutated oncogenes in CRC [94]. Interestingly, *KRAS* undergoes to isoform switch, where we observe that the isoform KRAS-201 (ENST00000256078) is upregulated in NE, while KRAS-202 is upregulated in both malignant conditions PT and MET (ENST00000311936). In our analysis, *KRAS* mutation showed a strong association with the Darkorange module, which exhibited enrichment of the Notch signaling pathway (**Table 3, Supplementary Figure 1**). Different studies have proposed the inhibition of Notch signaling as a therapeutic strategy in tumors driven by KRAS [95–97].

Associating the presence of the somatic variants from each sample with the modules found in the co-expression network can give us a clearer view of the processes that may be affected by the mutations. In addition, the intramodular analysis may suggest the most important isoforms that mediate the interplay between the presence of the mutation and the potentially affected processes. Thus, after the manual verification of the filtered somatic variants, only the variants that showed a strong significant correlation with the network modules from PT and MET conditions were selected. The results are present in **Supplementary Figure 1 and 2**.

The Ligthyellow module from PT network showed a high correlation with several mutations present identified in the genes *GATD1*, *H6PD*, *RBM39*, *TLCD2*, and *TNFAIP8L1* (**Table 3, Supplementary Figure 1**). This module is represented by the apoptotic process and the gene product from *TNFAIP8L1* is found in this module. TNFAIP8L1 is a regulator of apoptosis suggested being an important component of the necroptotic pathway [98]. In murine macrophage cell lines, TNFAIP8L1 promoted apoptosis by upregulating members of Bcl-2 family [99]. Interestingly, H6PD was also described to be related to the apoptosis pathway, where its knockdown increased apoptosis in triple-negative breast cancer [100]. Furthermore, RBM39 is related to alternative splicing and plays an important role in multiples processes, including G2/M transition, cellular response to DNA damage, adherent junctions, and endocytosis. [101]. RBM39 was already described in CRC, where it was suggested to have a "passenger" effect [102]. According to intramodular analysis, the most significant isoforms found in this module were RAB37-205 (ENST00000392614) that is upregulated in PT compared to NE, and the downregulated transcript TRAF1-201 (ENST00000373887) in PT relative to NE. RAB37 is involved in the exocytic pathway and in the suppression of cell invasion, macrophage polarization to M1 phenotype and Wnt inhibition [103–105]. However, its function sounds to be distinct in different types of cancers [104]. By its turns, TRAF1 was identified as a substrate of activated caspases in the TNF-induced apoptotic process, where one of the cleavage products is capable of preventing NF-κB induction [106]. TRAF1 activity is strongly related to the biological process found in this module, and it is possible that its activity is directly related to TNFAIP8L1. Remarkably, targeting TRAF1 with miR-483 was already observed to inhibit CRC cell proliferation and migration [107].

Another module highly correlated with several somatic variants was the Tan module from PT condition, which is involved with cell fate specification, type I interferon signaling and negative regulation of CD4+ (**Table 3**). The mutations related to this module were identified in *APAF1*, *BRI3BP*, *NECTIN1*, *PHAX*, *PKD1*, *RAD51*, and *ZNF468* (**Supplementary Figure 1**). PKD1 has multiple functions in several biological processes, including immune system regulation [108]. PKD1 also contributes to the activation of NF-κB, once that promote inhibitory kinase IKK phosphorylation that in turns results in the IkB degradation and releasing of NF-κB of the inhibitory complex [108]. The NF-κB signaling pathway was enriched in our analysis and has already been described as an important player in inflammation-associated carcinogenesis [88]. Furthermore, RAD51, which has an important role during DNA

replication, DNA repair, and recombination. Recently the downregulation of RAD51 was described to trigger innate immune response, linking the DNA damage response and replicative stress to the activation of the immune system [109]. Thus, the abnormal activity of variant RAD51 may be one of the many factors directly related with the strong immune response that we observed in our analysis. At least, APAF1 was described as an immunosuppressive by inhibiting the activation of T cells response [110]. The intramodular analysis of Tan module suggests that the most significant transcripts are the upregulated BCAR3-201 (ENST00000260502), KRT17-201 (ENST00000311208), and S100A2-202 (ENST00000368708), and the downregulated TNFAIP8-207 (ENST00000513374), and PTPRC-209 (ENST00000442510) transcripts. In gastric cancer, KRT17 was already related to tumor growth [111]. S100A2 was associated with a poor prognosis and to promote CRC cancer progression and metastasis through its activity in the cell cycle [112]. Also, S100A2 already was observed in CRC, being chemotactic for eosinophils and is associated with a poor outcome and recurrence with distant metastasis [112,113]. BCAR3 was suggested to be involved in metastasis in breast cancer by promoting cell motility and adhesion [114]. Furthermore, TNFAIP8 and PTPRC (also known as CD45) showed significant isoforms that are downregulated in PT relative to NE. PTPRC is abundantly expressed by lymphoid cells, where it acts in the regulation of external stimuli – thus, its abnormal expression may result in autoimmunity, immunodeficiency, and cancer [115]. TNFAIP8 also is broadly expressed in lymphoid tissues, and its upregulation promoted CD4+ T lymphocyte proliferation *in vitro* and affected CD4+ T lymphocyte polarization *in vivo* [98,116]. In this sense, the downregulation of TNFAIP8 in PT was consistent with the negative regulation of CD4+ ontology observed in this module. CD4+ T cells are important to the recognizing of the tumor antigens and activate macrophages involved in tumor clearance, therefore, the misregulation of CD4+ T cells affects the antitumor response triggered by the immune system [117]

In MET condition, the Black module was highly correlated with the mutation identified in *KRT80*, *MRE11*, *LPAL2*, and *TANGO2*, and was enriched in Wnt signaling pathway, DNA damage checkpoint, p53 signal transduction, and memory T cell differentiation (**Table 4, Supplementary Figure 2**). MRE11 has a role in DNA double-strand break repair, cell survival, proliferation, and tumor invasion, being described in breast cancer to generate a malignant phenotype and a poor response to therapeutics [118,119]. Another variant involved in tumor invasion is KRT80, whose transcript isoform KRT80-201 (ENST00000313234) is

overexpressed in MET. KRT80 was associated with tumor invasiveness and a poor cancer outcome and survival [120]. The Black module showed as the most significant isoforms resulted from intramodular analysis the upregulated MYO10-212 (ENST00000515803), WNT5B-201 (ENST00000310594), PTPRO-208 (ENST00000542557), and PTPRO-210 (ENST00000544244), and the downregulated transcript MBNL1-221 (ENST00000495875) Expression of PTPRO in pre-metastatic niche was described to prevent tumor growth and metastasis by increasing apoptosis and suppressing angiogenesis, preventing CTC entering the circulation [121]. However, in MET relative to NE there is a switch of the isoform PTPRO-203 to the isoforms PTPRO-208 and PTPRO-210; however, it is unclear whether this switch has a functional significance for metastasis. Also, related to suppressing cancer progression and metastasis was MBNL1, whose function involves RNA processing and expression regulation [122]. In contrast, WNT5B was related to cancer progression, where WNT5B-associated exosome increased migration and proliferation of malignant cells [123]. In our analysis, WNT5B undergoes an isoform switch, where WNT5B-201 is upregulated in MET while WNT5B-202 (ENST00000397196) is upregulated in NE, but none functional consequence was reported.

The Red module from MET network showed a strong correlation with the somatic variants presented in *APC*, *CYP4F11*, *DES*, *MIGA1*, and *SERF2* (**Supplementary Figure 2)**. This module is enriched in processes related to extracellular matrix disassembly, immune response and wound healing (**Table 4**). APC is involved in the regulation of β-catenin and is also frequently mutated in CRC [124]. Its role is related to cell adhesion, migration, control of cell growth and promotion of wound healing [125]. DES is an intermediate filament protein overexpressed in advanced stages of CRC and involved with tumor progression, cell adhesion and wound healing [126]. Our results showed the downregulated of protein-coding DES-201 (ENST00000373960) in MET (compared to both PT and NE), while the retained intron transcript DES-202 (ENST00000477226) is overexpressed in MET relative to NE. The intramodular analysis revealed as significant the overrepresented isoforms DUOXA2-202 (ENST00000350243), TPM1-206 (ENST00000358278), and S100A2-202 (ENST00000368708). The non-coding isoform DUOXA2-202 is upregulated in MET compared to NE, unlike DUOXA2-201. DUOXA2 is related to the formation of reactive oxygen species (ROS), contributing with DUOX2 to the conversion of $O_2$ in $H_2O_2$ as part of host defense in the gastrointestinal tract. However, excessive ROS generation promotes tissue

injury mediated by an inflammatory condition and has been related to chronic inflammation and inflammatory bowel disease [127]. In this sense, MET showed a non-coding isoform while NE expresses the main DUOXA2 isoform. This switch may suggest that MET population no longer trigger a defense response, but rather a response to evade the immune system and escape the cytotoxic effects, while NE condition produces a defense response that may be related with the inflammation found in pre-neoplastic conditions. S100A2-202 was also identified as significant isoform in the Tan module from the PT network, suggesting that this isoform has a prominent role in the CRC malignant condition and CRC liver metastasis. As mentioned before, S100A2 plays a role in the immune process by stimulating eosinophils chemotaxis, which may explain its importance in the Red module [112,113].

In our analysis, TPM1-206 was upregulated in MET relative to NE. TPM1 is considered a potential tumor-suppressor, inhibiting migration and promoting apoptosis [128]. Abnormal splicing of TPM1 is commonly observed in malignant transformations, including CRC [4,129]. The upregulated isoform TPM1-206, also called TPM1δ, is positively associated with stress fiber formation [130]. Stress fibers are necessary for adhesive structures and disrupted stress fibers are related to invasion and cell motility [128]. However, studies showed that in metastasis the stress fiber formation is associated with proliferative metastatic growth [131]. In breast cancer MCF-7 cells, overexpression of TPM1 suppresses anchorage-independent growth [132]. Besides heterogeneity found in metastatic populations, it is observed that several metastases express epithelial markers, indicating the reversion of mesenchymal to an epithelial phenotype through mesenchymal-to-epithelial transition [4] In this case, anchorage-dependent growth is reestablished and TPM1 may not exert a tumor-suppressor role. Thus, it is possible that the expression of TPM1-206 in MET condition was related to stress fiber formation, influencing adhesiveness and proliferation of metastatic cells.

Overall, these analyses indicate some isoform that may be influenced or contribute to mutational variants, affecting biological processes strongly related to malignant transformation.

## 4. Conclusion

In conclusion, this study illustrates the important role that alternative splicing has in the regulation of CRC and CRC liver metastasis. We show that several isoforms have differential usage in malignant processes compared to normal epithelium, as well as different functional

consequences and alternative splicing events were altered during the isoform switch in CRC. These results reinforce the importance of further studies exploring splicing mechanisms and its associated consequences. Our results also support the importance of the use of isoform expression in the regular analysis to obtain refined results, which will be more directly associated with the expression profile related to CRC. Large-scale studies have allowed us to integrate several levels of information, providing alternatives to fill the gaps in current knowledge and contribute to the discovery and development of new therapeutic targets.

## 5. Acknowledgments

This research was developed with the computational support of the Centro Nacional de Supercomputação (CESUP) and Laboratório de Bioinformática Estrutural e Biologia Computacional (SBCB), both from Universidade Federal do Rio Grande do Sul (UFRGS). We thank Dr Matthew Mort for kindly running MutPred Splice analysis. The fomenting agency CAPES and CNPq provided the financial support to this work under Grant 302969/2016–0.

## 6. Disclosure statement

No potential conflict of interest was reported by the authors

## 7. References

1. Granados-Romero JJ, Valderrama-Treviño AI, Contreras-Flores EH, Barrera-Mera B, Enríquez MH, Uriarte-Ruíz K, Ceballos-Villalba JC, Estrada-Mata AG, Rodríguez Alvarado C, Arauz-Peña G. Colorectal cancer : a review. Int J Res Med Sci. 2017; 5(11):4667–76.

2. Bogaert J, Prenen H. Molecular genetics of colorectal cancer. Ann Gastroenterol. 2014; 27(1):9–14.

3. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, Delrio P. Genetics, diagnosis and management of colorectal cancer (Review). Oncol Rep. 2015; 34(3):1087–96.

4. De Faria Poloni J, Bonatto D. Influence of transcriptional variants on metastasis. RNA Biol. 2018; 15(8):1006–24.

5. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, Zanovello P, Bortoluzzi S, Mandruzzato S. An integrative framework identifies alternative splicing events in colorectal cancer development. Mol Oncol. 2014; 8(1):129–41.

6. Yi Q, Tang L. Alternative spliced variants as biomarkers of colorectal cancer. Curr Drug Metab. 2011; 12(10):966–74.

7. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74.

8. Pagliarini V, Naro C, Sette C. Splicing Regulation: A Molecular Device to Enhance Cancer Cell Adaptation. Biomed Res Int. 2015; 2015:543067.

9. Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: The potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. Genes, Brain Behav. 2014; 13(1):13–24.

10. Kima S-K, Kima S-Y, Kima J-H, Roh SA, Cho D-H, Kim YS, Kim JC. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. Mol Oncol. 2014; 8(8):1653–66.

11. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20.

12. Li B, Dewey CN. RSEM : accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323.

13. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21.

14. Vitting-Seerup K, Sandelin A. The Landscape of Isoform Switches in Human Cancers. Mol Cancer Res. 2017; 15(9):1206–20.

15. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research. 2015; 4(2):1521.

16. Nowicka M, Robinson MD. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Research. 2016; 5:1356.

17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15(12):550.

18. Weischenfeldt J, Waage J, Tian G, Zhao J, Damgaard I, Jakobsen JS, Kristiansen K, Krogh A, Wang J, Porse BT. Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. Genome Biol. 2012; 13:R35.

19. Vitting-Seerup K, Porse BT, Sandelin A, Waage J. spliceR: An R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. BMC Bioinformatics. 2014; 15:81.

20. Finn RD, Bateman A, Clements J, et al. Pfam: The protein families database. Nucleic Acids Res. 2014; 42:D222–30.

21. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 2016; 44(D1):D279–85.

22. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Methods. 2011; 8(10):785–6.

23. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013; 41(6):e74.

24. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: A database for tissue-specific gene expression and regulation. BMC Bioinformatics. 2008; 9:271.

25. Rau A, Gallopin M, Celeux G, Jaffrézic F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics. 2013; 29(17):2146–52.

26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.

27. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol. 2007; 1:54.

28. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol. 2005; 4(1):Article 17.

29. Chalovich JM, Eisenberg E. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw. 2012; 46(11):pii: i11.

30. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. Bioinformatics. 2008; 24(5):719–20.

31. Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible ? PLoS Comput Biol. 2011; 7(1):e1001057.

32. DePristo M a., Banks E, Poplin RE, et al. A framework for variation discovery and genotyping using next- generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8.

33. Auwera GA Van Der, Carneiro MO, Hartl C, et al. From FastQ data to high confidence varant calls: the Genome Analysis Toolkit best practices pipeline. 2013.

34. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303.

35. Engström PG, Steijger T, Sipos B, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods. 2013; 10(12):1185–91.

36. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31(3):213–9.

37. Dayem Ullah AZ, Lemoine NR, Chelala C. A practical guide for the functional annotation of genetic variations using SNPnexus. Brief Bioinform. 2013; 14(4):437–47.

38. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17(1):1–14.

39. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: Assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res. 2018; 46(W1):W109–13.

40. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14(2):178–92.

41. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative Genome Viewer. Nat Biotechnol. 2011; 29(1):24–6.

42. Mort M, Sterne-Weiler T, Li B, Ball E V., Cooper DN, Radivojac P, Sanford JR, Mooney SD. MutPred Splice: Machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol. 2014; 15(1):1–20.

43. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005; 102(43):15545–50.

44. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27(12):1739–40.

45. Alexa A. Gene set enrichment analysis with topGO. 2007;

46. Durinck S, Spellman PT, Birney E, Huber W. Mapping identfiers for integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4(8):1184–91.

47. Safran M, Dalah I, Alexander J, et al. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010; 2010:baq020.

48. Apweiler R, Bairoch A, Wu CH, et al. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 2017; 45(D1):D158–69.

49. Pagès H, Carlson M, Falcon S, Li N. AnnotationDbi: Annotation Database Interface. R Packag version 1421. 2018;

50. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. R Packag version 360. 2018;

51. Carlson M. GO.db: A set of annotation maps describing the entire Gene Ontology. R Packag version 360. 2018;

52. Rodriguez JM, Carro A, Valencia A, Tress ML. APPRIS WebServer and WebServices. Nucleic Acids Res. 2015; 43(W1):W455–9.

53. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. 2018; 46(2):582–92.

54. Lai DP, Tan S, Kang YN, et al. Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. Hum Mol Genet. 2015; 24(12):3410–7.

55. Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 2015; 7(1):45.

56. Wong JJL, Au AYM, Ritchie W, Rasko JEJ. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. BioEssays. 2016; 38(1):41–9.

57. Jacob AG, Smith CWJ. Intron retention as a component of regulated gene expression programs. Hum Genet. 2017; 136:1043–57.

58. Stancikova J, Krausova M, Kolar M, et al. NKD1 marks intestinal and liver tumors linked to aberrant wnt signaling. Cell Signal. 2015; 27(2):245–56.

59. Schatoff EM, Leach BI, Dow LE. WNT Signaling and Colorectal Cancer. Curr Colorectal Cancer Rep. 2017; 13(2):101–10.

60. Wang J-Y, Xiao L, Wang J-Y. Posttranscriptional regulation of intestinal epithelial integrity by noncoding RNAs. Wiley Interdiscip Rev RNA. 2017; 8(2):1–15.

61. Jung B, Staudacher JJ, Beauchamp D. Transforming Growth Factor β Superfamily Signaling in Development of Colorectal Cancer. Gastroenterology. 2017; 152(1):36–52.

62. Al Obeed OA, Alkhayal KA, Al Sheikh A, Zubaidi AM, Vaali-Mohammed MA, Boushey R, Mckerrow JH, Abdulla MH. Increased expression of tumor necrosis factor-α is associated with advanced colorectal cancer stages. World J Gastroenterol. 2014; 20(48):18390–6.

63. Stanilov N, Miteva L, Dobreva Z, Stanilova S. Colorectal cancer severity and survival in correlation with tumour necrosis factor-alpha. Biotechnol Biotechnol Equip. 2014; 28(5):911–7.

64. Nishida N, Yano H, Nishida T, Kamura T, Kojiro M. Angiogenesis in cancer. Vasc Health Risk Manag. 2006; 2(3):213–9.

65. Menter DG, Kopetz S, Hawk E, Sood AK, Loree JM, Gresele P, Honn K V. Platelet "first responders" in wound response, cancer, and metastasis. Cancer Metastasis Rev. 2017; 36(2):199–213.

66. Golebiewska EM, Poole AW. Platelet secretion: From haemostasis to wound healing and beyond. Blood Rev. 2015; 29(3):153–62.

67. Kitamura T, Qian B-Z, Pollard JW. Immune cell promotion of metastasis. Nat Rev Immunol. 2015; 15(2):73–86.

68. Pancione M, Giordano G, Remo A, Febbraro A, Sabatino L, Manfrin E, Ceccarelli M, Colantuoni V. Immune Escape Mechanisms in Colorectal Cancer Pathogenesis and Liver Metastasis. 2014; 2014.

69. Van Den Eynden GG, Majeed AW, Illemann M, Vermeulen PB, Bird NC, Høyer-Hansen G, Eefsen RL, Reynolds AR, Brodt P. The multifaceted role of the microenvironment in liver metastasis: Biology and clinical implications. Cancer Res. 2013; 73(7):2031–43.

70. Brodt P. Role of the microenvironment in liver metastasis: From pre- to prometastatic niches. Clin Cancer Res. 2016; 22(24):5971–82.

71. Santaolalla R, Fukata M, Abreu M. Innate immunity in the small intestine. Curr Opin Gastroenterol. 2011; 27(2):124–31.

72. Chassaing B, Kumar M, Baker MT, Singh V, Kumar MV-. Mammalian gut immunity. Biomed J. 2014; 37(5):246–58.

73. Ullman TA, Itzkowitz SH. Intestinal inflammation and cancer. Gastroenterology. 2011; 140(6):1807–16.

74. Lucas C, Barnich N, Nguyen HTT. Microbiota, inflammation and colorectal cancer. Int J Mol Sci. 2017; 18(6):1310.

75. de la Motte CA. Hyaluronan in intestinal homeostasis and inflammation: implications for fibrosis. AJP Gastrointest Liver Physiol. 2011; 301(6):G945–9.

76. Pearce OMT, Läubli H. Sialic acids in cancer biology and immunity. Glycobiology. 2015; 26(2):111–28.

77. Park JJ, Lee M. Increasing the α 2, 6 sialylation of glycoproteins may contribute to metastatic spread and therapeutic resistance in colorectal cancer. Gut Liver. 2013; 7(6):629–41.

78. Varki A, Gagneux P. Multifarious roles of sialic acids in immunity. Ann N Y Acad Sci. 2012; 1253(1):16–36.

79. Kotredes KP, Thomas B, Gamero AM. The protective role of Type I interferons in the gastrointestinal tract. Front Immunol. 2017; 8:410.

80. Daulat AM, Borg JP. Wnt/Planar Cell Polarity Signaling: New Opportunities for Cancer Treatment. Trends in Cancer. 2017; 3(2):113–25.

81. Roussos ET, Condeelis JS, Patsialou A. Chemotaxis in cancer. Nat Rev Cancer. 2011; 11(8):573–87.

82. Mariotti A, Perotti A, Sessa C, Rüegg C. N-cadherin as a therapeutic target in cancer. Expert Opin Investig Drugs. 2007; 16(4):451–65.

83. Mlecnik B, Bindea G, Kirilovsky A, et al. The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis. Sci Transl Med. 2016; 8(327):327ra26.

84. Tonini T, Rossi F, Claudio PP. Molecular basis of angiogenesis and cancer. Oncogene. 2003; 22:6549–56.

85. Li Y, Wang L, Pappan L, Galliher-beckley A, Shi J. IL-1 β promotes stemness and invasiveness of colon cancer cells through Zeb1 activation. Mol Cancer. 2012; 11:87.

86. Tang D, Tao D, Fang Y, Deng C, Xu Q, Zhou J. TNF-Alpha Promotes Invasion and Metastasis via NF-Kappa B Pathway in Oral Squamous Cell Carcinoma. Med Sci Monit Basic Res. 2017; 23:141–9.

87. Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? Acta Pharmacol Sin. 2008; 29(11):1275–88.

88. Xia Y, Shen S, Verma IM. NF-κB, an Active Player in Human Cancers. Cancer Immunol Res. 2014; 2(9):823–30.

89. Venkatesh V, Nataraj R, Thangaraj GS, Karthikeyan M, Gnanasekaran A, Kaginelli SB, Kuppanna G, Kallappa CG, Basalingappa KM. Targeting Notch signalling pathway of cancer stem cells. Stem Cell Investig. 2018; 5:5.

90. Leach J, Morton JP, Sansom OJ. Neutrophils: Homing in on the myeloid mechanisms of metastasis. Mol Immunol. 2017; pii: S0161(17):30615–6.

91. Takeuchi O, Akira S. Pattern Recognition Receptors and Inflammation. Cell. 2010; 140(6):805–20.

92. Drickamer K, Taylor ME. Recent insights into structures and functions of C-type lectins in the immune system. Curr Opin Struct Biol. 2015; 34:26–34.

93. Sebestyen E, Zawisza M, Eyras E. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. Nucleic Acids Res. 2015; 43(3):1345–56.

94. Porru M, Pompili L, Caruso C, Biroccio A, Leonetti C. Targeting KRAS in metastatic colorectal cancer: current strategies and emerging opportunities. J Exp Clin Cancer Res. 2018; 37(1):57.

95. Kong G, You X, Wen Z, Chang YI, Qian S, Ranheim EA, Letson C, Zhang X, Zhou Y, Liu Y, Rajagopalan A, Zhang J, Stieglitz E, Loh M, Hofmann I, Yang D, Zhong X5, Padron E, Zhou L, Pear WS ZJ. Downregulating Notch counteracts KrasG12D-induced ERK activation and oxidative phosphorylation in myeloproliferative neoplasm. Leukemia. 2018;

96. Ambrogio C, Gómez-López G, Falcone M, et al. Combined inhibition of DDR1 and Notch signaling is a therapeutic strategy for KRAS-driven lung adenocarcinoma. Nat Med. 2016; 22(3):270–7.

97. Chalovich JM, Eisenberg E. Notch and Kras in pancreatic cancer: at the crossroads of mutation, differentiation and signaling. Magn Reson Imaging. 2013; 31(3):477–9.

98. Lou Y, Liu S. The TIPE (TNFAIP8) family in inflammation, immunity, and cancer. Mol Immunol. 2011; 49(1–2):4–7.

99. Wang Y, Liu Y, Hu C, Ni X, Huang X. Tumor necrosis factor α-induced protein 8-like 1 promotes apoptosis by regulating B-cell leukemia/lymphoma-2 family proteins in RAW264.7 cells. Oncol Lett. 2016; 12(5):3506–12.

100. Tsachaki M, Mladenovic N, Štambergová H, Birk J, Odermatt A. Hexose-6-phosphate dehydrogenase controls cancer cell proliferation and migration through pleiotropic effects on the unfolded-protein response, calcium homeostasis, and redox balance. FASEB J. 2018; 32(5):2690–705.

101. Mai S, Qu X, Li P, Ma Q, Cao C, Liu X. Global regulation of alternative RNA splicing by the SR-rich protein RBM39. Biochim Biophys Acta - Gene Regul Mech. 2016; 1859(8):1014–24.

102. Sillars-Hardebol AH, Carvalho B, Belién JAM, De Wit M, Delis-Van Diemen PM, Tijssen M, Van De Wiel MA, Pontén F, Meijer GA, Fijneman RJA. CSE1L, DIDO1 and RBM39 in colorectal adenoma to carcinoma progression. Cell Oncol. 2012; 35(4):293–300.

103. Tzeng H-T, Su C-C, Chang C-P, Lai W-W, Su W-C, Wang Y-C. Rab37 in lung cancer mediates exocytosis of soluble ST2 and thus skews macrophages toward tumor-suppressing phenotype. Int J Cancer. 2018; 143(7):1–31.

104. Tzeng HT, Wang YC. Rab-mediated vesicle trafficking in cancer. J Biomed Sci. 2016; 23:70.

105. Cho S-H, Kuo I-Y, Lu P-JF, Tzeng H-T, Lai W-W, Su W-C, Wang Y-C. Rab37 mediates exocytosis of secreted frizzled-related protein 1 to inhibit Wnt signaling and thus suppress lung cancer stemness. Cell Death Dis. 2018; 9(9):868.

106. Leo E, Deveraux QL, Buchholtz C, Welsh K, Matsuzawa SI, Stennicke HR, Salvesen GS, Reed JC. TRAF1 Is a Substrate of Caspases Activated during Tumor Necrosis Factor Receptor-alpha-induced Apoptosis. J Biol Chem. 2001; 276(11):8087–93.

107. Niu ZY, Li WL, Jiang DL, Li YS, Xie XJ. Mir-483 inhibits colon cancer cell proliferation and migration by targeting TRAF1. Kaohsiung J Med Sci. 2018; 34(9):479–86.

108. Sundram V, Chauhan SC, Jaggi M. Emerging roles of protein kinase D1 in cancer. Mol Cancer Res. 2011; 9(8):985–96.

109. Bhattacharya S, Srinivasan K, Abdisalaam S, et al. RAD51 interconnects between DNA replication, DNA repair and immunity. Nucleic Acids Res. 2017; 45(8):4590–605.

110. Tong H, Miyake Y, Mi-ichi F, Iwakura Y, Hara H, Yoshida H. Apaf1 plays a negative regulatory role in T cell responses by suppressing activation of antigen-stimulated T cells. PLoS One. 2018; 13(3):e0195119.

111. Hu H, Xu D, Huang X, Zhu C, Xu J, Zhang Z, Zhao G. Keratin17 Promotes Tumor Growth and is Associated with Poor Prognosis in Gastric Cancer. J Cancer. 2018; 9(2):346–57.

112. Masuda T, Ishikawa T, Mogushi K, Okazaki S, Ishiguro M, Iida S, Mizushima H, Tanaka H, Uetake H, Sugihara K. Overexpression of the S100A2 protein as a prognostic marker for patients with stage II and III colorectal cancer. Int J Oncol. 2016; 48(3):975–82.

113. Donato R, R. Cannon B, Sorci G, Riuzzi F, Hsu K, J. Weber D, L. Geczy C. Functions of S100 Proteins. Curr Mol Med. 2013; 13(1):24–57.

114. Wilson AL, Schrecengost RS, Guerrero MS, Thomas KS, Bouton AH. Breast Cancer Antiestrogen Resistance 3 (BCAR3) Promotes Cell Motility by Regulating Actin Cytoskeletal and Adhesion Remodeling in Invasive Breast Cancer Cells. PLoS One. 2013; 8(6):e65678.

115. Hermiston ML, Xu Z, Weiss A. CD45: A Critical Regulator of Signaling Thresholds in Immune Cells. Annu Rev Immunol. 2003; 21:107–37.

116. B Y, L X, M C, D Z, S L. Effect of tumor necrosis factor-α-induced protein 8 on the immune response of CD4+ T lymphocytes in mice following acute insult. Mol Med Rep. 2018; 17(5):6655–60.

117. Toes REM, Ossendorp F, Offringa R, Melief CJM. CD4 T Cells and Their Role in Antitumor Immune Responses. J Exp Mediicine. 1999; 189(5):753–6.

118. Fukuda T, Sumiyoshi T, Takahashi M, Kataoka T, Asahara T, Inui H, Watatani M, Yasutomi M, Kamada N, Miyagawa K. Alterations of the Double-Strand Break Repair Gene MRE11 in Cancer. Cancer Res. 2001; 61(1):23–6.

119. Yuan SSF, Hou MF, Hsieh YC, Huang CY, Lee YC, Chen YJ, Lo S. Role of MRE11 in cell proliferation, tumor invasion, and DNA repair in breast cancer. J Natl Cancer Inst. 2012; 104(19):1485–502.

120. Scurrah M, Haan S De, Olivera E, Ccanto R, Creed H. SREBP1 drives KRT80-dependent cytoskeletal changes and invasive behavior in endocrine resistant ERα breast cancer. bioRxiv. 2011;

121. Liu Z, Hou J, Ren L, He J, Sun B, Sun LZ, Wang S. Protein tyrosine phosphatase receptor type O expression in the tumor niche correlates with reduced tumor growth, angiogenesis, circulating tumor cells and metastasis of breast cancer. Oncol Rep. 2015; 33(4):1908–14.

122. Fish L, Pencheva N, Goodarzi H, Tran H, Yoshida M, Tavazoie SF. Muscleblind-like 1 suppresses breast cancer metastatic colonization and stabilizes metastasis suppressor transcripts. Genes Dev. 2016; 30(4):386–98.

123. Harada T, Yamamoto H, Kishida S, Kishida M, Awada C, Takao T, Kikuchi A. Wnt5b-associated exosomes promote cancer cell migration and proliferation. Cancer Sci. 2017; 108(1):42–52.

124. Narayan S, Roy D. Role of APC and DNA mismatch repair genes in the development of colorectal cancers. Mol Cancer. 2003; 2:41.

125. Jackson CJ, Xue M, Thompson P, Davey RA. Activated protein C prevents inflammation yet stimulates angiogenesis to promote cutaneous wound healing. Wound Repair Regen. 2005; 13(3):284–94.

126. Arentz G, Chataway T, Price TJ, Izwan Z, Hardi G, Cummins AG, Hardingham JE. Desmin expression in colorectal cancer stroma correlates with advanced stage disease and marks angiogenic microvessels [Internet]. In: Clinical Proteomics. BioMed Central Ltd; 2011. page 16.

127. Wu Y, Antony S, Hewitt SM, et al. Functional activity and tumor-specific expression of dual oxidase 2 in pancreatic cancer cells and human malignancies characterized with a novel monoclonal antibody. Int J Oncol. 2013; 42(4):1229–38.

128. Pan H, Gu L, Liu B, et al. Tropomyosin-1 acts as a potential tumor suppressor in human oral squamous cell carcinoma. PLoS One. 2017; 12(2):e0168900.

129. Gardina PJ, Clark T a, Shimada B, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics. 2006; 7:325.

130. Dube S, Yalamanchili S, Lachant J, Abbott L, Benz P, Mitschow C, Dube DK, Poiesz BJ. Expression of Tropomyosin 1 Gene Isoforms in Human Breast Cancer Cell Lines. Int J Breast Cancer. 2015; 2015:859427.

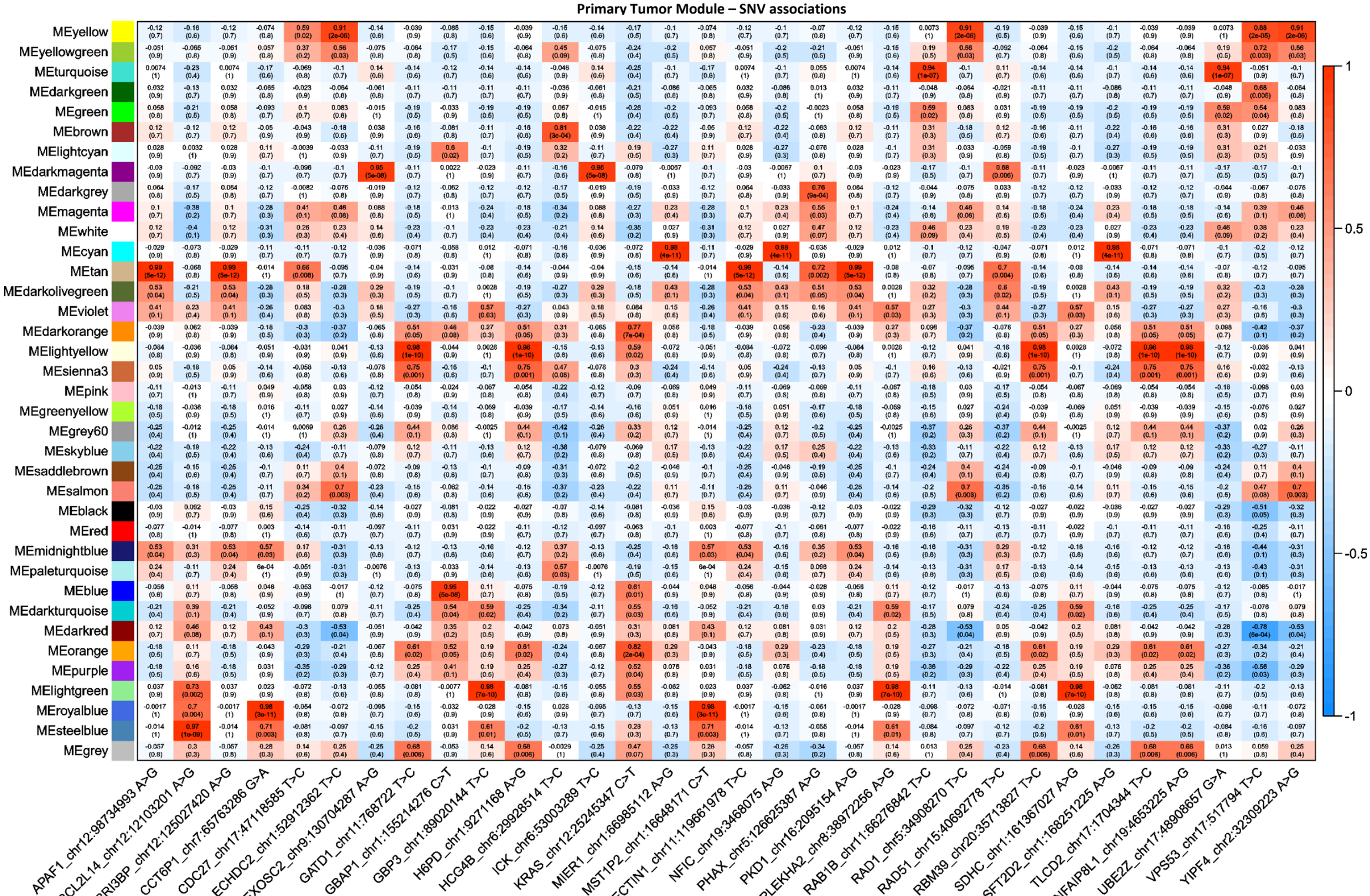
131. Tavares S, Vieira AF, Taubenberger AV, et al. Actin stress fiber organization promotes cell stiffening and proliferation of pre-invasive breast cancer cells. Nat Commun. 2017; 8(May).

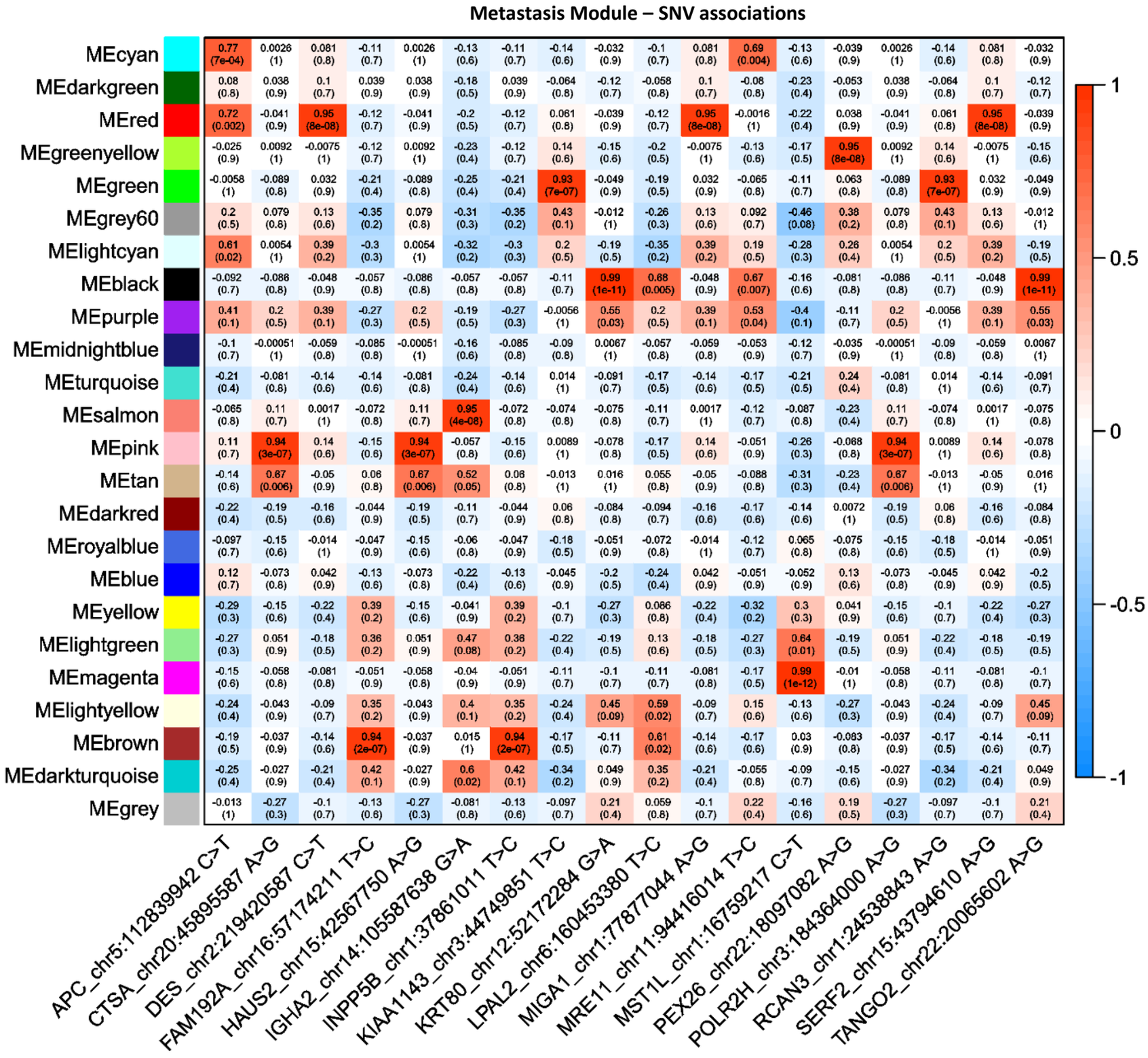
132. Zhu S, Si ML, Wu H, Mo YY. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J Biol Chem. 2007; 282(19):14328–36.

## 8. Supplemental figures legends

**Supplemental Figure 1**. Correlation of the somatic variants and module-trait relationship from PT condition. Each cell contains the test statistic value and its equivalent p-value.

**Supplemental Figure 2**. Correlation of the somatic variants and module-trait relationship from MET condition. Each cell contains the test statistic value and its equivalent p-value.


Primary Tumor Module – SNV associations
MEyellow
MEyellowgreen
MEturquoise
MEdarkgreen
MEgreen
MEbrown
MElightcyan
MEdarkmagenta
MEdarkgrey
MEmagenta
MEwhite
MEcyan
MEtan
MEdarkolivegreen
MEviolet
MEdarkorange
MElightyellow
MEsienna3
MEpink
MEgreenyellow
MEgrey60
MEskyblue
MEsaddlebrown
MEsalmon
MEblack
MEred
MEmidnightblue
MEpaleturquoise
MEblue
MEdarkturquoise
MEdarkred
MEorange
MEpurple
MElightgreen
MEroyalblue
MEsteelblue
MEgrey
1
0.5
0
-0.5
-1
APAF1_chr12:98734993 A>G
BCL2L14_chr12:12103201 A>G
BRI3BP_chr12:125027420 A>G
CCT6P1_chr7:65763286 G>A
CDC27_chr17:47118585 T>C
ECHDC2_chr1:52912362 T>C
EXOSC2_chr9:130704287 A>G
GATD1_chr11:768722 T>C
GBAP1_chr1:155214276 C>T
GBP3_chr1:89020144 T>C
H6PD_chr1:9271168 A>G
HCG4B_chr6:29928514 T>C
ICK_chr6:53003289 T>C
KRAS_chr12:25245347 C>T
MIER1_chr1:66985112 A>G
MST1P2_chr1:16648171 C>T
NECTIN1_chr11:119661978 T>C
NFIC_chr19:3468075 A>G
PHAX_chr5:126625387 A>G
PKD1_chr16:2095154 A>G
PLEKHA2_chr8:38972256 A>G
RAB1B_chr11:66276842 T>C
RAD1_chr5:34908270 T>C
RAD51_chr15:40692778 T>C
RBM39_chr20:35713827 T>C
SDHC_chr1:161367027 A>G
SFT2D2_chr1:168251225 A>G
TLCD2_chr17:1704344 T>C
TNFAIP8L1_chr19:4653225 A>G
UBE2Z_chr17:48908657 G>A
VPS53_chr17:517794 T>C
YIPF4_chr2:32309223 A>G


Metastasis Module – SNV associations
MEcyan
MEdarkgreen
MEred
MEgreenyellow
MEgreen
MEgrey60
MElightcyan
MEblack
MEpurple
MEmidnightblue
MEturquoise
MEsalmon
MEpink
MEtan
MEdarkred
MEroyalblue
MEblue
MEyellow
MElightgreen
MEmagenta
MElightyellow
MEbrown
MEdarkturquoise
MEgrey
1
0.5
0
-0.5
-1
APC_chr5:112839942 C>T
CTSA_chr20:45895587 A>G
DES_chr2:219420587 C>T
FAM192A_chr16:57174211 T>C
HAUS2_chr15:42567750 A>G
IGHA2_chr14:105587638 G>A
INPP5B_chr1:37861011 T>C
KIAA1143_chr3:44749851 T>C
KRT80_chr12:52172284 G>A
LPAL2_chr6:160453380 T>C
MIGA1_chr1:77877044 A>G
MRE11_chr11:94416014 T>C
MST1L_chr1:16759217 C>T
PEX26_chr22:18097082 A>G
POLR2H_chr3:184364000 A>G
RCAN3_chr1:24538843 A>G
SERF2_chr15:43794610 A>G
TANGO2_chr22:20065602 A>G

# DISCUSSÃO

Até o momento, poucos trabalhos têm sido realizados considerando as isoformas transcricionais como foco de estudo, sendo geralmente avaliada a expressão gênica para as análises de dados. Contudo, como o mecanismo de *splicing* alternativo é predominante em 95% dos genes humanos, e 70% deles apresentam isoformas com importantes mudanças funcionais e estruturais na proteína, é crucial compreender as alterações ao nível de isoformas. Um estudo considerando técnicas de *machine learning* mostrou que a quantificação da expressão das isoformas apresentou resultados superiores na maioria das situações para a classificação biológica de condições normais e tumorais em relação às análises considerando a quantificação da expressão do gene [60]. Essa informação ressalta a importância deste trabalho que abrange uma área de estudo ainda em crescimento. No entanto, ainda existem alguns desafios, entre os quais podemos citar a falta de atualizações de ferramentas e banco de dados específicos, anotações não consensuais e incompletas, informações disponíveis para poucos organismos, falta de ferramentas que possibilitem análises dos mecanismos de regulação do *splicing*, entre outras.

O CRC é uma das malignidades mais comuns no mundo, e uma das principais causas de morte por câncer, principalmente quando associado com a presença de metástase. Em torno de 30% dos pacientes apresentam CRC metastático, e destes, 50% exibem metástase hepática [3,38]. O desenvolvimento e a progressão do CRC ocorrem através de modificações genéticas e epigenéticas tanto em oncogenes como em genes supressores de tumores, proporcionando uma vantagem à malignidade, onde o epitélio da mucosa se torna hiperproliferativo dando origem ao adenoma benigno, que pode ou não formar o CRC e gerar metástases. As mutações sequenciais e cumulativas são as principais modificações que ocorrem nessa transformação, podendo levar à geração do CRC esporádico quando essas mutações são somáticas, ou levar às síndromes e doenças hereditárias envolvidas com o CRC quando existem mutações germinativas e predisposição genética.

A heterogeneidade celular é muito presente em tumores sólidos, como o CRC, sendo explicada principalmente pelos eventos que ocorrem em algumas células que conferem a elas a capacidade de formar um tumor [61]. Existem dois modelos que buscam explicar a progressão do tumor primário [61]. Esses modelos são a progressão linear, que sugere que o genótipo das células fundadoras da metástase é muito similar ao das células do tumor primário, mas estas

células sofrem uma adaptação aos sinais do microambiente ectópico; e o modelo de progressão paralela que prediz que existe uma disparidade genotípica entre as células do tumor primário e as células metastáticas, e alterações somáticas são adquiridas independentemente [61]. De acordo com nossos resultados, foram identificados menores números de mutações em MET em relação à PT (18 revisadas manualmente após a filtragem), enquanto que as diferenças entre PT e NE são maiores (32 revisadas manualmente após a filtragem). Além disso, a análise transcricional indicou que 606 isoformas foram trocadas na comparação de PT em relação à NE, e 336 isoformas em MET quando comparado à PT. Baseado nestas informações é possível dizer que mais alterações são encontradas quando há o desenvolvimento de uma lesão primária a partir de um tecido normal, enquanto que a metástase difere pouco do tumor primário. Contudo, para conseguirmos classificar nossos dados de acordo com os modelos, seria necessário entender se as mutações compartilhadas entre PT e MET seriam adquiridas no início da tumorigênese, ou ainda, se as mutações teriam surgido independentemente em cada condição, mas seriam compartilhadas por conferirem alguma vantagem ao tumor. O mesmo acontece para a mudança no uso das isoformas que são compartilhadas entre as condições malignas, sendo difícil estabelecer se a troca de isoformas ocorreu no início da tumorigênese, conferindo vantagens à progressão tumoral e à metástase, ou se essas trocas surgiram independentemente, como um reflexo do processo seletivo que a massa tumoral sofre. No entanto, é fato que os processos biológicos presentes no tumor primário e nas metástases são muito semelhantes, como adesão celular, migração, proliferação, e resposta imune. Em relação à resposta imune, apesar de ambos MET e PT apresentarem uma forte representatividade de processos relacionados ao sistema imunológico, cada condição parece atuar de uma forma um pouco diferente. Nossos resultados foram semelhantes aos observados em câncer de mama, onde a metástase apresentou menor expressão de citocinas quimiotáticas e ativadores do sistema imune, e maior expressão de genes relacionados à imunossupressão [62]. Além disso, possivelmente devido a uma reação imunológica relacionada à entrada das células disseminadas no ambiente hepático ou de fatores secretados pelo tumor primário, é possível que o recrutamento de neutrófilos tenha promovido a formação do nicho pré-metastático, contribuindo para o aumento da inflamação, migração e aderência das células tumorais no endotélio vascular [63].

Dentre as isoformas discutidas na revisão apresentada no Capítulo 1 por colaborarem com o processo metastático, foram identificadas a troca de isoformas em pelo menos uma das

condições malignas em relação à condição normal correspondente à CD44, CPEB2, CTTN, TPM1, CXCL12 e as integrinas ITGA5, ITGB1BP1 e ITGAE. Exceto por CXCL12 e CPEB2, todos estão relacionados com a motilidade celular e capacidade invasiva, destacando os elementos atuantes nesses processos entre os principais alvos alterados pelo mecanismo de *splicing* durante a tumorigênese.

Quando consideramos a troca de isoformas entre as condições, é possível observar isoformas superexpressas em ambas as condições malignas PT e MET que estão associados com a composição e modelagem da matriz extracelular, como FAP, TNC, COL11A1, COL12A1, COL4A2, ADAMTS2, e LAMA1 demonstrando mais uma vez que existe uma prevalência de troca de isoformas em elementos que contribuem com a invasividade [64–66].

No entanto, também foi observada a troca de isoformas em MET e PT em processos cruciais para o desenvolvimento tumoral. Um exemplo disso são os supressores de tumor DUSP10 e RB1, que da mesma forma que CDK14 e CDKN2C, estão envolvidas na proliferação e/ou regulação do ciclo celular [67–70].

Envolvidos com o processamento de RNA, encontramos a troca de isoformas em CLK1 e ADAR, além de MBNL1 descrita no Capítulo 2. Durante a transformação maligna, a isoforma CLK1-204 (ENST00000432425) expressa em NE é substituída pela superexpressão de CLK1-201 (ENST00000321356) em ambas as condições malignas. CLK1 é uma cinase que está envolvida na regulação do mecanismo de *splicing*, promovendo a fosforilação de fatores de *splicing* [71,72]. Já foi mostrado que a inibição de CLK1 suprime a proliferação celular e induz a apoptose, tornando CLK1 um interessante alvo terapêutico. A isoforma ADAR-202 (ENST00000368474) é uma proteína envolvida na edição de RNA, estabilidade de RNA e eficiência de tradução em resposta ao interferon tipo I. ADAR-202 é um *hub* positivamente regulado em NE em comparação com PT e o MET, contudo, em ambas as condições malignas existem uma troca de isoformas que favorece a expressão da isoforma canônica ADAR-201 (ENST00000368471). Como a edição de RNA é específica de acordo com a célula ou ambiente, essa troca de isoformas pode estar relacionada com a transformação maligna. De fato, foi demonstrado que a edição de RNA contribui para a diversidade proteômica do câncer, e propõe-se que alguns eventos de edição possam atuar como "condutores" para a iniciação do tumor [73,74].

Apesar de CLK1 ser a única cinase reguladora dos fatores de *splicing* que apresentou troca de isoformas, foi encontrado em PT comparado com NE a superexpressão da cinase

SRPK3 (log2FC 2,1), e das *RNA-binding proteins* KHDRBS2 (log2FC 4,4) e CELF4 (log2FC 5,1) que regulam o *splicing* alternativo, influenciando na escolha dos sítios de *splicing* e afetando a inclusão (ou exclusão) de exons e introns.

Em relação à predição de consequências funcionais associadas à troca de isoformas, em muitos casos não foram possíveis encontrar nenhum indicador de que as diferentes isoformas expressas apresentariam modificações que pudessem alterar a função ou o destino da isoforma. Nesses casos, frequentemente a troca das isoformas apresentavam transcritos que se diferenciavam por tamanho de pares de bases e aminoácidos na proteína final. Nesse caso, podemos especular que possivelmente grandes mudanças no tamanho poderiam levar a uma alteração da conformação proteica, mas devido a carência de dados e informações, nada indica que de fato a atuação ou a localização da proteína poderia ser distinta entre essas isoformas. Ou ainda, de forma especulativa, é possível que a diferença de tamanho em pares de bases da isoforma transcrita pode representar a presença de sequências regulatórias diferentes entre as isoformas.

No entanto, algumas modificações funcionais foram observadas em algumas isoformas, dentre os quais apresentam importante papel no desenvolvimento e progressão do tumor. Dentre elas podemos citar PDK1, CALD1, SFRP1 e MSR1.

PDK1-201 (ENST00000268673) é prevalente em NE, contudo, PDK1-206 (ENST00000461815) encontra-se superexpressa em PT e MET. A isoforma PDK1-206 perde o domínio PH_3 (*pleckstrin homology domain*, PF14593) e tem uma alteração na posição do domínio Pkinase (*protein kinase domain*, PF00069). PDK1 apresenta um importante papel na sinalização celular, sendo muito conhecida por fosforilar AKT e em estados patológicos está relacionada com o ciclo celular anormal, escape da apoptose, reprogramação metabólica, angiogênese, invasão, disseminação, e metástase [75]. A ausência de *pleckstrin homology domain* pode interferir na atuação e localização da proteína, uma vez que este domínio interage com fosfatidilinositol (3,4,5)-fosfato na bicamada lipídica e apresenta importante papel no recrutamento de proteínas e sinalização celular [76,77].

CALD1-201 (ENST00000361675) é uma isoforma presente em NE e substituída pela expressão de CALD1-202 (ENST00000361901) nas condições PT e MET. A isoforma CALD1-201 é do tipo *h*-CALD1 (*high molecular weight*, 89–93 kDa) e está presente no músculo liso, enquanto que CALD1-202 é não-muscular do tipo *l*-CALD1 (*low molecular weight*, 59–63 kDa). Ambas tem importante papel na dinâmica e arquitetura do citoesqueleto,

mas *l*-CALD1 já foi identificada por estar superexpressa em lesões primárias e metástase hepática de CRC e contribuir para a progressão tumoral e resistência à quimioterápicos [78].

Em NE, a isoforma SFRP1-201 (ENST00000220772) é superexpressa em relação a MET, enquanto que MET regula positivamente SFRP1-202 (ENST00000379845). SFRP1 já foi descrita por apresentar um papel importante na prevenção do câncer e é regulada negativamente em vários tipos de tumores [79,80]. SFRP1 atua na regulação negativa da β-catenina, c-Myc e ciclina D1, suprimindo o crescimento do tumor e a metástase [79,80]. A troca de SFRP1-201 para SFRP1-202 pode contribuir para a metástase, uma vez que a isoforma expressa na metástase (SFRP1-202) não apresenta domínio do peptídeo sinal e devido a isso, pode não realizar sua atividade supressora.

O receptor *scavenger* MSR1 atua como um receptor de reconhecimento de padrões, ligando-se a uma ampla gama de ligantes e reconhecendo moléculas alteradas, proteínas de estresse extracelular e padrões moleculares associados a patógenos [81]. É expresso principalmente em macrófagos e células dendríticas e desempenha um papel importante na depuração de células apoptóticas e outros ligantes [82]. No entanto, estudos revelaram que no tumor, o MSR1 está associado a um fenótipo de câncer mais agressivo e sua expressão é restrita a macrófagos associados a tumores, um componente crucial do microambiente tumoral que contribui para o crescimento, invasão e metástase de tumores [82,83]. As células dendríticas têm o papel essencial de detectar anormalidades através de receptores de reconhecimento de padrões e, induzindo a resposta imune [81]. Foi observado que o MSR1 pode realizar uma regulação negativa da resposta inflamatória e atenuar a ativação de células dendríticas no tumor [81,84]. Essa atividade pode estar relacionada com a tolerância imunológica ao tumor, uma vez que o comprometimento entre a tolerância e a imunidade das células T é definido em parte pelo estado de ativação das células dendríticas [84]. Nas nossas análises, a isoforma MSR1-204 (ENST00000381998) é superexpressa em PT e MET, enquanto MSR1-205 (ENST00000445506) é prevalente em NE. Interessantemente, a isoforma MSR1-204 perde o domínio SRCR (*scavenger receptor cysteine-rich*, PF00530) que está presente na membrana celular e é responsável pela função ligante da proteína [85]. Assim, a ausência deste domínio pode estar associada à atividade anormal no reconhecimento de padrões moleculares e possivelmente à tolerância imune.

Em resumo, os dados obtidos no Capítulo 2 mostram que existem ainda muitos potenciais alvos a serem estudados por apresentarem diferença na isoforma expressa entre as

condições normais, tumor primário e metástase. As observações das alterações do mecanismo de *splicing* observadas neste trabalho mostram a importância de se estudar mais detalhadamente à regulação dos mecanismos transcricionais, de forma a obter uma resposta mais precisa das possíveis alterações moleculares que são encontradas em diferentes condições e mesmo entre diferentes tecidos. Essa ideia é especialmente atraente devido a experimentos recentes que mostram que a utilização de dados de expressão ao nível de isoforma podem gerar resultados mais confiáveis. Adicionalmente, essas informações se enriquecem quando é possível associar a alteração do uso das isoformas com a predição de uma modificação funcional, porém, para isso ainda é necessário o aumento do conhecimento atual com novas informações confirmadas a partir de dados experimentais e armazenadas em banco de dados.

# CONCLUSÃO

A partir dos resultados pode-se concluir que:

- Existe a troca de isoformas de RNA entre as condições normais, lesão primária e metástase;
- Muitas isoformas de RNA que apresentam uso diferencial entre as condições apresentam consequências funcionais, indicando que uma possível alteração da atividade ou destino;
- Embora os eventos de *splicing* possam ser influenciados por mutações genéticas, nosso trabalho mostrou que a troca das isoformas não são diretamente afetadas pela presença de variantes somáticas;
- Apenas os genes *DES*, *MST1L* e *KRAS* foram afetados por variantes somáticas e troca de isoformas, porém, não foi possível associar a presença de mutações com os eventos de *splicing* ou geração de isoformas transcricionais específicas (testes feitos com MutPred não foram apresentados);
- Embora as mutações não tenham apresentado um efeito direto sobre as isoformas, foi possível observar que existe uma forte correlação de algumas variantes somáticas com determinados módulos na rede de correlação, bem como a presença de isoformas significantes nesta correlação sugere que existe uma conexão indireta do efeito da mutação, possivelmente afetando vias de sinalização que levam à alteração dos processos biológicos e favorecem do tumor;
- Dentre os principais processos relacionados com a troca das isoformas de RNA estão adesão celular, migração, sobrevivência e proliferação celular, apoptose e resposta imune.

# PERSPECTIVAS

- Considerar um novo *dataset* para confirmar as alterações observadas neste trabalho;
- Validar a expressão das isoformas de RNA mencionadas por qRT-PCR;
- Executar ensaios de silenciamento para investigar as possíveis funções desempenhadas pelas isoformas detectadas;
- Realizar análises de Kaplan-Meier para avaliar o possível impacto na sobrevida de pacientes que apresentam superexpressão das isoformas transcricionais selecionadas.

## REFERÊNCIAS


1. Granados-Romero JJ, Valderrama-Treviño AI, Contreras-Flores EH, Barrera-Mera B, Enríquez MH, Uriarte-Ruíz K, Ceballos-Villalba JC, Estrada-Mata AG, Rodríguez Alvarado C, Arauz-Peña G. Colorectal cancer : a review. Int J Res Med Sci. 2017; 5(11):4667–76.
2. Instituto Nacional do Câncer. http://www.inca.gov.br/estimativa/2018/2018;
3. Martini G, Troiani T, Cardone C, et al. Present and future of metastatic colorectal cancer treatment: A review of new candidate targets. World J Gastroenterol. 2017; 23(26):4675–88.
4. Deverakonda A. Diagnosis and Treatment of Colorectal Cancer: A Review. Res Rev J Med Heal Sci. 2016; 5(3):1–15.
5. Rodriguez-Salas N, Dominguez G, Barderas R, Mendiola M, García-Albéniz X, Maurel J, Batlle JF. Clinical relevance of colorectal cancer molecular subtypes. Crit Rev Oncol Hematol. 2017; 109:9–19.
6. De Rosa M, Pace U, Rega D, Costabile V, Duraturo F, Izzo P, Delrio P. Genetics, diagnosis and management of colorectal cancer (Review). Oncol Rep. 2015; 34(3):1087–96.
7. Winawer SJ. The History of Colorectal Cancer Screening: A Personal Perspective. Dig Dis Sci. 2015; 60(3):596–608.
8. Morson B. Precancerous and early malignant lesions of the large intestine. Br J Surg. 1968; 55(10):725–31.
9. Winawer S, Zauber A, Ho M, O'Brien M, Gottlieb L, Sternberg S, Waye J, Schapiro M, Bond J, Panish J. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med. 1993; 329:1977–1981.
10. Mueller JD, Bethke B, Stolte M. Colorectal de novo carcinoma: a review of its diagnosis, histopathology, molecular biology, and clinical relevance. Virchows Arch. 2002; 440(5):453–60.
11. Walsh JME, Terdiman JP. Colorectal Cancer Screening Scientific Review. JAMA. 2003; 289(10):1288–96.
12. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Smits AMM, Bos JL. Genetic Alterations during Colorectal-Tumor Development. N Engl J

Med. 1988; 319(9):525–32.

13. Fearon EF, Vogelstein B. A Genetic Model for Colorectal Tumorigenesis. Cell. 1989; 61(1):759–67.
14. Bogaert J, Prenen H. Molecular genetics of colorectal cancer. Ann Gastroenterol. 2014; 27(1):9–14.
15. Tomasetti C, Marchionni L, Nowak MA, Parmigiani G, Vogelstein B. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci. 2015; 112(1):118–23.
16. Tariq K, Ghias K. Colorectal cancer carcinogenesis: a review of mechanisms. Cancer Biol Med. 2016; 13(1):120–35.
17. Coli AP, Journal S, Gatroenterology O, Crcs KCRC, Arabia S, Europe E. KRAS and TP53 Mutations in Colorectal Carcinoma. October. 2009; 15(4):2009–11.
18. Ballester V, Rashtak S, Boardman L. Clinical and molecular features of young-onset colorectal cancer. World J Gastroenterol. 2016; 22(5):1736–44.
19. Colussi D, Brandi G, Bazzoli F, Ricciardiello L. Molecular Pathways Involved in Colorectal Cancer: Implications for Disease Behavior and Prevention. Int J Mol Sci. 2013; 14:16365–85.
20. Molinari F, Frattini M. Functions and Regulation of the PTEN Gene in Colorectal Cancer. Front Oncol. 2014; 3:326.
21. Schumacher V, Vogel T, Leube B, Driemel C, Goecke T, Möslein G, Royer-Pokora B. STK11 genotyping and cancer risk in Peutz-Jeghers syndrome. J Med Genet. 2005; 42(5):428–35.
22. Tauriello DVF, Batlle E. Targeting the Microenvironment in Advanced Colorectal Cancer. Trends in Cancer. 2016; 2(9):495–504.
23. Hisamuddin IM, Yang VW. Molecular Genetics of Colorectal Cancer: An Overview. Curr Color Cancer Rep. 2006; 2(2):53–9.
24. Beumer J, Clevers H. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. Development. 2016; 143(20):3639–49.
25. Morin PJ, Kinzler KW, Sparks AB. β-Catenin mutations: Insights into the APC pathway and the power of genetics. Cancer Res. 2016; 76(19):5587–9.
26. Medema JP, Vermeulen L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. Nature. 2011; 474(7351):318–26.

27. Zenonos K. RAS signaling pathways, mutations and their role in colorectal cancer. World J Gastrointest Oncol. 2013; 5(5):97–101.
28. Júnior RAA, de Souza CM, de Azevedo MLV, Netto MRM, Baldin RKS, Sebastião APM, Soares LFP, Tullio LF, de Noronha L. O papel de PI3K e COX2 na carcinogênese de pólipos colorretais. J Coloproctology. 2018; 38(1):1–8.
29. Aoki T, Narumiya S. Prostaglandin E2-EP2 signaling as a node of chronic inflammation in the colon tumor microenvironment. Inflamm Regen. 2017; 37(1):4.
30. Dingzhi Wang, Lingchen Fu, Haiyan Sun, Lixia Guo and RND. Prostaglandin E2 Promotes Colorectal Cancer Stem Cell Expansion and Metastasis in Mice. Gastroenterology. 2015; 149(7):1884–1895.
31. Park M, Hong J. Roles of NF-κB in Cancer and Inflammatory Diseases and Their Therapeutic Approaches. Cells. 2016; 5(2):15.
32. Hassanzadeh P. Colorectal cancer and NF-κB signaling pathway. Gastroenterol Hepatol from Bed to Bench. 2011; 4(3):127–32.
33. Liu Q, Zhang H, Jiang X, Qian C, Liu Z, Luo D. Factors involved in cancer metastasis: A better understanding to “seed and soil” hypothesis. Mol Cancer. 2017; 16(1):1–19.
34. Guan X. Cancer metastases: Challenges and opportunities. Acta Pharm Sin B. 2015; 5(5):402–18.
35. Nguyen DX, Bos PD, Massagué J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009; 9(4):274–84.
36. Kaplan RN, Rafii S, Lyden D. Preparing the “soil”: The premetastatic niche. Cancer Res. 2006; 66(23):11089–93.
37. Brodt P. Role of the microenvironment in liver metastasis: From pre- to prometastatic niches. Clin Cancer Res. 2016; 22(24):5971–82.
38. Vatandoust S, Price TJ, Karapetis CS. Colorectal cancer: Metastases to a single organ. World J Gastroenterol. 2015; 21(41):11767–76.
39. Fiorentini G, Sarti D, Aliberti C, Carandina R, Mambrini A, Guadagni S. Multidisciplinary approach of colorectal cancer liver metastases. World J Clin Oncol. 2017; 8(3):190–202.
40. Keirsse J, Van Damme H, Geeraerts X, Beschin A, Raes G, Van Ginderachter JA. The role of hepatic macrophages in liver metastasis. Cell Immunol. 2018; 330:202–15.
41. Van Den Eynden GG, Majeed AW, Illemann M, Vermeulen PB, Bird NC, Høyer-

Hansen G, Eefsen RL, Reynolds AR, Brodt P. The multifaceted role of the microenvironment in liver metastasis: Biology and clinical implications. Cancer Res. 2013; 73(7):2031–43.

42. Jin K, Gao W, Lu Y, Lan H, Teng L, Cao F. Mechanisms regulating colorectal cancer cell metastasis into liver (review). Oncol Lett. 2012; 3(1):11–5.
43. Obenauf AC, Massagué J. Surviving at a Distance: Organ-Specific Metastasis. Trends in Cancer. 2015; 1(1):76–91.
44. Kurokawa T, Ohkohchi N. Platelets in liver disease, cancer and regeneration. World J Gastroenterol. 2017; 23(18):3228–39.
45. Dehm SM. mRNA splicing variants: Exploiting modularity to outwit cancer therapy. Cancer Res. 2013; 73(17):5309–14.
46. Chen FC. Alternative RNA structure-coupled gene regulations in tumorigenesis. Int J Mol Sci. 2015; 16(1):452–75.
47. Kim E, Goren A, Ast G. Insights into the connection between cancer and alternative splicing. Trends Genet. 2008; 24(1):7–10.
48. Kim Y-J, Kim H-S. Alternative Splicing and Its Impact as a Cancer Diagnostic Marker. Genomics Inform. 2012; 10(2):74–80.
49. Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer. EMBO Rep. 2008; 9(11):1087–93.
50. Luz FAC, Brígido PC, Moraes AS, Silva MJB. Aberrant Splicing in Cancer: Mediators of Malignant Progression through an Imperfect Splice Program Shift. Oncology. 2017; 92(1):3–13.
51. Shilo A, Siegfried Z, Karni R. The role of splicing factors in deregulation of alternative splicing during oncogenesis and tumor progression. Mol Cell Oncol. 2015; 2(1):e970955.
52. Anczuków O, Krainer AR. Splicing-factor alterations in cancers. RNA. 2016; 22(9):1285–301.
53. Niu Z, Jin W, Zhang L, et al. The role of splicing factors in deregulation of alternative splicing during oncogenesis and tumor progression. Nat Struct Mol Biol. 2012; 586(11):333–49.
54. Cohen-Eliav M, Golan-Gerstl R, Siegfried Z, Andersen CL, Thorsen K, Ørntoft TF, Mu D, Karni R. The splicing factor SRSF6 is amplified and is an oncoprotein in lung

and colon cancers. J Pathol. 2013; 229(4):630–9.

55. Carpenter B, Mckay M, Dundas SR, Lawrie LC, Telfer C, Murray GI. Heterogeneous nuclear ribonucleoprotein K is over expressed, aberrantly localised and is associated with poor prognosis in colorectal cancer. Br J Cancer. 2006; 95(7):921–7.
56. Sveen A, Kilpinen S, Ruusulehto A, Lothe R, Skotheim R. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene. 2015; 35(19):2413–27.
57. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74.
58. Ladomery M. Aberrant alternative splicing is another hallmark of cancer. Int J Cell Biol. 2013; 2013.
59. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene. 2013; 33(46):5311–8.
60. Johnson NT, Dhroso A, Hughes KJ, Korkin D. Biological classification with RNA-Seq data: Can alternative splicing enhance machine learning classifier? RNA. 2017; 24(9):1119–32.
61. Stoecklein NH, Klein CA. Genetic disparity between primary tumours, disseminated tumour cells, and manifest metastasis. Int J Cancer. 2010; 126(3):589–98.
62. Author T, Biology M. Immunological differences between primary and metastatic breast cancer. Ann Oncol. 2018; 10.1093/an.
63. Leach J, Morton JP, Sansom OJ. Neutrophils: Homing in on the myeloid mechanisms of metastasis. Mol Immunol. 2017; pii: S0161(17):30615–6.
64. Przemyslaw L, Oguslaw HA, Elzbieta S, Malgorzata SM. ADAM and ADAMTS family proteins and their role in the colorectal cancer etiopathogenesis. BMB Rep. 2013; 46(3):139–50.
65. Wan L, Pantel K, Kang Y. Tumor metastasis: moving new biological insights into the clinic. Nat Med. 2013; 19(11):1450–64.
66. Fischer H, Salahshor S, Stenling R, Björk J, Lindmark G, Iselius L, Rubio C, Fap C. COL11A1 in FAP polyps and in sporadic colorectal tumors. BMC Cancer. 2001; 1:17.
67. Png CW, Weerasooriya M, Guo J, James SJ, Poh HM, Osato M, Flavell RA, Dong C, Yang H, Zhang Y. DUSP10 regulates intestinal epithelial cell growth and colorectal tumorigenesis. Oncogene. 2016; 35(2):206–17.

68. Indovina P, Pentimalli F, Casini N, Vocca I, Giordano A. RB1 dual role in proliferation and apoptosis: cell fate control and implications for cancer therapy. Oncotarget. 2015; 6(20):17873–90.
69. Ou-Yang J, Huang L-H, Sun X-X. Cyclin-Dependent Kinase 14 Promotes Cell Proliferation, Migration and Invasion in Ovarian Cancer by Inhibiting Wnt Signaling Pathway. Gynecol Obstet Invest. 2017; 82(3):230–9.
70. Lim S, Kaldis P. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development. 2013; 140(15):3079–93.
71. Duncan PI, Stojdl DF, Marius RM, Bell JC. In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. Mol Cell Biol. 1997; 17(10):5996–6001.
72. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK. RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer. 2016; 16(7):413–30.
73. Peng X, Xu X, Wang Y, et al. A-to-I RNA Editing Contributes to Proteomic Diversity in Cancer. Cancer Cell. 2018; 33(5):817–828.e7.
74. Han L, Diao L, Yu S, et al. The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. Cancer Cell. 2015; 28(4):515–28.
75. Gagliardi PA, Puliafito A, Primo L. PDK1: At the crossroad of cancer signaling pathways. Semin Cancer Biol. 2018; 48(March 2017):27–35.
76. Yamamoto E, Kalli AC, Yasuoka K, Sansom MSP. Interactions of Pleckstrin Homology Domains with Membranes: Adding Back the Bilayer via High-Throughput Molecular Dynamics. Structure. 2016; 24(8):1421–31.
77. Lngley E, Hemmings BA. Pleckstrin Homology (PH) Domains in Signal Transduction. J Cell Biochem. 1994; 56:436–43.
78. Kim KH, Yeo SG, Kim WK, et al. Up-regulated expression of l-caldesmon associated with malignancy of colorectal cancer. BMC Cancer. 2012; 12(1):601.
79. Ren XY, Zhou GQ, Jiang W, et al. Low SFRP1 expression correlates with poor prognosis and promotes cell invasion by activating the Wnt/β-catenin signaling pathway in NPC. Cancer Prev Res. 2015; 8(10):968–77.
80. Vincent KM, Postovit LM. A pan-cancer analysis of secreted Frizzled-related proteins: Re-examining their proposed tumour suppressive function. Sci Rep. 2017; 7(February):1–9.
81. Guo C, Yi H, Yu X, Hu F, Zuo D, Subjeck JR, Wang XY. Absence of scavenger

receptor A promotes dendritic cell-mediated cross-presentation of cell-associated antigen and antitumor immune response. Immunol Cell Biol. 2012; 90(1):101–8.

82. Yu X, Guo C, Fisher PB, Subjeck JR, Genetics M. Scavenger Receptors: Emerging Roles in Cancer Biology and Immunology. 2016; 1990:1–45.
83. Neyen C, Pluddemann A, Mukhopadhyay S, Maniati E, Bossard M, Gordon S, Hagemann T. Macrophage Scavenger Receptor A Promotes Tumor Progression in Murine Models of Ovarian and Pancreatic Cancer. J Immunol. 2013; 190(7):3798–805.
84. Wang XY, Facciponte J, Chen X, Subjeck JR, Repasky EA. Scavenger receptor-A negatively regulates antitumor immunity. Cancer Res. 2007; 67(10):4996–5002.
85. Finn RD, Bateman A, Clements J, et al. Pfam: The protein families database. Nucleic Acids Res. 2014; 42:D222–30.