U.PORTO

DOUTORAMENTO

CIÊNCIAS BIOMÉDICAS

Establishing the glycoprofile of CD44 for novel immunotherapies in bladder cancer

Cristiana Milhazes Gaiteiro

D

2024

Establishing the glycoprofile of CD44 for novel immunotherapies in bladder cancer

Cristiana Milhazes Gaiteiro

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

CRISTIANA MILHAZES GAITEIRO

Establishing the glycoprofile of CD44 for novel immunotherapies in bladder cancer

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador: Lúcio José de Lara Santos
Categoria: Professor Catedrático convidado, Diretor do Serviço dos Cuidados Intermédios e Coordenador
Afiliação: Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto e Instituto Português de Oncologia do Porto

Co-orientador: José Alexandre Ribeiro de Castro Ferreira
Categoria: Investigador Principal
Afiliação: Instituto Português de Oncologia do Porto

Co-orientador: Jesús María Prieto Valtueña
Categoria: Professor Emérito
Afiliação: Universidade de Navarra

I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.

Marie Curie

NOTA PRELIMINAR

Declaro que a presente tese é integralmente da minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As fontes de informação e as obras de outros autores encontram-se devidamente referenciadas e identificadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tendo consciência de que a prática de plágio constitui um ilícito académico, esta tese não contém, por isso, qualquer tipo de plágio de textos publicados ou trabalhos académicos.

Os trabalhos de investigação desta tese foram realizados no Grupo de Patologia e Terapêutica Experimental do Centro de Investigação do Instituto Português Oncologia do Porto Francisco Gentil, EPE (IPO-Porto).

Declaro ainda que o presente trabalho foi financiado pela Fundação para a Ciência e a Tecnologia (FCT), pela atribuição de uma bolsa individual de doutoramento, com a referência SFRH/BD/127327/2016. A FCT é co-financiada pelo Fundo Europeu de Desenvolvimento Regional da União Europeia, através do Programa Operacional Fatores de Competitividade do Quadro de Referência Estratégico Nacional. Reconheço também o financiamento, pela FCT, do Centro de Investigação do IPO-Porto (UIDB-00776-2020 e UIDP-00776-2020; CI-IPOP-29-2014; CI-IPOP-58-2015, RESOLVE-PTDC/MED-OUT/2512/2021). Por fim, declaro que esta tese teve o suporte financeiro do projeto TeamUp4Cancer no âmbito do consórcio *Porto Comprehensive Cancer Centre* Raquel Seruca (Porto.CCC) pelo financiamento do Programa Operacional Regional Norte (NORTE 2020), através do Fundo Europeu de Desenvolvimento Regional (NORTE-01-0145-FEDER-072678), e ainda do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

DECLARAÇÃO DE HONRA

Declaro que a presente tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico. O artigo integralmente reproduzido nesta tese, no qual a candidata ao grau de Doutor é primeira autora, encontra-se publicado em regime de acesso aberto ao abrigo da Licença Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/) permitindo partilhar e adaptar os documentos para qualquer fim, mesmo que comercial, desde que seja atribuído o devido crédito ao autor, fornecida uma hiperligação para a licença, e indicando se foram feitas alterações. Mais declaro que não foram feitas alterações ao artigo já publicado.

Porto, 30 de Março de 2024



Cristiana Milhazes Gaiteiro

Em obediência ao disposto no nº 2 do Artigo 8º do Decreto-Lei nº 388/70, o autor declara que participou na conceção e na execução do trabalho experimental, bem como na interpretação dos resultados e na redação do trabalho a seguir referenciado, e que integra a tese aqui apresentada:

I. **Gaiteiro C**, Soares J, Relvas-Santos M, Peixoto A, Ferreira D, Paulo P, Brandão A, Fernandes E, Azevedo R, Palmeira C, Freitas R, Miranda A, Osório H, Prieto J, Lima L, Silva AMN, Santos LL, Ferreira JA. Glycoproteogenomics characterizes the CD44 splicing code associated with bladder cancer invasion. Theranostics. 2022 Mar 28;12(7):3150-3177. doi: 10.7150/thno.67409. PMID: 35547758; PMCID: PMC9065180. **Impact Factor in 2022: 11.6; Q1**

Para além do trabalho científico anteriormente mencionado e que integra esta tese, durante o período de doutoramento, o autor da presente tese participou nos seguintes artigos complementares:

I. Lima L*, **Gaiteiro C***, Peixoto A*, Soares J, Neves M, Santos LL, Ferreira JA. Reference Genes for Addressing Gene Expression of Bladder Cancer Cell Models under Hypoxia: A Step Towards Transcriptomic Studies. PLoS One. 2016;11(11):e0166120. **Impact Factor in 2016: 2.806; Q1**

II. Peixoto A*, Fernandes E*, **Gaiteiro C***, Lima L, Azevedo R, Soares J, Cotton S, Parreira B, Neves M, Amaro T, Tavares A, Teixeira F, Palmeira C, Rangel M, Silva AM, Reis CA, Santos LL, Oliveira MJ, Ferreira JA. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget. 2016. **Impact Factor in 2016: 5.168; Q1**

III. Cotton S*, Azevedo R*, **Gaiteiro C**, Ferreira D, Lima L, Peixoto A, Fernandes E, Neves M, Neves D, Amaro T, Cruz R, Tavares A, Rangel M, Silva AMN, Santos LL, Ferreira JA. Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumors Mol Oncol. 2017;11(8):895-912. **Impact Factor in 2017: 5.264; Q1**

IV. Lima L*, Neves M*, Oliveira MI, Dieguez L, Freitas R, Azevedo R, **Gaiteiro C**, Soares J, Ferreira D, Peixoto A, Fernandes E, Montezuma D, Tavares A, Ribeiro R, Castro A, Oliveira M, Fraga A, Reis CA, Santos LL, Ferreira JA. Sialyl-Tn identifies muscle-invasive bladder cancer basal and luminal subtypes facing decreased survival, being expressed by circulating tumor cells and metastases. Urol Oncol. 2017;35(12):675 e1- e8. **Impact Factor in 2017: 2.88; Q1**

V. Azevedo R*, Peixoto A*, **Gaiteiro C**, Fernandes E, Neves M, Lima L, Santos LL, Ferreira JA. Over forty years of bladder cancer glycobiology: Where do glycans stand facing precision oncology? Oncotarget. 2017;5(0). **Impact Factor in 2017: 5.168; Q1**

VI. Azevedo R*, Soares J*, **Gaiteiro C**, Peixoto A, Lima L, Ferreira D, Relvas-Santos M, Fernandes E, Tavares A, Cotton S, Daniel-da-Silva AL, Santos LL,

Vitorino R, Amado F, Ferreira JA. Glycan affinity magnetic nanoplatforms for urinary glycobiomarkers discovery in bladder cancer. Talanta. 2018;184:347-55. **Impact Factor in 2018: 5.043; Q1**

VII. Azevedo R, **Gaiteiro C**, Peixoto A, Relvas-Santos M, Lima L, Santos LL, Ferreira JA. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clin Proteomics. 2018;15:22. **Impact Factor in 2018: 3.01; Q1**

VIII. Neves M*, Azevedo R*, Lima L, Oliveira MI, Peixoto A, Ferreira D, Soares J, Fernandes E, **Gaiteiro C**, Palmeira C, Cotton S, Mereiter S, Campos D, Afonso LP, Ribeiro R, Fraga A, Tavares A, Mansinho H, Monteiro E, Videira PA, Freitas PP, Reis CA, Santos LL, Dieguez L, Ferreira JA. Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. N Biotechnol. 2019;49:77-87. **Impact Factor in 2019: 4.81; Q1**

IX. Madeira D, Mendonca V, Madeira C, **Gaiteiro C**, Vinagre C, Diniz MS. Molecular assessment of wild populations in the marine realm: Importance of taxonomic, seasonal and habitat patterns in environmental monitoring. Sci Total Environ. 2019;654:250-63. **Impact Factor in 2019: 6.97; Q1**

X. Fernandes E, Freitas R, Ferreira D, Soares J, Azevedo R, **Gaiteiro C**, Peixoto A, Oliveira S, Cotton S, Relvas-Santos M, Afonso LP, Palmeira C, Oliveira MJ, Ferreira R, Silva AMN, Lara Santos L, Ferreira JA. Nucleolin-Sle A Glycoforms as E-Selectin Ligands and Potentially Targetable Biomarkers at the Cell Surface of Gastric Cancer Cells. Cancers (Basel). 2020;12(4). **Impact Factor in 2020: 6.10; Q1**

XI. Peixoto A, Freitas R, Ferreira D, Relvas-Santos M, Paulo P, Cardoso M, Soares J, **Gaiteiro C**, Palmeira C, Teixeira F, Ferreira R, José Oliveira M, Silva AMN, Lara Santos L, Ferreira JA. Metabolomics, Transcriptomics and Functional Glycomics Reveals Bladder Cancer Cells Plasticity and Enhanced Aggressiveness Facing Hypoxia and Glucose Deprivation. bioRxiv. 2021:2021.02.14.431133. **No impact factor (preprint)**

XII. Peixoto A, Ferreira D, Azevedo R, Freitas R, Fernandes E, Relvas-Santos M, **Gaiteiro C**, Soares J, Cotton S, Teixeira B, Paulo P, Lima L, Palmeira C, Martins G, Oliveira MJ, Silva AMN, Santos LL, Ferreira JA. Glycoproteomics identifies HOMER3 as a potentially targetable biomarker triggered by hypoxia and

glucose deprivation in bladder cancer. J Exp Clin Cancer Res. 2021;40(1):191. **Impact Factor in 2021:** 11.38; Q1

XIII. Santos, D.A.R., **Gaiteiro C**, Santos M, Santos L, Dinis-Ribeiro M, Lima L. MicroRNA Biomarkers as Promising Tools for Early Colorectal Cancer Screening-A Comprehensive Review. Int J Mol Sci, 2023. 24(13). **Impact Factor in 2023:** 6.2; Q1

XIV. Peixoto A, Ferreira D, Miranda A, Freitas R, Relvas-Santos M, Veth TS, Brandão A, Ferreira E, Paulo P, Cardoso M, **Gaiteiro C**, Cotton S, Soares J, Lima L, Teixeira F, Ferreira R, Palmeira C, Heck AJ, Oliveira MJ, Silva AMN., Santos LL, Ferreira JA. Multilevel Plasticity and Altered Glycosylation Drive Aggressiveness in Hypoxic and Glucose-Deprived Bladder Cancer Cells. bioRxiv, 2023: p. 2023.10.21.561355. **No impact factor (preprint)**

XV. Ferreira JA, Silva AMN., Freitas R, Santos LL, Relvas-Santos M, **Gaiteiro C**, Peixoto A, Miranda A, Ferreira D, Castro F, Sarmento B, Oliveira M J. (2023). CD44 glycoepitopes and chimeric vaccine glycoconjugates for cancer therapy and synthesis methods thereof. **International Patent Number: WO/2023/033664**; Patent Application Number: PCT/PT2021/050025

XVI. Freitas, R, Miranda, A; Ferreira, D; Relvas-Santos, M, Castro, F, Ferreira, E, **Gaiteiro C**, Soares J, Cotton S, Goncalves M, Eiras M, Santos B, Palmeira C, Correia MP, Oliveira MJ, Sarmento B, Peixoto A, Santos LL, Silva AMN, and Ferreira JA; A multivalent CD44 glycoconjugate vaccine candidate for cancer immunotherapy. J Control Release. 2024; 367: 540-556. **Impact Factor in 2024:** 10.8; Q1

*** Equal contribution.**

AGRADECIMENTOS

Foi um caminho árduo chegar até aqui. Um caminho feliz e prazeroso, mas com percalços e adversidades, muitas vezes com solidão e instrospeção. No entanto, foi um caminho do qual me orgulho e com o sentimento de dever cumprido, merecedora em alcançar esta meta.

A elaboração deste trabalho de doutoramento nunca teria sido possível sem a colaboração e entusiasmo de diversas pessoas, que me são muito queridas, e das instituições que me acolheram. Por isso, um sentido OBRIGADA:

À Universidade do Porto e ao Instituto de Ciências Biomédicas Abel Salazar por me terem aceite como aluna do Programa Doutoral em Ciências Biomédicas, em particular ao Diretor, o Professor Doutor Eduardo Rocha. Agradecer também por todo o trabalho profícuo dos membros responsavéis deste programa doutoral, nomeadamente à Helena Martins e Ana Paula Pereira.

Ao Instituto Português de Oncologia do Porto Francisco Gentil (IPO-Porto) e ao seu excelente centro de investigação devo profundos agradecimentos por terem sido a minha casa nestes últimos anos, e por me terem dado as estruturas físicas e humanas para o desenvolvimento do meu trabalho. Em particular, o meu agradecimento sincero aos antecessor e atual diretores do centro de investigação, os Professores Doutores Manuel Teixeira e Carmen Jerónimo.

Ao meu orientador e mentor, Professor Doutor Lúcio de Lara Santos. Obrigada por me ter dado a oportunidade de pertencer à sua equipa e a este grupo fantástico, que é o Grupo de Patologia e Terapêutica Experimental (GPTE). Obrigada por todo o estímulo e transmissão de conhecimentos, e todas oportunidades em explorar novas áreas e projetos paralelos a esta tese. Com isso, eu amadureci e aprendi muito mais além. Serei eternamente grata, e foi um gosto tremendo ter trabalhado para si, e consigo, durante quase 9 anos.

Ao meu co-orientador, Doutor José Alexandre Ferreira, pela orientação inagualável, pelo apoio incondicional ao longo de todo este trabalho. Sem você nada disto teria sido possível alcançar. Obrigada por todos os ensinamentos e constantes desafios, e por ter contribuído para o meu crescimento enquanto investigadora.

Ao meu co-orientador, Professor Eméritus Jesús Prieto, muito obrigada pela confiança depositada no meu trabalho e por todo o apoio e orientação ao longo destes anos.

Ao Professor Doutor Carlos Palmeira, muito obrigada por toda a sua ajuda e amizade, e pelo carinho que sempre me tratou. Obrigada pelos projetos que trabalhamos juntos, pelas noitadas de trabalho, sempre bem acompanhadas por umas boas fatias de pizza.

Ao Luís Lima, agradecer por seres meu amigo e companheiro nestes últimos anos. Obrigada por todas as boas e tantas conversas (e conselhos também). Obrigada por me teres permitido ajudar e trabalhar nos teus projetos também, fez-me crescer enquanto profissional.

À Andreia, minha amiga desde o Mestrado em Bioquímica. Foste quem me abriu o caminho até ao GPTE, e me permitiu conhecer todos os bons companheiros que dele fizeram e fazem parte deste grupo. Obrigada por isso, e obrigada por toda a tua ajuda, que sabes que foi fundamental. Obrigada pelas nossas conversas e risadas.

Ao Rui e à Janine, agradecer-vos por serem meus amigos e por se lembrarem sempre de mim. Rui, foste quem eu gostei mais de ensinar aquilo que eu (pouco) sei. Empenhado, muito trabalhador, justo e companheiro. Obrigada!

À Sofia, sempre muito discreta e de poucas palavras, mas sempre acertadas quando conversávamos. Obrigada pela tua amizade e por todos os momentos divertidos que tivemos.

À Catarina. Não estávamos diariamente juntas, mas foste sempre muito especial. Obrigada pelo teu apoio, obrigada pelas noitadas a dentro que fazíamos no biotério e em citometria. Obrigada pelos húngaros, que me aqueciam a alma (e a barriga também)!

A todos os elementos que fizeram e fazem parte do GPTE, muito muito obrigada por todos os momentos, tanto de risos, como de muito trabalho. Em especial, à à Ana Tavares, à Marta, à Rita, à Rafaela, à Daniela, às Beatrizes, à Mariana, à Martina e ao Eduardo.

Sem intenções de ferir os demais, os obrigadas mais especiais são para os meus queridos amigos Elisabete e Dylan. Amizade que nasceu no GPTE, mas que extravasou para a minha vida, para sempre! Elisabete, minha *labmate*, as palavras

são insuficientes para te agradecer a amizade que construímos ao longo destes 9 anos. Foste, e és o meu suporte, minha ouvinte, conselheira e companheira. Obrigada por estares sempre presente, tanto nos bons como nos maus momentos. Obrigada pelas chamadas intermináveis contigo a ouvir, sempre com muita paciência, as minhas tristezas, mas também as minhas alegrias. Mesmo quando estavas longe, fazias-te sempre perto, e isso nunca vou esquecer. Quero-te para sempre na minha vida, minha amiga querida. Dylan, meu querido, talentoso e brilhante amigo. O meu obrigada nunca será suficiente. Sempre ao meu lado sem me deixar cair. Sempre ao meu lado em todos os momentos. Obrigada por todos os lanchinhos e chocolatinhos, pelas nossas maravilhosas conversas, pelas nossas saídas sempre tão divertidas, pelas nossas risadas. Obrigada por partilhares comigo sempre o teu bacalhau à brás, era um gesto de muito carinho, e aquecia o meu coração. Obrigada por toda a tua ajuda, que foi essencial. Que a vida nos permita ser amigos para sempre, porque nunca te vou esquecer. Adoro-vos aos dois meus fofos!!

À minha amiga Flávia, agradecer por toda a amizade desde a nossa licenciatura em 2009. Amizade que atravessou todos estes anos (e cidades também) e que perdura até hoje. Minha fiel companheira, que nunca se esqueceu de mim. Obrigada por estares sempre presente, mesmo com as nossas vidas sempre a mil. Obrigada por todas as nossas conversas, desabafos e choros. Adoro-te para sempre!

Aos meus fiéis e muito queridos amigos, de sempre e para sempre: Cláudia, Mellissa, Alex e Zé. Obrigada por me darem momentos leves e alegres que me faziam esquecer, por momentos, as fases menos boas deste caminho. Obrigada por estarem sempre presentes. Obrigada pela vossa amizade. Adoro-vos e quero-vos para sempre na minha vida!

Aos meus sogros, Maria José e Carlos Alberto. Obrigada por todo o vosso apoio, principalmente por cuidarem do meu Vasco enquanto eu ia trabalhar. Sem vocês isto teria sido muito mais dificil. Obrigada! Obrigada também à minha cunhada Carla, e as minhas queridas sobrinhas Bruna e Gabriela, que tanto mimo me deram.

Ao meu irmão Pedro, homem de poucas palavras, mas que eu sei que tem muito orgulho em todo o meu percurso. Cuidou sempre de mim. És o melhor irmão que eu podia ter. Amo-te! Obrigada por me teres dado os meus sobrinhos, Rodrigo e Francisca, que mesmo sem saberem, me davam amor e sorrisos quando eu mais precisava. Um obrigada também à minha cunhada Tina por todo o carinho e amor.

Aos meus queridos e amados pais, Arminda e Manuel. Sem eles nada disto teria sido possivel alcançar. Impulsionadores dos meus estudos, da minha procura por mais e melhor. Sempre me permitiram voar e sonhar. Sempre na retaguarda quando tudo parecia difícil. O meu maior apoio, os meus maiores apoiantes! Obrigada pelo vosso amor, pela vossa paciência, pelo vosso esforço para me conseguirem deixar alcançar este doutoramento. O meu obrigada nunca será suficiente para retribuir tudo o que já me deram. Amo-vos muito.

Ao meu eterno amor, desde os 16 anos, Bruno. Obrigada por teres sido o meu maior suporte ao longo deste doutoramento. Sem ti, isto teria sido muito duro. Obrigada pela tua compreensão, pelos teus conselhos, pelo teu companheirismo, pela tua paciência, por estares sempre na primeira fila a apoiar-me. Obrigada por não desistires de mim e não me deixares desistir. Obrigada por compreenderes todas as minhas ausências. Obrigada por ouvires os meus desaforos e dissabores. Obrigada por ouvires as minhas apresentações, mesmo não percebendo nada. Não podia ter melhor companheiro ao meu lado, certa que escolhi o melhor pai para o nosso maior tesouro. Amo-te muito!

Por último e o mais importante de todos, agradecer ao meu filho Vasco, que mesmo sem saber, foi ele a minha força motriz nestes últimos 2 anos. Filho, ainda não sabes ler, mas um dia quero muito que leias estas palavras. Ter-te gerado foi a melhor, e mais bem feita ,coisa que fiz na minha vida! Foste o projeto mais bem escrito de todos. Quero que saibas, que foi por tua causa que eu consegui terminar este doutoramento, porque quero ser um exemplo de resiliência, determinação e superação para ti. Quero mostrar-te, que mesmo quando achamos que não somos capazes, conseguimos arranjar sempre uma maneira de alcançar as metas a que nos propomos. Amo-te eternamente filho. Já és o meu maior orgulho, e tenho a certeza que me darás muitas muitas alegrias.

RESUMO

O cancro da bexiga (CB) destaca-se como um dos tumores geniturinários mais prevalentes e mortais. O seu tratamento é complexo devido à heterogeneidade característica desses tumores e à escassez de alvos terapêuticos para uma abordagem de medicina de precisão. O CD44, uma proteína altamente glicosilada, desempenha um papel amplo na agressividade e na progressão do CB. No entanto, a sua diversidade molecular resultante de *splicing* alternativo e da glicosilação resulta numa vasta panóplia de glicoproteoformas com papéis funcionais distintos. As interpretações erróneas, bem como a ambiguidade da nomenclatura, aliadas à falta de ferramentas para uma caraterização molecular precisa, têm conduzido à informação científica contraditória, atrasando assim o progresso clínico.

Para ultrapassar estes obstáculos, utilizou-se neste trabalho uma abordagem glicoproteogenómica avançada para analisar uma grande coorte de doentes com CB, com o objetivo de identificar assinaturas específicas de CD44 clinicamente relevantes. Foi empregue uma série retrospetiva de tumores da bexiga, assim como um painel de linhas celulares de CB, para avaliar os inúmeros transcritos de CD44 através da técnica de Reação em Cadeia da Polimerase em tempo real. A análise *TCGA* de casos músculo-invasivos foi integrada com resultados gerados internamente para inferir a relevância clinicopatológica das isoformas de CD44. A *O*-glicómica das células de CB permitiu uma abordagem glicoproteómica orientada das glicoproteoformas de CD44. O CD44 foi isolado de modelos celulares e de tecidos tumorais para identificação das glicoproteoformas de CD44 por espetrometria de massa *nano-LC-tandem* e ferramentas bioinformáticas. Conjuntos de dados personalizados de transcriptómica e glicómica foram integrados e utilizados para anotação de proteínas numa abordagem inovadora de glicoproteogenómica. A co-localização do CD44 e dos *O*-glicanos curtos foi avaliada através das técnicas de Ensaio de ligação de proximidade, imunohistoquímica e imunofluorescência dupla. Adicionalmente, foi determinado o impacto funcional da expressão de CD44 nas linhas celulares de CB, por meio de ensaios de proliferação e invasão, bem como em *arrays* de anticorpos para identificar alterações em vias de sinalização associadas ao CB.

A análise transcriptómica revelou uma heterogeneidade significativa nas isoformas de CD44, com a isoforma de *splicing* padrão CD44 (CD44s) associada tanto à invasão como a um mau prognóstico. Os péptidos de diagnóstico e as

glicoformas associadas de CD44s foram inequivocamente identificados em linhas celulares de CB, células de glico-engenharia e tecidos tumorais de bexiga através da abordagem inovadora de glicoproteogenómica. O silenciamento do CD44 em células de CB e em modelos celulares de glico-engenharia resultou numa redução significativa da capacidade de invasão das mesmas, especialmente no modelo com expressão aumentada do glicano truncado Sialil-Tn (STn), evidenciando assim a relevância funcional da *O*-glicosilação curta do CD44.

Em suma, o CD44s destacou-se como um biomarcador de prognóstico, com a presença de *O*-glicanos truncados no CD44s identificada como uma assinatura molecular promissora para aplicações clínicas direcionadas. Além disso, este trabalho proporcionou também um guia abrangente para investigar o papel funcional do CD44 tanto em condições saudáveis como em situações patológicas.

ABSTRACT

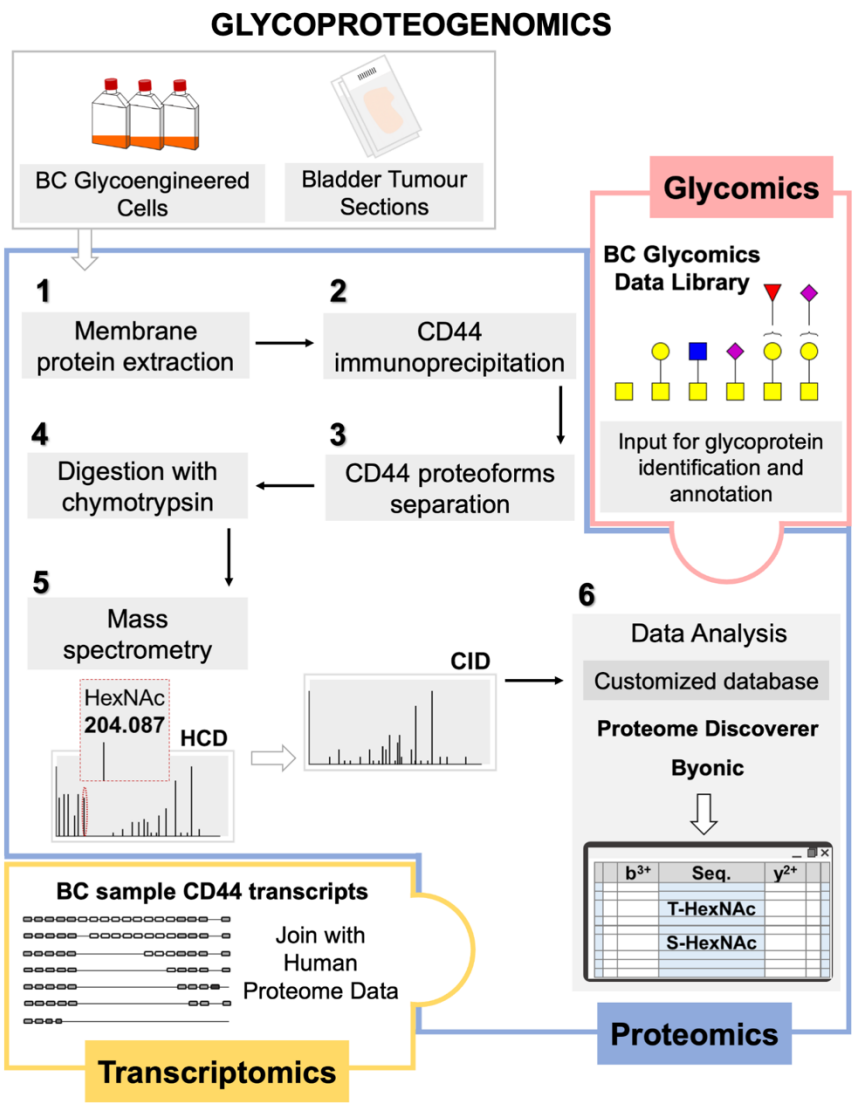
Bladder cancer (BlCa) is one of the most common and deadliest genitourinary tumors Its management is hampered by tumor heterogeneity and the lack of therapeutic targets for precision medicine. Cell-surface CD44, a heavily glycosylated protein, is widely implicated in BlCa aggressiveness and stemness. However, its molecular diversity resulting from alternative splicing and glycosylation yields a vast panoply of glycoproteoforms, potentially with distinct functional roles. Misinterpretation as well as ambiguity of nomenclature and the lack of tools for accurate molecular characterization has led to conflicting reports, delaying clinical progress.

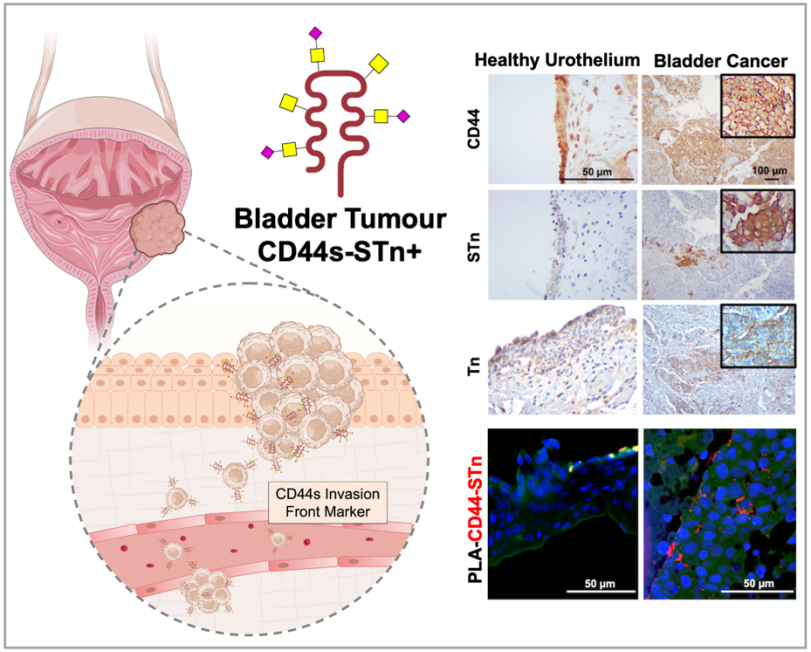
To address these hurdles, an advanced glycoproteogenomics approach was employed in this work to interrogate a large cohort of BlCa patients, aiming to identify clinically relevant CD44 splicing signatures. A retrospective series of bladder tumors as well as a panel of BlCa cell lines were used to evaluate CD44 transcripts through real-time PCR. TCGA analysis of muscle-invasive cases was integrated with *in-house* generated results to infer the clinicopathological relevance of CD44 isoforms. BlCa cells *O*-glycomics has allowed the targeted glycoproteomics of CD44 glycoproteoforms. CD44 was isolated from cell models and tumor tissues for identification of CD44 glycoproteoforms by nano-LC-tandem mass spectrometry and bioinformatics tools. Transcriptomics and glycomics custom datasets were integrated and used for protein annotation in an innovative glycoproteogenomics approach. The co-localization of CD44 with short *O*-glycans was assessed using proximity ligation assay, immunohistochemistry and double-immunofluorescence. Additionally, the functional impact of CD44 expression on BlCa cell lines was determined through proliferation and invasion assays, as well as Phospho-Kinase antibody arrays.

Transcriptomic analysis revealed significant heterogeneity in CD44 isoforms, with CD44 standard splicing isoform (CD44s) being associated with both muscle invasion and poor prognosis. Diagnostic peptides and associated glycoforms of CD44s have been unequivocally identified in BlCa cell lines, glycoengineered cells, and tumor tissues through glycoproteogenomics. Knocking down CD44 in BlCa cells and glycoengineered cell models significantly reduced invasion capacity, particularly in an Sialyl-Tn-dependent manner, shedding a light on the functional relevance of CD44 short *O*-glycosylation.

In summary, CD44s has emerged as a prognostic biomarker, with CD44s carrying truncated *O*-glycans identified as promising molecular signatures for targeted clinical applications. This study also offers a comprehensive roadmap for exploring the functional role of CD44 in both health and disease.

GRAPHICAL ABSTRACT





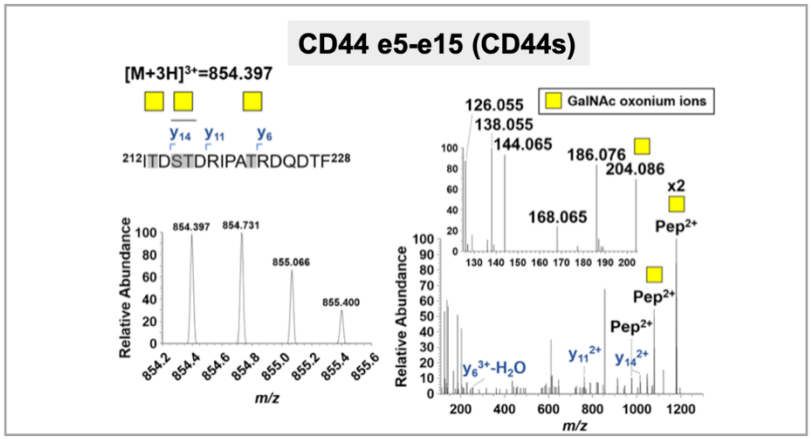


TABLE OF CONTENTS

NOTA PRELIMINAR ..... **v**

DECLARAÇÃO DE HONRA ..... **vii**

AGRADECIMENTOS ..... **xv**

RESUMO ..... **xix**

ABSTRACT ..... **xxi**

GRAPHICAL ABSTRACT ..... **xxiii**

LIST OF FIGURES ..... **xxvii**

LIST OF TABLES ..... **xxix**

ABBREVIATIONS & ACRONYMS ..... **xxxi**

INTRODUCTION **36**

AIMS & STUDY OUTLINE **103**

CHAPTER 1: GLYCOPROTEOGENOMICS CHARACTERIZES THE CD44 SPLICING CODE ASSOCIATED WITH BLADDER CANCER INVASION. **109**

Abstract 110

Introduction 112

Material and Methods 113

Results and discussion 125

Conclusions 155

References 160

Supplementary figures and tables 170

CONCLUDING REMARKS & FUTURE PERSPECTIVES **195**

LIST OF FIGURES

INTRODUCTION

Figure 1. Staging of Bladder Cancer 40

Figure 2. Schematic representation illustrating *N*- and *O*-glycan biosynthesis pathways..... 49

Figure 3. Terminal Lewis and sialylated Lewis structures on glycans..... 52

Figure 4. Schematic representation of experimentally validated human CD44 pre-mRNA and its isoforms..... 59

CHAPTER 1- GLYCOPROTEOGENOMICS CHARACTERIZES THE CD44 SPLICING CODE ASSOCIATED WITH BLADDER CANCER INVASION

Figure 1. CD44s is increased in more aggressive bladder tumors in comparison to superficial lesions, is not present in the healthy urothelium and shows limited expression in other healthy human tumors..... 129

Figure 2. CD44-Tn and STn glycoproteoforms present high cancer specificity..... 133

Figure 3. Highly invasive T24 bladder cancer cells express high levels of CD44s..... 137

Figure 4. Glycoproteogenomics, building on RNAseq-customized databases and glycomics for protein annotation, enables CD44 proteoforms identification in bladder cancer cell lines and tumors..... 145

Figure 5. CD44 silencing inhibits proliferation, invasion, and activation of relevant proteins in MAPK/ERK, and PI3K/AKT pathways in T24 CD44s^{high} cells..... 149

Figure 6. Glycoengineered T24 cells express CD44s-Tn/STn glycoproteoforms in mimicry of human tumors..... 151

Figure 7. CD44 glycosylation in CD44s-enriched cells shapes its functional contribution to proliferation and invasion..... 154

Figure S1. A) Schematic representation of human CD44 pre-mRNA and experimentally confirmed isoforms resulting from alternative splicing and extensively explored in this study. **B)** Amino acid sequence evidencing potential *O*- and *N*-glycosites..... 170

Figure S2. CD44 variants expression according to the histological subtypes of TCGA muscle invasive bladder tumors…….. 171
Figure S3. CD44 splice variants expression in healthy human tissues……. 172
Figure S4. Reverse Transcriptase-Polymerase Chain Reaction and Sanger Sequencing for the *CD44* gene using random primer (RP) and e16 specific primer amplification……. 174
Figure S5. Glycoproteogenomics workflow for identification of CD44 glycoproteoforms……. 176
Figure S6. CID spectra for CD44s glycopeptides carrying *O*-glycans identified in T24 cells…….. 177
Figure S7. A) CD44-T and **B)** Tn glycopeptides belonging to the conserved extracellular domain in different isoforms identified in muscle invasive bladder tumors by nanoLC-HCD-MS/MS……. 178
Figure S8. The human phospho-receptor tyrosine kinase (Phospho-RTK) array coordinates…….. 179
Figure S9. The human phospho-kinase array coordinates……. 181
Figure S10. Western Blots show co-expression of CD44 and short-chain *O*-glycans, namely STn and Tn, in glycoengineered T24 cells…….. 183
Figure S11. CD44-Tn glycopeptides belonging to the conserved extracellular domain identified in T24 *C1GALT1*KO cells by nanoLC-HCD-MS/MS……. 184

LIST OF TABLES

CHAPTER 1- GLYCOPROTEOGENOMICS CHARACTERIZES THE CD44 SPLICING CODE ASSOCIATED WITH BLADDER CANCER INVASION

Table 1. Peptides corresponding to CD44 exon and exon-exon junctions identified by SDS-PAGE-nanoLC-MS/MS..... 143

Table S1. Tailored-made Taqman Gene Expression Assay ID's to specifically detect mRNA encoding for total CD44 and 5 splicing variants..... 185

Table S2. List of antibodies and lectins used in this manuscript..... 186

Table S3. CD44 isoforms identified in 5637 and T24 cells by RNAseq..... 187

Table S4. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for 5637 cells..... 188

Table S5. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for T24 cell line..... 189

Table S6. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for CD44shigh MIBC showing areas of CD44 and STn co-localization..... 190

Table S7. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for T24 C1GALT1 KO cell model..... 191

ABBREVIATIONS & ACRONYMS

ATG7: Autophagy-related gene 7

B2M: β-2-microglobulin

BCG: *Bacillus Calmette-Guérin*

BlCa: bladder cancer

BSA: Bovine Serum Albumin

C1GalT1: core 1 synthase or N-Acetylgalactosamine 3-Beta-Galactosyltransferase 1

C2GnT: Core 2 β1,6-*N*-acetylglucosaminyltransferase

CD44: cluster of differentiation 44

CD44s: CD44 standard splicing isoform

CID: collision induced fragmentation

CIS: Carcinoma *in situ*

COSMC: Core 1 β1-3-galactosyltransferase-specific chaperone 1

CR: Complete responses

CSC: cancer stem cells

CT: computed tomography

CTCs: Circulating tumor cells

DAPI: 4',6'-diamidino-2-phenylindole dihydrochloride

DFS: Disease-free survival

DTT: 1,4-dithiothreitol

EGFR: epidermal growth factor receptor

EGFR: Epidermal growth factor receptor

EMT: epithelial-to-mesenchymal transition

ER: Endoplasmic reticulum

FBS: fetal bovine serum

FDA: Food and Drug Administration

FFPE: formalin-fixed paraffin-embedded

FGFR3: Fibroblast growth factor 3

Fuc: Fucose

FUT: fucosyltransferases

GA: Golgi apparatus

Gal: galactose

GalNAc*: O-N*-acetylgalactosamine

GlcNAc: *N*-acetylglucosamine

GSTM: glutathione S-transferase mu 1
GTEx: genotype-tissue expression
GTs: Glycosyltransferases
HA: Hyaluronic acid
HABD: HA-binding domain
HPRT: Hypoxanthine-guanine phosphoribosyltransferase
HRP: horseradish peroxidase
ICIs: Immune checkpoint inhibitors
IF: immunofluorescence
IGF-1R: insulin-like growth factor 1 receptor
IHQ: Immunohistochemistry
IL-6: Interleukin-6
IP: immunoprecipitation
IR: insulin receptor
KI: knock-in;
KO: knock-out
KRT: Keratin
Le: Lewis
LVI: Lymphovascular invasion
MALT: mucosa-associated lymphoid tissue
MIBC: muscle invasive bladder cancer
MRI: magnetic resonance imaging
MUC1: Mucin 1
MUC16: Mucin 16
NAC: Neoadjuvant chemotherapy
nanoLC-MS/MS: nano liquid chromatography mass spectrometry
NAT2: N-acetyltransferase 2 slow-acetylator
NCE: normalized collision energy
Neu5Ac: sialic acid
NMIBC: non-muscle invasive bladder cancer
OPN: Osteopontin
OS: Overall survival
OST: Oligosaccharyltransferase
PD-1: Programmed cell death-1 protein
PD-L1: Programmed cell death ligand 1
PFA: paraformaldehyde

PLA: proximity ligation assay

PTMs: Post-Translational Modifications

PUNLMP: papillary urothelial neoplasm of low malignant potential

RB1: Retinoblastoma Transcriptional Corepressor 1

RC: Radical cystectomy

RFS: Recurrence-free survival

RT-PCR: real time-polymerase chain reaction

RT: room temperature

RTK: receptor tyrosine kinase

SCC: Squamous cell carcinoma

SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SDS: sodium dodecyl sulphate

SiRNA: small interfering RNA

SLe: Sialyl-Lewis

ST: sialyl-T

ST6GalNAc: α-N-acetylgalactosaminide α-2,6-sialyltransferase

STn: sialyl-Tn

T: Thomsen–Friedenreich antigen

TCC: Transitional cell carcinoma

TCGA: The Cancer Genome Atlas Program

TP53: Tumor Protein P53

TUR: transurethral resection

TURBT: transurethral resection of the bladder tumor

UCC: urothelial cell carcinoma

VVA*: vicia villosa* lectin

WB: western blotting

β3GnT: β1,3-*N*-acetylglucosaminyltransferase

INTRODUCTION

INTRODUCTION

1. Overview of Bladder Cancer

1.1. Epidemiology and main risk factors

Bladder Cancer (BlCa) stands as the most common malignant disease of the urinary tract and ranks as the tenth most frequent cancer worldwide [1]. Globally, approximately 573,000 new cases of BlCa are diagnosed each year, resulting in 213,000 deaths, making it the sixth most prevalent cancer among men and the seventeenth most incident in women [1]. In Portugal, BlCa is among the most prevalent cancers with 2,608 new cases and 1,210 deaths reported in 2020 [2].

The incidence and pathophysiology of BlCa are significantly impacted by different risk factors, namely genetic predispositions and exposures to several carcinogens. Genetic risks are notably associated with the N-acetyltransferase 2 (*NAT2*) slow-acetylator and glutathione S-transferase mu 1 (*GSTM1*)–null genotypes, which play a significant influence on BlCa development, especially by enhancing susceptibility to other risk factors [3-6]. Remarkably, tobacco smoke, the major contributor to BlCa development, is responsible for 50–65% of cases in males and 20–30% in female patients [7, 8]. One of the associated reasons is the fact that tobacco is a rich source of aromatic amines and N-Nitroso compounds known for their carcinogenic nature [3]. Occupational exposures to substances like aromatic amines, benzene dyes and factory chemicals, as well as environmental factors such as nitrates and arsenic compounds, further increase BlCa risk [3]. Furthermore, the global incidence of this neoplasia is intimately linked with country-specific socioeconomic development [9]. Particularly, in developing countries, urogenital *Schistosoma haematobium* infection and chronic inflammation in the urothelium also severely contribute to BlCa occurrence [10]. Additional factors such as exposure to ionizing pelvic radiation (e.g. for prostate, cervical and rectal cancer treatment), use of cyclophosphamide (a cytotoxic chemotherapy agent), and obesity are associated with a higher risk of BlCa [11-13].

Even though there are well-known risk factors associated to BlCa development, standardizing the analysis of BlCa epidemiological data is warranted to fully disclose the impact of environmental and socioeconomic factors on BlCa-associated mortality.

1.2. Clinical presentation and diagnosis

Gross or visible hematuria is the most frequent BlCa-presenting symptom, which usually appears promptly and painlessly [14]. Less commonly, some patients can also present microscopic hematuria (≥ 3 red blood cells per high-power field) on urinalysis, unexplained urine frequency, urgency and dysuria, irritative voiding symptoms [15, 16]. More advanced tumors are commonly associated with pelvic pain and symptoms related to urinary tract obstruction [17]. Of note, about 10-20% of patients with macroscopic hematuria are diagnosed with BlCa while only 2-5% of patients with microhematuria present a malignant bladder tumor [16].

The current guideline work-up in patients with gross or high-risk microscopic hematuria includes cystoscopy, upper tract imaging, and urine cytology [11]. Cystoscopy is the standard gold method during the initial diagnosis which is an outpatient endoscopy performed with a flexible cystoscope and local anesthesia [11]. Any atypical finding such as reddish flat, papillary, or solid lesions requires histological evaluation and can be obtained by transurethral biopsy or resection. Despite being a specific and sensitive test, conventional cystoscopy is not effective in discerning Carcinoma *in situ* (CIS)/papillary Ta/T1 from normal bladder tissue [16]. Therefore, in recent years, cystoscopy detection was enhanced by narrow-band imaging or blue light cystoscopy (or photodynamic diagnosis) improving the differentiation of bladder malignant lesions from normal tissue [18]. Narrow-band imaging provides increased contrast in blood vessels architecture, augmenting the detection rate in approximately 20% on per-lesion basis [16, 17, 19]. Regarding blue light cystoscopy, this takes advantage of the increased metabolic activity (by the administration of hexaminolevulinate hydrochloride) that occurs in cells, improving the detection up to 14% of papillary Ta/T1 lesions and 40% of CIS missed by conventional cystoscopy [16-18, 20]. Concerning upper urinary tract (i.e. renal collecting system and the ureter) imaging, computed tomography (CT) or magnetic resonance imaging (MRI) urography plays an important role in the evaluation of patients with hematuria, however it constitutes a minor part of BlCa diagnosis [17]. Nevertheless, MRI and CT could improve detection and diagnosis accuracy, and the latter is mandatory for staging and assessing the risk of distant spreading, when patients are diagnosed with MIBC [11, 16]. In addition, ultrasound (US) permits the evaluation of renal and/or intraluminal bladder masses, and could be useful for patients with low-risk microscopic hematuria [11, 21, 22]. Lastly, non-invasive urinary cytology, the microscopic

inspection of voided urine or bladder-washing specimens for exfoliated tumor cells, is used as adjunct method to catch missed cancer cells from a malignant injury [11, 16]. Despite its high specificity, this method is limited by its low sensitivity for BlCa detection, being more sensitive only for high-grade lesions and CIS [11, 23]. Thereupon, in order to augment its sensitivity, urine biomarkers have been implemented to complement urinary cytology. Currently, there are several urinary-based tumor biomarkers tests approved by the Food and Drug Administration (FDA) to improve BlCa diagnosis, prognosis, and post-treatment surveillance [24]. These molecular assays have been mainly developed based on differential expression of tumor-related proteins, DNA, RNA, or cellular markers [18]. Bladder tumor-associated antigen (BTA) stat, BTA TRAK, Cx Bladder, Nuclear Matrix Protein (NMP) 22 ELISA, and UroVysion are examples of promising tools for BlCa detection [24, 25]. Some of these molecular assays might be useful adjuncts in the selection of patients for immunotherapies or to stratify patients for enrolment in clinical trials [18]. However, none of these markers can replace cystoscopy in routine practice once they lack accuracy and suffer from high false positive rates [11, 24].

Thus, the new era of diagnosis will go through the development of non-invasive, highly accurate and sensitive, and cost-effective tests based on biomarkers from blood or urine specimens [26]. Notably, sequencing and gene expression studies as well as innovative glycoproteogenomics analysis have been increasingly explored, allowing the discovery of numerous proteins, RNA or DNA markers of BlCa [26, 27]. This offers the opportunity to develop useful biomarker panels for early diagnosis, patient stratification, prognosis, and patient therapeutic management.

1.3. Bladder cancer classification and pathophysiology

Bladder Cancer is defined as a carcinoma of the urothelial “umbrella” cells that line the lumen of the urinary bladder [16]. It comprehends three main types of bladder cancerous lesions: squamous cell carcinoma (SCC), adenocarcinoma, and urothelial carcinoma (UCC, also called Transitional cell carcinoma [TCC]), accounting for approximately 90% of cases [17, 28, 29].

Bladder urothelial carcinomas are classified as low-grade and high-grade lesions according to histomorphologic features defined by the World Health Organization (WHO) classification [11, 30]. Namely, papillary urothelial neoplasm of low malignant potential (PUNLMP), low-grade (LG) papillary urothelial carcinoma, and high-grade (HG) papillary urothelial carcinoma [11, 30]. In turn, BlCa stages are

assigned by the tumor, lymph node, metastasis (TMN) system according to tumor invasion status through the bladder tissue layers (Figure 1) [31]. Carcinoma *in situ* (Stage Tis, Figure 1) is a flat non-invasive and high-grade lesion, located in the urothelium. These lesions often recurrent and prone to progression [11]. Tumors confined to the urothelium (Stage Ta) and rarely invading the lamina propria (Stage T1) are considered non-muscle invasive bladder cancer (NMIBC) [32]. Muscle-invasive bladder cancers (MIBC; Stages T2 to T4) can invade the *muscularis propria* (Stage T2), perivesical tissue (Stage T3), extravesical layers into surrounding tissues or organs outside the bladder (Stage T4)[32]. Approximately 25% of patients with BlCa present tumors >T2 [7].

Bladder tumors can be further classified into papillary and non-papillary based on their genetic alterations [18]. Briefly, the most notable alterations for papillary tumors are the activating mutations in the Fibroblast growth factor 3 (*FGFR3*) gene [18, 33]. In contrast, non-papillary and invasive neoplasms commonly exhibit inactivating mutations in Tumor Protein P53 (*TP53*) and retinoblastoma Transcriptional Corepressor 1 (*RB1*) genes [34].

Recently, bladder neoplasms were molecularly categorized into basal and luminal subtypes, presenting different clinical behaviors and responses to therapy [18, 35]. Luminal subtypes, enriched with papillary signatures, overexpress keratin (*KRT)20* as well as *FOXA1*, and *GATA3* genes, while basal-like subtypes (squamous features) present high expression of *KRT5* and/or *KRT14* and/or *KRT6A* [18, 36, 37]. Building on this preliminary classification, a new consensus set of six molecular classes of MIBC was proposed in late 2020 [38]. Namely, luminal papillary, luminal non-specified, luminal unstable, stroma-rich, basal/squamous, and neuroendocrine-like. These consensus classification accounts for different oncogenic mechanisms, degree of immune and stromal infiltration of the tumor, histological and clinical features, as well as different clinical outcomes [38]. Even though protein immunohistochemistry and transcriptomic signatures have been demonstrating potential in stratifying MIBC patients, current treatment decisions are not based on any predictive molecular information. This is mainly driven by the fact that molecular subtypes of MIBC do not fully correlate with the morphological characteristics of the tumors, and it fails to entirely reflect the molecular heterogeneity of bladder tumors [39]. However, pre-clinical and clinical research is ongoing to refine current molecular panels and several clinical trials are currently exploiting predictive and prognostic markers [40, 41].

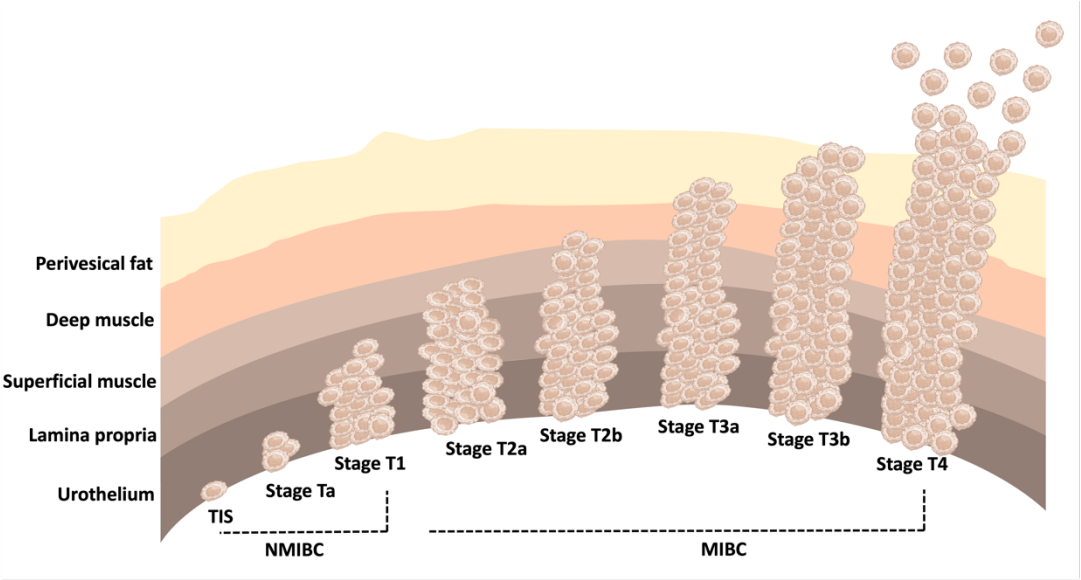


Figure 1- Staging of Bladder Cancer. The staging system for urothelial carcinoma, according to the Tumor, Node, Metastasis (TNM) classification, involves categorizing bladder cancer based on the extent of tumor infiltration into bladder tissue layers. Carcinoma in situ (Cis), indicated as Tis in the TNM system, denotes a flat, high-grade lesion confined to the urothelium. Ta papillary tumors are confined to the mucosa and T1 have invaded the lamina propria. Stage 2 (T2) tumors infiltrate the muscle layer, either superficially (T2a) or deeply (T2b). Stage 3 (T3) tumors extend beyond the muscularis propria into perivesical fat (T3a invasion is microscopic, T3b is macroscopic). T4 tumors invade adjacent structures such as the prostate, uterus, vagina, or pelvic or abdominal walls. Tumors with stages Tis, Ta and T1 are classified as non-muscle invasive bladder cancer (NMIBC), while neoplasms >T2 are considered muscle invasive bladder cancer (MIBC).

1.4. Bladder Cancer management

Currently, all newly diagnosed BlCa patients undergo transurethral resection of bladder tumor (TURBT) for diagnosis, staging and either total or partial resection of the mass [42]. Non-Muscle Invasive BlCa cases are typically managed with TURBT and intravesical instillation with chemo- or immunotherapeutic agents, while MIBCs are treated with cystectomy, with or without chemotherapy [11, 43-45]. Additionally, radiation and targeted immunotherapy have been alternative treatments when conventional therapies fail [46, 47]. The management of BlCa is intricately tied to factors such as disease stage and grade, the presence of metastases, overall health, comorbidities, age, and patient preferences [48].

1.4.1. Non-muscle invasive bladder cancer management

Transurethral resection is an endoscopic technique performed with a resectoscope through the urethra, allowing the removal of all visible tumors with

an adequate margin and depth, including the muscularis propria [16, 18]. During the initial TURBT, NMIBC patients are pathologically stratified into three risk groups (i.e. low, intermediate and high-risk groups) based on multifocality, tumor size, recurrence rate, stage, grade, lymphovascular invasion (LVI), and the presence or absence of CIS [49-52]. For low-risk NMIBCs (single low-grade (Ta) tumor <3cm), initial transurethral resection is curative, showing 5-year relapse- and progression-free survivals of 43% and 93%, respectively [53]. However, surveillance protocols involving cystoscopy are recommended due to the risk of new recurrences [17]. Accurate staging through initial TURBT is crucial in BlCa to determine the appropriate treatment based on the risk group, and thereby increase recurrence-free survival (RFS) [54, 55]. Therefore, for patients with a known incomplete tumor resection or neoplasia invading the lamina propria (T1), a second TURBT is recommended [56]. In addition, high- and intermediate-risk NMIBC patients are advised to undergo adjuvant intravesical therapy, namely chemotherapy or *Bacillus Calmette-Guérin* (BCG) immunotherapy [57, 58].

After TURBT, a single dose of intravesical instillation with chemotherapy should be administered within 24 hours to target free-floating/circulating tumor cells (CTCs) and eliminate residual BlCa cells at the resection site [11, 59]. Chemotherapeutic agents like mitomycin or doxorubicin in the United States, as well as epirubicin or pirarubicin in Europe, can reduce the recurrence rate by 40% at 1 year and 15% at 5 years [60, 61]. Treatment with mitomycin, the first-line agent, may lead to irritative voiding symptoms and, rarely, bladder necrosis [62]. Therefore, gemcitabine has emerged as a viable alternative with similar efficacy and improved tolerance at a lower cost [63]. Despite these advancements, there is a need for upgraded therapies for NMIBC to achieve better efficacy and/or tolerability [64]. Multiagent intravesical chemotherapy regimens have been implemented and tested in recent years [65]. Notably, intravesical combinations include gemcitabine + docetaxel and cabazitaxel + gemcitabine + cisplatin, demonstrating complete responses (CR) of 66% and 88% (seven of eight patients), respectively [65-67]. Furthermore, gemcitabine + docetaxel treatment exhibits a disease-free survival (DFS) rate of 54% at 1 year and 34% at 2 years [66].

In 1990, the BCG vaccine, originally design for tuberculosis, became the first FDA-approved immunotherapeutic treatment for BlCa [68]. This vaccine acts as a potent immunomodulator, recruiting cytotoxic T-cells and other immune cells capable of inducing apoptosis in BlCa cells [68]. Several studies have confirmed the efficacy of BCG instillation, decreasing both recurrence and progression rates,

highlighting the superiority of TURBT followed by BCG over TURBT alone, particularly in high-risk disease [69-72]. Despite being the first-line treatment for NMIBC, 30%-40% of patients fail to respond, and adverse side-effects, such as cystitis, hypersensitivity pneumonitis, granulomatous lesions of prostate and testicles, and other emergent symptoms, have been reported [73, 74]. For BCG non-responder patients, gemcitabine-docetaxel treatment has demonstrated RFS rates of 60% at 1 year and 46% at 2 years [75]. Another FDA-approved alternative is valrubicin for BCG-unresponsive patients with CIS, yet only 20% of patients remain tumor-free at 12 months, and less than 10% beyond 12 months [76]. In a recent phase II clinical trial based on pembrolizumab, an immune checkpoint inhibitor, a CR rate of 41% was observed and maintained for at least 12 months [77]. Other novel intravesical therapies, such as nadoferagene firadenovec (a nonreplicating recombinant adenovirus carrying the *INFα2b* gene), CG0070 (an oncocytic adenovirus that targets BlCa cells), and ALT-803 (an IL-15 superagonistic), have shown favorable outcomes in patients with BCG-unresponsive disease [78-81]. Despite these promising results, the development of targeted predictive tools and treatments for a broad spectrum of NMIBCs, where BCG fails, remains a critical clinical research priority to prevent the morbidity and life-altering procedure of cystectomy [82].

1.4.2. Muscle invasive bladder cancer management

Radical cystectomy (RC) stands as the mainstay treatment for MIBC patients, preceded by neoadjuvant platinum-based chemotherapy [83]. It is also considered for NMIBC patients who do not respond to the above-mentioned intravesical therapies [84]. Early RC may be pondered to T1 NMIBC with lamina propria invasion and for patients assigned as high-risk on TURBT [85, 86]. The procedure typically entails bladder removal, lymphadenectomy (partial or extended), prostatectomy in men, hysterectomy, partial resection of the vagina and urethra in women, followed by urinary diversion to reroute urine flow [16]. Post-surgery, survival outcomes are directly linked to final pathological staging [87]. The 10-year RFS for patients without lymphadenectomy is 86% for T0 tumors, 76% for T1-pT3a, 61% for T3b, and 45% for T4 [88]. Notably, a significant 34% drop in RFS is observed when lymph nodes are dissected during RC, regardless of the final tumor stage [88]. In a prospective study, comparing standard versus extended lymph node dissection during RC, a 5-year DFS of 54.7% and 66.6% was observed, respectively [89].

However, in the lymph node-positive group, extended lymphadenectomy was associated with better DFS (48%) compared to the standard lymphadenectomy approach (28.2%) [89]. Controversy persists regarding extended lymphadenectomy, with many surgeons advocating for a minimum standard dissection, necessitating further comparative studies [90, 91]. While RC yields favorable clinical outcomes, a significant proportion of patients may experience recurrences or postoperative complications, leading to death, denoting that BlCa remains a lethal disease [16]. Some studies showed 90-day complication rates in approximately two-thirds of patients, and mortality rates of 1.5%-2% at 30 days post-surgery [92, 93]. The most common complications associated with RC include gastrointestinal (29%), infectious (25%), wound-related (15%), and genitourinary (11%) detrimental effects [92]. Muscle-Invasive BlCa patients face tremendous challenges post-RC that impact their quality of life, overall well-being, sex life, and urinary and bowel function [16]. For patients unwilling or unfit for surgery, bladder-sparing options have been explored, such as maximal TURBT followed by radiosensitizing chemoradiotherapy, known as “trimodal therapy” (TMT) [94]. Trimodal therapy often employs chemotherapeutic agent combinations like cisplatin with fluorouracil or paclitaxel, and fluorouracil plus mitomycin C, or single agents (e.g. cisplatin and gemcitabine) [17]. This approach has shown favorable 5-year cancer-specific survival rates of 50%-82%, but 25%-30% of patients fail to respond, requiring RC [95]. Current guidelines support TMT as an alternative to RC, highlighting the need for precise patient selection and close surveillance protocols, even in patients achieving complete remission, due to the high risk of recurrence [96].

Neoadjuvant chemotherapy (NAC) before RC plays a crucial role in enhancing patient survival, downstaging the primary tumor, and possibly eradicating unrecognized micro-metastases [97, 98]. Several randomized clinical trials have provided robust data supporting the benefits of platinum-based NAC, such as methotrexate + vinblastine + doxorubicin + cisplatin (MVAC) and cisplatin + methotrexate + vinblastine (CMV) and gemcitabine + cisplatin (GC), preceding locoregional treatment alone [98-102]. A meta-analysis of 15 randomized clinical trials (a total of 3,285 patients) showed that cisplatin-based NAC was associated with significant overall survival (OS) benefit (Hazard Ratio (HZ), 0.87 [95% confidence interval (CI), 0.79-0.96]; $p<0.01$) compared with locoregional treatment alone [103]. Another example is the SWOG trial (n=317 patients) which estimated a reduction of 33% in risk of death in MVAC plus cystectomy group compared to

cystectomy alone [98]. Improvement in median OS (77 months and 46 months in the NAC and surgery groups, respectively) were also observed [98]. Additionally, other clinical study proved the value of adding CMV regimen before cystectomy or radiotherapy, and revealed a 16% reduction in the overall risk of death, as well as a 6% increase in survival at 10 years [104]. Furthermore, a retrospective study with MIBC patients who received neoadjuvant GC or MVAC prior cystectomy showed similar clinical outcomes for both agents [101]. Notably, GC-treated patients showed a pathologic downstaging to either pT0 (26%) or <pT2 (36%), while pT0 was achieved in 28% and <pT2 in 35% of MVAC group at cystectomy [101]. Another clinical study, which main goal was to compare adjuvant versus neoadjuvant therapies in MIBCs, reported a median DFS of 24.9 and 34.6 months, respectively, as well as shown that 21% of patients treated with NAC had reduced risk of tumor progression [105]. Altogether, the better outcomes in NAC-treated patients might be due to a decrease in micro-metastases and a reduction in disease burden, leading to improved prognosis [105]. However, cisplatin-based therapies tend to fail and not be well-tolerated by some MIBC patients, delaying surgery and potentiating fatal outcomes [106]. The most commonly reported NAC-related toxicities included nausea and vomiting, hematological toxicities, and impaired renal function [107]. Therefore, it is necessary to have a more precise stratification of patients who undergo NAC to identify those who are more likely to respond and to redirect those who are likely to fail towards other therapeutic strategies [106]. If platinum-based NAC cannot be tolerated or adequate, MIBC patients should proceed to other therapeutic options, including up-front RC, alternative chemotherapy and immunotherapy [16, 45].

Patients with adverse pathologic features, such as an extravesical extension or node-positive disease, that have undergone RC without prior NAC might be indicated for adjuvant therapy [17]. Nevertheless, the role of adjuvant treatment remains controversial because individual studies have been underpowered, not clarifying which patients would benefit from adjuvant therapy and which regimens are the most adequate [18, 99, 108]. Data from a meta-analysis from 10 randomized controlled trials [(RCT); 1183 patients] showed encouraging information about the value of cisplatin-based adjuvant therapy on OS (HR=0.82, 95% CI = 0.70–0.96, $p = 0.004$) [109]. In more detail, the analysis demonstrated an absolute improvement in survival of 6% at 5-years, as well as in RFS, locoregional RFS, and metastasis-free survival with absolute rates of 11%, 11%, and 8%, respectively [109]. Additionally, a phase III trial of platinum-based therapy, which

main goal was to compare immediate adjuvant chemotherapy and delayed therapy after RC in MIBC patients, revealed a 5-year PFS of 47.6% and 31.8%, respectively [110]. However, no significant improvement in 5-year OS (57% vs 47%; p=0.13) was shown comparing immediate and deferred groups [110]. A retrospective study about the role of adjuvant radiotherapy after RC in MIBC patients (170 patients T3 and T4) demonstrated a significant 5-year DFS rates of 65% for post-operative radiotherapy (PORT) group and 40% for de non-PORT group [111]. Nevertheless, no differences between groups were observed regarding OS and overall distant metastases rate [111]. Altogether, most experts believe these analyses are immature to allow drawing definite conclusions, and thus further validation and clinical studies are required. Notwithstanding, adjuvant therapy might be considered as a valid option for patients ineligible for NAC after RC to improve outcomes for MIBC [7].

1.4.3. Metastatic Bladder Cancer Management

Only 4% of BlCa patients present metastatic disease at diagnosis, however, these patients have a poor prognosis with a median survival with standard chemotherapy of 13 to 15 months [112, 113]. For decades, there have been modest advances in metastatic BlCa therapeutics, being systemic cisplatin-based cytotoxic chemotherapy the first-line of treatment [114]. Namely the MVAC, CMV, GC, or GC plus paclitaxel combinations [16]. For instance, a phase III RCT with 405 patients showed similar OS for GC and MVAC treatments (13.8 and 14.8 months, respectively), but substantially fewer adverse-effects, including mucositis (1% vs 22%; p<0.001) and neutropenic sepsis (1% vs 12%; p<0.00) for GC treatment [113]. In another phase III trial, dose-dense of MVAC demonstrated improved outcomes in comparison to standard MVAC, especially in CR rates (21% vs 9%, p=0.009) and median PFS (9.1 vs 8.2 months; p=0.037) [115]. Subsequently, metastatic patients who are not eligible for cisplatin-based therapy (primarily driven by renal insufficiency) might be managed with carboplatin-based regimens, although data have suggested overall worse outcomes compared to cisplatin-based management [116, 117]. Up to this point, there are no successful therapeutic options for those who are not candidates for RC or are not eligible for standard therapies. Remarkably, since 2016 the treatment landscape for metastatic BlCa patients has shifted dramatically with the introduction of systemic immunotherapy based on immune checkpoint inhibitors (ICIs) [118-120].

The characterization of the molecular profile of bladder tumors enhanced the identification of potential immunotherapeutic targets such as the programmed cell death-1 protein (PD-1) and its ligand, programmed cell death ligand 1 (PD-L1) [120-122]. Currently, five ICIs against these targets were approved by the FDA for BlCa therapy, namely atezolizumab, avelumab, durvalumab, nivolumab, and pembrolizumab, which are monoclonal antibodies that block the interaction between PD-1 and PD-L1 [123-126]. A KEYNOTE-045 phase III trial (n= 542 patients) demonstrated the benefit of pembrolizumab (PD-1 inhibitor) over chemotherapy for the treatment of advanced/metastatic BlCa [127]. This study reported an improvement in the median 2-year OS rate in patients treated with pembrolizumab (26.9%) over patients receiving chemotherapy (14.3%) [127]. Moreover, the PD-1 inhibitor also showed lower rates of treatment-related adverse events compared to standard chemotherapy [127]. The phase III IMvigor211 clinical trial revealed statistically significant 2-year OS rates of 23% and 13% for atezolizumab (PD-L1 inhibitor) and chemotherapy groups, respectively [128]. Thus, both pembrolizumab and atezolizumab were approved as second-line agents for patients with positive for PD-L1tumors/infiltrating immune cells or as first-line therapies for patients ineligible for cisplatin [123, 129]. Furthermore, avelumab, durvalumab, and nivolumab are second-line treatments, exhibiting favorable clinical outcomes in the metastatic BlCa patients [124-126]. While the field of cancer immunotherapy is rapidly evolving, tremendous efforts to identify novel agents or rationale combinations as later-line treatment are underway[130]. Recent research has identified various targets for BlCa therapy based on altered signaling pathways or epigenetic modifications. Accordingly, the enfortumab vedotin, an antibody-drug conjugate (ADC) targeting Nectin-4 conjugated to monomethyl auristatin E (MMAE) was approved by FDA for advanced or metastatic BlCa patients previously treated with platinum chemotherapy and ICIs [129, 131]. In more detail, a phase II study (EV-201) with 125 patients treated with platinum and anti-PD-1 therapies demonstrated a confirmed objective response of 44%, including 12% of CR with a manageable and tolerable safety profile [132]. Other recent clinical trials with ADC enfortumab vedotin also showed improved OS and clinical benefits for advanced MIBC patients who had received prior platinum-based chemotherapy or immunotherapy with ICIs, as well as for cisplatin-ineligible patients previously treated with other therapies [133-135]. Several fibroblast growth factor receptor (FGFR) inhibitors, such as erdafitinib (the first FDA-approved) and infigratinib, have also shown encouraging results in clinical trials with advanced BlCa patients [136,

137]. Collectively, emergent BlCa therapeutics hold promise for improving the outcomes of BlCa patients. Ongoing clinical trials will further elucidate the efficacy and safety of novel agents, as well as the potential combinations between them, and their integration into the standard of care is anticipated soon.

In spite of the promising benefit of novel immunotherapies for platinum-ineligible/resistant and metastatic BlCa patients, response rates remain modest, typically ranging from 15% to 25% [118, 119, 125, 138, 139]. As such, innovative therapeutics are required to offer the necessary opportunity to advanced BlCa patients faced with palliative care. Classically, tumor-associated and tumor-specific antigens were deemed to arise from genomic mutations[140]. However, this notion has progressed to the acknowledgement that transcript and proteomic variants, including post-translationally modified proteins can provide the necessary neo-antigenicity for novel immunotherapies [141-143]. In line with this, glycosylation is one of the most common Post-Translational Modifications (PTMs) of proteins, being present in virtually all living cells [144]. Accordingly, glycomics, glycoproteomics, and glycoproteogenomics are expected to be the next generation fields fostering unique molecular signatures to target aggressive tumors [27].

2. Glycosylation patterns in Bladder Cancer

2.1. General features of glycosylation

The cellular proteome comprises approximately 100,000 proteins that are exposed to several PTMs, including phosphorylation, ubiquitination, SUMOylation, acetylation, lipidation, and glycosylation [145, 146]. It is estimated that up to one-fifth of all proteins may be glycosylated to some extent, making glycosylation one of the most frequent events occurring in cells [146].

Briefly, glycosylation is defined as the enzymatic process where carbohydrates (also known as saccharides or glycans) are covalently linked to other sugars, proteins, or lipids, by glycosyltransferases (GTs), forming glycoconjugates [147]. This mechanism also depends on the coordinated action of several nucleotide sugar transporters, glycosidases, and molecular chaperones in the endoplasmic reticulum (ER) and Golgi apparatus (GA) of mammalian cells [148-150].

There are two main classes of glycans: *N*-glycans, covalently attached to a polypeptide backbone via the nitrogen atom of asparagine (Asn) residues, and *O*-glycans, attached to a polypeptide chain via the hydroxyl group of serine (Ser), threonine (Thr) or tyrosine (Tyr) residues [151-153]. In the ER, *N*-glycans synthesis

is initiated by *en bloc* transferring of the preassembled oligosaccharide $Glc_3Man_9GlcNAc_2$ from the dolichol-phosphate lipid carrier to the Asn residues of an Asn-X-Ser/Thr ("X" stands for any amino acid except proline) sequence of the nascent protein [154]. This event is catalyzed by the oligosaccharyltransferase (OST) complex [154]. Once transferred to the protein, the oligosaccharide core is trimmed and modified in the ER and GA by several GTs and glycosidases, yielding a diverse array of *N*-glycans [155]. All mammalian *N*-glycans (Figure 2) share a common core sequence, $Man_3GlcNAc_2Asn$, and can be classified into three major types as high-mannose *N*-glycans, complex *N*-glycans and hybrid *N*-glycans [156]. High-mannose *N*-glycans, the simplest and most conserved type, consists of the core structure with up to nine mannose (Man) residues [156]. Complex *N*-glycans, the most structurally diverse type, have a core sequence customized with several monosaccharides, such as *N*-acetylglucosamine (GlcNAc), fucose (Fuc), galactose (Gal), and sialic acid (Neu5Ac), forming complex tree-like structures through branching and elongation by the action of several GlcNActransferases (GnTs) [156]. Hybrid *N*-glycans, which have the combination of features from both high-mannose and complex *N*-glycans, exhibit a core structure of Man and GlcNAc residues modified with various monosaccharides and complex structures [156]. Moreover, mature *N*-glycans could often exhibit specific terminal epitopes such as Lewis (Le) blood group-related antigens (Le^a, Le^x, Le^b, and Le^y; Figure 3) and their corresponding sialylated structures (Figure 3) or ABO(H) blood group determinants [157]. Further *N*-glycome complexity can be increased by other structural modifications including *O*-acetylation of sialic acids, as well as *O*-sulfation of galactose and *N*-acetylglucosamine residues [158]. Thus, the diversity of *N*-glycans reflects the importance of this type of glycosylation in a wide range of biological processes, namely in protein folding, protein-protein interactions, cell signaling, and immune recognition [155].

In contrast with *N*-glycosylation, the bulk of *O*-glycan biosynthesis (Figure 2) occurs in the GA and does not always require the consensus glycan core to initiate their synthesis [159]. In fact, *O*-glycans can be modified with *O*-fucose, *O*-mannose, *O*-glucose, *O*-galactose, and *O*-xylose in proteoglycans [159]. Nucleoplasmic proteins can also carry *O*-GlcNAc glycosylation [160]. However, the most abundant type of *O*-glycans are *O*-*N*-acetylgalactosamine (GalNAc) glycans, also called mucin-type *O*-glycans, given their high abundance in mucins [161].

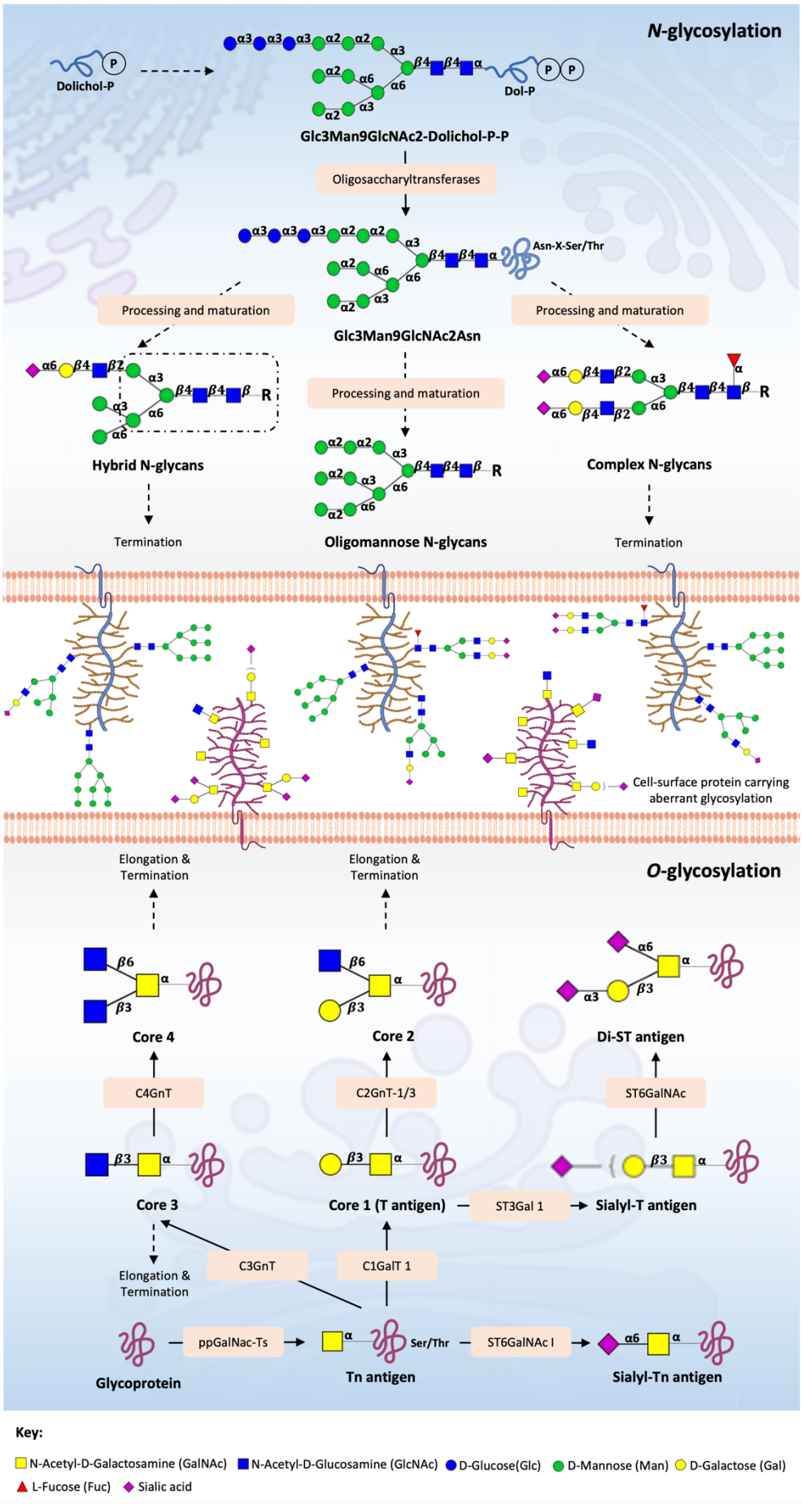

N-glycosylation
Dolichol-P
Dol-P
Glc3Man9GlcNAc2-Dolichol-P-P
Oligosaccharyltransferases
Asn-X-Ser/Thr
Glc3Man9GlcNAc2Asn
Processing and maturation
Processing and maturation
Processing and maturation
Hybrid N-glycans
Oligomannose N-glycans
Complex N-glycans
Termination
Termination
Cell-surface protein carrying aberrant glycosylation
O-glycosylation
Elongation & Termination
Elongation & Termination
Core 4
Core 2
Di-ST antigen
C4GnT
C2GnT-1/3
ST6GalNAc
Core 3
Core 1 (T antigen)
ST3Gal 1
Sialyl-T antigen
C3GnT
C1GalT 1
Elongation & Termination
ppGalNac-Ts
Ser/Thr
ST6GalNAc I
Glycoprotein
Tn antigen
Sialyl-Tn antigen
Key:
N-Acetyl-D-Galactosamine (GalNAc)
N-Acetyl-D-Glucosamine (GlcNAc)
D-Glucose(Glc)
D-Mannose (Man)
D-Galactose (Gal)
L-Fucose (Fuc)
Sialic acid

Figure 2- Schematic representation illustrating *N*- and *O*-glycan biosynthesis pathways. Proteins undergo glycosylation via the covalent attachment of saccharides to polypeptide backbones, occurring through *N*-linkage to Aspartic acid (Asn) or *O*-linkage to Serine (Ser)/Threonine (Thr) residues. *N*-glycosylation primarily occurs in consensus peptide sequences, Asn-X-Ser/Thr, where X represents any amino acid. *N*-glycans share a common pentasaccharide core region (depicted by a dotted line box) which can undergo further maturation into hybrid, oligomannose, or complex types. These structures are subject to modification with terminal monosaccharides as *N*-Acetylglucosamine (GlcNAc), Galactose (Gal), or sialic acid (Neu5Ac). Mucin-type *O*-glycans biosynthesis initiates with the transfer of *N*-acetylgalactosamine (GalNAc) onto Ser or Thr residues of the polypeptide by GalNAc transferases (GalNAc-Ts). This results in the formation of the Tn antigen (GalNAcα1-Ser/Thr), subsequently converted into the T antigen or Core1 (Galβ1-3GalNAcα1-Ser/Thr) by Core1 synthase (C1GalT1). Core1 may undergo further modification into Core2 by C2GnT, while the Tn antigen can also be transformed into Core3 by Core3 synthase (C3GnT), and further into Core4 by C4GnT. Concurrently, the Tn antigen may be converted into Sialyl-Tn (STn) antigen by ST6GalNAc1, while Core1 can be transformed into Sialyl-T (ST) antigen by ST3Gal1. Moreover, ST antigen can be elongated with another sialic acid, yielding in di-ST glycoforms. These core structures act as precursors for a diverse array of more complex and elongated glycans, regulated by tissue-specific glycosyltransferases in a microenvironment-dependent manner.

Particularly, mucins are a family of heavily glycosylated transmembrane proteins that have large repetitive domains rich in proline (Pro), Thr, and Ser (PTS domains), which are targets for *O*-GalNAc glycosylation (hereinafter referred to as *O*-glycosylation) [162].

O-glycosylation (Figure 2) is a multi-step pathway that begins with the transfer of GalNAc residue from a UDP-GalNAc donor to either a Ser, Thr or Tyr residues on the protein backbone, catalyzed by UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases (GalNAc-Ts) [152, 163]. Following this step, the simplest *O*-glycan (Tn antigen) is formed, and can then be branched, elongated, and finally capped with terminal residues. Succinctly, the Tn antigen (GalNAc-*O*-Ser/Thr structure) can be modified by the enzyme Core 1 β1,3-galactosyltransferase (C1GalT1) that catalyzes the attachment of a Gal residue to GalNAc, forming the Core 1 structure (also known as T or Thomsen-Friedenreich antigen) [164]. This step is reliant on the specific chaperone, the Core 1 β1-3-galactosyltransferase-specific chaperone 1 (COSMC) [165]. Subsequently, core 1 (Galβ1-3GalNAcα1-Ser/Thr structure) is branched with GlcNAc residue by the enzyme Core 2 β1,6-*N*-acetylglucosaminyltransferase (C2GnT), synthetizing the core 2 antigen, which consists of a GlcNAcβ1-6(Galβ1-3)GalNAcα1-Ser/Thr structure [164]. Alternatively, a GlcNAc moiety can be added to the Tn antigen under the action of the β1,3-*N*-acetylglucosaminyltransferase (β3GnT) enzyme, yielding the

core 3 (GlcNAcβ1-3GalNAcα1-Ser/Thr) structure, which can further be elongated with another GlcNAc, originating the core 4 structure (GlcNAcβ1-3(Galβ1-3(GlcNAcβ1-6)GalNAcα1-Ser/Thr) [164]. All core structures can then be further extended with Gal, GalNAc, or GlcNAc residues, and terminally capped with sialic acid and fucose monosaccharides [164]. In homeostasis, the resulting *O*-glycoproteins can then be secreted or transferred to the plasma membrane, thereby ensuring their standard function in a panoply of biological events, including cell signaling, cell–cell communication, immune recognition and surveillance, signal transduction, and host-pathogen interactions [164]. However, in malignant conditions, the *O*-glycans synthesis is profoundly deregulated, which leads to significant alterations in membrane protein's glycosylation patterns and critical alterations of biological processes, ultimately creating unique variants with implication in disease course [159, 166].

Particularly, one of the most described cancer-related glycosylation alterations consists in the premature stop of *O*-glycans biosynthesis (Figure 2), resulting in the formation of truncated forms, and the expression *de-novo* of glycosidic antigens of foetal type [167]. During *O*-glycans biosynthesis, Tn antigen can be catalytically capped with a sialic acid by ST6GalNAc 1 (α2,6-sialyltransferase), forming the Sialyl-Tn (STn) antigen [166, 168]. In addition, under the action of ST3Gal1 enzyme, T antigen can be α2,3-sialylated giving rise to Sialyl-T (ST) antigen, which can be further extended with another sialic acid, yielding Di-ST forms [166, 168]. Similarly to *N*-glycans, terminal fucosylated Lewis antigens H-type 1 (Le^b, Le^a and Sialyl-Le^a [SLe^a]; Figure 3) and H-type 2 (Le^y, Le^x and Sialyl-Le^x [SLe^x]; Figure 3) can also frequently occur in *O*-glycans [169, 170]. The H-type 1 chain (Galβ1,3GlcNAc) GlcNAc fucosylation is dependent of the enzymatic activity of fucosyltransferases (FUT) 3 and 5 to synthetized Le^b and Le^a, whereas H-type 2 chain (Galβ1,4GlcNAc) modification with α1,3-fucose is mediated by FUT3, FUT4, FUT7 and FUT9, forming Le^y, Le^x (Figure 3) [169, 170]. Following fucosylation, the terminal Gal residue of type 1 and type 2 chains are capped with a sialic acid under the action of sialyltransferases, generating SLe^a and SLe^x antigens (Figure 3), respectively [171]. Upon *O*-glycans terminal capping with sialic acid and fucose residues, which culminate in these truncated glycoforms, any further processing of the oligosaccharide chain is stopped [166]. This process results in numerous proteins, including enzymes, transcription factors, receptors, and structural proteins, with abnormal *O*-glycosylation patterns, which will determine their

specific roles in cancer [167]. Indeed, aberrant *O*-glycosylation carried by many cell-surface proteins, such as Mucin 1 (MUC1), Cluster of Differentiation 44 (CD44) and Epidermal growth factor receptor (EGFR), among several others, is involved in tumor progression and metastasis by altering cell adhesion, migration, invasion, and resistance to therapy [172-177].

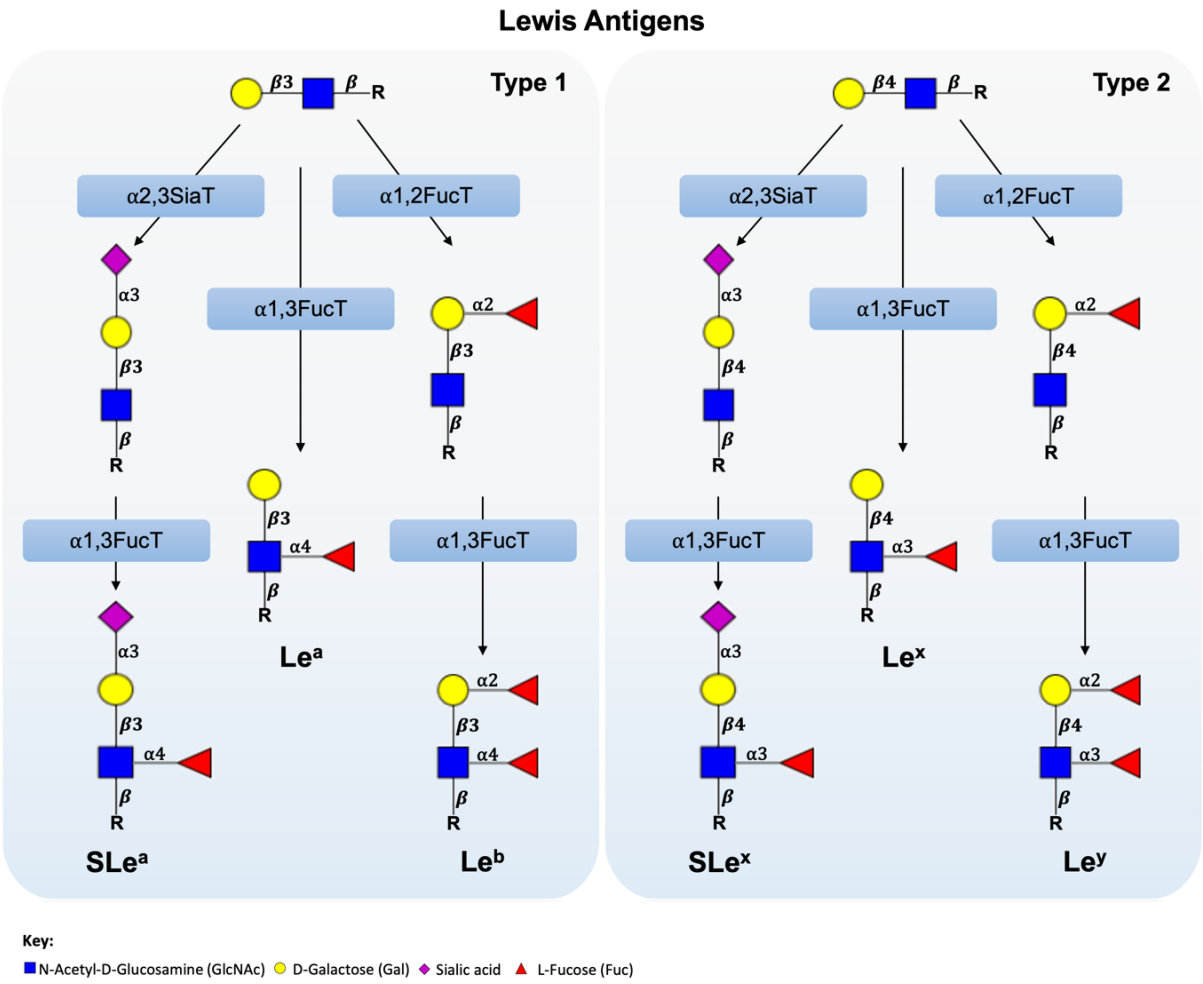


Figure 3- Terminal Lewis and sialylated Lewis structures on glycans. Type 1 Lewis antigen is formed by a Gal residue attached to a GlcNAc residue through a β1,3 linkage, while Type 2 Lewis antigen features a β1,4-linked Gal residue attached to the GlcNAc residue. The α1,3/4 fucosylation of type 1 and type 2 chains results in Lewis^a (Le^a) and Lewis^x (Le^x), respectively. Additionally, H1- and H2-determinants can undergo further modifications to yield Lewis^b (Le^b) and Lewis^y (Le^y) di-fucosylated antigens, achieved through α1,4 or α1,3-fucosylation, respectively. In cancer cells, α2,3-sialylation of Type 1 and Type 2 chains, followed by α1,4 or α1,3-fucosylation, leads to the expression of Sialyl-Lewis^a (SLe^a)and Sialyl-Lewis^x (SLe^x) antigens, respectively. These terminal antigens can be found in *O*-glycans, *N*-glycans, and glycolipids. Abbreviations: FucTs- fucosylstransferases; SiaTs-Sialyltransferases.

2.2. *O*-glycans relevance in bladder cancer

Alterations in the *O*-glycosylation pathway and the glycan patterns on cell-surface proteins are frequent hallmarks of malignant transformations in BlCa [178]. In recent decades, glycosylation status has been associated with diverse BlCa scenarios, including bladder carcinogenesis, disease progression, invasion, metastasis, and immune evasion [179-181]. Furthermore, aberrant *O*-glycoproteomic signatures could provide a promising tool for BlCa detection, treatment assignment and prognosis definition in BlCa patients [37, 182, 183].

In general, urinary bladder malignant transformation is accompanied by the loss of terminal ABO(H) blood group antigens on cell-surface proteins, an increase or *de novo* expression of truncated *O*-glycans, overexpression of Lewis-type terminal structures, overall hypersialylation and fucosylation of *O*-glycans, as well as altered expression and/or activity of GTs [157, 181].

The loss of ABO(H) blood group determinants in bladder carcinomas was the first change in glycosylation patterns reported in BlCa [184, 185]. Bergman and Javadpour, comparing normal bladder epithelium with low-grade and early-stage TCC, demonstrated that these antigens were present in normal tissues but not in the majority of TCC cases, highlighting their role as a predictor of malignant potential [184]. Also, It showed that initially expressing tumors that recurred locally or progressed to invasion or metastization, lose these cell-surface antigens [184]. Other reports seem to corroborate these findings, revealing a reduction in expression in ABO(H) antigens in TCC and its correlation with an invasive phenotype [186-188]. Orntoft and Wolf established the correlation between blood-group determinants expression and GTs, demonstrating that their catalytical activity was drastically reduced in tumors that lost the A antigen expression [189]. Moreover, hypermethylation and allelic loss in the promoter region of the ABO(H) gene have also been suggested as an explanation for the deletion of these epitopes in bladder tumors [190]. By contrast, despite biosynthetic and structural similarities between the ABO(H) and Lewis antigens, several authors demonstrated a reverse pattern of expression of Le^x epitope in BlCa in comparison to ABO(H) determinants. As such, a significantly higher expression of Le^x in cancerous tissues compared to healthy urothelium was shown [191, 192]. This finding was supported by later studies demonstrating that Le^x was detected in 90% of bladder tumors by immunohistochemistry (IHC), as well as in exfoliated cells of urine samples, denoting its potential as a diagnostic biomarker [193-195]. In addition, the

sialylated form of Le^x antigen seems to play a key role as a predictor of invasive and metastatic potential in BlCa [196]. Nevertheless, a different investigation reported there was no link between SLe^x and stage or grade of urothelial carcinoma in the renal pelvis, ureter, and bladder [197]. Concerning H-type 1 Lewis antigens, elevated levels of SLe^a (also called CA19-9) expression were found in bladder tumors that had metastasized and spread [198]. Monitoring its presence in patients' urine appears to be a useful diagnostic tool for detecting BlCa [199]. In this context, the augmentation of SLe^a and SLe^x expression on the surface of cancer cells helps them to interact with their particular ligands, namely E-selectin and P-selectin, which are expressed on endothelial cells [200, 201]. This interaction promotes the adhesion of cancer cells to the endothelium lining blood vessels and eventually leads to metastasis [200, 201]. Besides this, the overall increase of sialic acid content on cell-surface glycans can also have a significant impact on BlCa progression and metastasis by masking tumor-associated antigens and preventing recognition by immune cells, avoiding their destruction and favoring metastatic spread [202]. Moreover, sialylation can modulate the function of adhesion molecules such as selectins and integrins, which are involved in tumor cell adhesion and metastasis [202].

The overall increase of sialylated glycans in advanced bladder tumors arises not only from sialylated Lewis antigens but also from a substantial rise of STn antigens expression. STn is expressed by more than 80% of human carcinomas, being invariably associated with unfavorable outcomes and reduced overall survival rates in patients [203, 204]. Particularly in BlCa, the overexpression of STn antigen was associated with the augmentation of ST6GalNAc1 sialyltransferase expression levels in high-grade cases [179]. This antigen was found in over 70% of high-grade NMIBC and MIBC but was absent in healthy urothelium, highlighting its cancer-associated nature [205]. Moreover, the same study demonstrated that STn was mostly expressed in non-proliferative tumor areas, which are known to be highly resistant to cytostatic agents currently used to treat advanced-stage BlCa [205, 206].

More recently, the functional impact of the STn in BlCa has been investigated through the use of glycoengineered tumor cells [207]. This approach is necessary because STn expression is uncommonly observed in *in vitro* cell models, indicating a dependence on the tumor microenvironment for its biosynthesis [208, 209]. As such, some reports have demonstrated that microenvironmental factors, such as hypoxia and glucose deprivation, are significant triggers for STn expression in BlCa

cells. This underscores the dynamic nature of this PTM, which promptly mirrors cellular stress. [207, 210, 211]. In this context, STn has been shown to enhance resistance to platinum-based chemotherapy and anoykis [207]. This indicates that STn-positive cells that are likely resistant to chemotherapy may have a higher potential for metastasis [207]. However, the correlation between STn overexpression and chemoresistance in BlCa has yet to be fully investigated. Additionally, a different study has reported that STn expression is significantly elevated in MIBC compared to NMIBC, suggesting its association with muscle invasion and poor prognosis [212]. The presence of STn in CTCs and distant metastases of BlCa patients also reinforces its link between STn-associated glycoproteome and increased metastatic potential [37]. Incorporating STn into molecular prognostic and stratification models based on keratins expression (KRT14+, and/or KRT5+, and KRT20-) for advanced-stage bladder tumors has allowed the discrimination of patients with the worst survival, further supporting its association with aggressiveness [37]. On the same note, the expression of STn and sialyl-6-T (s6T), a sialylated form of T antigen, exhibit independent prognostic value for predicting the response to BCG treatment in BlCa patients, thus serving as potential biomarkers for identifying patients who may benefit the most from this immunotherapy [183]. Cotton *et al.* glycoproteomics analysis of advanced bladder tumors showed several cancer-related glycoproteins carrying altered glycosylation, such as CD44, Mucin 16 (MUC16), and numerous integrins [182]. Specifically, this study demonstrated that MUC16-STn glycoproteoforms hold potential as biomarkers for poor prognosis and designing targeted therapeutics [182]. Importantly, neither MUC-16 or STn expression alone in the evaluated bladder tumors had the capacity to stratify patients prognosis in opposition to MUC-16-STn glycoproteoforms [182]. The role of STn in modulating immune tolerance in BlCa was also explored. Interestingly, BlCa cells expressing STn harm the maturation of dendritic cells (DCs) and bestow them with a tolerogenic function, ultimately hindering their ability to initiate effective anti-tumor T-cell responses [179]. Overall, these findings have significant implications for understanding the mechanisms underlying the development and progression of BlCa, as well as the potential theragnostic targeting of STn antigen.

In turn, additional cancer-associated glycoforms can be generated from the premature stop of *O*-glycosylation, namely the T and ST antigens [171]. An early study conducted in BlCa reported an association between T antigen expression and tumor invasion [213]. Other reports have linked the presence of T glycoforms with

higher grade, stage, and poor prognosis in BlCa, implying that these antigens may act as surrogate markers of profound cellular changes [214, 215]. Moreover, Videira *et al.* highlighted the correlation between the overexpression of ST3Gal I, the sialyltransferase responsible for T antigen sialylation, and higher stage and poor prognosis in BlCa [216]. In line with these findings, the sialylated T antigens, specifically the S3T and S6T glycoforms, were found in bladder tumors, but not in the healthy urothelium, underscoring their cancer-associated nature [182]. Of note, the same study demonstrated that these sialylated glycans were overexpressed in advanced lesions, particularly in cases with muscle invasion [182]. Altogether, these observations emphasize the potential of these short glycoforms as therapeutic targets for BlCa, as well as predictors of disease prognosis.

As previously described, *O*-glycans play a crucial role in the tumor metastasis, but there is some duality regarding their role in this process, being highly dependent on both their structure and the origin of the tumor. Interestingly, MUC1 carrying core 2 *O*-glycans on the surface of BlCa cells acts as a molecular barrier against natural killer (NK) cells, thereby facilitating bladder tumor metastasis [217, 218]. Overall, structurally different *O*-glycans have distinct roles in tumor metastasis, and further investigation is necessary to understand the molecular mechanism underlying the effect of *O*-glycans and corresponding GTs on BlCa metastization.

Finally, in recent years, there has been a growing recognition of the potential of these cancer glycosignatures as targets for the development of innovative targeted therapies, given their cancer-associated and cell-surface nature [219-221]. Consequently, a range of therapeutic strategies targeting the Tn antigen are being developed, including glycovaccines, cellular immunotherapies, and monoclonal antibodies, which are currently in pre-clinical or clinical trial stages [222-227]. Of note, most cancer-associated glycosylation signatures have residual expression in non-malignant conditions related to tissue inflammation events and mucinous secretions of respiratory and gastrointestinal tissues [203]. In line with this, next generation glycan-targeted therapeutics are expected to explore the bi-specificity of glycan and carrier proteins together in accordance with the cancer-specificity associated to glycoproteoforms found in BlCa. However, BlCa remains mostly an "orphan disease" in terms of targeted therapeutics, mainly in the field of glycan-based therapies, demanding the identification of novel molecular targets [228, 229]. In this context, a study has shown that the glycosylation patterns of CD44 is altered in BlCa, leading to the formation of specific glycosignatures that are not

present in normal cells [182]. Specifically, the authors demonstrated a co-expression of STn and CD44 in the same tumor area using IHQ and proximity ligation assay (PLA) techniques. Nevertheless, the specific *N*- and *O*-glycan patterns of CD44 in BlCa remained largely unknown, but represent a promising area for further investigation. Thus, narrowing down CD44 isoforms to its clinically relevant glycoepitopes may provide key molecular insights for highly specific BlCa targeting therapies.

3. CD44: a multifaceted glycoprotein

3.1. Overview of human CD44 and its isoforms

CD44 was initially described in the 80's as a lymphocyte-homing receptor, and subsequently cloned and identified as a member of the cartilage link protein family [230-232]. The human *CD44* gene is located on chromosome 11 on the short arm p13 and consists of 19 exons [233, 234]. Mouse *CD44* gene (located on chromosome 2) has 20 exons, containing an additional exon, the exon V1 (V stands for variant exon), which is absent in humans [235, 236]. CD44, also known as In (Lu)-related protein p80, Pgp-1/Ly-24, ECMRIII, HUTCH-1, Hermes antigen, and hyaluronic acid (HA) receptor, is composed of a variable extracellular domain that includes the stalk region, a transmembrane domain, and an intracellular cytoplasmic tail [237]. In humans, exons 1 to 16 encode the extracellular domain, with exons 15-16 encoding the stalk region, while exon 17 and exons 18-19 encode the transmembrane and cytoplasmic domains, respectively (Figure 4)[234]. Exons 1-5, constitutively expressed across all known isoforms, encode the binding sites on the extracellular domain for several ligands, such as HA, osteopontin (OPN), collagen, laminin, and fibronectin [235, 238-241]. On the other hand, the variable exons 6-14, also known as v2-v10, are alternatively spliced, giving rise to numerous CD44 variant isoforms with distinct structures and functions [235]. The complexity of these isoforms is further amplified by PTMs with *N*-glycans, *O*-glycans, and glycosaminoglycans, such as chondroitin sulphate and heparan sulphate [242]. In general, the majority of potential glycosylation sites are located at the CD44 ectodomain [243]. For the canonical CD44 isoform, a total of 146 *O*-glycosylation sites and 8 *N*-glycosylation sites have been predicted [235, 243]. Despite being often neglected, glycosylation can potentially increase (sometimes even double) the molecular weight of isoforms, leading to significant modulation of the biophysical, biochemical, and functional properties of the CD44 receptor

[235]. In light of these particularities, the discrimination of CD44 isoforms at the protein level represents a significant challenge due to the lack of standardized protocols. As above mentioned, the density of CD44 glycosylation constitutes an analytical hurdle to CD44 isoform identification, while considerably increasing molecular heterogeneity [235]. Furthermore, inconsistency and ambiguity in the nomenclature of CD44 isoforms leads to confusion and bias in the interpretation of CD44 isoforms studies [235]. This lack of nomenclature standardization also hampers inter-study comparisons and makes clinical translation almost impossible. To address these challenges, Azevedo *et al.* recently proposed a logical terminology building on pre-existing designations to improve the interpretation and comparison of studies on CD44 isoforms [235]. Currently, according to the Ensembl and UniProt databases, the *CD44* gene exhibits 38 splicing variants, out of which 24 transcripts are possibly translated into proteins. However, among these, only 8 isoforms have been experimentally validated (Figure 4), and according to the proposed nomenclature are the following: CD44v2-10, which is the canonical form of CD44 and is composed of a peptide sequence encoded by exons 6-14, with exon 18 being spliced out; CD44v3-10 or epican, arises from retaining exons 7-14 (v3-v10) and splicing out exon 18; CD44v8-10, also known as CD44E or CD44R1, is an isoform that contains the variant exons 12-14 (v8-v10) and splices out exon 18; CD44v10, also called gp116 or CD44R2, retains only the variable exon 14 (v10) while splicing out exon 18; CD44s, also known as CD44H or gp85, is the standard form of CD44 and does not contain any splicing exons or exon 18; CD44st, or short-tail, is a smaller isoform that splices out the variable region and exon 19, while retaining exon 18. Notably, exon 18 harbors a stop codon that leads to a truncated cytoplasmic tail, resulting in the loss of intracellular domains and signaling motifs required for interaction with cytoskeletal components; CD44s-exon15 is a homolog of CD44s that lacks the peptide sequence encoded by exon 15; Lastly, CD44sol is a soluble isoform of CD44 that retains only exons 1-4 with truncated forms of exons 3 and 4, resulting in a smaller extracellular domain and loss of transmembrane and cytoplasmic domains [235].

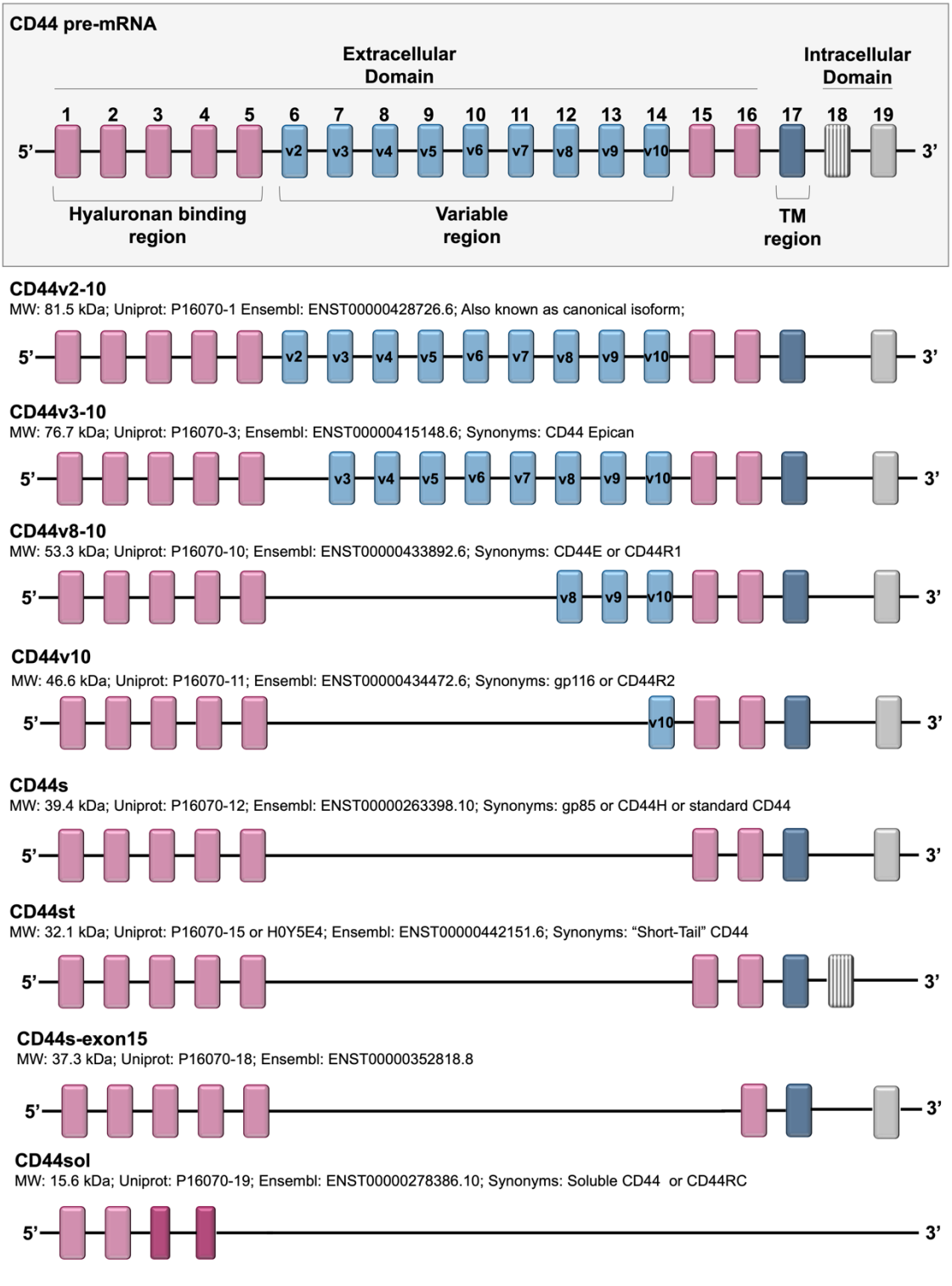


Figure 4- Schematic representation of experimentally validated human CD44 pre-mRNA and its isoforms. Conserved exons, namely exons 1-5 and 15-16, are presented in pink, while variable exons 6-14 (v2-v10) are shown in light blue. The transmembrane region is highlighted in dark blue (exon 17), and the intracellular tail is depicted in grey (exons 18-19). Exon 18, indicated by gray stripes, contains an early 3'UTR and is exclusive to the CD44st isoform. Truncated exons from the constant

region and are denoted by dark pink filled boxes with reduced size. A black line indicates the absence of exon(s). Additionally, the estimated molecular weight (in kDa) for each variant (excluding posttranslational modifications), as well as Uniprot and Ensemble codifications, are emphasized. According to the NCBI and Uniprot databases, only eight CD44 isoforms have been experimentally validated, with their mRNA coding for the protein. These isoforms include CD44-v2-10, CD44v3-10, CD44v8-10, CD44v10, CD44s, CD44st, CD44s-exon15, and CD44sol.*The isoforms not represented in this figure are predicted based on data from NCBI protein database.

Under physiological conditions, CD44 exhibits widespread expression in various cell types, including lymphocytes, smooth muscle, fibroblasts, and various types of epithelia [237]. Specifically, the standard form of CD44, initially identified in hematopoietic cells, is also found in other tissues such as the liver, lung, pancreas, skin, and central nervous system [244]. Conversely, variant isoforms of CD44 are confined to specific cell types, encompassing macrophages, activated lymphocytes, keratinocytes, and epithelial cells from the stomach, bladder, and uterine cervix [237, 244, 245].

This multi-structural glycoprotein plays a pivotal role in a myriad of biological processes, with its functionality intricately tied to the nature of the host cell or tissue and the specific pathophysiological context. Notably, CD44 is involved in cell adhesion and aggregation, cell migration, lymphocyte homing, leukocyte activation, lymphopoiesis, myelopoiesis, angiogenesis, and cytokine release [244, 246-248]. In a pathological context, particularly in cancer, this glycoprotein is overexpressed in a spectrum of tumor tissues, namely lymphoma, breast, prostate, ovarian, gallbladder, and bladder malignancies [249-255]. However, the precise role of CD44 in malignant tumors remains complex and not well-defined. Nevertheless, some studies have highlighted its association with tumor progression, migration, invasion, metastasis, therapy resistance, and modulation of immune responses [256-261].

The upcoming section will delve into a comprehensive exploration of CD44 and its isoforms, specifically scrutinizing their pivotal roles and functions in the context of BlCa.

3.2. CD44 and its isoforms in Bladder Cancer: focusing on clinical relevance

The involvement of CD44 in cancer is complex and context-dependent, with both tumor-promoting and tumor-suppressing functions reported across different cancer types [262-265]. Its versatility arises from its capacity to undergo alternative

splicing, leading to the generation of multiple isoforms with distinct, but often overlapping, functional properties. Notably, the discovery of CD44 splicing variants in the 1990s sparked a new era in CD44 research [266, 267].

3.2.1. Prognostic value of CD44 and its isoforms in Bladder Cancer

Particularly in BlCa, CD44 has garnered increasing attention due to its potential as a prognostic biomarker. One of the most described spliced variants is the CD44 isoforms containing exon v9 (hereafter referred to as CD44v9). High levels of these specific isoforms have been associated with a worse prognosis in BlCa. In a study conducted by Kobayashi *et al.,* it was established that CD44v9 holds promise as a potential prognostic biomarker for assessing tumor progression and predicting cancer-specific death in MIBC and High Risk-NMIBC patients [268]. Another work, involving 63 MIBC patients who underwent RC, reinforced the prognostic value of CD44v9. It demonstrated that the presence of CD44v9 expression within tumor specimens stands as an independent prognostic marker for identifying unfavorable survival outcomes [269]. Furthermore, in a retrospective study with a cohort of 110 patients treated surgically for muscle invasive upper tract urothelial cancer (UTUC), CD44v9 emerged as a novel biomarker for poor clinical outcomes. The study revealed that CD44v9 expression was independently associated with both tumor recurrence and cancer death [270]. In the same note, CD44v8-10 isoform, which contains the variant exon v9, was also identified as a novel prognostic predictor of the BlCa by Miyake *et al.* [271].These authors revealed that this particular isoform was found overexpressed in urinary exfoliated cells of patients with invasive disease compared with those with superficial urothelial cancer (UC). Moreover, patients with UC who had an increased urinary CD44v8-10/CD44s ratio experienced a shorter DFS [271]. Over the past few decades, there has been extensive research conducted on CD44 isoforms containing the exon v6 (CD44v6),as well as lesser-explored variant exons (v5, v7, and v10), along with CD44s within this context. An in-depth multivariate survival analysis involving 173 cases of TCC suggested that CD44v6 expression level was linked to a favorable outcome in muscle-invasive tumors, in contrast to superficial tumors, where CD44s emerged as an independent prognostic factor [272]. In another study including a panel of 13 TCC cell lines and 55 TCC tumor tissues, the expression of CD44 and its isoforms, namely CD44s, CD44v5, v6, v7, v7-8 and v10 was assessed by IHQ/IF and flow cytometry [273]. The findings revealed an inverse correlation between the

expression of CD44v5 isoforms and tumor recurrence, while CD44v10 expression displayed a positive association with tumor recurrence [273]. Conversely, Hong *et al.* did not find a significant difference in survival or identified CD44v6 as significant prognostic factor for bladder tumor recurrence according to Cox's multiple variant regression analysis, as confirmed by later studies conducted by Chuang and colleagues [273, 274]. Also, in a recent study that assessed the CD44s expression in 38 specimens of endoscopically resected UC, it was found that there was no significant correlation between this specific isoform levels and tumor recurrence and progression [275]. On the other hand, some reports described the absence or loss of CD44v6 and CD44s as independent poor prognosis predictors of BlCa. For instance, in a comprehensive study including 410 tumors from BlCa patients, the researchers concluded that the absence of CD44v6 is associated with a 2.3-fold increased risk of tumor recurrence following TUR [276]. Moreover, this absence significantly influences the OS rates, as patients with tumors lacking these isoforms tend to experience a shorter median OS (30 months) compared to those expressing CD44v6 (75 months) [276]. Similarly, Toma *et al.* demonstrated that in context of non-invasive TCC, patients with tumors characterized by the loss of both CD44v6 and CD44 isoforms containing exon v3 (CD44v3) presented a significantly shorter RFS time [277]. On the same note, a study involving patients with superficial bladder tumors suggested that the loss of CD44s expression could complement established prognostic factors, and serve as a useful predictor of short-term tumor progression [256]. However, the prognostic role of CD44 and its isoforms in BlCa still remains controversial. In a recent meta-analysis and The Cancer Genome Atlas Program (TCGA) data, It was found that CD44 expression was not associated with tumor-specific survival, OS, or recurrence/relapse-free survival [278]. Nonetheless, it was notably linked to disease failure (HR = 2.912, 95% CI = 1.51-5.61) [279]. Furthermore, in another report aimed to investigate the expression of three stem cell markers, including CD44, as prognostic predictors in patients undergoing RC, it was found that CD44 expression was correlated with decreased cancer-specific survival. Nevertheless, its expression could not be considered as independent prognostic parameter for these patients [278].

The prognostic value of CD44 and its isoforms in BlCa has been refuted by conflicting results. This is mainly due to the lack of standardized detection methods of both CD44 and its isoforms. Namely, several studies resort to commercial antibodies against CD44 isoforms, neglecting that several isoforms can contain the immunogen present in the antibodies rather than only the intended

one, which extensively biases studies outcomes. In addition, the nomenclature conundrum of CD44 seriously hinders study interpretation and reproducibility. Altogether, these issues compromise CD44 reliability as a prognostic marker of BlCa.

3.2.2. Unravel the Diagnostic significance of CD44 and its isoforms in Bladder Cancer

Beyond its prognostic character, this glycoprotein and its isoforms also play a critical role in the diagnosis of BlCa, being promising biomarkers in the detection, assessment, and clinical management of this neoplasia. In 1995, Sugiyama and colleagues evaluated the expression of CD44s and CD44v6 in voided urine samples and tumor tissues from 47 patients with BlCa and 43 healthy individuals [280]. This study showed key findings including the discovery of high molecular weight CD44 isoforms (>160 KDa) in urine cell lysates from 75% of BlCa patients, conspicuously absent in the control group. Moreover, immunohistochemical analysis revealed that normal bladder mucosa exhibited reduced CD44 isoforms expression, whereas carcinoma cells consistently overexpressed both CD44s and CD44v6 throughout their progression, including shedding into the urine. Therefore, the researchers concluded that the atypical presence of CD44 proteins in exfoliated cancer cells holds potential as a diagnostic biomarker for the non-invasive detection of BlCa [280]. Subsequently, in a different work, a highly effective sandwich-ELISA assay was developed for the detection of CD44 protein in urothelial cancer cells from voided urine samples, introducing a novel and accurate diagnostic approach for the detection and monitoring BlCa [281]. This assay showed remarkable sensitivity and specificity when compared to conventional methods such as urine cytology, thereby emerging as promising tool for non-invasive BlCa detection. Importantly, this assay accurately diagnosed BlCa in 81.1% of patients with histologically confirmed tumors, while maintaining a 100% specificity rate, showcasing its potential to bring substantial clinical advantages [281]. Furthermore, in a study by Hattori *et al.,* the assessment of CD44v6 expression RNA through quantitative Real-Time PCR (qRT-PCR) in urine samples unveiled a significant higher levels of these isoforms in BlCa patients in comparison to healthy controls [282]. These findings highlight the potential of urinary CD44v6 total RNA as a non-invasive and precise marker for BlCa detection [282]. Contrastingly, a few authors have cast doubt on the usefulness of CD44 and its isoforms as diagnostic markers for identifying BlCa.

Müller and colleagues, for example, noted the presence of CD44 isoforms containing exon v2 (CD44v2) in both healthy and malignant urinary tissues, as well as in urine samples [283]. Their research ultimately led to the conclusion that these isoforms, at the protein level, do not offer significant diagnostic value for BlCa [283]. Additionally, another study investigated the role of CD44v5 in serum samples collected from male patients with urological carcinomas [284]. This research showed significantly lower CD44v5 concentrations in male patients with BlCa compared to healthy men. However, the authors did not identify any other statistically significant differences, concluding that serum CD44 proteins may not serve as valuable markers for urological neoplasms [284].

The CD44 potential as a non-invasive diagnostic marker for BlCa has been compromised by the poor cancer specificity of CD44 isoforms. However, several CD44 isoforms as CD44v6 and CD44s have a clear cancer-associated nature. Improvements in the discriminatory capacity of CD44 isoforms are expected to arise by exploring aberrant post-translational modifications of these proteoforms, giving rise to potential theragnostic applications.

3.2.3. Clinical value of CD44 in bladder cancer: from tumor differentiation to therapeutic response

CD44 expression has been correlated with tumor grade, differentiation status and prediction of therapeutic response in BlCa. Namely, a study by Kuncová *et al*. suggested that CD44 expression patterns can potentially be used to categorize TCC according to their differentiation status [285]. A pattern of positive staining for CD44v6 and CD44s, extending toward the luminal surface was identified in well-differentiated tumors (grades 1 and 2). Poorly differentiated tumors (grades 3-4) exhibited heterogeneous expression of these isoforms [285]. Another report corroborated these findings by demonstrating a positive correlation between CD44v6 and CD44s expression and increasing grade and stage of TCC [286]. Invasive TCC exhibited higher CD44 expression levels than normal urothelium, while SCC showed lower CD44 expression [286]. Conversely, the absence of CD44v6 emerged as an effective discriminator between poorly differentiated UC and SCC. A report employed immunohistochemical staining targeting isoforms containing the exon v6 on SCC and TCC specimens. CD44v6 was absent in 93% of SCC, in sharp contrast to TCC, where CD44v6 was present in all cases [287]. Moreover, Wu *et al.* established a correlation between CD44

overexpression and increased clinical tumor stage and lymph node involvement in patients with MIBC treated with chemoradiotherapy [288]. Also, this study, employing *in vitro* and *in vivo* experiments, demonstrated that CD44-positive cells exposed to radiation exhibited higher radioresistance, showing reduced radiation-induced cell death compared to CD44-negative cells. This suggests that CD44 may have functional implications in BlCa radioresistance, while predicting tumor response to radiotherapy [288]. In fact, CD44 has been associated with increased chemoresistance in various cancers, including BlCa [177, 289-292]. A study has highlighted the potential of CD44v8-10 isoform expression as a novel biomarker for identifying BlCa patients with resistance to cisplatin chemotherapy [292]. The research revealed a significant increase in CD44v8–10 expression in cisplatin-resistant T24 cells compared to the parental T24 cell line, underscoring the correlation between CD44v8–10 expression and the development of cisplatin resistance in BlCa cells. Notably, silencing CD44v8-10 in cisplatin-resistant T24 cells led to the restoration of sensitivity to cisplatin, suggesting a potential therapeutic strategy for overcoming cisplatin resistance in BlCa [292].

There had been few attempts to correlate CD44 isoforms expression and tumors morphological status. Most studies focused on CD44s and CD44v6 isoforms, highlighting a possible correlation with more differentiated and advanced tumors stages; however, more studies are required to validate these findings.

3.2.4. CD44 impact on bladder cancer progression

Several reports in the literature emphasize the involvement of CD44 and its isoforms in driving progression, migration, invasion, and metastasis in BlCa. Indeed, some of these malignancy-related processes rely on the close interaction between CD44 and its classical ligands, such as OPN and HA, triggering the activation of specific intracellular signaling pathways. In accordance with this, Ahmed et al. demonstrated that OPN secreted by macrophages binds to CD44s on BlCa cells, promoting invasion, clonal growth, and metastasis [293]. These malignant events require the binding of CD44s to TIAM1, subsequently activating Rac1. Therefore, targeting the OPN-CD44s/TIAM1/Rac1 axis could present a promising therapeutic avenue for BlCa treatment [293]. Moreover, down-regulation of Hyaluronic Acid Synthase-1 (HAS1) in BlCa cells, an enzyme responsible for HA production, has resulted in a reduction in CD44 variant isoforms, specifically CD44v3, CD44v6, and CD44v8-10, both at the protein and mRNA levels [294]. The

introduction of HA led to an increase in CD44 expression and, consequently, enhanced BlCa cell growth [294]. Similarly, Anand and colleagues evaluated the expression of CD44, HA and Hyaluronic Acid Synthase-2 (HAS2) in a cohort of 50 patients undergoing RC for UC [295]. Increased levels of HA, HAS2, and CD44 was observed in UC patients, and these expressions were found to be correlated with the severity of the disease. Moreover, the study demonstrated that silencing CD44v3 in HT1376 cell line resulted in cell cycle arrest, apoptosis, and downregulation of pAkt, pERK, and pSTAT3 signaling [295]. These studies highlight the regulatory role of HA in modulating CD44 expression, offering potential for targeted interventions. In contrast, other report described that the overexpression of CD44v8-10 in UM-UC-3 BlCa cells reduced their ability to bind to HA when compared to control cells, with no impact on *in vitro* growth rates [255]. Nevertheless, CD44v8-10-overexpressing cells exhibited a significant increase in tumor growth when injected into athymic nude mice. Furthermore, these cells showed enhanced lymph node metastasis after orthotopic injection, in comparison to UM-UC-3/C control cells, suggesting the role of CD44v8-10 in potentiating malignant progression [255]. This demonstrates that CD44 isoforms functional impact has a tumor microenvironment dependence, urging the employment of translational models to fully disclose CD44 impact on BlCa progression.

Several reports highlight the impact of proinflammatory cytokines in BlCa pathogenesis, including in the regulation and maintenance of cancer stem cells (CSCs) properties [296-298]. Accordingly, an attempt to investigate the correlation CSCs marker CD44 and interleukins expression has been ongoing. A study using both *in vitro* experiments and orthotopic tumor models showed that CD44-positive BlCa cells presented higher invasion capacity and increased epithelial-mesenchymal transition (EMT) compared with CD44-negative cells [299]. Moreover, these authors revealed a significant correlation between Interleukin-6 (IL-6) and CD44 levels, *in vitro* and in clinical samples. Notably, blocking IL-6 led to reduced CD44 expression, and diminished cancer stem-cell-like properties and tumor aggressiveness *in vitro* and *in vivo* [299]. Another cytokine identified to interact with CD44 is CXCL5. A recent study demonstrated, using a lung metastatic BlCa cell line (T24-L cells), that knocking down CD44 expression led to inhibited invasion and migration [300]. However, the most significant impact was observed when both CXCL5 and CD44 were simultaneously knocked down; not only were invasion and migration abilities reduced, but there was also a decrease in CSCs characteristics. This suggests an interaction between CXCL5 and CD44, with both molecules

playing roles in influencing the EMT and stem cell capacity of T24-L cells [300]. Additionally, Zhu *et al.* reported an interaction between autophagy-related gene 7 (ATG7) and CD44 in the invasion and lung metastasis of human BlCa cells [301]. The authors observed that the knockdown of ATG7 led to a reduction in CD44s levels, enhancing the stem-like characteristics of human BlCa cells. Intriguingly, their findings demonstrated that artificially introducing CD44s restored invasive and metastatic capabilities in BlCa cells where ATG7 was suppressed. This suggests that CD44s functions not only as a biomarker for CSCs but also acts as a ATG7 downstream regulator of invasion and metastasis of BlCa cells [301]. In another study focusing on MicroRNA-34a (miR-34a) in BlCa, the authors reported that miR-34a plays a suppressive role in metastasis and angiogenesis by directly targeting CD44 [302]. The report also suggests that the functions mediated by CD44 in BlCa cells can be reversed by miR-34a, offering a potential therapeutic approach to hamper human BlCa metastasis [302]. Furthermore, the silencing of CD44v9 splicing isoforms in HT1376 and 5637 BlCa cells resulted in elevated E-cadherin levels and reduced expression of N-cadherin, snail, and slug proteins associated with EMT [268]. Also, this knockdown significantly decreased the migration and invasion capacity of human BlCa cells [268].

The complex interplay between CD44 and other key oncogenic molecular players underscores its multifaceted role in influencing cell cycle, invasion, metastasis, and the maintenance of CSCs-like properties in BlCa. While the specific interactions and outcomes vary in different experimental setups, the cumulative evidence supports the notion that targeting the CD44 axis holds significant therapeutic potential. However, as mentioned above, CD44 is not exclusive to cancer cells, diminishing its suitability for targeted therapeutics with less deleterious effects on normal tissues. Exploring this protein through its glycoproteoforms presents a promising avenue to overcome this limitation, thereby paving the way for the development of highly targeted and cancer-specific therapies.

3.3. Exploring CD44 glycosylation for cancer-specific therapies

3.3.1. The impact of glycosylation on CD44 biological functions

Deciphering the glycosylation patterns on CD44 has become a hot topic in cancer research, marked by significant advancements [243]. As previously mentioned, CD44 exhibits both *N*- and *O*-glycans mostly composed of sulfated,

sialylated, and Lewis antigens moieties. Several studies have investigated the influence of CD44 glycosylation in the context of cancer [173, 175, 303-305]. Given the highly glycosylated nature of CD44, it is expected that under non-homeostatic conditions this protein constitutes an important carrier of altered glycosylation, which can potentially alter its function, while providing therapeutic targets for intervention. Namely, CD44 *N*-glycosylation, for instance at Asn25 [306] and Asn120 [307], regulates its binding to HA, with potential impact on proliferation, motility, invasion and chemoresistance [308, 309]. On the same note, the level of sialylation on *N*-glycans also alters the binding capacity of CD44 to HA by generating additional repulsive negatively charged moieties [310, 311]. CD44 *N*-glycosylation also modulates its interaction with other hub proteins to regulate tumor progression. To state a representative example, the interaction between TM4SF5 and CD44 and consequent activation of the c-Src/STAT3/TWIST1/Bmi1 pathway occurs through *N*-glycan modification of the extracellular domain of both proteins to govern tumor cell migration and anticancer drug resistance [312, 313]. Of note, distinct *N*-glycosylation patterns of CD44 were found in NMIBC compared to MIBC with potential implications in disease progression [314, 315]. This highlights that targeting aberrant *N*-glycosylation on CD44 may offer the necessary means to hamper pro-tumor protein interaction and effectively inhibit tumor progression. These findings suggest that CD44 *N*-glycan/HA axis as well as the CD44 *N*-glycan/TM4SF5 axis constitute potential targets for therapeutic intervention. On a similar note, CD44 terminal *O*-glycan sialic acids can regulate its interactions and binding to sialic acid-binding immunoglobulin-type lectins (SIGLECs) on infiltrating immune cells, as siglec-15 [316]. This interaction results in prevention of CD44 lysosome-mediated degradation [317], which impacts on cell migration, while probably negatively impacting anti-tumor immune responses, given the immune-inhibitory role of SIGLECs on antigen presenting cells. Moreover, CD44 *O*-glycosylation was also found to regulate cancer aggressiveness. Namely, increased levels of CD44 carrying truncated *O*-glycans, as the STn antigen, have been found increased in gastric cancer patients compared to controls [318, 319]. In addition, blockade of CD44 STn has precluded dendritic cells immune tolerance to BlCa cells, reinforcing that targeting of CD44 glycosylation constitutes a promising strategy to circumvent tumor-induced tolerogenic mechanisms [179]. Also, CD44 *O*-glycans truncation and Tn antigen overexpression have been implicated in changes in tumor stem cell characteristics in pancreatic cancer by altering *NANOG* expression [173]. Fucosylation modifications of CD44 protein has

also been implicated in improving stemness of mouse bone marrow mesenchymal stem cells [320, 321]. These findings highlight that both *O*-glycan shortening and increased terminal fucosylation may contribute to the functional character of CD44 as a stem cell marker.

Of note, despite the extensive body of literature implicating CD44 expression alterations as well as aberrant glycosylation in BlCa, the study of CD44 associated glycome in the context of BlCa remains a poorly explored field. However, evidence collected from other models indicate that the interaction between CD44 and HA is influenced by the amount of *N*-glycans and sialic acids on CD44 glycosites, with increased levels negatively impacting binding [309, 310]. Notwithstanding, the role of *O*-glycosylation on this interaction requires further investigation. CD44's role typically involves associating with other membrane proteins or the extracellular matrix. *N*- and *O*-glycosylation on the extracellular domain of CD44 either cover or form the respective binding sites, thereby influencing the regulation of the tumor microenvironment and intracellular signal transduction. Thereby, targeting this glycoprotein aberrant glycosylation or specific aberrant glycoproteoforms of CD44 may constitute a valuable strategy for clinical intervention. In this context, the exploration of cancer glyconeoantigens for the development of cancer vaccines and targeted therapies holds promise for enhancing the effectiveness and specificity of tumor targeting.

3.3.2. CD44 glycoprofile for targeted interventions

Effective targeted therapeutics and immunotherapy, including cancer vaccines, is dependent on the identification of cancer neoantigens [322]. Glyconeoantigens identification relies on complex structural information provided by high-throughput mass spectrometry for the identification of glycosites number, location, density, and complexity [27]. To date, CD44 *N*-glycosites are more extensively studied and well-known, potentially due to their relatively lower abundance on CD44 compared to *O*-glycosylation sites. Han et al., employing mass spectrometry-based techniques, revealed five conserved N-glycosites located at CD44s amino-terminal (N25, N57, N100, N110 and N120) [323]. Moreover, quantitative analysis highlighted the prevalence of non-sialylated, fucosylated complex-type *N*-glycans in CD44s. The site-specific *N*-glycan distribution indicated that most glycosites carry complex-type glycans, with the exception of glycosite N100, which features high-mannose-type *N*-glycans [323]. Although this study

focused on the CD44s isoform, the *N*-glycosylation patterns identified herein may be applicable across all potential isoforms of CD44. This is because the *N*-glycosylation sites are located in the conserved extracellular region, the HA-binding domain (HABD) [324]. Regarding *O*-glycosylation patterns of CD44 and its isoforms, there is limited information available in the literature. Employing a sialoglycopeptide trapping approach, Halim and colleagues identified CD44 glycopeptides exhibiting simple *O*-glycans on either Thr-637 or Thr-638 (Net-*O*-Glyc predicts Thr-638) in urine of healthy male individuals [325]. Furthermore, another study, using glycoengineered chronic myelogenous leukemia, colonic adenocarcinoma, and pancreatic adenocarcinoma SimpleCells expressing truncated *O*-glycosylation, detected a total of 22 *O*-glycopeptides of CD44, representing a minimum of 39 occupied *O*-glycosites [151]. Notably, these researchers reported robust glycoproteomics data revealing *O*-glycopeptides from CD44 carrying glycans at Tyr-560 along with Ser-653, Thr-654, Thr-655, Thr-661, Ser-662, and Thr-667 [151]. Additionally, Campos *et al*. identified a CD44 glycopeptide (IPVTSAKTGSF^{584}), carrying aberrant STn, in gastric cancer patients serum, and not found in healthy controls [319]. This glycopeptide, located in variable exon v9 of CD44, also includes glycosites identified in gastric cancer simple cells [319]. Moreover, all four potential Ser/Thr glycosylation sites within this peptide have been described as glycosylated in another study, which also used glycoproteomics-based techniques in pancreatic adenocarcinoma SimpleCells [326]. This highlights the potential pancarcinomic nature of targetable CD44 glycopeptides. However, significantly more structural studies are required to identify specific glycopeptides not found in healthy cells, opening an avenue for targeted interventions. Moreover, it is expected that different CD44 isoforms possess distinct associated glycomes given rise to increasingly complex CD44 associated glycoproteomes. In turn, these can be both tissue and physiological context-dependent, possibly precluding of the shelf therapeutic options targeting CD44 glycoproteoforms. Also, the majority of structural studies exploring cellular *O*-glycoproteome have relied on the SimpleCell strategies. However, the glycoproteomics analytical workflows employed in simple cells may serve as a foundation for future approaches exploring more complex samples, as non-simplified cellular models and tissues.

Based on molecular information on CD44 glycosignatures in prostate and skin cancer, both monoclonal antibodies and CAR-T cells immunotherapies were proposed. Namely, the monoclonal antibody F77 against glycosylated CD44v10 has

been demonstrated to induce CD44-dependent prostate cancer cells apoptosis, providing a new strategy to control disease spread [315]. However this antibody was produced after immunization of mice with the androgen-independent prostate cancer cell line PC3, which expands antibody recognition to several other glycosylated lipid and proteomic signatures, hampering its usefulness as an antibody based targeted therapy [327]. In line with this, a cancer-specific monoclonal antibody (mAb), 4C8, selectively binding Tn-glycosylated CD44v6 gave rise to the development of Chimeric Antigen Receptor T cells (CAR-Ts) [328]. These were able to selectively kill target cells in a mixed organotypic skin cancer model overexpressing CD44v6 without affecting healthy keratinocytes [328]. This targeted approach has demonstrated remarkable selectivity, being highly tolerable *in vivo* [328]. These therapeutic tools showcase the outstanding potential of CD44 glycoproteoforms for precision oncology. The pre-clinical success of such approaches reinforces the importance of addressing CD44 glycoproteoforms in BlCa, also characterized by a profound deregulation of glycosylation pathways able to foster cancer specific glycosignatures. Notwithstanding, BlCa remains an orphan disease regarding CD44 glycoproteoforms-targeted interventions, urging the establishment of CD44 glycoprofile for novel immunotherapies.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71: 209-49.
2. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. 2020.
3. Cumberbatch MGK, Jubber I, Black PC, Esperto F, Figueroa JD, Kamat AM, et al. Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018. Eur Urol. 2018; 74: 784-95.
4. Pesch B, Gawrych K, Rabstein S, Weiss T, Casjens S, Rihs HP, et al. N-acetyltransferase 2 phenotype, occupation, and bladder cancer risk: results from the EPIC cohort. Cancer Epidemiol Biomarkers Prev. 2013; 22: 2055-65.
5. Lukas C, Selinski S, Prager HM, Blaszkewicz M, Hengstler JG, Golka K. Occupational bladder cancer: Polymorphisms of xenobiotic metabolizing enzymes, exposures, and prognosis. J Toxicol Environ Health A. 2017; 80: 439-52.
6. An Y, Li H, Wang KJ, Liu XH, Qiu MX, Liao Y, et al. Meta-analysis of the relationship between slow acetylation of N-acetyl transferase 2 and the risk of bladder cancer. Genet Mol Res. 2015; 14: 16896-904.
7. Witjes JA, Bruins HM, Cathomas R, Comperat EM, Cowan NC, Gakis G, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. Eur Urol. 2021; 79: 82-104.
8. Freedman ND, Silverman DT, Hollenbeck AR, Schatzkin A, Abnet CC. Association between smoking and risk of bladder cancer among men and women. JAMA. 2011; 306: 737-45.
9. Wong MCS, Fung FDH, Leung C, Cheung WWL, Goggins WB, Ng CF. The global epidemiology of bladder cancer: a joinpoint regression analysis of its incidence and mortality trends and projection. Sci Rep. 2018; 8: 1129.
10. Ishida K, Hsieh MH. Understanding Urogenital Schistosomiasis-Related Bladder Cancer: An Update. Front Med (Lausanne). 2018; 5: 223.
11. Babjuk M, Burger M, Capoun O, Cohen D, Comperat EM, Dominguez Escrig JL, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). Eur Urol. 2022; 81: 75-94.

12. Moschini M, Zaffuto E, Karakiewicz PI, Andrea DD, Foerster B, Abufaraj M, et al. External Beam Radiotherapy Increases the Risk of Bladder Cancer When Compared with Radical Prostatectomy in Patients Affected by Prostate Cancer: A Population-based Analysis. Eur Urol. 2019; 75: 319-28.
13. Knight A, Askling J, Granath F, Sparen P, Ekbom A. Urinary bladder cancer in Wegener's granulomatosis: risks and relation to cyclophosphamide. Ann Rheum Dis. 2004; 63: 1307-11.
14. Lokeshwar SD, Lopez M, Sarcan S, Aguilar K, Morera DS, Shaheen DM, et al. Molecular Oncology of Bladder Cancer from Inception to Modern Perspective. Cancers (Basel). 2022; 14.
15. Ramirez D, Gupta A, Canter D, Harrow B, Dobbs RW, Kucherov V, et al. Microscopic haematuria at time of diagnosis is associated with lower disease stage in patients with newly diagnosed bladder cancer. BJU Int. 2016; 117: 783-6.
16. Sanli O, Dobruch J, Knowles MA, Burger M, Alemozaffar M, Nielsen ME, et al. Bladder cancer. Nat Rev Dis Primers. 2017; 3: 17022.
17. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder Cancer: A Review. JAMA. 2020; 324: 1980-91.
18. Kamat AM, Hahn NM, Efstathiou JA, Lerner SP, Malmstrom PU, Choi W, et al. Bladder cancer. Lancet. 2016; 388: 2796-810.
19. Xiong Y, Li J, Ma S, Ge J, Zhou L, Li D, et al. A meta-analysis of narrow band imaging for the diagnosis and therapeutic outcome of non-muscle invasive bladder cancer. PLoS One. 2017; 12: e0170819.
20. Burger M, Grossman HB, Droller M, Schmidbauer J, Hermann G, Dragoescu O, et al. Photodynamic diagnosis of non-muscle-invasive bladder cancer with hexaminolevulinate cystoscopy: a meta-analysis of detection and recurrence based on raw data. Eur Urol. 2013; 64: 846-54.
21. Wong VK, Ganeshan D, Jensen CT, Devine CE. Imaging and Management of Bladder Cancer. Cancers (Basel). 2021; 13.
22. Hilton S, Jones LP. Recent advances in imaging cancer of the kidney and urinary tract. Surg Oncol Clin N Am. 2014; 23: 863-910.
23. Sugeeta SS, Sharma A, Ng K, Nayak A, Vasdev N. Biomarkers in Bladder Cancer Surveillance. Front Surg. 2021; 8: 735868.
24. Ng K, Stenzl A, Sharma A, Vasdev N. Urinary biomarkers in bladder cancer: A review of the current landscape and future directions. Urol Oncol. 2021; 39: 41-51.

25. Ahmadi H, Duddalwar V, Daneshmand S. Diagnosis and Staging of Bladder Cancer. Hematol Oncol Clin North Am. 2021; 35: 531-41.
26. Tran L, Xiao JF, Agarwal N, Duex JE, Theodorescu D. Advances in bladder cancer biology and therapy. Nat Rev Cancer. 2021; 21: 104-21.
27. Ferreira JA, Relvas-Santos M, Peixoto A, A MNS, Lara Santos L. Glycoproteogenomics: Setting the Course for Next-generation Cancer Neoantigen Discovery for Cancer Vaccines. Genomics Proteomics Bioinformatics. 2021; 19: 25-43.
28. Kantor AF, Hartge P, Hoover RN, Fraumeni JF, Jr. Epidemiological characteristics of squamous cell carcinoma and adenocarcinoma of the bladder. Cancer Res. 1988; 48: 3853-5.
29. Saginala K, Barsouk A, Aluru JS, Rawla P, Padala SA, Barsouk A. Epidemiology of Bladder Cancer. Med Sci (Basel). 2020; 8.
30. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO Classification of tumors of the Urinary System and Male Genital Organs-Part B: Prostate and Bladder tumors Eur Urol. 2016; 70: 106-19.
31. Brierley J, Gospodarowicz MK, Wittekind C. TNM classification of malignant tumors Eighth edition. ed. Chichester, West Sussex, UK ;: John Wiley & Sons, Inc.; 2017.
32. Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer. 2015; 15: 25-41.
33. Billerey C, Chopin D, Aubriot-Lorton MH, Ricol D, Gil Diez de Medina S, Van Rhijn B, et al. Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. Am J Pathol. 2001; 158: 1955-9.
34. Dinney CP, McConkey DJ, Millikan RE, Wu X, Bar-Eli M, Adam L, et al. Focus on bladder cancer. Cancer Cell. 2004; 6: 111-6.
35. Garczyk S, Bischoff F, Schneider U, Golz R, von Rundstedt FC, Knüchel R, et al. Intratumoral heterogeneity of surrogate molecular subtypes in urothelial carcinoma in situ of the urinary bladder: implications for prognostic stratification of high-risk non-muscle-invasive bladder cancer. Virchows Arch. 2021; 479: 325-35.
36. Volkmer JP, Sahoo D, Chin RK, Ho PL, Tang C, Kurtova AV, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. Proc Natl Acad Sci U S A. 2012; 109: 2078-83.
37. Lima L, Neves M, Oliveira MI, Dieguez L, Freitas R, Azevedo R, et al. Sialyl-Tn identifies muscle-invasive bladder cancer basal and luminal subtypes facing

decreased survival, being expressed by circulating tumor cells and metastases. Urol Oncol. 2017; 35: 675 e1- e8.

38. Kamoun A, de Reyniès A, Allory Y, Sjödahl G, Robertson AG, Seiler R, et al. A Consensus Molecular Classification of Muscle-invasive Bladder Cancer. Eur Urol. 2020; 77: 420-33.

39. Warrick JI, Sjödahl G, Kaag M, Raman JD, Merrill S, Shuman L, et al. Intratumoral Heterogeneity of Bladder Cancer by Molecular Subtypes and Histologic Variants. Eur Urol. 2019; 75: 18-22.

40. Mooi JK, Wirapati P, Asher R, Lee CK, Savas P, Price TJ, et al. The prognostic impact of consensus molecular subtypes (CMS) and its predictive effects for bevacizumab benefit in metastatic colorectal cancer: molecular analysis of the AGITG MAX clinical trial. Ann Oncol. 2018; 29: 2240-6.

41. Stintzing S, Wirapati P, Lenz HJ, Neureiter D, Fischer von Weikersthal L, Decker T, et al. Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. Ann Oncol. 2019; 30: 1796-803.

42. Kim LHC, Patel MI. Transurethral resection of bladder tumour (TURBT). Transl Androl Urol. 2020; 9: 3056-72.

43. Suarez-Ibarrola R, Soria F, Abufaraj M, D'Andrea D, Preto M, Gust KM, et al. Surgical checklist impact on recurrence-free survival of patients with non-muscle-invasive bladder cancer undergoing transurethral resection of bladder tumour. BJU Int. 2019; 123: 646-50.

44. Oosterlinck W, Kurth KH, Schroder F, Bultinck J, Hammond B, Sylvester R. A prospective European Organization for Research and Treatment of Cancer Genitourinary Group randomized trial comparing transurethral resection followed by a single intravesical instillation of epirubicin or water in single stage Ta, T1 papillary carcinoma of the bladder. J Urol. 1993; 149: 749-52.

45. Gakis G. Management of Muscle-invasive Bladder Cancer in the 2020s: Challenges and Perspectives. Eur Urol Focus. 2020; 6: 632-8.

46. Donin NM, Lenis AT, Holden S, Drakaki A, Pantuck A, Belldegrun A, et al. Immunotherapy for the Treatment of Urothelial Carcinoma. J Urol. 2017; 197: 14-22.

47. James ND, Hussain SA, Hall E, Jenkins P, Tremlett J, Rawlings C, et al. Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. N Engl J Med. 2012; 366: 1477-88.

48. Spiess PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, Clark PE, et al. Bladder Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2017; 15: 1240-67.

49. Cambier S, Sylvester RJ, Collette L, Gontero P, Brausi MA, van Andel G, et al. EORTC Nomograms and Risk Groups for Predicting Recurrence, Progression, and Disease-specific and Overall Survival in Non-Muscle-invasive Stage Ta-T1 Urothelial Bladder Cancer Patients Treated with 1-3 Years of Maintenance Bacillus Calmette-Guérin. Eur Urol. 2016; 69: 60-9.

50. Gontero P, Sylvester R, Pisano F, Joniau S, Vander Eeckt K, Serretta V, et al. Prognostic factors and risk groups in T1G3 non-muscle-invasive bladder cancer patients initially treated with Bacillus Calmette-Guérin: results of a retrospective multicenter study of 2451 patients. Eur Urol. 2015; 67: 74-82.

51. Martin-Doyle W, Leow JJ, Orsola A, Chang SL, Bellmunt J. Improving selection criteria for early cystectomy in high-grade t1 bladder cancer: a meta-analysis of 15,215 patients. J Clin Oncol. 2015; 33: 643-50.

52. Sylvester RJ, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffioux C, Denis L, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. Eur Urol. 2006; 49: 466-5; discussion 75-7.

53. Ritch CR, Velasquez MC, Kwon D, Becerra MF, Soodana-Prakash N, Atluri VS, et al. Use and Validation of the AUA/SUO Risk Grouping for Nonmuscle Invasive Bladder Cancer in a Contemporary Cohort. J Urol. 2020; 203: 505-11.

54. Jurewicz M, Soloway MS. Approaching the optimal transurethral resection of a bladder tumor. Turk J Urol. 2014; 40: 73-7.

55. Chamie K, Ballon-Landa E, Bassett JC, Daskivich TJ, Leventhal M, Deapen D, et al. Quality of diagnostic staging in patients with bladder cancer: a process-outcomes link. Cancer. 2015; 121: 379-85.

56. Cumberbatch MGK, Foerster B, Catto JWF, Kamat AM, Kassouf W, Jubber I, et al. Repeat Transurethral Resection in Non-muscle-invasive Bladder Cancer: A Systematic Review. Eur Urol. 2018; 73: 925-33.

57. Arrabal Polo M, Melgarejo Segura MT, Yáñez Castillo Y, Morales Martínez A, Pareja Vílchez M, Arrabal Martín M. Adjuvant intravesical treatment in patients with intermediate and high-risk non-muscle-invasive bladder cancer with BCG versus MMC applied with COMBAT or EMDA. Results of a prospective study. J Cancer Res Clin Oncol. 2023; 149: 7453-9.

58. Huncharek M, Geschwind JF, Witherspoon B, McGarry R, Adcock D. Intravesical chemotherapy prophylaxis in primary superficial bladder cancer: a meta-analysis of 3703 patients from 11 randomized trials. J Clin Epidemiol. 2000; 53: 676-80.

59. Chang SS, Boorjian SA, Chou R, Clark PE, Daneshmand S, Konety BR, et al. Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline. J Urol. 2016; 196: 1021-9.

60. Sylvester RJ, Oosterlinck W, van der Meijden AP. A single immediate postoperative instillation of chemotherapy decreases the risk of recurrence in patients with stage Ta T1 bladder cancer: a meta-analysis of published results of randomized clinical trials. J Urol. 2004; 171: 2186-90, quiz 435.

61. Perlis N, Zlotta AR, Beyene J, Finelli A, Fleshner NE, Kulkarni GS. Immediate post-transurethral resection of bladder tumor intravesical chemotherapy prevents non-muscle-invasive bladder cancer recurrences: an updated meta-analysis on 2548 patients and quality-of-evidence review. Eur Urol. 2013; 64: 421-30.

62. Doherty AP, Trendell-Smith N, Stirling R, Rogers H, Bellringer J. Perivesical fat necrosis after adjuvant intravesical chemotherapy. BJU Int. 1999; 83: 420-3.

63. Messing EM, Tangen CM, Lerner SP, Sahasrabudhe DM, Koppie TM, Wood DP, Jr., et al. Effect of Intravesical Instillation of Gemcitabine vs Saline Immediately Following Resection of Suspected Low-Grade Non-Muscle-Invasive Bladder Cancer on Tumor Recurrence: SWOG S0337 Randomized Clinical Trial. JAMA. 2018; 319: 1880-8.

64. Matulewicz RS, Steinberg GD. Non-muscle-invasive Bladder Cancer: Overview and Contemporary Treatment Landscape of Neoadjuvant Chemoablative Therapies. Rev Urol. 2020; 22: 43-51.

65. Steinberg RL, Thomas LJ, O'Donnell MA. Combination Intravesical Chemotherapy for Non-muscle-invasive Bladder Cancer. Eur Urol Focus. 2018; 4: 503-5.

66. Steinberg RL, Thomas LJ, O'Donnell MA, Nepple KG. Sequential Intravesical Gemcitabine and Docetaxel for the Salvage Treatment of Non-Muscle Invasive Bladder Cancer. Bladder Cancer. 2015; 1: 65-72.

67. DeCastro GJ, Sui W, Pak JS, Lee SM, Holder D, Kates MM, et al. A Phase I Trial of Intravesical Cabazitaxel, Gemcitabine and Cisplatin for the Treatment of Nonmuscle Invasive bacillus Calmette-Guerin Unresponsive or Recurrent/Relapsing Urothelial Carcinoma of the Bladder. J Urol. 2020; 204: 247-53.

68. Mukherjee N, Julián E, Torrelles JB, Svatek RS. Effects of Mycobacterium bovis Calmette et Guérin (BCG) in oncotherapy: Bladder cancer and beyond. Vaccine. 2021; 39: 7332-40.

69. Malmstrom PU, Sylvester RJ, Crawford DE, Friedrich M, Krege S, Rintala E, et al. An individual patient data meta-analysis of the long-term outcome of randomised studies comparing intravesical mitomycin C versus bacillus Calmette-Guerin for non-muscle-invasive bladder cancer. Eur Urol. 2009; 56: 247-56.

70. Shelley MD, Kynaston H, Court J, Wilt TJ, Coles B, Burgon K, et al. A systematic review of intravesical bacillus Calmette-Guérin plus transurethral resection vs transurethral resection alone in Ta and T1 bladder cancer. BJU Int. 2001; 88: 209-16.

71. Bohle A, Bock PR. Intravesical bacille Calmette-Guerin versus mitomycin C in superficial bladder cancer: formal meta-analysis of comparative studies on tumor progression. Urology. 2004; 63: 682-6; discussion 6-7.

72. Sylvester RJ, van der MA, Lamm DL. Intravesical bacillus Calmette-Guerin reduces the risk of progression in patients with superficial bladder cancer: a meta-analysis of the published results of randomized clinical trials. J Urol. 2002; 168: 1964-70.

73. Zlotta AR, Fleshner NE, Jewett MA. The management of BCG failure in non-muscle-invasive bladder cancer: an update. Can Urol Assoc J. 2009; 3: S199-205.

74. Han J, Gu X, Li Y, Wu Q. Mechanisms of BCG in the treatment of bladder cancer-current understanding and the prospect. Biomed Pharmacother. 2020; 129: 110393.

75. Steinberg RL, Thomas LJ, Brooks N, Mott SL, Vitale A, Crump T, et al. Multi-Institution Evaluation of Sequential Gemcitabine and Docetaxel as Rescue Therapy for Nonmuscle Invasive Bladder Cancer. J Urol. 2020; 203: 902-9.

76. Steinberg G, Bahnson R, Brosman S, Middleton R, Wajsman Z, Wehle M. Efficacy and safety of valrubicin for the treatment of Bacillus Calmette-Guerin refractory carcinoma in situ of the bladder. The Valrubicin Study Group. J Urol. 2000; 163: 761-7.

77. Balar AV, Kamat AM, Kulkarni GS, Uchio EM, Boormans JL, Roumiguié M, et al. Pembrolizumab monotherapy for the treatment of high-risk non-muscle-invasive bladder cancer unresponsive to BCG (KEYNOTE-057): an open-label, single-arm, multicentre, phase 2 study. Lancet Oncol. 2021; 22: 919-30.

78. Al Hussein Al Awamlh B, Chang SS. Novel Therapies for High-Risk Non-Muscle Invasive Bladder Cancer. Curr Oncol Rep. 2023; 25: 83-91.

79. Lee A. Nadofaragene Firadenovec: First Approval. Drugs. 2023; 83: 353-7.
80. Grandi P, Darilek A, Moscu A, Pradhan A, Li R. Intravesical Infusion of Oncolytic Virus CG0070 in the Treatment of Bladder Cancer. Methods Mol Biol. 2023; 2684: 303-17.
81. Chen W, Liu N, Yuan Y, Zhu M, Hu X, Hu W, et al. ALT-803 in the treatment of non-muscle-invasive bladder cancer: Preclinical and clinical evidence and translational potential. Front Immunol. 2022; 13: 1040669.
82. Kim HS, Seo HK. Emerging treatments for bacillus Calmette-Guérin-unresponsive non-muscle-invasive bladder cancer. Investig Clin Urol. 2021; 62: 361-77.
83. Yamamoto Y, Kawashima A, Morishima T, Uemura T, Yamamoto A, Yamamichi G, et al. Comparative Effectiveness of Radiation Versus Radical Cystectomy for Localized Muscle-Invasive Bladder Cancer. Adv Radiat Oncol. 2023; 8: 101157.
84. Haas CR, Barlow LJ, Badalato GM, DeCastro GJ, Benson MC, McKiernan JM. The Timing of Radical Cystectomy for bacillus Calmette-Guérin Failure: Comparison of Outcomes and Risk Factors for Prognosis. J Urol. 2016; 195: 1704-9.
85. Fritsche HM, Burger M, Svatek RS, Jeldres C, Karakiewicz PI, Novara G, et al. Characteristics and outcomes of patients with clinical T1 grade 3 urothelial carcinoma treated with radical cystectomy: results from an international cohort. Eur Urol. 2010; 57: 300-9.
86. Chalasani V, Kassouf W, Chin JL, Fradet Y, Aprikian AG, Fairey AS, et al. Radical cystectomy for the treatment of T1 bladder cancer: the Canadian Bladder Cancer Network experience. Can Urol Assoc J. 2011; 5: 83-7.
87. Aminoltejari K, Black PC. Radical cystectomy: a review of techniques, developments and controversies. Transl Androl Urol. 2020; 9: 3073-81.
88. Stein JP, Lieskovsky G, Cote R, Groshen S, Feng AC, Boyd S, et al. Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1,054 patients. J Clin Oncol. 2001; 19: 666-75.
89. Abol-Enein H, Tilki D, Mosbah A, El-Baz M, Shokeir A, Nabeeh A, et al. Does the extent of lymphadenectomy in radical cystectomy for bladder cancer influence disease-free survival? A prospective single-center study. Eur Urol. 2011; 60: 572-7.
90. Gschwend JE, Heck MM, Lehmann J, Rubben H, Albers P, Wolff JM, et al. Extended Versus Limited Lymph Node Dissection in Bladder Cancer Patients Undergoing Radical Cystectomy: Survival Results from a Prospective, Randomized Trial. Eur Urol. 2019; 75: 604-11.

91. Katims AB, Bochner BH. Extended pelvic lymph node dissection in muscle invasive bladder cancer. Curr Opin Urol. 2023; 33: 252-7.
92. Shabsigh A, Korets R, Vora KC, Brooks CM, Cronin AM, Savage C, et al. Defining early morbidity of radical cystectomy for patients with bladder cancer using a standardized reporting methodology. Eur Urol. 2009; 55: 164-74.
93. Quek ML, Stein JP, Daneshmand S, Miranda G, Thangathurai D, Roffey P, et al. A critical analysis of perioperative mortality from radical cystectomy. J Urol. 2006; 175: 886-9; discussion 9-90.
94. Knoedler JJ, Boorjian SA, Kim SP, Weight CJ, Thapa P, Tarrell RF, et al. Does partial cystectomy compromise oncologic outcomes for patients with bladder cancer compared to radical cystectomy? A matched case-control analysis. J Urol. 2012; 188: 1115-9.
95. Ploussard G, Daneshmand S, Efstathiou JA, Herr HW, James ND, Rodel CM, et al. Critical analysis of bladder sparing with trimodal therapy in muscle-invasive bladder cancer: a systematic review. Eur Urol. 2014; 66: 120-37.
96. Tholomier C, Souhami L, Kassouf W. Bladder-sparing protocols in the treatment of muscle-invasive bladder cancer. Transl Androl Urol. 2020; 9: 2920-37.
97. Peyton CC, Tang D, Reich RR, Azizi M, Chipollini J, Pow-Sang JM, et al. Downstaging and Survival Outcomes Associated With Neoadjuvant Chemotherapy Regimens Among Patients Treated With Cystectomy for Muscle-Invasive Bladder Cancer. JAMA Oncol. 2018; 4: 1535-42.
98. Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, Trump DL, et al. Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. N Engl J Med. 2003; 349: 859-66.
99. Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration. Eur Urol. 2005; 48: 202-5; discussion 5-6.
100. Plimack ER, Hoffman-Censits JH, Viterbo R, Trabulsi EJ, Ross EA, Greenberg RE, et al. Accelerated methotrexate, vinblastine, doxorubicin, and cisplatin is safe, effective, and efficient neoadjuvant treatment for muscle-invasive bladder cancer: results of a multicenter phase II study with molecular correlates of response and toxicity. J Clin Oncol. 2014; 32: 1895-901.

101. Dash A, Pettus JAt, Herr HW, Bochner BH, Dalbagni G, Donat SM, et al. A role for neoadjuvant gemcitabine plus cisplatin in muscle-invasive urothelial carcinoma of the bladder: a retrospective experience. Cancer. 2008; 113: 2471-7.
102. Choueiri TK, Jacobus S, Bellmunt J, Qu A, Appleman LJ, Tretter C, et al. Neoadjuvant dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin with pegfilgrastim support in muscle-invasive urothelial cancer: pathologic, radiologic, and biomarker correlates. J Clin Oncol. 2014; 32: 1889-94.
103. Yin M, Joshi M, Meijer RP, Glantz M, Holder S, Harvey HA, et al. Neoadjuvant Chemotherapy for Muscle-Invasive Bladder Cancer: A Systematic Review and Two-Step Meta-Analysis. Oncologist. 2016; 21: 708-15.
104. International Collaboration of T, Medical Research Council Advanced Bladder Cancer Working P, European Organisation for R, Treatment of Cancer Genito-Urinary Tract Cancer G, Australian Bladder Cancer Study G, National Cancer Institute of Canada Clinical Trials G, et al. International phase III trial assessing neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: long-term results of the BA06 30894 trial. J Clin Oncol. 2011; 29: 2171-7.
105. Del Bene G, Calabro F, Giannarelli D, Plimack ER, Harshman LC, Yu EY, et al. Neoadjuvant vs. Adjuvant Chemotherapy in Muscle Invasive Bladder Cancer (MIBC): Analysis From the RISC Database. Front Oncol. 2018; 8: 463.
106. Motterle G, Andrews JR, Morlacco A, Karnes RJ. Predicting Response to Neoadjuvant Chemotherapy in Bladder Cancer. Eur Urol Focus. 2020; 6: 642-9.
107. Neoadjuvant chemotherapy for invasive bladder cancer. Cochrane Database Syst Rev. 2005; 2004: Cd005246.
108. Leow JJ, Martin-Doyle W, Rajagopal PS, Patel CG, Anderson EM, Rothman AT, et al. Adjuvant chemotherapy for invasive bladder cancer: a 2013 updated systematic review and meta-analysis of randomized trials. Eur Urol. 2014; 66: 42-54.
109. Advanced Bladder Cancer Meta-analysis Collaborators G. Adjuvant Chemotherapy for Muscle-invasive Bladder Cancer: A Systematic Review and Meta-analysis of Individual Participant Data from Randomised Controlled Trials. Eur Urol. 2022; 81: 50-61.
110. Sternberg CN, Skoneczna I, Kerst JM, Albers P, Fossa SD, Agerbaek M, et al. Immediate versus deferred chemotherapy after radical cystectomy in patients with pT3-pT4 or N+ M0 urothelial carcinoma of the bladder (EORTC 30994): an intergroup, open-label, randomised phase 3 trial. Lancet Oncol. 2015; 16: 76-86.

111. Bayoumi Y, Heikal T, Darweish H. Survival benefit of adjuvant radiotherapy in stage III and IV bladder cancer: results of 170 patients. Cancer Manag Res. 2014; 6: 459-65.

112. Sternberg CN, de Mulder P, Schornagel JH, Theodore C, Fossa SD, van Oosterom AT, et al. Seven year update of an EORTC phase III trial of high-dose intensity M-VAC chemotherapy and G-CSF versus classic M-VAC in advanced urothelial tract tumors Eur J Cancer. 2006; 42: 50-4.

113. von der Maase H, Hansen SW, Roberts JT, Dogliotti L, Oliver T, Moore MJ, et al. Gemcitabine and cisplatin versus methotrexate, vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large, randomized, multinational, multicenter, phase III study. J Clin Oncol. 2000; 18: 3068-77.

114. Park I, Lee JL. Systemic treatment for advanced urothelial cancer: an update on recent clinical trials and current treatment options. Korean J Intern Med. 2020; 35: 834-53.

115. Sternberg CN, de Mulder PH, Schornagel JH, Theodore C, Fossa SD, van Oosterom AT, et al. Randomized phase III trial of high-dose-intensity methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC) chemotherapy and recombinant human granulocyte colony-stimulating factor versus classic MVAC in advanced urothelial tract tumors: European Organization for Research and Treatment of Cancer Protocol no. 30924. J Clin Oncol. 2001; 19: 2638-46.

116. Freshwater T, Li H, Valiathan C, Li M, Perini R, Bracco OL, et al. Systematic Literature Review and Meta-Analysis of Response to First-Line Therapies for Advanced/Metastatic Urothelial Cancer Patients Who Are Cisplatin Ineligible. Am J Clin Oncol. 2019; 42: 802-9.

117. Mertens LS, Meijer RP, Kerst JM, Bergman AM, van Tinteren H, van Rhijn BW, et al. Carboplatin based induction chemotherapy for nonorgan confined bladder cancer--a reasonable alternative for cisplatin unfit patients? J Urol. 2012; 188: 1108-13.

118. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016; 387: 1909-20.

119. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally

advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017; 389: 67-76.

120. Lopez-Beltran A, Cimadamore A, Blanca A, Massari F, Vau N, Scarpelli M, et al. Immune Checkpoint Inhibitors for the Treatment of Bladder Cancer. Cancers (Basel). 2021; 13.

121. Nakanishi J, Wada Y, Matsumoto K, Azuma M, Kikuchi K, Ueda S. Overexpression of B7-H1 (PD-L1) significantly associates with tumor grade and postoperative prognosis in human urothelial cancers. Cancer Immunol Immunother. 2007; 56: 1173-82.

122. Inman BA, Sebo TJ, Frigola X, Dong H, Bergstralh EJ, Frank I, et al. PD-L1 (B7-H1) expression by urothelial carcinoma of the bladder and BCG-induced granulomata: associations with localized stage progression. Cancer. 2007; 109: 1499-505.

123. Suzman DL, Agrawal S, Ning YM, Maher VE, Fernandes LL, Karuri S, et al. FDA Approval Summary: Atezolizumab or Pembrolizumab for the Treatment of Patients with Advanced Urothelial Carcinoma Ineligible for Cisplatin-Containing Chemotherapy. Oncologist. 2019; 24: 563-9.

124. Sharma P, Retz M, Siefker-Radtke A, Baron A, Necchi A, Bedke J, et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. Lancet Oncol. 2017; 18: 312-22.

125. Powles T, O'Donnell PH, Massard C, Arkenau HT, Friedlander TW, Hoimes CJ, et al. Efficacy and Safety of Durvalumab in Locally Advanced or Metastatic Urothelial Carcinoma: Updated Results From a Phase 1/2 Open-label Study. JAMA Oncol. 2017; 3: e172411.

126. Apolo AB, Infante JR, Balmanoukian A, Patel MR, Wang D, Kelly K, et al. Avelumab, an Anti-Programmed Death-Ligand 1 Antibody, In Patients With Refractory Metastatic Urothelial Carcinoma: Results From a Multicenter, Phase Ib Study. J Clin Oncol. 2017; 35: 2117-24.

127. Fradet Y, Bellmunt J, Vaughn DJ, Lee JL, Fong L, Vogelzang NJ, et al. Randomized phase III KEYNOTE-045 trial of pembrolizumab versus paclitaxel, docetaxel, or vinflunine in recurrent advanced urothelial cancer: results of >2 years of follow-up. Ann Oncol. 2019; 30: 970-6.

128. van der Heijden MS, Loriot Y, Duran I, Ravaud A, Retz M, Vogelzang NJ, et al. Atezolizumab Versus Chemotherapy in Patients with Platinum-treated Locally Advanced or Metastatic Urothelial Carcinoma: A Long-term Overall Survival and Safety Update from the Phase 3 IMvigor211 Clinical Trial. Eur Urol. 2021; 80: 7-11.

129. Cathomas R, Lorch A, Bruins HM, Comperat EM, Cowan NC, Efstathiou JA, et al. The 2021 Updated European Association of Urology Guidelines on Metastatic Urothelial Carcinoma. Eur Urol. 2022; 81: 95-103.
130. Liu J, Gao Y, Song C, Liao W, Meng L, Yang S, et al. Immunotherapeutic prospects and progress in bladder cancer. J Cell Mol Med. 2024; 28: e18101.
131. Chang E, Weinstock C, Zhang L, Charlab R, Dorff SE, Gong Y, et al. FDA Approval Summary: Enfortumab Vedotin for Locally Advanced or Metastatic Urothelial Carcinoma. Clin Cancer Res. 2021; 27: 922-7.
132. Rosenberg JE, O'Donnell PH, Balar AV, McGregor BA, Heath EI, Yu EY, et al. Pivotal Trial of Enfortumab Vedotin in Urothelial Carcinoma After Platinum and Anti-Programmed Death 1/Programmed Death Ligand 1 Therapy. J Clin Oncol. 2019; 37: 2592-600.
133. Yu EY, Petrylak DP, O'Donnell PH, Lee JL, van der Heijden MS, Loriot Y, et al. Enfortumab vedotin after PD-1 or PD-L1 inhibitors in cisplatin-ineligible patients with advanced urothelial carcinoma (EV-201): a multicentre, single-arm, phase 2 trial. Lancet Oncol. 2021; 22: 872-82.
134. Powles T, Rosenberg JE, Sonpavde GP, Loriot Y, Duran I, Lee JL, et al. Enfortumab Vedotin in Previously Treated Advanced Urothelial Carcinoma. N Engl J Med. 2021; 384: 1125-35.
135. Rosenberg JE, Flaig TW, Friedlander TW, Milowsky MI, Srinivas S, Petrylak DP, et al. Study EV-103: Preliminary durability results of enfortumab vedotin plus pembrolizumab for locally advanced or metastatic urothelial carcinoma. Journal of Clinical Oncology. 2020; 38: 441-.
136. Loriot Y, Necchi A, Park SH, Garcia-Donas J, Huddart R, Burgess E, et al. Erdafitinib in Locally Advanced or Metastatic Urothelial Carcinoma. N Engl J Med. 2019; 381: 338-48.
137. Pal SK, Rosenberg JE, Hoffman-Censits JH, Berger R, Quinn DI, Galsky MD, et al. Efficacy of BGJ398, a Fibroblast Growth Factor Receptor 1-3 Inhibitor, in Patients with Previously Treated Advanced Urothelial Carcinoma with FGFR3 Alterations. Cancer Discov. 2018; 8: 812-21.
138. Balar AV, Castellano D, O'Donnell PH, Grivas P, Vuky J, Powles T, et al. First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. Lancet Oncol. 2017; 18: 1483-92.
139. Apolo AB, Ellerton JA, Infante JR, Agrawal M, Gordon MS, Aljumaily R, et al. Avelumab as second-line therapy for metastatic, platinum-treated urothelial

carcinoma in the phase Ib JAVELIN Solid Tumor study: 2-year updated efficacy and safety analysis. J Immunother Cancer. 2020; 8.

140. Smith CC, Selitsky SR, Chai S, Armistead PM, Vincent BG, Serody JS. Alternative tumour-specific antigens. Nat Rev Cancer. 2019; 19: 465-78.

141. Mnatsakanyan R, Shema G, Basik M, Batist G, Borchers CH, Sickmann A, et al. Detecting post-translational modification signatures as potential biomarkers in clinical mass spectrometry. Expert Rev Proteomics. 2018; 15: 515-35.

142. Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell. 2018; 34: 211-24.e6.

143. Wirth TC, Kühnel F. Neoantigen Targeting-Dawn of a New Era in Cancer Immunotherapy? Front Immunol. 2017; 8: 1848.

144. Bobalova J, Strouhalova D, Bobal P. Common Post-translational Modifications (PTMs) of Proteins: Analysis by Up-to-Date Analytical Techniques with an Emphasis on Barley. J Agric Food Chem. 2023; 71: 14825-37.

145. Oo HZ, Seiler R, Black PC, Daugaard M. Post-translational modifications in bladder cancer: Expanding the tumor target repertoire. Urol Oncol. 2020; 38: 858-66.

146. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep. 2011; 1.

147. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol. 2019; 15: 346-66.

148. Cummings RD. The repertoire of glycan determinants in the human glycome. Mol Biosyst. 2009; 5: 1087-104.

149. Joshi HJ, Hansen L, Narimatsu Y, Freeze HH, Henrissat B, Bennett E, et al. Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases. Glycobiology. 2018; 28: 284-94.

150. Varki A. Biological roles of glycans. Glycobiology. 2016; 27: 3-49.

151. Vakhrushev SY, Steentoft C, Vester-Christensen MB, Bennett EP, Clausen H, Levery SB. Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. Mol Cell Proteomics. 2013; 12: 932-44.

152. Halim A, Brinkmalm G, Ruetschi U, Westman-Brinkmalm A, Portelius E, Zetterberg H, et al. Site-specific characterization of threonine, serine, and tyrosine

glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. Proc Natl Acad Sci U S A. 2011; 108: 11848-53.

153. Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, et al. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. Nat Methods. 2011; 8: 977-82.

154. Schwarz F, Aebi M. Mechanisms and principles of N-linked protein glycosylation. Curr Opin Struct Biol. 2011; 21: 576-82.

155. Bieberich E. Synthesis, Processing, and Function of N-glycans in N-glycoproteins. Adv Neurobiol. 2014; 9: 47-70.

156. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY); 2009.

157. Azevedo R, Peixoto A, Gaiteiro C, Fernandes E, Neves M, Lima L, et al. Over forty years of bladder cancer glycobiology: Where do glycans stand facing precision oncology? Oncotarget. 2017; 8: 91734-64.

158. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol. 2019; 9: 380.

159. Magalhães A, Duarte HO, Reis CA. The role of O-glycosylation in human disease. Mol Aspects Med. 2021; 79: 100964.

160. Vosseller K, Wells L, Hart GW. Nucleocytoplasmic O-glycosylation: O-GlcNAc and functional proteomics. Biochimie. 2001; 83: 575-81.

161. Zhang Y, Sun L, Lei C, Li W, Han J, Zhang J, et al. A Sweet Warning: Mucin-Type O-Glycans in Cancer. Cells. 2022; 11.

162. Bansil R, Turner BS. The biology of mucus: Composition, synthesis and organization. Adv Drug Deliv Rev. 2018; 124: 3-15.

163. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology. 2012; 22: 736-56.

164. Brockhausen I, Wandall HH, Hagen KGT, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. 4th ed. Cold Spring Harbor (NY); 2022. p. 117-28.

165. Ju T, Aryal RP, Kudelka MR, Wang Y, Cummings RD. The Cosmc connection to the Tn antigen in cancer. Cancer Biomark. 2014; 14: 63-81.

166. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015; 15: 540-55.

167. Bellis SL, Reis CA, Varki A, Kannagi R, Stanley P. Glycosylation Changes in Cancer. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. 4th ed. Cold Spring Harbor (NY); 2022. p. 631-44.
168. Brockhausen I. Biosynthesis and functions of O-glycans and regulation of mucin antigen expression in cancer. Biochem Soc Trans. 1997; 25: 871-4.
169. Holgersson J, Löfling J. Glycosyltransferases involved in type 1 chain and Lewis antigen biosynthesis exhibit glycan and core chain specificity. Glycobiology. 2006; 16: 584-93.
170. Miyoshi E, Moriwaki K, Nakagawa T. Biological function of fucosylation in cancer biology. J Biochem. 2008; 143: 725-9.
171. Ferreira JA, Magalhaes A, Gomes J, Peixoto A, Gaiteiro C, Fernandes E, et al. Protein glycosylation in gastric and colorectal cancers: Toward cancer detection and targeted therapeutics. Cancer Lett. 2017; 387: 32-45.
172. Cascio S, Farkas AM, Hughey RP, Finn OJ. Altered glycosylation of MUC1 influences its association with CIN85: the role of this novel complex in cancer cell invasion and migration. Oncotarget. 2013; 4: 1686-97.
173. Leon F, Seshacharyulu P, Nimmakayala RK, Chugh S, Karmakar S, Nallasamy P, et al. Reduction in O-glycome induces differentially glycosylated CD44 to promote stemness and metastasis in pancreatic cancer. Oncogene. 2022; 41: 57-71.
174. Huang T, Meng F, Huang H, Wang L, Wang L, Liu Y, et al. GALNT8 suppresses breast cancer cell metastasis potential by regulating EGFR O-GalNAcylation. Biochem Biophys Res Commun. 2022; 601: 16-23.
175. Mereiter S, Martins Á M, Gomes C, Balmaña M, Macedo JA, Polom K, et al. O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. FEBS Lett. 2019; 593: 1675-89.
176. Tajadura-Ortega V, Gambardella G, Skinner A, Halim A, Van Coillie J, Schjoldager KTG, et al. O-linked mucin-type glycosylation regulates the transcriptional programme downstream of EGFR. Glycobiology. 2021; 31: 200-10.
177. Hu Z, Gao J, Zhang D, Liu Q, Yan L, Gao L, et al. High expression of Lewis y antigen and CD44 is correlated with resistance to chemotherapy in epithelial ovarian cancers. PLoS One. 2013; 8: e57250.
178. Wilczak M, Surman M, Przybyło M. Altered Glycosylation in Progression and Management of Bladder Cancer. Molecules. 2023; 28.

179. Carrascal MA, Severino PF, Guadalupe Cabral M, Silva M, Ferreira JA, Calais F, et al. Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. Mol Oncol. 2014; 8: 753-65.
180. Lynch TP, Ferrer CM, Jackson SR, Shahriari KS, Vosseller K, Reginato MJ. Critical role of O-Linked β-N-acetylglucosamine transferase in prostate cancer invasion, angiogenesis, and metastasis. J Biol Chem. 2012; 287: 11070-81.
181. Jian Y, Xu Z, Xu C, Zhang L, Sun X, Yang D, et al. The Roles of Glycans in Bladder Cancer. Front Oncol. 2020; 10: 957.
182. Cotton S, Azevedo R, Gaiteiro C, Ferreira D, Lima L, Peixoto A, et al. Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumors Mol Oncol. 2017; 11: 895-912.
183. Lima L, Severino PF, Silva M, Miranda A, Tavares A, Pereira S, et al. Response of high-risk of recurrence/progression bladder tumors expressing sialyl-Tn and sialyl-6-T to BCG immunotherapy. Br J Cancer. 2013; 109: 2106-14.
184. Bergman S, Javadpour N. The cell surface antigen A, B or O(H) as an indicator of malignant potential in stage A bladder carcinoma: preliminary report. J Urol. 1978; 119: 49-51.
185. Decenzo JM, Howard P, Irish CE. Antigenic deletion and prognosis of patients with stage A transitional cell bladder carcinoma. J Urol. 1975; 114: 874-8.
186. Newman AJ, Jr., Carlton CE, Jr., Johnson S. Cell surface A, B, or O(H) blood group antigens as an indicator of malignant potential in stage A bladder carcinoma. J Urol. 1980; 124: 27-9.
187. Abel PD, Thorpe SJ, Williams G. Blood group antigen expression in frozen sections of presenting bladder cancer: 3-year prospective follow-up of prognostic value. Br J Urol. 1989; 63: 171-5.
188. Yamada T, Fukui I, Kobayashi T, Sekine H, Yokogawa M, Yamada T, et al. The relationship of ABH(O) blood group antigen expression in intraepithelial dysplastic lesions to clinicopathologic properties of associated transitional cell carcinoma of the bladder. Cancer. 1991; 67: 1661-6.
189. Orntoft TF, Wolf H. Blood group ABO and Lewis antigens in bladder tumors: correlation between glycosyltransferase activity and antigen expression. APMIS Suppl. 1988; 4: 126-33.
190. Chihara Y, Sugano K, Kobayashi A, Kanai Y, Yamamoto H, Nakazono M, et al. Loss of blood group A antigen expression in bladder cancer caused by allelic loss and/or methylation of the ABO gene. Lab Invest. 2005; 85: 895-907.

191. Sheinfeld J, Reuter VE, Fair WR, Cordon-Cardo C. Expression of blood group antigens in bladder cancer: current concepts. Semin Surg Oncol. 1992; 8: 308-15.
192. Cordon-Cardo C, Reuter VE, Lloyd KO, Sheinfeld J, Fair WR, Old LJ, et al. Blood group-related antigens in human urothelium: enhanced expression of precursor, LeX, and LeY determinants in urothelial carcinoma. Cancer Res. 1988; 48: 4113-20.
193. Pode D, Golijanin D, Sherman Y, Lebensart P, Shapiro A. Immunostaining of Lewis X in cells from voided urine, cytopathology and ultrasound for noninvasive detection of bladder tumors. J Urol. 1998; 159: 389-92; discussion 93.
194. Friedrich MG, Hellstern A, Hautmann SH, Graefen M, Conrad S, Huland E, et al. Clinical use of urinary markers for the detection and prognosis of bladder carcinoma: a comparison of immunocytology with monoclonal antibodies against Lewis X and 486p3/12 with the BTA STAT and NMP22 tests. J Urol. 2002; 168: 470-4.
195. Dal Moro F, Valotto C, Guttilla A, Zattoni F. Urinary markers in the everyday diagnosis of bladder cancer. Urologia. 2013; 80: 265-75.
196. Numahata K, Satoh M, Handa K, Saito S, Ohyama C, Ito A, et al. Sialosyl-Le(x) expression defines invasive and metastatic properties of bladder carcinoma. Cancer. 2002; 94: 673-85.
197. Kajiwara H, Yasuda M, Kumaki N, Shibayama T, Osamura Y. Expression of carbohydrate antigens (SSEA-1, sialyl-Lewis X, DU-PAN-2 and CA19-9) and E-selectin in urothelial carcinoma of the renal pelvis, ureter, and urinary bladder. Tokai J Exp Clin Med. 2005; 30: 177-82.
198. Jha DK, Mittal A, Gupta SP, Sathian B. Significance of CA19-9 in predicting the prognosis of urothelial carcinoma: a hospital based study from Nepal. Asian Pac J Cancer Prev. 2013; 14: 4067-9.
199. Pal K, Roy S, Mondal SA, Chatterjee U, Tiwari P, Bera M. Urinary level of CA19-9 as a tumor marker in urothelial carcinoma of the bladder. Urol J. 2011; 8: 203-8.
200. St Hill CA. Interactions between endothelial selectins and cancer cells regulate metastasis. Front Biosci (Landmark Ed). 2011; 16: 3233-51.
201. Kohler S, Ullrich S, Richter U, Schumacher U. E-/P-selectins and colon carcinoma metastasis: first in vivo evidence for their crucial role in a clinically relevant model of spontaneous metastasis formation in the lung. Br J Cancer. 2010; 102: 602-9.
202. Dobie C, Skropeta D. Insights into the role of sialylation in cancer progression and metastasis. Br J Cancer. 2021; 124: 76-90.

203. Julien S, Videira PA, Delannoy P. Sialyl-tn in cancer: (how) did we miss the target? Biomolecules. 2012; 2: 435-66.
204. Munkley J. The Role of Sialyl-Tn in Cancer. Int J Mol Sci. 2016; 17: 275.
205. Ferreira JA, Videira PA, Lima L, Pereira S, Silva M, Carrascal M, et al. Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumors Mol Oncol. 2013; 7: 719-31.
206. Broxterman HJ, Lankelma J, Hoekman K. Resistance to cytotoxic and anti-angiogenic anticancer agents: similarities and differences. Drug Resist Updat. 2003; 6: 111-27.
207. Peixoto A, Freitas R, Ferreira D, Relvas-Santos M, Paulo P, Cardoso M, et al. Metabolomics, Transcriptomics and Functional Glycomics Reveals Bladder Cancer Cells Plasticity and Enhanced Aggressiveness Facing Hypoxia and Glucose Deprivation. bioRxiv. 2021: 2021.02.14.431133.
208. Freitas D, Campos D, Gomes J, Pinto F, Macedo JA, Matos R, et al. O-glycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. EBioMedicine. 2019; 40: 349-62.
209. Marcos NT, Pinho S, Grandela C, Cruz A, Samyn-Petit B, Harduin-Lepers A, et al. Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. Cancer Res. 2004; 64: 7050-7.
210. Peixoto A, Fernandes E, Gaiteiro C, Lima L, Azevedo R, Soares J, et al. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget. 2016; 7: 63138-57.
211. Peixoto A, Ferreira D, Azevedo R, Freitas R, Fernandes E, Relvas-Santos M, et al. Glycoproteomics identifies HOMER3 as a potentially targetable biomarker triggered by hypoxia and glucose deprivation in bladder cancer. J Exp Clin Cancer Res. 2021; 40: 191.
212. Costa C, Pereira S, Lima L, Peixoto A, Fernandes E, Neves D, et al. Abnormal Protein Glycosylation and Activated PI3K/Akt/mTOR Pathway: Role in Bladder Cancer Prognosis and Targeted Therapeutics. PLoS One. 2015; 10: e0141253.
213. Langkilde NC, Wolf H, Clausen H, Kjeldsen T, Orntoft TF. Nuclear volume and expression of T-antigen, sialosyl-Tn-antigen, and Tn-antigen in carcinoma of the human bladder. Relation to tumor recurrence and progression. Cancer. 1992; 69: 219-27.
214. Limas C, Lange P. T-antigen in normal and neoplastic urothelium. Cancer. 1986; 58: 1236-45.

215. Yokoyama M, Ohoka H, Oda H, Oda T, Utsumi S, Takeuchi M. Thomsen-Friedenreich antigen in bladder cancer tissues detected by monoclonal antibody. Hinyokika Kiyo. 1988; 34: 255-8.
216. Videira PA, Correia M, Malagolini N, Crespo HJ, Ligeiro D, Calais FM, et al. ST3Gal.I sialyltransferase relevance in bladder cancer tissues and cell lines. BMC Cancer. 2009; 9: 357.
217. Suzuki Y, Sutoh M, Hatakeyama S, Mori K, Yamamoto H, Koie T, et al. MUC1 carrying core 2 O-glycans functions as a molecular shield against NK cell attack, promoting bladder tumor metastasis. Int J Oncol. 2012; 40: 1831-8.
218. Tsuboi S, Sutoh M, Hatakeyama S, Hiraoka N, Habuchi T, Horikawa Y, et al. A novel strategy for evasion of NK cell immunity by tumors expressing core2 O-glycans. EMBO J. 2011; 30: 3173-85.
219. Berois N, Pittini A, Osinaga E. Targeting Tumor Glycans for Cancer Therapy: Successes, Limitations, and Perspectives. Cancers (Basel). 2022; 14.
220. Posey AD, Jr., Schwab RD, Boesteanu AC, Steentoft C, Mandel U, Engels B, et al. Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. Immunity. 2016; 44: 1444-54.
221. Freitas R, Miranda A, Ferreira D, Relvas-Santos M, Castro F, Ferreira E, et al. A multivalent CD44 glycoconjugate vaccine candidate for cancer immunotherapy. J Control Release. 2024; 367: 540-56.
222. Lo-Man R, Vichier-Guerre S, Perraut R, Deriaud E, Huteau V, BenMohamed L, et al. A fully synthetic therapeutic vaccine candidate targeting carcinoma-associated Tn carbohydrate antigen induces tumor-specific antibodies in nonhuman primates. Cancer Res. 2004; 64: 4987-94.
223. Freire T, Zhang X, Deriaud E, Ganneau C, Vichier-Guerre S, Azria E, et al. Glycosidic Tn-based vaccines targeting dermal dendritic cells favor germinal center B-cell development and potent antibody response in the absence of adjuvant. Blood. 2010; 116: 3526-36.
224. Scheid E, Major P, Bergeron A, Finn OJ, Salter RD, Eady R, et al. Tn-MUC1 DC Vaccination of Rhesus Macaques and a Phase I/II Trial in Patients with Nonmetastatic Castrate-Resistant Prostate Cancer. Cancer Immunol Res. 2016; 4: 881-92.
225. Zhai X, You F, Xiang S, Jiang L, Chen D, Li Y, et al. MUC1-Tn-targeting chimeric antigen receptor-modified Vgamma9Vdelta2 T cells with enhanced antigen-specific anti-tumor activity. Am J Cancer Res. 2021; 11: 79-91.

226. Sharma P, Marada V, Cai Q, Kizerwetter M, He Y, Wolf SP, et al. Structure-guided engineering of the affinity and specificity of CARs against Tn-glycopeptides. Proc Natl Acad Sci U S A. 2020; 117: 15148-59.
227. Welinder C, Baldetorp B, Borrebaeck C, Fredlund BM, Jansson B. A new murine IgG1 anti-Tn monoclonal antibody with in vivo anti-tumor activity. Glycobiology. 2011; 21: 1097-107.
228. Peng M, Chu X, Peng Y, Li D, Zhang Z, Wang W, et al. Targeted therapies in bladder cancer: signaling pathways, applications, and challenges. MedComm (2020). 2023; 4: e455.
229. Azevedo R, Ferreira JA, Peixoto A, Neves M, Sousa N, Lima A, et al. Emerging antibody-based therapeutic strategies for bladder cancer: A systematic review. J Control Release. 2015; 214: 40-61.
230. Gallatin WM, Weissman IL, Butcher EC. A cell-surface molecule involved in organ-specific homing of lymphocytes. Nature. 1983; 304: 30-4.
231. Stamenkovic I, Amiot M, Pesando JM, Seed B. A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. Cell. 1989; 56: 1057-62.
232. Goldstein LA, Zhou DF, Picker LJ, Minty CN, Bargatze RF, Ding JF, et al. A human lymphocyte homing receptor, the hermes antigen, is related to cartilage proteoglycan core and link proteins. Cell. 1989; 56: 1063-72.
233. Goodfellow PN, Banting G, Wiles MV, Tunnacliffe A, Parkar M, Solomon E, et al. The gene, MIC4, which controls expression of the antigen defined by monoclonal antibody F10.44.2, is on human chromosome 11. Eur J Immunol. 1982; 12: 659-63.
234. Weng X, Maxwell-Warburton S, Hasib A, Ma L, Kang L. The membrane receptor CD44: novel insights into metabolism. Trends Endocrinol Metab. 2022; 33: 318-32.
235. Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clin Proteomics. 2018; 15: 22.
236. Colombatti A, Hughes EN, Taylor BA, August JT. Gene for a major cell surface glycoprotein of mouse macrophages and other phagocytic cells is on chromosome 2. Proc Natl Acad Sci U S A. 1982; 79: 1926-9.
237. Hassn Mesrati M, Syafruddin SE, Mohtar MA, Syahir A. CD44: A Multifunctional Mediator of Cancer Progression. Biomolecules. 2021; 11.

238. Misra S, Hascall VC, Markwald RR, Ghatak S. Interactions between Hyaluronan and Its Receptors (CD44, RHAMM) Regulate the Activities of Inflammation and Cancer. Front Immunol. 2015; 6: 201.
239. Weber GF, Ashkar S, Glimcher MJ, Cantor H. Receptor-ligand interaction between CD44 and osteopontin (Eta-1). Science. 1996; 271: 509-12.
240. Jalkanen S, Jalkanen M. Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin. J Cell Biol. 1992; 116: 817-25.
241. Hibino S, Shibuya M, Engbring JA, Mochizuki M, Nomizu M, Kleinman HK. Identification of an active site on the laminin alpha5 chain globular domain that binds to CD44 and inhibits malignancy. Cancer Res. 2004; 64: 4810-6.
242. Chen KL, Li D, Lu TX, Chang SW. Structural Characterization of the CD44 Stem Region for Standard and Cancer-Associated Isoforms. Int J Mol Sci. 2020; 21.
243. Liao C, Wang Q, An J, Chen J, Li X, Long Q, et al. CD44 Glycosylation as a Therapeutic Target in Oncology. Front Oncol. 2022; 12: 883831.
244. Mishra MN, Chandavarkar V, Sharma R, Bhargava D. Structure, function and role of CD44 in neoplasia. J Oral Maxillofac Pathol. 2019; 23: 267-72.
245. Kugelman LC, Ganguly S, Haggerty JG, Weissman SM, Milstone LM. The core protein of epican, a heparan sulfate proteoglycan on keratinocytes, is an alternative form of CD44. J Invest Dermatol. 1992; 99: 886-91.
246. Huet S, Groux H, Caillou B, Valentin H, Prieur AM, Bernard A. CD44 contributes to T cell activation. J Immunol. 1989; 143: 798-801.
247. Trochon V, Mabilat C, Bertrand P, Legrand Y, Smadja-Joffe F, Soria C, et al. Evidence of involvement of CD44 in endothelial cell proliferation, migration and angiogenesis in vitro. Int J Cancer. 1996; 66: 664-8.
248. Sneath RJ, Mangham DC. The normal structure and function of CD44 and its role in neoplasia. Mol Pathol. 1998; 51: 191-200.
249. Turi M, Anilkumar Sithara A, Hofmanová L, Žihala D, Radhakrishnan D, Vdovin A, et al. Transcriptome Analysis of Diffuse Large B-Cell Lymphoma Cells Inducibly Expressing MyD88 L265P Mutation Identifies Upregulated CD44, LGALS3, NFKBIZ, and BATF as Downstream Targets of Oncogenic NF-κB Signaling. Int J Mol Sci. 2023; 24.
250. Xu H, Wu K, Tian Y, Liu Q, Han N, Yuan X, et al. CD44 correlates with clinicopathological characteristics and is upregulated by EGFR in breast cancer. Int J Oncol. 2016; 49: 1343-50.
251. Kataki A, Giannakoulis VG, Derventzi A, Papiris K, Koniaris E, Konstadoulakis M. Membranous CD44v6 is upregulated as an early event in colorectal cancer:

Downregulation is associated with circulating tumor cells and poor prognosis. Oncol Lett. 2021; 22: 820.

252. Shang Z, Cai Q, Zhang M, Zhu S, Ma Y, Sun L, et al. A switch from CD44$^+$ cell to EMT cell drives the metastasis of prostate cancer. Oncotarget. 2015; 6: 1202-16.

253. Gao Y, Foster R, Yang X, Feng Y, Shen JK, Mankin HJ, et al. Up-regulation of CD44 in the development of metastasis, recurrence and drug resistance of ovarian cancer. Oncotarget. 2015; 6: 9313-26.

254. He Y, Xue C, Yu Y, Chen J, Chen X, Ren F, et al. CD44 is overexpressed and correlated with tumor progression in gallbladder cancer. Cancer Manag Res. 2018; 10: 3857-65.

255. Muramaki M, Miyake H, Kamidono S, Hara I. Over expression of CD44V8-10 in human bladder cancer cells decreases their interaction with hyaluronic acid and potentiates their malignant progression. J Urol. 2004; 171: 426-30.

256. Stavropoulos NE, Filliadis I, Ioachim E, Michael M, Mermiga E, Hastazeris K, et al. CD44 standard form expression as a predictor of progression in high risk superficial bladder tumors. Int Urol Nephrol. 2001; 33: 479-83.

257. Ouhtit A, Rizeq B, Saleh HA, Rahman MM, Zayed H. Novel CD44-downstream signaling pathways mediating breast tumor invasion. Int J Biol Sci. 2018; 14: 1782-90.

258. Primeaux M, Gowrikumar S, Dhawan P. Role of CD44 isoforms in epithelial-mesenchymal plasticity and metastasis. Clin Exp Metastasis. 2022; 39: 391-406.

259. Zhang H, Cao H, Luo H, Zhang N, Wang Z, Dai Z, et al. RUNX1/CD44 axis regulates the proliferation, migration, and immunotherapy of gliomas: A single-cell sequencing analysis. Front Immunol. 2023; 14: 1086280.

260. Boulbes DR, Chauhan GB, Jin Q, Bartholomeusz C, Esteva FJ. CD44 expression contributes to trastuzumab resistance in HER2-positive breast cancer cells. Breast Cancer Res Treat. 2015; 151: 501-13.

261. Liu S, Liu Z, Shang A, Xun J, Lv Z, Zhou S, et al. CD44 is a potential immunotherapeutic target and affects macrophage infiltration leading to poor prognosis. Sci Rep. 2023; 13: 9657.

262. Paulis YW, Huijbers EJ, van der Schaft DW, Soetekouw PM, Pauwels P, Tjan-Heijnen VC, et al. CD44 enhances tumor aggressiveness by promoting tumor cell plasticity. Oncotarget. 2015; 6: 19634-46.

263. Ludwig N, Szczepanski MJ, Gluszko A, Szafarowski T, Azambuja JH, Dolg L, et al. CD44(+) tumor cells promote early angiogenesis in head and neck squamous cell carcinoma. Cancer Lett. 2019; 467: 85-95.
264. Gao AC, Lou W, Dong JT, Isaacs JT. CD44 is a metastasis suppressor gene for prostatic cancer located on human chromosome 11p13. Cancer Res. 1997; 57: 846-9.
265. Herrlich P, Morrison H, Sleeman J, Orian-Rousseau V, Konig H, Weg-Remers S, et al. CD44 acts both as a growth- and invasiveness-promoting molecule and as a tumor-suppressing cofactor. Ann N Y Acad Sci. 2000; 910: 106-18; discussion 18-20.
266. Stamenkovic I, Aruffo A, Amiot M, Seed B. The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. EMBO J. 1991; 10: 343-8.
267. Gunthert U, Hofmann M, Rudy W, Reber S, Zoller M, Haussmann I, et al. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell. 1991; 65: 13-24.
268. Kobayashi K, Matsumoto H, Matsuyama H, Fujii N, Inoue R, Yamamoto Y, et al. Clinical significance of CD44 variant 9 expression as a prognostic indicator in bladder cancer. Oncol Rep. 2016; 36: 2852-60.
269. Ogihara K, Kikuchi E, Okazaki S, Hagiwara M, Takeda T, Matsumoto K, et al. Sulfasalazine could modulate the CD44v9-xCT system and enhance cisplatin-induced cytotoxic effects in metastatic bladder cancer. Cancer Sci. 2019; 110: 1431-41.
270. Hagiwara M, Kikuchi E, Kosaka T, Mikami S, Saya H, Oya M. Variant isoforms of CD44 expression in upper tract urothelial cancer as a predictive marker for recurrence and mortality. Urol Oncol. 2016; 34: 337 e19-26.
271. Miyake H, Eto H, Arakawa S, Kamidono S, Hara I. Over expression of CD44V8-10 in urinary exfoliated cells as an independent prognostic predictor in patients with urothelial cancer. J Urol. 2002; 167: 1282-7.
272. Lipponen P, Aaltoma S, Kosma VM, Ala-Opas M, Eskelinen M. Expression of CD44 standard and variant-v6 proteins in transitional cell bladder tumors and their relation to prognosis during a long-term follow-up. J Pathol. 1998; 186: 157-64.
273. Chuang CK, Liao SK. Differential expression of CD44 variant isoforms by cell lines and tissue specimens of transitional cell carcinomas. Anticancer Res. 2003; 23: 4635-9.

274. Hong RL, Pu YS, Hsieh TS, Chu JS, Lee WJ. Expressions of E-cadherin and exon v6-containing isoforms of CD44 and their prognostic values in human transitional cell carcinoma. J Urol. 1995; 153: 2025-8.
275. Bacha D, Ben Rejeb S, Saidani A, Dhaoui A, Khouni H, Bellil K. Prognostic value of CD44s expression in urothelial bladder tumors. Tunis Med. 2020; 98: 816-22.
276. Klatte T, Seligson DB, Rao JY, Yu H, de Martino M, Garraway I, et al. Absent CD44v6 expression is an independent predictor of poor urothelial bladder cancer outcome. J Urol. 2010; 183: 2403-8.
277. Toma V, Hauri D, Schmid U, Ackermann D, Maurer R, Alund G, et al. Focal loss of CD44 variant protein expression is related to recurrence in superficial bladder carcinoma. Am J Pathol. 1999; 155: 1427-32.
278. Hofner T, Macher-Goeppinger S, Klein C, Schillert A, Eisen C, Wagner S, et al. Expression and prognostic significance of cancer stem cell markers CD24 and CD44 in urothelial bladder cancer xenografts and patients undergoing radical cystectomy. Urol Oncol. 2014; 32: 678-86.
279. Hu Y, Zhang Y, Gao J, Lian X, Wang Y. The clinicopathological and prognostic value of CD44 expression in bladder cancer: a study based on meta-analysis and TCGA data. Bioengineered. 2020; 11: 572-81.
280. Sugiyama M, Woodman A, Sugino T, Crowley S, Ho K, Smith J, et al. Non-invasive detection of bladder cancer by identification of abnormal CD44 proteins in exfoliated cancer cells in urine. Clin Mol Pathol. 1995; 48: M142-7.
281. Woodman AC, Goodison S, Drake M, Noble J, Tarin D. Noninvasive diagnosis of bladder carcinoma by enzyme-linked immunosorbent assay detection of CD44 isoforms in exfoliated urothelia. Clin Cancer Res. 2000; 6: 2381-92.
282. Hattori S, Kojima K, Minoshima K, Yamaha M, Horie M, Sawamura T, et al. Detection of bladder cancer by measuring CD44v6 expression in urine with real-time quantitative reverse transcription polymerase chain reaction. Urology. 2014; 83: 1443 e9-15.
283. Muller M, Heicappell R, Habermann F, Kaufmann M, Steiner U, Miller K. Expression of CD44V2 in transitional cell carcinoma of the urinary bladder and in urine. Urol Res. 1997; 25: 187-92.
284. Lein M, Jung K, Weiss S, Schnorr D, Loening SA. Soluble CD44 variants in the serum of patients with urological malignancies. Oncology. 1997; 54: 226-30.
285. Kuncova J, Urban M, Mandys V. Expression of CD44s and CD44v6 in transitional cell carcinomas of the urinary bladder: comparison with tumour grade,

proliferative activity and p53 immunoreactivity of tumour cells. APMIS. 2007; 115: 1194-205.
286. Omran OM, Ata HS. CD44s and CD44v6 in diagnosis and prognosis of human bladder cancer. Ultrastruct Pathol. 2012; 36: 145-52.
287. Iczkowski KA, Shanks JH, Bostwick DG. Loss of CD44 variant 6 expression differentiates small cell carcinoma of urinary bladder from urothelial (transitional cell) carcinoma. Histopathology. 1998; 32: 322-7.
288. Wu CT, Lin WY, Chang YH, Chen WC, Chen MF. Impact of CD44 expression on radiation response for bladder cancer. J Cancer. 2017; 8: 1137-44.
289. Gerardo-Ramirez M, Keggenhoff FL, Giam V, Becker D, Groth M, Hartmann N, et al. CD44 Contributes to the Regulation of MDR1 Protein and Doxorubicin Chemoresistance in Osteosarcoma. Int J Mol Sci. 2022; 23.
290. Pothuraju R, Rachagani S, Krishn SR, Chaudhary S, Nimmakayala RK, Siddiqui JA, et al. Molecular implications of MUC5AC-CD44 axis in colorectal cancer progression and chemoresistance. Mol Cancer. 2020; 19: 37.
291. Jogo T, Oki E, Nakanishi R, Ando K, Nakashima Y, Kimura Y, et al. Expression of CD44 variant 9 induces chemoresistance of gastric cancer by controlling intracellular reactive oxygen spices accumulation. Gastric Cancer. 2021; 24: 1089-99.
292. Hagiwara M, Kikuchi E, Tanaka N, Kosaka T, Mikami S, Saya H, et al. Variant isoforms of CD44 involves acquisition of chemoresistance to cisplatin and has potential as a novel indicator for identifying a cisplatin-resistant population in urothelial cancer. BMC Cancer. 2018; 18: 113.
293. Ahmed M, Sottnik JL, Dancik GM, Sahu D, Hansel DE, Theodorescu D, et al. An Osteopontin/CD44 Axis in RhoGDI2-Mediated Metastasis Suppression. Cancer Cell. 2016; 30: 432-43.
294. Golshani R, Lopez L, Estrella V, Kramer M, Iida N, Lokeshwar VB. Hyaluronic acid synthase-1 expression regulates bladder cancer growth, invasion, and angiogenesis through CD44. Cancer Res. 2008; 68: 483-91.
295. Anand V, Khandelwal M, Appunni S, Gupta N, Seth A, Singh P, et al. CD44 splice variant (CD44v3) promotes progression of urothelial carcinoma of bladder through Akt/ERK/STAT3 pathways: novel therapeutic approach. J Cancer Res Clin Oncol. 2019; 145: 2649-61.
296. Zhu Z, Shen Z, Xu C. Inflammatory pathways as promising targets to increase chemotherapy response in bladder cancer. Mediators Inflamm. 2012; 2012: 528690.

297. Smyth MJ, Cretney E, Kershaw MH, Hayakawa Y. Cytokines in cancer immunity and immunotherapy. Immunol Rev. 2004; 202: 275-93.
298. Chen MF, Lin PY, Wu CF, Chen WC, Wu CT. IL-6 expression regulates tumorigenicity and correlates with prognosis in bladder cancer. PLoS One. 2013; 8: e61901.
299. Wu CT, Lin WY, Chen WC, Chen MF. Predictive Value of CD44 in Muscle-Invasive Bladder Cancer and Its Relationship with IL-6 Signaling. Ann Surg Oncol. 2018; 25: 3518-26.
300. Wang W, Zhang M, Huang Z, Wang L, Yue Y, Wang X, et al. Knockdown of CXCL5 inhibits the invasion, metastasis and stemness of bladder cancer lung metastatic cells by downregulating CD44. Anticancer Drugs. 2022; 33: e103-e12.
301. Zhu J, Huang G, Hua X, Li Y, Yan H, Che X, et al. CD44s is a crucial ATG7 downstream regulator for stem-like property, invasion, and lung metastasis of human bladder cancer (BC) cells. Oncogene. 2019; 38: 3301-15.
302. Yu G, Yao W, Xiao W, Li H, Xu H, Lang B. MicroRNA-34a functions as an anti-metastatic microRNA and suppresses angiogenesis in bladder cancer by directly targeting CD44. J Exp Clin Cancer Res. 2014; 33: 779.
303. Zhang J, Wang H, Wu J, Yuan C, Chen S, Liu S, et al. GALNT1 Enhances Malignant Phenotype of Gastric Cancer via Modulating CD44 Glycosylation to Activate the Wnt/β-catenin Signaling Pathway. Int J Biol Sci. 2022; 18: 6068-83.
304. Gao T, Wen T, Ge Y, Liu J, Yang L, Jiang Y, et al. Disruption of Core 1-mediated O-glycosylation oppositely regulates CD44 expression in human colon cancer cells and tumor-derived exosomes. Biochem Biophys Res Commun. 2020; 521: 514-20.
305. Moreira IB, Pinto F, Gomes C, Campos D, Reis CA. Impact of Truncated O-glycans in Gastric-Cancer-Associated CD44v9 Detection. Cells. 2020; 9.
306. Teriete P, Banerji S, Noble M, Blundell CD, Wright AJ, Pickford AR, et al. Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. Mol Cell. 2004; 13: 483-96.
307. Lesley J, Hascall VC, Tammi M, Hyman R. Hyaluronan binding by cell surface CD44. J Biol Chem. 2000; 275: 26967-75.
308. Pratt RL. Hyaluronan and the Fascial Frontier. Int J Mol Sci. 2021; 22.
309. Vuorio J, Škerlová J, Fábry M, Veverka V, Vattulainen I, Řezáčová P, et al. N-Glycosylation can selectively block or foster different receptor-ligand binding modes. Sci Rep. 2021; 11: 5239.

310. Faller CE, Guvench O. Terminal sialic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine sidechains. Proteins. 2014; 82: 3079-89.
311. Zhang X, Dou P, Akhtar ML, Liu F, Hu X, Yang L, et al. NEU4 inhibits motility of HCC cells by cleaving sialic acids on CD44. Oncogene. 2021; 40: 5427-40.
312. Lee D, Na J, Ryu J, Kim HJ, Nam SH, Kang M, et al. Interaction of tetraspan(in) TM4SF5 with CD44 promotes self-renewal and circulating capacities of hepatocarcinoma cells. Hepatology. 2015; 61: 1978-97.
313. Lee D, Lee JW. Self-renewal and circulating capacities of metastatic hepatocarcinoma cells required for collaboration between TM4SF5 and CD44. BMB Rep. 2015; 48: 127-8.
314. Sathe G, George IA, Deb B, Jain AP, Patel K, Nayak B, et al. Urinary glycoproteomic profiling of non-muscle invasive and muscle invasive bladder carcinoma patients reveals distinct N-glycosylation pattern of CD44, MGAM, and GINM1. Oncotarget. 2020; 11: 3244-55.
315. Azevedo R, Soares J, Gaiteiro C, Peixoto A, Lima L, Ferreira D, et al. Glycan affinity magnetic nanoplatforms for urinary glycobiomarkers discovery in bladder cancer. Talanta. 2018; 184: 347-55.
316. Chang L, Chen YJ, Fan CY, Tang CJ, Chen YH, Low PY, et al. Identification of Siglec Ligands Using a Proximity Labeling Method. J Proteome Res. 2017; 16: 3929-41.
317. Liu W, Ji Z, Wu B, Huang S, Chen Q, Chen X, et al. Siglec-15 promotes the migration of liver cancer cells by repressing lysosomal degradation of CD44. FEBS Lett. 2021; 595: 2290-302.
318. Oliveira FMS, Mereiter S, Lönn P, Siart B, Shen Q, Heldin J, et al. Detection of post-translational modifications using solid-phase proximity ligation assay. N Biotechnol. 2018; 45: 51-9.
319. Campos D, Freitas D, Gomes J, Magalhaes A, Steentoft C, Gomes C, et al. Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. Mol Cell Proteomics. 2015; 14: 1616-29.
320. Shi H, Jiang C, Yao H, Zhang Y, Zhang Q, Hou X, et al. CD44 fucosylation on bone marrow-derived mesenchymal stem cells enhances homing and promotes enteric nervous system remodeling in diabetic mice. Cell Biosci. 2021; 11: 118.
321. Xiao F, Chen L. Effects of extracorporeal fucosylation of CD44 on the homing ability of rabbit bone marrow mesenchymal stem cells. J Orthop Sci. 2019; 24: 725-30.

322. Xie N, Shen G, Gao W, Huang Z, Huang C, Fu L. Neoantigens: promising targets for cancer therapy. Signal Transduction and Targeted Therapy. 2023; 8: 9.
323. Han H, Stapels M, Ying W, Yu Y, Tang L, Jia W, et al. Comprehensive characterization of the N-glycosylation status of CD44s by use of multiple mass spectrometry-based techniques. Anal Bioanal Chem. 2012; 404: 373-88.
324. English NM, Lesley JF, Hyman R. Site-specific de-N-glycosylation of CD44 can activate hyaluronan binding, and CD44 activation states show distinct threshold densities for hyaluronan binding. Cancer Res. 1998; 58: 3736-42.
325. Halim A, Nilsson J, Ruetschi U, Hesse C, Larson G. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. Mol Cell Proteomics. 2012; 11: M111 013649.
326. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 2013; 32: 1478-88.
327. Zhang G, Zhang H, Wang Q, Lal P, Carroll AM, de la Llera-Moya M, et al. Suppression of human prostate tumor growth by a unique prostate-specific monoclonal antibody F77 targeting a glycolipid marker. Proc Natl Acad Sci U S A. 2010; 107: 732-7.
328. Aasted MKM, Groen AC, Keane JT, Dabelsteen S, Tan E, Schnabel J, et al. Targeting Solid Cancers with a Cancer-Specific Monoclonal Antibody to Surface Expressed Aberrantly O-glycosylated Proteins. Mol Cancer Ther. 2023; 22: 1204-14.

AIMS & STUDY OUTLINE

AIMS & STUDY OUTLINE

Bladder cancer (BlCa) is one of the most common genitourinary malignancies [1]. Currently, the five-year overall survival of advanced-stage bladder cancer does not exceed 38% [2]. The mainstay treatments frequently fail to prevent disease relapse and progression, being also associated with significant side effects, urging novel and more effective treatment options [2]. Cell-surface glycoprotein CD44 has been widely studied as a potential biomarker of BlCa aggressiveness and cancer-stem cells [3-6]. However, CD44 constitutes a highly heterogeneous protein family composed of several splicing proteoforms, which are further diversified by multiple glycosylation variants [6]. The lack of tools to tackle CD44 molecular heterogeneity has led to conflicting results, delaying clinical applications [6].

This work envisaged the establishment of CD44 glycoprofile for novel immunotherapies in BlCa. This thesis starts with a brief introductory section focusing on BlCa epidemiologic and pathophysiologic factors. Given that BlCa is characterized by a profound deregulation of glycosylation pathways, the most common *O*-glycosylation alterations found in BlCa are described. The CD44 glycoprotein is then explored from a clinical and functional standpoint in BlCa, with emphasis on novel CD44 glycoproteoforms-based targeted immunotherapies. The original section (Chapter 1) of this dissertation consists of a research paper published in Theranostics in early 2022, which resorts to a multi-omics approach for the unequivocal characterization of clinically relevant CD44 glycoproteoforms for future targeted therapeutics. As such, this paper interrogates the transcriptome and glycoproteome of a bladder cancer patient cohort for CD44 splicing signatures in glycoproteogenomics settings. Orthogonal validation of research findings was pursued in glycoengineered cell models. Phosphoarrays were used to tackle the role of CD44 in relevant oncogenic pathways.

Several specific tasks were established to address the main objective, encompassing the following:

i) Screen a broad series of bladder tumors and healthy urothelium, using gene expression assays, to identify experimentally confirmed CD44 splicing isoforms.

ii) Interrogate a larger patient series from the Cancer Genome Atlas with a detailed clinical history to CD44 splicing isoforms.
iii) Assess CD44 transcripts in the context of transcriptome molecular subtypes of bladder cancer.
iv) Evaluate a wide number of healthy tissues for CD44 splicing isoforms.
v) Screen bladder cancer tissues with different histological backgrounds to evaluate the co-localization of CD44 and common BlCa glycoantigens such as Tn and STn.
vi) Investigate the cancer specificity of CD44-Tn/STn proteoforms.
vii) Characterize CD44 glycoproteoforms in both tumor tissues and BlCa cell lines through a combined approach involving transcriptomics, glycomics, and glycoproteomics in glycoproteogenomics settings.
viii) Explore the functional role of CD44 and its altered glycosylation by silencing its expression in BlCa cells and glycoengineered BlCa cell models.
ix) Conduct antibody phosphoarrays screening on BlCa cells and glyco-edited BlCa cell lines to investigate the role of CD44 and its glycoforms in relevant oncogenic pathways.

Overall, chapter 1 sets a roadmap for cancer-specific CD44 glycoproteoforms discovery in BlCa, providing the molecular foundation for targeted interventions.
Finally, the Concluding Remarks and Future Perspectives section offers a comprehensive summary of the dissertation's key findings, while also discussing future approaches and critical milestones for results exploitation for immunotherapy design.

References

1. Dobruch J, Oszczudłowski M. Bladder Cancer: Current Challenges and Future Directions. Medicina (Kaunas). 2021; 57.

2. Bilim V, Kuroki H, Shirono Y, Murata M, Hiruma K, Tomita Y. Advanced Bladder Cancer: Changing the Treatment Landscape. J Pers Med. 2022; 12.

3. Yang YM, Chang JW. Bladder cancer initiating cells (BCICs) are among EMA-CD44v6+ subset: novel methods for isolating undetermined cancer stem (initiating) cells. Cancer Invest. 2008; 26: 725-33.

4. Hofner T, Macher-Goeppinger S, Klein C, Schillert A, Eisen C, Wagner S, et al. Expression and prognostic significance of cancer stem cell markers CD24 and CD44 in urothelial bladder cancer xenografts and patients undergoing radical cystectomy. Urol Oncol. 2014; 32: 678-86.

5. Muramaki M, Miyake H, Kamidono S, Hara I. Over expression of CD44V8-10 in human bladder cancer cells decreases their interaction with hyaluronic acid and potentiates their malignant progression. J Urol. 2004; 171: 426-30.

6. Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clin Proteomics. 2018; 15: 22.

CHAPTER 1

GLYCOPROTEOGENOMICS CHARACTERIZES THE CD44 SPLICING CODE ASSOCIATED WITH BLADDER CANCER INVASION

The information provided in this section is based in the following publication:

Gaiteiro C, Soares J, Relvas-Santos M, Peixoto A, Ferreira D, Paulo P, Brandão A, Fernandes E, Azevedo R, Palmeira C, Freitas R, Miranda A, Osório H, Prieto J, Lima L, Silva AMN, Santos LL, Ferreira JA. Glycoproteogenomics characterizes the CD44 splicing code associated with bladder cancer invasion. Theranostics. 2022 Mar 28;12(7):3150-3177. doi: 10.7150/thno.67409. PMID: 35547758; PMCID: PMC9065180. **Impact Factor in 2022: 11.6; Q1**

CHAPTER 1: GLYCOPROTEOGENOMICS CHARACTERIZES THE CD44 SPLICING CODE ASSOCIATED WITH BLADDER CANCER INVASION

Cristiana Gaiteiro[1,2,3,4,5], Janine Soares[1,2,3,4,6], Marta Relvas-Santos[1,2,3,4,7,8,9], Andreia Peixoto[1,2,3,7,8], Dylan Ferreira[1,2,3,4,5,7,8,], Paula Paulo[2,3,10], Andreia Brandão[2,3,10], Elisabete Fernandes[1,2,3,11], Rita Azevedo[12], Carlos Palmeira[1,2,3,11,13], Rui Freitas[1,2,3,4,7,8], Andreia Miranda[1,2,3,14], Hugo Osório[7,14,15], Jesús Prieto[5], Luís Lima[1,2,3], André M. N. Silva[9,16], Lúcio Lara Santos[1,2,3,4,11,16,17], José Alexandre Ferreira[1,2,3,4,16,a]

[1]Experimental Pathology and Therapeutics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; [2]RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto), 4200-072 Porto, Portugal; [3]Porto Comprehensive Cancer Center (P.ccc), 4200-072 Porto, Portugal; [4]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-013 Porto, Portugal; [5]Center for Applied Medical Research (Centro de Investigación Médica Aplicada, CIMA), University of Navarra, 31008 Pamplona, Navarra, Spain; [6]REQUIMTE-LAQV, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal; [7]Institute for Research and Innovation in Health (i3S), University of Porto, 4200-135 Porto, Portugal; [8]Institute for Biomedical Engineering (INEB), University of Porto, 4200-135 Porto, Portugal; [9]REQUIMTE-LAQV, Department of Chemistry and Biochemistry, Faculty of Sciences of the University of Porto, 4169-007 Porto, Portugal; [10]Cancer Genetics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; [11]FP-I3ID, University Fernando Pessoa, 4249-004 Porto, Portugal; [12]Laboratoire d'Etude du Métabolisme des Médicaments (LEMM), CEA, INRA, Université Paris Saclay, F-91191, Gif-sur-Yvette cedex, France; [13]Immunology Department, Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; [14]Faculty of Medicine of the University of Porto, 4200-319 Porto, Portugal; [15]Ipatimup—Institute of Molecular Pathology and Immunology of the University of Porto, University of Porto, 4200-135 Porto, Portugal; [16]GlycoMatters Biotech, 4500-162 Espinho, Portugal; [17]Department of Surgical Oncology, Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal

[a]**Corresponding author:**
José Alexandre Ferreira (jose.a.ferreira@ipoporto.min-saude.pt)
ORCID ID: https://orcid.org/0000-0002-0097-6148
Experimental Pathology and Therapeutics Group, Research Centre, Portuguese Oncology Institute of Porto, R. Dr. António Bernardino de Almeida 62, 4200-072 Porto, Portugal; Tel. +351 225084000 (ext. 5111).

Abstract

Rationale: Bladder cancer (BlCa) management demands the introduction of novel molecular targets for precision medicine. Cell surface glycoprotein CD44 has been widely studied as a potential biomarker of BlCa aggressiveness and cancer stem cells. However, significant alternative splicing and multiple glycosylation generate a myriad of glycoproteoforms with potentially distinct functional roles. The lack of tools for precise molecular characterization has led to conflicting results, delaying clinical applications. Addressing these limitations, we have interrogated the transcriptome of a large BlCa patient cohort for splicing signatures.

Methods: *CD44* gene and its splicing variants were assessed by Real Time-Polymerase Chain Reaction (RT-PCR) and RNAseq in tumor tissues. The co-localization of CD44 and short *O*-glycans was evaluated by proximity ligation assay (PLA), immunohistochemistry and double-immunofluorescence. An innovative glycoproteogenomics approach, integrating transcriptomics-customized datasets and glycomics for protein annotation from nanoLC-ESI-MS/MS experiments, was developed and implemented to identify CD44 variants and associated glycosignatures. The impact of CD44 silencing on proliferation and invasion of BlCa cell lines and glycoengineered cells was determined by BrdU ELISA and Matrigel invasion assays, respectively. Antibody phosphoarrays were used to investigate the role of CD44 and its glycoforms in the activation of relevant oncogenic signaling pathways.

Results: Transcriptomics analysis revealed remarkable CD44 isoforms heterogeneity in bladder cancer tissues, as well as associations between short CD44 standard splicing isoform (CD44s), invasion and poor prognosis. We further demonstrated that targeting short *O*-glycoforms such as the Tn and sialyl-Tn antigens was key to overcome the lack of cancer specificity presented by CD44. Glycoproteogenomics allowed, for the first time, the comprehensive characterization of CD44 splicing code at the protein level. The concept was applied to invasive human BlCa cell lines, glycoengineered cells, and tumor tissues, enabling unequivocal CD44s identification as well as associated glycoforms. Finally, we confirmed the link between CD44 and invasion in CD44s-enriched cells *in vitro* by small interfering RNA (siRNA) knockdown, supporting findings from BlCa tissues. The key role played by short-chain *O*-glycans in CD44-mediated invasion was also demonstrated through glycoengineered cell models.

Conclusions: Overall, CD44s emerged as biomarker of poor prognosis and CD44-Tn/ Sialyl-Tn (STn) as promising molecular signatures for targeted interventions. This study materializes the concept of glycoproteogenomics and provides a key vision to address the cancer splicing code at the protein level, which may now be expanded to better understand CD44 functional role in health and disease.

Keywords: glycomics; proteogenomics; glycoproteogenomics; bladder cancer; CD44

Introduction

Bladder cancer (BlCa) remains a pressing health concern and encompasses significant mortality, especially when diagnosed at advanced stages [1]. Cluster of Differentiation 44 (CD44) is a multifunctional and heavily glycosylated transmembrane protein [2] involved in cell-cell and cell-extracellular matrix adhesion [3], immune functions [4, 5], lymphocyte homing [6, 7], hematopoiesis [8], and oncogenic signaling. By interacting with several downstream effector proteins, it dictates cell migration and adhesion [9, 10], tumor invasion [11], and metastasis [12, 13]. Hence, this glycoprotein has been found overexpressed in more aggressive bladder tumors [14, 15], being widely adopted as a biomarker of bladder cancer stem cells (CSC) [16, 17].

The human CD44 gene is located on chromosome 11p13 and consists of 19 exons [2, 18]; however, the full-length protein has never been detected due to the occurrence of intense alternative splicing. Exons 1 to 16 encode the extracellular domain, exon 17 the transmembrane region, and exons 18 and 19 the intracellular domains [2]. The first 5 exons are constitutively transcribed across all currently known isoforms, encoding hyaluronic acid [19], osteopontin [20], collagen [21], laminin [22], and fibronectin [23] binding sites. On the other hand, exons 6 to 14 are subjected to alternative splicing, generating a myriad of different variants whose functional implications are yet to be fully understood [2]. The number of proteoforms generated by mRNA translation and processing is greatly amplified by *O*-GalNAc glycosylation, mostly occurring in CD44 variable regions that are rich in serine and threonine residues. Moreover, a multiplicity of different but closely related glycan structures may be found in the same protein, exponentiating molecular micro-, macro-, and meta-heterogeneity [2, 24]. Still, few studies have addressed the CD44 glycocode in cancer and its functional implications for disease progression [25-27]. Moreover, conflicting results have been generated concerning its role in disease and clinical value, which are directly linked to analytical limitations for unequivocal molecular characterization, mostly due to the lack of high-throughput approaches to characterize CD44 at the protein level [2]. Misguiding nomenclature has also been posing a major limitation for comprehensive data mining and definitive molecular characterization, leading us to propose an uniformization (summarized in Figure S1) [2].

High-throughput proteomics constitutes the gold standard approach to tackle CD44 molecular heterogeneity; however, the success of current workflows

is limited by the capacity of databases for protein annotation. The high degree of sequence similarity amongst proteoforms, many times differing by short and potentially heavily glycosylated peptide sequences, poses a significant limitation. As such, CD44 molecular characterization has been mainly inferred from transcripts analysis, supported by immunoassays based on antibodies that lack isoform specificity. In summary, a single omics cannot portrait its molecular complexity, delaying clinical applications. Herein, we hypothesize that multi-omics settings may be required to unequivocally characterize CD44 glycoproteoforms, combining transcriptomics, glycomics and glycoproteomics in glycoproteogenomics settings [28] for precise identification of CD44 signatures of clinical relevance, foreseeing the design of novel targeted therapies.

Material and Methods

Patients sampling and healthy human tissues

A retrospective series of 75 formalin-fixed paraffin-embedded (FFPE) bladder tumor tissues from the Portuguese Institute of Oncology of Porto (IPO-Porto) biobank were used for this study. Bladder tumors were surgically removed from 61 male and 14 female, ranging from 26 to 85 years of age (67 ± 11.7), admitted and treated at the IPO-Porto between 2000 and 2017. The tumors were classified as non-muscle invasive (≤T1, NMIBC; n=34) and muscle-invasive (≥T2; MIBC; n=41) bladder cancers. Eleven histologically normal urothelium tissue sections from healthy individuals were included. Additionally, a broad library of healthy tissues (liver, colon, small intestine, gallbladder, pancreas, thyroid, stomach, appendix, testicle, skin, breast, kidney, lung, mucosa-associated lymphoid tissue-MALT, and white blood cells) were also considered. Bladder tumor and urothelial sections were characterized in terms of CD44 isoforms by RT-PCR and glycoproteogenomics. Additionally, tumors and healthy tissues were screened for CD44 and altered glycosylation (Tn and STn antigens) by different immunoassays (immunohistochemistry, WB; PLA), double staining immunofluorescence with lectins and antibodies). All procedures were performed under the approval of the hospital's ethics committee (project reference: CES 86/017) after obtaining informed patient's consent, being the clinicopathological information obtained from patient's clinical records.

TCGA Dataset

Updated clinical information for 413 TCGA Bladder Carcinoma (BLCA) cases corresponding to muscle invasive lesions of different stages (T2, T3, T4) and histopathological natures (papillary and non-papillary), including overall survival and disease-free survival information, were obtained from cBioPortal database (https://www.cbioportal.org/). The level-3 RNA-Seq data from 408 TCGA-BLCA cases was retrieved from Broad GDAC FIREHOSE (http://gdac.broadinstitute.org/) in March 2021.

Cell lines and cell culture conditions

BC cell lines, RT4, 5637 and T24, were acquired from ATCC and cultured in RPMI 1640 GlutaMAX™ medium (Gibco, Thermo Fisher Scientific) supplemented with 10% heat-inactivated FBS (Gibco, Thermo Fisher Scientific) and 1% penicillin–streptomycin (10,000 Units/mL penicillin; 10,000 μg/mL streptomycin; Gibco, Thermo Fisher Scientific). T24 glycoengineered cell models (T24 *C1GALT1* knock-out (KO) and T24 *C1GALT1* KO*/ST6GALNAC1* knock-in (KI)) and corresponding controls (T24 *C1GALT1* control carrying a silent mutation and T24 *C1GALT1* control/*ST6GALNAC*1 mock carrying a silent mutation and a mock vector) were generated as described by Peixoto A. *et al.[29]*. Transfected cells selection was performed based on puromycin (2 μg/mL, EMD Millipore) resistance. BlCa cell lines and glycoengineered models were cultured at 37 ˚C in a 5% CO_2 humidified atmosphere.

CD44 transcripts identification by Real-time polymerase chain reaction

TriPure isolation reagent (Roche Diagnostics GmbH) was used to extract total RNA from BlCa cells, while RNA from tissues was extracted using the Absolutely RNA FFPE Kit (Agilent), according to the manufacturer's instructions. Complementary DNA (cDNA) reverse transcription and mRNA expression were performed as previously described [30]. The relative expression of total *CD44* and its isoforms was determined by RT-PCR analysis using TaqMan Gene Expression Assays (total *CD44*: Hs01075864_m1; *CD44e5-v2*: Hs01075866_m1; *CD44e5-v3*: Hs01081480_m1; *CD44e5-v8*: Hs01081475_m1, *CD44e5-e15:* Hs01081473_m1; Applied Biosystems) in a 7500 Sequence Detector (Applied Biosystems). β-2-microglobulin *(B2M)* and Hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) were used for normalization, also as previously described. Details on TaqMan gene

expression assays can be found in Table S1. All samples were run in duplicate and relative mRNA gene expression was calculated with the $2^{-\Delta Ct}$ formula.

Reverse Transcriptase-Polymerase Chain Reaction and Sanger Sequencing

For CD44 transcripts characterization, RNA from T24 and 5637 cells was converted into cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with random primers according to the manufacturer's instructions, and specific primer for CD44. For the latter, 1 μg of total RNA was added to 4 mM dNTP Mix and 20 pmol of a CD44 specific primer (5'-CCTTATAGGACCAGAGGTTGTGTTT-3') annealing in exon 16 (NM_000610.4) in reverse orientation, and converted to cDNA in the presence of 1 μl of MultiScribe™ Reverse Transcriptase, 1μl of RNase Inhibitor and 1x RT Buffer, in a 20 μl reaction mix. Reaction took place at 37 ºC for 120 min, followed by enzyme inactivation at 85 ºC for 5 min. For amplification of CD44 transcripts, cDNA was used in two different PCR reactions with a Forward primer for exon 5 (5'-CCCCAGCAACCCTACTGATG-3') and different Reverse primers, one to exon 8 (5'-TGAATGGCTTGGGTTCCACT-3') and one to exon 16 (same used in cDNA synthesis). Briefly, 2 μl of cDNA were mixed with 400 μM dNTPs, 200 nM of each primer and 2.5 U of Takara LA Taq DNA Polymerase in a 50 μl reaction containing 1x LA PCR Buffer II (Mg^{2+} Plus) (Takara). PCR program was the following: initial denaturation at 94 ºC for 1 min; 30 cycles of annealing at 64 ºC for 1 min and extension at 68 ºC for 4 min; and final extension at 72 ºC for 10 min. Amplified products were separated in a 2% agarose gel electrophoresis and bands were excised and purified using the GFX PCR DNA and Gel Band Purification Kit (Cytiva), according to the manufacturer´s instructions. For Sanger sequencing of purified PCR products, the BigDye Terminator V3.1 chemistry and the 3500 Genetic Analyzer (Applied Biosystems) were used, following general instructions.

CD44 transcripts identification by RNAseq

CD44 variants were identified based on a protocol described in detail by Peixoto et al. [31]. Briefly, RNA was extracted from 5637 and T24 cell pellets from three independent replicates using the RNeasy Plus Mini kit (Qiagen), pooled together for each cell line, and quantified using Qubit 2.0 Fluorometer (Thermo Fisher Scientific). RNA integrity was evaluated with Agilent TapeStation (Agilent Technologies). The NEBNext Ultra RNA Library Prep Kit was used to prepare RNA

sequencing library for Illumina following manufacturer's recommendations. After validation and quantification, the sequencing libraries were clustered on one lane of a flow cell, loaded on an Illumina HiSeq 4000 instrument and sequenced using a 2 × 150 Paired End configuration. The HiSeq Control Software was used to analyze the obtained results. The gene hit counts were assessed using the feature Counts from the Subread package v.1.5.2, counting only unique reads that belong to exon regions, and then used for downstream differential expression analysis. A SNP/INDEL and a gene fusion analysis were performed using the Samtools v.1.3.1 program followed by VarScan v.2.3.9 and STAR Fusion v.1.1.0, respectively. For novel CD44 transcripts discovery, Stringtie software was used to extract the transcripts expressed in each cell line. Novel transcripts were pinpointed through the comparison between the reference annotation file and resulting gft file.

Flow cytometry

Cells were detached using Accutase™ Cell Detachment Solution (BD™), fixed with 2% paraformaldehyde (PFA; Sigma-Aldrich), and incubated with rabbit anti-CD44 polyclonal antibody (ab157107, Abcam; Table S2) using a 1:100 dilution in PBS/2% FBS for 1 hour at room temperature (RT). Goat anti-rabbit IgG (H + L) cross-adsorbed secondary antibody Alexa Fluor 488 (Invitrogen) was used for CD44 detection at a 1:300 dilution in PBS/2% FBS for 15 minutes (min) at RT. Data analysis was performed through CXP Software in a FC500 Beckman Coulter flow cytometer. Results represent the standard deviation of three independent experiments.

Cell models and tumor proteins extraction

Plasma membrane proteins from BlCa wild type and glycoengineered cell lines were extracted by subcellular fractionation using ultracentrifugation, as previously described [32]. Briefly, cultured cells were detached by scrapping with fractionation buffer (20 mM HEPES buffer (pH = 7.4), 10 mM KCl, 2 mM $MgCl_2$, 1 mM EDTA and 1 mM EGTA) on ice. The cell suspension was then passed through a 27G needle, left on ice for 20 min, and then centrifuged at 720 *g* for 5 min at 4 ˚C to remove the nuclei. Supernatants were transferred to a new tube and recentrifuged at 10,000 *g* for 5 min at 4 ˚C to remove mitochondria. Samples were then transferred to polycarbonate centrifuge bottles with cap assemblies and centrifuged for 1 hour at 100,000 *g* at 4 ˚C. The pellets were recovered, resuspended in the fractionation buffer, and passed through a 25G needle before

a new centrifugation for 45 min at 100,000 *g* at 4 ˚C. Finally, the plasma membrane-enriched fraction was resuspended in an appropriate volume of TBS with 0.3% SDS. Regarding bladder tumors, total protein was extracted from CD44-STn/Tn expressing areas excised from formalin fixed paraffin embedded tumors using the Qproteome FFPE tissue kit (Qiagen), according to the manufacturer's instructions.

CD44 isolation and proteolytic digestion

The Pierce™ Protein G Agarose (Thermo Fisher Scientific) was used to immunoprecipitate CD44 from membrane protein extracts of 5637 and T24 wild type cells and T24 glycoengineered models (100 mg or 500 mg of starting material for WB and glycoproteomics, respectively) and tumor protein extracts (500 mg of starting material). Briefly, agarose beads were blocked with 1% Bovine Serum Albumin (BSA; Sigma-Aldrich) for 1 hour at 4 ºC. Prior to immunoprecipitation (IP), membrane protein extracts were cleared with blocked agarose beads for 2 hours at 4 ºC, and then, incubated at 4 ºC for 2 hours with 3 µg (WB) or 6 µg (glycoproteomics) of polyclonal anti-CD44 antibody (ab157107; Abcam). The protein-antibody complexes were incubated overnight with newly blocked agarose beads at 4 ºC and eluted with SDS Sample Loading Buffer (250 mM Tris–HCl pH 6.8, 8% (*w/v*), SDS, 0.2% (*w/v*) bromophenol blue, 40% (*v/v*) glycerol, 20% (*v/v*) β-mercaptoethanol). Immunoprecipitated CD44 was used for further WB and glycoproteomics analysis. CD44 IP products were resolved by electrophoresis under denaturing conditions using 4-20% gradient precast polyacrylamide gels (Bio-Rad) and bands were excised from gels. Proteins were then reduced with 10 mM 1,4-dithiothreitol (DTT; Sigma-Aldrich) for 45 min at 56 ºC, alkylated with 10 mM iodoacetamide (Sigma-Aldrich) for 30 min in the dark, desialylated with 10 U α-neuraminidase [*Clostridium perfringens* neuraminidase Type VI (Sigma-Aldrich)][8] for 2 hours at 37 ºC and digested overnight at 37 ºC with chymotrypsin (25 µg/mL; Promega). The proteolytic digests were then analyzed by nanoLC-MS/MS.

Glycomics

Bladder cancer cells *O*-glycome was characterized using the Reporter/Amplification method [33], as previously described by us [31, 32]. Briefly, cell culture media of semi-confluent cells was supplemented with peracetylated benzyl 2-acetamido-2-deoxy-α-D-galactopyranoside (Sigma-Aldrich, St. Louis, MO, USA) to a final concentration of 150 µM. Following 24 hours incubation, glycans

were isolated from the conditioned media by filtration using 10 kDa centrifugal filter (Amicon Ultra-4; Merck KGaA, Darmstadt, Germany), followed by solid-phase extraction in Sep-Pak 3 cc C18 cartridges (Waters, Milford, MA, USA). The isolated Bn-*O*-glycosides were then permethylated and analyzed by MALDI-TOF-MS on a Bruker UltrafleXtreme mass spectrometer (Bruker Daltonics). Dried samples were resuspended in methanol, mixed (1:1 sample:matrix ratio) with 2,5-dihydroxybenzoic acid (DHB; 10 mg/ml in 50% methanol and 0.1% trifluoroacetic acid; Sigma-Aldrich) and spotted onto a MTP 384 polished steel target plate (Bruker Daltonics). Spectra were acquired in positive ion reflector mode, for a mass range from 540 to 2000 kDa. Then, spectra were subjected to external calibration, using the Peptide Calibration Standard II (Bruker Corporation) combined with α-cyano-4-hydroxycinnamic acid (5 mg/ml in 50% acetonitrile and 0.1% trifluoroacetic acid; Sigma-Aldrich), and internal calibration, using a mass control list constructed by us, considering previous knowledge on bladder cancer *O*-glycosylation.

Nano-Liquid chromatography-Tandem mass spectrometry

CD44 IP digests were analyzed by nano liquid chromatography mass spectrometry (nanoLC-MS/MS), exploring an HCD-triggered CID approach. nanoLC-HCD-MS2 was carried out in a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific) coupled to a Ultimate 3000 RSLCnano system (Dionex, Thermo Scientific). Briefly, chymotrypsin digested samples were pre-concentrated in an Acclaim PepMap C18 column (100 Å, 5 mm × 300 µm, i.d. 160454, Thermo Fisher Scientific). Peptide separation was performed in an analytical EASY-Spray column (C18, 100 Å, 2 µm, 75 µm × 500 mm, Thermo Fisher Scientific) with a flow rate of 0.25 µl/min, by mixing the eluent A: 0.1% aqueous formic acid (FA) and eluent B: 0.1% FA in 80% acetonitrile (ACN), with the following gradient: 2 min (2.5% B to 10% B), 50 min (10% B to 35% B), 8 min (35% B to 99% B), and 10 min (hold at 99% B). The column was equilibrated with 2.5% B for 17 min. The mass spectrometer was operated in the positive ion mode over the *m/z* range 380-1580, and spray voltage was set at 1.9 kV. Full MS settings were the following: 70k resolution (*m/z*=200), AGC target 3×10^6, maximum injection time 100 milliseconds (ms). The data-dependent parameters were: minimum AGC target 7×10^3, intensity threshold 6.4×10^4, charge state exclusion: unassigned, 1, 8, >8, peptide match preferred, exclude isotopes on, and dynamic exclusion of 20 seconds (s). The top 10 peaks were selected for HCD fragmentation, using the

following MS/MS settings: normalized collision energy (NCE) of 27%, 35k resolution (*m/z*=200), AGC target 2×10^5, maximum injection time 110 ms, isolation window 2.0 *m/z*, isolation offset 0.0 *m/z*, HCD first mass at 110 *m/z*. Mass spectrometer was controlled by Xcalibur 4.0 and Tune 2.9 software (Thermo Scientific). The samples were then run on a nanoLC system coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) to allow characterization by collision induced fragmentation (CID). Liquid chromatography was performed on an EASY-Spray C18 PepMap, 100 Å, 2 µm, 150 mm × 75 µm (Thermo Fisher Scientific) using the same gradient mentioned above. The mass spectrometer was operated in the positive ion mode over the *m/z* range 380-1580. Nanospray voltage was set at 1.9 kV and full scan nominal resolution was 60k (*m/z*=400). CID was triggered from a precursor ion list containing the *m/z* values that presented the HexNAc oxonium ion (*m/z* 204.087 within a ± 0.01 range) in the HCD-MS/MS spectra previously acquired. The 6 most intense ions from the customized parent list were selected for CID fragmentation with NCE = 35%, and MS/MS spectra were acquired in the linear ion trap with an isolation width of 2 Da. Specific parameters were: MS maximum injection time of 500 ms; MS/MS maximum injection time of 50 ms; AGC target 1×10^6 for the Orbitrap and 1×10^4 for LTQ MS^n analysis; dynamic exclusion 45 s; charge rejection: unassigned and 1. Mass spectrometer was controlled by Xcalibur 3.1 software (Thermo Scientific).

Bioinformatics for CD44 glycoproteoforms identification

MS data were first converted to peak lists using Proteome Discoverer version 2.5.0.400 (Thermo Scientific), and then searched against the UniProt Homo sapiens proteome (May 5 2020; 75069 entries), using the MSPepSearch and SequestHT search engines for protein identification and the Percolator algorithm v3.05.0 for statistical validation. Prior to the search, the human proteome FASTA database was edited to include CD44 sequences inferred from RNAseq characterization of 5637 and T24 cell lines (Table S2). Searches for HCD tandem spectra were performed with a tolerance of 5 ppm for precursor and 0.02 Da for fragment ions. For CID, a tolerance of 5 ppm was admitted for precursor and 0.6 Da for fragment ions. Chymotrypsin was selected as the proteolytic enzyme and up to two missed cleavages were allowed. A new customized databased composed of high confidence identifications resulting from the initial search was then constructed for definitive CD44 glycoproteoforms identification in glycoproteogenomics settings.

For cell lines, carbamidomethylcysteine (+57.0215 Da) was set as a fixed modification, while methionine oxidation (+15.9949 Da), protein N-terminal formylation (+27.994 Da), N-terminal acetylation (+42.0106 Da), Asparagine deamidation (+0.9840 Da), ammonia-loss of Cysteine N-terminal (-17.0265 Da) and glutamine to pyro-glutamine modification (-17.0265 Da) were considered as variable modifications. For tumor samples, carbamidomethylcysteine (+57.0215 Da) was elected as a fixed modification and, based on previous reports concerning the analysis of FFPE tissues [34], the following oxidative modifications were also included in the variable modifications list: lysine to aminoadipic semialdehyde (-1.0316 Da), arginine to glutamic semialdehyde (-43.0534 Da), proline to pyroglutamic acid (+13.9794 Da), tryptophan to hydroxykynurenin (+19.9898 Da), tryptophan to kynurenine (+3,9949 Da), tryptophan to N-formylkynurenine (+31.9898 Da), threonine to 2-amino-3-ketobutyric Acid (-2.0156 Da), lysine methylation (+14.0156 Da), phenylalanine, Proline, Histidine and Tryptophan hydroxylation (+15.9949 Da) and Phenylalanine, Proline, Histidine and Tryptophan carbonylation (+13.9794 Da) [35]. For identification of CD44 glycoproteoforms in cell lines and tumors, the following variable modifications of serine and threonine were also considered: HexNac (+203.0794 Da), NacHexHex 365,1322 Da), HexNac(2) (+406.1588 Da), HexNacNeuAc (494,1748 Da), HexNac(2)Hex (+568.2116 Da), HexNacHexNeuAc (+656,2276 Da), HexNacHexdHex 511,1901 Da), HexNacHexNeuAc(2) (947,3230 Da), HexNac(2) Hex(2)NeuAc (1021.3598 Da). The presence of sialic acids was included to contemplate the possibility of incomplete de-sialylation. Glycoproteomics data was also analyzed using ByonicTM version 2.13.2 (Protein Metrics, Cupertino, CA, USA) using default settings [36]. Glycopeptide assignment was confirmed by manual spectra interpretation using the Xcalibur™ software (Thermo Scientific).

Western and lectin blotting

Protein extracts and CD44 immunoprecipitates from 5637 and T24 wild type, glycoengineered cells and bladder tumors were separated in 4–20% precast polyacrylamide gels (Bio-Rad) and transferred onto a nitrocellulose membrane (GE Healthcare Life Sciences). STn and Tn expressions were evaluated using the anti-tag-72 antibody [B72.3 + CC49] (1 μg/mL, ab199002, Abcam; Table S2) and biotinylated *Vicia Villosa* lectin (VVA lectin, 1:1000, Vector Laboratories; Table S2), respectively. CD44 expression was screened with the anti-CD44 antibody (1:5000,

ab157107, Abcam; Tables S2). Proteins were blotted with the primary antibody or lectin during 1 hour at RT. The peroxidase affiniPure goat anti-mouse IgG (H+L) polyclonal antibody (1:90,000, ImmunoResearch) was used as a secondary antibody for anti-tag-72 antibody detection, and the goat anti-rabbit IgG (H+L) HPR conjugate antibody (1:60,000; Thermo Fisher Scientific) was used for anti-CD44 antibody detection, both incubated for 30 min at RT. The VECTASTAIN® Elite ABC-HRP Reagent (1:10; Vector Laboratories) was used for 15 min at RT for analysis of Tn expression. Detection of B2M with the recombinant anti-B2M antibody [EP2978Y] (ab75853, Abcam; Table S2) followed by incubation with goat anti-rabbit IgG (H+L) HPR conjugate antibody (1:60,000; 30 min at RT) was performed as loading control.

Immunohistochemistry

FFPE bladder tumors and healthy tissue sections were screened by immunohistochemistry for CD44 and Tn and STn antigens, as previously described by Peixoto, A *et al*. [31]. Tn antigen expression was evaluated using the biotinylated VVA lectin (Vector Laboratories, 40 mg/mL, 1 hour at 37 ºC; Table S2) and the detection of STn and CD44 antigens were performed using the anti-tag-72 (B72.3 + CC49; Abcam, 0.5 mg/mL, overnight at 4 ºC; Table S2) and anti-CD44 (1:5000, ab157107, Abcam; Table S2) antibodies, respectively. Lack of cross-reactivity of VVA for blood group A and AB antigens was confirmed using the anti-blood group A monoclonal antibody (HE-193, Thermo Fisher, 1:5, overnight at 4 ºC; Table S2). Sialidase treatment of tissue samples prior to anti-STn probing was also performed to confirm the presence of the glycan (Sigma-Aldrich, 0.2 mg/mL, overnight at 37 ºC). CD44 and anti-tag-72 were detected using Novolink Polymer Detection System (Leica) according to manufacturer guidelines. Biotinylated VVA was detected using Streptavidin, Horseradish Peroxidase Conjugate (Thermo Fisher, ready-to-use, 30 min, RT) followed by incubation with ImmPACT® DAB Substrate, Peroxidase (Vector, 30:1000, 5 min, RT). All images were acquired on a Motic BA310E microscope (Motic) using the Motic Images Plus 3.0 software (Motic).

Double staining immunofluorescence

A selection of FFPE tissue sections positive for Tn and CD44 were screened for both antigens through double immunofluorescence to determine colocalization of both epitopes. Briefly, FFPE tissues were deparaffined, hydrated, and exposed to

antigen retrieval with EDTA 1mM pH8. Tn antigens were detected using 40 μg/mL FITC-labeled VVA lectin for 2 hours at RT. CD44 antigen detection was achieved using an unlabeled rabbit polyclonal CD44 antibody (Abcam; Table S2) at 1:250 for 1 hour at RT. An Alexa Fluor 594 anti-rabbit was used for 30 min at RT in the dark as a secondary antibody. T24 wild type cells and T24 glycoengineered models were evaluated for T, ST, Tn, STn, and CD44 to detect simultaneous expression between CD44 and these *O*-glycans. Shortly, cells were fixed with 4% PFA for 15 min, and then incubated for 1 hour with FITC-labeled VVA lectin (Tn, 0.02 μg/μL; Table S2), or FITC-labeled PNA lectin (T and ST after desialylation with 70 mU α-neuraminidase, 0.02 μg/μL; Table S2), or anti-tag-72 (STn, 5 μL/well; Table S2) and CD44 (1:100; Table S2). Alexa Fluor 594 anti-rabbit and Alexa Fluor 488 anti-mouse (ThermoFisher Scientific, 1:100) were used to detect CD44 and STn primary antibodies. Nuclear counterstain was performed with 4',6'-diamidino-2-phenylindole dihydrochloride (DAPI, 2.3×10^{-3} μg/μL; Thermo Scientific); for 10 min at RT in the dark. Fluorescence images were acquired on a Leica DMI6000 FFW microscope using Las X software (Leica).

Proximity Ligation Assay

In situ PLA were used for simultaneous detection of CD44 and STn antigens whenever in close spatial proximity at the cell surface in tumors and healthy tissues and glycoengineered bladder cancer cells. Cells were cultured in μ-Chamber 8-well slides (ibidi), fixed with 4% PFA for 15 min. Bladder tumors showing co-localization of CD44 and STn by immunohistochemistry, the healthy urothelium and the other human tissues described in the "Patients sampling and healthy human tissues" sections suggesting CD44 and STn co-localization were evaluated. PFA-fixed cancer cells and tissue sections first undergone antigen retrieval for 15 min with boiling citrate buffer pH=6.0 (Vector Laboratories), followed by incubation overnight at 4 °C with the conjugated primary antibodies mentioned in the immunohistochemistry section. Ligation and amplification of the PLA signal were achieved using the Duolink PLA Technology kit (Sigma-Aldrich). All slides were incubated with DAPI and mounted using Duolink Mounting Medium. Sialidase treatment prior to antibody probing was used to confirm the specificity of the PLA signals. STn and CD44 negative bladder tumors were used as negative controls. MCR-STn+ glycoengineered cell lines expressing CD44-STn [37, 38] and/or tumor tissues were

used as positive controls. The images were acquired on a Leica DMI6000 FFW microscope (Leica Microsystems) using the Las X software (Leica Microsystems).

siRNA Silencing Assay

siRNA reverse transfection was applied to silence *CD44* in 5637, T24, and T24 glycoengineered cells *in vitro*, using two different Silencer® Select siRNAs targeting CD44 (siRNAs ID: s2681 [target exon: 2] and s2682 [target exons: 1 and 2], Invitrogen). A Silencer® Select siRNA negative control (4390843, Invitrogen) was also included. Cells were detached and seeded (100,000 cells/well) in a 24-well plate before transfection with lipofectamine RNAiMAX (Invitrogen), according to the manufacturer's instructions. In brief, Silencer® Select siRNAs and lipofectamine RNAiMAX were diluted in Opti-MEM reduced serum medium (Gibco) and incubated for 5 min at RT. Subsequently, cells were incubated with siRNA-lipofectamine complexes for 72 hours at 37 ˚C. Cells were plated in duplicates for each experiment, and the CD44 silencing was confirmed by RT-PCR using the TaqMan Gene Expression Assay Hs01075864_m1.

Cell proliferation Assay

The proliferation of 5637, T24 cells and T24 glycoengineered cell models was evaluated in basal and silenced-CD44 conditions, using the cell proliferation ELISA BrdU Kit (Roche Diagnostics GmbH), according to the manufacturer's instructions. The immunoassay results were monitored at 450 nm using the iMARK™ microplate reader (Bio-Rad). Cell death negative controls composed of 1% Triton-X in a complete cell culture medium were used. The results are presented as the average and standard deviation of three independent assays with three replicates each.

Invasion Assay

The invasive capacity of 5637, T24 cells and T24 glycoengineered models, in basal and CD44-silenced conditions, was assessed using Corning BioCoat Matrigel Invasion Chambers (Corning). Briefly, 5×10^4 cells/mL were plated onto rehydrated invasion inserts according to the manufacturer's instructions and incubated at 37 ˚C for 24 hours. After removal of non-invasive cells, membranes were washed, and invasive cells were fixed with 4% PFA for 15 min. Membranes

were then mounted with VECTASHIELD mounting medium with DAPI, and invasive cells were counted in a Leica DM2000 microscope (Leica Microsystems). Three independent experiments were performed, and cells were seeded five times for each experiment. Invasion assays were normalized to cell proliferation average (Invasion Rate).

Phospho-Kinase Antibody Array

The relative phosphorylation levels of 37 phosphorylation sites by kinases and 49 receptor tyrosine kinases (RTK) (Figures S7-8) were determined with the Human Phospho-Kinase/RTK Array kits (ARY003C and ARY001B, respectively, R&D Systems), according to manufacturer's instructions. Briefly, T24 wild type, glycoengineered cells and corresponding controls were firstly transfected with Silencer Select siRNA targeting CD44 as described above, and then lysed to extract protein from whole cells. For each cell line 300 µg and 600 µg of protein were used for Human Phospho-Kinase and RTK Array kits, respectively. The Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) was used as developing reagent. Data analysis was performed through Image Lab Software (Bio-Rad) in a ChemiDoc XRS (Bio-Rad).

Statistical Analysis

One-away and Two-way ANOVA followed by Tukey's multiple comparisons tests and Unpaired T tests were used to determine the different expression patterns of CD44 and its isoforms in BlCa cells and tissues, as well as in healthy tissues (RT-PCR and flow cytometry), and to test the effect of CD44 silencing in BlCa cells and glycoengineered cell models in functional responses (invasion, proliferation). Differences were considered significant for $p<0.05$. All experiments were performed at least in triplicates and three replicates were conducted for each independent experiment. The results are presented as the average and standard deviation of these independent assays. For the TCGA series, the Shapiro–Wilk normality test was used to determine variable normality. CD44 isoforms log2 (normalized RSEM+1) transformed expression levels were visualized as boxplots. Statistically significant differences between two groups were evaluated using the nonparametric Wilcoxon test, whereas Kruskal-Wallis tests were used for comparison of multiple groups. Spearman correlation analyses were performed to calculate correlation coefficients. Patients were separated into two groups

according to the expression levels of each CD44 isoform, using the 25th and 75th percentiles as the cut-off point. Kaplan-Meier (K-M) survival curves were generated to compare the survival between patients with high and low expression levels. The statistical significance between the curves was determined using the log-rank test. The univariate and multivariate Cox proportional hazard regression models were performed to determine independent factors associated with prognosis. All statistical analysis were performed using R software (3.6.2), and *p*-values < 0.05 were considered statistically significant.

Data Availability

Proteomics datasets and tables describing protein assignment glycosylation annotations have been deposited in the Proteomics Identifications Database (PRIDE; https://www.ebi.ac.uk/pride/; Project accession: PXD028307) and RNAseq was deposited in the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) and are freely available. All other study data are included in the article and/or SI Appendix. Transcriptomics analysis of healthy tissues for CD44 isoforms were obtained from the Genotype-Tissue Expression (GTEx) Portal (https://gtexportal.org) on 03/06/21 and/or dbGaP accession number phs000424.v8.p2.

Results and discussion

The CD44 glycoproteocode remains poorly characterized in BlCa and other tumors, frustrating expectations for precise clinical interventions. This is closely related to CD44 high molecular heterogeneity, resulting from intense alternative splicing, close sequence similarity between multiple proteoforms, and dense and diverse *O*-glycosylation, amplified by conflicting nomenclature. Herein, we show that multi-omics, combining transcriptomics, glycomics, and glycoproteomics in glycoproteogenomics settings, are required for precise identification of CD44 signatures of clinical relevance.

High CD44s/st mRNA is associated with muscle invasion and worst prognosis in bladder cancer

We have started by screening by gene expression using Taqman probes the healthy urothelium (11 cases) and a broad series of bladder tumors (34 non-muscle

invasive and 41 muscle-invasive bladder tumors) representative of all disease stages for total CD44 mRNA and 5 exon-exon junctions (*e5-v2*; *e5-v3*; *e5-v8*; *e5-e15*; e2(partial)-e3(partial); Table S1) frequently found in cancer models [39-43]. The e5-v2, e5-v3 and e5-v8 junctions are also frequently observed in experimentally confirmed splicing isoforms such as *CD44v2-10*, *CD44v3-10*, *CD44v8-10* (Figure S1) but may also be present in many others showing variable exon combinations. On the other hand, e5-e15 and e2(partial)-e3(partial) detect shorter *CD44s/st* and *CD44sol*, respectively. For simplification and easy follow-up, we have adopted the nomenclature proposed by Azevedo, R. *et al.* [2], which reflects the nature of the transcribed exons. Notably, due to high similarity, this approach was unable to differentiate *CD44*s from the close related *CD44st* isoform containing a shorter cytoplasmic tail (Figure S1).

In general, *CD44* mRNA was expressed more abundantly in the healthy urothelium than in bladder tumors and showed a trend to decrease with the severity of the disease (Figure 1A). Furthermore, healthy tissues and tumors presented mRNAs encoding for all targeted exon-exon junctions, supporting high microheterogeneity. This hypothesis was further reinforced by the fact that the studied isoforms only account for approximately 30% of the total *CD44* mRNA in the healthy urothelium and do not exceed 60% of *CD44* transcripts found in tumors The existence of a more complex mixture of variable isoforms not fully captured by this approach is likely, especially for the healthy urothelium. Zooming in on the nature of the transcripts, we found that superficial tumors were enriched for variants presenting e5-v2 and e5-v3 junctions generally found in longer isoforms (namely *CD44v2-10 and/or CD44v3-10*) in comparison to muscle-invasive tumors (Figure 1A). The e5-v8 junction was elevated in cancer, irrespectively of disease stage. On the other hand, invasive tumors expressed higher percentages of shorter CD44 mRNAs lacking the variable region (*CD44s/st*). *CD44sol* mRNA was either vestigial or undetected in most samples and, therefore, has not been represented in Figure 1A**.** Collectively, these observations suggest that muscle-invasive tumors may be enriched for shorter mRNA isoforms missing variable exons in comparison to other isoforms.

We then comprehensively interrogated a larger and more homogeneous patient series comprehending over 400 muscle-invasive tumors from The Cancer Genome Atlas (TCGA) with a detailed clinical history (Figures 1B-D). We found that muscle-invasive tumors presented *CD44v2-10*, *CD44v3-10*, *CD44v8-10* and *CD44s/st* variants containing the exon-exon junctions accessed in bladder tumors

by RT-PCR, as well as *CD44v10*. Even though the presence of other transcripts resulting from multiple variable exon combinations cannot be excluded, these observations reinforced the complex nature of the splice code. We started by observing that the relative abundance of these isoforms in relation to the total CD44 mRNA did not vary with the stage of disease (T2-T4; data not shown). As such, focus was on the total isoform expression foreseeing the identification of more abundant targetable signatures. *In silico* analysis revealed that late-stage invasive disease (>T2) and more aggressive non-papillary lesions [44], presented significantly higher levels of short *CD44* transcripts (*CD44sol; CD44s; CD44v10*) in comparison to T2 (Figure 1B and S2) and papillary tumors (Figure S2), respectively. Furthermore, increased *CD44*s (p=0.029) and *CD44*v10 (p=0.049) were significantly associated with decreased overall survival (Figure 1C). However, none of them were independent predictors of poor prognosis when adjusted to disease stage, which was a relevant variable in this context.

Finally, CD44 transcripts were evaluated in the context of transcriptome-based molecular subtypes of BlCa. Namely, the luminal subtype, with favorable prognosis, overexpressing *KRT20* as well as *FOXA1*, and *GATA3*, and the basal-like subtype, displaying worst prognosis, high *KRT5* and/or *KRT14* and/or *KRT6A* [45, 46]. In the analysis, we also included known epithelial-to-mesenchymal transition (EMT) markers closely linked to invasion (*FOXF1*, *CDH2*, *ZEB1*, *ZEB2*, *SNAI1*, *TWIST1*). The Spearman correlation plot in Figure 1D highlights three clear clusters corresponding to these groups. Some proximity between the basal-like and EMT molecular phenotypes could also be observed, supporting the existence of a basal subgroup enriched for more invasive traits. The luminal molecular subtype was not characterized by *CD44* expression. On the other hand, high *CD44v2-10, v3-10, and v8-10* were characteristic of the basal-like group, whereas *CD44*v10 and *CD44*s were characteristic of the EMT, reinforcing the close link between shorter isoforms resulting from mRNA processing and invasive traits. Notably, in this cohort, the three subgroups presented similar survivals explained by the aggressive nature of the tumors. Nevertheless, we showed the existence of different molecular subtypes for tumors of apparently similar histology and outcome, including CD44 signatures that should be confirmed foreseeing precise cancer targeting.

Collectively, transcriptomics strongly supports the co-existence of multiple CD44 isoforms in the healthy urothelium and cancer. It also highlights the close link between the presence of shorter isoforms missing all variable exons and muscle-invasive disease, characterized by unfavorable outcomes. Nevertheless,

confirmation by high-throughput proteomics is required for definitive confirmation.

CD44s variant is expressed in healthy human tissues

Precise targeting requires significant cancer specificity. Therefore, we also evaluated a wide number of healthy tissues that exhibited variable but always detectable *CD44* mRNA expression (Figure 1E). This highlighted the complex mosaicism presented by CD44 in human organs, later confirmed by consultation of the GTEx database (Figure S3). In addition, RT-PCR (Figure 1E) and whole transcriptome analysis (Figure S3) show that, despite overexpressed in bladder cancer, CD44s may also be strongly expressed by most healthy cells, which significantly challenges its cancer specificity. Therefore, addressing CD44s post-translational modifications, namely glycosylation, poses as the next logical step towards this objective.

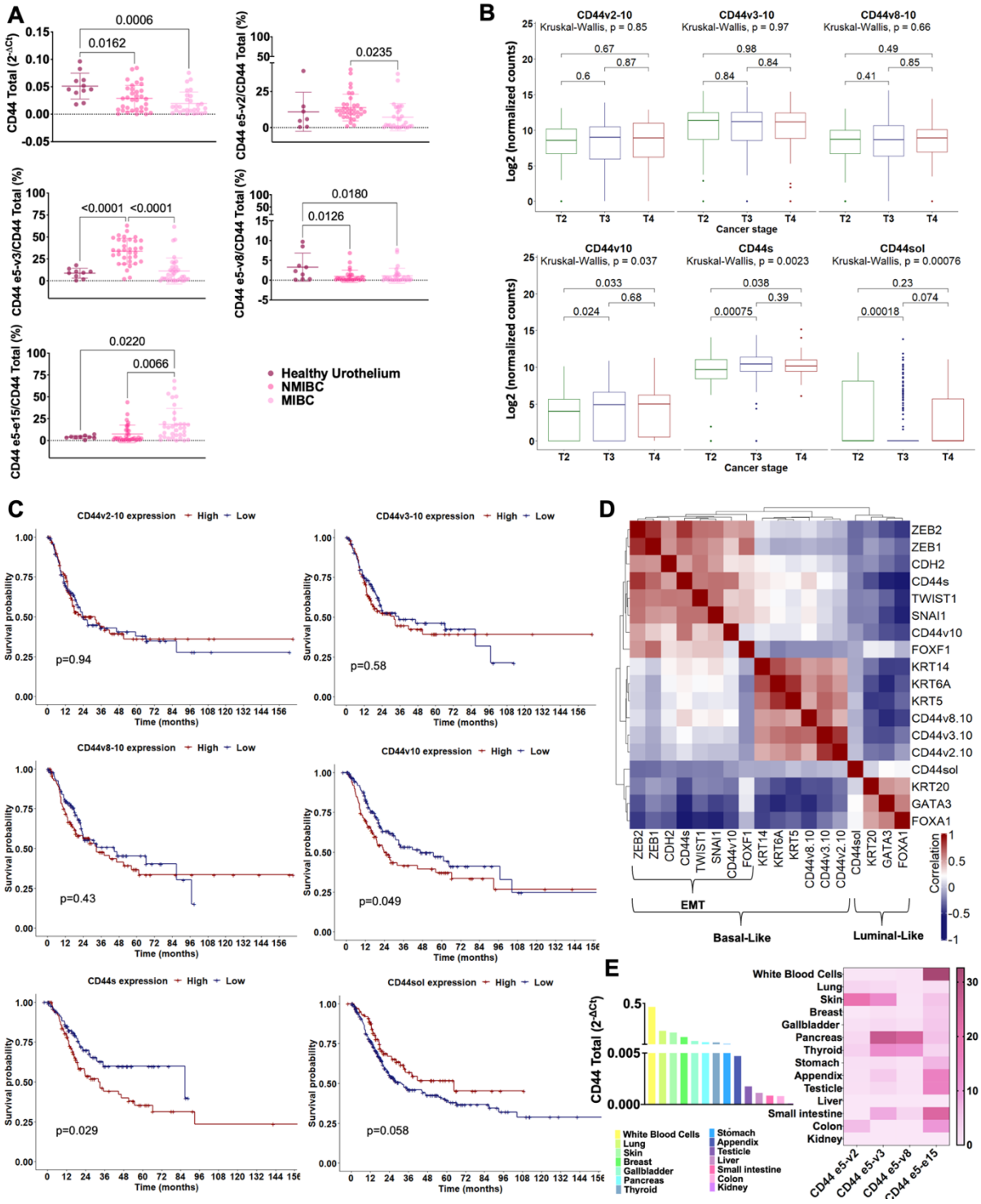


Figure 1. *CD44s* is increased in more aggressive bladder tumors in comparison to superficial lesions, is not present in the healthy urothelium and shows limited expression in other healthy human tumors. A) Expression of CD44 and cancer relevant exon-exon junctions in the healthy urothelium and bladder tumors. *CD44* gene expression significantly decreased in bladder tumors in comparison to the healthy urothelium. Furthermore, the nature of the variants changed with the severity of disease, with non-muscle invasive bladder cancer (NMIBC), showing higher expression of *e5-v2* and *e5-v3* generally found in lengthier isoforms (*CD44v2-10*; *CD44v3-10*) and a significantly lower abundance of *e5-e15* characteristic of CD44s/st in comparison to muscle-invasive bladder cancer (MIBC). MIBC presented an opposite pattern, supporting enrichment for shorter CD44s

isoform. The *CD44sol* isoform was also evaluated but its expression was vestigial and not linked to any tumor type, as such it was not represented in the panel A. **B) TCGA analysis of 413 MIBC cases confirmed the association of CD44s with more aggressive late-stage disease**. Briefly, *CD44s* mRNA was elevated in T3/4 tumors in comparison to T2 tumors, supporting its association with invasion, confirming the observations from our patient's dataset. **C) Elevated *CD44v10* and *CD44s* mRNA significantly associate with worst prognosis in MIBC**. The Kaplan-Meier curves highlight a clear link between the expression of shorter CD44 isoforms and worst prognosis in bladder cancer. Also, despite its low expression, patients with tumors presenting high *CD44sol* mRNA presented better prognosis. **D) CD44s is expressed by a subgroup of basal-like tumors enriched for genes defining mesenchymal traits.** In general, *CD44sol* associated with more differentiated and less aggressive luminal tumors, whereas the other forms of CD44 related with basal phenotypes, frequently less cohesive and poorly differentiated lesions. However, *CD44s and CD44v10* were characteristic of a subgroup of basal tumors enriched for *ZEB1/2, CDH2, TWIST1, SNAI1* that define mesenchymal phenotypes. Collectively, these observations link shorter CD44 isoforms to bladder cancer invasion and poor prognosis. **E) Total *CD44* and isoforms expressions in relevant healthy cells and organs.** The *CD44* gene presents a heterogeneous expression pattern in healthy tissues, and multiple isoforms may coexist in the same organ. Notably, *CD44s* was expressed by most of the studied tissues, showing significantly high mRNA levels in white blood cells. The gastrointestinal and colorectal tracts as well as the testicle also present high *CD44s*, however, with low total *CD44* levels. Collectively, this result shows that CD44s is not a cancer specific signature. The results correspond to the mean and standard deviation for three independent experiments. Triplicate measurements were conducted for each experiment. P values are presented for one-way ANOVA, nonparametric Wilcoxon, and Kruskal-Wallis tests.

BC expresses CD44 glycoforms not observed in relevant healthy cells/organs

In the past, we and other groups have demonstrated how glycosylation may increase the cancer specificity of proteins, which could be explored for precise cancer-targeting [25, 31, 32, 34, 47, 48]. CD44 presents multiple potential *O*-glycosylation sites in its extracellular domain, mostly in the variable region, which could be explored towards this objective [2]. We hypothesized that CD44 may reflect the profound alterations occurring in glycosylation pathways in BlCa, which have been directly linked to aggressive traits. As such, we have screened BlCa sections of different histological natures by immunohistochemistry for signs of co-localization between CD44 and the STn antigen, an immature *O*-glycan promoter of invasion [49, 50], immune escape [38], and an independent predictor of poor prognosis [51]. We have also studied the Tn antigen, the shortest *O*-glycan, whose expression in BlCa has been linked to cancer aggressiveness in explorative studies. To avoid any possible cross-reactivity between the VVA lectin used to detect the Tn antigen and blood group A antigens, tissue sections positive for the A antigen were excluded from the analysis. According to Figures 2A-B, neither the Tn nor the STn

could be observed in healthy urothelium, except for some Tn positivity in the cytoplasm of upper stratum umbrella cells. STn antigen levels were significantly elevated in tumors in comparison to the healthy urothelium, being statistically significant for invasive lesions (Figure 2A), as previously reported by us [37, 46, 51, 52]. It was mostly detected at the surface of cancer cells in both superficial and invasive tumor layers, being more intense in invasive fronts in accordance with its functional role in disease progression [46, 47]. The less studied Tn antigen was not significantly overexpressed in bladder tumors, except for a subgroup of patients at more advanced stages. In BlCa, the Tn antigen was predominantly cytoplasmic, most likely due to the detection of immature glycans in protein secretory pathways. For 47% of the cases (data not shown), membrane expression was also evident, but no pattern could be established regarding location, proximity to vessels or any other histopathological characteristics. Nevertheless, all tumors exhibiting abnormal glycosylation, also presented areas of CD44-Tn and/or STn co-expression (Figure 2B), irrespectively of their histological nature. Also, *in situ* PLA (Figure 2C) and double staining immunofluorescence (Figure 2D) for CD44 and STn and Tn antigens, respectively, revealed close spatial proximity between CD44 and the glycans in tumors but not in the healthy urothelium. Notably, positive PLA for CD44-STn was mainly observed in invasive fronts (Figure 2C) consistent with the role played individually by CD44 and STn in cancer invasion [11, 29].

We have further investigated the cancer specificity of the CD44-Tn/STn proteoforms in healthy tissues. STn expression was either absent or low in secretions and cells facing the lumen of the respiratory, gastrointestinal, and colorectal tracts (Figure 2E), in accordance with our previous reports [31, 47]. Amongst STn-positive tissues, we found possible co-localization between the glycan and CD44 in the stomach, appendix, small intestine, colon, the gallbladder, and white blood cells in MALT (Figure 2E) and in peripheral blood mononuclear cells from healthy donors (data not shown). Interestingly, these organs also showed higher levels of cancer associated CD44s. However, orthogonal validation by double staining immunofluorescence and PLA did not confirm possible CD44-STn glycoproteoforms in healthy tissues, suggesting cancer specificity (Figure 2E). On the other hand, the Tn antigen was circumscribed to the cytoplasm of goblet cells at the intestinal tract, Leydig cells at testicle, pancreatic acini, hepatocytes, mucinous cells at gastric epithelium, alveolar macrophages at lung and gallbladder epithelium. Immunohistochemistry suggested some degree of overlap between Tn and CD44 in the pancreas, testicle, and gallbladder, which showed low CD44s

expression (Figure 2E). However, the presence of CD44-Tn glycoproteoforms was not supported by immunofluorescence (Figure 2E). Collectively, according to our observations, short-chain *O*-glycans provide cancer-specificity to CD44 and the means for precise cancer targeting. We also identified a clear CD44s/st^{high}/STn^{high} phenotype in 15% of MIBC patients, characterized by high *e5-15* transcripts (>50% in relation to total CD44; Figure 1A) and STn overexpression (>75% of the tumor area) in co-localization with CD44. These observations strongly suggest that this isoform maybe carrying altered glycosylation not observed in the healthy urothelium and mostly absent from healthy tissues.

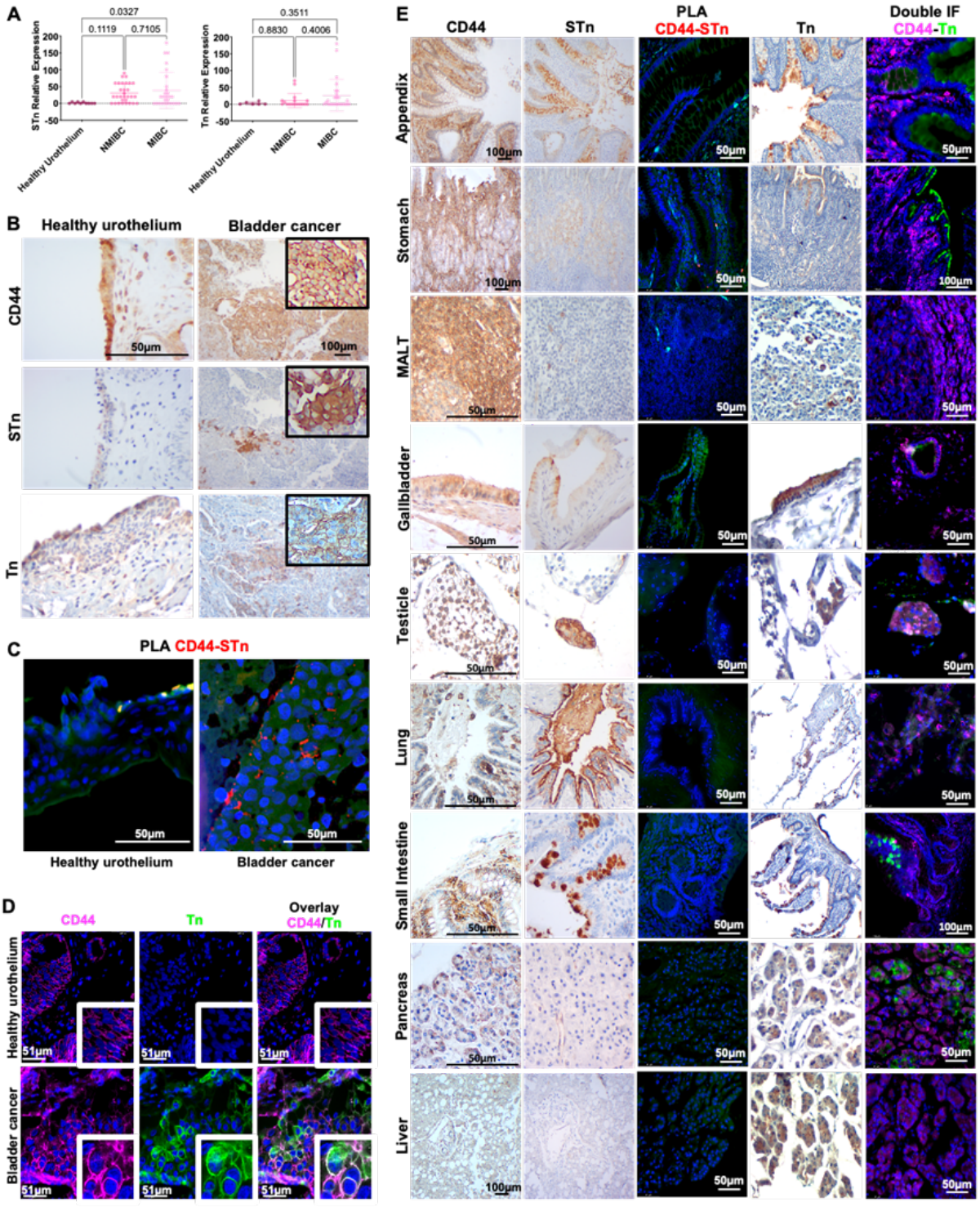


Figure 2. CD44-Tn and STn glycoproteoforms present high cancer specificity. A) Tn and STn antigens are not expressed in the healthy urothelium, are elevated in cancer, and increased in invasive tumors. The STn antigen, defined by affinity for the anti-tag-72 antibody [B72.3+CC49], is not expressed in the healthy urothelium and is elevated in cancer, being significantly overexpressed in muscle invasive bladder cancer (MIBC). The Tn antigen, based on VVA lectin immunoaffinity, is also not expressed in the healthy urothelium and is elevated in a subset of patients, specially at more advanced stages. **B) CD44s colocalizes with Tn and STn antigens in MIBC.** CD44 was diffusively

expressed across the tumor section without a defined pattern. The STn antigen was observed both in superficial and invasive layers and the Tn antigen was found in scattered niches without a defined expression pattern. Immunohistochemistry also showed the co-localization of CD44 with Tn and STn positive areas in CD44s^{high} tumors, exhibiting low amounts of other isoforms (according to RT-PCR, data not shown). The healthy urothelium expressed high amounts of CD44 and the Tn antigen is present in the cytoplasm of upper stratum umbrella cells while the STn antigen was not detected. **C) *In situ* proximity ligation assays (PLA) supports CD44s-STn glycoproteoforms in MIBC and its presence within tumor invasive fronts.** *In situ* proximity ligation assay showed close spatial proximity between CD44 and STn in the same cells, strongly supporting CD44-STn glycoproteoforms in tumors. This phenotype was mostly observed in invasive fronts of CD44s^{high} tumors and was not detected in the healthy urothelium, suggesting cancer-specificity. **D) Double staining immunofluorescence supports CD44s-Tn glycoproteoforms in bladder cancer.** CD44s^{high} tumors presented niches of cells co-expressing CD44 and Tn antigen, strongly suggesting CD44-Tn glycoproteoforms. This was not observed in the healthy urothelium. **E) Glycosylation with Tn and STn antigens provides cancer specificity to CD44.** CD44 was abundantly expressed in all studied healthy tissues. STn expression was either absent or low in secretions and cells facing the lumen of the respiratory, gastrointestinal, and colorectal tracts. Immunohistochemistry suggested STn and CD44 co-localization in the stomach, appendix, small intestine, colon, gallbladder, and white blood cells in MALT. PLA did not confirm these hypotheses, suggesting cancer specificity of CD44-STn glycoproteoforms. In healthy tissues, the Tn was restricted to the cytoplasm of goblet cells in the intestinal tract, Leydig cells in testicular tissue, pancreatic acini, hepatocytes, mucinous cells of the gastric epithelium, alveolar macrophages, and gallbladder epithelium. CD44 and Tn antigen co-expression was suggested in pancreatic tissue, testicle, and gallbladder, which was not confirmed by double staining immunofluorescence.

Glycoproteogenomics identifies multiple glycoproteoforms in BlCa cells

Transcriptomics

Transcripts analysis strongly supported that the splicing code changes with cancer and varies with the different stages of the disease, which require precise identification by proteomics. Therefore, we started by choosing cell models to support the implementation of a mass-spectrometry based roadmap to address this challenge. Three well established BlCa cell lines reflecting distinct stages of the disease (derived from a grade I (RT4), a grade II (5637) and grade III (T24) carcinomas) were evaluated for the capability to invade Matrigel *in vitro* and screened for CD44 by real time PCR using Taqman assays. These assays were latter complemented with random and exon 16 primer specific RT-PCR followed by Sanger sequencing, RNAseq, flow cytometry, and western blotting. CD44 expression increased alongside with cell lines invasive capacity (T24>5637>RT4; Figures 3A-B), in accordance with the histopathological nature of the carcinomas from which cell lines derived. We also found that less invasive RT4 and 5637 cells

expressed lower levels of CD44 in comparison to T24 cells (Figures 3B-C), being also enriched for longer CD44 variants (Figures 3C-E and S4). Interestingly, RT4 cells presented a higher number of transcripts corresponding to e5-v2 and e5-v3 junctions, whereas e5-v8 was the dominant signature in 5637 cells (Figure 3D and S4). Transcripts containing exon e15-e15 junctions characteristic of CD44s/st were much less abundant in these cell lines (>10% of total *CD44* expression; Figure 3D and S4), even though it increased in 5637 in relation to RT4 cells. On the other hand, T24 cells expressed high levels of CD44s and/or st (approximately 75% of total *CD44*) and residual amounts of other exon-exon junctions (Figures 3C-E and S4), in agreement with patient samples observations linking this variant to invasive traits. Taken together with patient samples data, these observations suggest that aggressiveness may be accompanied by a progressive loss of variable exons and, consequently, a decrease in CD44 chain length, which warrants definitive confirmation. Interestingly and in agreement with the analysis of human tissues, approximately 20-30% of *CD44* transcripts did not present any of the studied exon-exon junctions. It is probable that these cell lines may express other less abundant molecular motifs confirming the complexity of CD44 splicing code, as also suggested by exon 16 primer specific RT-PCR followed by Sanger sequencing (Figure S4) and RNAseq (Figure 3E). Namely, Figure S4 shows transcripts containing e5-v4 and e5-v10 junctions in T24 and e5-v8 in both T24 and 5637 cells, all also presenting exon 16. Alongside, RNAseq analysis also considered probable transcripts containing the following junctions: e5-v6; e5-v7; e5-v9, e3-v8, and e3-v16, to name a few (Figure 3E), which still warrant definitive confirmation by other methods. Collectively, these findings support the complex nature of the bladder cancer splicing code suggested by the analysis of tissues. Nevertheless, we considered T24 cell line as a suitable model to study the influence of CD44s/st in BlCa, since the other variants are practically vestigial. Moreover, based on the close similarity between RT4 and 5637, we selected 5637 and T24 cells for downstream proteomics-based studies.

In explorative studies using conventional bottom-up proteomics approaches, we concluded that the precise characterization of CD44 glycoproteoforms could only be accomplished by the customization of the databases used for protein annotation and by upfront knowledge of the cellular glycome. Through transcriptome analysis of 5637 and T24 cell lines, we discovered 38 different RNA transcripts of CD44, 18 of which have been previously described (Figures 3D-E; Table S3). Amongst these, 8 transcripts had already been

experimentally validated (including CD44v2-10, CD44v3-10, CD44v8-10 and CD44s herein addressed by RT-PCR; Figure 3D) and 3 were related to non-coding RNA (Figure 3E; Table S3). The remaining 7 transcripts correspond to computationally inferred short transcripts (Figure 3E; Table S3). However, half lacked open-reading frames (ENST00000526000.6 [Uniprot:H0YDW7]; ENST00000279452.10 [Uniprot:H0Y2P0]; ENST00000528455.5 [Uniprot:H0YD17]; ENST00000531873.5 [Uniprot:H0YD90]) are unlikely to be translated into proteins, while another contained a premature translation-termination codon (ENST00000425428.6 [Uniprot:Q86UZ1]). Transcript ENST00000526669.6 (Uniprot:H0YD13) encoded an incomplete extracellular domain, lacked membrane anchoring and/or intracellular region, and presented low transcript levels, suggesting reduced probability of occurrence. Finally, transcript ENST00000442151.6 (Uniprot:H0Y5E4) presents almost 100% homology with CD44st with the exception of a minor 5'truncation corresponding to the first amino acid in the protein sequence. In summary, only the computationally determined variant ENST00000442151.6 was considered for downstream validation by mass spectrometry. The remaining 20 transcripts corresponded to previously undescribed RNA sequences, probability corresponding to short non-coding transcripts, whose biological role should be investigated in the future. Collectively, transcriptomics confirmed the intense alternative splicing of CD44 in T24 cells, leading to several short transcripts encoding proteins lacking a variable extracellular region. After excluding ambiguous transcripts, a total of 9 sequences were considered plausible to generate proteins and were initially selected to construct the CD44 protein database (Table S3).

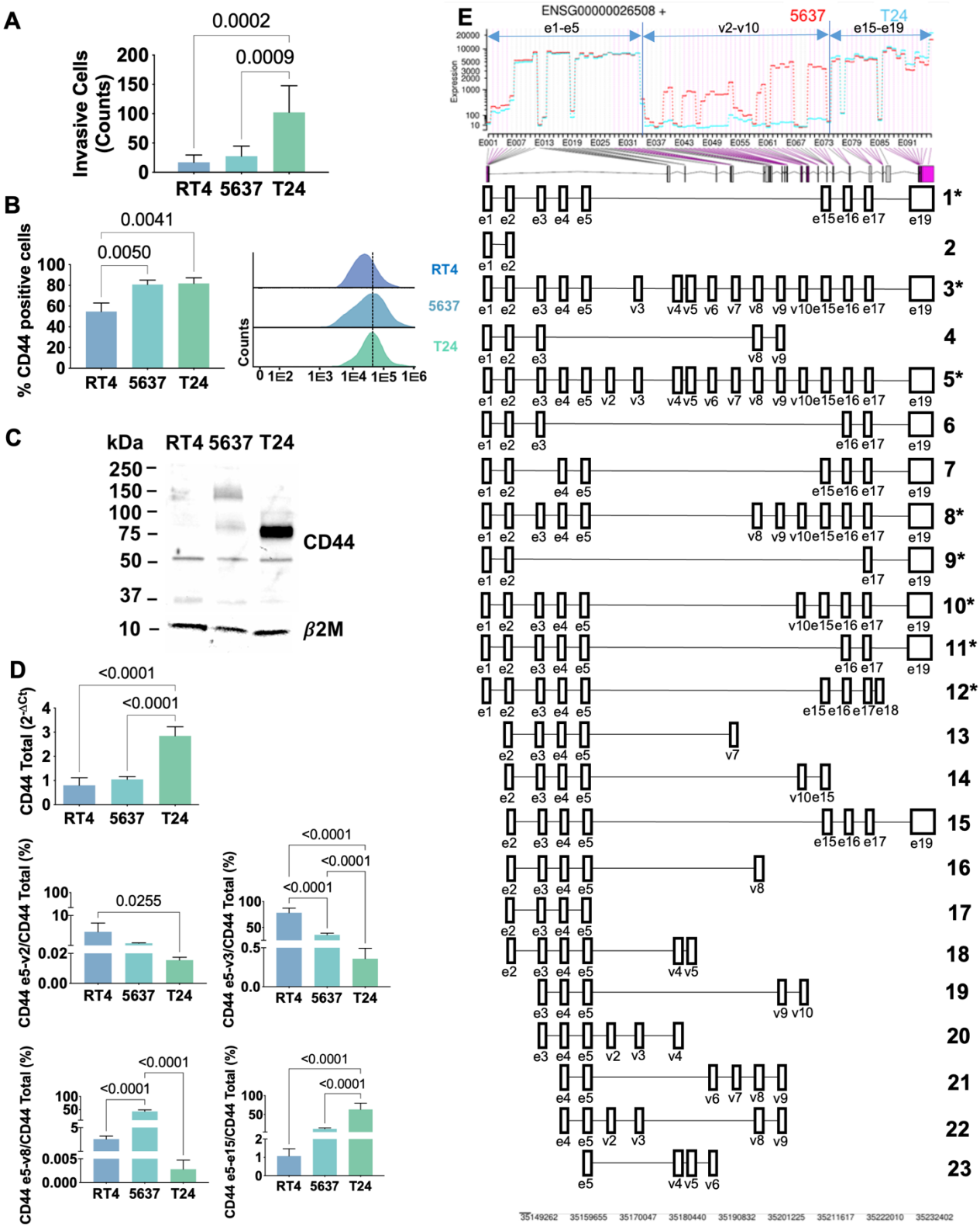


Figure 3. Highly invasive T24 bladder cancer cells express high levels of CD44s. A) Capacity to invade Matrigel *in vitro* for RT4, 5637 and T24 cells. Grades I/II cell lines (RT4 and 5637) are significantly less capable of invading Matrigel *in vitro*. **B) CD44 expression is higher for grade II (5637) and III (T24) cells.** Flow cytometry analysis showed significantly higher CD44 levels in 5637 and T24 compared to RT4 cells. **C) Western blots show different CD44 expression patterns according to cell grade, with T24 cells showing mostly shorter proteoforms.** WB confirmed the

overexpression of shorter CD44 proteoforms (at approximately 75 and 50 kDa) in T24 cells, and the presence of heavier proteoforms in the other cell lines (above 150 kDa). **D) Characterization of CD44 isoforms by RT-PCR showing isoforms shortening with cell lines aggressiveness and the marked CD44s^{high} phenotype of T24 cells.** *CD44* gene expression was significantly higher in T24 in comparison to RT4 and 5637 cells, in agreement with protein analysis. RT-PCR also revealed increasing mRNA shortening with cells aggressiveness. Accordingly, RT4 showed higher *e5-v2* and *e5-v3* transcripts in comparison to the other cell lines, 5637 predominantly presented *e5-v8*, whereas T24 cells mainly expressed *e5-v15* characteristic of CD44s/st. **E) RNAseq confirmed the marked difference between T24 and 5637 cells, the first expressing shorter CD44 mRNAs in opposition to lengthier mRNAs in 5637 cells.** The top image is a DESeq2's plot for visualization of alternative splicing events. The plot shows the coverage of constitutive and variable exons for 5637 (in red) and T24 (in blue) cells. The bottom part of the image shows the gene structure and depicts differentially expressed exons (purple lines linked to the x-axis of the coverage plot). The main differences are found in the variable region spanning v2-v10, which was mostly missing in T24 cells. Predicted transcripts are shown in the bottom. Asterisks highlight those showing transcripts higher probability of generating proteins and, therefore, used for glycoprotegenomics. Legend for panel E with Uniprot and Ensemble codes: **1**- P16070-12 (CD44-201); **2**- E9PKC6 (CD44-222); **3**- P16070-4 (CD44-206); 4-CD44-227 (no Uniprot code); **5**- P16070-1 (CD44-208); **6**- H0YD13 (CD44-224); **7**- Q86UZ1 (CD44-207); **8**- P16070-10 (CD44-209); **9**- P16070-19 (CD44-203); **10**- P16070-11 (CD44-210); **11**- P16070-18 (CD44-205); **12**- H0Y5E4 (CD44-211); **13**- H0YDW7 (CD44-221); **14**- H0Y2P0 (CD44-204); **15**-CD44-232 (no Uniprot code); **16**- H0YD17 (CD44-228); **17**-H0YD90 (CD44-234); **18**- CD44-213 (no Uniprot code); **19**- H0YEU1 (CD44-231); **20**- H0YEV3 (CD44-219); **21**- H0YES0 (CD44-212). The results for A-D correspond to the mean and standard deviation for three independent experiments. Triplicate measurements were conducted for each experiment. *P* values are presented for one-way ANOVA tests.

Glycomics

In addition, we have characterized the 5637 and T24 cells *O*-glycome by MALDI-MS. Adding to distinct clinicopathological features, these cells exhibited slightly different *O*-glycomes in terms of glycan abundance (Figure 4A). Both cells predominantly expressed core 1 derived structures, namely fucosyl-T (*m/z* 768.38; $[M+Na]^+$) and mono- (*m/z* 955.46) and di-sialylated T antigens (*m/z* 1316.64) (Figure 4A). However, in 5637, fucosylated and mono-sialylated T antigens were more abundant, whereas in T24 it was disialyl-T. The presence of several extended core 2 glycans (*m/z* 1217.60; m/z 1391.69; 1404.69; 1578.78; 1765.86) were also observed, being however more pronounced in 5637 cells. Low amounts of core 3 (*m/z* 635.32) and T (*m/z* 594.29) antigens were found in both cells as well as STn antigens (*m/z* 751.36) in T24. Overall, there were no significant differences comparing these cells *O*-glycome that could account for alteration in invasion. Namely, no significant changes were observed in terms of sialylation, fucosylation, glycosidic chains length or overexpression of immature *O*-glycans such as the Tn

and STn antigens, generally associated with changes in invasion. Notably, these cells missed high levels of Tn and STn antigens highly expressed in tissue samples linked to invasion and CD44 (Figure 2), which can be explained by strong dependent on microenvironmental cues not reflected *in vitro*. Nevertheless, key information was generated to guide CD44 glycosites identification, contributing to increase protein coverage in glycoproteogenomics settings.

Glycoproteogenomics

For glycoproteoforms identification (Figure S5) we first immunoprecipitated CD44 from plasma membrane-enriched protein extracts isolated by differential ultracentrifugation. To ensure broad CD44 representation, we adopted an antibody targeting the cytoplasmatic tail. This avoided glycosylated regions of the protein, which may experience significant structural variations due to the dynamic nature of this post-translational modification, thus affecting antibody recognition. We first separated different proteoforms by gradient SDS-PAGE supported by western blot, excised the bands from gels (Figure 4B) and applied a proteomics workflow contemplating digestion with chymotrypsin and protein identification by nanoLC-HCD-MS/MS. This approach was complemented by CID-MS/MS triggered by the presence of the oxonium ion HexNAc (*m/z* 204.087) in the HCD-MS/MS. To render the glycosylation more homogeneous and facilitate downstream assignments by MS/MS, the glycoproteins were desialylated prior to proteolytic digestion. The western blots in Figure 4B started by highlighting multiple proteoforms across a wide range of molecular weights in BlCa cell lines (Figures 4C-D), as suggested by transcripts analysis. Western blots also reflected the abundance of CD44 in the cell lines, previously suggested by RT-PCR and later confirmed by flow cytometry. CD44 molecules bigger than 150 kDa predominated in 5637 cells, whereas T24 cells presented a major band at 75 kDa. Interestingly, both cell lines presented several well defined proteoforms below 37 kDa.

In explorative settings, we observed that unambiguous identification of peptides corresponding to exon-exon junctions was not possible without considering *O*-glycosylation, even when using transcriptome-customized databases. An integrative glycoproteogenomics approach was then applied, which enabled the identification of several of these junctions (detailed in Figure 4B) with high degree of confidence, based on specific diagnostic glycopeptides (Figures 4C, S6 and ST; Tables 1, S4 and S5). Using HCD fragmentation, we could identify several product ions consistent with the presence of glycosylated moieties, together with

some *y*- and *b*-type peptide fragments supporting these reporter glycopeptides, even though with some ambiguity in terms of glycosites assignment (Figure 4C). The presence of GalNAc, the first sugar residue in *O*-glycans, could be further confirmed by several HexNAc cross ring fragments (*m/z* 126.055,168.066,186.076, and 204.087; Figures 4C-D and Figures S6-7 and 11) and, in some cases, high 144.066/138.055 oxonium ions ratios [53, 54]. We also explored HexNAc-triggered CID fragmentations for further glycopeptide validation, including the characterization of extended glycan chains (Figure S6). For certain species, CID-based fragmentation retrieved a significant number of peptide backbone fragmentations, enabling unequivocal glycosites characterization, thus in accordance with our previous reports using a similar approach [55]. Notably, we found CD44 glycoproteoforms carrying both short- and elongated glycans, as well as the co-existence of different glycoforms in the same peptide sequence, even though in distinct glycosites (Figure 4C; Tables S3 and S4; Figure S6). Also, despite subjected to desialylation, some glycopeptides carrying sialic acids were still observed due to incomplete enzymatic digestion (Figures 4C and S6). Nevertheless, such observations, reinforce the tremendous molecular micro-, macro-, and meta-heterogeneity presented by CD44, previously reported for other human glycoproteins [24], portraying the relevance of glycosylation towards unequivocal molecular characterization.

SDS-PAGE-nanoLC-MS/MS analysis confirmed CD44 above 75 kDa in 5637 cells, whereas in T24 it spanned from 150 to 25 kDa (Figure 4B, Table 1), thus in accordance with the high *vs* low molecular weight splicing codes suggested by transcript analysis and western blots (Figures 3, 4B and S4). The intense band at 50 kDa in the blots may, at least partially, result from the anti-CD44 antibody used for IP, since no CD44 isoforms were identified in 5637 cells. Several CD44 glycopeptides were also found at between 37 and 25 kDa, including v2-v3 or e5-v3 exon-exon junctions. Interestingly, the corresponding variants containing all the constitutive and variable exons (CD44v2-10 and CD44v3-10, respectively) present predicted molecular weights at 80 and 75 kDa. We hypothesize that these bands may instead contain CD44 short isoforms presenting combinations of variant exons or constitutive exons which can become variable. Products of proteolytic cleavage of CD44 ectodomain by membrane-associated metalloproteases can also be found at these molecular weights and are common in cancer [56], further contributing to what we believe to be a mixture of CD44 products. The possibility of proteolysis occurring at the variable region mediated by other extracellular proteases such as

metalloproteinases and ADAM has also been suggested and may also account for these findings [57, 58]. In the future, efforts should be devoted to resolving these bands towards comprehensive identification.

The remaining bands also suggested that multiple isoforms could coexist in the same lane. However, the lack of MS/MS data supporting several exon-exon junctions hampers unequivocal confirmation. Nevertheless, we were able to characterize with significant detail multiple (glyco)peptides belonging to the constitutive region encoded by exons e1-e5 as well as exon e19 in both cell lines (Table 1; Figure S7). Exon-exon junctions linking the constitute region to v3 and e15 (e5-v3 and e5-e15; Table 1, Figures 4B-C and S6) were also observed in 5637 and T24. Even though we could not find e5-v2 we also successfully identified v2-v3 junctions (Figure 4C). The e2-v3 and e5-v3 glycopeptides were found many times in association with v2-v3, v3-v4 v4-v5 and (Table 1, S4 and S5), suggesting, but not unequivocally demonstrating, multiple sequential exons. The v10-e15 junction was also frequently observed together with these structural motifs at high molecular weights (>75-250 kDa), reinforcing this hypothesis. However, we were unable to identify exon-exon junctions within the v6-v10 region. In addition, we could not identify e5-v8 junctions, even in 5637 cells that presented high e5-v8 transcripts (Figure 3D and S4). We hypothesize that the density of glycosites in this region of the protein (Figure S1) may interfere with proteolytic digestion, generating high molecular weight glycopeptides not easily identified by the adopted analytical strategy. Complementary middle-down proteomics should be considered in the future for better coverage of multiple exons, specially across densely glycosylated areas. Another relevant finding was the identification of glycopeptides for the e5-e15 junction (Table 1, S4-S5Figures 4C and S6), diagnostic of CD44s. This was associated with high CD44s coverage (>60% of the CD44s sequence; Table 1). In T24 cells, CD44s was identified in bands spanning from 150 to 37 kDa (its non-glycosylated form), including major bands at 75 and 50 kDa, in agreement with its higher abundance in these cells (Figures 3D and 4B, Table 1). The fact that CD44s and potentially other isoforms could be found spanning different molecular weights may derived from differences in glycosites occupancy and glycosidic chains, as confirmed by Table S4 and S5. Interestingly, and despite low transcription (data not shown), our glycoproteogenomics approach also identified with high coverage the CD44sol isoform (65-100%; Table 1). However, this isoform does not present nor a membrane anchoring nor a cytoplasmatic domain that enables targeting by the anti-CD44 antibody chosen for IP. Moreover,

it has a predicted molecular weight of 15 kDa but has been consistently found in all bands and including at very high molecular weights. Cross-contamination is the most probable explanation, further reinforced by the fact that soluble forms of CD44 are known to bind vimentin [59] and proteoglycans [60, 61] and potentially many other yet unknown proteins. The presence of this polypeptide in most bands maybe an indicator of un-optimal SDS-PAGE [62] and/or nanoLC conditions, which should be carefully investigated and optimized in future studies.

Finally, we compared the number of glycosites identified for 5637 and T24 cells and found suggestions of isoform specific glycosylation according to the nature of the cell line. A higher number of glycosites were identified at the e5-e15 junction for T24 in comparison to 5637 cells. On the other hand, a higher number of glycosites were identified at the v2-v3 and e5-v3 junctions in 5637 cells. These findings are summarized together with the nature of the glycosidic chains in Figure 4D, but their functional meaning should carefully require evaluation in the future. We must also emphasize that the study was not designed to provide a precise identification of glycan chains, which should also be addressed in the future using dedicated protocols.

In summary, even though missing definitive confirmation regarding the composition of long isoforms presenting multiple variable exons, this strategy provide enabled the identification of CD44s diagnostic peptides. We have shown that CD44 characterization cannot be achieved by a single omics, requiring glycoproteogenomics settings. Moreover, addressing glycosylated domains is key for precise CD44 characterization. Blueprints for improving this glycoproteogenomics strategy towards more definitive CD44 splicing code characterization have been provided.

Table 1. Peptides corresponding to CD44 exon and exon-exon junctions identified by SDS-PAGE-nanoLC-MS/MS.

	5637		T24	
Mw (kDa)	**Exons and exon-exon junctions[1]**	**Assignment[2]** (Protein coverage)	**Exon and exon-exon junctions[1]**	**Assignment[2]** (Protein coverage)
>200-250	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3, v3-v4; v4-v5; v6 (partial); v10-e15; e19; CD44sol[3]	CD44v2-10 (46%) CD44 sol (96%)	No ID	-
200-150	No ID	-	No ID	-
150-100	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3, v3-v4; v4-v5; v6 (partial); v10-e15; e19; e5-e15	CD44v2-10 (49%) CD44s (59%)	CD44sol[3]	CD44sol (96%)
100-75	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3; v3-v4; v6; v10-e15; e5-e15; e19, CD44sol[3]	CD44v2-10 (43%) CD44s (68%) CD44sol (95%)	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3, v3-v4; v4-v5; v7 (partial); v10-e15; e19; e5-e15, CD44sol[3]	CD44v2-10 (48%) CD44s (58%) CD44sol (96%)
75-50	No ID	-	e1-e2; e2-e3; e3-e4; e5-v3; v2-v3, v3-v4, v4-v5; v10-e15, e19, e5-e15; CD44sol[3]	CD44v2-10 (47%) CD44v3-10 (49%) CD44s (73%) CD44sol (96%)
50-37	No ID	-	e1-e2, e2-e3, e3-e4, e4-e5, e5-e15	CD44s (59%)
37-25	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3; e5-v3; v3-v4; v10-e15, e19; e5-e15, CD44sol	CD44v2-10 (49%) CD44v3-10 (50%) CD44s (59%) CD44sol (97%)	e1-e2; e2-e3; e3-e4; e4-e5; v2-v3, v3-v4; v4-v5; v6 (partial); v7 (partial); v10-e15; e19; e5-e15	CD44v2-10 (49%) CD44s (58%)
25-20	No ID	-	No ID	-
20-15	No ID	-	No ID	-
15-10	No ID	-	No ID	-

Glycoproteogenomics identified cancer specific CD44s glycoproteoforms in bladder tumours

Transcriptomics together with different immunoassays strongly suggested but cannot present unequivocal demonstrations regarding the existence of CD44s-STn/Tn glycoproteoforms in BlCa. To address this limitation, our glycoproteogenomics approach was further employed to interrogate tumors for short glycoproteoforms specifically linked to BlCa aggressiveness. A protein pool from invasive tumors of five different patients showing high e5-e15 transcripts, CD44 western blot patterns resembling T24 cells (Figure 4D), and high STn and Tn antigens expressions was elected for this study. Immunofluorescence assays suggesting the presence of CD44-STn and Tn glycoproteoforms in the same tumor area led to consider these glycans during protein annotation. CD44-STn/Tn co-expressing areas were then excised from paraffin-embedded tumor sections after deparaffination, antigen retrieval, proteins were then extracted, and CD44 was isolated by immunoprecipitation and characterized as demonstrated for BlCa cell lines. Despite the significant degree of protein alterations induced by the conservation of the tumors during histological procedure, namely through multiple oxidations, we successfully identified e5-e15 glycopeptides in tumors (Figure 4D). Furthermore, we could identify the presence of short-chain cancer-associated *O*-glycans in both the CD44 constitutive region (Figure S7) as well as in CD44s reporter peptide sequences (Figure 4D). These findings provide proof of aberrant CD44s glycosylation in BlCa, since glycosylation with immature *O*-glycans such as the Tn and T antigens are not found in proteins of the healthy urothelium [31, 34, 52] (Figure 2) and rarely observed in healthy human tissues (Figure 2).

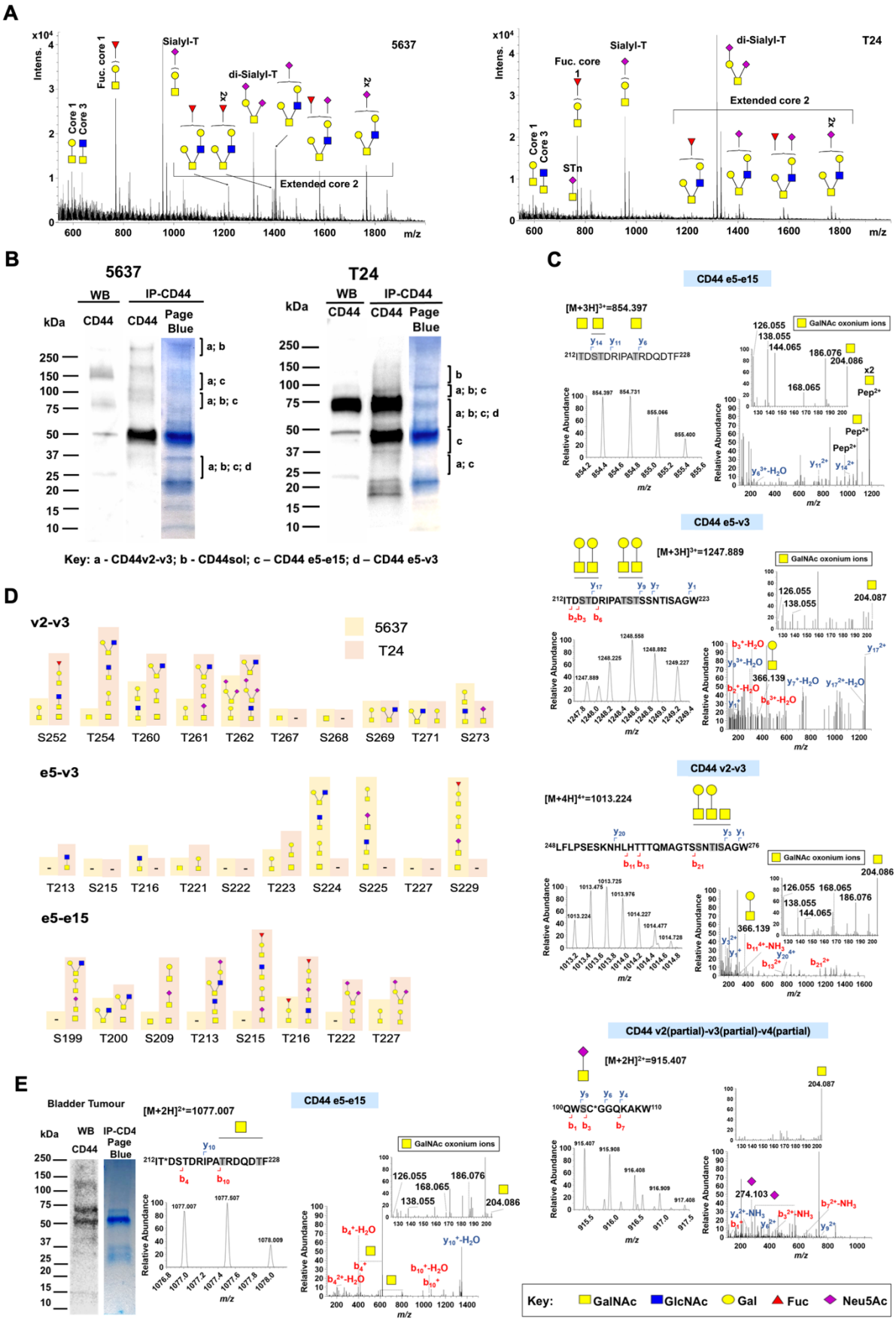

A
5637
T24
Sialyl-T
di-Sialyl-T
Fuc. core 1
Core 1
Core 3
Extended core 2
STn
Intens.
m/z
B
WB
IP-CD44
CD44
Page Blue
kDa
Key: a - CD44v2-v3; b - CD44sol; c – CD44 e5-e15; d – CD44 e5-v3
C
CD44 e5-e15
CD44 e5-v3
CD44 v2-v3
CD44 v2(partial)-v3(partial)-v4(partial)
GalNAc oxonium ions
D
v2-v3
e5-v3
e5-e15
S252 T254 T260 T261 T262 T267 S268 S269 T271 S273
T213 S215 T216 T221 S222 T223 S224 S225 T227 S229
S199 T200 S209 T213 S215 T216 T222 T227
E
Bladder Tumour
Key: GalNAc GlcNAc Gal Fuc Neu5Ac

Figure 4. Glycoproteogenomics, building on RNAseq-customized databases and glycomics for protein annotation, enables CD44 proteoforms identification in bladder cancer cell lines and tumors. A) Glycomics characterization of 5637 and T24 cells showed fucosylated and sialylated T antigens as main glycospecies. MALDI-MS analysis of permethylated benzyl-GalNAc glycans revealed $[M+Na]^+$ main ions for fucosylated and sialylated T antigens in both cell lines. However, 5637 predominantly expressed mono-sialylated T antigens whereas di-sialylated structures were more abundant in T24 cells. Moreover, 5637 cells presented higher abundance and diversity of glycoforms extended beyond core 2 with different degrees of sialylation and fucosylation. **B) SDS-PAGE gels and western blots for CD44 IPs highlighting the nature of the isoforms identified by nano-LC-M/MS in glycoproteogenomics settings.** Briefly, CD44 immunoprecipitated from membrane extracts was separated by gradient SDS-PAGE and bands were excised and analyzed by nanoLC-HCD/CID-MS/MS using RNAseq-customized databases and glycomics data for isoforms annotation. The nature of the isoforms translated by reporter ions for constitute-variable junctions identified in each band and cell line was highlighted. CD44 heterogeneity was evident for the two cell lines. CD44s was found to be the main isoform in T24 cells. **C) nano-LC-MS and HCD-MS/MS spectra for reporter ions v2-v3, e5-v3, e5-e15 (CD44s/st), and e2(partial)-e3(partial) CD44sol in T24 cells.** For reporter ions we show the MS isotopic envelope, the corresponding HCD product ion spectra highlighting GalNAc oxonium ions, GalNAc cross-ring fragments, other glycan fragments, and y- and b-series peptide backbone ions that support protein annotation. The predicted glycopeptide sequence, including the nature of the glycans and glycosites annotation (whenever possible) was also presented. **D) Glycosites identified for v2-v3, e5-v3 and e5-e15 reporter glycopeptides and corresponding glycans for 5637 and T24 cells.** A higher number of glycosites were identified at the e5-e15 junction for T24 in comparison to 5637 cells. On the other hand, more glycosites were identified at the v2-v3 and e5-v3 junctions in 5637 cells. **E) nano-LC-MS and HCD-MS/MS spectra for CD44 immunoprecipitated from CD44s^{high} tumors and areas of CD44-STn co-expression.** SDS-PAGE and western blots showed a pattern similar to T24 cells, characterized by bands bellow 75 kDa. An HCD product ion spectrum for an CD44s-Tn specific glycopeptide sequence is presented. Identified and possible glycosites are identified in grey. The symbol * corresponds to amino acid modifications: C – carbamidomethyl and T – 2-amino-3-ketobutyric acid.

CD44 is a driver of invasion

We then investigated the functional role played by CD44 in proliferation and invasion by silencing its expression in 5637 and T24 cells using siRNA targeting the constitutive region of the protein (80-90% CD44 decrease according to RT-PCR and WB; Figures 5A-B). This significantly reduced cell proliferation in both cell lines (Figure 5C) but impacted differently in terms of invasion, potentiating 5637 and decreasing the number of T24 invasive cells (Figure 5D). These are novel findings linking CD44 to cell proliferation in BlCa, supporting previous observations for breast [63] and colorectal [64] cancers, and suggesting that intense splicing of extracellular domains may play a key role in driving invasion, in agreement with observations from patient samples. Since the percentage of CD44s is significantly

higher in CD44-dependent invasive cells (75% in T24 *vs* 5% in 5637; Figure 3D and S4) and significantly associated with invasion (Figure 1A-B), we also concluded that this isoform maybe used to specifically target more aggressive cancer cells subpopulations.

To disclose how CD44 could mediate proliferation and invasion, we used a human Phospho-RTK array to detect possible alterations in the phosphorylation levels of 49 different proteins (Figure S8) from T24 cells after CD44 siRNA-mediated knockdown. Epidermal growth factor receptor (EGFR), a cancer-associated member of the ErbB family, was significantly less activated in CD44 silenced T24 cells when compared to controls, which is consistent with CD44/EGFR interaction and synergic oncogenic signaling towards proliferation and invasion [65, 66]. Similar behavior was observed for the highly homologous insulin (IR) and insulin-like growth factor 1 (IGF-1R) receptors [67] after CD44 silencing (Figure 5E). Interestingly, increased IGF-1R phosphorylation in CD44 expressing cells has been linked to concomitant PI3K/AKT/mTOR pathway activation, EMT molecular traits and stem cells maintenance in solid tumors [68]. These observations support that CD44-mediated IR and IGF-1R activation, including downstream AKT and MAPK signaling transduction governing proliferation and invasion [69, 70] may also occur in BlCa. Altogether, CD44s, the most prevalent variant in T24 cells, seems to promote the activity of relevant RTK, culminating in biological cues that drive cancer cell proliferation and invasion. However, the exact mechanism by which CD44 triggers RTK activation remains poorly understood and should be addressed in future studies.

We further investigated how CD44 silencing would influence the activation by phosphorylation of several phospho-kinases and related downstream targets, using a human Phospho-Kinase array (Figure S9). As shown in Figure 5F, the activating phosphorylation of the transcription factor CREB as well as Src family non-RTK Yes was significantly decreased in T24 cells upon CD44 knock-down. CREB has been previously reported as a downstream target of CD44, mediating tumor proliferation, progression, and invasion in several tumor models [71, 72]. CD44 knock-down also resulted in HSP27, MSK1/2, p38α, PDGFRβ, c-Jun, STAT5a/b, Yes and p53 (S15) phospho-kinase activation. In agreement with these observations, several studies have described CD44 as a negative regulator of PDGFRβ signaling by mediating the recruitment of a tyrosine phosphatase to PDGFR and by destabilizing PDGFR heteroreceptor complexes, ultimately modulating growth factor signaling and proliferation [73, 74]. Moreover, CD44 has been previously

described as a negative regulator of p53 mediated responses, allowing cells to resist programmed death and senescence as well as respond to proliferative signals [75, 76]. The regulatory effect of CD44 on HSP27, MSK1/2, STAT5a/b and Yes are being described for the first time and warrant deeper investigation. Notably, Yes is the most widely expressed member of the Src kinase family and has a known regulatory role in growth factor signaling, cell proliferation, and invasion in different cancer models [77, 78]. Our observations support that CD44 may act as a Yes regulator in BlCa, which also merits future validation. On the other hand, CD44 has been described as a positive regulator of p38α [79] and c-Jun [80] in other tumor models, which potentially conflicts with our current observations and suggests cell-dependent responses that may be governed by the nature of CD44 isoforms. Collectively, our observations emphasize that we are still far from fully understanding CD44 regulation of oncogenic pathways, which may now be comprehensively addressed in the context of its isoforms. Nevertheless, we provide evidence that CD44 acts as an oncogenic modulator of phospho-kinases involved in several signaling pathways that support proliferation and invasion, including ERK/MAPK, PI3K/AKT signaling pathways. The dominant nature of CD44s in relation to other isoforms in T24 cells together with observations in patient samples suggest that it may play a functional role in bladder cancer aggressiveness, including the modulation of the above-described molecular events. However, this remains still a hypothesis which warrants confirmation through specific silencing of this isoform in a close future.

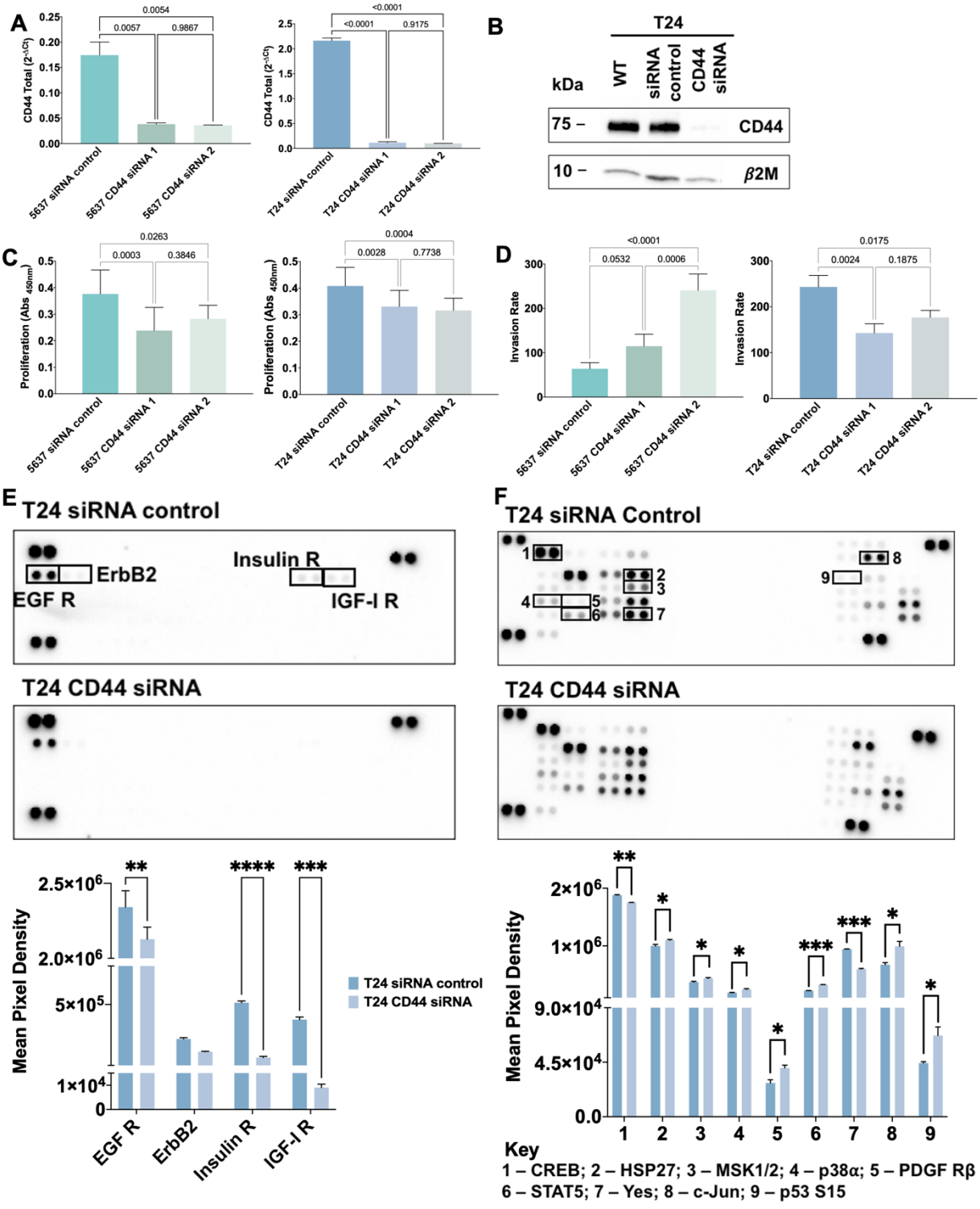


Figure 5. CD44 silencing inhibits proliferation, invasion, and activation of relevant proteins in MAPK/ERK, and PI3K/AKT pathways in T24 CD44s^{high} cells. A and B) CD44 siRNA promotes a massive decrease in *CD44* gene expression and complete abrogation of CD44 protein levels in 5637 and T24 cell lines. According to Figure A, *CD44* gene expression after siRNA is decreased by approximately 80% and 96% in 5637 and T24 cells, respectively, in comparison to controls, which translated into significantly decreased protein expression (exemplified in Figure B for T24 cells). **C) CD44 is a promoter of proliferation in 5637 and T24 cells.** Proliferation was reduced by more than

50% after CD44 siRNA silencing in both cell lines compared to controls. **D) CD44 silencing decreases the invasive capacity of CD44s^{high} T24 cells and promoted invasion in 5637 cells.** The number of T24 invading cells decreased approximately 1.7-fold after CD44 knock-down comparing to controls. On the other hand, the number of cells invading matrigel *in vitro* was 1.7-fold higher for 5637 cells under the same stimuli. **E) Analysis of the phosphorylation status of 49 RTK and F) Effect of CD44 silencing in phosphorylation of 37 proteins activated by different kinases.** CD44s seems to be a modulator and a co-receptor of different RTKs, including EGFR, Insulin, and IGF-1 receptors, as well as transcription factors such CREB, involved in tumor proliferation and invasion. The results correspond to mean and standard deviation for three independent experiments. Triplicate measurements were conducted for each experiment. *P* values are presented for Unpaired T tests.

CD44 glycosylation modulates proliferation and invasion

To assess the functional impact of altered CD44 glycosylation, we explored glycoengineered T24 cells overexpressing Tn (T24 *C1GALT1* KO) and STn antigens (T24 *C1GALT1* KO/*ST6GALNAC1* KI in mimicry of aggressive bladder tumors. Tn antigen expression was accomplished by complete abrogation of *O*-glycans extension by *C1GALT1* KO, as translated by a complete loss of sialylated T antigens (Figure 6A). Alongside the Tn antigen, these cells also expressed low amounts of STn, denoting low levels of sialylation in comparison to wild type cells. Then, *ST6GALNAC1* KI was introduced to promote Tn sialylation, generating a cell line expressing high STn levels and, to less extent, Tn antigens (Figure 6A). According to immunofluorescence, all models co-expressed CD44 together with altered glycosylation at the cell surface (Figure 6A). Furthermore, changes in glycosylation did not alter *CD44* gene expression or the nature of the variants, except for CD44s, which was found increased in cells overexpressing Tn but not for cells overexpressing STn (Figure 6B). Western blots supported little changes in CD44 levels in glycoengineered models, as suggested by gene expression. However, the blots for glycoengineered cell models showed a slight decrease in the main molecular band at 75 kDa, most likely from changes in glycans complexity (Figure 6C). This hypothesis was later confirmed by CD44 immunoprecipitation assays (Figure S10) and MS/MS analysis for T24 *C1GALT1* KO models (Figure 6D; Figure S11; Table S6). A less intense band above 37 kDa (close to CD44s estimated molecular weight) was also obvious in glycoengineered models (Figure 6C), suggesting that *O*-glycans shortening may also impact on glycosites occupancy.

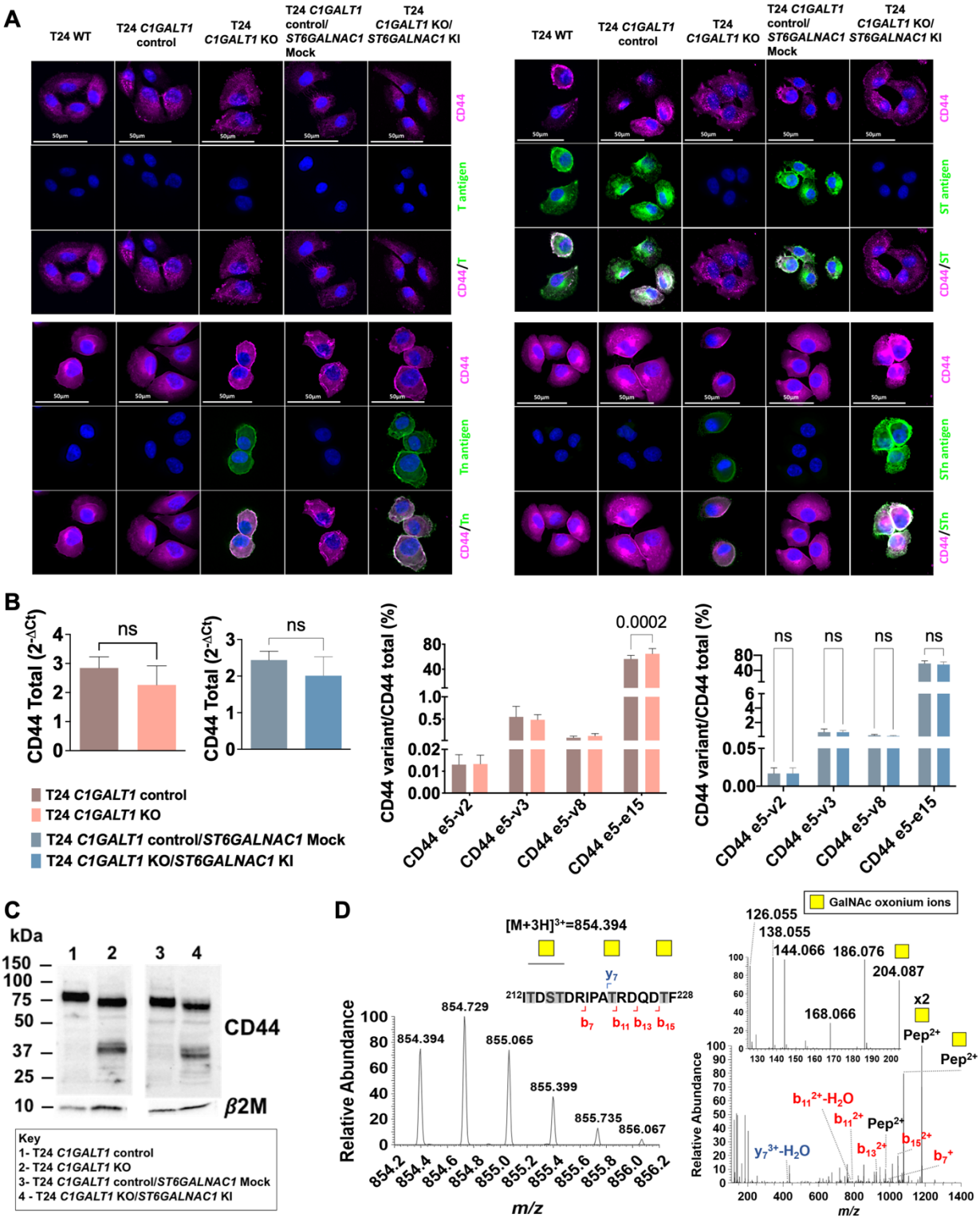


Figure 6. Glycoengineered T24 cells express CD44s-Tn/STn glycoproteoforms in mimicry of human tumors. A) Immunofluorescence studies on glycoengineered T24 bladder cancer cells demonstrated the overexpression of immature short-chain *O*-glycans together with CD44 at the cell surface of the cells. Immunoassays show high levels of sialylated T antigens and no Tn and STn antigens in wild type cells. T24 *C1GALT1* KO lacked extended glycosylation translated by the presence of sialylated T antigens and presented high Tn levels. *ST6GALNAC1* KI induced an overexpression of

STn. These glycans were co-localized with CD44 at the cells surface. **B) *CD44* gene expression and the nature of the variants do not change with alterations in glycosylation induced by *C1GALT1* KO and *ST6GALNAC1* KI, except for *CD44s*, which increased in Tn-expressing cells. C) Western blots show a different CD44 expression pattern for T24 glycoengineered cells in comparison to controls.** The main molecular band of CD44 at 75kDa shows a slight decrease, possibly due to glycans complexity reduction. A new band at 37kDa emerged, suggesting that glycans truncation may impact on CD44 molecular weight. **D) MS and HCD-MS/MS for a CD44s specific glycopeptide (corresponding to the e5-e15 junction) from T24 *C1GALT1* KO cells evidencing short-chain *O*-glycosylation.** The product ion spectrum shows a typical GalNAc fragmentation pattern, characterized by the presence of an evident ion at *m/z* 144.066 and a high 144.066/138.055 oxonium ions ratio, as well as the presence of the others HexNAc oxonium ions (*m/z* 126.055,168.066, 186.076 and 204.087). The spectrum also shows major fragment glycopeptides with GalNAc losses. The results correspond to the mean and standard deviation for three independent experiments. Triplicate measurements were conducted for each experiment. *P* values are presented for two-way ANOVA and Unpaired T tests.

To disclose the influence of CD44 glycosylation, we induced a transient inhibition of the protein by siRNA in glycoengineered cell models and controls, which led to almost complete abrogation of *CD44* mRNA expression (Figure 7C), as previously observed for wild type cells. First, we confirmed that proliferation was not affected by alterations in glycosylation induced by *C1GALT1* KO and *ST6GALNAC1* KI (Figure 7A). Also, Tn overexpression did not change invasion, whereas STn-overexpression increased by 3-fold the number of T24 cells invading Matrigel (Figure 7B). The link between STn and invasion is in close agreement with our previous report for glycoengineered BlCa cells [29], and other cancer models showing increased cell motility and acquisition of invasive traits resulting from the presence of this antigen [47, 49, 81]. Nevertheless, subsequent studies showed that T24 *C1GALT1* KO and *C1GALT1* KO*/ST6GALNAC1* KI cells proliferation was not significantly affected by CD44s silencing (Figure 7D). This contrasted with the anti-proliferative effect observed for T24 cells expressing more mature *O*-glycans under the same conditions (Figure 5C). Such findings suggest that proliferation may be fine-tuned by the nature of CD44s *O*-glycosylation, being activated by *O*-glycans extension. On the other hand, invasion induced by STn overexpression was reversed after CD44 inhibition (Figure 7E), as previously observed for wild type cells (Figure 5D). However, this effect was not so evident for Tn-expressing cells, suggesting that sialylation may play a role in CD44-mediated invasion, which warrants future confirmation. Collectively, these finds suggest that CD44-mediated invasion in T24 cell may be mostly driven by the nature of the splicing code.

We have then attempted to understand the functional impact of STn-overexpression on the activation of different RTKs upon siRNA-mediated knockdown of CD44. As shown in Figure 7F, CD44 silencing in STn positive cells induced a significant increase in activation of EGFR and Insulin receptors (IR) as well as of downstream proteins ERK 1/2, MSK 1/2 and RSK 1/2/3, all of which intermediates of the ERK/MAPK signaling pathway. Interestingly, wild type cells presented an opposite phenotype regarding activation of EGFR and insulin receptors under the same stimuli (Figure 5E). These findings again portrait the key role played by the glycocode in the regulation of oncogenic pathways governed by CD44. It further suggests that CD44-STn driven invasion may be supported by other molecular mechanisms. Additionally, we showed increased phosphorylation of Lck, a member of Src kinase family, which is intimately involved in T-cell activation in the absence of CD44 [82], suggesting a broader role of CD44 in immune regulation that should be addressed in future studies. Furthermore, we observed a significant downregulation of STAT3, STAT2 and HSP60 phosphorylation in STn-expressing cells upon CD44 inhibition (Figure 7G). STAT3 activation via CD44 has been linked to the acquisition of invasive traits and metastasis in different types of tumors [83, 84], suggesting a similar role in this context intimately linked to CD44. Altogether, these observations provide glances on the intricate nature of oncogenic pathways regulation by CD44, which should be carefully addressed by dedicated phosphoproteomics studies in the future. Moreover, it portrayed the key role played by glycosylation in CD44-mediated invasion of cancer cells. It suggests that these events may be modulated by sialylation of glycan chains, since this was not observed in Tn expressing cells. On the other hand, CD44 glycosylation appears to be key for defining the nature of oncogenic pathways supporting cancer aggressiveness.

In summary, we observed that invasion is mediated by the nature of CD44 splicing code and that cancer cells and tumors enriched for CD44s show higher invasion and worst prognosis. Furthermore, the splicing code presented by T24 cells was implicated in an onset of relevant oncogenic pathways that support these functional traits. Finally, we found that the nature of the CD44 glycocode is also crucial to enhance invasion and that CD44s-sTn/Tn glycoproteoforms hold tremendous potential to precisely target more aggressive subpopulation of cancer cells.

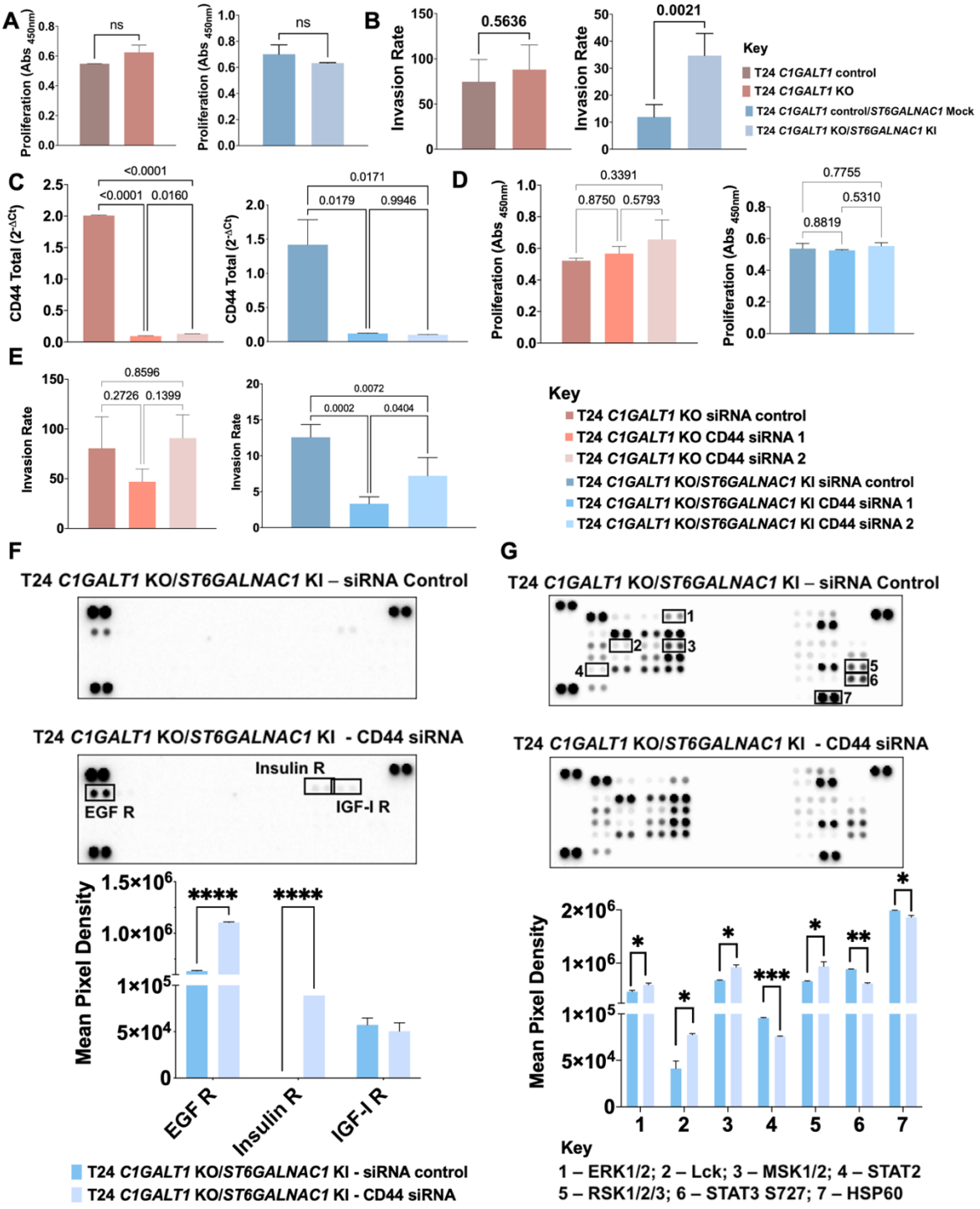


Figure 7. CD44 glycosylation in CD44s-enriched cells shapes its functional contribution to proliferation and invasion. A) *O*-glycans shortening by *C1GALT1* KO and *C1GALT1* KO/*ST6GALNAC*1 KI does not change the proliferation of T24 cells. T24 cells overexpressing Tn or STn antigens maintain proliferation rates compared to controls. **B) Increased STn antigen at the cell surface, driven by *C1GALT1* KO and *ST6GALNAC*1 KI, enhances the invasive capacity of T24 cells.** T24 cells overexpressing the Tn antigen do not change their capacity to invade Matrigel in comparison to the control *in vitro*. On the other hand, STn overexpression leads to a significant

increase in invasion. **C) CD44 silencing promotes a notable decrease in *CD44* gene expression in glycoengineered T24 cells expressing Tn and STn.** *CD44* mRNA expression after siRNA is significantly decreased by approximately 95% and 91% in *C1GALT1* KO and *C1GALT1* KO*/ST6GALNAC1* KI T24 cells, respectively, in comparison to controls. **D) CD44-Tn and STn glycoproteoforms do not play a role in T24 cells proliferation.** After CD44s inhibition, *C1GALT1* KO and *C1GALT1* KO*/ST6GALNAC1* KI T24 cells maintained their proliferation in relation to controls, showing that CD44-Tn/STn glycoproteoforms do not influence cell proliferation. **E) CD44-STn but not Tn glycoproteoforms drive invasion.** After CD44s silencing, *C1GALT1* KO T24 cells maintained their capacity to invade Matrigel *in vitro*, whereas *C1GALT1* KO*/ST6GALNAC1* KI cells showed decreased invasion. Similar effects were observed in wild-type cells but not in Tn-expressing cells, suggesting that sialylation may have a role in tumor invasion. **F) Influence of STn-overexpression on the response of tyrosine kinases receptors to CD44 silencing.** We show an activation of RTK proteins such EGFR and Insulin receptor in the absence of CD44 for *C1GALT1* KO*/ST6GALNAC1* KI cells, opposite to the effect observed in wild-type cells**. G) CD44-STn glycoproteoforms seem to promote the activation of relevant transcription factors involved in cell proliferation and invasion.** In the absence of CD44, a significant downregulation of STAT3 S727 and STAT2 phosphorylation was observed in *C1GALT1* KO*/ST6GALNAC1* KI cells. Results correspond to the mean and standard deviation for three independent experiments. Triplicate measurements were conducted for each experiment. *P* values are presented for Unpaired T tests.

Conclusions

CD44 is a pivotal glycoprotein in cancer, intricately related to most of its hallmarks, and widely accepted as a stem cell biomarker in BlCa [3, 16]. The biological role of CD44 in cancer is intimately linked to the nature of protein isoforms [85] that, in the absence of dedicated analytical workflows, have been inferred from transcripts analysis complemented with immunoassays, lacking the necessary specificity for isoform distinction. Together with conflicting nomenclature, this has constituted a barrier for definitive CD44 exploitation in clinical settings. In fact, CD44 mRNA alternative splicing, which generates highly homologous polypeptide sequences and dense and heterogeneous glycosylation, has constituted a major obstacle for precise identification of cancer-specific glycoproteoforms and the establishment of solid structure-function relationships [2]. Using BlCa as a model, we have started by exploring transcriptomics to provide the first glance at CD44 mosaicism in BlCa. This showed significant heterogeneity in terms of expressed isoforms and identified CD44s as a potential biomarker of invasion and poor prognosis. In agreement with these observations, CD44s was also strongly correlated with a subgroup of tumors displaying gene signatures characteristic of the basal BlCa molecular subtype but enriched for typical epithelial-to-mesenchymal transition markers. Later, we demonstrated that cells

significantly enriched for CD44s displayed CD44-mediated invasion *in vitro*, reinforcing the close link between this isoform and disease progression and its potential for targeting more aggressive subpopulations of cancer cells. The focus should now be on providing unequivocal demonstrations about the functional role of CD44s in these cells. Interestingly, basal bladder tumors express high levels of genes typical of more undifferentiated urothelial basal cells (*KRT5*, *KRT6*, and *KRT14)* together with several transcription factors that support stem cell homeostasis and cancer progression also presented a CD44s^{high} phenotype. The association of EMT markers with CSC properties has been reported for more aggressive tumors [86], which may likely be also the case for BlCa. In support of this hypothesis, in a recent study, Zhu *et al.* has elegantly demonstrated the pivotal role played by CD44s as regulator of stem-like properties, invasion, and lung metastasis in BlCa, providing solid molecular grounds to our current observations in patient samples [87] Similar reports have been found for breast [88], lung [89], hepatic [90], and pancreatic [11] cancers, which supports the pan-carcinogenic nature of CD44s and its role in regulating EMT and adaptive plasticity of CSC. Nevertheless, transcriptomics also showed that CD44s expression may not be exclusive of tumors, as many healthy human organs may also express high levels of this isoform. This apparent lack of cancer specificity raises a major obstacle for precise cancer targeting, casting doubts on therapeutic development strategies. However, we showed that targeting CD44s *O*-glycoforms, such as the Tn and STn antigens associated with aggressive forms of cancer, may allow to overcome this limitation, since these glycoforms are rarely found in healthy tissues.

Additionally, we tackled the limitations of transcriptomics and proteomics for definitive CD44 splicing code characterization, bringing together these different omics and glycomics. We have demonstrated that CD44 characterization cannot be achieved by a single omics and highlighted the importance of adding glycosylation for definitive isoforms identification, materializing the concept of glycoproteogenomics [28]. Overall glycoproteogenomics analysis of different cell models and bladder tumors supported transcriptomics data and provided a mass spectrometry-based approach for validation of several CD44 exon-exon junctions, moving one step forward in the characterization of this glycoprotein. It also confirmed CD44s as a major carrier of altered glycosylation linked to cancer aggressiveness, since we have demonstrated by mass spectrometry the presence of CD44s-(S)Tn in CD44s^{high}/STn^{high} tumors A major limitation of our glycoproteogenomics strategy was the incapacity to identify (glyco)peptides

spanning multiple exon-exon junctions, which would be key for more definitive isoform characterization. Future studies should complement this approach with middle-down proteomics adopting partial protein digestions to generate larger and potentially more informative peptide fragments.

Then, exploring the functional implications of CD44 in cancer cells, we demonstrated that this glycoprotein may play a pivotal role driving cell proliferation and invasion. We also hypothesize that shorter splicing codes characterized by high CD44s levels may be a key factor driving invasion, which still warrants confirmation. Phosphoarray-based assays further reinforced the link between CD44 and key phospho-tyrosine kinases and related transcription factors, mainly associated to EGFR dependent signaling, supporting proliferation and invasion and providing relevant molecular grounds to understand its role in BlCa progression. Similar observations have been made for other cancer models including gastric [91], breast [92], head and neck squamous cell carcinoma [65] and lung [66] tumors, reinforcing the influence of CD44 as a co-regulator of RTK signaling and EGFR mediated pathways, thereby contributing to cell proliferation and invasion. Our study does not unequivocally demonstrate that CD44s is directly implicated in these processes, however it shows that it may be explored to target cells presenting these molecular features. Moreover, it this isoform maybe useful for identifying patients benefiting from therapies targeting these relevant oncogenic pathways, paving the way towards precision oncology.

Finally, exploring a library of glycoengineered cell models we have demonstrated that the nature of CD44 glycosylation in cells significantly enriched for CD44s is key to fine-tune its functional behavior, namely in terms of proliferation and invasion. CD44 is a heavily *O*-glycosylated protein, and it is likely that dramatic changes in glycans structures may induce major conformational changes implicating interactions with extracellular matrix components and cell surface receptors, which are yet to be fully understood [93]. We found that CD44 inhibition in cells expressing high levels of sialylated *O*-glycans, including the STn antigen, had a major impact on cell invasion. Sialylation has long been described as a driver of cancer cells invasion, namely through the promotion of less cohesive cell phenotypes due to repulsive effects between proteins at the cell surface [94, 95]. In fact, some reports suggest that sialylation could decrease the interaction of CD44 with its ligand hyaluronic acid, whose interaction contributes to cancer invasiveness and metastasis [25]. By contrast, it has been previously demonstrated that CD44 *O*-glycans shortening increased hyaluronan binding capacity and may

promote migratory pro-invasive cancer cell phenotypes in gastric cancer [93]. Taken together, we propose that CD44 has a crucial role in cell invasion, however, it might be independent of the presence of STn moieties at its surface. On the other hand, a broader screening of main oncogenic pathways reflects another level of regulation dictated by the type of CD44 glycosylation, which appears to be key for defining oncogenic pathways adopted by cancer cells to support cancer aggressiveness. These findings, illustrate the relevance of CD44 glycocode characterization towards true precision oncology and casts important research topics for addressing CD44 role in cancer. It also supports the implementation of systems biology approaches, namely through an interrogation of the phosphoproteome for more comprehensive understanding about the role of this glycoprotein in cancer.

In summary, we have provided molecular and functional contexts linking CD44s and immature *O*-glycosylation, to cancer aggressiveness, in agreement with our previous reports linking both the protein and STn with invasion, metastasis and poor prognosis in BlCa [14, 34, 46]. Therefore, we hypothesize that targeting CD44-Tn/STn glycoproteoforms may constitute a key strategy to control BlCa progression, which will be addressed in future studies. Given that CD44s and STn are frequently found in many human tumors linked to unfavorable prognosis [46, 96], we also believe that this may constitute a valuable cancer signature for future clinical interventions at different levels. Moreover, according to the analysis of healthy tissues, CD44s glycosignatures are highly cancer specific, holding tremendous potential to overcome some of the off-target limitations associated with glycans. Therefore, this study may come to contribute significantly to the development of antibodies and other ligands against cancer cells, enabling the creation of an onset of novel theranostics interventions (cancer detection, targeted therapeutics) and decisively aid patient stratification. For instance, encouraging pre-clinical studies were presented concerning CAR-Ts targeting Tn and STn antigens in MUC1 [97], which may now be adapted for more cancer specific proteoforms such as CD44-Tn/STn. Moreover, it may pave the way for glycan-based cancer vaccines capable of educating the immune system to respond to abnormally glycosylated cancer cells, avoiding cancer spread and relapse through protective memory. Overall, we have provided a roadmap to characterize the functional role of glycosylation in CD44, paving the way for more educated clinical interventions. This analytical strategy sets the grounds for a comprehensive characterization of many different types of tumors, constituting a crucial step towards understanding

of the role of CD44 and other clinically relevant cell surface proteins in health and disease.

Abbreviations

B2M: β-2-microglobulin; BC: bladder cancer; BLCA: Bladder Carcinoma; BSA: bovine serum albumin; CD44: cluster of differentiation 44; CD44s: CD44 standard splicing isoform; CID: collision induced fragmentation; CSC: cancer stem cells; DAPI: 4',6'-diamidino-2-phenylindole dihydrochloride; DTT: 1,4-dithiothreitol; EGFR: epidermal growth factor receptor; EMT: epithelial-to-mesenchymal transition; FFPE: formalin-fixed paraffin-embedded; GTEx: genotype-tissue expression; HPRT: Hypoxanthine-guanine phosphoribosyltransferase; IGF-1R: insulin-like growth factor 1 receptor; IP: immunoprecipitation; IR: insulin receptor; KI: knock-in; KO: knock-out; MALT: mucosa-associated lymphoid tissue; MIBC: muscle invasive bladder cancer; nanoLC-MS/MS: nano liquid chromatography mass spectrometry; NCE: normalized collision energy; NMIBC: non-muscle invasive bladder cancer; PFA: paraformaldehyde; PLA: proximity ligation assay; RT-PCR: real time-polymerase chain reaction; RT: room temperature; RTK: receptor tyrosine kinase; SiRNA: small interfering RNA; STn: sialyl-Tn; VVA*: vicia villosa* lectin; WB: western blotting.

Acknowledgments

The authors wish to acknowledge the Portuguese Foundation for Science and Technology (FCT) for the human resources grants: PhD grants SFRH/BD/127327/2016 (CG), SFRH/BD/142479/2018 (JS), SFRH/BD/146500/2019 (MRS), SFRH/BD/111242/2015 (AP), 2020.08708.BD (RF), SFRH/BD/105355/2014 (RA), 2020.09384.BD (DF), and FCT assistant researcher grant CEECIND/03186/2017 (JAF). FCT is co-financed by European Social Fund (ESF) under Human Potential Operation Programme (POPH) from National Strategic Reference Framework (NSRF). The authors also acknowledge PhD Programs in Biomedical Sciences and Pathology and Molecular Genetics of ICBAS-University of Porto. The authors thank the financial support of the Portuguese Oncology Institute of Porto - Research Centre (CI-IPOP-29-2017-2020; CI-IPOP-58-2017-2021). This work was also financed by national funds through FCT/MCTES within the scope of the project UIDB/50006/2020 | UIDP/50006/2020, and by the Portuguese Mass Spectrometry Network, integrated in the National Roadmap of Research Infrastructures of Strategic Relevance (ROTEIRO/0028/2013; LISBOA-01-0145-FEDER-022125). The authors also wish to acknowledge the GTEx Project, supported

by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Authors Contributions

CG, AMNS, JAF designed research; CG, JS, MRS, AP, DF, AB, EF, RA, PP, CP, LL, RF, AM, HO, AMNS performed research; CG, JS, MRS, AP, DF, AB, RA, LL, AMNS, LLS, JAF analyzed data; PP, HO, AMNS, LLS, JAF contributed new reagents/analytic tools; CG and JAF wrote the paper; all authors revised the paper.

Competing interests

The other authors declare no conflicts of interest.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71: 209-49.

2. Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clin Proteomics. 2018; 15: 22.

3. Ponta H, Sherman L, Herrlich PA. CD44: from adhesion molecules to signalling regulators. Nat Rev Mol Cell Biol. 2003; 4: 33-45.

4. Teder P, Vandivier RW, Jiang D, Liang J, Cohn L, Pure E, et al. Resolution of lung inflammation by CD44. Science. 2002; 296: 155-8.

5. Nadella V, Wang Z, Johnson TS, Griffin M, Devitt A. Transglutaminase 2 interacts with syndecan-4 and CD44 at the surface of human macrophages to promote removal of apoptotic cells. Biochim Biophys Acta. 2015; 1853: 201-12.

6. Jalkanen S, Bargatze RF, de los Toyos J, Butcher EC. Lymphocyte recognition of high endothelium: antibodies to distinct epitopes of an 85-95-kD glycoprotein antigen differentially inhibit lymphocyte binding to lymph node, mucosal, or synovial endothelial cells. J Cell Biol. 1987; 105: 983-90.

7. Pals ST, Hogervorst F, Keizer GD, Thepen T, Horst E, Figdor CC. Identification of a widely distributed 90-kDa glycoprotein that is homologous to the Hermes-1 human lymphocyte homing receptor. J Immunol. 1989; 143: 851-7.

8. Cao H, Heazlewood SY, Williams B, Cardozo D, Nigro J, Oteiza A, et al. The role of CD44 in fetal and adult hematopoietic stem cell regulation. Haematologica. 2016; 101: 26-37.

9. Subramaniam V, Vincent IR, Gardner H, Chan E, Dhamko H, Jothy S. CD44 regulates cell migration in human colon cancer cells via Lyn kinase and AKT phosphorylation. Exp Mol Pathol. 2007; 83: 207-15.

10. Wolf KJ, Shukla P, Springer K, Lee S, Coombes JD, Choy CJ, et al. A mode of cell adhesion and migration facilitated by CD44-dependent microtentacles. Proc Natl Acad Sci U S A. 2020; 117: 11432-43.

11. Zhao S, Chen C, Chang K, Karnad A, Jagirdar J, Kumar AP, et al. CD44 Expression Level and Isoform Contributes to Pancreatic Cancer Cell Plasticity, Invasiveness, and Response to Therapy. Clin Cancer Res. 2016; 22: 5592-604.

12. 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.

13. Gunthert U, Hofmann M, Rudy W, Reber S, Zoller M, Haussmann I, et al. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell. 1991; 65: 13-24.

14. Azevedo R, Soares J, Gaiteiro C, Peixoto A, Lima L, Ferreira D, et al. Glycan affinity magnetic nanoplatforms for urinary glycobiomarkers discovery in bladder cancer. Talanta. 2018; 184: 347-55.

15. Wu CT, Lin WY, Chen WC, Chen MF. Predictive Value of CD44 in Muscle-Invasive Bladder Cancer and Its Relationship with IL-6 Signaling. Ann Surg Oncol. 2018; 25: 3518-26.

16. Hofner T, Macher-Goeppinger S, Klein C, Schillert A, Eisen C, Wagner S, et al. Expression and prognostic significance of cancer stem cell markers CD24 and CD44 in urothelial bladder cancer xenografts and patients undergoing radical cystectomy. Urol Oncol. 2014; 32: 678-86.

17. Yang YM, Chang JW. Bladder cancer initiating cells (BCICs) are among EMA-CD44v6+ subset: novel methods for isolating undetermined cancer stem (initiating) cells. Cancer Invest. 2008; 26: 725-33.

18. Naor D, Sionov RV, Ish-Shalom D. CD44: structure, function, and association with the malignant process. Adv Cancer Res. 1997; 71: 241-319.

19. Misra S, Hascall VC, Markwald RR, Ghatak S. Interactions between Hyaluronan and Its Receptors (CD44, RHAMM) Regulate the Activities of Inflammation and Cancer. Front Immunol. 2015; 6: 201.

20. Fnu G, Agrawal P, Kundu GC, Weber GF. Structural Constraint of Osteopontin Facilitates Efficient Binding to CD44. Biomolecules. 2021; 11.

21. Ishii S, Ford R, Thomas P, Nachman A, Steele G, Jr., Jessup JM. CD44 participates in the adhesion of human colorectal carcinoma cells to laminin and type IV collagen. Surg Oncol. 1993; 2: 255-64.

22. Hibino S, Shibuya M, Engbring JA, Mochizuki M, Nomizu M, Kleinman HK. Identification of an active site on the laminin alpha5 chain globular domain that binds to CD44 and inhibits malignancy. Cancer Res. 2004; 64: 4810-6.

23. Iczkowski KA, Omara-Opyene AL, Shah GV. The predominant CD44 splice variant in prostate cancer binds fibronectin, and calcitonin stimulates its expression. Anticancer Res. 2006; 26: 2863-72.

24. Caval T, Heck AJR, Reiding KR. Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation Between Sites on the Same Glycoprotein. Mol Cell Proteomics. 2020; 20: 100010.

25. Zhang X, Dou P, Akhtar ML, Liu F, Hu X, Yang L, et al. NEU4 inhibits motility of HCC cells by cleaving sialic acids on CD44. Oncogene. 2021.

26. Singh R, Subramanian S, Rhodes JM, Campbell BJ. Peanut lectin stimulates proliferation of colon cancer cells by interaction with glycosylated CD44v6 isoforms and consequential activation of c-Met and MAPK: functional implications for disease-associated glycosylation changes. Glycobiology. 2006; 16: 594-601.

27. Du T, Jia X, Dong X, Ru X, Li L, Wang Y, et al. Cosmc Disruption-Mediated Aberrant O-glycosylation Suppresses Breast Cancer Cell Growth via Impairment of CD44. Cancer Manag Res. 2020; 12: 511-22.

28. Ferreira JA, Relvas-Santos M, Peixoto A, Silva AMN, Santos LL. Glycoproteogenomics: Setting the Course for Next Generation Cancer Neoantigens Discovery for Cancer Vaccines. Genomics Proteomics Bioinformatics. 2021.

29. Peixoto A, Freitas R, Ferreira D, Relvas-Santos M, Paulo P, Cardoso M, et al. Metabolomics, Transcriptomics and Functional Glycomics Reveals Bladder Cancer Cells Plasticity and Enhanced Aggressiveness Facing Hypoxia and Glucose Deprivation. bioRxiv. 2021: 2021.02.14.431133.

30. Lima L, Gaiteiro C, Peixoto A, Soares J, Neves M, Santos LL, et al. Reference Genes for Addressing Gene Expression of Bladder Cancer Cell Models under Hypoxia: A Step Towards Transcriptomic Studies. PLoS One. 2016; 11: e0166120.

31. Peixoto A, Ferreira D, Azevedo R, Freitas R, Fernandes E, Relvas-Santos M, et al. Glycoproteomics identifies HOMER3 as a potentially targetable biomarker triggered by hypoxia and glucose deprivation in bladder cancer. J Exp Clin Cancer Res. 2021; 40: 191.

32. Fernandes E, Freitas R, Ferreira D, Soares J, Azevedo R, Gaiteiro C, et al. Nucleolin-Sle A Glycoforms as E-Selectin Ligands and Potentially Targetable Biomarkers at the Cell Surface of Gastric Cancer Cells. Cancers (Basel). 2020; 12.

33. Kudelka MR, Antonopoulos A, Wang Y, Duong DM, Song X, Seyfried NT, et al. Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. Nat Methods. 2016; 13: 81-6.

34. Cotton S, Azevedo R, Gaiteiro C, Ferreira D, Lima L, Peixoto A, et al. Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advanced-stage bladder tumors Mol Oncol. 2017; 11: 895-912.

35. Uchida Y, Sasaki H, Terasaki T. Establishment and validation of highly accurate formalin-fixed paraffin-embedded quantitative proteomics by heat-compatible pressure cycling technology using phase-transfer surfactant and SWATH-MS. Sci Rep. 2020; 10: 11271.

36. Zeng J, Eljalby M, Aryal RP, Lehoux S, Stavenhagen K, Kudelka MR, et al. Cosmc controls B cell homing. Nat Commun. 2020; 11: 3990.

37. Ferreira JA, Videira PA, Lima L, Pereira S, Silva M, Carrascal M, et al. Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumors Mol Oncol. 2013; 7: 719-31.

38. Carrascal MA, Severino PF, Guadalupe Cabral M, Silva M, Ferreira JA, Calais F, et al. Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. Mol Oncol. 2014; 8: 753-65.

39. Hao J, Chen H, Madigan MC, Cozzi PJ, Beretov J, Xiao W, et al. Co-expression of CD147 (EMMPRIN), CD44v3-10, MDR1 and monocarboxylate transporters is associated with prostate cancer drug resistance and progression. Br J Cancer. 2010; 103: 1008-18.

40. Dastych M, Hubatka F, Turanek-Knotigova P, Masek J, Kroupa R, Raska M, et al. Overexpression of CD44v8-10 in Colon Polyps-A Possible Key to Early Diagnosis. Pathol Oncol Res. 2021; 27: 614281.

41. Chen DD, Ji JA, Yan HC, Huang GH, Fang XJ. Effect of CD44st and HER2 expression on the postoperative prognosis of breast cancer patients. Onco Targets Ther. 2019; 12: 577-85.

42. Ozawa M, Ichikawa Y, Zheng YW, Oshima T, Miyata H, Nakazawa K, et al. Prognostic significance of CD44 variant 2 upregulation in colorectal cancer. Br J Cancer. 2014; 111: 365-74.

43. Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh LA, Yang J, 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: 1064-74.

44. Amin MB. Histological variants of urothelial carcinoma: diagnostic, therapeutic and prognostic implications. Mod Pathol. 2009; 22 Suppl 2: S96-S118.

45. Volkmer JP, Sahoo D, Chin RK, Ho PL, Tang C, Kurtova AV, et al. Three differentiation states risk-stratify bladder cancer into distinct subtypes. Proc Natl Acad Sci U S A. 2012; 109: 2078-83.

46. Lima L, Neves M, Oliveira MI, Dieguez L, Freitas R, Azevedo R, et al. Sialyl-Tn identifies muscle-invasive bladder cancer basal and luminal subtypes facing decreased survival, being expressed by circulating tumor cells and metastases. Urol Oncol. 2017; 35: 675 e1- e8.

47. Cotton S, Ferreira D, Soares J, Peixoto A, Relvas-Santos M, Azevedo R, et al. Target Score-A Proteomics Data Selection Tool Applied to Esophageal Cancer Identifies GLUT1-Sialyl Tn Glycoforms as Biomarkers of Cancer Aggressiveness. Int J Mol Sci. 2021; 22.

48. Rodrigues JG, Duarte HO, Gomes C, Balmana M, Martins AM, Hensbergen PJ, et al. Terminal alpha2,6-sialylation of epidermal growth factor receptor modulates antibody therapy response of colorectal cancer cells. Cell Oncol (Dordr). 2021.

49. Freitas D, Campos D, Gomes J, Pinto F, Macedo JA, Matos R, et al. O-glycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. EBioMedicine. 2019; 40: 349-62.

50. Peixoto A, Fernandes E, Gaiteiro C, Lima L, Azevedo R, Soares J, et al. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget. 2016; 7: 63138-57.

51. Costa C, Pereira S, Lima L, Peixoto A, Fernandes E, Neves D, et al. Abnormal Protein Glycosylation and Activated PI3K/Akt/mTOR Pathway: Role in Bladder Cancer Prognosis and Targeted Therapeutics. PLoS One. 2015; 10: e0141253.

52. Santos J, Fernandes E, Ferreira JA, Lima L, Tavares A, Peixoto A, et al. P53 and cancer-associated sialylated glycans are surrogate markers of cancerization of the bladder associated with Schistosoma haematobium infection. PLoS Negl Trop Dis. 2014; 8: e3329.

53. Pirro M, Mohammed Y, de Ru AH, Janssen GMC, Tjokrodirijo RTN, Madunic K, et al. Oxonium Ion Guided Analysis of Quantitative Proteomics Data Reveals Site-Specific O-Glycosylation of Anterior Gradient Protein 2 (AGR2). Int J Mol Sci. 2021; 22.

54. Halim A, Westerlind U, Pett C, Schorlemer M, Ruetschi U, Brinkmalm G, et al. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. J Proteome Res. 2014; 13: 6024-32.

55. Freitas R, Relvas-Santos M, Azevedo R, Soares J, Fernandes E, Teixeira B, et al. Single-pot enzymatic synthesis of cancer-associated MUC16 O-glycopeptide libraries and multivalent protein glycoconjugates: a step towards cancer glycovaccines. New Journal of Chemistry. 2021; 45: 9197-211.

56. Okamoto I, Tsuiki H, Kenyon LC, Godwin AK, Emlet DR, Holgado-Madruga M, et al. Proteolytic cleavage of the CD44 adhesion molecule in multiple human tumors. Am J Pathol. 2002; 160: 441-7.

57. Medrano-Gonzalez PA, Rivera-Ramirez O, Montano LF, Rendon-Huerta EP. Proteolytic Processing of CD44 and Its Implications in Cancer. Stem Cells Int. 2021; 2021: 6667735.

58. Nakamura H, Suenaga N, Taniwaki K, Matsuki H, Yonezawa K, Fujii M, et al. Constitutive and induced CD44 shedding by ADAM-like proteases and membrane-type 1 matrix metalloproteinase. Cancer Res. 2004; 64: 876-82.

59. Pall T, Pink A, Kasak L, Turkina M, Anderson W, Valkna A, et al. Soluble CD44 interacts with intermediate filament protein vimentin on endothelial cell surface. PLoS One. 2011; 6: e29305.

60. Fujimoto T, Kawashima H, Tanaka T, Hirose M, Toyama-Sorimachi N, Matsuzawa Y, et al. CD44 binds a chondroitin sulfate proteoglycan, aggrecan. Int Immunol. 2001; 13: 359-66.

61. Ahrens T, Sleeman JP, Schempp CM, Howells N, Hofmann M, Ponta H, et al. Soluble CD44 inhibits melanoma tumor growth by blocking cell surface CD44 binding to hyaluronic acid. Oncogene. 2001; 20: 3399-408.

62. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci U S A. 2009; 106: 1760-5.

63. Guo Q, Liu Y, He Y, Du Y, Zhang G, Yang C, et al. CD44 activation state regulated by the CD44v10 isoform determines breast cancer proliferation. Oncol Rep. 2021; 45.

64. Gao L, Liu H, Yin N, Zuo S, Jin G, Hu Y, et al. BNIPL2 expression is correlated with the prognosis and regulates the proliferation of colorectal cancer through CD44. Mol Med Rep. 2019; 20: 4073-80.

65. Perez A, Neskey DM, Wen J, Pereira L, Reategui EP, Goodwin WJ, et al. CD44 interacts with EGFR and promotes head and neck squamous cell carcinoma initiation and progression. Oral Oncol. 2013; 49: 306-13.

66. Yin J, Zhang H, Wu X, Zhang Y, Li J, Shen J, et al. CD44 inhibition attenuates EGFR signaling and enhances cisplatin sensitivity in human EGFR wildtype nonsmallcell lung cancer cells. Int J Mol Med. 2020; 45: 1783-92.

67. De Meyts P, Wallach B, Christoffersen CT, Urso B, Gronskov K, Latus LJ, et al. The insulin-like growth factor-I receptor. Structure, ligand-binding mechanism and signal transduction. Horm Res. 1994; 42: 152-69.

68. Chang WW, Lin RJ, Yu J, Chang WY, Fu CH, Lai A, et al. The expression and significance of insulin-like growth factor-1 receptor and its pathway on breast cancer stem/progenitors. Breast Cancer Res. 2013; 15: R39.

69. Zhou G, Yang J, Song P. Correlation of ERK/MAPK signaling pathway with proliferation and apoptosis of colon cancer cells. Oncol Lett. 2019; 17: 2266-70.

70. He Y, Ge Y, Jiang M, Zhou J, Luo D, Fan H, et al. MiR-592 Promotes Gastric Cancer Proliferation, Migration, and Invasion Through the PI3K/AKT and MAPK/ERK Signaling Pathways by Targeting Spry2. Cell Physiol Biochem. 2018; 47: 1465-81.

71. De Falco V, Tamburrino A, Ventre S, Castellone MD, Malek M, Manie SN, et al. CD44 proteolysis increases CREB phosphorylation and sustains proliferation of thyroid cancer cells. Cancer Res. 2012; 72: 1449-58.

72. Ouhtit A, Madani S, Gupta I, Shanmuganathan S, Abdraboh ME, Al-Riyami H, et al. TGF-beta2: A Novel Target of CD44-Promoted Breast Cancer Invasion. J Cancer. 2013; 4: 566-72.

73. Li L, Heldin CH, Heldin P. Inhibition of platelet-derived growth factor-BB-induced receptor activation and fibroblast migration by hyaluronan activation of CD44. J Biol Chem. 2006; 281: 26512-9.

74. Porsch H, Mehic M, Olofsson B, Heldin P, Heldin CH. Platelet-derived growth factor beta-receptor, transforming growth factor beta type I receptor, and CD44 protein modulate each other's signaling and stability. J Biol Chem. 2014; 289: 19747-57.

75. Godar S, Ince TA, Bell GW, Feldser D, Donaher JL, Bergh J, et al. Growth-inhibitory and tumor- suppressive functions of p53 depend on its repression of CD44 expression. Cell. 2008; 134: 62-73.

76. Dhar D, Antonucci L, Nakagawa H, Kim JY, Glitzner E, Caruso S, et al. Liver Cancer Initiation Requires p53 Inhibition by CD44-Enhanced Growth Factor Signaling. Cancer Cell. 2018; 33: 1061-77 e6.

77. Hamamura K, Tsuji M, Hotta H, Ohkawa Y, Takahashi M, Shibuya H, et al. Functional activation of Src family kinase yes protein is essential for the enhanced malignant properties of human melanoma cells expressing ganglioside GD3. J Biol Chem. 2011; 286: 18526-37.

78. Sirvent A, Benistant C, Pannequin J, Veracini L, Simon V, Bourgaux JF, et al. Src family tyrosine kinases-driven colon cancer cell invasion is induced by Csk membrane delocalization. Oncogene. 2010; 29: 1303-15.

79. Ishimoto T, Nagano O, Yae T, Tamada M, Motohara T, Oshima H, et al. CD44 variant regulates redox status in cancer cells by stabilizing the xCT subunit of system xc(-) and thereby promotes tumor growth. Cancer Cell. 2011; 19: 387-400.

80. Chen L, Bourguignon LY. Hyaluronan-CD44 interaction promotes c-Jun signaling and miRNA21 expression leading to Bcl-2 expression and chemoresistance in breast cancer cells. Mol Cancer. 2014; 13: 52.

81. Julien S, Lagadec C, Krzewinski-Recchi MA, Courtand G, Le Bourhis X, Delannoy P. Stable expression of sialyl-Tn antigen in T47-D cells induces a decrease of cell adhesion and an increase of cell migration. Breast Cancer Res Treat. 2005; 90: 77-84.

82. Lefebvre DC, Lai JC, Maeshima N, Ford JL, Wong AS, Cross JL, et al. CD44 interacts directly with Lck in a zinc-dependent manner. Mol Immunol. 2010; 47: 1882-9.

83. So JY, Smolarek AK, Salerno DM, Maehr H, Uskokovic M, Liu F, et al. Targeting CD44-STAT3 signaling by Gemini vitamin D analog leads to inhibition of invasion in basal-like breast cancer. PLoS One. 2013; 8: e54020.

84. Anand V, Khandelwal M, Appunni S, Gupta N, Seth A, Singh P, et al. CD44 splice variant (CD44v3) promotes progression of urothelial carcinoma of bladder through Akt/ERK/STAT3 pathways: novel therapeutic approach. J Cancer Res Clin Oncol. 2019; 145: 2649-61.

85. Jordan AR, Racine RR, Hennig MJ, Lokeshwar VB. The Role of CD44 in Disease Pathophysiology and Targeted Treatment. Front Immunol. 2015; 6: 182.

86. Tanabe S, Quader S, Cabral H, Ono R. Interplay of EMT and CSC in Cancer and the Potential Therapeutic Strategies. Front Pharmacol. 2020; 11: 904.

87. Zhu J, Huang G, Hua X, Li Y, Yan H, Che X, et al. CD44s is a crucial ATG7 downstream regulator for stem-like property, invasion, and lung metastasis of human bladder cancer (BC) cells. Oncogene. 2019; 38: 3301-15.

88. Afify A, Purnell P, Nguyen L. Role of CD44s and CD44v6 on human breast cancer cell adhesion, migration, and invasion. Exp Mol Pathol. 2009; 86: 95-100.

89. Ko YH, Won HS, Jeon EK, Hong SH, Roh SY, Hong YS, et al. Prognostic significance of CD44s expression in resected non-small cell lung cancer. BMC Cancer. 2011; 11: 340.

90. Asai R, Tsuchiya H, Amisaki M, Makimoto K, Takenaga A, Sakabe T, et al. CD44 standard isoform is involved in maintenance of cancer stem cells of a hepatocellular carcinoma cell line. Cancer Med. 2019; 8: 773-82.

91. Koyama S, Maruyama T, Adachi S. Expression of epidermal growth factor receptor and CD44 splicing variants sharing exons 6 and 9 on gastric and esophageal carcinomas: a two-color flow-cytometric analysis. J Cancer Res Clin Oncol. 1999; 125: 47-54.

92. Morath I, Jung C, Leveque R, Linfeng C, Toillon RA, Warth A, et al. Differential recruitment of CD44 isoforms by ErbB ligands reveals an involvement of CD44 in breast cancer. Oncogene. 2018; 37: 1472-84.

93. Mereiter S, Martins AM, Gomes C, Balmana M, Macedo JA, Polom K, et al. O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. FEBS Lett. 2019; 593: 1675-89.

94. Suzuki O, Abe M, Hashimoto Y. Sialylation and glycosylation modulate cell adhesion and invasion to extracellular matrix in human malignant lymphoma: Dependency on integrin and the Rho GTPase family. Int J Oncol. 2015; 47: 2091-9.

95. Pally D, Pramanik D, Hussain S, Verma S, Srinivas A, Kumar RV, et al. Heterogeneity in 2,6-Linked Sialic Acids Potentiates Invasion of Breast Cancer Epithelia. ACS Cent Sci. 2021; 7: 110-25.

96. Hou H, Ge C, Sun H, Li H, Li J, Tian H. Tunicamycin inhibits cell proliferation and migration in hepatocellular carcinoma through suppression of CD44s and the ERK1/2 pathway. Cancer Sci. 2018; 109: 1088-100.

97. Posey AD, Jr., Schwab RD, Boesteanu AC, Steentoft C, Mandel U, Engels B, et al. Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. Immunity. 2016; 44: 1444-54.

Supplementary figures and tables

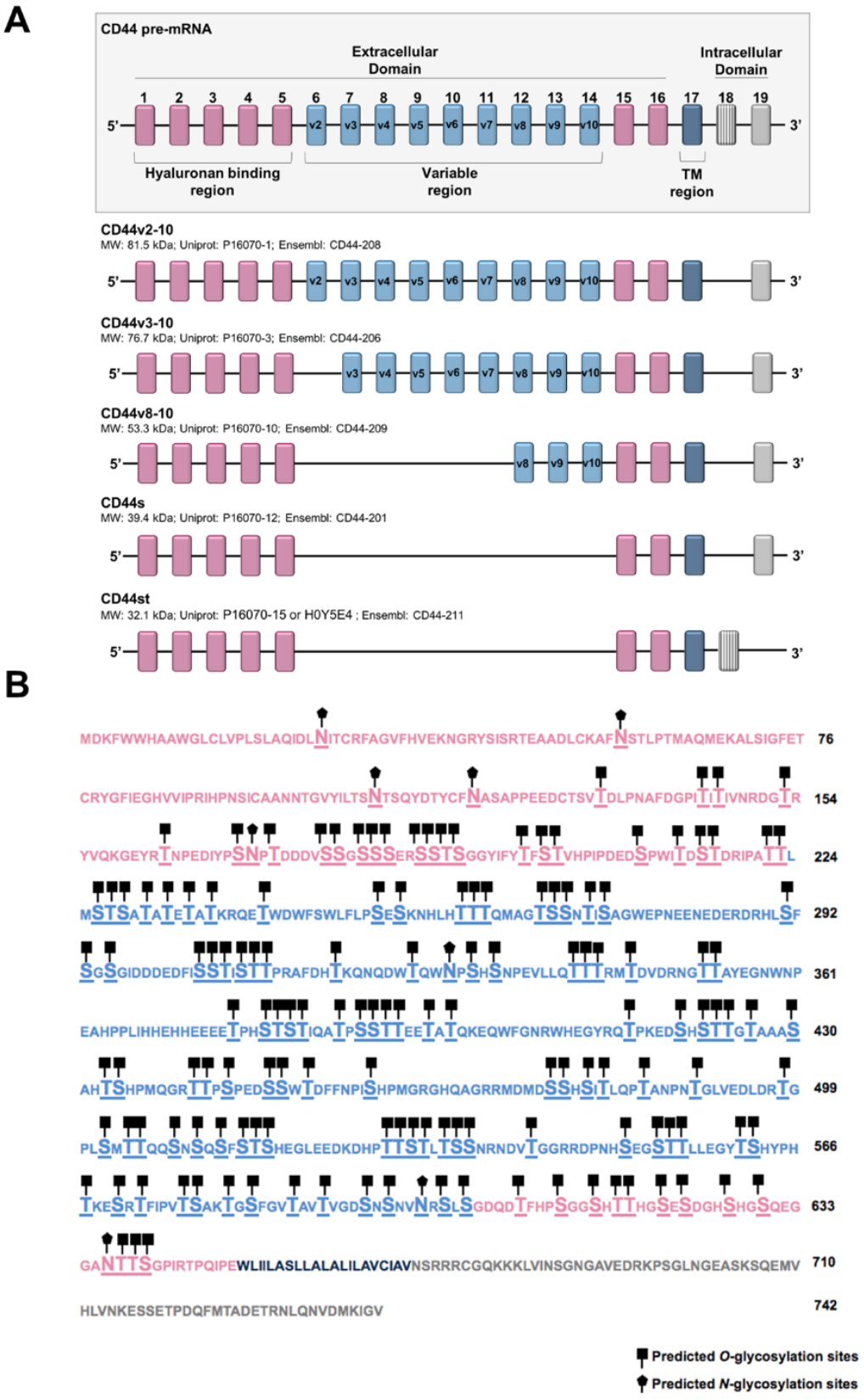


Figure S1. A) Schematic representation of human CD44 pre-mRNA and experimentally confirmed isoforms resulting from alternative splicing and extensively explored in this study. Exons generally regarded as conserved are represented in pink (constitutive exons e1-e5 and e15-e16), variable exons are represented in blue (exons v6-v14), the transmembrane region is represented in dark blue (exon e17), and the intracellular tail is represented in grey (exons e18-e19). Exon 18, filled with gray stripes, contains an early 3'UTR and is only present in the CD44st isoform. The estimated molecular weight for each variant (without posttranslational modifications), Uniprot and Ensemble codifications are also highlighted. **B) Amino acid sequence evidencing potential *O*- and *N*-glycosites.** *O*-glycosylation sites were predicted using the Net*O*Glyc 4.0 server (http://www.cbs.dtu.dk/services/NetOGlyc/) and *N*-glycosylation sites with the Net*N*Glyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). The different domains of the protein are highlighted using the same color code as in panel A and evidences the high density of *O*-glycosites in comparison to *N*-glycosites, which is particularly evident in the variable region.

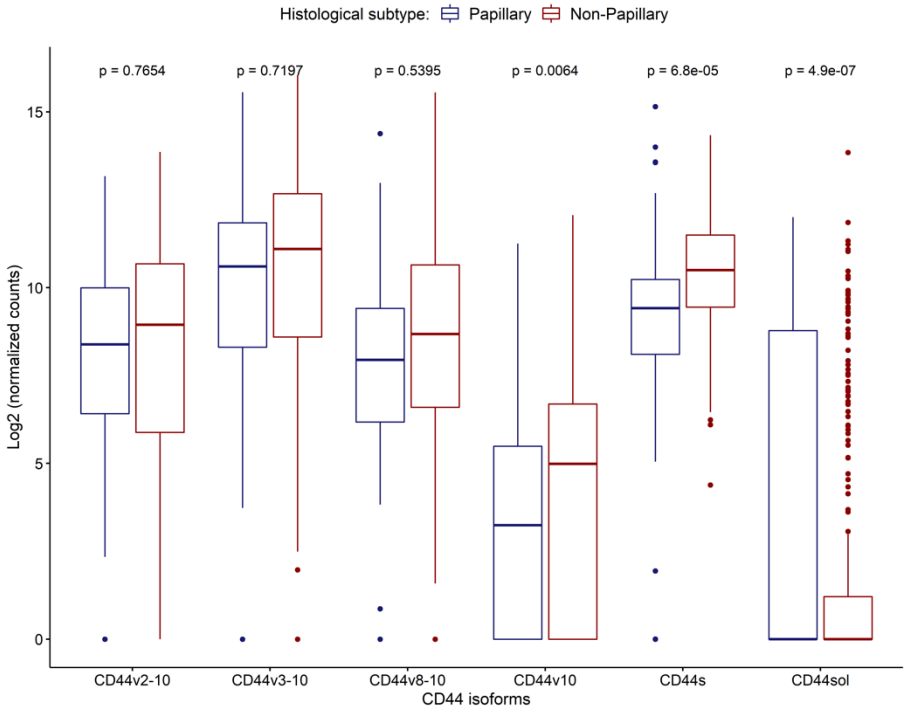
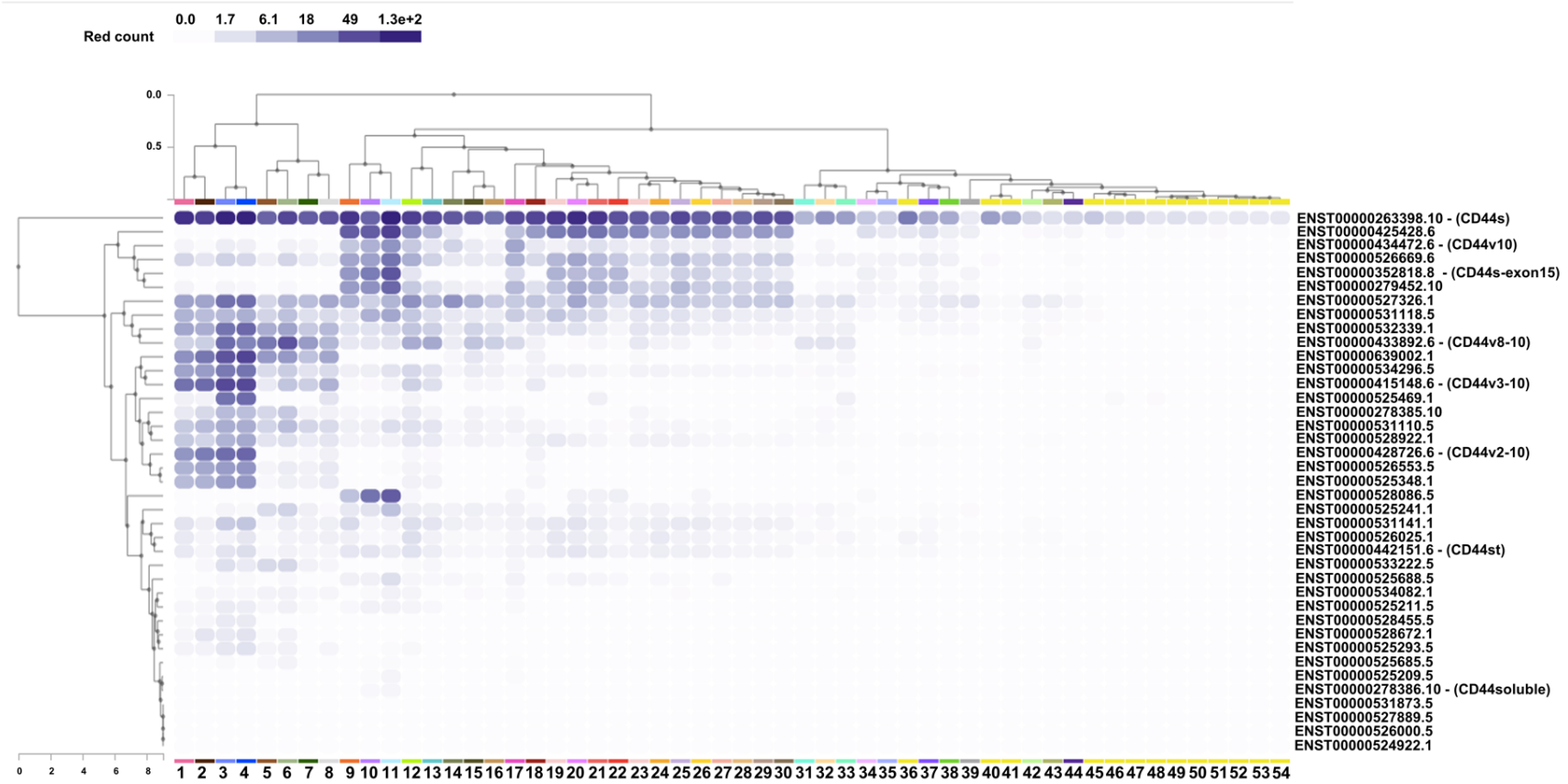


Figure S2. CD44 variants expression according to the histological subtypes of TCGA muscle invasive bladder tumors. Wilcoxon test revealed a significant overexpression of CD44s in more aggressive non-papillary lesions.



KEY

1- Vagina; 2- Esophagus-Mucosa; 3- Skin – Sun Exposed (Lower leg); 4- Skin – Not Sun Exposed (Suprapubic); 5- Pancreas; 6- Minor Salivary Gland; 7- Thyroid; 8- Prostate; 9- Adipose – subcutaneous; 10- Cells- EBV-transformed lymphocytes; 11- Cells- Cultured fibroblasts; 12- Lung: 13- Breast – Mammary Tissue; 14- Spleen; 15- Small Intestine – Terminal Ileum; 16- Colon – Transverse; 17- Whole Blood; 18- Bladder; 19- Cervix – Ectocervix; 20- Uterus; 21- Artery – Aorta; 22- Artery – Tibial; 23- Fallopian Tube; 24- Adipose – Visceral (Omentum); 25- Cervix – Endocervix; 26- Nerve – Tibial; 27- Artery – Coronary;28- Colon – Sigmoid; 29- Esophagus – Muscularis; 30- Esophagus – Gastroesophageal Junction; 31- Kidney – Cortex; 32- Stomach; 33- Kidney – Medulla; 34- Ovary; 35- Muscle – Skeletal; 36- Brain – Spinal cord (cervical c-1); 37- Heart – Atrial Appendage; 38- Adrenal Gland; 39- Testis; 40- Brain – Substantia nigra; 41- Brain – Hypothalamus; 42- Pituitary; 43- Liver; 44- Heart – Left Ventricle; 45- Brain – Hippocampus; 46- Brain – Cerebellum; 47- Brain – Amygdala; 48- Brain – Cortex; 49- Brain – Cerebellar Hemisphere; 50- Brain – Caudate (basal ganglia); 51- Brain – Frontal Cortex (BA9); 52- Brain – Nucleus accumbens (basal ganglia); 53- Brain – Putamen (basal ganglia); 54- Brain – Anterior cingulate cortex (BA24).

Figure S3. CD44 splice variants expression in healthy human tissues. Transcriptomics analysis of healthy tissues for CD44 isoforms were obtained and adapted from the GTEx Portal on 03/26/21 and/or dbGaP accession number phs000424.v8.p2. The Figure shows the complex mosaicism of CD44 mRNA in healthy tissues, including the co-expression of multiple isoforms in the same tissue. It highlights that CD44s is not expressed in healthy organs. Moreover, CD44st that differs from CD44s in terms of cytoplasmic tail but presents a similar extracellular domain was detected in low abundance in a restricted number of cells/organs, mostly from the skin, fibroblasts, subcutaneous adipocytes, the female reproductive system, the esophagus and breast. On the other hand, CD44s related structures lacking exon 15 (CD44s-exon 15) was present in high abundance in adipose tissues, fibroblasts, lungs, blood cells, across the female reproductive system and the gastrointestinal tract.

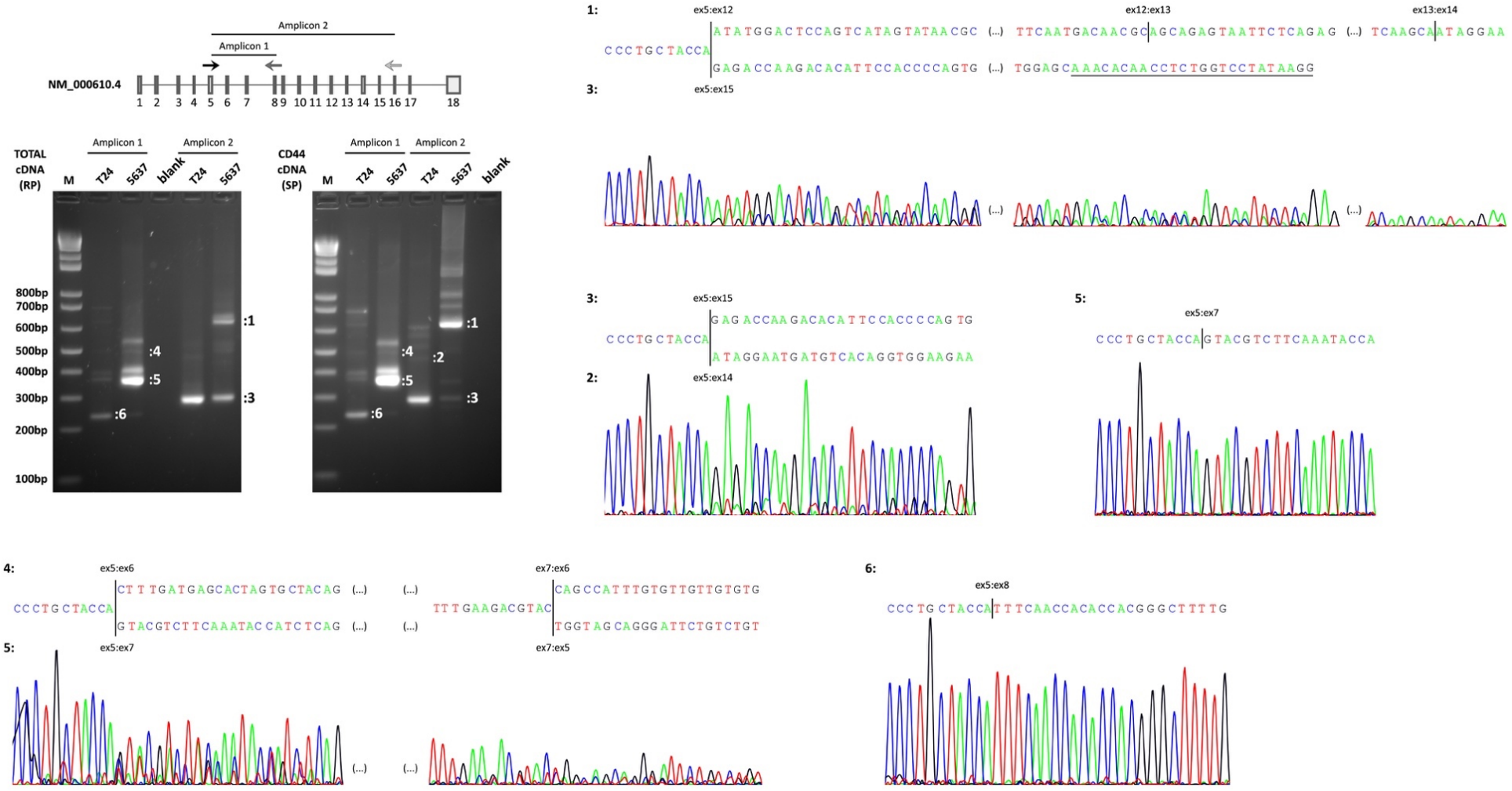

Amplicon 2
Amplicon 1
NM_000610.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
TOTAL cDNA (RP)
CD44 cDNA (SP)
Amplicon 1
Amplicon 2
M
T24
5637
blank
800bp
700bp
600bp
500bp
400bp
300bp
200bp
100bp
:1
:2
:3
:4
:5
:6
1:
ex5:ex12
ex12:ex13
ex13:ex14
ATATGGACTCCAGTCATAGTATAACGC (...) TTCAATGACAACGCAGCAGAGTAATTCTCAGAG (...) TCAAGCAATAGGAA
CCCTGCTACCA
GAGACCAAGACACATTCCACCCCAGTG (...) TGGAGCAAACACAACCTCTGGTCCTATAAGG
3:
ex5:ex15
2:
ex5:ex14
5:
ex5:ex7
CCCTGCTACCAGTACGTCTTCAAATACCA
4:
ex5:ex6
CTTTGATGAGCACTAGTGCTACAG (...)
GTACGTCTTCAAATACCATCTCAG (...)
ex7:ex6
CAGCCATTTGTGTTGTTGTGTG
TGGTAGCAGGGATTCTGTCTGT
ex7:ex5
6:
ex5:ex8
CCCTGCTACCATTTCAACCACACCACGGGCTTTTG

Figure S4. Reverse Transcriptase-Polymerase Chain Reaction and Sanger Sequencing for the *CD44* gene using random primer (RP) and e16 specific primer amplification. Two amplicons have been used: amplicon 1: e5-v4; amplicon 2: e5-e16. According to these experiments, 5637 were enriched longer isoforms in relation to T24 cells. However, we could not identify the precise nature of the long isoforms by sanger sequencing, most likely because of mixture of different isoforms. Supporting the distinct splicing codes presented by these cell lines, the following differences were observed: **e5-v8-v9-v10-(**…**)-e16** (**band 1**)**:** 5637 (main band)>T24; **e5-v10-(**…**)-e16** (**band 2**)**:** only in T24; **e5-e15-(**…**)-e16** (**band 3**)**:** T24 (main band; most likely corresponding to CD44s and/or st)>5637; **e5-v2-v3-(**…**)-e16** (**band 4**)**:** only in 5367; **e5-v3-v4-(**…**)-e16** (**band 5**)**:** 5637>T24; **e5-v4-(**…**)-e16** (**band 6**)**:** only in T24.

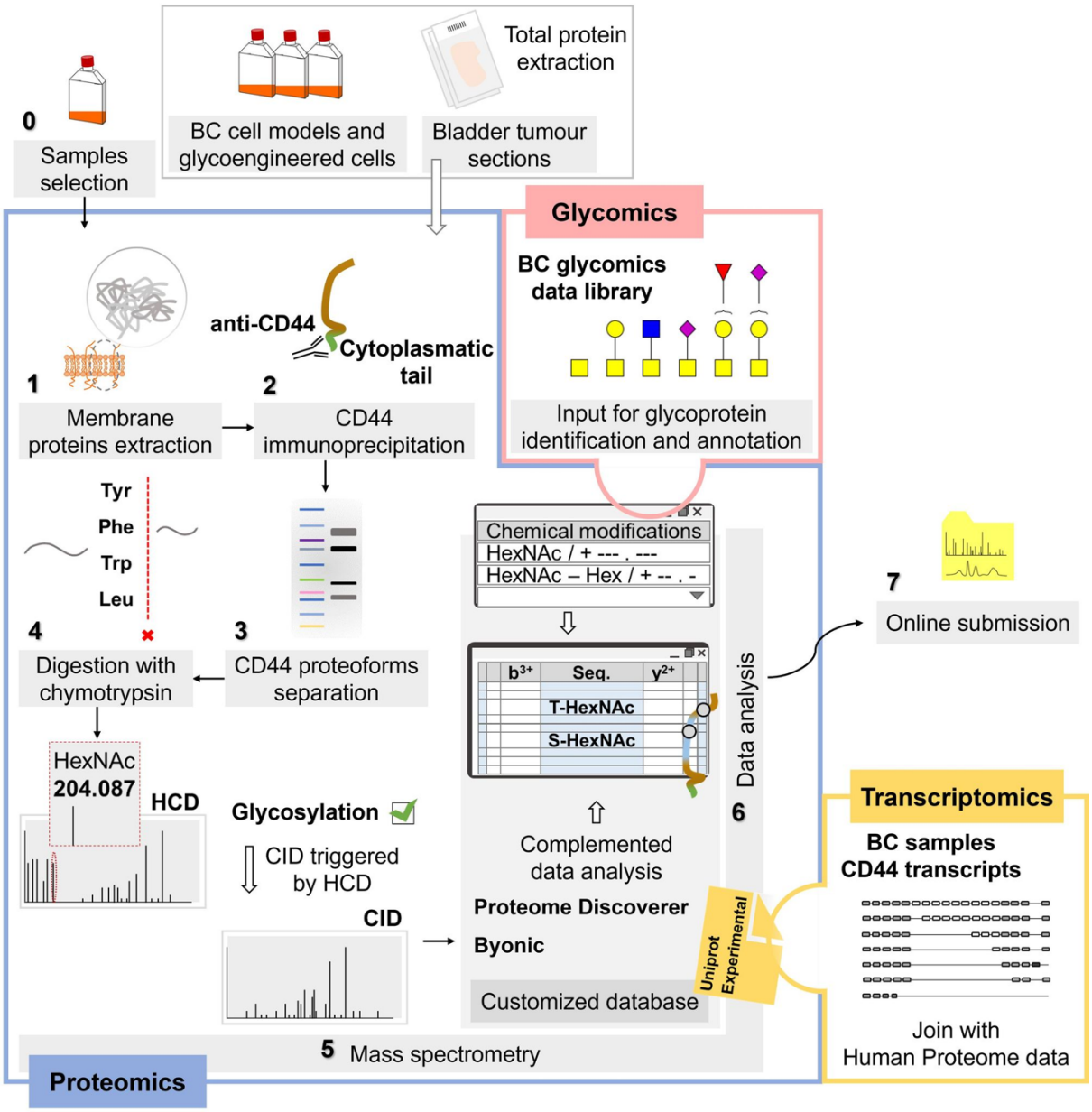


Figure S5. Glycoproteogenomics workflow for identification of CD44 glycoproteoforms. Briefly, bladder cancer cell models, glycoengineered cells and tumor tissue sections were used as starting material. **1.** Membrane glycoproteins from cell lines are isolated by differential ultracentrifugation. **2.** CD44 is immunoprecipitated using a monoclonal antibody targeting the cytoplasmatic tail domain expressed by most CD44 isoforms. **3.** CD44-IP enriched extracts were then digested with sialidase, separated by gradient SDS-PAGE electrophoresis. **4.** Bands are excised from the gels, proteins are reduced, alkylated, and digested with chymotrypsin *in* gel. **5.** Chymotrypsin digests are analyzed by nanoLC-MS/MS (HCD-MS2 and CID-MS2 with a dependent acquisition based on HCD detection of HexNAc oxonium ion at *m/z* 204.087. **6.** Glycoproteoforms identification using a glycoproteogenomics approach. RNAseq was used to identify CD44 isoforms expressed by 5637 and T24 cells. Predicted protein sequences were included in a database together with the human proteome from Uniprot database and used for protein identification in the Proteome Discoverer bioinformatics platform. Glycomics analysis was performed in parallel to determine the glycan structures that may

be present as variable CD44 post-translational modification. This information was included upon protein identification and glycopeptides annotation by MS/MS.

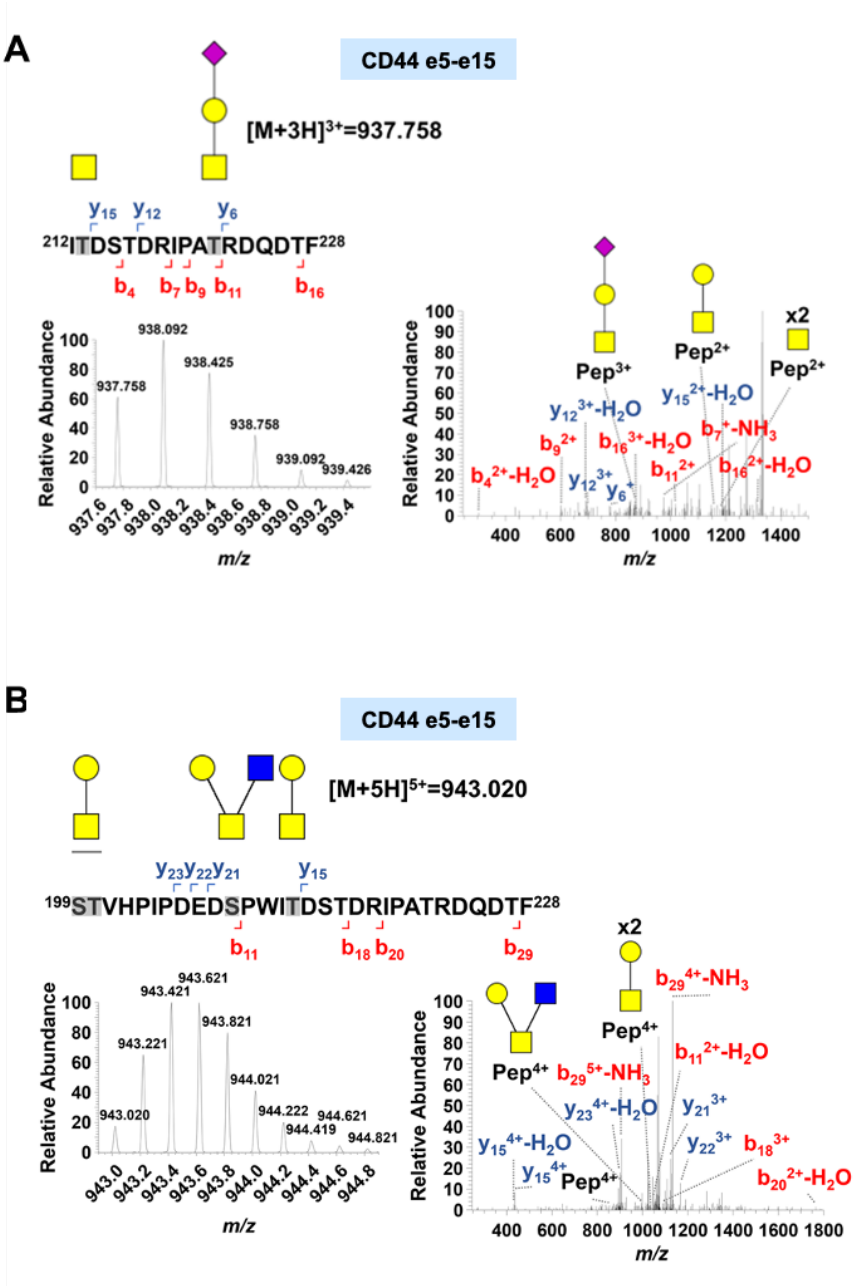


Figure S6. CID spectra for CD44s glycopeptides carrying *O*-glycans identified in T24 cells. A) CID spectrum for CD44s-Tn/ST and B) CD44s-T/core 2 glycopeptides. CD44s glycopeptides carrying these *O*-glycans were identified by CID based on an HCD data dependent acquisition. Briefly, species presenting HCD product ion spectrum with an HexNAc fragment at m/z 204.087 were elected for CID fragmentation. CID spectra confirmed the glycan structures and presented several *y*- and *b*-series peptide fragments that helped supporting glycosites assignment (highlighted in grey in the peptide sequence). Collectively, these spectra confirm the presence of the illustrated *O*-glycans in CD44s from T24 cells and the heterogeneous nature of glycosylation at the peptide level.

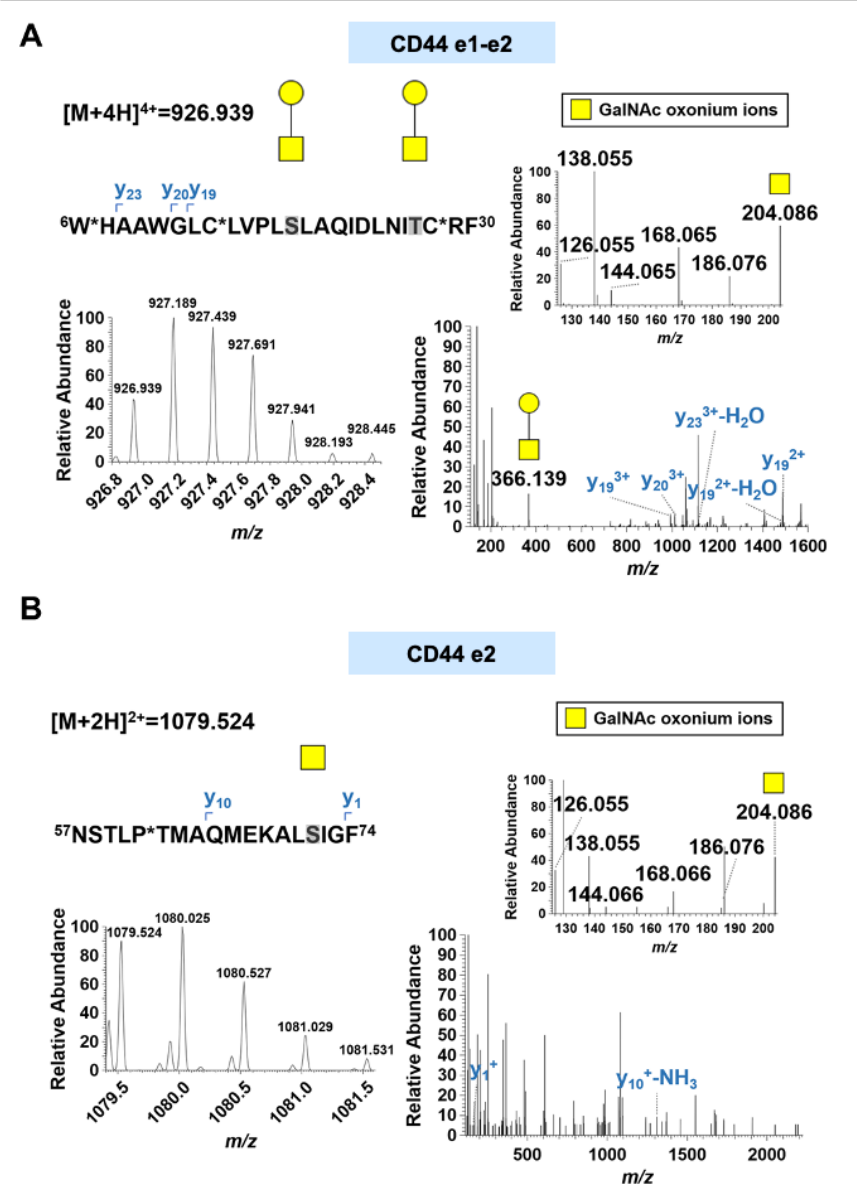
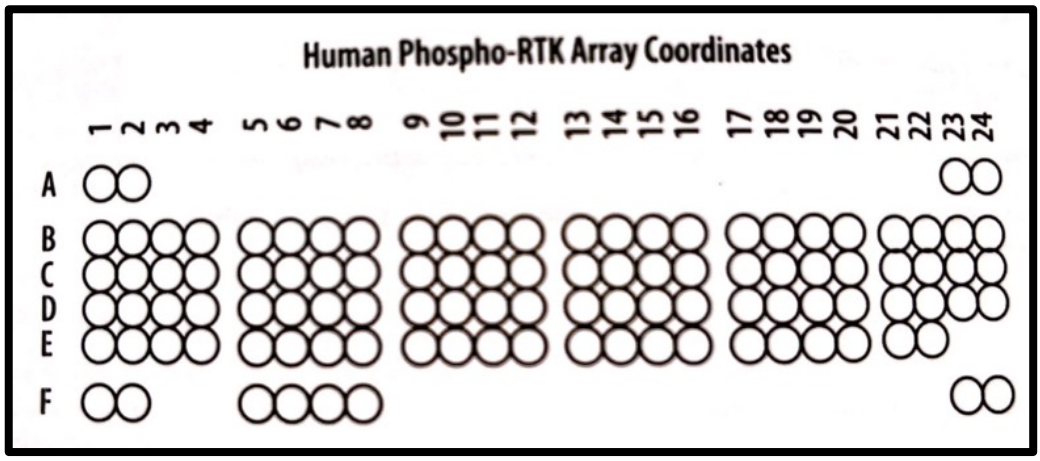


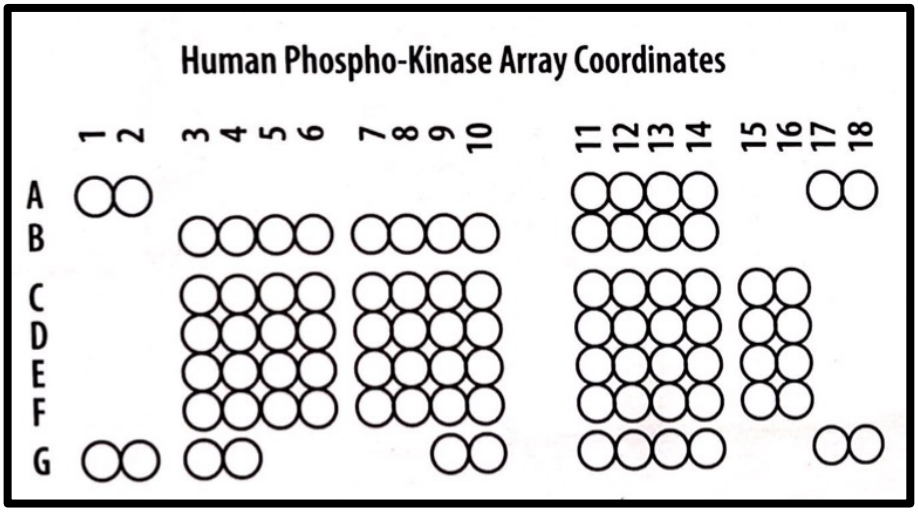
Figure S7. A) CD44-T and **B)** Tn glycopeptides belonging to the conserved extracellular domain in different isoforms identified in muscle invasive bladder tumors by nanoLC-HCD-MS/MS. Both MS/MS spectra present the HexNAc oxonium ions, namely the ion at m/z 204.086. The MS/MS spectrum of the CD44-T glycopeptide also presents an evident ion at m/z 366.139, corresponding to the oxonium ion for the T antigen. y-series ions resulting from peptide backbone fragmentation were also highlighted. The sequence of the glycopeptide presents the glycosites identified in grey. The symbol *corresponds to amino acid modifications: W – hydroxykynurenine; C – carbamidomethyl; P – glutamic semialdehyde.



A1, A2	Reference Spots	-
B1, B2	Reference Spots	-
B3, B4	EGF R	EGF R
B5, B6	EGF R	ErbB2
B7, B8	EGF R	ErbB4
B9, B10	FGF R	FGF R1
B11, B12	FGF R	FGF R2α
B13,B14	FGF R	FGF R3
B15, B16	FGF R	FGF R4
B17, B18	Insulin R	Insulin R
B19, B20	Insulin R	IGF-I R
B21, B22	Axl	Axl
B23, B24	Axl	Dtk
C1, C2	Axl	Mer
C3, C4	HGF R	HGF R
C5, C6	HGF R	MSP R
C7, C8	PDGF R	PDGF Rα
C9, C10	PDGF R	PDGF Rβ
C11, C12	PDGF R	SCF R
C13, C14	PDGF R	Flt-3
C15, C16	PDGF R	M-CSF R
C17, C18	RET	c-Ret
C19, C20	ROR	ROR1
C21, C22	ROR	ROR2
C23, C24	Tie	Tie-1
D1, D2	Tie	Tie-2
D3, D4	NGF R	TrkA
D5, D6	NGF R	TrkB
D7, D8	NGF R	TrkC
D9, D10	VEGF R	VEGF R1
D11, D12	VEGF R	VEGF R2
D13, D14	VEGF R	VEGF R3
D15, D16	MuSK	MuSK
D17, D18	Eph R	EphA1
D19, D20	Eph R	EphA2
D21, D22	Eph R	EphA3
D23, D24	Eph R	EphA4
E1, E2	Eph R	EphA6
E3, E4	Eph R	EphA7

E5, E6	Eph R	EphB1
E7, E8	Eph R	EphB2
E9, E10	Eph R	EphB4
E11, E12	Eph R	EphB6
E13, E14	Insulin R	ALK
E15, E16	-	DDR1
E17, E18	-	DDR2
E19, E20	Eph R	EphA5
E21, E22	Eph R	EphA10
F1, F2	Reference spots	-
F5, F6	Eph R	Ephb3
F7, F8	-	RYK
F23, F24	Control (-)	PBS

Figure S8. The human phospho-receptor tyrosine kinase (Phospho-RTK) array coordinates. The phospho-array presents capture and control antibodies spotted in duplicates for the 49 different phosphorylated human RTKs, which are represented by alpha-numeric coordinates.



A1, A2	Reference Spot	-
A11,A12	Akt 1/2/3	T308
A13, A14	Akt 1/2/3	S473
A17, A18	Reference Spot	-
B3, B4	CREB	S133
B5, B6	EGF R	Y1086
B7, B8	eNOS	S1177
B9, B10	ERK 1/2	T202/Y204, T185/Y187
B11, B12	Chk-2	T68
B13, B14	c-Jun	S63
C3, C4	Fgr	Y412
C5, C6	GSK-3α/β	S21/S9
C7, C8	GSK-3β	S9
C9, C10	HSP27	S78/S82
C11, C12	p53	S15
C13, C14	p53	S46
C15, C16	p53	S392
D3, D4	JNK 1/2/3	T183/Y185, T221/Y223
D5, D6	Lck	Y394
D7, D8	Lyn	Y397
D9, D10	MSK 1/2	S376/S360
D11, D12	p70 S6 Kinase	T389
D13, D14	p70 S6 Kinase	T421/S424
D15, D16	PRAS40	T246
E3, E4	p38α	T180/Y182
E5, E6	PDGF Rβ	Y751
E7, E8	PLC-γ1	Y783
E9, E10	Src	Y419
E11, E12	PYK2	Y402
E13, E14	RSK 1/2	S221/S227
E15, E16	RSK 1/2/3	S380/S386/S377
F3, F4	STAT2	Y689
F5, F6	STAT5a/b	Y694/Y699
F7, F8	WNK1	T60
F9, F10	Yes	Y426
F11, F12	STAT1	Y701
F13, F14	STAT3	Y705

F15, F16	STAT3	S727
G1, G2	Reference Spot	-
G3, G4	β-Catenin	-
G9, G10	PBS (Control -)	-
G11, G12	STAT6	Y641
G13, G14	HSP60	-
G17, G18	PBS (Control -)	-

Figure S9. The human phospho-kinase array coordinates. The phospho-array presents capture and control antibodies spotted in duplicates for the detection of the relative levels of phosphorylation of 37 kinase phosphorylation sites and 2 related total proteins represented by alpha-numeric coordinates.

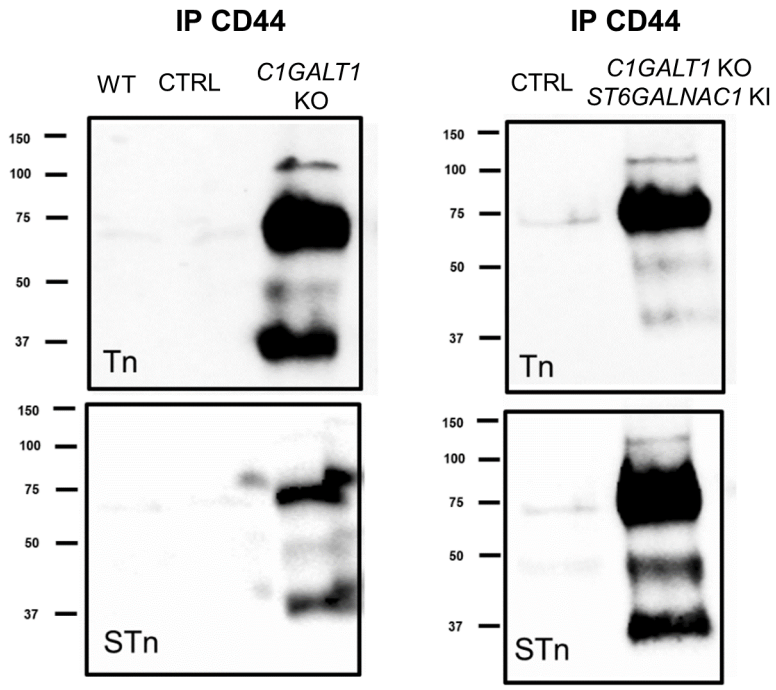


Figure S10. Western Blots show co-expression of CD44 and short-chain *O*-glycans, namely STn and Tn, in glycoengineered T24 cells. CD44 was immunoprecipitated from membrane protein extracts of T24 *C1GALT1* KO, T24 *C1GALT1* KO/*ST6GALNAC1* KI and corresponding controls and then blotted for STn, Tn, and CD44. A demarked expression of CD44-STn and CD44-Tn was detected for both cell model mainly in bands at 75 kDa and bellow.

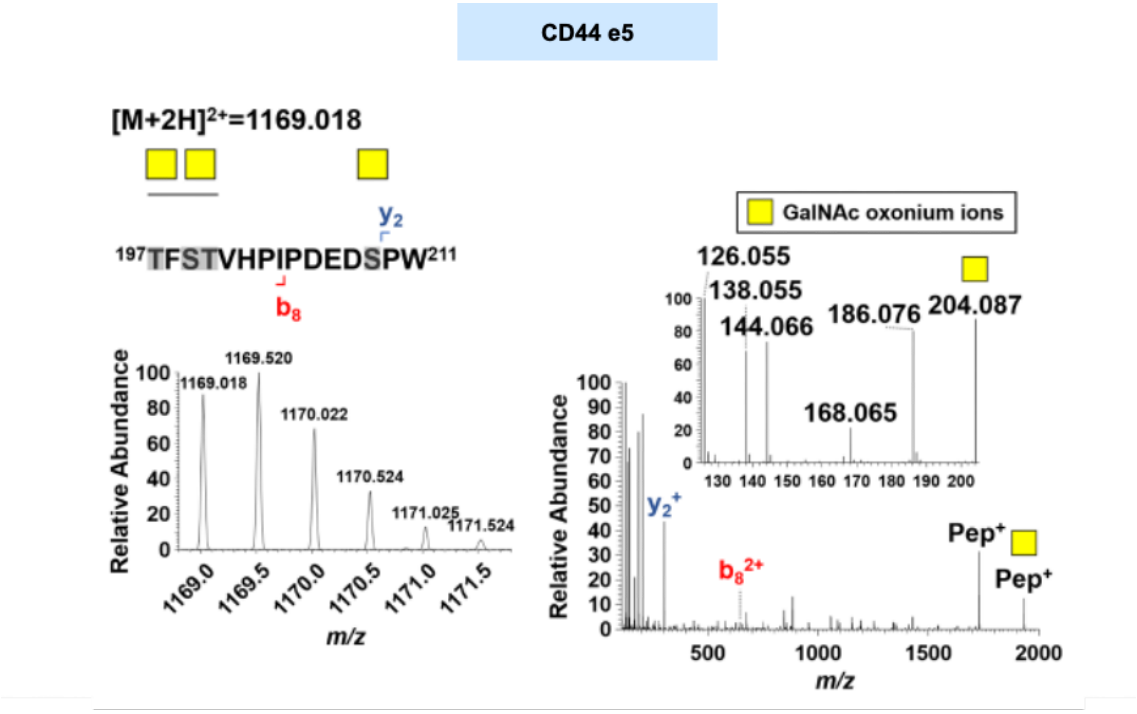


Figure S11. CD44-Tn glycopeptides belonging to the conserved extracellular domain identified in T24 *C1GALT1*KO cells by nanoLC-HCD-MS/MS. MS/MS spectrum highlights the typical HexNAC oxonium ion (*m/z* 204.087) as well as other oxonium ion fragments consistent with GalNAc (highlighted at *m/z* 138.055 and 144.066). The MS/MS spectrum of the CD44-Tn glycopeptide also highlights peptide ions with GalNAc losses. The predicted and possible glycosites are identified in grey.

Table S2. Tailored-made Taqman Gene Expression Assay ID's to specifically detect mRNA encoding for total CD44 and 5 splicing variants.

CD44 Total	Hs01075864_m1	e3-e4	NM_000610.3 (CD44 v2-10: ENST00000428726.8) NM_001001389.1 (CD44 v3-10: ENST00000415148.6) NM_001001390.1 (CD44 v8-10: ENST00000433892.6) NM_001001391.1(CD44s: ENST00000263398.11) NM_001202555.1 (CD44v10: ENST00000434472.6) NM_001202556.1(CD44s-exon15: ENST00000352818.8) NM_001202557.1 (CD44st: ENST00000442151.6)
CD44 e5-v2	Hs01075866_m1	e5-v2	NM_000610.3 (CD44 v2-10: ENST00000428726.8)
CD44 e5-v3	Hs01081480_m1	e5-v3	NM_001001389.1 (CD44 v3-10: ENST00000415148.6)
CD44 e5-v8	Hs01081475_m1	e5-v8	NM_001001390.1 (CD44 v8-10: ENST00000433892.6)
CD44 e5-e15	Hs01081473_m1	e5-e15	NM_001001391.1 (CD44s: ENST00000263398.11) NM_001202557.1 (CD44st: ENST00000442151.6)
CD44sol	Hs01081469_m1	v2(partial)-v3(partial)	NM_001001392.1 (CD44sol: ENST00000278386.10)

Table S2. List of antibodies and lectins used in this manuscript.

CD44	Rabbit polyclonal to CD44 antibody	Abcam	ab157107	Not provided	Synthetic peptide within Human CD44 a.a 650-750
STn	Mouse monoclonal to tag 72	Abcam	ab199002	B72.3+CC49	Membrane-enriched fraction of a human breast carcinoma liver metastasis [B72.3]; Purified TAG-72 protein [CC49]
Tn	Biotinylated and FITC-labeled VVA lectin	Vector Laboratories	B-1235-2 FL-1231-2	Not provided	α- or β-linked terminal N-acetylgalactosamine structures
T and ST after desialylation with α-neuraminidase	FITC-labeled PNA lectin	Vector Laboratories	FL-1071	Not provided	Gal (β-1,3) *GalNAc* structures
B2M	Rabbit monoclonal to beta 2 Microglobulin	Abcam	ab75853	EP2978Y	Not provided

Table S3. CD44 isoforms identified in 5637 and T24 cells by RNAseq.

ENST00000263398.11	P16070-12	**CD44s**
ENST00000526025.2	E9PKC6	**CD44splice E2**
ENST00000415148.6	P16070-4	**CD44v3-10**
ENST00000428726.8	P16070-1	**CD44v2-10**
ENST00000433892.6	P16070-10	**CD44v8-10**
ENST00000278386.10	P16070-19	**CD44sol**
ENST00000434472.6	P16070-11	**CD44v10**
ENST00000352818.8	P16070-18	**CD44s-exon15**
ENST00000442151.6	P16070-15 or H0Y5E4	**CD44st**
ENST00000526669.6	H0YD13	-
ENST00000425428.6	Q86UZ1	-
ENST00000528086.5	-	-
ENST00000526000.6	H0YDW7	-
ENST00000279452.10	H0Y2P0	-
ENST00000531118.5	-	-
ENST00000528455.5	H0YD17	-
ENST00000531873.5	H0YD90	-
ENST00000525209.5	-	-

Table S4. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for 5637 cells.

Peptide sequences	5637								
	Glycan chains								
	HexNAc	HexNAcNeuAc	HexHexNAc	HexHexNAcNeuAc	HexHexNAcNeuAc(2)	HexNAc(2)	HexNAc(2)Hex	HexNAc(2)Hex(2)NeuAc	HexNAc(1)Hex(1)Fuc
CD44v2-10									
LFLPSESKNHLHTTTQMAGTSsNtISAGW			x				x		
lPsEsKNHLHTTTQMAGTSSNTISAGW	x		x						
LFLPSESKNHLHtTTQMAGTsSnTlsAGW	x		x						
LFLPSESKNHLHtttQMAGTSSnTISAGW			x						
DWFsW									x
FsWLF									x
LFLPSESKNHLHTTtQMAGTSSNTISAGW					x				
dWFsW					x				
FsWLF						x			
LFLPSESKnHLHTTTQmAGtSSnTlSAGW	x								
fsWLF								x	
FsWLF							x		
LFLPSESKNHLHtTTQmAGtSSNTlsAGW	x					x			
DWFsW			x						
LFLPsEsKNHLHTTTQmAGTSSNTISAGW						x			
LPsEsKNHLHTTTQMAGTSSNTISAGW									
LPSESKNHLHtTTQMAGTSSNTlsAGW									
LFLPsESKNHLHTTTQmAGTSSNTISAGW									
LFLPSESKnHLHtttQMAGTSSNTISAGW									
LPSESKnHLHtttQMAGTSSNTISAGW									
LPSESKnHLHTttQMAGTSSNTISAGW									
LPSESKNHLHTtTQmAGTSSNTISAGW									
LFLPsESKNHLHTTtQMAGTSSNTISAGW									
LFLPSEsKNHLHTTTQmAGTsSNTISAGW									
LFLPsEsKNHLHTTTQmAGTSSNTlSAGW									
LPSESKNHLHTtTQmAGTSSNTlSAGW									
sWLF									
sWLF									
FsWLF									
LFLPsESKnHLHTtTQMAGTSSNTlSAGW									
STVHPIPDEDSPWITDStDRIPAtRDQDtF									
dWFsW		x							
sWLF									
FsWLF									
DWFsW									
FsWLF									
DWFsW									
CD44v3-10									
ltDSTDRIPATSTSSnTISAGW									
ltDSTDRIPAtStSSNTISAGW									
ITDSTDRIPATStSSNTISAGW									
ITDSTDRIPATSTSsNTIsAGW		x							x
ITDStDRIPATSTssNTlsAGW	x		x			x			
ITDStDRIPATSTsSNTlsAGW			x			x	x		
iTDSTDRIPATSTsSNTISAGW			x						
iTDSTDRIPATSTSsnTISAGW	x								
ITDSTDRIPAtStSSnTlsAGW	x		x						
ITDSTDRIPATSTssnTlsAGW	x					x			
ITDSTDRIPATSTssnTlsAGW									
ITDSTDRIPATSTssnTlsAGW									
ITDSTDRIPATSTssnTlsAGW									
CD44s									
StVHPIPDEDsPWITDSTDRIPATRDQDtF	x		x				x		
stVHPIPDEDsPWITDsTDRIPATRDQDTF	x		x						
ITDStDRIPATRDQDTF									x
sTVHPIPDEDSPWITDSTDRIPAtRDQDtF									
STVHPIPDEDSPWItDsTDRIPATRDQDTF									
StVHPIPDEDSPWItDSTDRIPAtRDQDTF									
ltDsTDRIPATRDQDTF									
sTVHPIPDEDsPWITDsTDRIPATRDQDTF									
lTDstDRIPAtRDQDtF									
ltDStDRIPAtRDQDTF									
STVHPIPDEDSPWItDStDRIPAtRDQDTF									
sTVHPIPDEDSPWItDStDRIPATRDQDTF									
ltDStDRIPAtRDQDTF									
lTDStDRIpAtRDQDtF									
STVhPIPDEDSpWITDsTDRIPATRDQDTf									
STVHPIPDEDsPWITDstDRIpATRDQDTF									
CD44sol									
qWScGGQKAKWtQRRGQQVSGnGAF						x			
GEQGVVRNsRPVY						x			
GEQGVVRNsRPVY	x								
GEQGVVRnsRPVYDs	x		x						
QWscGGQKAKW	x								
tQRRGQQVsGNGAF			x						
TQRRGQQVsGnGAF	x								
etcsLHcSQQSKKVWAEEKASDQQW						x	x		
scGGQKAKWTQRRGQQVsGnGAF	x								
ETcSLHcsQQsKKVW						x			
tQRRGQQVSGNGAFGEQGVVRNsRPVYDs						x	x		
TQRRGQQVsGNGAF		x							
gEQGVVRNsRPVY		x							
scGGQKAKWTQRRGQQVSGNGAF									x
TQRRGQQVsGNGAF								x	
EtcSLHcSQQSKKVW									
AEEKAsDQQWQW									
ETcsLHcSQQSKKVWAEEKAsDQQWQW									
qWScGGQKAKWTQRRGQQVsGNGAF									
QWScGGQKAKWTQRRGQQVsGNGAF									

Table S5. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for T24 cell line.

Peptide sequence	T24								
	Glycan chains								
	HexNAc	HexNAcNeuAc	HexHexNAc	HexHexNAcNeuAc	HexHexNAcNeuAc(2)	HexNAc(2)	HexNAc(2)Hex	HexNAc(2)Hex(2)NeuAc	HexNAc(1)Hex(1)Fuc
CD44v2-10									
LFLPSESKNHLHTTTQMAGTSsNtlSAGW			x				x		
lPsEsKNHLHTTTQMAGTSSNTlSAGW	x		x						
LFLPSESKNHLHtTTQMAGTsSnTlsAGW									
LFLPSESKNHLHtttQMAGTSSnTlSAGW									
DWFsW									
FsWLF									x
LFLPSESKNHLHTTtQMAGTSSNTlSAGW					x				
dWFsW									
FsWLF						x			
LFLPSESKnHLHTTTQmAGtSSnTlSAGW									
fsWLF								x	
FsWLF							x		
LFLPSESKNHLHtTTQmAGtSSNTlsAGW									
DWFsW									
LFLPsEsKNHLHTTTQmAGTSSNTlSAGW						x			
LPsEsKNHLHTTTQMAGTSSNTlSAGW									
LPSESKNHLHtTTQMAGTSSNTlsAGW					x				x
LFLPsESKNHLHTTTQmAGTSSNTlSAGW		x							
LFLPSESKnHLHtttQMAGTSSNTlSAGW						x	x		
LPSESKnHLHtttQMAGTSSNTlSAGW	x		x				x		
LPSESKnHLHTttQMAGTSSNTlSAGW							x		
LPSESKNHLHTtTQmAGTSSNTlSAGW									x
LFLPsESKNHLHTTtQMAGTSSNTlSAGW		x							x
LFLPSEsKNHLHTTTQmAGTsSNTlSAGW			x						
LFLPsEsKNHLHTTTQmAGTSSNTlSAGW	x								
LPSESKNHLHTtTQmAGTSSNTlSAGW		x							
sWLF									x
sWLF								x	
FsWLF				x					
LFLPsESKnHLHTtTQMAGTSSNTlSAGW									
STVHPIPDEDSPWITDStDRIPAtRDQDtF									
dWFsW									
sWLF									
FsWLF									
DWFsW									
FsWLF									
DWFsW									
CD44v3-10									
ltDSTDRIPATSTSSnTlSAGW						x			
ltDSTDRIPAtStSSNTlSAGW	x						x		
lTDSTDRIPATStSSNTlSAGW			x						
lTDSTDRIPATSTSsNTlsAGW									
lTDStDRIPATSTssNTlsAGW									
lTDStDRIPATSTsSNTlsAGW									
iTDSTDRIPATSTsSNTlSAGW									
iTDSTDRIPATSTSsnTlSAGW									
lTDSTDRlPAtStSSnTlsAGW									
lTDSTDRIPATSTssnTlsAGW									
lTDSTDRIPATSTssnTlsAGW									
lTDSTDRIPATSTssnTlsAGW									
lTDSTDRIPATSTssnTlsAGW									
CD44s									
StVHPIPDEDsPWITDSTDRIPATRDQDtF	x		x				x		
stVHPIPDEDsPWITDsTDRIPATRDQDTF	x		x						
lTDStDRIPATRDQDTF									x
sTVHPIPDEDSPWITDSTDRlPAtRDQDtF			x				x		
STVHPIPDEDSPWItDsTDRIPATRDQDTF		x						x	
StVHPIPDEDSPWItDSTDRIPAtRDQDTF	x		x				x		
ltDsTDRIPATRDQDTF						x			
sTVHPIPDEDsPWITDsTDRIPATRDQDTF		x							x
lTDstDRIPAtRDQDtF		x			x				x
ltDStDRIPAtRDQDTF									
STVHPIPDEDSPWItDStDRIPAtRDQDTF									
sTVHPIPDEDSPWItDStDRIPATRDQDTF	x					x			
ltDStDRIPAtRDQDTF									
lTDStDRlpAtRDQDtF									
STVhPIPDEDSpWITDsTDRIPATRDQDTf									
STVHPIPDEDsPWITDstDRlpATRDQDTF									
CD44sol									
qWScGGQKAKWtQRRGQQVSGnGAF									
GEQGVVRNsRPVY									
GEQGVVRNsRPVY	x								
GEQGVVRnsRPVYDs									
QWscGGQKAKW									
tQRRGQQVsGNGAF									
TQRRGQQVsGnGAF									
etcsLHcSQQSKKVWAEEKASDQQW						x	x		
scGGQKAKWTQRRGQQVsGnGAF	x								
ETcSLHcsQQsKKVW									
tQRRGQQVSGNGAFGEQGVVRNsRPVYDs									
TQRRGQQVsGNGAF		x							
gEQGVVRNsRPVY		x							
scGGQKAKWTQRRGQQVSGNGAF									x
TQRRGQQVsGNGAF								x	
EtcSLHcSQQSKKVW				x					
AEEKAsDQQWQW		x							
ETcsLHcSQQSKKVWAEEKAsDQQWQW			x			x			
qWScGGQKAKWTQRRGQQVsGNGAF									x
QWScGGQKAKWTQRRGQQVsGNGAF		x							

Table S6. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for CD44s^{high} MIBC showing areas of CD44 and STn co-localization.

Peptide sequence	MIBC	
	Glycan chains	
	HexNAc	HexNAcNeuAc
CD44v2-10		
LFLPSESKNHLHTTTQMAGTSsNtISAGW		
lPsEsKNHLHTTTQMAGTSSNTISAGW		
LFLPSESKNHLHtTTQMAGTsSnTlsAGW		
LFLPSESKNHLHtttQMAGTSSnTISAGW		
DWFsW		
FsWLF		
LFLPSESKNHLHTTtQMAGTSSNTISAGW		
dWFsW		
FsWLF		
LFLPSESKnHLHTTTQmAGtSSnTISAGW		
fsWLF		
FsWLF		
LFLPSESKNHLHtTTQmAGtSSNTlsAGW		
DWFsW		
LFLPsEsKNHLHTTTQmAGTSSNTISAGW		
LPsEsKNHLHTTTQMAGTSSNTISAGW		
LPSESKNHLHtTTQMAGTSSNTlsAGW		
LFLPsESKNHLHTTTQmAGTSSNTISAGW		
LFLPSESKnHLHtttQMAGTSSNTISAGW		
LPSESKnHLHtttQMAGTSSNTISAGW		
LPSESKnHLHTttQMAGTSSNTISAGW		
LPSESKNHLHTtTQmAGTSSNTISAGW		
LFLPsESKNHLHTTtQMAGTSSNTISAGW		
LFLPSEsKNHLHTTTQmAGTsSNTISAGW		
LFLPsEsKNHLHTTTQmAGTSSNTISAGW	x	
LPSESKNHLHTtTQmAGTSSNTISAGW	x	
sWLF		
sWLF		
FsWLF		
LFLPsESKnHLHTtTQMAGTSSNTISAGW	x	x
STVHPIPDEDSPWITDStDRIPAtRDQDtF		
dWFsW		
sWLF		x
FsWLF	x	
DWFsW	x	
FsWLF		x
DWFsW		x
CD44v3-10		
ItDSTDRIPATSTSSnTISAGW		
ItDSTDRIPAtStSSNTISAGW		
ITDSTDRIPATStSSNTISAGW		
ITDSTDRIPATSTSsNTlsAGW		
ITDStDRIPATSTssNTlsAGW		
ITDStDRIPATSTsSNTlsAGW		
iTDSTDRIPATSTsSNTISAGW		
iTDSTDRIPATSTSsnTISAGW		
ITDSTDRIPAtStSSnTlsAGW		
ITDSTDRIPATSTssnTlsAGW		
ITDSTDRIPATSTssnTlsAGW	x	x
ITDSTDRIPATSTssnTlsAGW		x
ITDSTDRIPATSTssnTlsAGW	x	
CD44s		
StVHPIPDEDsPWITDSTDRIPATRDQDtF		
stVHPIPDEDsPWITDsTDRIPATRDQDTF		
ITDStDRIPATRDQDTF		
sTVHPIPDEDSPWITDSTDRIPAtRDQDtF		
STVHPIPDEDSPWItDsTDRIPATRDQDTF		
StVHPIPDEDSPWItDSTDRIPAtRDQDTF		
ItDsTDRIPATRDQDTF		
sTVHPIPDEDsPWITDsTDRIPATRDQDTF		
ITDstDRIPAtRDQDtF		
ItDStDRIPAtRDQDTF	x	
STVHPIPDEDSPWItDStDRIPAtRDQDTF	x	
sTVHPIPDEDSPWItDStDRIPATRDQDTF		
ItDStDRIPAtRDQDTF		x
ITDStDRIpAtRDQDtF	x	x
STVhPIPDEDSpWITDsTDRIPATRDQDTf		x
STVHPIPDEDsPWITDstDRIpATRDQDTF	x	x
CD44sol		
qWScGGQKAKWtQRRGQQVSGnGAF		
GEQGVVRNsRPVY		
GEQGVVRNsRPVY	x	
GEQGVVRnsRPVYDs		
QWscGGQKAKW		
tQRRGQQVsGNGAF		
TQRRGQQVsGnGAF		
etcsLHcSQQSKKVWAEEKASDQQW		
scGGQKAKWTQRRGQQVsGnGAF	x	
ETcSLHcsQQsKKVW		
tQRRGQQVSGNGAFGEQGVVRNsRPVYDs		
TQRRGQQVsGNGAF		
gEQGVVRNsRPVY		x
scGGQKAKWTQRRGQQVSGNGAF		
TQRRGQQVsGNGAF		
EtcSLHcSQQSKKVW		
AEEKAsDQQWQW		
ETcsLHcSQQSKKVWAEEKAsDQQWQW		
qWScGGQKAKWTQRRGQQVsGNGAF		
QWScGGQKAKWTQRRGQQVsGNGAF		x

Table S7. Glycopeptides identified by nanoLC-HCD/CID-MS/MS in glycoproteogenomics settings for T24 *C1GALT1* KO cell model.

Peptide sequence	T24 *C1GALT1* KO		
	Glycan chains		
	HexNAc	HexNAcNeuAc	HexNAc(2)
CD44v2-10			
LFLPSESKNHLHTTTQMAGTSsNtISAGW			
lPsEsKNHLHTTTQMAGTSSNTISAGW			
LFLPSESKNHLHtTTQMAGTsSnTlsAGW			
LFLPSESKNHLHtttQMAGTSSnTlSAGW			
DWFsW			
FsWLF			
LFLPSESKNHLHTTtQMAGTSSNTISAGW			
dWFsW			
FsWLF			
LFLPSESKnHLHTTTQmAGtSSnTlSAGW			
fsWLF			
FsWLF			
LFLPSESKNHLHtTTQmAGtSSNTlsAGW			
DWFsW			
LFLPsEsKNHLHTTTQmAGTSSNTISAGW			
LPsEsKNHLHTTTQMAGTSSNTISAGW			
LPSESKNHLHtTTQMAGTSSNTlsAGW			
LFLPsESKNHLHTTTQmAGTSSNTISAGW			
LFLPSESKnHLHtttQMAGTSSNTISAGW			
LPSESKnHLHtttQMAGTSSNTlSAGW			
LPSESKnHLHTttQMAGTSSNTISAGW			
LPSESKNHLHTtTQmAGTSSNTISAGW			
LFLPsESKNHLHTTtQMAGTSSNTISAGW			
LFLPSEsKNHLHTTTQmAGTsSNTISAGW			
LFLPsEsKNHLHTTTQmAGTSSNTISAGW	x		
LPSESKNHLHTtTQmAGTSSNTISAGW			
sWLF			
sWLF			
FsWLF			
LFLPsESKnHLHTtTQMAGTSSNTISAGW	x	x	
STVHPIPDEDSPWITDStDRIPAtRDQDtF	x		
dWFsW			
sWLF			
FsWLF			
DWFsW			
FsWLF			
DWFsW			
CD44v3-10			
ltDSTDRIPATSTSSnTISAGW			
ltDSTDRIPAtStSSNTISAGW			
ITDSTDRIPATStSSNTISAGW			
ITDSTDRIPATSTSsNTIsAGW			
ITDStDRIPATSTssNTIsAGW			
ITDStDRIPATSTsSNTIsAGW			
iTDSTDRIPATSTsSNTISAGW			
iTDSTDRIPATSTSsnTISAGW			
ITDSTDRIPAtStSSnTIsAGW			
ITDSTDRIPATSTssnTlsAGW			
ITDSTDRIPATSTssnTlsAGW			
ITDSTDRIPATSTssnTlsAGW			
ITDSTDRIPATSTssnTlsAGW			
CD44s			
StVHPIPDEDsPWITDSTDRIPATRDQDtF			
stVHPIPDEDsPWITDsTDRIPATRDQDTF			
ITDStDRIPATRDQDTF			
sTVHPIPDEDSPWITDSTDRIPAtRDQDtF			
STVHPIPDEDSPWItDsTDRIPATRDQDTF			
StVHPIPDEDSPWItDSTDRIPAtRDQDTF			
ltDsTDRIPATRDQDTF			
sTVHPIPDEDsPWITDsTDRIPATRDQDTF			
ITDstDRIPAtRDQDtF			
ltDStDRIPAtRDQDTF	x		
STVHPIPDEDSPWItDStDRIPAtRDQDTF	x		
sTVHPIPDEDSPWItDStDRIPATRDQDTF			
ltDStDRIPAtRDQDTF			
ITDStDRlpAtRDQDtF			
STVhPIPDEDSpWITDsTDRIPATRDQDTf			
STVHPIPDEDsPWITDstDRIpATRDQDTF			
CD44sol			
qWScGGQKAKWtQRRGQQVSGnGAF			
GEQGVVRNsRPVY			
GEQGVVRNsRPVY			
GEQGVVRnsRPVYDs			
QWscGGQKAKW			
tQRRGQQVsGNGAF			
TQRRGQQVsGnGAF	x		
etcsLHcSQQSKKVWAEEKASDQQW			
scGGQKAKWTQRRGQQVsGnGAF	x		
ETcSLHcsQQsKKVW			
tQRRGQQVSGNGAFGEQGVVRNsRPVYDs			
TQRRGQQVsGNGAF		x	
gEQGVVRNsRPVY			
scGGQKAKWTQRRGQQVSGNGAF			
TQRRGQQVsGNGAF			
EtcSLHcSQQSKKVW			
AEEKAsDQQWQW			
ETcsLHcSQQSKKVWAEEKAsDQQWQW			
qWScGGQKAKWTQRRGQQVsGNGAF			
QWScGGQKAKWTQRRGQQVsGNGAF			

CONCLUDING REMARKS & FUTURE PERSPECTIVES

CONCLUDING REMARKS & FUTURE PERSPECTIVES

This work confronts the pressing issue of limited therapeutic options for advanced-stage BlCa patients refractory or resistant to mainstay therapies, which are currently faced with palliative care. While recent immune checkpoint-based immunotherapies have offered renewed hope for extended life expectancies, the response rate remains modest at 15% to 25% of patients, underscores the imperative for innovative approaches[1-5].

CD44 is a heavily glycosylated protein, being widely adopted as a biomarker of BlCa stem cells [6-8]. Moreover, CD44 is characterized by a complex landscape of proteoforms, provided by alternative splicing, exponentiating the complexity of workflows necessary to address this protein [8]. This molecular heterogeneity is further amplified by its highly glycosylated nature [8]. Despite its significance, few studies have explored the CD44 glycocode in BlCa or studied its functional implications in disease progression. Accordingly, this work has resorted to a multi-omics approach to unequivocally characterize cancer-associated CD44 glycoproteoforms for identification of signatures that hold clinical relevance for the development novel targeted therapies. Acknowledging the absence of well-established high-throughput methods for CD44 protein-level characterization, a novel glycoproteogenomics approach integrating transcriptomics, glycomics and glycoproteomics was developed.

Profiling a large BlCa patient cohort for CD44 splicing signatures, a significant CD44 isoforms array was revealed, with confirmed associations between short-CD44s, invasion and poor prognosis. Resorting to glycoengineered BlCa cell models, CD44s-Tn/STn glycoproteoforms emerged, providing the necessary cancer specificity to foster precision oncology solutions. This study not only materializes the concept of glycoproteogenomics but also lays the foundation to address the cancer splicing code, which could be expanded to further explore CD44's functional role in health and disease.

The molecular insights generated by this project has given rise to one international Patent (Ferreira, J. A., Silva, A. M. N., Freitas, R, Santos, L. L., Relvas-Santos., M, Gaiteiro, C., Peixoto, A., Miranda, A., Ferreira, D., Castro, F., Sarmento, B., Oliveira, M. J. (2023). CD44 glycoepitopes and chimeric vaccine glycoconjugates for cancer therapy and synthesis methods thereof. WO/2023/033664; Application number: PCT/PT2021/050025), anticipating future economic exploitation of

cancer-specific signatures for BlCa targeted therapeutics. In line with this, a CD44s-Tn/STn multimeric glycovaccine is being currently pursued, and gathered significant momentum to attract considerable funding for development and exploitation. In parallel, cell-based immunotherapies focusing on CD44s-Tn/STn CAR-T technologies will also be pursued. Given the pancarcinomic nature of CD44 and Tn/STn glycosignatures, the CD44 glycocode will also be addressed in other solid tumors marked by these molecular alterations, envisaging the transferability of developed immunotherapies across solid tumors

Overall, this work provides a comprehensive roadmap for characterizing the functional role of glycosylation and the clinical relevance of CD44 proteoforms, paving the way for more informed and tailored clinical interventions.

References

1. Balar AV, Castellano D, O'Donnell PH, Grivas P, Vuky J, Powles T, et al. First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. Lancet Oncol. 2017; 18: 1483-92.

2. Balar AV, Galsky MD, Rosenberg JE, Powles T, Petrylak DP, Bellmunt J, et al. Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. Lancet. 2017; 389: 67-76.

3. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016; 387: 1909-20.

4. Powles T, O'Donnell PH, Massard C, Arkenau HT, Friedlander TW, Hoimes CJ, et al. Efficacy and Safety of Durvalumab in Locally Advanced or Metastatic Urothelial Carcinoma: Updated Results From a Phase 1/2 Open-label Study. JAMA Oncol. 2017; 3: e172411.

5. Apolo AB, Ellerton JA, Infante JR, Agrawal M, Gordon MS, Aljumaily R, et al. Avelumab as second-line therapy for metastatic, platinum-treated urothelial carcinoma in the phase Ib JAVELIN Solid Tumor study: 2-year updated efficacy and safety analysis. J Immunother Cancer. 2020; 8.

6. Hofner T, Macher-Goeppinger S, Klein C, Schillert A, Eisen C, Wagner S, et al. Expression and prognostic significance of cancer stem cell markers CD24 and CD44 in urothelial bladder cancer xenografts and patients undergoing radical cystectomy. Urol Oncol. 2014; 32: 678-86.

7. Yang YM, Chang JW. Bladder cancer initiating cells (BCICs) are among EMA-CD44v6+ subset: novel methods for isolating undetermined cancer stem (initiating) cells. Cancer Invest. 2008; 26: 725-33.

8. Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clin Proteomics. 2018; 15: 22.