



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”



*Instituto de
Biociências Campus
do Litoral Paulista*

Davi de Mattos Ribeiro Lima

**“DETERMINAÇÃO DE UMA ASSINATURA MOLECULAR RENAL DE
REGENERAÇÃO APÓS TERAPÊUTICA COM CÉLULAS ESTROMAIS
MESENQUIMAIS MULTIPOTENTES”**

São Vicente SP
2025

Instituto de Biociência - Câmpus do Litoral Paulista
Praça Infante D. Henrique s/nº - CEP 11330-900 – São Vicente (SP)
Brasil Tel. (13) 3569-7100 - diretor.clp@unesp.br

Davi de Mattos Ribeiro Lima

**“DETERMINAÇÃO DE UMA ASSINATURA MOLECULAR RENAL DE
REGENERAÇÃO APÓS TERAPÊUTICA COM CÉLULAS ESTROMAIS
MESENQUIMAIS MULTIPOTENTES”**

Trabalho de conclusão de curso
apresentado à Universidade Estadual
Paulista (UNESP), Instituto de
Biociências, Câmpus do Litoral
Paulista, como parte dos requisitos para
obtenção do Grau de Bacharel em
Ciências Biológicas, com habilitação
em Biologia Marinha.

Orientador:

Prof. Dr. Roberto Fioravanti Carelli
Fontes

Co-orientadores:

Prof. Dr. Danilo Cândido de Almeida
Msc. Gabriel Pereira

L732d

Lima, Davi de Mattos Ribeiro

DETERMINAÇÃO DE UMA ASSINATURA MOLECULAR RENAL DE
REGENERAÇÃO APÓS TERAPÊUTICA COM CÉLULAS ESTROMAIS
MESENQUIMAIS MULTIPOTENTES / Davi de Mattos Ribeiro Lima. -- São Vicente,
2025

40 p. : il., fotos

Trabalho de conclusão de curso (Bacharelado - Ciências Biológicas) - Universidade
Estadual Paulista (UNESP), Instituto de Biociências, São Vicente

Orientador: Roberto Fioravanti Carelli Fontes

Coorientador: Danilo Cândido de Almeida

1. Células-tronco mesenquimais. 2. Bioinformática. 3. Regeneração (Biologia). I.
Título.

Agradecimentos

Gostaria de prestar os agradecimentos primeiro à minha mãe. Pois ela é a mecenas não só de minha carreira acadêmica, como o farol que guia minha vida.

Também a meus irmãos, Leonardo pois me ensinou a se dedicar ao próprio sonho independente de circunstâncias, e Artur por me ensinar a seguir o caminho que lhe satisfaz todos os dias.

Meus avós, pois são uma eterna fonte cultural e acadêmica.

A todos professores que já tive na vida, de aulas reais e fora de sala de aula, por vezes em um tatame com o usual odor de álcool e suor que me remete à disciplina.

Destes, especialmente, agradeço ao Professor Roberto “Bob” Fontes, por incentivar mesmo as ideias desorganizadas de um rapaz que queria unir programação e biologia e pela paciência nesta empreitada.

Também ao professor Danilo da UNIFESP de São Paulo, por me iniciar na bioinformática e por ser um mentor não apenas acadêmico.

E também à toda a equipe da UNIFESP que fez eu sentir que meu futuro está seguro e certo.

Epígrafes

“Om bhaiśajye bhaiśajye bhaiśajya-samudgate svāhā”

(Em sânscrito: Om, à medicina, à medicina, à suprema medicina, saudações!) Mantra do
Buddha Bhaişajyaguru, o Buda da Medicina.

Lista de Abreviações:

AKI: Acute Kidney Injury (Injúria Renal Aguda)

CKD: Chronic Kidney Disease (Doença Renal Crônica)

CTMs: Células-Tronco Estromais Mesenquimais Multipotentes

DEGs: Differentially Expressed Genes (Genes Diferencialmente Expressos)

FC: Fold Change (Mudança na Expressão)

FDR: False Discovery Rate (Taxa de Descoberta Falsa)

GFR: Glomerular Filtration Rate (Taxa de Filtração Glomerular)

GO: Gene Ontology (Ontologia Gênica)

GSEA: Gene Set Enrichment Analysis (Análise de Enriquecimento de Conjuntos Gênicos)

IL-6: Interleucina-6

KEGG: Kyoto Encyclopedia of Genes and Genomes

MAPK: Mitogen-Activated Protein Kinase (Proteína Quinase Ativada por Mitógeno)

MCP-1/CCL2: *Monocyte Chemoattractant Protein-1*

MSCs: Mesenchymal Stromal Cells (Células Estromais Mesenquimais - equivalente em inglês para CTMs)

NF- κ B: Nuclear Factor kappa B (Fator Nuclear-kappa B)

PCA: Principal Component Analysis (Análise de Componentes Principais)

PPI: Protein-Protein Interaction (Interação Proteína-Proteína)

RNA-Seq: RNA Sequencing (Sequenciamento de RNA)

ROS: Reactive Oxygen Species (Espécies Reativas de Oxigênio)

TNF- α : Tumor Necrosis Factor alpha (Fator de Necrose Tumoral alfa)

VEGF: Vascular Endothelial Growth Factor (Fator de Crescimento do Endotélio Vascular)

WGCNA: Weighted Gene Co-expression Network Analysis (Análise de Rede de Co-expressão Gênica Ponderada)

Resumo

A doença renal pode variar de um declínio súbito das funções renais (injúria renal aguda, IRA) até a perda progressiva e irreversível da taxa de filtração (doença renal crônica, DRC). A progressão e a relação causal entre IRA e DRC pode envolver uma cascata complexa de sinalização molecular, tornando-se alvos terapêuticos das células-tronco estromais mesenquimais multipotentes (CTMs), as quais possuem uma multiplicidade de propriedades renoprotetoras. Biomoléculas podem agir como biomarcadores e classificadores em processos patofisiológicos como a IRA e DRC. Padrões ou assinaturas moleculares podem ser ferramentas úteis na disfunção renal auxiliando na detecção, estadiamento ou no prognóstico precoce a alterações clínicas. Portanto, este presente estudo identificou uma assinatura transcricional associada à regeneração/proteção renal após terapêutica com as CTMs. Por meio de ferramentas computacionais específicas e de bioinformática utilizamos dois bancos de dados transcriptômicos (Gene Expression Omnibus, GEO); o primeiro identificou e treinou a assinatura, (i) banco de treino; e um segundo banco (GEO) foi utilizado para ensaio de validação, (ii) banco de validação. Os resultados mostraram a existência de 137 genes diferencialmente expressos em comum a pelo menos 2 bancos de dados de treino (GSE275449-DRC, GSE153217-ND, GSE59958-IRA). Em seguida, por nível de expressão/regulação e significância estatística, identificamos 4 potenciais genes alvos que poderiam ser usados em nossa assinatura (Pak3, Cll2, Tnn, Gjd2). Ao realizarmos combinações por pares detectamos que Tenascina N (Tnn) e Conexina 36 (Gjd2) se destacaram com expressão invertidas em rins tratado com CTMs (regeneração) e rins doentes, sendo assim utilizados na assinatura de treino. Como esperado a assinatura de treino (z-score Tnn-Gjd2) mostrou-se eficiente em classificar corretamente 100% dos rins doentes e rins tratados com CTMs (saudável/regeneração) com ambos apresentando um FDR = 0. Dessa forma testamos a nossa assinatura (z-score Tnn-Gjd2) em um segundo banco de dados chamado de validação (GSE66494-DRC, GSE30528-ND, GSE30718-IRA) e identificamos uma classificação, porém com falsos positivos para rins doentes (38%) e rins saudáveis (34%). No intuito de aprimorar a nossa assinatura realizamos o cálculo classificatório nos bancos de dados de treino, mas agora separado por patologias e observamos para IRA falsos positivos para rins doentes (50%) e rins saudáveis (37%), para DRC falsos positivos para rins doentes (32%) e rins saudáveis (0%), e para ND falsos positivos para rins doentes (55%) e rins saudáveis (53%). Nossos resultados demonstraram que o uso de Tnn e Gjd2 podem ser utilizados para classificar rins doentes e rins saudáveis e/ou em regeneração, no entanto seu poder classificatório mostrou-se mais efetivo para DRCs, sendo que Tnn pode também ser utilizado como um excelente marcador de rins

saudáveis. Esperamos redefinir, treinar e validar a nossa assinatura molecular transcricional em bancos de dados maiores abrangendo maior número de patologias renais e amostras. Acreditamos que Tnn e Gjd2 possuem um potencial de serem utilizados para múltiplas aplicações futuras, não só como assinaturas de classificação, mas também como biomarcadores de estadiamento e prognóstico renal em diversas condições de patologias renais.

Palavras-chaves: Células-tronco Estromais Mesenquimais Multipotentes, Regeneração, Assinatura Molecular e Bioinformática.

Abstract

Renal disease ranges from an abrupt decline in kidney function (acute kidney injury, AKI) to the progressive and irreversible loss of filtration rate (chronic kidney disease, CKD). The progression and causal relationship between AKI and CKD may involve a complex cascade of molecular signaling, thereby constituting therapeutic targets for multipotent mesenchymal stromal cells (MSCs), which exhibit a multiplicity of renoprotective properties. Biomolecules can act as biomarkers and classifiers in pathophysiological processes such as AKI and CKD. Molecular patterns or signatures may serve as useful tools in renal dysfunction, aiding early detection, staging, or prognostication prior to overt clinical alterations. Accordingly, the present study identified a transcriptional signature associated with renal regeneration/protection following MSC therapy. Using dedicated computational and bioinformatics approaches, we analyzed two transcriptomic datasets (Gene Expression Omnibus, GEO): the first to identify and train the signature (i) training set; and a second GEO dataset for validation (ii) validation set. The results revealed 137 differentially expressed genes shared by at least two training datasets (GSE275449-CKD, GSE153217–non-diseased [ND], GSE59958-AKI). Next, based on expression/regulation level and statistical significance, we identified four potential target genes for inclusion in our signature (Pak3, Ccl2, Tnn, Gjd2). Pairwise combinations indicated that Tenascin N (Tnn) and Connexin 36 (Gjd2) were most discriminative, showing inverse expression in MSC-treated kidneys (regeneration) versus diseased kidneys, and were therefore used in the training signature. As expected, the training signature (z-score Tnn–Gjd2) correctly classified 100% of diseased kidneys and MSC-treated (healthy/regenerating) kidneys, both with FDR = 0. We then tested our signature (z-score Tnn–Gjd2) in a second dataset designated for validation (GSE66494-CKD, GSE30528-ND, GSE30718-AKI) and obtained classification, albeit with false positives for diseased kidneys (38%) and healthy kidneys (34%). To refine the signature, we recalculated the classification score in the training datasets stratified by pathology, observing for AKI false positives for diseased kidneys (50%) and healthy kidneys (37%); for CKD, false positives for diseased kidneys (32%) and healthy kidneys (0%); and for ND, false positives for diseased kidneys (55%) and healthy kidneys (53%). Our findings demonstrate that Tnn and Gjd2 can be leveraged to classify diseased and healthy and/or regenerating kidneys; however, their classificatory power was more effective for CKD, with Tnn additionally emerging as an excellent marker of healthy kidneys. We aim to redefine, train, and validate our transcriptional molecular signature in larger datasets encompassing a broader range of renal pathologies and samples. We posit that Tnn and Gjd2 hold promise for multiple future applications, not only as classificatory

signatures but also as biomarkers for renal staging and prognosis across diverse renal disease conditions.

Keywords: Multipotent Mesenchymal Stromal Cells; Regeneration; Molecular Signature; Bioinformatics.

1. Introdução

1.1. *Células-tronco Estromais Mesenquimais Multipotentes*

No século 19, os precursores dos estudos com células-tronco evidenciaram a origem medular de células estromais que atuavam na regeneração. Esse fato iniciou-se com Conheim em 1867 estudando a regeneração de feridas com foco observacional em suas extremidades (1). Após um “lapso temporal”, Maximow correlacionou componentes sanguíneos e a mesoderme durante a embriogênese (1). Então, em 1976, Friedstein e colaboradores isolaram e identificaram pela primeira vez células mesenquimatosas provenientes de medula óssea em modelos murinos (2). Finalmente, no fim da década de 80, Maureen Owen e Arnold Caplan ampliaram ainda mais a pesquisa de Friedstein e classificaram essas células como células-tronco mesenquimais, propondo a existência de uma célula-tronco adulta multipotente responsável pela formação e desenvolvimento de tecidos mesodérmicos (1).

Por definição funcional, essas células tronco-adultas receberam o nome de células-tronco estromais mesenquimais multipotentes (CTMs) e são caracterizadas por serem: células aderentes, com capacidades de auto renovação, com potencial de diferenciação mesodermal (adipócitos, osteócitos e condrócitos) e por serem obtidas de diferentes tecidos como: tecido adiposo, medula óssea, placenta, cordão umbilical, polpa dentária (2). Recentemente, foi realizado um estudo para determinar um consenso para quais características definitivas utilizar na classificação das CTMs (RENESME et al., 2025), neste artigo, os marcadores de superfície celular incluídos como sendo essenciais para caracterização definitiva foram: “CD73+, CD90+ e CD105+” com expressão positiva e “CD45-” com expressão negativa (3). As capacidades terapêuticas das CTMs estão ligadas tanto à sua habilidade de diferenciação quanto à sua sinalização parácrina, as quais incluem propriedades regulatórias e modulatórias. Algumas dessas capacidades podem compreender efeitos pró-angiogênico, anti-fibróticos, anti-apoptóticos, anti-estresse oxidativo e anti-inflamatório, além de proteção tecidual como neuroproteção, regeneração muscular, renal, pulmonar, cardíaca, cartilaginosa e óssea (3). Essas capacidades estão ligadas a fatores moleculares e particulados dos mais diversos (exossomos e microvesículas) que para maior especificidade, serão abordados no próximo tópico.

1.2. *Doença renal e Células-tronco Estromais Mesenquimais Multipotentes*

A injúria renal aguda (IRA) consiste em uma perda ou declínio súbito das funções renais. É caracterizada pelo aumento da creatinina sérica ($\geq 0,3$ mg/dL dentro de 48 hs ou $\geq$

1,5 vezes o valor de referência), diminuição do volume urinário ($< 0,5$ mL/kg/hs dentro de 6 hs), ou ambos simultaneamente (4). Apesar dessas alterações, a IRA é uma condição clinicamente silenciosa, não causa sintomas ou dores anormais, além de ser negligenciada em muitos casos (5). Outros parâmetros para sua detecção incluem a Taxa de Filtração Glomerular (TFG) e os níveis de Cistatina C (4, 5). A TFG, apesar de medida em condições de pesquisa, é trabalhosa e consome tempo, e por isso, a diurese e monitoramento de solutos liberados pelos rins como creatinina, ureia e cistatina C, são os métodos comumente utilizados.

As mudanças na creatinina sérica, porém, calham a sensibilidade, pois em uma pessoa saudável aproximadamente 50% da TFG precisa ser perdida para detecção de mudança em seu nível basal para patológico, além de ser alterada por fatores como idade e massa muscular (5). A Cistatina C é uma inibidora de protease com peso molecular de 13 kDa. Ela é filtrada nos glomérulos, reabsorvida quase que totalmente e metabolizada pelas células do túbulo proximal. Apesar de não ser produzida apenas pelos rins, a cistatina C é um marcador importante para a função renal, pois sericamente reflete da TFG. Quando a função renal está comprometida, a cistatina C é filtrada de forma menos eficiente, o que causa um aumento de seus níveis no sangue, sendo mais específica que a creatinina sérica (5). No que tange suas causas, múltiplas condições podem ser associadas ao desenvolvimento da IRA como: hipoperfusão renal, síndrome cardiorrenal tipo 1, exposição à nefrotoxicidade, sepse, procedimentos cirúrgicos, hipertensão intra-abdominal, glomerulonefrite rapidamente progressiva, nefrite intersticial aguda e IRA pós-renal (5).

A relação causal direta entre a IRA e a Doença Renal Crônica (DRC) ainda precisa ser compreendida, porém, múltiplos estudos mostram uma forte ligação epidemiológica entre ambas as doenças. Muitos mecanismos, que quando engatilhados por IRA, já foram propostos para levarem à DRC. Tais mecanismos envolvem um processo chamado reparação maladaptativa, que consiste na transformação de células tubulares/endoteliais em fibroblastos, os quais conduzem para um processo fibrótico e disfuncional.

A DRC é uma doença altamente prevalente (10-13% da população), irreversível, progressiva e associada com maiores riscos cardiovasculares (6). É caracterizada por uma TFG abaixo de 60 mL/min/ $1,73\text{m}^2$, por um período igual ou maior a noventa dias, precedida ou não por episódios pretéritos de IRA (5, 6). Além da TFG, um indicador valioso é a Albuminúria, definida pela presença de mais de 30 mg de albumina no exame de urina de 24 horas ou mais de 30 mg/g de albumina em uma amostra de urina ajustada pela creatinina urinária (6).

Algumas das causas principais da DRC são: diabetes, hipertensão, glomerulonefrite crônica, litíase, pielonefrite crônica, uso crônico de medicação anti-inflamatória, doença

autoimune, doença do rim policístico, doença de Alport, malformações congênitas e injúria renal aguda prolongada (6). A manifestação final em comum de todas DRCs é a fibrose renal, a qual é representada pela perda de reparação tecidual e é caracterizada pela presença de glomerulosclerose, atrofia tubular e fibrose intersticial (7).

No Brasil, segundo o Censo Brasileiro de Diálise de 2023, são 157,357 pacientes submetidos à diálise (8). Embora seja incerta qual parcela deste número é tomada por pacientes com DRC (ou mesmo IRA), a doença é associada a maiores riscos de doenças cardiovasculares com alto grau de letalidade. Dados de 2013 associaram a DRC com 4% da mortalidade mundial (7). Dessa forma as doenças renais como a IRA e DRCs compreendem uma escala de enfermidades de grande impacto público, e devido a suas fisiopatologias de acometimento, tornam-se alvos terapêuticos eminentes das CTMs. Nesse contexto, as propriedades pró-regenerativas das CTMs supracitadas como diferenciação tecidual, e modulação por secreção trófica ou por particulados, tornam-se mecanismos fundamentais que promovem a renoproteção (9).

Como exemplo disso podemos destacar, os processos inflamatórios que são de grande importância para os processos patogênicos nos rins. Em estudos recentes, a aplicação das CTMs e seu exossoma (conteúdo de exossomos) promoveram em tecidos renais a inibição de fatores inflamatórios como IL-6, fator de necrose tumoral alfa (TNF- α), fator nuclear-kappa B (NF- κ B) e da proteína quimioatrativa de monócitos-1 (MCP-1 ou CCL2). Além disso, foi reportado a diminuição da sinalização via proteína quinase ativada por mitógeno quinase (MAPK) e a infiltração de macrófagos (9).

A terapêutica com CTMs e seus derivados, especialmente exossomos, tem demonstrado efeitos antiapoptóticos relevantes no contexto da morte tecidual renal, tanto em modelos de IRA quanto em DRC. O mecanismo central envolve a modulação de vias de sinalização intracelulares e a transferência de microRNAs funcionais, os quais atuam no balanço de vias pró- e anti-apoptóticas (caspases, Bax, BCL-2, ERK1/2), desempenhando assim um papel importante na senescência e morte celular (9).

Adicionalmente, a falta de suprimento de oxigênio induzido por mecanismos de hipóxia é uma importante causa de dano renal. O transplante de CTMs, por ação direta ou trófica, parece aumentar a expressão do fator de crescimento endotelial vascular (VEGF) nos tecidos renais, promovendo a angiogênese e reduzindo a perda de capilares peritubulares (9).

De forma complementar a fibrose túbulo-intersticial é uma marca registrada na progressão para a DRC. Já foi demonstrado que CTMs de medula óssea bloqueiam a sinalização Akt/GSK3 β /Snail em células/tecidos renais, a qual é essencialmente pró-

fibrótica. Além disso, o estresse oxidativo, presente na fisiopatologia da DRC, também é mitigado nos tecidos renais através da diminuição de espécies reativas de oxigênio (ROS) após o tratamento com CTMs. Finalmente, a autofagia e a senescência são eventos, que aparecem aumentados e diminuídos, respectivamente, após a aplicação das CTMs em tecidos renais, podendo prevenir a disfunção renal por modulações de vias moleculares específicas (9, 10).

1.3. Assinaturas Moleculares

As biomoléculas são moléculas presentes nas células e que possuem importância nos processos bioquímicos e estruturais dos seres vivos. Enquanto isso, os biomarcadores são quaisquer medições dessas biomoléculas que podem ser estimadas e mensuradas como um índice em potencial para a diferenciação de um estado normal ou anormal celular, ou mesmo de estadiamento de um processo patofisiológico, ou de respostas toxicológicas à fármacos (11). Logo, biomarcadores moleculares são marcadores à nível molecular nas células, com alguns deles sendo: proteínas, ácidos nucleicos, metabólitos e genes, que mensuram a homeostase e são capazes de serem utilizados para distinguir um estado “fisiologicamente normal” de um estado “alterado” (12).

Diversos alvos podem ser utilizados em conjunto para atingir uma rede de marcadores, assim, correlacionando dois ou mais biomarcadores simultâneos. Dessa forma, é possível estabelecer uma fidelidade maior, pois considerando uma injúria tecidual, especialmente as renais, esses marcadores podem representar uma multitude de disfunções coordenadas de forma sinérgicas e antagônicas, e não apenas questões singulares isoladas (12). A relevância desse método se dá pelo fato já discutido de que, sozinhos, os parâmetros de filtração glomerular, creatinina e volume urinário são indicadores insuficientes para determinação de um estado patológico, *per si*. Porém um conjunto de biomarcadores compondo uma “assinatura” podem traduzir de forma mais eficiente as funções celulares e teciduais nas condições que se encontram (4-6, 11).

Além da aplicação em diagnósticos, os biomarcadores também podem ser utilizados para diferenciar populações celulares e estados de ativação e/ou alteração celular. Por meio de tecnologias computacionais que integram um conjunto de ferramentas estatístico e matemático podemos identificar assinaturas ou conjunto de “padrões” de modulação diferencial de moléculas entre amostras e tecidos em diversos estados fisiológicos, revelando-se assim diferenças transcriptômicas, proteômicas entre outras, que se comportem como “perfis” únicos e específicos para uma determinada população amostral de análise (13).

1.4. Assinaturas moleculares na Doença Renal

No que tange às doenças renais, a aplicabilidade das técnicas que utilizam classificadores por padrões ou assinaturas moleculares é importante. Por exemplo, podemos utilizar bioclassificadores de lesão tecidual que surgem precocemente a alterações de albumina, as quais mensuram somente o dano glomerular, ou seja a TFG, não abrangendo danos tubulares que muitas vezes passam despercebidos na IRA (14). Também é possível através da mensuração de bioclassificadores alcançar uma maior precisão na observação do dano renal relativo à localização da porção em que ocorre a lesão renal, isso se dá pois existem marcadores diferentes para distintas partes no néfron (unidade funcional renal), além de servirem do mesmo propósito para diversas outras patologias renais específicas (14).

Adicionalmente, no caso renal, é possível aferir parâmetros de forma não invasiva, utilizando como amostragem apenas a urina como objeto de análise. Tendo esta visão, a aplicação das Ômicas (estudo que analisam grandes conjuntos de dados moleculares para entender sistemas biológicos complexos) em matrizes líquidas mostra um novo horizonte nas tecnologias diagnósticas atuais para as doenças renais (14).

Dentre os biomarcadores discutidos para patologias renais, alguns são exclusivos da IRA como: Clusterina, Cyr61, π -Glutathione S-Transferase (GST- π), N-acetyl- β -D glucosaminidase (NAG), Na⁺/H⁺ exchanger isoform 3 (NHE3), Proteína de ligação ao retinol (RBP) e β 2-microglobulina (β 2M). Enquanto outros são preferivelmente abordados na DRC como: COL IV e α 1-microglobulina. E outros ainda são comuns à ambas as patologias IRA e DRC: Fetuína A, Proteína de ligação a ácidos graxos do tipo cardíaco (H-FABP), Fator de crescimento de hepatócito (HGF), Proteína 7 de ligação ao fator de crescimento semelhante à insulina (IGFBP-7), Interleucina-18 (IL-18), Molécula da injúria renal-1 (KIM-1), Proteína de ligação a ácidos graxos do tipo hepático (L-FABP), Netrina-1, Lipocalina associada à gelatinase de neutrófilos (NGAL), Osteopontina, Inibidor tecidual de metaloproteinases-2 (TIMP-2), α 1-antitripsina, α 1-microglobulina (14). Todas essas moléculas são biomarcadores que diferem como indicadores ou classificadores de função, segmento néfrico e/ou patologia estudada.

Portanto, neste presente estudo identificamos uma assinatura de expressão gênica associada à terapêutica das CTMs na regeneração/proteção renal que podem servir de biomarcadores ou classificadores de estadiamento na IRA ou DRC. Investigamos por ferramentas computacionais específicas bancos de dados transcriptômicos, os quais estão disponíveis publicamente no repositório Gene Expression Omnibus (GEO), de tecidos

renais doentes em modelos experimentais, e comparamos aos tecidos renais tratados com CTMs. Por meio de análises estatísticas e de bioinformática, definimos e validamos uma assinatura molecular transcricional de regeneração/proteção renal que pode ser explorada para múltiplas aplicações futuras em diversas condições de patologias renais.

2. Objetivo

Identificar, caracterizar e validar uma assinatura molecular transcricional de regeneração/proteção renal associada à terapêutica com células-tronco estromais mesenquimais multipotentes (CTMs).

3. Materiais e Métodos

3.1. Configuração dos Bancos de Dados

Inicialmente para compor o banco de dados de Treino, Banco (i), selecionamos dados de expressão gênica a partir de bancos de dados públicos do *Gene Expression Omnibus*, (GEO) com os números de acesso GSE275449-DRC, GSE153217-ND e GSE59958-IRA, os quais correspondem a experimentos de RNA-Seq em tecidos renais de animais (ratos sprague-dawley, *Rattus norvegicus* e camundongos SCID, *Mus musculus*) (lesionado/doente). Para o Banco (i), Treino, todas as análises foram realizadas utilizando a comparação rins tratados com CTMs versus rins acometidos por patologias (ND, DRC, IRA, lesionado/doente). Para o Banco (ii) (GSE30528-ND, GSE 66494-DRC, GSE 30718-IRA), Validação, todas as análises foram realizadas utilizando a comparação rins saudáveis (sem patologias) versus rins acometidos por patologias renais (ND, DRC, IRA, doente). Os arquivos brutos (FASTQ) foram baixados, e submetidos aos programas FASTQC e fastp, O FastQC consiste em um software para a verificação da qualidade de sequências biológicas brutas, já o fastp atua como um pre-processador *all-in-one* para arquivos FASTQ, executando simultaneamente a aparagem (*trimming*) de adaptadores, a correção de bases e o controle de qualidade dos dados antes das etapas de alinhamento.". Os outputs foram então alinhados ao genoma de referência (e.g. rat - Rnor\6.0, mouse - GRCm38) utilizando o programa Salmon, todos os processos foram feitos na plataforma Galaxy, uma plataforma *open source* desenvolvida por equipes de múltiplas universidades internacionais. Em seguida, a quantificação gênica foi realizada obtendo-se matrizes de contagem bruta “*raw counts*” por gene. Essas contagens não-normalizadas são os dados de entrada “*inputs*” para os métodos de expressão diferencial. Todas as análises seguintes foram implementadas com o software R (versão 4.5.2) usando pacotes do Bioconductor.

Especificamente o pacote DESeq2 foi utilizado para análise de expressão diferencial no R (15). Para cada banco de dados, foi construído um objeto DESeqDataSet com o design experimental apropriado ($\sim$ condição de tratamento), filtrando genes de baixa expressão (e.g. excluindo genes com contagem total ≤ 1) para melhorar a robustez estatística. Em seguida, foi aplicada a função DESeq() para estimar fatores de normalização, dispersões e calcular estatísticas de teste (testes de Wald) para cada comparação “Rim+CTMs vs Rim lesado” em cada experimento específico. Para cada análise de expressão diferencial, os genes diferencialmente expressos (DEGs) foram selecionados segundo limiares clássicos: valor de $|\log_2FC| \geq 1,25$ (razão de pelo menos 2,5 vezes e p valor ajustado (ajuste de p-value pelo método de Benjamini–Hochberg) $\leq 0,05$). O corte de $\log_2FC$ foi aplicado bilateralmente (tanto para genes supra-regulados quanto infra-regulados). A significância estatística foi

considerada após ajuste de múltiplos testes, reportando-se sempre ao valor de p ajustado (adjust p). A lista total de DEGs para cada banco de dados foi extraída e armazenada para análises subsequentes.

3.2. *Análise Exploratória e Visualização*

Para avaliar a qualidade dos dados e a variação global entre amostras, foi aplicada a transformação de estabilização da variância (rlog) aos dados de contagem a fim de homogeneizar a variância ao longo dos níveis de expressão. Com os dados transformados, foi calculada a matriz de distâncias amostrais (distâncias euclidianas entre os vetores de expressão de cada par de amostras) para aferir similaridades globais. Um mapa de calor *heatmap* de distância amostral (mapa de calor das distâncias) foi construído usando pacotes como *pheatmap* ou *ComplexHeatmap*, ilustrando agrupamentos hierárquicos (dendrogramas) baseados na semelhança das amostras. No mapa de calor hierárquico de distâncias entre as amostras baseado nas contagens rlog-transformadas, as cores mais frias (azul) indicam amostras mais semelhantes e as cores quentes (vermelho) mais distantes, em termos de expressão. Cada célula do mapa de calor refletiu a distância (euclidiana) entre pares de amostras.

Adicionalmente, foi realizada uma análise de componentes principais (PCA) para redução de dimensionalidade e visualização dos principais eixos de variação nos dados. O PCA foi gerado utilizando os pacotes *FactoMineR* e *Factoextra* para cálculo e expressão gráfica e também foi customizado com *ggplot2*. Os dois primeiros componentes principais (PC1 e PC2) ilustram a dispersão amostral e frequentemente destacam a separação entre grupos experimentais, onde cada ponto foi representado por uma amostra (cores/dimensões codificam grupo tratamento).

3.3. *Análise de expressão diferencial*

Após a exploração inicial, para cada comparação “tecido renal + CTMs vs. tecido renal doente/lesado”, foram extraídos os resultados com a função *results()*, filtrando-se pelos critérios de significância estabelecidos como $|\log_2FC| \geq 1,25$, p valor ajustado $\leq 0,05$. Os genes diferencialmente expressos (DEGs) foram organizados em tabelas contendo $\log_2FC$, valor de p simples e p valor ajustado. Para visualização dos resultados, foram confeccionados gráficos de vulcão (volcano) e *heatmaps*. No vulcano, cada ponto representa um gene plotado em $\log_2FC$ (eixo x) versus $-\log_{10}(p\text{-valor})$ (eixo y). A região nas extremidades do gráfico (cantos superiores esquerdo e direito) identifica genes com altas mudanças de expressão e significância estatística. Genes supra- e infra-regulados foram

destacados por cores distintas. Por exemplo, genes com $\log_2FC > 1,25$ e valor de p-ajustado $< 0,05$ foram marcados como “seta para cima”, enquanto $\log_2FC < -1,25$ foram marcados como “seta para baixo”. No presente estudo, os gráficos de vulcano facilitaram a inspeção visual dos DEGs em cada banco de dado estudado (i) e (ii).

3.4. Identificação de Genes Comuns

Para definir uma assinatura molecular conservada que classificaria corretamente um rim como saudável/regeneração ou doente/patológico, foram identificados os DEGs comuns a pelo menos 2 bancos de dados do banco (i) treino. A partir desse momento, foi realizada a interseção dos conjuntos de DEGs de GSE275449-DRC, GSE153217-ND e GSE59958-IRA. Essa análise de interseção foi efetuada com a ajuda do pacote Biovenn (v1.1.3) e da função `intersect()` em R, que permitiu visualizar conjuntos de genes compartilhados entre múltiplos experimentos (datasets). Genes presentes nos três conjuntos de bases de dados (interseção tripla) e genes presentes em dois dos conjuntos (interseção dupla) foram considerados na assinatura final. Essa etapa reduziu o ruído e destacou genes consistentes associados ao tratamento regenerativo das MSCs.

3.5 Análise de Co-Expressão

No final, os DEGs mais representativos em termos e módulos gênicos (maior fold change, menor variância e menor p valor) foram selecionados para a obtenção de uma assinatura molecular de regeneração renal que foi aplicada em novos bancos de dados (validação) de doenças renais diversas Banco (ii) (GSE30528-ND, GSE 66494-DRC, GSE 30718-IRA) para a verificação se, de fato, os DEGs modulados e que compunham a assinatura tinham poder de classificação entre rim saudável/regenerado e rim doente/patológico.

4. Desenho experimental

Em resumo, abaixo encontra-se a figura do fluxograma esquemático do pipeline de análise computacional para obtenção da assinatura renal de classificação.

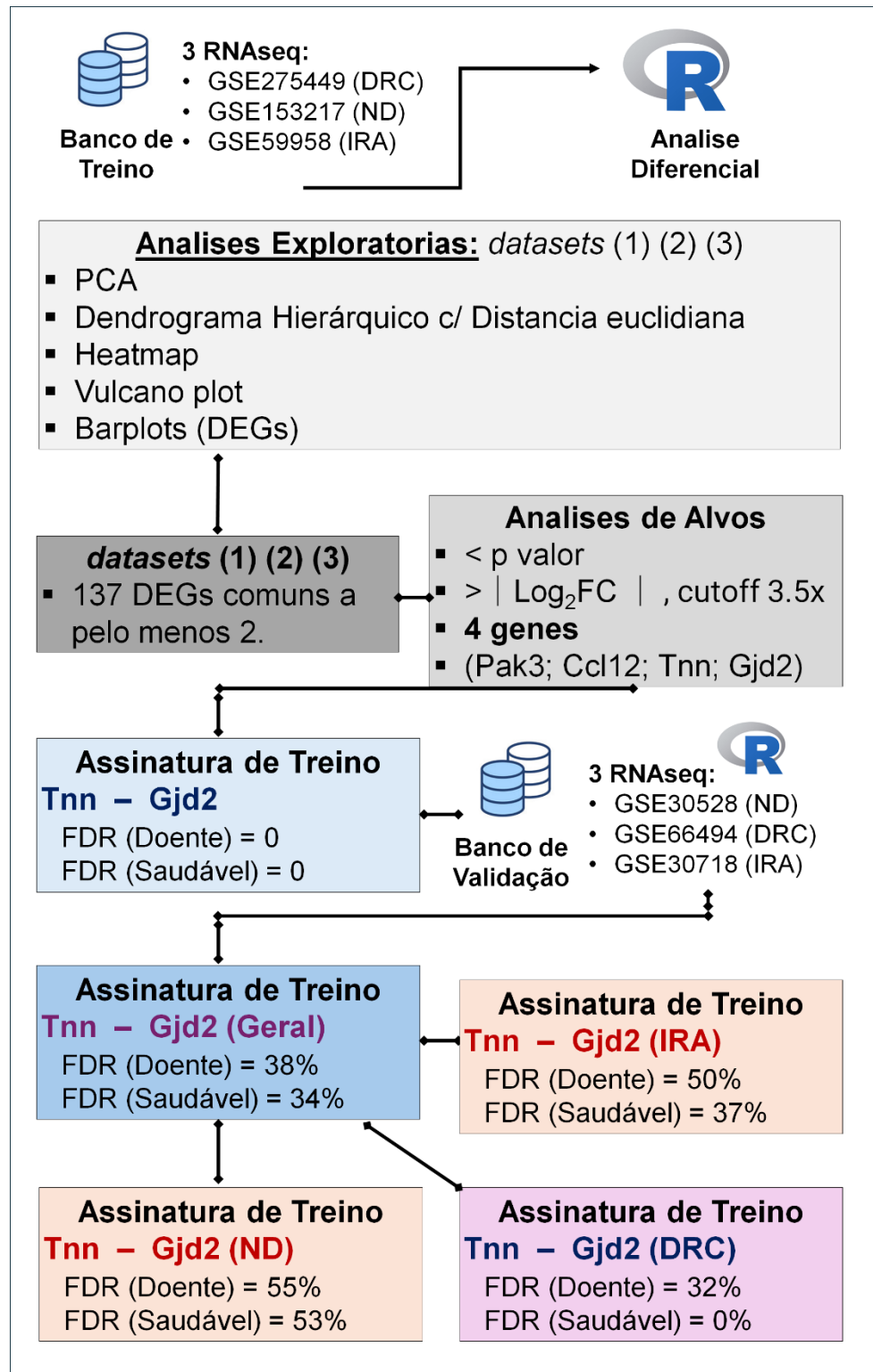


Figura 1. Desenho Experimental da assinatura de classificação renal. Os bancos de dados de treino (GSE275449-DRC, GSE153217-ND e GSE59958-IRA) foram submetidos a análises que incluem etapas de pré-processamento (obtenção de dados do GEO e quantificação), análise exploratória (PCA, correlação amostral), identificação de DEGs (DESeq2, volcano plot, heatmap), interseção de DEGs (UpSetR) e análises subsequentes para a obtenção da assinatura renal de classificação com posterior validação em bancos de dados

5. Resultados

5.1. *Análise nos Bancos de Dados de Treinamento*

Inicialmente foram compilados 3 bancos de dados para a formação do Banco (i) designado como de treinamento (GSE59958-IRA, GSE275449-DRC, GSE153217-ND), respectivamente, o primeiro representa *Mus musculus* cuja administração das CTMs foi através da veia caudal, e segundo e terceiro ratos *Sprague-dawley*, onde também foram injetados através da veia caudal. Primeiramente os dados brutos foram subidos ou baixados direto na plataforma do Galaxy.org, onde sofreram o primeiro controle de qualidade através do FASTQC; seguindo então para a fase de *trimming*, ou fase de corte de leituras desnecessárias como adaptadores, primers e outras regiões que poderiam gerar “ruídos”, ou seja, dados indesejados às análises. Após isso, foi realizado o Salmon, que quantifica a abundância de transcritos a partir de dados de sequenciamento de RNA (RNA-Seq).

O principal resultado do Salmon é um arquivo que contém uma lista de todos os transcritos do transcriptoma de referência (ex: genes e isoformas) e suas respectivas abundâncias estimadas. Tais arquivos foram então lidos no R (4.5.2) utilizando a função ‘`transcripts()`’ do pacote GenomicFeatures. Os dados de contagem foram então exportados como arquivos .csv com a função ‘`write.csv()`’. Com as contagens e os metadados (no caso do interesse do estudo, os grupos que representam cada amostra) foi possível realizar a normalização dos resultados com o ‘DESeq2’, antes, porém, foi realizada a análise exploratória, gerando os gráficos de PCA, Matriz de correlação e Dendrograma representados nas Figuras 2-4 de cada banco de dados do treino Banco (i).

Para os três bancos de dados do treino, a análise do Componente Principal (PCA) mostrou a separação consistente dos grupos Rins tratados com CTMs e grupos Rins doentes, porém no banco de dados GSE3 de Nefropatia Diabética mostrou uma interseção entre as elipses, demonstrando uma variabilidade maior nas amostras desse banco (Figura 2, 3,4. A). De uma forma geral, nos três bancos de dados do treino foi identificada uma correlação positiva consistente entre as amostras de cada grupo (Figuras 2,3,4. A). A visualização da Distância Euclidiana revelou uma distinção hierárquica clara entre os grupos e maior homogeneidade entre as amostras para os bancos de dados do treino GSE1 (IRA) e GSE3 (ND), porém o banco de dados GSE2 (DRC) apresentou uma não distinção com maior variabilidade entre as amostras dentro de cada grupo (Figuras 2,3,4. C).

Após a análise diferencial com a função ‘DESeq(dds)’ , por meio do gráfico de vulcão

foi possível visualizar genes diferencialmente expressos (DEGs, visualização de genes não significantes em cinza, genes infra-expressos em rosa, e genes supraexpressos em azul) nos 3 bancos do treino. Isso também foi corroborado pelo mapa de calor que revelou um padrão de expressão diferencial (áreas supraexpressas em vermelho, e áreas infraexpressos em azul) das amostras dentro de cada grupo de cada um dos bancos de dado utilizados no treino, Banco (i): GSE59958-IRA, GSE275449-DRC, GSE153217-ND). Para cada banco de dados do treino o número total de DEGs foi representado através de um gráfico de barras, o banco GSE59958-IRA exibiu 167 genes infra-regulados e 223 genes supra-regulados na comparação Rim + CTMs *versus* Rim Lesado. Na mesma comparação o banco GSE275449-DRC mostrou 1043 genes infra-regulados e 311 genes supra-regulados; e por fim, seguindo o mesmo perfil de comparação o banco GSE153217-ND revelou 218 genes infra-regulados e 76 genes supra-regulados. (Figuras 2,3,4. D-F).

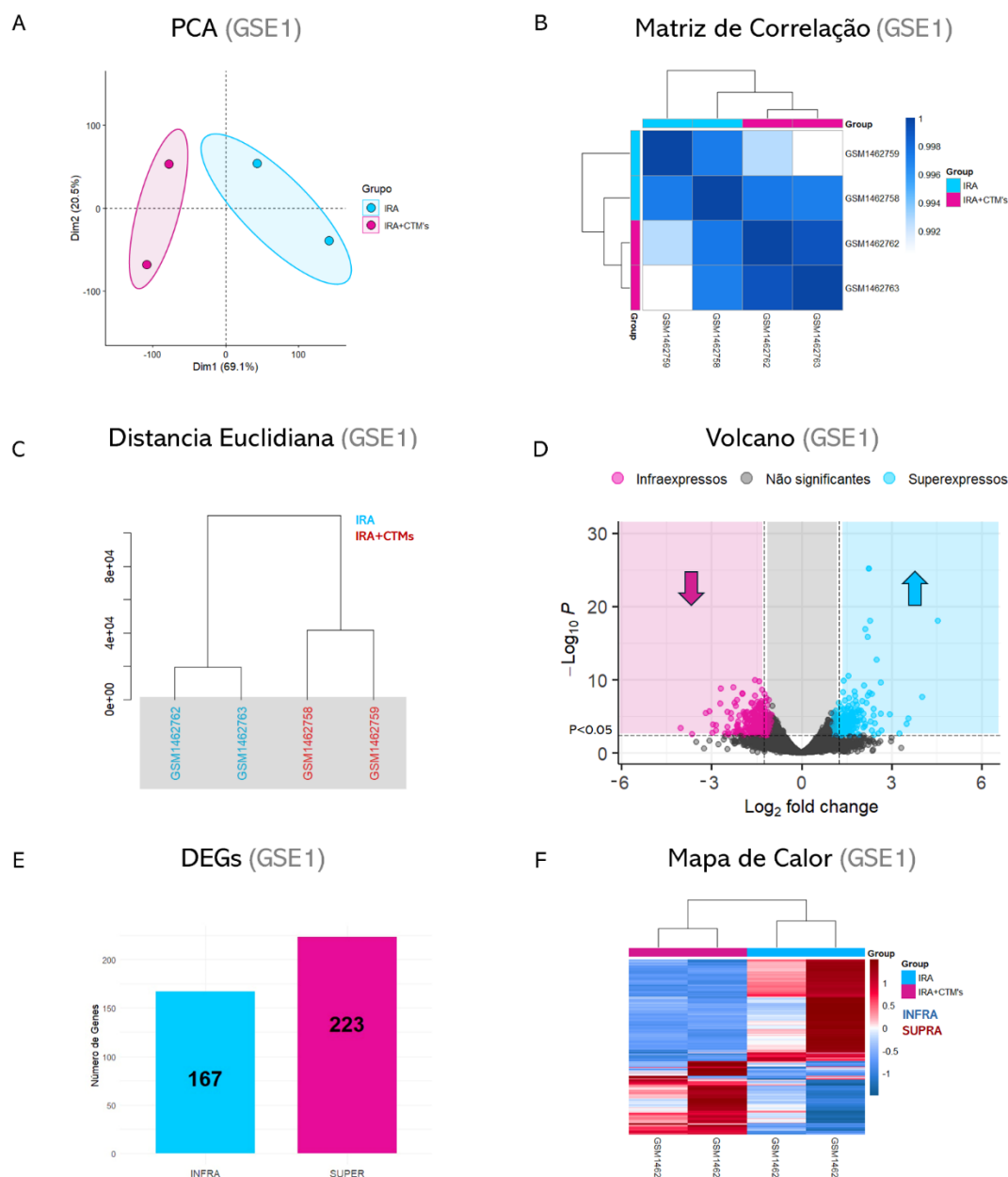


Figura 2. Perfil transcricional do primeiro banco de treinamento (GSE59958 IRA). (A) Gráfico de PCA,

(B) Matriz de correlação, (C) Dendograma baseado na distância euclidiana, (D) Gráfico de vulcão (Volcanoplot) baseado nos limiares definidos ($P < 0,05$ e $\text{Log}_2\text{FoldChange}$, a expressão relativa dos genes, maior ou igual a 1,25), (E) Gráfico de barras mostrando a quantidade de genes diferencialmente expressos, e (F) Mapa de calor (Heatmap) dos perfis transcricionais distintos entre as amostras dos grupos em comparação.

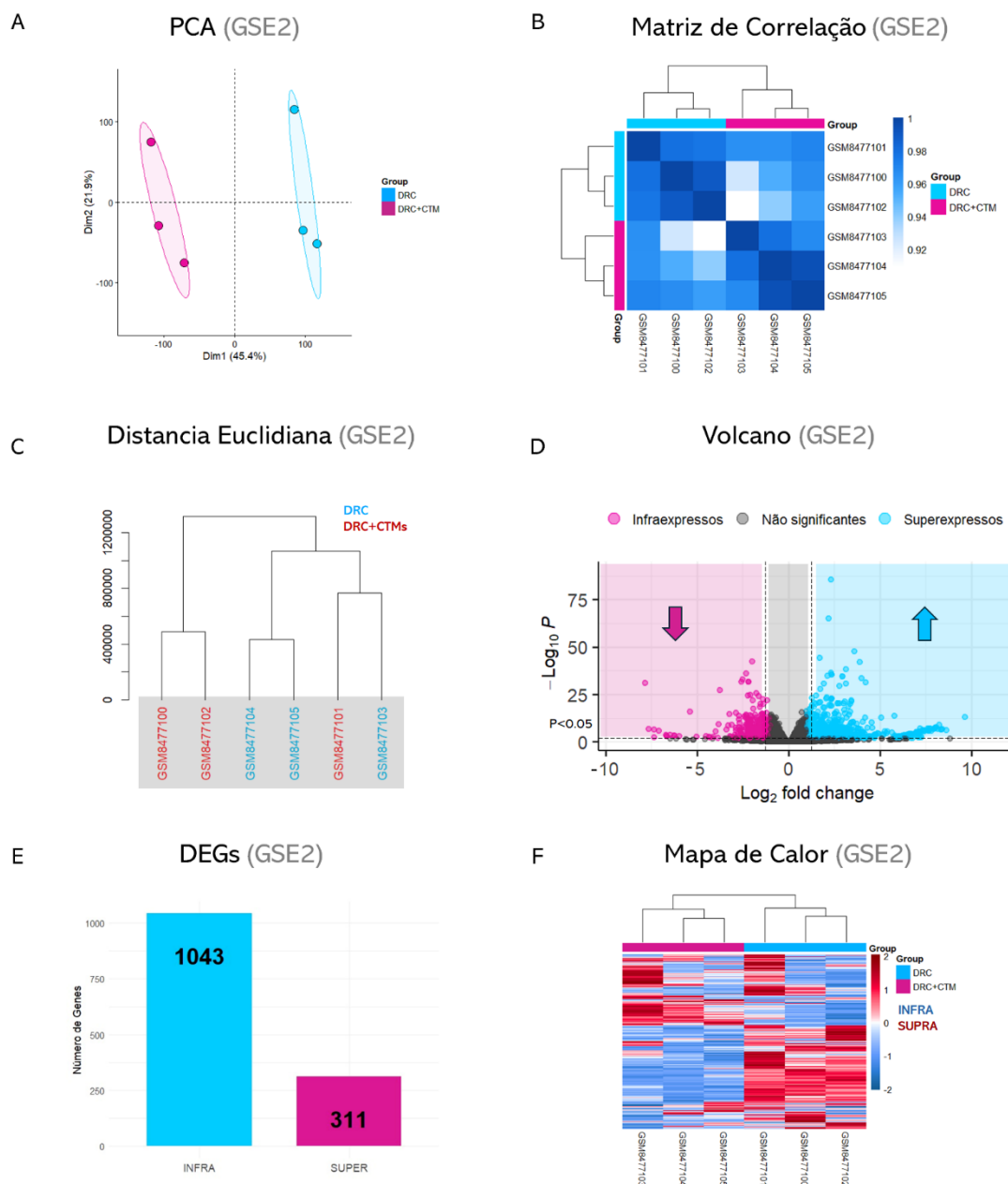


Figura 3. Perfil transcricional do segundo banco de treinamento (GSE275449 DRC). (A) Gráfico de PCA, (B) Matriz de correlação, (C) Dendograma baseado na distância euclidiana, (D) Gráfico de vulcão (Volcanoplot) baseado nos limiares definidos ($P < 0,05$ e $\text{Log}_2\text{FoldChange}$, a expressão relativa dos genes, maior ou igual a 1,25), (E) Gráfico de barras mostrando a quantidade de genes diferencialmente expressos, e (F) Mapa de calor (Heatmap) dos perfis transcricionais distintos entre as amostras dos grupos em comparação.

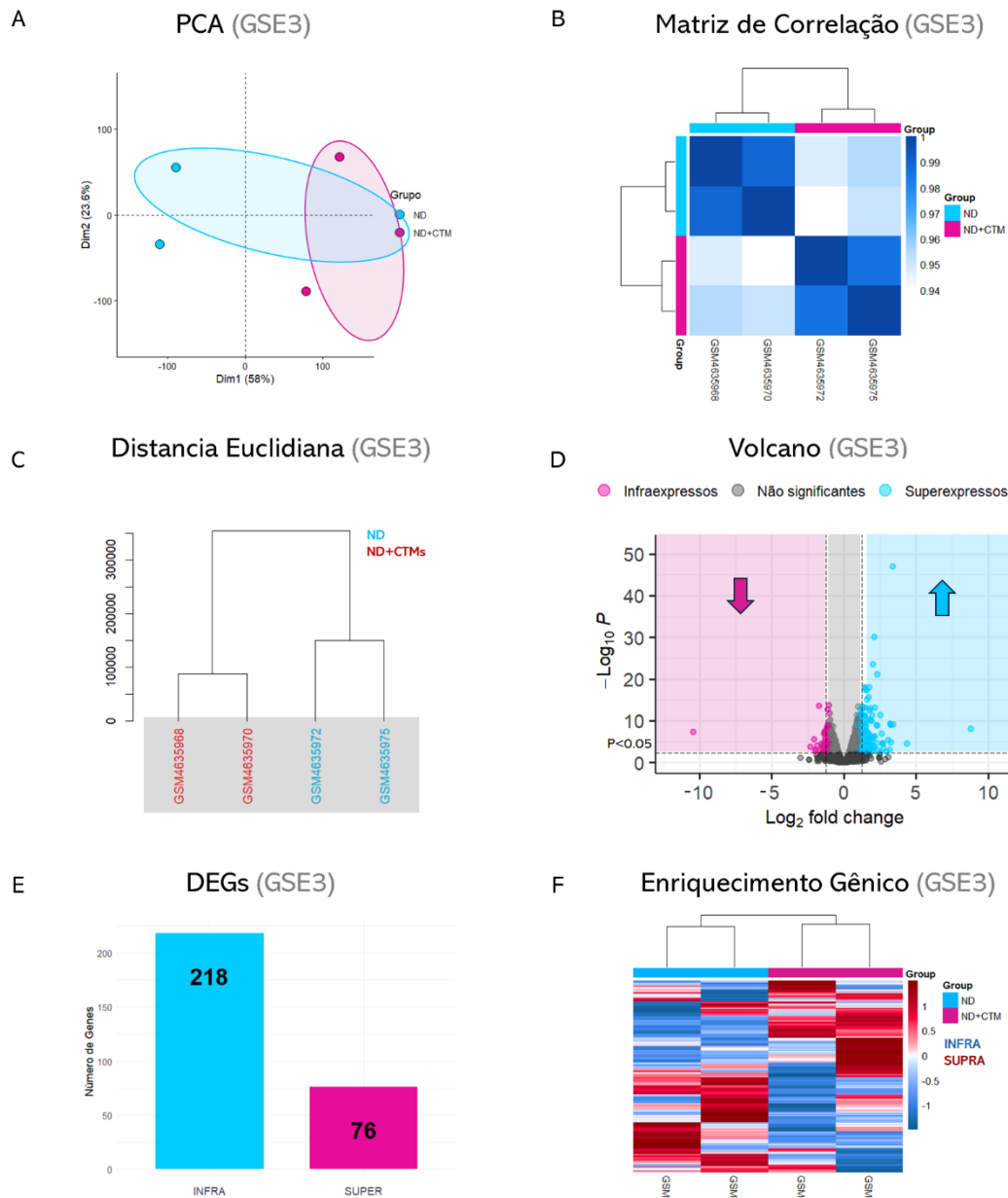


Figura 4. Perfil transcricional do primeiro banco de treinamento (GSE153217 ND). (A) Gráfico de PCA, (B) Matriz de correlação, (C) Dendograma baseado na distância euclidiana, (D) Gráfico de vulcão (Volcanoplot) baseado nos limiares definidos ($P < 0,05$ e $\log_2$ fold change, a expressão relativa dos genes, maior ou igual a 1,25), (E) Gráfico de barras mostrando a quantidade de genes diferencialmente expressos, e (F) Mapa de calor (Heatmap) dos perfis transcricionais distintos entre as amostras dos grupos em comparação.

5.2 Identificação de genes alvos e Assinatura Renal de Treino

Após a visualização dos genes diferencialmente expressos em cada banco de treino, selecionamos os genes comuns a pelo menos 2 bancos de dados entre os três analisados. Os genes ditos como “objetos” de resultados filtrados foram exportados para uma tabela (.csv) para análise conjunta, esta foi aberta no R e sujeita ao pacote ‘BioVenn’ que tornou possível a visualização dos genes presentes em dois ou mais bancos através de um diagrama de Venn (Figura 5. A). Os genes comuns à pelo menos dois bancos de dados foram plotados em um diagrama de dispersão de pontos por força de regulação modular ($|\log_2 \text{fold change}|$) maior

que 3,5 e significância estatística ($< p$ valor ajustado) (Figura 5. B). Essa análise identificou 4 potenciais genes com um perfil de modulação diferencial (Pak3, Ccl12, Tnn, Gjd2) entre as patologias associadas aos bancos de dados do treino (IRA, DRC, ND), os quais serviriam para a construção da assinatura renal de classificação (Figura 5 C).

Pouco antes da criação dos objetos de resultado, foi utilizada a função 'vst()' ou *Variance Stabilizing Transformation*, para permitir a comparação do nível de expressão referencial dos genes entre os bancos de dados do treino. Foi calculado o Z-score, uma medida que indica quantos desvios padrão um determinado ponto de dados está distante da média de um conjunto de dados. Este foi definido pela equação $Z_i = ((x_i - \bar{x}) / s)$, Onde Z_i é o Zscore para a observação i , x_i o valor da observação, $\bar{x}$ denota a média amostral e s corresponde ao desvio padrão amostral. Na prática o Z-score não mostra a expressão absoluta, mas sim como a expressão de um gene qualquer se desvia do "normal" (a média) em um conjunto de amostras, permitindo identificar padrões de supra- e infra-expressão de forma relativa. Sua visualização revelou que dentre os 4 genes selecionados destacou-se um perfil dicotômico para os genes Tnn (tenascina N) e Gjd2 (conexina 36), sugerindo que maior expressão de Tnn seja positiva para um perfil renal saudável/regenerado e Gjd2, o contrário, sendo conseqüentemente mais expresso nos rins lesados/patológicos (Figura 5 D).

Como próxima etapa, o perfil antagônico renal entre o estado renal regenerado e patológico de Tnn e Gjd2 permitiu-nos realizar uma subtração entre os Z-scores obtendo-se assim uma assinatura de treino de classificação renal (Z-score Tnn - Gjd2). Por meio desse cálculo foi possível classificar corretamente com 100% de sensibilidade, especificidade e precisão (FDR=0 para ambos) as amostras previamente rotuladas nos bancos de dados como Rim doente e Rim regenerado (Figura 5 E).

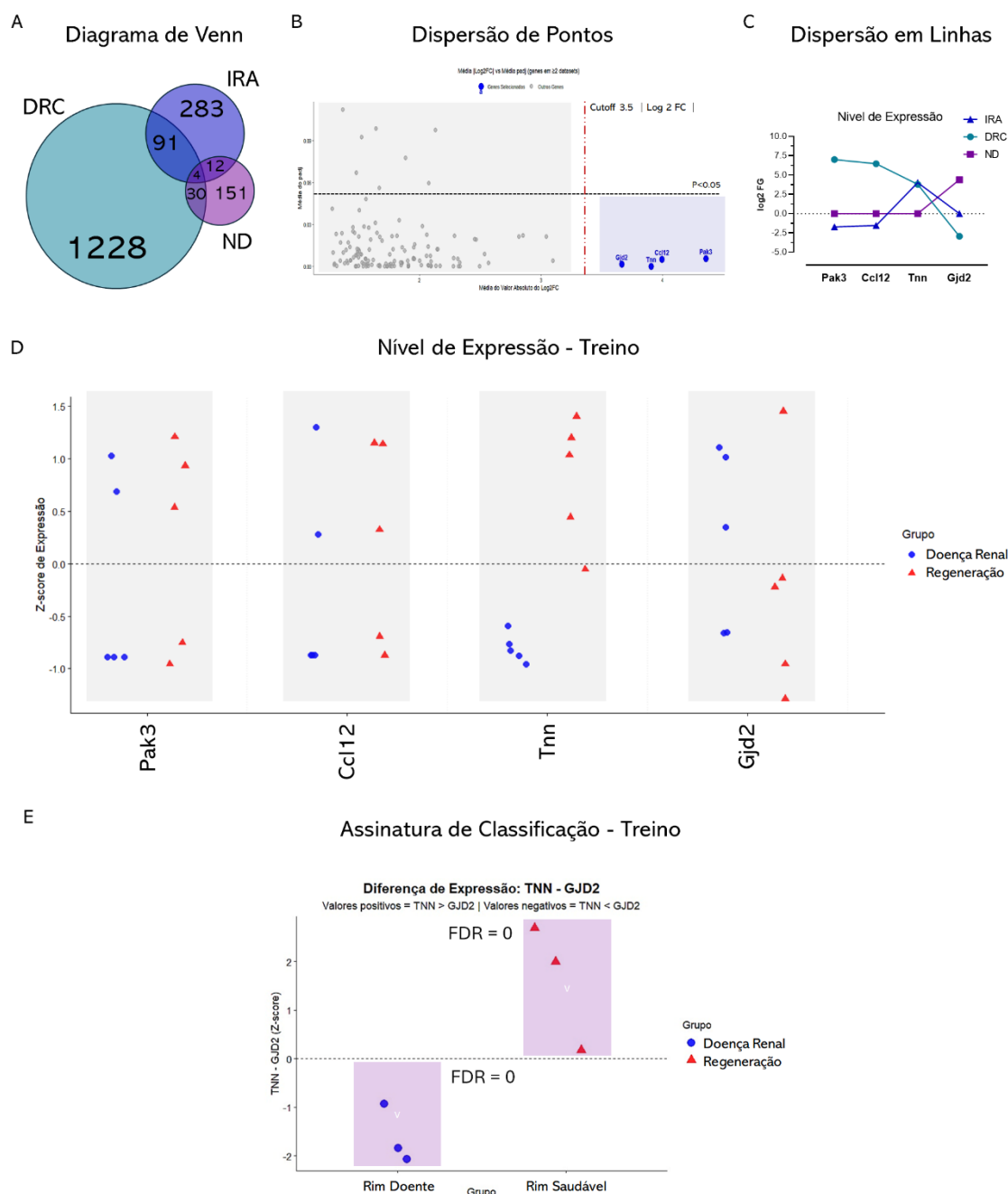


Figura 5. Obtenção da assinatura de treinamento renal de classificação. (A) Diagrama de Venn, mostrando as interseções entre os genes diferencialmente expressos (DEGs) nos bancos de dados de treinamento, (B) Gráfico de dispersão de pontos (scatterplot), evidenciando o perfil distinto dos quatro genes selecionados (Cutoff de Log2FoldChange de $<3,5$ e p -valor $< 0,05$), (C) Gráfico de linhas, que aponta a diferença dos níveis de expressão dos 4 genes selecionados entre as patologias renais do banco de treinamento, (D) Comparação do Z-score de expressão, indiciando um perfil dicotômico entre os genes Tnn e Gjd2, e (E) Teste da assinatura de treinamento baseada na diferenças dos Z-scores de Tnn e Gjd2 ($Z(TNN) - Z(GJD2)$), denotando um perfil de classificação entre Rim doente e Rim regenerado.

5.3 Validação da Assinatura Renal de Classificação

Para a etapa de validação de nossa assinatura de classificação renal, nós escolhemos três banco de dados de RNAseq em humanos que correspondessem as mesmas patologias renais observadas no treino, portanto foram escolhidos com base na qualidade dos dados os bancos de validação GSE30528-ND, GSE66494-DRC, GSE30718-IRA, que respectivamente representam: Nefropatia diabética em glomérulos dissecados de rins humanos, Doença renal crônica em rins humanos e Injúria renal aguda em biopsias. Tais bancos de dados de validação, Banco (ii), foram escolhidos pela excelência em seu controle de qualidade, dessa forma ao contrário do banco de treino, não foram necessárias muitas ferramentas de processamento, com exceção de uma checagem dos metadados para certificação de que os dados estavam condizentes com as amostras.

Dessa foram, verificamos individualmente o nível de expressão referencial pelo Z-score dos 4 genes selecionados na análise de treino (Pak3, Ccl12, Tnn, Gjd2) nas patologias renais do banco validação e novamente verificamos um perfil dicotômico para TNN e GJD2, porém dessa vez com menor representatividade (Figura 6. A). De forma semelhante a análise de treinamento, testamos o poder de classificação de nossa assinatura (Z-score Tnn - Gjd2) de forma conjunta para os dados existentes nos 3 bancos de dados de validação. Essa análise identificou uma classificação parcial na comparação “doente” *versus* “saudável” entre as amostras previamente rotuladas como Rim doente e Rim saudável, com uma sensibilidade de 84%, especificidade de 35%, precisão de 61% e FDR de 38% (Figura 6. A).

Na tentativa de melhorar o poder de classificação e identificar se algum banco de dado de alguma patologia renal específica poderia interferir no potencial discriminatório geral de nossa assinatura, partimos para uma análise separada em cada uma das patologias renais do banco de dados de validação. A análise separada mostrou um perfil de expressão referencial distinto para cada gene em seu grupamento, no entanto na doença renal crônica (GSE66494) o perfil de expressão foi evidentemente mais dicotômico para TNN e GJD2, o que se aproximou dos achados nas análises de treinamento (Figura 6. B). Assim, quando analisamos a eficiência classificatória de nossa assinatura na comparação “doente” *versus* “saudável” e de forma separada identificamos para Nefropatia Diabética uma classificação parcial entre amostras rotuladas como Rim doente e Rim saudável, com sensibilidade de 36%, especificidade de 54%, precisão de 44% e FDR de 55%. O mesmo resultado foi encontrado para Injúria Renal Aguda na mesma comparação com sensibilidade de 82%, especificidade de 26%, precisão de 50% e FDR de 50%. Porém, de forma interessante para Doença Renal Crônica, a classificação foi 100% eficiente para amostras rotuladas como

Rim saudável e parcial para amostras nomeadas de Rim doentes, com sensibilidade de 100%, especificidade de 32%, precisão de 67% e FDR de 32% (Figura 6. C).

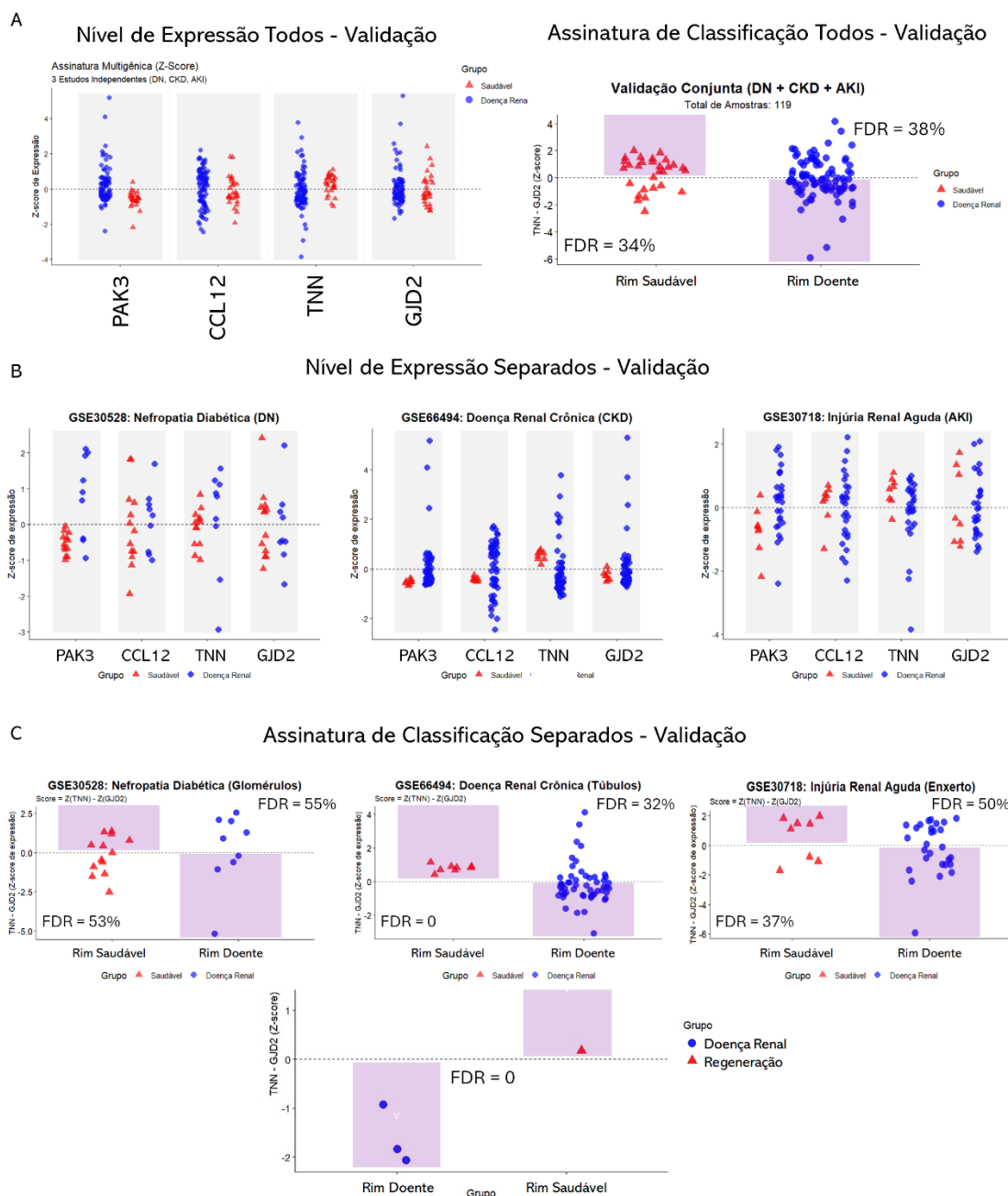


Figura 6. Validação da assinatura renal de classificação. Análise comparativa dos padrões de expressão nos datasets de validação com e sem o agrupamento dos dados. (A) Comparação do Z-score de expressão dos 4 genes selecionados (Pak3, Ccl12, Tnn, Gjd2) nas análises de treino no banco de validação, de maneira agrupada. Denotando somente uma dicotomia, porém não tão clara entre TNN e GJD2. Na diferença dos Z-scores (assinatura renal), todavia, houve classificação parcial entre os grupos saudável e doente, (B) Comparação entre Z-scores dos 4 genes selecionados (Pak3, Ccl12, Tnn, Gjd2) nas análises de treino, de forma separada entre os 3 bancos de validação (GSE30528-ND, GSE66494-DRC, GSE30718-IRA), e (C) Diferença entre Z-scores dos genes da assinatura ($Z(\text{TNN}) - Z(\text{GJD2})$) em cada patologia renal separada.

6. Discussão

De acordo com o Ministério da Saúde, a doença renal crônica (DRC) consiste em: “um termo geral para alterações heterogêneas que afetam tanto a estrutura quanto a função renal, com múltiplas causas e múltiplos fatores de risco.” Segundo Francis et al. (2024), A DRC afeta principalmente populações carentes, marginalizadas e vulneráveis. Os autores revelam também uma relação entre mudanças climáticas e dinâmicas populações e a doença. Atualmente, estima-se que existam 700.000.000 de casos de DRC no mundo. No Brasil, em 2023, houve 140.648 internações pela enfermidade segundo boletim epidemiológico Vol.55 Nº12 do Ministério da Saúde (Brasil, 2024). Além desses dados preocupantes, Francis e colaboradores também mostraram que é possível e provável que haja uma subestimação nesses números, dada a falta global de programas de detecção e *screening* de doenças nefrológicas.

Devido a essa grande necessidade de um diagnóstico precoce, além de ações de prevenção (especialmente em grupos de risco), o boletim epidemiológico do Ministério da Saúde mostra que parte significativa dos custos relacionados à patologia são advindos da diálise peritoneal e hemodiálise. Adicionalmente, também se levanta a potencial perda de produtividade e qualidade de vida geral desses pacientes. Segundo os dados do Sistema de Internação Hospitalar (SIH), no Brasil, de 2022 a 2023 foram gastos R\$ 758.778.769,71 no tratamento das DRCs. Finalmente, a mortalidade decorrente de enfermidades renais foi crescente na última década, tendo 8.429 óbitos somente em 2022.

Considerando sua importância em diversas patologias renais, a descoberta de marcadores moleculares, sendo eles advindos de uma investigação decorrente de múltiplas camadas de regulação biológica (como por exemplo, genética, transcricional, proteica), esta, carrega um potencial inestimável, especialmente relevante no contexto da multiômica integrada. Diversos estudos mostram a importância dos marcadores moleculares em estratégias de diagnóstico e estadiamento de patologias nefrológicas.

Woroniecka *et al.* (2011) identificou milhares de genes alterados na nefropatia diabética, cujas alterações mostraram especificidade tecidual entre glomérulos e túbulos, revelando vias críticas para a patologia. Ju *et al.* (2015) usou dados transcriptômicos para buscar biomarcadores para DRC, identificando o Fator de Crescimento Epidermal (EGF) como um promissor candidato que se correlacionou inversamente com a gravidade da doença. Ju *et al.* (2013) anteriormente também desenvolveu um método para definir a especificidade do tipo celular em doenças, mostrando que processos biológicos específicos de células tubulares renais são fundamentais na nefropatia diabética (ND). Flechner *et al.*

analisou perfis de expressão genica em biopsias, o que lhe permitiu inferir assinaturas moleculares para determinar uma possível rejeição de transplante renal e lesão tecidual. Hodgin *et al.* (2013) identificou assinaturas transcricionais específicas para compartimentos renais em diferentes estágios da ND.

Dessa forma, a capacidade de caracterizar corretamente células e tecidos ou seus estados fisiopatológicos através de marcadores moleculares, também é de suma importância para identificação e rastreamento de danos nas estruturas dos néfrons. Também é possível caracterizar células com base em assinaturas simples de perfil/expressão transcricional ou epigenético, como reportado por Almeida *et al.* em um estudo pioneiro que demonstrou através de uma ferramenta precisa e simples uma assinatura molecular para classificar a origem tecidual de células-tronco mesênquimais (CTMs).

Nesse sentido o nosso trabalho segue a mesma linha de raciocínio onde, através de análises computacionais robustas de perfis transcricionais em bancos de dados e RNAseq, identificamos potenciais genes alvos que conferiram em uma assinatura molecular simples e rápida de classificação renal em um estado saudável e patológico.

Dentre os genes encontrados destacamos a Tenascina-N (TNN), ou Tenascina-W, localizada no cromossomo 1 braço longo é considerada uma glicoproteína da matriz extracelular com função integrativa e estrutural celular, permitindo a ligação à integrina de membrana com a matriz (Figura Suplementar 1. A). A TNN também está envolvida na regulação positiva da angiogênese por brotamento; regulação da adesão celular; e regulação da migração celular. Entre suas vias relacionadas estão a via de sinalização PI3K-Akt e a trimerização da cadeia de colágeno. Anotações de Gene Ontology (GO) relacionadas a este gene incluem ligação a proteínas de ligação à integrina (*Gene Cards*, acesso novembro/2025).

Estudos sobre regeneração renal enfatizam que o rim adulto possui capacidade regenerativa limitada, mas significativa, especialmente em células tubulares proximais. A super-expressão do gene *Tnn* em contextos regenerativos pode indicar seu papel como mediador de processos reparativos, possivelmente através da modulação da matriz extracelular ou sinalização celular (28). O que faz todo sentido frente aos nossos dados onde identificamos a *Tnn* altamente expressas em tecidos renais tratados com CTMs e em rins saudáveis.

A TNN é relatada participando da homeostase óssea e na progressão tumoral. Neidhardt *et al.* (2003) caracterizaram a Tenascina-N como um novo membro da família das tenascinas que mediava a repulsão de neuritos (inibição do crescimento neuronal) em explantes de hipocampo. O artigo de Degen *et al.* é uma importante revisão que resume essa descoberta, evolução e funções da Tenascina-N, destacando seu papel na modulação da resposta imune, progressão de câncer e regeneração de tecidos. De forma adicional,

Chiovaro, Chiquet-Ehrismann e Chiquet (2015) fornecem uma revisão profunda sobre os mecanismos transcricionais que controlam a expressão dos genes da família das tenascinas em diversas condições. Os autores destacam que a expressão das tenascinas é rigorosamente controlada por sinais específicos durante o desenvolvimento, especificamente em processos de reparo tecidual e no microambiente tumoral. A regulação do seu gene é rigorosamente controlada, possuindo uma caixa TATA em seu promotor, uma característica associada a genes que são ativados sob estímulos precisos. A TNN atua como uma proteína de modulação, não promove adesão celular de forma robusta como a fibronectina, e em vez disso, frequentemente exerce efeitos anti-adesivos e modula as respostas celulares a fatores de crescimento, influenciando diretamente a migração, proliferação e diferenciação celular.

Estudos com tenascina em âmbito renal restringe-se a Tenascina-C, a qual está intimamente associada a processos de lesão e fibrose renal (29-31). Em contraste, a TNN já foi encontrada em bancos de dados especializados como o Nephroseq como altamente expressa em tecidos renais saudáveis e apresenta correlação negativa com índices de creatinina sérica (em hipertensão e nefroesclerose) e proteinúria (na GESF), e ainda demonstra correlação positiva com o aumento da taxa de filtração glomerular em pacientes DRCs, aumento ainda mais o impacto de nosso achado (*nephroseq.org*) (Figura Suplementar 2).

No contexto do presente estudo, acredita-se que claramente ela esteja influenciando a regeneração renal. Em investigação paralela, ao analisar a expressão em Log2FC da TNN, em um banco de dados “irmão” de um dos bancos de validação, onde ao invés de amostras glomerulares, se utilizou de amostras tubulares, assim percebeu-se uma possibilidade de especificidade maior de expressão de TNN para tais estruturas.

Outro gene que se destacou-se em nossos estudos foi o GJD2 (Gap Junction Protein Delta 2) que se localiza no cromossomo 15 braço longo e codifica a proteína conexina 36 (Cx36), um membro da família de proteínas conexinas (Figura Suplementar 1. B). As conexinas são proteínas de junção comunicante que se organizam em grupos de 6 ao redor de um poro central para formar um *conexon* ou gap junctions, um componente do canal intercelular da junção comunicante, permitindo a troca de moléculas catiônicas entre as células humanas. Entre suas vias relacionadas estão o tráfego de junções comunicantes e o transporte mediado por vesículas. As anotações da Gene Ontology (GO) relacionadas a este gene incluem atividade de canal de junção comunicante (*Gene Cards*, acesso novembro/2025).

Van der Sande et al. (2022) investigaram o papel do gene GJD2 no desenvolvimento de distúrbios de refração ótica ocular, concluindo que esse gene é um fator de risco significativo para a miopia. Sendo essencial para a transmissão de sinais visuais na retina, modulando a relação sinal-ruído e contribuindo para a visão noturna. No cérebro, regula a sincronização de redes neuronais. Sua expressão é enriquecida no cérebro (hipocampo,

bulbo olfatório), glândula adrenal, pâncreas, hipófise, e de forma crucial, na retina. Quint et al. (2021) forneceram evidências causais ao demonstrar que a perda dos ortólogos do GJD2 no peixe-zebra resulta diretamente em erros de refração. Kunceviciene et al. (2022) analisaram polimorfismos no gene GJD2 em uma população humana, reforçando sua associação com a susceptibilidade a erros refrativos como miopia e hipermetropia. Yang et al. (2019) observaram uma diminuição significativa na expressão do RNA mensageiro do gene GJD2 e da sua proteína, a Conexina 36, em um modelo de miopia por privação de forma (induzindo a condição) em cobaias.

Porém evidências renais da expressão de GJD2 em rins e suas associações com patologias são praticamente escassas. Os estudos de transcriptômica e de expressão gênica em populações com doença renal não identificam GJD2 como um gene diferencialmente expresso ou relevante para processos de regeneração ou patologia renal. A expressão aumentada de conexinas em rins doentes está relacionada principalmente à conexina 43 (Gja1), e não à conexina 36 (Gjd2) (32). Investigando novamente os dados do Nephroseq é possível identificar a expressão elevada de GJD2 na DRC, com correlação positiva com BUN (ureia), índice de IFTA (fibrose intersticial e atrofia tubular) e aumento de pressão sanguínea em nefrite lúpica, IFTA, e DMB (doença de membrana basal fina), respectivamente (*nephroseq.org*) (Figura Suplementar 3).

Em sequência, a análise dos dados estatísticos de nossa assinatura revela um modelo que, embora tenha atingido a perfeição teórica no Treinamento (Sensibilidade, Especificidade e Precisão de 100%). No ensaio de validação com o conjunto de dados agregado, a nossa assinatura (z-score Tnn–Gjd2), na comparação ‘doente’ versus ‘saúdável’, apresentou sensibilidade de 84,6% (identifica a maioria dos rins doentes), especificidade de 35,1% (alta taxa de falsos positivos, isto é, 64,9% dos rins saudáveis classificados como doentes) e precisão (PPV) de 61% (entre os classificados como ‘doente’, 61% eram de fato doentes).

Ao investigarmos a força de nossa assinatura em classificar diferentes etiologias patológicas renais, a assinatura apresentou seu melhor desempenho na doença renal crônica (DRC), com sensibilidade de 100% (sem falsos negativos no conjunto avaliado) e especificidade de 32% (alta taxa de falsos positivos). Na injúria renal aguda (IRA), observou-se sensibilidade de 82,35% e especificidade de 26,31%, reproduzindo o padrão de boa detecção de doentes, porém baixa discriminação de saudáveis. Em contraste, na nefropatia diabética (ND) o desempenho foi inferior, com sensibilidade de 36,36% (elevada proporção de falsos negativos) e especificidade de 54,54%.

Em suma, acreditamos que a nossa assinatura possui um grande potencial, porém a mesma necessita de uma melhor base para treino e uma melhor ensaio de validação. No entanto, isso não diminui neste trabalho o achado de 2 genes importantes (TNN e GJD2)

que possuem ausências de evidências funcionais em condições de regeneração e lesão renal. Assim, a identificação de genes diferencialmente expressos em contextos de regeneração renal versus patologia representa um avanço significativo na compreensão dos mecanismos moleculares subjacentes à homeostase renal. Os achados apresentados aqui, com super-expressão do gene TNN em contextos de regeneração e saúde renal, e alta expressão do gene GDJ2 em rim patológico, sugerem papéis funcionais distintos desses marcadores moleculares. Por fim, consideramos que os mecanismos moleculares de regulação renal envolvem complexas redes de sinalização que controlam a proliferação, diferenciação e manutenção da homeostase tecidual.

7. Conclusão

Conseguimos padronizar e executar uma análise de dados complexa de expressão gênica diferencial a partir de dados brutos de RNAseq utilizando um número variado de ferramentas computacionais

Identificamos e validamos uma assinatura simples e rápida (z-score Tnn-Gjd2) de classificação de estado renal regenerativo ou patológico em bancos de dados de DRC, IRA e ND, porém ainda com ajustes a serem realizados.

E por fim, o nosso estudo identificou de forma intrigante a expressão de dois genes, a Tenascina-N e a Conexina 36, o qual ainda não tinham sido descritos e de forma dicotômica em rins saudáveis e em rins doentes com potencial de serem considerados biomarcadores promissores de prognóstico e estadiamento em doenças renais crônicas.

8. Referências Bibliográficas

1. PAREKKADAN, B.; MILWID, J. M. Mesenchymal stem cells as therapeutics. *Annual Review of Biomedical Engineering*, [S. l.], v. 12, p. 87-117, 2010. DOI:<https://doi.org/10.1146/annurev-bioeng-070909-105309>.
2. WANG, Y.; WANG, H.; TAN, J. et al. Therapeutic effect of mesenchymal stem cells and their derived exosomes in diseases. *Molecular Biomedicine*, [S. l.], v. 6, artigo 34, 2025. DOI:<https://doi.org/10.1186/s43556-025-00277-4>.
3. RENESME, L. et al. Delphi-driven consensus definition for mesenchymal stromal cells and clinical reporting guidelines for mesenchymal stromal cell-based therapeutics. *Cytotherapy*, [S. l.], v. 27, n. 2, p. 146-168, fev. 2025. DOI:<https://doi.org/10.1016/j.jcyt.2024.10.008>.
4. TURGUT, F.; AWAD, A. S.; ABDEL-RAHMAN, E. M. Acute Kidney Injury: Medical Causes and Pathogenesis. *Journal of Clinical Medicine*, [S. l.], v. 12, n. 1, artigo 375, 2023. DOI:<https://doi.org/10.3390/jcm12010375>.
5. RONCO, C.; BELLOMO, R.; KELLUM, J. A. Acute kidney injury. *The Lancet*, [S. l.], v. 394, n. 10212, p. 1949-1964, 2019. DOI:[https://doi.org/10.1016/S0140-6736\(19\)32563-2](https://doi.org/10.1016/S0140-6736(19)32563-2).
6. AMMIRATI, Adriano Luiz. Chronic Kidney Disease. *Revista da Associação Médica Brasileira*, São Paulo, v. 66, Suppl. 1, p. s03-s09, 2020. DOI:<https://doi.org/10.1590/1806-9282.66.S1.3>.
7. WEBSTER, Angela C. et al. Chronic kidney disease. *The lancet*, v. 389, n. 10075, p. 1238-1252, 2017. DOI:[https://doi.org/10.1016/S0140-6736\(16\)32064-5](https://doi.org/10.1016/S0140-6736(16)32064-5).
8. NERBASS, FABIANA B.; LIMA, HELBERT DO NASCIMENTO; STROGOFF-DE-MATOS, JORGE P.; ZAWADZKI, BRUNO; MOURA-NETO, JOSÉ A.; LUGON, JOCEMIR RONALDO; SESSO, RICARDO. Brazilian Dialysis Survey 2023. *Braz. J. Nephrol.*, v. 47, n. 1, e20240081, Jan. 2025. DOI:<https://doi.org/10.1590/2175-8239-JBN-2024-0081pt>.
9. CHEN, F.; CHEN, N.; XIA, C. et al. Mesenchymal Stem Cell Therapy in Kidney Diseases: Potential and Challenges. *Cell Transplantation*, [S. l.], v. 32, 2023. [online]. DOI:<https://doi.org/10.1177/09636897231164251>.
10. HEJAZIAN, S. M.; HEJAZIAN, S. S.; MOSTAFAVI, S. M. et al. Targeting cellular senescence in kidney diseases and aging: A focus on mesenchymal stem cells and their paracrine factors. *Cell Communication and Signaling*, [S. l.], v. 22, artigo 609, 2024. DOI:<https://doi.org/10.1186/s12964-024-01968-1>.
11. AHMAD, A.; IMRAN, M.; AHSAN, H. Biomarkers as Biomedical Bioindicators: Approaches and Techniques for the Detection, Analysis, and Validation of Novel Biomarkers of Diseases. *Pharmaceutics*, [S. l.], v. 15, n. 6, artigo 1630, 2023. DOI:<https://doi.org/10.3390/pharmaceutics15061630>.
12. LIU, R.; WANG, X.; AIHARA, K. et al. Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers. *Medicinal Research Reviews*, [S. l.], v. 34, p. 455-478, 2014. DOI:<https://doi.org/10.1002/mrd.21293>.
13. KUÇI, S.; KUÇI, Z.; SCHÄFER, R. et al. Molecular signature of human bone marrow-derived mesenchymal stromal cell subsets. *Scientific Reports*, [S. l.], v. 9, artigo 1774, 2019. DOI:<https://doi.org/10.1038/s41598-019-38517-7>.

14. DA VEIGA, G. L. et al. Kidney Diseases: The Age of Molecular Markers. In: SANTOS, H. M. B. C. (ed.). Translational Urinomics. Cham: Springer, 2021. p. [xx-xx]. (Advances in Experimental Medicine and Biology, v. 1306). DOI:https://doi.org/10.1007/978-3-030-63908-2_2.
15. LOVE M. I.; HUBER W.; ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. DOI:<https://doi.org/10.1186/s13059-014-0550-8>.
16. FRANCIS, Anna et al. Chronic kidney disease and the global public health agenda: an international consensus. Nature Reviews Nephrology, [S. l.], v. 20, n. 7, p. 473–485, jul. 2024. DOI: 10.1038/s41581-024-00820-6.
17. WORONIECKA, K. I. et al. Transcriptome analysis of human diabetic kidney disease. Diabetes, v. 60, n. 9, p. 2354-2369, set. 2011. DOI: <https://doi.org/10.2337/db10-1181>.
18. JU, W. et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. Science Translational Medicine, v. 7, n. 316, p. 316ra193, 2 dez. 2015. DOI: <https://doi.org/10.1126/scitranslmed.aac7071>.
19. JU, W. et al. Defining cell-type specificity at the transcriptional level in human disease. Genome Research, v. 23, n. 11, p. 1862-1873, nov. 2013. DOI: <https://doi.org/10.1101/gr.155697.113>.
20. FLECHNER, S. M. et al. Kidney transplant rejection and tissue injury by gene profiling of biopsies and peripheral blood lymphocytes. American Journal of Transplantation, v. 4, n. 9, p. 1475-1489, set. 2004. DOI: <https://doi.org/10.1111/j.1600-6143.2004.00526.x>.
21. HODGIN, J. B. et al. Transcriptional analysis of diabetic nephropathy in human subjects. Journal of the American Society of Nephrology, v. 24, n. 8, p. 1235-1244, ago. 2013. DOI: <https://doi.org/10.1681/ASN.2012080843>.
22. DEGEN, Martin; SCHERBERICH, Arnaud; TUCKER, Richard P. Tenascin-W: Discovery, Evolution, and Future Prospects. Frontiers in Immunology, v. 11, 2021. DOI: <https://doi.org/10.3389/fimmu.2020.623305>.
23. NEIDHARDT, John et al. Tenascin-N: characterization of a novel member of the tenascin family that mediates neurite repulsion from hippocampal explants. Molecular and Cellular Neuroscience, v. 23, n. 2, p. 193-209, 2003. DOI: [https://doi.org/10.1016/S1044-7431\(03\)00012-5](https://doi.org/10.1016/S1044-7431(03)00012-5).
24. VAN DER SANDE, E. et al. The Role of GJD2(Cx36) in Refractive Error Development. Investigative Ophthalmology & Visual Science, v. 63, n. 3, p. 5, 2 mar. 2022. DOI: <https://doi.org/10.1167/iovs.63.3.5>.
25. QUINT, W. H. et al. Loss of Gap Junction Delta-2 (GJD2) gene orthologs leads to refractive error in zebrafish. Communications Biology, v. 4, n. 1, p. 676, 3 jun. 2021. DOI: <https://doi.org/10.1038/s42003-021-02185-z>.
26. KUNCEVICIENE, E. et al. Association of CX36 Protein

- Encoding Gene GJD2 with Refractive Errors. *Genes*, v. 13, n. 7, p. 1166, 2022. DOI: <https://doi.org/10.3390/genes13071166>.
27. YANG, G. Y. et al. Decreased expression of gap junction delta-2 (GJD2) messenger RNA and connexin 36 protein in form-deprivation myopia of guinea pigs. *Chinese Medical Journal*, v. 132, n. 14, p. 1700–1705, 20 jul. 2019. DOI: <https://doi.org/10.1097/CM9.0000000000000319>.
 28. Roye, Y., Bhattacharyya, S., & Gruenwald, A. (2021). Molecular mechanisms of renal progenitor regulation. *Cells*, 10(1), 59. <https://doi.org/10.3390/cells10010059>
 29. Zhu H, Liao J, Zhou X, et al. Tenascin-C Promotes Acute Kidney Injury to Chronic Kidney Disease Progression by Impairing Tubular Integrity via $\text{Av}\beta 6$ Integrin Signaling. *Kidney International*. 2020;97(5):1017-1031. [doi:10.1016/j.kint.2020.01.026](https://doi.org/10.1016/j.kint.2020.01.026).
 30. Fu H, Tian Y, Zhou L, et al. Tenascin-C Is a Major Component of the Fibrogenic Niche in Kidney Fibrosis. *Journal of the American Society of Nephrology : JASN*. 2017;28(3):785-801. [doi:10.1681/ASN.2016020165](https://doi.org/10.1681/ASN.2016020165).
 31. Xie Q, Zhang M, Mao X, et al. Matrix Protein Tenascin-C Promotes Kidney Fibrosis via STAT3 Activation in Response to Tubular Injury. *Cell Death & Disease*. 2022;13(12):1044. [doi:10.1038/s41419-022-05496-z](https://doi.org/10.1038/s41419-022-05496-z).
 32. Roger E, Boutin L, Chadji-christos CE. The Role of Connexin 43 in Renal Disease: Insights From in Vivo Models of Experimental Nephropathy. *International Journal of Molecular Sciences*. 2022;23(21):13090. [doi:10.3390/ijms232113090](https://doi.org/10.3390/ijms232113090).

9. Material Suplementar

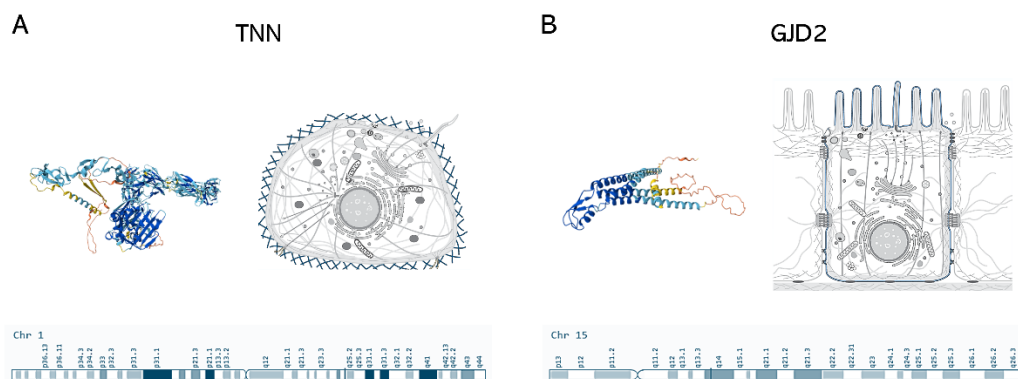


Figura Suplementar 1. Estrutura e localização de TNN e Gjd2. (A) o gene TNN localiza-se no cromossomo 1 braço longo, possui uma localização extracelular de interação com a matriz e uma estrutura tridimensional em “T” com alfa-hélices e dobras, e (B) o gene GJD2 localiza-se no cromossomo 15 braço longo, possui uma localização na membrana apical, e basolateral celular de interação célula-célula e uma estrutura tridimensional em “V” predominando alfa-hélices e uma dobra central.

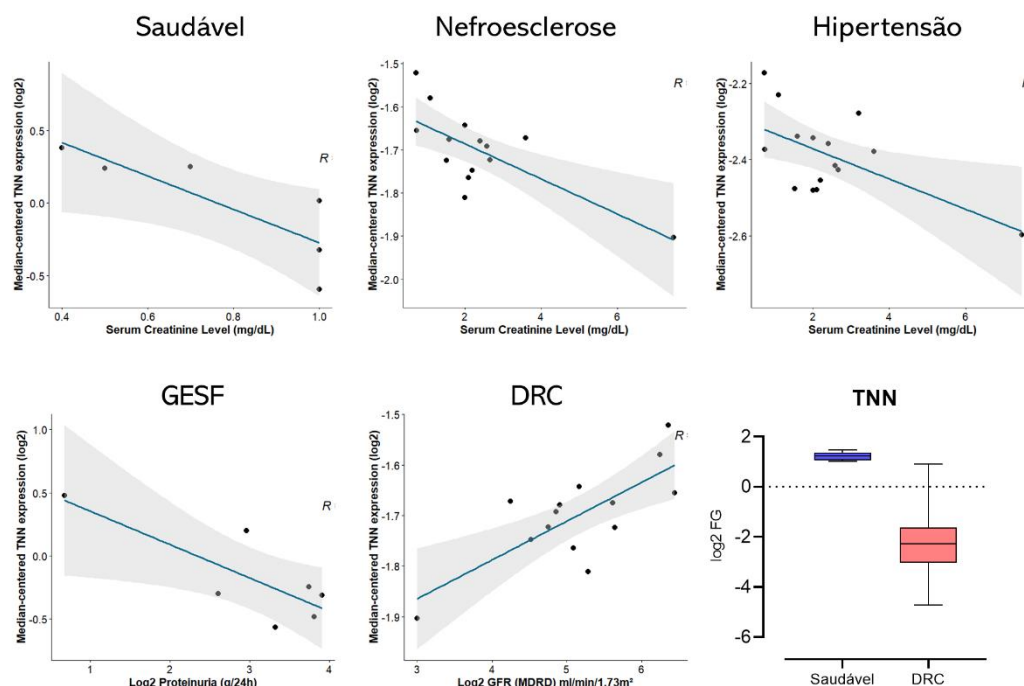


Figura Suplementar 2. Expressão de TNN em pacientes saudáveis e em DRC e correlações funcionais. TNN possui alta expressão em tecidos renais saudáveis e apresenta correlação negativa com índices de creatinina sérica (em hipertensão e nefroesclerose) e proteinúria (na GESF), e ainda demonstra correlação positiva com o aumento da taxa de filtração glomerular em pacientes DRCs (nephroseq.or).

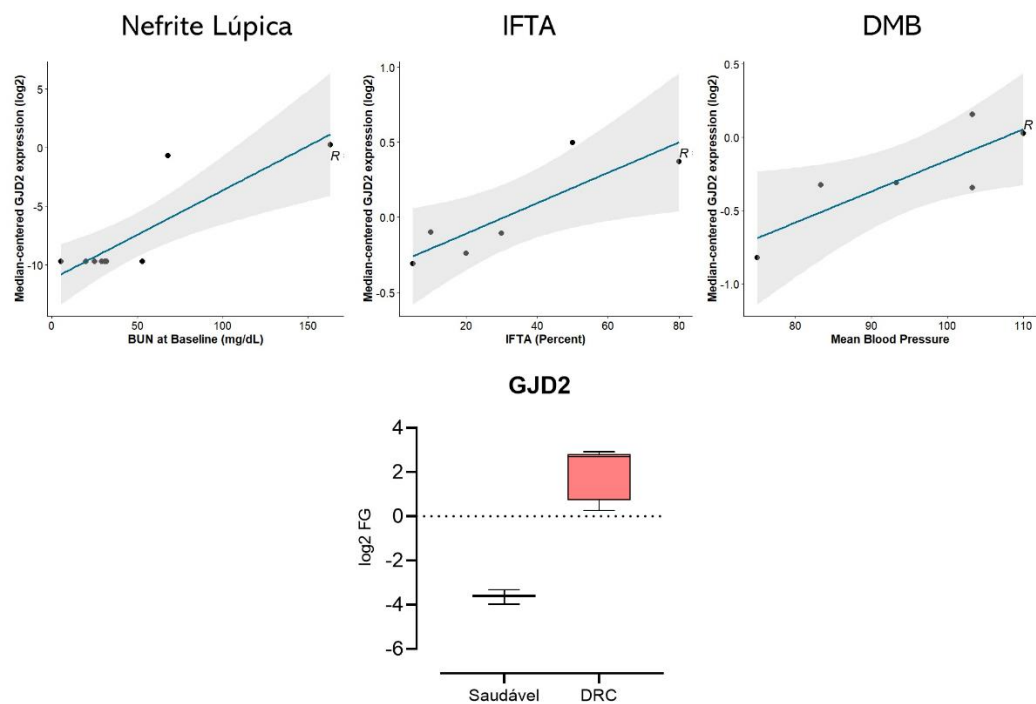


Figura Suplementar 3. Expressão de GJD2 em pacientes saudáveis e em DRC e correlações funcionais. expressão elevada de GJD2 na DRC, com correlação positiva com BUN (ureia), índice de IFTA (fibrose intersticial e atrofia tubular) e aumento de pressão sanguínea em nefrite lúpica, IFTA, e DMB (doença de membrana basal fina), respectivamente (nephroseq.or).

PARECER FINAL DO TRABALHO DE CONCLUSÃO DE CURSO

Discente: DAVI DE MATTOS RIBEIRO LIMA

Título: "DETERMINAÇÃO DE UMA ASSINATURA MOLECULAR RENAL DE REGENERAÇÃO APÓS TERAPÊUTICA COM CÉLULAS ESTROMAIS MESENQUIMAIS MULTIPOTENTES"

Orientador: Prof. Dr. Roberto Fioravanti Carelli Fontes

Curso/Habilitação: Bacharelado em Ciências Biológicas/Biologia Marinha

COMISSÃO EXAMINADORA	CONCEITO
Prof. Dr. Roberto Fioravanti Carelli Fontes	APROVADO
Prof. Dr. Danilo Candido	APROVADO
MSC. Gabriel Pereira	APROVADO

PARECER:

O discente demonstrou excelente domínio do tema, utilizando adequadamente os recursos disponíveis e expondo o seu trabalho com clareza e organização. Respondeu com precisão a todas as questões da banca e comprometeu-se a realizar os pequenos ajustes solicitados no texto final. Pela qualidade da apresentação e maturidade acadêmica demonstrada, recomendamos sua APROVAÇÃO.

CONCEITO FINAL:

A Comissão Examinadora abaixo assinada conclui que o discente **Davi de Mattos Ribeiro Lima** obteve o seguinte conceito:

☒ APROVADO

☐ REPROVADO

São Vicente, 03 de dezembro de 2025.

Documento assinado digitalmente
gov.br ROBERTO FIORAVANTI CARELLI FONTES
Data: 03/12/2025 13:28:11-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Roberto Fioravanti Carelli Fontes

Documento assinado digitalmente
gov.br DANILO CANDIDO DE ALMEIDA
Data: 03/12/2025 15:06:07-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Danilo Candido

Documento assinado digitalmente
gov.br GABRIEL PEREIRA
Data: 03/12/2025 14:55:42-0300
Verifique em <https://validar.iti.gov.br>

MSC. Gabriel Pereira