

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS:
BIOQUÍMICA

Guilherme Povala

**APLICAÇÃO DE ESTRATÉGIAS NEUROCOMPUTACIONAIS PARA A
IDENTIFICAÇÃO DE ALTERAÇÕES FUNCIONAIS NA DOENÇA DE ALZHEIMER**

Porto Alegre

2022

Guilherme Povala

**APLICAÇÃO DE ESTRATÉGIAS NEUROCOMPUTACIONAIS PARA A
IDENTIFICAÇÃO DE ALTERAÇÕES FUNCIONAIS NA DOENÇA DE ALZHEIMER**

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas: Bioquímica do Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de doutor em Bioquímica.

Orientador: Prof. Dr. Eduardo Rigon Zimmer

Coorientador: Prof. Dr. Bruno Zatt

Porto Alegre

2022

CIP - Catalogação na Publicação

Povala, Guilherme

Aplicação de estratégias neurocomputacionais para a identificação de alterações funcionais na doença de alzheimer / Guilherme Povala. -- 2022.

110 f.

Orientador: Eduardo Rigon Zimmer.

Coorientador: Bruno Zatt.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Porto Alegre, BR-RS, 2022.

1. Doença de Alzheimer. 2. Biomarcadores. 3. Aprendizado de Máquina. I. Rigon Zimmer, Eduardo, orient. II. Zatt, Bruno, coorient. III. Título.

SUMÁRIO

PARTE I	1
RESUMO	2
RESUMO EM INGLÊS (ABSTRACT)	3
LISTA DE ABREVIACÕES	4
INTRODUÇÃO	6
A Doença de Alzheimer	7
O Sistema AT(N)	8
Biomarcadores de imagem	9
Tomografia por emissão de pósitrons (PET)	10
Biomarcadores de fluído	11
Líquido cefalorraquidiano (LCR)	11
Sangue	12
Técnicas neurocomputacionais na Doença de Alzheimer	12
Biologia de Sistemas	14
Justificativa	14
OBJETIVOS	16
Objetivo principal	16
Objetivos específicos	16
PARTE II	17
Capítulo 1	18
<i>Soluble amyloid-beta isoforms predict downstream Alzheimer's disease pathology</i>	18
Capítulo 2	58
<i>Serine/threonine kinase activity regulates brain glucose metabolism in Alzheimer's Disease</i>	58
PARTE III	89
DISCUSSÃO	90
CONCLUSÃO	100
PERSPECTIVAS	101
REFERÊNCIAS	102

PARTE I

Na Parte I, são descritos o resumo, a lista de abreviaturas, a introdução e os objetivos desta tese de doutorado.

RESUMO

A deposição cerebral de placas de beta-amiloide ($A\beta$) e emaranhados neurofibrilares de tau hiperfosforilada são características patológicas da doença de Alzheimer (DA) que acabam por levar ao declínio cognitivo e, conseqüentemente, à neurodegeneração. Recentemente, foi proposto pelo *National Institute on Aging and Alzheimer's Association* um critério de pesquisa clínica onde imagens [ressonância nuclear magnética estrutural (RM) e tomografia por emissão de pósitrons (PET)] e medidas baseadas no líquido cefalorraquidiano (LCR) são utilizadas para classificar o paciente no sistema A, T e (N), que indica respectivamente positividade para amiloide, tau e neurodegeneração. A imagem PET pode ser usada para identificar e quantificar os agregados de $A\beta$ e tau, enquanto a redução dos níveis de $A\beta_{1-42}/A\beta_{1-40}$ e aumento de tau [tanto tau total (t-tau) quanto tau fosforilada (p-tau)] podem ser medidos no LCR. No entanto, se os níveis de $A\beta$ podem prever características patológicas da DA em indivíduos cognitivamente saudáveis (CU) permanece incerto. Acredita-se que esses biomarcadores sigam uma cascata temporal onde as anormalidades seguem o modelo a seguir: $A \rightarrow T \rightarrow N$. Recentes avanços no campo da neuroinformática, possibilitaram a expansão de estratégias analíticas e computacionais para estudar dados multimodais na DA. A integração de imagens PET e dados ômicos pode fornecer novos *insights* sobre a fisiopatologia da DA. Assim, nesta tese, primeiro investigamos o papel das isoformas solúveis de $A\beta$ na predição de T e (N), usando modelos de aprendizado de máquina (ML). Em segundo lugar, usamos a biologia de sistemas para ajudar a elucidar os fundamentos biológicos e destacar novos biomarcadores potenciais da fisiopatologia da DA usando um método de integração de imagens de PET com fluorodeoxiglicose (FDG-PET), um marcador de atividade sináptica, e dados de transcriptômica no sangue. Em resumo, mostramos que as isoformas de $A\beta$, especialmente as espécies menores, podem prever T e (N) com alta acurácia em indivíduos CU. Além disso, a integração de imagens mostrou uma forte associação entre um *cluster* relacionado à regulação da atividade da proteína serina/treonina cinase e a unidade reguladora ZNF653 com o sinal do PET-FDG cerebral. Para finalizar, mostramos o potencial das assinaturas gênicas associadas ao metabolismo cerebral como novos biomarcadores de DA e que a combinação de metodologias utilizadas neste trabalho pode ajudar a entender melhor a heterogeneidade da patologia da DA.

RESUMO EM INGLÊS (ABSTRACT)

Brain deposition of beta-amyloid ($A\beta$) plaques and neurofibrillary tangles of hyperphosphorylated tau are pathological features of Alzheimer's disease (AD) that ultimately lead to cognitive decline and, consequently, to neurodegeneration. Recently, a clinical research criterion was proposed by the National Institute on Aging and Alzheimer's Association where imaging [structural magnetic resonance imaging (MRI) and positron emission tomography (PET)] and measurements based on cerebrospinal fluid (CSF) are used to classify the patient in the A, T and (N) system, which respectively indicate positivity for amyloid, tau and neurodegeneration. PET imaging can be used to identify and quantify $A\beta$ and tau aggregates, while reducing $A\beta_{1-42}/A\beta_{1-40}$ levels and increasing tau [both total tau (t-tau) and phosphorylated tau (p-tau)] can be measured in the CSF. However, whether $A\beta$ levels can predict pathological features of AD in cognitively healthy individuals (CU) remains unclear. These biomarkers are believed to follow a temporal cascade where the abnormalities follow the following pattern: $A \rightarrow T \rightarrow N$. Recent advances in the field of neuroinformatics have enabled the expansion of analytical and computational strategies to study multimodal data in AD. The integration of PET images and omics data may provide new insights into the pathophysiology of AD. Thus, in this thesis, we first investigate the role of soluble $A\beta$ isoforms in the prediction of T and (N), using machine learning (ML) models. Second, we use systems biology to help elucidate the biological underpinnings and highlight potential new biomarkers of AD pathophysiology using a method of integrating PET images with fluorodeoxyglucose (FDG-PET), a marker of synaptic activity, and transcriptomics data in blood. In summary, we show that $A\beta$ isoforms, especially the smaller species, can predict T and (N) with high accuracy in CU individuals. Furthermore, image integration showed a strong association between a cluster related to the regulation of protein serine/threonine kinase activity and the regulatory unit ZNF653 with the brain PET-FDG signal. Finally, we show the potential of gene signatures associated with brain metabolism as new biomarkers of AD and that the combination of methodologies used in this work can help to better understand the heterogeneity of AD pathology.

LISTA DE ABREVIACÕES

A - Amiloide

A+ - Positividade de A β

APP - Proteína precursora de amiloide

A β - Beta-amiloide

BACE - Enzima de clivagem beta-amiloide

DA - Doença de Alzheimer

DP - Doença de Parkinson

ELA - Esclerose Lateral Amiotrófica

FDG - [^{18}F]-fluorodesoxiglicose

GFAP - Proteína glial fibrilar ácida

IP-MS - Imunoprecipitação acoplada à espectrometria de massa

LCR - Líquido Cefalorraquidiano

ML - Aprendizado de Máquina

N - Neurodegeneração

(N)+ - Positividade de neurodegeneração

NFT - Emaranhados neurofibrilares

NIA-AA - *National Institute on Aging-Alzheimer's Association*

NfL - Neurofilamento de cadeia leve

PET - Tomografia por emissão de pósitrons

RM - Ressonância Magnética

SNC - Sistema Nervoso Central

SPECT - Tomografia computadorizada por emissão de fóton único

T - Tau

T+ - Positividade de tau

TC - Tomografia Computadorizada

fMRI - Ressonância magnética funcional

p-tau - Tau fosforilada

p-tau181 - Tau hiperfosforilada em treonina 181

p-tau217 - Tau hiperfosforilada em treonina 217

p-tau231 - Tau hiperfosforilada em treonina 231

t-tau - Tau total

1. INTRODUÇÃO

O crescimento da população mundial tem sido acompanhado por um aumento progressivo do número de idosos. Em países desenvolvidos, a expectativa de vida já atinge valores acima dos 80 anos. Embora as causas de morte predominantes em pessoas idosas ainda sejam as doenças cardiovasculares e diferentes tipos de câncer, a ocorrência de doenças neurodegenerativas está fortemente relacionada com a idade e estão entre as dez principais doenças que ocasionam óbito (G7, 2017, p. 7).

No entanto, as mortes causadas por doenças cardiovasculares e câncer têm decrescido anualmente (ALZHEIMER'S & DEMENTIA, 2019) o que contribuiu ainda mais para o crescimento no número de cidadãos idosos, provocando um aumento expressivo no número de pessoas afetadas por doenças neurodegenerativas (GOOCH; PRACHT; BORENSTEIN, 2017).

Doenças neurodegenerativas são definidas por condições nas quais há disfunção neuronal progressiva e atrofia cerebral. Essas doenças são condições, até o momento, incuráveis e debilitantes que resultam em degeneração progressiva, o que reflete em disfunções motoras e/ou cognitivas. Temos como doenças neurodegenerativas clássicas, em que a neurodegeneração é a principal marca registrada, a doença de Alzheimer (DA), a doença de Parkinson (DP) e a esclerose lateral amiotrófica (ELA) (DUGGER; DICKSON, 2017; GBD 2017 US NEUROLOGICAL DISORDERS COLLABORATORS et al., 2021).

1.1. A Doença de Alzheimer

A DA é a doença neurodegenerativa mais prevalente no mundo (GBD 2017 US NEUROLOGICAL DISORDERS COLLABORATORS et al., 2021). Suas principais características neuropatológicas envolvem a deposição de duas proteínas, beta-amiloide ($A\beta$) e tau, em agregados insolúveis no cérebro (PERL, 2010; SERRANO-POZO et al., 2011).

A proteína $A\beta$ denota peptídeos de 36 a 43 aminoácidos que são o principal componente das placas amiloides encontradas no cérebro de pessoas com DA. As placas amiloides são depósitos extracelulares da proteína $A\beta$ principalmente na substância cinzenta do cérebro (CRAS et al., 1991; DICKSON, 1997). Os peptídeos de $A\beta$ são produzidos pela clivagem proteolítica da proteína precursora de amiloide (APP) (HAMLEY, 2012). A APP é uma proteína transmembrana e é produzida por muitos tipos de células do corpo, mas é especialmente abundante nos neurônios (HAASS et al., 2012). Para liberação de $A\beta$, a APP é clivada sequencialmente por duas enzimas: pela beta secretase [ou enzima de clivagem beta-amiloide (BACE)] fora da membrana, e pela gama secretase (γ -secretase), um complexo enzimático dentro da membrana celular (HAASS et al., 2012). São as ações sequenciais dessas secretases que fazem com que fragmentos da proteína $A\beta$ sejam liberados no espaço extracelular (HAASS et al., 2012; SUH; CHECLER, 2002). Esses fragmentos geralmente são compostos por peptídeos de $A\beta$ com 40 ou 42 aminoácidos. Porém, vários fragmentos $A\beta$ de outros tamanhos e menos abundantes também são gerados (DUNYS; VALVERDE; CHECLER, 2018; KUMMER; HENEKA, 2014). O comprimento desses peptídeos e modificações químicas aplicadas a eles podem influenciar tanto sua tendência de agregação quanto sua toxicidade (WALKER,

2020). O segundo componente da patologia da DA são os emaranhados neurofibrilares (NFT), derivados do acúmulo anormal da proteína tau. Neurônios saudáveis, em parte, são sustentados internamente por estruturas chamadas microtúbulos, que ajudam a guiar nutrientes e moléculas do corpo celular para o axônio e os dendritos. Em neurônios saudáveis, normalmente a tau se liga e estabiliza os microtúbulos. Na DA, no entanto, excesso de fosforilação em sítios aceptores de fosfato faz com que a tau se desestabilize e se separe dos microtúbulos, se agregando intracelularmente. Esses emaranhados bloqueiam o sistema de transporte neuronal, o que prejudica a comunicação sináptica entre os neurônios, que mais tarde, acabam sofrendo morte celular (CHUN; JOHNSON, 2007; LAKHAN, 2016).

1.2. O Sistema AT(N)

O modelo teórico de DA mais aceito sugere que o dismetabolismo do A β desencadeia uma cascata de eventos patológicos, incluindo a patologia de tau, disfunção sináptica e neurodegeneração, que leva ao declínio cognitivo e, finalmente, à demência. Ou seja, uma cadeia linear que leva à positividade de A β (A+) \rightarrow positividade de tau (T+) \rightarrow positividade de neurodegeneração [(N)+] \rightarrow sintomas cognitivos (HARDY; HIGGINS, 1992; SELKOE, 1991). Essa cascata previamente proposta baseada em estudos em cérebros *post mortem*, mais recentemente passou a poder ser identificada in vivo graças aos avanços no área de biomarcadores. Os biomarcadores são medidas objetivas de um processo biológico ou patogênico que podem ser usados para avaliar o risco ou prognóstico da doença, para orientar o diagnóstico clínico ou para monitorar as intervenções terapêuticas (BLENNOW et al., 2010).

A compreensão da progressão longitudinal desses biomarcadores serviram como base para a definição biológica da DA e constituem a base do *National Institute on Aging-Alzheimer's Association (NIA-AA) Research Framework* proposto para estudos clínicos, que adotou o sistema AT(N) para biomarcadores amiloide (A), tau (T) e neurodegeneração (N). Em cada categoria, os biomarcadores são dicotomizados para indicar um estado normal (negativo para biomarcadores) ou anormal (positivo para biomarcadores) (JACK et al., 2016). Esse modelo teórico baseia-se em dados derivados de estudos multicêntricos transversais e longitudinais usando vários biomarcadores. Atualmente, os biomarcadores de DA são divididos em duas classes principais: medida em biofluidos [sangue e líquido cefalorraquidiano (LCR)] e medida por neuroimagem [ressonância magnética (RM) e tomografia por emissão de pósitrons (PET)] (BLENNOW; ZETTERBERG, 2018).

1.3. Biomarcadores de imagem

Biomarcadores de neuroimagem têm um papel importante tanto no diagnóstico precoce, bem como no acompanhamento periódico da DA (VARGHESE et al., 2013). Podemos classificar as técnicas de neuroimagem em duas grandes categorias: estruturais e funcionais.

As principais técnicas de imagem estrutural são a tomografia computadorizada (TC) e a RM. A TC permite distinguir estruturas diferentes. A RM, devido à sua alta resolução espacial, pode ser usada para distinguir a diferença entre dois tecidos arbitrariamente semelhantes, mas não idênticos (PERRIN; FAGAN; HOLTZMAN, 2009; WOLZ et al., 2011).

No contexto de técnicas funcionais, temos o PET, tomografia computadorizada por emissão de fóton único (SPECT) e ressonância magnética

funcional (fMRI) (WALHOVD et al., 2010). Embora a técnica de imagem funcional forneça algumas informações estruturais, sua resolução espacial é menor do que a técnica de imagem estrutural. No campo da DA, a RM é uma das técnicas de imagem não invasivas para a análise estrutural de cérebros (JACK et al., 2008), utilizada para a identificação de neurodegeneração, ou seja, para a classificação da positividade de N no sistema AT(N).

1.4. Tomografia por emissão de pósitrons (PET)

O PET, por outro lado, tem sido amplamente utilizado para investigar alterações funcionais do cérebro tanto em indivíduos saudáveis como no contexto patológico. O PET é uma técnica de imagem molecular que permite obter imagens tridimensionais do que está acontecendo no cérebro de um paciente em nível molecular e celular (ELLIS et al., 2011). Estudos do metabolismo cerebral com PET utilizam [¹⁸F]-fluorodesoxiglicose (FDG) como marcador metabólico no diagnóstico diferencial da DA. Com o FDG-PET é possível avaliar o metabolismo da glicose em diferentes regiões cerebrais, o que permite, em última instância, assim como a RM, medir a positividade para a neurodegeneração no sistema AT(N) (YANASE et al., 2005).

No entanto, diferentes radiotraçadores podem ser utilizados com a técnica de imagem por PET. Com isso, é possível visualizar o acúmulo de placa A β em um cérebro com DA e monitorar a progressão da doença através de imagens PET de A β (HAMEED et al., 2020). Dentre os radiotraçadores de A β , os mais conhecidos são Florbetapir, Flutemetamol e Florbetaben. Estes, por sua vez, podem ser utilizados para classificar a categoria A do sistema AT(N).

Também temos diversos estudos utilizando radiotraçadores de tau, possibilitando identificar pacientes com acúmulo de NFT. A imagem por tau PET demonstrou ser sensível e detecta alterações cognitivas precoces (pré-clínica) na DA (BRIER et al., 2016). Atualmente, os radiotraçadores de tau ainda não estão disponíveis para uso clínico. Imagens obtidas por tau PET podem ser utilizados para classificar a categoria T do sistema AT(N)

1.5. Biomarcadores de fluido

Alternativamente aos biomarcadores de imagens, os biomarcadores de fluídos podem ser utilizados como critério diagnóstico de cada categoria de biomarcadores do sistema AT(N). Atualmente, são aceitos, para indicar a positividade de cada categoria, dentre os biomarcadores de fluídos, aqueles obtidos a partir do LCR (JACK et al., 2016).

1.5.1. Líquido cefalorraquidiano (LCR)

O LCR está em contato direto com o espaço extracelular do cérebro e, portanto, as alterações bioquímicas no cérebro são refletidas nele, fazendo com que ele seja uma das fontes ideais para biomarcadores da DA (BLENNOW et al., 2010). Atualmente sabe-se que os sistemas de clearance das proteínas clássicas da patologia da DA, A β e tau, envolvem a liberação dessas proteínas para o LCR com posterior absorção para o sistema linfático (TARASOFF-CONWAY et al., 2015). No LCR, podemos medir inúmeros biomarcadores com significado clínico para a DA. Os mais utilizados, mas não limitados a eles, incluem a razão A β_{1-42} /A β_{1-40} , tau hiperfosforilada em treonina 181 (p-tau181) e neurofilamento de cadeia leve (NfL) (BLENNOW; ZETTERBERG; FAGAN, 2012; JACK et al., 2016).

O $A\beta_{1-42}$ e tau demonstraram ser componentes primários de placas amiloides e NFT, respectivamente. Por isso, os níveis dessas proteínas no LCR foram avaliados como potenciais biomarcadores dessas características patológicas (BLENNOW; ZETTERBERG; FAGAN, 2012).

1.5.2. Sangue

No entanto, em busca de exames com menor custo e maior acessibilidade, vem sendo aplicado um grande esforço na descoberta de exames sanguíneos no campo de pesquisa em biomarcadores de fluídos para a DA. A otimização de ferramentas analíticas permitindo uma sensibilidade na magnitude de picogramas na detecção de biomarcadores sanguíneos, incluindo a razão $A\beta_{1-42}/A\beta_{1-40}$, tau fosforilada em diversos sítios aceptores (p-tau) e NfL (CULLEN et al., 2021) trouxe expectativas promissoras para a área.

1.6. Técnicas neurocomputacionais na Doença de Alzheimer

Quando consideramos os avanços obtidos na medicina nas últimas décadas, é inegável que a computação foi uma grande contribuinte para que esses pudessem ser atingidos. Ademais, a computação se mostra uma grande aliada para ajudar no diagnóstico de doenças, uma vez que a Inteligência Artificial (IA) tem sido cada vez mais aplicada na área médica (HAMET; TREMBLAY, 2017).

Além disso, é crescente a preocupação na medicina para obter a maior quantidade possível de dados sobre a saúde ou doença dos pacientes e tomar decisões com base nisso. Antes, os médicos tinham a necessidade de confiar em sua experiência, capacidade de julgamento e habilidade de resolução de problemas enquanto usavam ferramentas rudimentares e recursos limitados. No entanto, nos

dias de hoje, as tecnologias disruptivas começaram a disponibilizar métodos avançados não só para profissionais da medicina, mas também para seus pacientes. Dentre estas técnicas neurocomputacionais, temos a ômica, biotecnologia, sensores portáteis e a IA (MESKO, 2017).

Ao invés de desenvolver tratamentos para populações e tomar as mesmas decisões médicas com base em algumas características físicas semelhantes entre os pacientes, a medicina tem avançado para a prevenção, personalização e precisão, o que fará com que soluções médicas tradicionais *one-fits-all* migrem para tratamentos direcionados, terapias personalizadas e drogas exclusivas. Em outras palavras: medicina de precisão. Nesta mudança e transformação cultural, técnicas neurocomputacionais são uma tecnologia chave que pode trazer esta oportunidade à prática diária (MESKO, 2017).

No aspecto da IA, a sua aplicação na medicina se divide em dois ramos principais: virtual e físico. O ramo físico envolve o uso de robôs enquanto o ramo virtual, representado pelo Aprendizado de Máquina (ML), utiliza algoritmos matemáticos que melhoram a aprendizagem através da experiência (HAMET; TREMBLAY, 2017).

Os algoritmos de ML podem ser classificados em três tipos: (i) supervisionados, (ii) não supervisionados e (iii) aprendizagem de reforço (HAMET; TREMBLAY, 2017). Dentre esses tipos, o mais comum e utilizado é o aprendizado supervisionado (LECUN; BENGIO; HINTON, 2015). Os algoritmos de aprendizado supervisionado constroem modelos a partir de um conjunto de dados que contém tanto as entradas quanto as saídas desejadas (RUSSELL; NORVIG; DAVIS, 2010).

Dentre as principais abordagens de ML supervisionada podemos citar as árvores de decisão, regressão logística e redes neurais artificiais (RNA).

1.7. Biologia de Sistemas

A natureza multifacetada da DA exige a incorporação de paradigmas complexos para estudar suas origens, mecanismos, desenvolvimento e possíveis tratamentos. Assim, o uso de métodos de biologia de sistemas se torna crucial na pesquisa da DA (VILLOSLADA; STEINMAN; BARANZINI, 2009; WANG et al., 2019). Uma visão sistêmica da biologia descreve uma abordagem multidisciplinar e holística para entender os fenômenos biológicos, focando nas interações de muitos elementos simultaneamente (KIRSCHNER, 2005; OLTVAI; BARABÁSI, 2002). Nesse contexto, tecnologias ômicas, como a transcriptômica, são os métodos de escolha para avaliar propriedades e dinâmicas de sistemas de forma rápida, ampla e confiável. A transcriptômica é especialmente atraente porque os RNAs representam um nível intermediário entre um contexto genômico estático e a dinâmica espaço-temporal da complexidade proteômica (VERHEIJEN; SLEEGERS, 2018).

1.8. Justificativa

No contexto deste trabalho, podemos considerar que esforços realizados nas pesquisas da área de biomarcadores da DA sejam divididos em dois grandes tipos: i) identificação de biomarcadores em diferentes compartimentos biológicos (cérebro, LCR e sangue) para classificação de cada uma das etapas do atual sistema AT(N) e ii) descoberta de novos biomarcadores capazes de prever alterações biológicas na DA. Desta forma, propomos duas estratégias neste trabalho. A primeira, apresentada no Capítulo 1, para a classificação de eventos patológicos da DA

através do uso de técnicas de ML com a utilização de diferentes isoformas de A β e a segunda, exposta no Capítulo 2, com o objetivo de descobrir potenciais novos biomarcadores sanguíneos da DA.

2. OBJETIVOS

2.1. Objetivo principal

Aplicar estratégias neurocomputacionais para a identificação de alterações patofisiológicas na Doença de Alzheimer.

2.2. Objetivos específicos

- Revisar a literatura de biomarcadores de fluídos e imagem para definir potenciais biomarcadores de estágios iniciais da DA;
- Investigar se a combinação de isoformas solúveis de A β no LCR podem prever a positividade de patologia de tau e neurodegeneração utilizando aprendizado de máquina;
- Desenvolver um método de integração de tecnologia ômica e imagens PET para descoberta de potenciais novos biomarcadores sanguíneos da DA;

PARTE II

Na Parte II, são apresentados dois capítulos que representam artigos científicos que compõem esta tese de doutorado.

Capítulo 1

Soluble amyloid-beta isoforms predict downstream Alzheimer's disease pathology

No capítulo 1 apresentamos o artigo publicado no periódico *Cell & Bioscience*.

Sabe-se que alterações nos níveis de A β solúvel no LCR são detectáveis nos estágios pré-clínicos da DA. No entanto, não se sabe se os níveis de A β podem prever os eventos patológicos seguintes na DA em indivíduos cognitivamente saudáveis. Para isso, propusemos o uso de aprendizado de máquina para criar modelos a partir da combinação de isoformas solúveis de A β para prever T+ e (N)+ nesses indivíduos. Além disso, utilizamos dados de proteômica do LCR para investigar o enriquecimento funcional de processos biológicos alterados em indivíduos T+ e (N)+. Com este trabalho, demonstramos que a combinação de isoformas de A β pode prever T+ e (N)+ com alta AUC. A análise proteômica do LCR destacou um grupo promissor de proteínas que podem ser exploradas para melhorar a previsão de T+ e (N)+.

Soluble amyloid-beta isoforms predict downstream Alzheimer's disease pathology

Guilherme Povala^{1,2}, Bruna Bellaver¹, Marco Antônio De Bastiani¹, Wagner S. Brum¹, Pamela C. L. Ferreira^{1,3}, Andrei Bieger¹, Tharick A. Pascoal³, Andrea L. Benedet^{4,5}, Diogo O. Souza^{1,6}, Ricardo M. Araujo², Bruno Zatt², Pedro Rosa-Neto^{4,5,7}, Eduardo R. Zimmer^{1,8,9*} and for the Alzheimer's Disease Neuroimaging Initiative

¹ Graduate Program in Biological Sciences: Biochemistry, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

² Graduate Program in Computing, Universidade Federal de Pelotas (UFPEL), Pelotas, Brazil.

³ Department of Neurology and Psychiatry, University of Pittsburgh, Pittsburgh, US.

⁴ Translational Neuroimaging Laboratory, The McGill University Research Centre for Studies in Aging, 6825 LaSalle Boulevard, Verdun, QC, H4H 1R3, Canada.

⁵ Montreal Neurological Institute, 3801 University Street, Montreal, QC, H3A 2B4, Canada.

⁶ Department of Biochemistry, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

⁷ Douglas Hospital, McGill University, 6875 La Salle Blvd-FBC room 3149, Montreal, QC, H4H 1R3, Canada.

⁸ Department of Pharmacology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

⁹ Graduate Program in Biological Sciences: Pharmacology and Therapeutics, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

* Corresponding author. Email: erzimmer@gmail.com

Guilherme Povala: guilhermepovala@gmail.com

Bruna Bellaver: brunabellaver90@gmail.com

Marco Antônio De Bastiani: tyrev@hotmail.com

Wagner S. Brum: wagnersbrum@gmail.com

Pamela C. L. Ferreira: pamlukasewicz@gmail.com

Andrei Bieger: andreibieger@gmail.com

Tharick A. Pascoal: pascoal@pitt.edu

Andrea L. Benedet: andrea.benedet@mail.mcgill.ca

Diogo O. Souza: diogo@ufrgs.br

Ricardo M. Araujo: ricardo.araujo@gmail.com

Bruno Zatt: zatt@inf.ufpel.edu.br

Pedro Rosa-Neto: pedro.rosa@mcgill.ca

Eduardo R. Zimmer: erzimmer@gmail.com

Abstract

Background: Changes in soluble amyloid-beta ($A\beta$) levels in cerebrospinal fluid (CSF) are detectable at early preclinical stages of Alzheimer's disease (AD). However, whether $A\beta$ levels can predict downstream AD pathological features in cognitively unimpaired (CU) individuals remains unclear. With this in mind, we aimed at investigating whether a combination of soluble $A\beta$ isoforms can predict tau pathology (T+) and neurodegeneration (N+) positivity.

Methods: We used CSF measurements of three soluble $A\beta$ peptides ($A\beta_{1-38}$, $A\beta_{1-40}$ and $A\beta_{1-42}$) in CU individuals ($n = 318$) as input features in machine learning (ML) models aiming at predicting T+ and N+. Input data was used for building 2046 tuned predictive ML models with a nested cross-validation technique. Additionally, proteomics data was employed to investigate the functional enrichment of biological processes altered in T+ and N+ individuals.

Results: Our findings indicate that $A\beta$ isoforms can predict T+ and N+ with an area under the curve (AUC) of 0.929 and 0.936, respectively. Additionally, proteomics analysis identified 17 differentially expressed proteins (DEPs) in individuals wrongly classified by our ML algorithm. More specifically, enrichment analysis of gene ontology biological processes revealed an upregulation in myelination and glucose metabolism-related processes in CU individuals wrongly predicted as T+. A significant enrichment of DEPs in pathways including biosynthesis of amino acids, glycolysis/gluconeogenesis, carbon metabolism, cell adhesion molecules and prion disease was also observed.

Conclusions: Our results demonstrate that, by applying a refined ML analysis, a combination of $A\beta$ isoforms can predict T+ and N+ with a high AUC. CSF proteomics

analysis highlighted a promising group of proteins that can be further explored for improving T+ and N+ prediction.

Keywords: Alzheimer's disease; amyloid-beta; tau pathology; neurodegeneration; machine learning; proteomics.

Background

Alzheimer's disease (AD) is the most prevalent neurodegenerative disease worldwide (1). Its main neuropathological features involve the deposition of two proteins, amyloid- β (A β) and tau, into insoluble aggregates in the brain (2, 3). Indeed, the most accepted AD theoretical model suggests that A β dysmetabolism triggers a cascade of downstream pathological events, including tau pathology, synaptic dysfunction, and neurodegeneration, which leads to cognitive decline and, ultimately, to dementia (4, 5).

This theoretical model relies on data derived from cross-sectional and longitudinal multicentric studies using multiple biomarkers. Currently, AD biomarkers are divided into two main classes: biofluid-based [blood and cerebrospinal fluid (CSF)] and neuroimaging [magnetic resonance imaging (MRI) and positron emission tomography (PET)] (6). These biomarkers constitute the basis of the National Institute on Aging-Alzheimer's Association (NIA-AA) Research Framework proposed for clinical studies, which adopted the A/T/(N) system for amyloid, tau, and neurodegeneration biomarkers (7). In each category, biomarkers are dichotomized to indicate a normal or abnormal status (7).

Importantly, this system relies on the amyloid cascade hypothesis, i.e., the linear chain A β positivity (A+) \rightarrow tau positivity (T+) \rightarrow neurodegeneration positivity (N+) \rightarrow cognitive symptoms (4, 5). However, around 30% of cognitively unimpaired (CU) individuals are A+ but do not present any other AD pathological features (8-10). Thus, A+, usually indexed by CSF A β ₁₋₄₂ or PET, does not infer *per se* if an individual presents or will develop tau pathology or neurodegeneration. Therefore, it is clear

that other biological processes are also critical in the progression toward clinical symptoms.

In this study, we asked (i) whether a combination of A β isoforms, measured in the CSF, would be capable of predicting downstream pathological biomarkers and (ii) what biological processes are related to an increase in A β isoforms' prediction power over downstream AD pathology. To answer these inquiries, we aimed at predicting T+ and N+ using a combination of demographics and A β isoforms levels in the CSF (A β ₁₋₃₈, A β ₁₋₄₀, and A β ₁₋₄₂) as input features in machine learning models (ML). We also evaluated whether CSF proteomic analyses could reveal altered biological processes heterogeneity in individuals wrongly classified in ML models.

Methods

ADNI description

Data used in this article are available at the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI is a longitudinal multicentric study launched in 2004, as a result of a public-private partnership, including the Foundation for the National Institutes of Health and the National Institute on Aging alongside contributors from many other sources. The study is currently in its 4th phase (ADNI1, ADNI GO, ADNI2, and ADNI3) and has recruited over 2300 participants in North America, to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. More information on the study design can be found in adni.loni.usc.edu/about/.

Eligibility criteria

In this study, data from 318 CU subjects were collected from ADNI1 and ADNI2 database. Specific criteria for inclusion in this study were the availability of

CSF levels of $A\beta_{1-38}$, $A\beta_{1-40}$, and $A\beta_{1-42}$ proteins measured by 2D-ultra-performance liquid chromatography-tandem mass spectrometry (2D-UPLC-MS/MS). ADNI's inclusion and diagnostic criteria have been described elsewhere (11).

CSF biomarker collection and analysis

CSF $A\beta_{1-38}$, $A\beta_{1-40}$, and $A\beta_{1-42}$ peptide levels were measured using the 2D-UPLC-MS/MS method (Waters® XEVO-TQ-S), which had been previously described (12) and has been recently revalidated. This updated technique has been recognized as an accepted analytical reference by the Joint Committee for Traceability in Laboratory Medicine (JCTLM), in whose database it was published under the JCTLM Identification Number C12RMP1. For defining T+ and N+, p-tau (Thr-181) and t-tau levels used in this study were measured by the Elecsys® immunoassay, with T+ defined as CSF p-tau (181-Thr) > 19.2 pg/mL and N+ defined as CSF t-tau > 242 pg/mL (13). Data for the 2D-UPLC-MS/MS and Elecsys® methods are available, respectively, at the ADNI database under the file names "UPENNMSMSABETA.csv" and "UPENNBIOMK9_04_19_17.csv".

Statistical analysis

All statistical analyses were performed in GraphPad Prism 8. Data are expressed as mean \pm standard deviation (SD). Normality was evaluated using histograms and quantile plots. Because samples did not have Gaussian distributions, comparisons between groups were carried out using Mann-Whitney test. P-values of less than 0.05 were reported as statistically significant.

Machine learning framework

We developed a ML framework that combines multiple techniques and models to predict T+ and N+ with the use of CSF $A\beta$ isoform levels, demographic information

and APOE $\epsilon 4$ status. The framework was coded in Python (version 3.6.8, <https://www.python.org/>), using the scikit-learn (version 0.20.2, <https://scikit-learn.org/>) and xgboost (version 0.81, <https://xgboost.readthedocs.io/>) libraries. The supervised ML algorithms used in our framework are composed of Logistic Regression, Naive Bayes, k-Nearest Neighbors (kNN), Support Vector Classifier (SVC), Decision Trees, Random Forest, Gradient Boosting, XGBoost, and AdaBoost.

As input features for our framework, we used A β peptide levels (A β_{1-38} , A β_{1-40} , and A β_{1-42}), demographic information (age, sex and years of education), and APOE $\epsilon 4$ status. For feature selection, we evaluated all possible feature combinations, generating 1023 subsets. For each feature subset, we performed the nested cross-validation (CV) technique. Here, we used the nested CV since we needed to train different ML models together with its hyperparameter optimization. The nested CV has an inner CV loop nested in outer CV. The inner loop is composed of a 2-fold CV, and it is responsible for model selection and hyperparameter tuning, which is similar to a validation set. The outer loop, however, is composed of a 5-fold CV and it is used for error estimation, as a test set. The nested cross-validation uses the area under de curve (AUC) metric to select the best hyperparameters and models. Then, an independent test set is used to test the overall performance of the best model and to generate the AUC result. The hyperparameters evaluated for each ML algorithm used in this work are shown in **Table 1**. After obtaining the AUC results for tuned ML algorithms with the nested cross-validation, only the model that presented the best performance is chosen for each feature subset. Among all these models, we selected the best one and then extracted the AUC for the independent test set.

CSF proteomics analysis

Processed CSF proteomics data were collected from the ADNI database. Samples were measured using the LC/MS-MRM method (12). Proteins and peptides were selected based upon their previous detection in CSF, relevance to AD, and previous results from the Rules Based Medicine (RBM) multiplex immunoassay analysis of ADNI CSF. The final MRM panel consisted of 567 peptides representing 221 proteins. From these 567 peptides, 320 were detectable in > 10% of ADNI samples and are available in the file “CSFMRM.csv”.

From the previously included CU individuals, only 76 presented CSF proteomics data in the ADNI database and were included in further analyses. CSF proteomics analysis was performed comparing T- (n = 55) and T+ (n = 21) individuals and N- (n = 57) and N+ (n = 19). All proteomic analyses were implemented in an R statistical environment. Differentially expressed analysis was computed for T-/T+ and N-/N+ groups independently, using the LIMMA (version 3.46.0) package (14), and considering FDR-adjusted p-value < 0.05 as differentially expressed proteins (DEP) criteria. Finally, functional enrichment analyses of gene ontology (GO) biological processes and KEGG pathways were computed and visualized using the clusterProfiler (version 3.18.1) and Goplot (version 1.0.2) packages (15, 16).

Results

Sample characteristics

For this study, we included 318 CU individuals from ADNI, whose CSF had been analyzed with 2D-UPLC-MS/MS. Characteristics of the ADNI cohort and the different A, T, and N status of samples are provided in **Table 2**. Population

characteristics were compared between positive and negative group status for each of the above-mentioned biomarker categories. A+ and T+ showed significantly more APOE ϵ 4 carriers than A β negative (A-) and tau negative (T-) groups. As already observed in previous studies, APOE ϵ 4 carriers are associated with decreased A β_{1-42} and elevated CSF p-tau in the CSF (14, 15). T+ and N+ presented elevated age, when compared with T- and neurodegeneration negative (N-) groups, respectively. No significant differences were observed in sex, years of education, Mini-Mental State Examination (MMSE), and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) among groups.

Changes in A β soluble isoforms in T+ and N+ CU individuals

Fig. 1 compares A β isoform levels and their respective ratios between T+ and T- (**Fig. 1a**), and N+ and N- (**Fig. 1b**). When comparing T status, T+ group presented higher levels of A β_{1-38} (**Fig. 1c**, T- = 1764 \pm 496.1 pg/mL, T+ = 2411 \pm 566.95 pg/mL, $p < 0.0001$) and A β_{1-40} (**Fig. 1d**, T- = 7617 \pm 2052 pg/mL, T+ = 10424 \pm 2529 pg/mL, $p < 0.0001$). Additionally, a decrease in A β_{1-42} /A β_{1-40} (**Fig. 1f**, T- = 0.1749 \pm 0.05, T+ = 0.1381 \pm 0.06, $p < 0.0001$) and A β_{1-42} /A β_{1-38} ratios (**Fig. 1g**, T- = 0.7610 \pm 0.22, T+ = 0.6014 \pm 0.25, $p < 0.0001$) was observed in T+ individuals. However, we did not observe any significant difference in A β_{1-42} levels (**Fig. 1e**, T- = 1353 \pm 559.4 pg/mL, T+ = 1492 \pm 784 pg/mL, $p = 0.41$) and A β_{1-40} /A β_{1-38} ratio (**Fig. 1h**, T- = 4.354 \pm 0.42, T+ = 4.329 \pm 0.35, $p = 0.60$) between T+ and T- groups.

For N+ individuals, A β_{1-38} (**Fig. 1i**, N- = 1760 \pm 469.6 pg/mL, N+ = 2503 \pm 567.2 pg/mL, $p < 0.0001$), A β_{1-40} (**Fig. 1j**, N- = 7593 \pm 1945 pg/mL, N+ = 10838 \pm 2503 pg/mL, $p < 0.0001$), and A β_{1-42} (**Fig. 1k**, N- = 1328 \pm 565.1 pg/mL, N+ = 1575 \pm 778.8 pg/mL, $p = 0.03$) measures were significantly elevated when compared to N-,

along with a decrease in $A\beta_{1-42}/A\beta_{1-40}$ ratio (**Fig. 1l**, $N^- = 0.1720 \pm 0.05$, $N^+ = 0.1411 \pm 0.05$, $p < 0.0001$) and $A\beta_{1-42}/A\beta_{1-38}$ ratio (**Fig. 1m**, $N^- = 0.7483 \pm 0.23$, $N^+ = 0.6146 \pm 0.25$, $p < 0.0001$). By contrast, $A\beta_{1-40}/A\beta_{1-38}$ ratio (**Fig. 1n**, $N^- = 4.350 \pm 0.41$, $N^+ = 4.336 \pm 0.35$, $p = 0.78$) does not differ between N^+ and N^- groups.

To test whether single $A\beta$ isoforms or its ratios can predict downstream AD pathological processes in CU individuals, we used logistic regression models. The AUC results for predicting T^+ and N^+ individuals are shown in **Table 3**. Among all results, $A\beta_{1-38}$ and $A\beta_{1-40}$ seem to be the most reliable features to predict T^+ , with an AUC of 0.811 for both $A\beta$ isoforms. For predicting N^+ , $A\beta_{1-38}$ and $A\beta_{1-40}$ showed similar results, with AUCs of 0.847 and 0.855, respectively. On the other hand, $A\beta_{1-42}$ presented an AUC of 0.580 for predicting N^+ and 0.529 for T^+ .

Machine learning framework

Aiming at better predictive models, we proposed a ML framework, which is presented in **Fig. 2**. $A\beta$ isoforms in the CSF ($A\beta_{1-38}$, $A\beta_{1-40}$, and $A\beta_{1-42}$; measured by 2D-UPLC-MS/MS), APOE $\epsilon 4$ carrier status, and demographic information (age, sex, and years of education) were used as input features. Besides, for feature generation, $A\beta$ isoforms were used either alone or combined in ratios (**Fig. 2a**). In the feature subset generation step (**Fig. 2b**), all possible combinations of features were created (1023 different subsets). Then, for each subset, two models were selected using the nested CV technique (**Fig. 2c**): one for T^+ prediction and another to predict N^+ (**Fig. 2d**).

In our ML framework, to choose the best model for each subset to classify T^+ and N^+ , we evaluated the use of the following ML algorithms: Logistic Regression, Naïve Bayes, kNN, SVC, Decision Trees, Random Forest, Gradient Boosting,

XGBoost, and AdaBoost within the nested CV technique. For each subset, the best model was defined based on the model's AUC obtained from the validation set. The top 1 model among the 1023 models (one for each subset) was evaluated using an independent test set and was defined as the best model to predict T+ or N+.

Tau pathology positivity prediction

From our proposed ML framework, 1023 tuned ML models were generated for predicting T+ (**Additional file 1**). **Fig. 3a** shows the AUC results for predicting T+ horizontally ordered by AUC – SD. In **Fig. 3b**, the best 10 models are ranked. Among the 10 models, all of them presented similar results, ranging from 0.877 to 0.887.

The top 1 model was a logistic regression model using $A\beta_{1-42}$, $A\beta_{1-42}/A\beta_{1-40}$, $A\beta_{1-42}/A\beta_{1-38}$, $A\beta_{1-40}/A\beta_{1-38}$, and years of education as input features. The AUC result obtained for the validation set was 0.881 +/- 0.024. For the independent test set, we achieved an AUC of 0.929 (**Fig. 3c**).

Neurodegeneration positivity prediction

For N+ prediction, we generated another 1023 models using the same method (**Additional file 2**). The AUC results for the N+ predictions are shown in **Fig. 3d** horizontally ordered by AUC – SD. The best 10 models were ranked and plotted on the graph represented in **Fig. 3e**. The best 10 models presented similar results, ranging from 0.909 to 0.915.

A kNN generated the best results, which had $A\beta_{1-42}$, $A\beta_{1-40}$, $A\beta_{1-42}/A\beta_{1-40}$, $A\beta_{1-42}/A\beta_{1-38}$, and $A\beta_{1-40}/A\beta_{1-38}$ as input features. The AUC result for the validation set for this model was 0.915 +/- 0.018. The independent test set achieved an AUC of 0.936 (**Fig. 3f**).

CSF proteomics of T+ and N+ CU individuals

To address T+ and N+ CU individuals' functional changes in biological processes, we performed CSF-based proteomics analyses. A total of 112 DEPs were observed in the CSF of CU T+ compared to T- subjects (**Additional file 3**). The enrichment analysis of GO biological processes in T+ individuals evidenced processes related to myelinization, synapse and neurogenesis regulation, immune response, carbohydrate metabolism, memory and learning, and glial cell differentiation (**Fig. 4a**). **Fig. 4b** depicts top 20 GO terms enriched in T+ subjects compared to T-. To identify the most affected pathways related to changes in proteomics profile of T+, we performed an enrichment analysis using canonical pathways described in the KEGG pathway database (17). This revealed a significant enrichment of 112 DEPs in 4 signaling pathways: “cell adhesion molecules”, “biosynthesis of amino acids”, “carbon metabolism”, and “prion disease” (**Fig. 4c-g**). Regarding proteomics analysis of N+, we identified 123 DEPs when compared to N- individuals (**Additional file 4**). Of note, T+ and N+ subjects share 101 DEPs. Functional enrichment analyses revealed an overlap of enriched GO terms in N+ individuals and T+ individuals (**Fig. 5a**). Synapse organization, learning and memory processes, and APP metabolic processes are among the top 20 GO terms enriched in N+ (**Fig. 5b**). Interestingly, the same 4 KEGG pathways enriched for T+ were found enriched for N+ individuals (**Fig. 5c-g**).

CSF proteomics analysis for ML wrong predictions

Because A β isoforms predicted T+ and N+ outcomes with an AUC of up to 0.936, we next aimed, with a second proteomics analysis, at identifying differences in biological processes occurring in CU individuals that were wrongly classified by our

ML algorithm in the test set. First, we stratified the ML predictions for T+ in false-positive (n = 17), false-negative (n = 23), true-positive (n = 51), and true-negative (n = 147). Proteomic analyses for N+ prediction model was not carried out, since few wrong predictions were generated, leading to a small sample size.

We identified 17 upregulated DEPs between true-positive and false-positive (**Fig. 6a**) and 67 upregulated DEPs between true-negative and false-negatives for T+ individuals (**Fig. 7a**). Interestingly, enrichment analysis of GO biological processes revealed that processes related to myelination, and glucose metabolism are enriched when comparing false-positive and true-positive predictions for T+ (**Fig. 6a-b**). When considering the false-negative and true-negative predictions for T+, DEPs related to glucose metabolism, synapse transmission, gliogenesis, and axogenesis appeared among the enriched GO terms (**Fig. 7a-b**). Finally, to recognize the most affected pathways related to changes in proteomics profile of individuals that were wrongly predicted, we performed an enrichment analysis using canonical pathways described in the KEGG pathway database. This revealed a significant enrichment of DEPs in pathways including “biosynthesis of amino acids”, “glycolysis/gluconeogenesis”, “carbon metabolism”, “cell adhesion molecules”, and “prion disease” (**Fig. 6c-g** and **7c-l**).

Discussion

In the present study, we demonstrated that ML models using combined A β soluble isoforms can predict downstream AD pathological processes, T+ and N+, in CU individuals with better results than A β isoforms independently. In the generated models, a higher AUC was achieved for predicting N+ when comparing with the T+. Our proteomics analysis identified several biological processes and signaling

pathways altered at pre-symptomatic phase of AD. These findings are especially relevant when considering that AD pathological processes initiate around 20-30 years before the occurrence of the first clinical symptoms (18-22). Finally, we identified DEPs among individuals wrongly classified as T+ by ML that can be further explored to improve prediction performance of the models.

The notion that A β triggers tau hyperphosphorylation and neurodegeneration has been corroborated by multiple experimental studies (23-26). In fact, Höglund and colleagues demonstrated that CU individuals with amyloidosis presented increased levels of p-tau181 and t-tau in the CSF (27). However, the diagnostic value of A β_{1-42} has been explored in the literature delivering, though, only modest accuracy for AD prediction (28, 29). Accordingly, here we demonstrated a poor AUC of 0.580 for N+ and 0.529 for T+ prediction modeled using the A β_{1-42} isoform by itself, the most used CSF biomarker in the diagnosis of AD. *Per se*, the poorly explored isoform A β_{1-38} (AUC of 0.847) along with A β_{1-40} (AUC of 0.811) were the most accurate predictors for both T+ and N+, respectively. In clinical studies, the A β_{1-42} /A β_{1-38} ratio has been capable of significantly discriminating AD from other forms of dementia (30-32) and shown to be negatively correlated with CSF p-tau levels in AD patients (31). Additionally, a slight increase in A β_{1-38} levels was found in a disease-specific manner in the CSF of AD subjects (32, 33). Nevertheless, a meta-analysis pointed no significant difference in A β_{1-38} levels between AD individuals and control group after comparing eight studies (34). Cullen and colleagues more recently demonstrated that higher CSF A β_{1-38} levels are negatively associated with cognitive decline and risk of developing AD (35). In this context, it is evident that the potential of this isoform to add information in the preclinical stage of the disease remains under-explored.

In this work, we showed that a logistic regression model could predict T+ using multiple input features, with an AUC of 0.929. It has been demonstrated that A β dysmetabolism is capable of triggering the conversion from a normal to a toxic state of tau-dependent synaptic (23). As well, a synergistic interaction between A β and tau pathology is likely to occur in AD, rather than the sum of their independent effects (36-38). Bilgel and colleagues showed that a higher baseline amyloid load in CU individuals was associated with steeper cognitive decline (39). In parallel, we hereby demonstrated that amyloid isoforms levels can predict N+ in CU individuals with an AUC of 0.936 using a kNN model. The combination of A β isoforms, especially those including smaller A β isoforms, seems to help to deliver the best results to predict N+. Indeed, we only found one *in vivo* study showing significant correlations between A β_{1-42} levels in the CSF and neurodegeneration in CU individuals (27). On the other hand, the importance of A β_{1-42} isoform as a toxic amyloid specie has been extensively demonstrated (23-26). In the context of isoform production, literature indicates that A β_{1-38} is partially formed by cleavage of the A β_{1-42} isoform (40). Also, it seems that no further cleavage of A β_{1-38} occurs, resulting in a “more stable” isoform of A β , easier to detect (40). One could argue that a more prominent amyloid dysmetabolism, with higher rates of cleavage of A β_{1-42} into A β_{1-38} , might be a crucial process that seems to drive tau pathology and neurodegeneration. However, the already described (41) faster turnover of A β_{1-42} might be accounting for its poor predictive value in our model. Accordingly, our model shows an important role for less explored A β isoforms as indicators of emerging tau pathology and neurodegeneration. In addition to CSF, plasma biomarkers have been gaining attention in the recent years for their valuable contribution to AD diagnosis.

A β isoforms used in combination seems key for predicting T+ and N+, but do not completely explain all the aspects of AD downstream events. Thus, it is believed that simultaneous phenomena, that account for AD heterogeneity, are taking place in the brains of these individuals. In this context, CSF proteomics has been increasingly applied in the attempt to discover novel biomarkers for AD. However, it is mainly focused in comparing CU and AD individuals (42, 43). Here, we showed A β pathology-dependent changes at protein level occurring in the CSF of CU individuals. Similarly, Whelan and colleagues performed a multiplex proteomics analysis in the CSF of CU A+ and A- patients and found two DEPs significantly altered: Chitinase 3-like protein (YKL-40) and SPARC-related modular calcium binding protein 2 (SMOC2) (44). The great number of DEPs between CU T+ and T- subjects identified in our study allowed the further determination of biological processes and signaling pathways significantly enriched in these individuals. Additionally, significant differences in DEPs and its associated biological processes and signaling pathways were observed when comparing right and wrong ML predictions for T+. Interestingly, DEPs identified in other studies comparing CU and AD were also found in our analysis of ML wrong predictions for T+ (43). In specific, YKL-40, SOD1, PKM, and glucose metabolism related proteins are among the DEPs found in both studies. The degree of similarity between studies seems to highlight a robust pattern of change rather than a cohort-specific effect. These results might shed light to key proteins that can be further explored to improve ML performance for predicting T+ and N+.

Conclusions

Our findings indicate that the use of ML models with A β isoforms as input features might help to predict individuals with AD downstream pathology. In addition,

CSF proteomics analysis highlighted a promising group of proteins potentially driving tau pathology, which can be further explored for improving future T+ and N+ prediction. Finally, the combination of methodologies used here - ML and proteomics - may help to further understand AD pathology heterogeneity.

List of abbreviations

2D-UPLC-MS/MS: 2D-ultra-performance liquid chromatography-tandem mass spectrometry; A+: amyloid-beta positivity; A-: A β negative; AD: Alzheimer's disease; ADAS-Cog: Alzheimer's Disease Assessment Scale-Cognitive Subscale; ADNI: Alzheimer's Disease Neuroimaging Initiative; AUC: area under the curve; A β : amyloid-beta; CSF: cerebrospinal fluid; CU: cognitively unimpaired; DEP: differentially expressed proteins; GO: gene ontology; JCTLM: Joint Committee for Traceability in Laboratory Medicine; ML: machine learning; MMSE: Mini-Mental State Examination; MRI: magnetic resonance imaging; N+: neurodegeneration positivity; N-: neurodegeneration negative; NIA-AA: National Institute on Aging - Alzheimer's Association; p-tau: phosphorylated tau; PET: positron emission tomography; RBM: Rules Based Medicine; SD: standard deviation; SMOC2: SPARC-related modular calcium binding protein 2; T+: tau pathology positivity; T-: tau pathology negative; t-tau: total tau; YKL-40: Chitinase 3-like protein.

Declarations

Ethics approval and consent to participate

ADNI was ethically approved by the institutional review board of all participating sites, subjects provided written informed consent.

Consent for publication

Not applicable.

Availability of data and materials

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The dataset supporting the conclusions of this manuscript is available at the ADNI website (<http://adni.loni.usc.edu/>).

Competing interests

The authors declare that they have no competing interests.

Funding

GP receives financial support from CAPES [88882.345577/2019-01]. BB receives financial support from CAPES [88887.336490/2019-00]. PRN receive grants from CIHR [MOP-11-51-31; FRN, 152985], Alzheimer's Association [NIRG-12-92090; NIRP-12-259245] and FRQS [2020-VICO-279314]. ERZ receives grants from CNPq [435642/2018-9; CNPq 312410/2018-2] Instituto Serrapilheira [Serra-1912-31365], FAPERGS/MS/CNPq/SESRS–PPSUS [30786.434.24734.23112017] and ARD/FAPERGS [54392.632.30451.05032021].

Authors' contributions

Conceptualization: GP, BB, WSB, BZ, EZ.

Methodology: GP, BB, WSB, MADB.

Software: GP.

Investigation: GP.

Visualization: GP, MADB.

Supervision: BB, BZ, RMA, EZ.

Writing—original draft: GP, BB, WSB, PCLF, EZ.

Writing—review & editing: GP, BB, WSB, MADB, PCLF, TAP, ALB, PRN, DOS, BZ, EZ.

All authors read and approved the final manuscript.

Acknowledgements

Data used in preparation of this manuscript were obtained from the ADNI database (adni.loni.usc.edu). The list of ADNI investigators can be found online at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

References

1. Collaborators GBDD. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019;18(1):88-106.
2. Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med.* 2011;1(1):a006189.
3. Perl DP. Neuropathology of Alzheimer's disease. *Mt Sinai J Med.* 2010;77(1):32-42.
4. Selkoe DJ. The molecular pathology of Alzheimer's disease. *Neuron.* 1991;6(4):487-98.
5. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science.* 1992;256(5054):184-5.
6. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. *J Intern Med.* 2018;284(6):643-63.
7. Jack CR, Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology.* 2016;87(5):539-47.
8. Aizenstein HJ, Nebes RD, Saxton JA, Price JC, Mathis CA, Tsopelas ND, et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. *Arch Neurol.* 2008;65(11):1509-17.
9. Jack CR, Jr., Lowe VJ, Weigand SD, Wiste HJ, Senjem ML, Knopman DS, et al. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. *Brain.* 2009;132(Pt 5):1355-65.
10. Pike KE, Savage G, Villemagne VL, Ng S, Moss SA, Maruff P, et al. Beta-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. *Brain.* 2007;130(Pt 11):2837-44.
11. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology.* 2010;74(3):201-9.
12. Korecka M, Waligorska T, Figurski M, Toledo JB, Arnold SE, Grossman M, et al. Qualification of a surrogate matrix-based absolute quantification method for

amyloid-beta(4)(2) in human cerebrospinal fluid using 2D UPLC-tandem mass spectrometry. *J Alzheimers Dis.* 2014;41(2):441-51.

13. Schindler SE, Gray JD, Gordon BA, Xiong C, Batrla-Utermann R, Quan M, et al. Cerebrospinal fluid biomarkers measured by Elecsys assays compared to amyloid imaging. *Alzheimers Dement.* 2018;14(11):1460-9.

14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.

15. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284-7.

16. Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015;31(17):2912-4.

17. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27-30.

18. Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med.* 2012;367(9):795-804.

19. Braak H, Braak E. Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiol Aging.* 1997;18(4 Suppl):S85-8.

20. Fagan AM, Xiong C, Jasielec MS, Bateman RJ, Goate AM, Benzinger TL, et al. Longitudinal change in CSF biomarkers in autosomal-dominant Alzheimer's disease. *Sci Transl Med.* 2014;6(226):226ra30.

21. Morris JC, Price JL. Pathologic correlates of nondemented aging, mild cognitive impairment, and early-stage Alzheimer's disease. *J Mol Neurosci.* 2001;17(2):101-18.

22. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid beta deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* 2013;12(4):357-67.

23. Bloom GS. Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol.* 2014;71(4):505-8.

24. Nisbet RM, Polanco JC, Ittner LM, Gotz J. Tau aggregation and its interplay with amyloid-beta. *Acta Neuropathol.* 2015;129(2):207-20.

25. Jacobs HIL, Hedden T, Schultz AP, Sepulcre J, Perea RD, Amariglio RE, et al. Structural tract alterations predict downstream tau accumulation in amyloid-positive older individuals. *Nat Neurosci*. 2018;21(3):424-31.
26. Jeong S. Molecular and Cellular Basis of Neurodegeneration in Alzheimer's Disease. *Mol Cells*. 2017;40(9):613-20.
27. Hoglund K, Kern S, Zettergren A, Borjesson-Hansson A, Zetterberg H, Skoog I, et al. Preclinical amyloid pathology biomarker positivity: effects on tau pathology and neurodegeneration. *Transl Psychiatry*. 2017;7(1):e995.
28. Hampel H, Toschi N, Baldacci F, Zetterberg H, Blennow K, Kilimann I, et al. Alzheimer's disease biomarker-guided diagnostic workflow using the added value of six combined cerebrospinal fluid candidates: Aβ₁₋₄₂, total-tau, phosphorylated-tau, NFL, neurogranin, and YKL-40. *Alzheimers Dement*. 2018;14(4):492-501.
29. Khoonsari PE, Shevchenko G, Herman S, Remnestal J, Giedraitis V, Brundin R, et al. Improved Differential Diagnosis of Alzheimer's Disease by Integrating ELISA and Mass Spectrometry-Based Cerebrospinal Fluid Biomarkers. *J Alzheimers Dis*. 2019;67(2):639-51.
30. Mulugeta E, Londos E, Ballard C, Alves G, Zetterberg H, Blennow K, et al. CSF amyloid β₃₈ as a novel diagnostic marker for dementia with Lewy bodies. *J Neurol Neurosurg Psychiatry*. 2011;82(2):160-4.
31. Welge V, Fiege O, Lewczuk P, Mollenhauer B, Esselmann H, Klafki HW, et al. Combined CSF tau, p-tau₁₈₁ and amyloid-β_{38/40/42} for diagnosing Alzheimer's disease. *J Neural Transm (Vienna)*. 2009;116(2):203-12.
32. Wiltfang J, Esselmann H, Bibl M, Smirnov A, Otto M, Paul S, et al. Highly conserved and disease-specific patterns of carboxyterminally truncated Aβ peptides 1-37/38/39 in addition to 1-40/42 in Alzheimer's disease and in patients with chronic neuroinflammation. *J Neurochem*. 2002;81(3):481-96.
33. Bibl M, Mollenhauer B, Lewczuk P, Esselmann H, Wolf S, Trenkwalder C, et al. Validation of amyloid-β peptides in CSF diagnosis of neurodegenerative dementias. *Mol Psychiatry*. 2007;12(7):671-80.
34. Olsson B, Lautner R, Andreasson U, Ohrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15(7):673-84.

35. Cullen NC, Janelidze S, Palmqvist S, Stomrud E, Mattsson-Carlgrén N, Hansson O. CSF A β 38 levels are associated with Alzheimer-related decline: implications for γ -secretase modulators. medRxiv. 2021:2021.01.31.21250702.
36. Pascoal TA, Mathotaarachchi S, Mohades S, Benedet AL, Chung CO, Shin M, et al. Amyloid-beta and hyperphosphorylated tau synergy drives metabolic decline in preclinical Alzheimer's disease. *Mol Psychiatry*. 2017;22(2):306-11.
37. Pascoal TA, Mathotaarachchi S, Shin M, Benedet AL, Mohades S, Wang S, et al. Synergistic interaction between amyloid and tau predicts the progression to dementia. *Alzheimers Dement*. 2017;13(6):644-53.
38. Busche MA, Hyman BT. Synergy between amyloid-beta and tau in Alzheimer's disease. *Nat Neurosci*. 2020;23(10):1183-93.
39. Bilgel M, An Y, Helphrey J, Elkins W, Gomez G, Wong DF, et al. Effects of amyloid pathology and neurodegeneration on cognitive change in cognitively normal adults. *Brain*. 2018;141(8):2475-85.
40. Okochi M, Tagami S, Yanagida K, Takami M, Kodama TS, Mori K, et al. gamma-secretase modulators and presenilin 1 mutants act differently on presenilin/gamma-secretase function to cleave Abeta42 and Abeta43. *Cell Rep*. 2013;3(1):42-51.
41. Patterson BW, Elbert DL, Mawuenyega KG, Kasten T, Ovod V, Ma S, et al. Age and amyloid effects on human central nervous system amyloid-beta kinetics. *Annals of neurology*. 2015;78(3):439-53.
42. Sathe G, Na CH, Renuse S, Madugundu AK, Albert M, Moghekar A, et al. Quantitative Proteomic Profiling of Cerebrospinal Fluid to Identify Candidate Biomarkers for Alzheimer's Disease. *Proteomics Clin Appl*. 2019;13(4):e1800105.
43. Bader JM, Geyer PE, Muller JB, Strauss MT, Koch M, Leyboldt F, et al. Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol Syst Biol*. 2020;16(6):e9356.
44. Whelan CD, Mattsson N, Nagle MW, Vijayaraghavan S, Hyde C, Janelidze S, et al. Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun*. 2019;7(1):169.

Figure legends

Fig. 1. A β isoforms levels discriminate tau pathology positivity (T+) and neurodegeneration positivity (N+) in CU individuals. (A) T+ indicated by a high level of p-tau in the CSF (p-tau > 19.2 pg/mL). (B) N+ indicated by a high level of t-tau in the CSF (t-tau > 242 pg/mL). (C) A β_{1-38} , (D) A β_{1-40} and (E) A β_{1-42} levels for T- and T+ individuals. (F) A β_{1-42} /A β_{1-40} , (G) A β_{1-42} /A β_{1-38} and (H) A β_{1-40} /A β_{1-38} ratios for T- and T+ individuals. (I) A β_{1-38} , (J) A β_{1-40} and (K) A β_{1-42} levels for N- and N+ individuals. (L) A β_{1-42} /A β_{1-40} , (M) A β_{1-42} /A β_{1-38} and (N) A β_{1-40} /A β_{1-38} ratios for N- and N+ individuals. Boxplots are displayed as median (center line) and extend from the 25th to 75th percentiles. The whiskers go down to the smallest value and up to the largest. Statistical differences were tested using Mann-Whitney test (*p \leq 0.05, ***p \leq 0.001, ****p \leq 0.0001).

Fig. 2. Machine learning framework delivers models that predict tau pathology and neurodegeneration. (A) cognitively unimpaired (CU) individual's cerebrospinal fluid (CSF) levels of A β_{1-38} , A β_{1-40} and A β_{1-42} , demographics data and APOE ϵ 4 status were used for feature generation. (B) All possible combinations of features were generated using the feature set. (C) The subsets were used for generating tuned machine learning models validated with nested cross-validation aiming to (D) identify tau pathology (T+) and neurodegeneration (N+) positivity.

Fig. 3. Results for predicting tau pathology (T) and neurodegeneration (N) status. (A) Area under the ROC curve (AUC) results (vertical axis) for all 1023 subsets to predict T status ordered by AUC – standard deviation (SD). (B) AUC results (horizontal axis) for the top 10 models (vertical axis) to predict T status. (C) ROC curve for the best model to predict T status using the independent test set. (D)

AUC results (vertical axis) for all 1023 subsets to predict N status ordered by AUC – SD. (E) AUC results (horizontal axis) for the top 10 models (vertical axis) to predict N status. (F) ROC curve for the best model to predict N status using the independent test set.

Fig. 4. Proteome analyses results of cerebrospinal fluid (CSF) cells between T- and T+ cognitively unimpaired (CU) patients. (A) Gene ontology (GO) network of enriched terms were constructed from differentially expressed proteins mapping the node sizes to GO term significance and edge width to shared protein proportions (Jaccard coefficient). (B) Radial plot of top 20 enriched GO terms. (C) Enriched pathways obtained from functional enrichment of KEGG terms. (D-G) Pie charts of enriched KEGG pathways showing the proportion of proteins upregulated in T+ vs T- comparison.

Fig. 5. Proteome analyses results of cerebrospinal fluid (CSF) cells between N- and N+ cognitively unimpaired (CU) patients. (A) Gene ontology (GO) network of enriched terms were constructed from differentially expressed proteins mapping the node sizes to GO term significance and edge width to shared protein proportions (Jaccard coefficient). (B) Radial plot of top 20 enriched GO terms. (C) Enriched pathways obtained from functional enrichment of KEGG terms. (D-G) Pie charts of enriched KEGG pathways showing the proportion of proteins upregulated in N+ vs N- comparison.

Fig. 6. Proteome analyses results of cerebrospinal fluid (CSF) cells between true positive (TP) and false positive (FP) predictions for tau pathology positivity (T+) in cognitively unimpaired (CU) individuals. (A) Gene ontology (GO) network of enriched terms were constructed from differentially expressed

proteins mapping the node sizes to GO term significance and edge width to shared protein proportions (Jaccard coefficient). **(B)** Radial plot of top 15 enriched GO terms. **(C)** Enriched pathways obtained from functional enrichment of KEGG terms. **(D-G)** Pie charts of enriched KEGG pathways showing the proportion of proteins upregulated in TP vs FP comparison.

Fig. 7. Proteome analyses results of cerebrospinal fluid (CSF) cells between true negative (TN) and false negative (FN) predictions for tau pathology positivity (T+) in cognitively unimpaired (CU) individuals. **(A)** Gene ontology (GO) network of enriched terms were constructed from differentially expressed proteins mapping the node sizes to GO term significance and edge width to shared protein proportions (Jaccard coefficient). **(B)** Radial plot of top 15 enriched GO terms. **(C)** Enriched pathways obtained from functional enrichment of KEGG terms. **(D-L)** Pie charts of enriched KEGG pathways showing the proportion of proteins upregulated in TN vs FN comparison.

Additional files

Additional file 1

File format: .docx

Title of data: Machine learning results for predicting tau pathology positivity (T+).

Description of data: Table containing features, AUC and standard deviation results for all 1023 models for predicting tau pathology positivity.

Additional file 2

File format: .docx

Title of data: Machine learning results for predicting neurodegeneration positivity (N+).

Description of data: Table containing features, AUC and standard deviation results for all 1023 models for predicting neurodegeneration positivity.

Additional file 3

File format: .docx

Title of data: Differentially expressed proteins (DEPs) in the cerebrospinal fluid (CSF) of cognitively unimpaired (CU) tau pathology positive (T+) compared to negative (T-) subjects.

Description of data: Table containing Protein ID, p-value, adjusted p-value, t-value and logFC for differentially expressed proteins in the cerebrospinal fluid of cognitively unimpaired tau pathology positive compared to negative subjects.

Additional file 4

File format: .docx

Title of data: Differentially expressed proteins (DEPs) in the cerebrospinal fluid (CSF) of cognitively unimpaired (CU) neurodegeneration positive (N+) compared to negative (N-) subjects.

Description of data: Table containing Protein ID, p-value, adjusted p-value, t-value and logFC for differentially expressed proteins in the cerebrospinal fluid of cognitively unimpaired neurodegeneration positive compared to negative subjects.

Figure 1

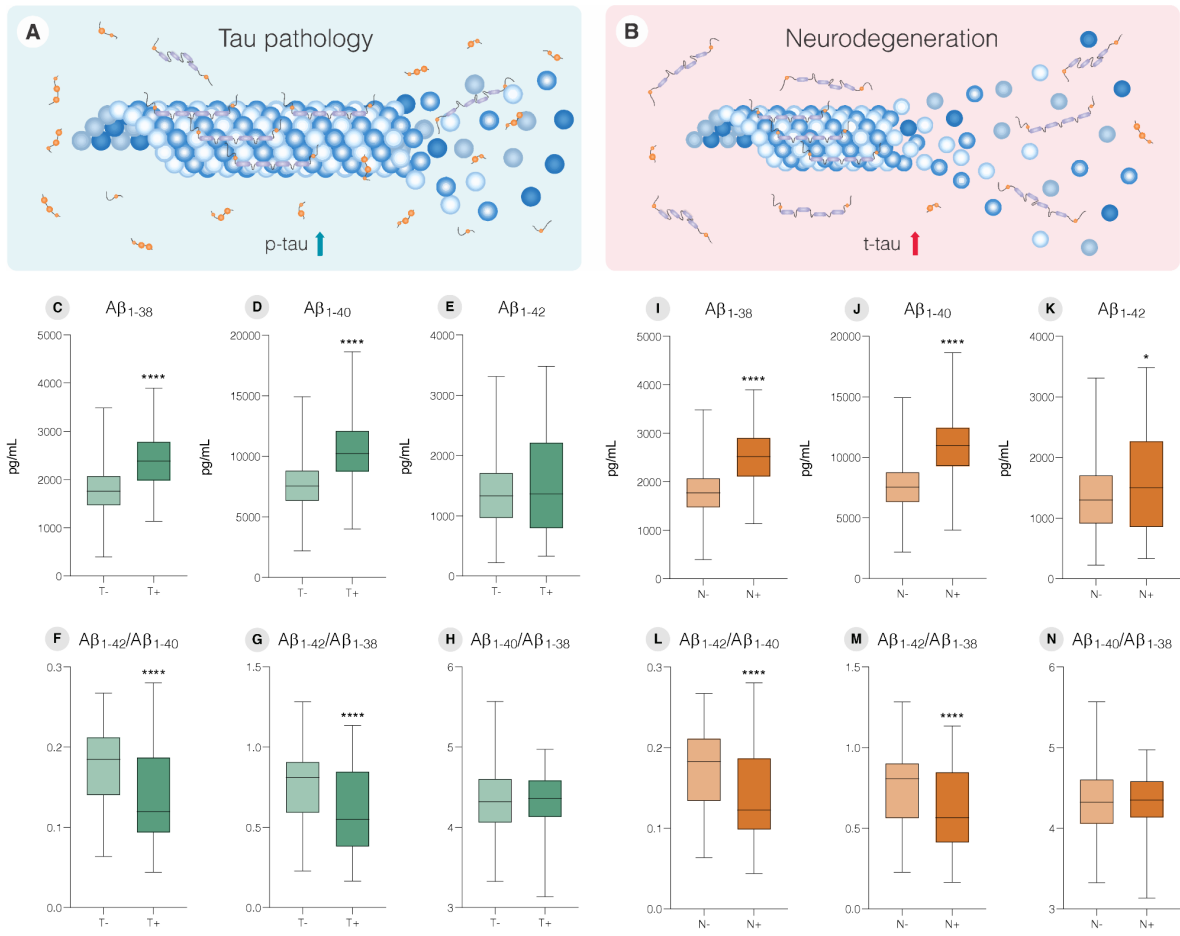


Figure 2

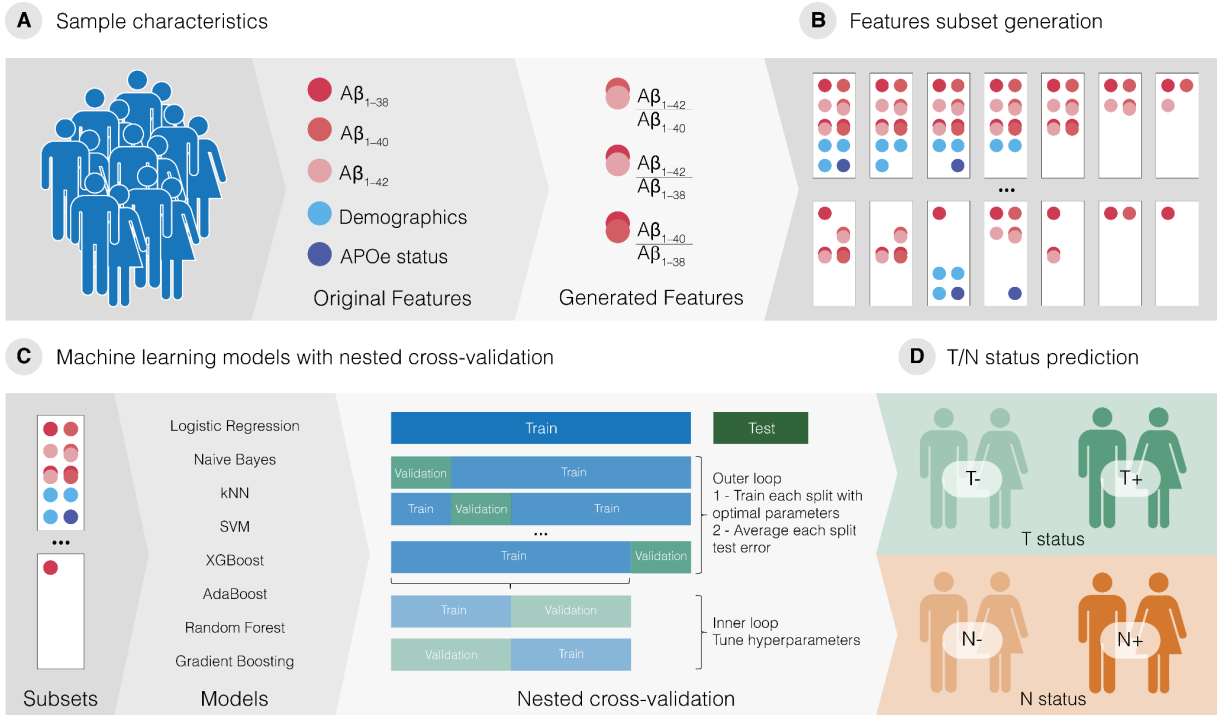
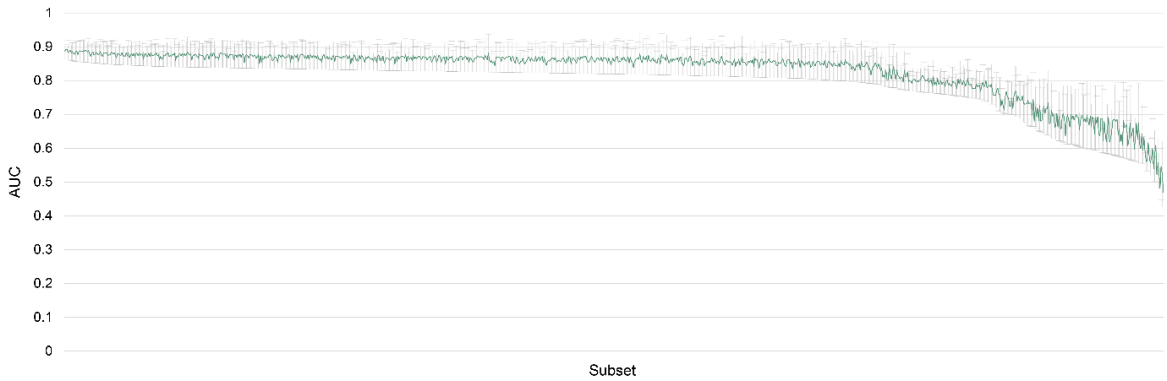
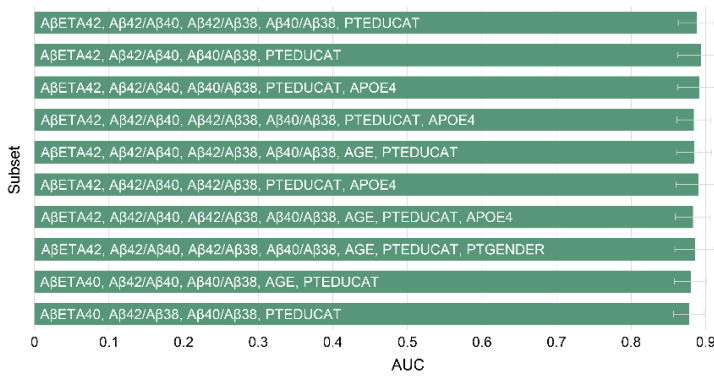


Figure 3

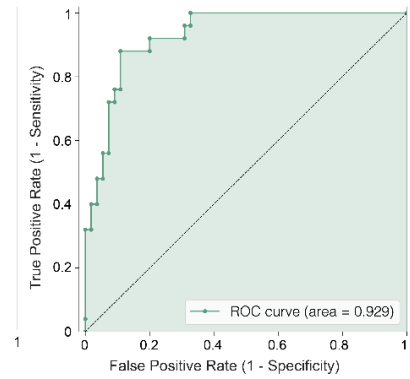
A Overall prediction results for T status



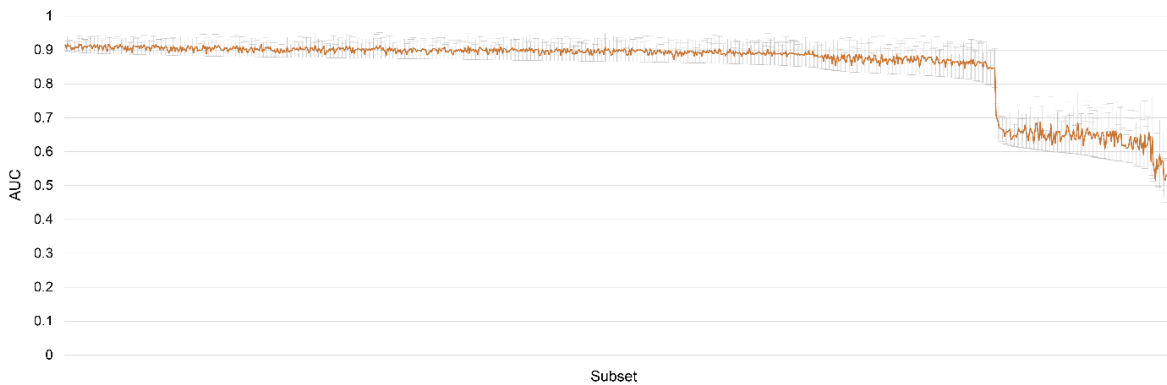
B Top 10 prediction results for T status



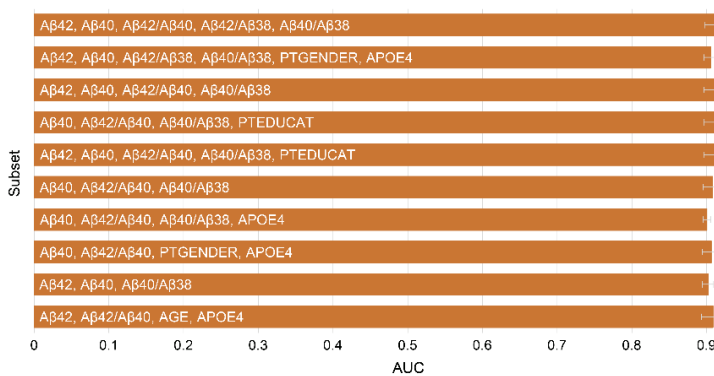
C ROC curve for T status - testing set



D Overall prediction results for N status



E Top 10 prediction results for N status



F ROC curve for N status - testing set

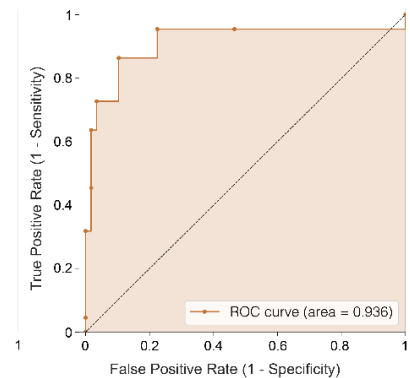


Figure 4

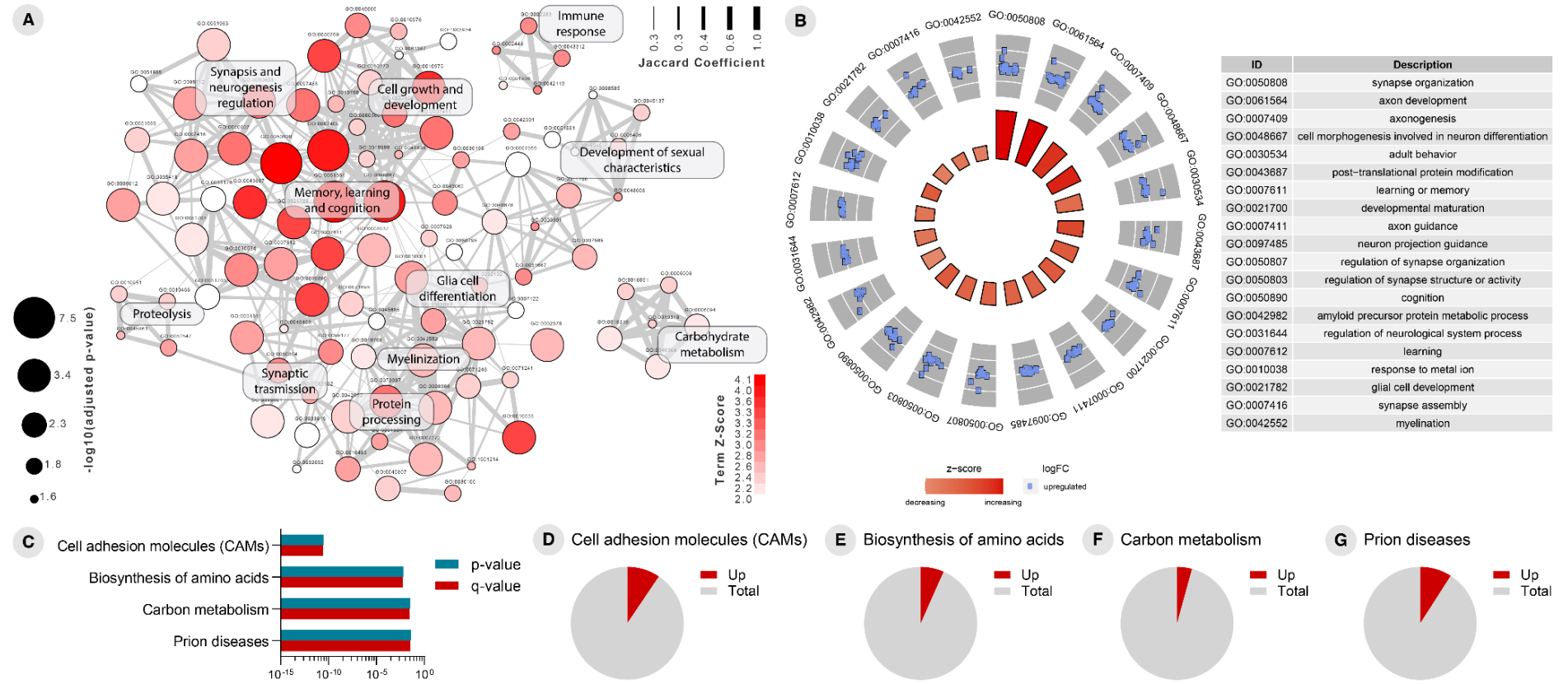


Figure 5

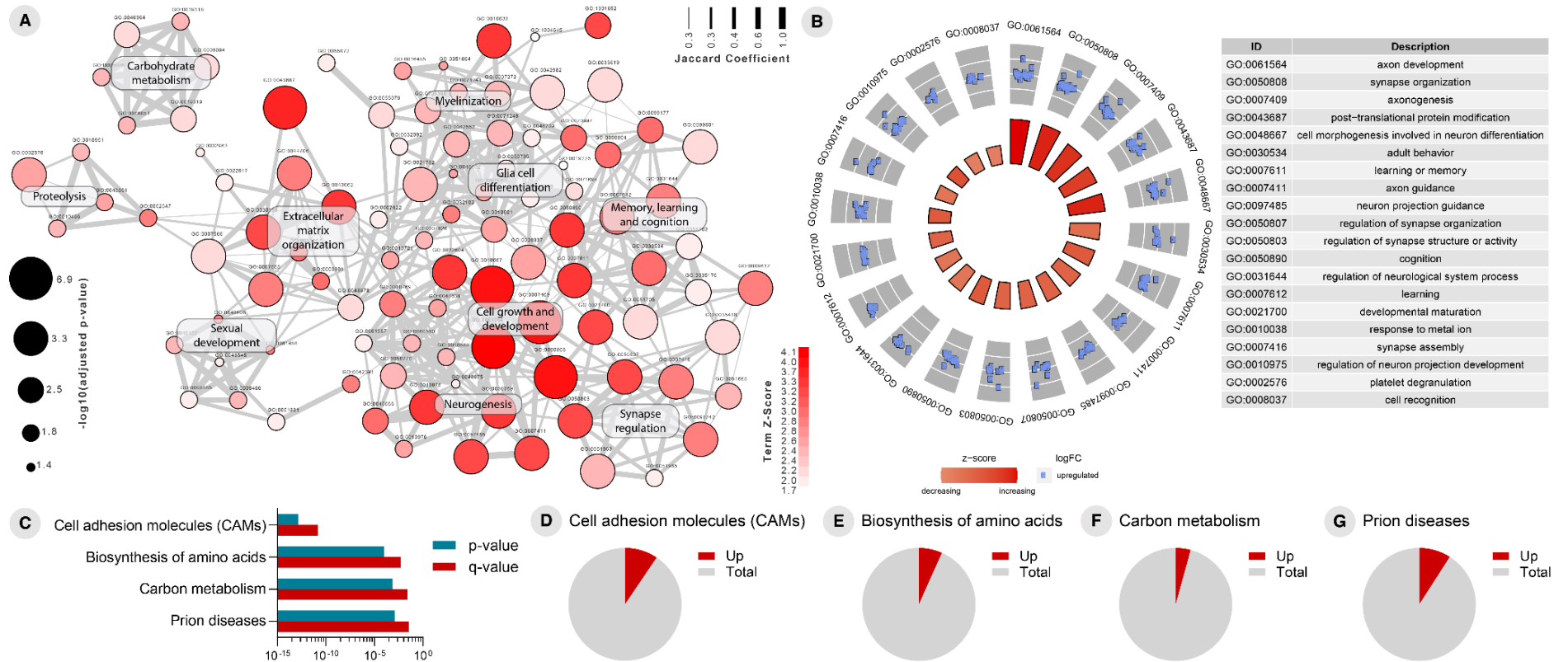


Figure 6

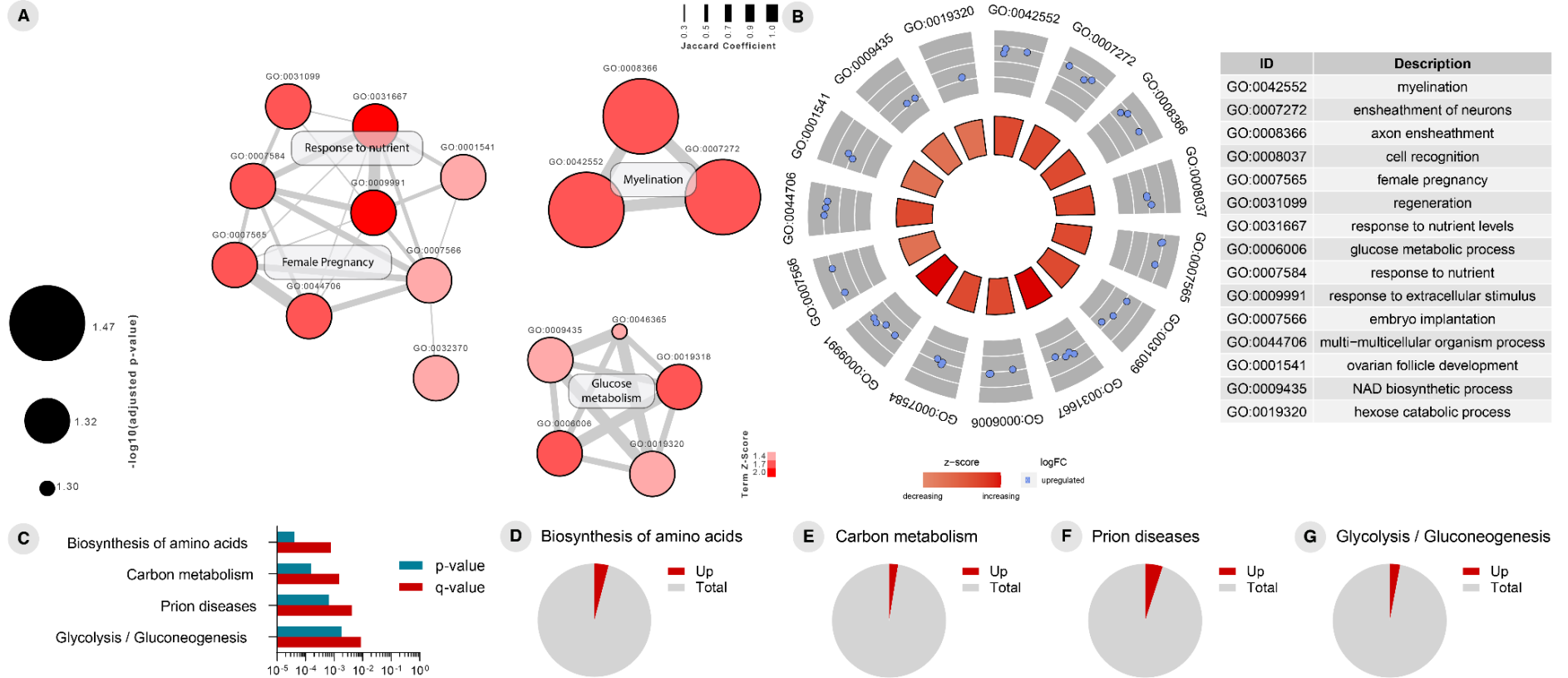


Figure 7

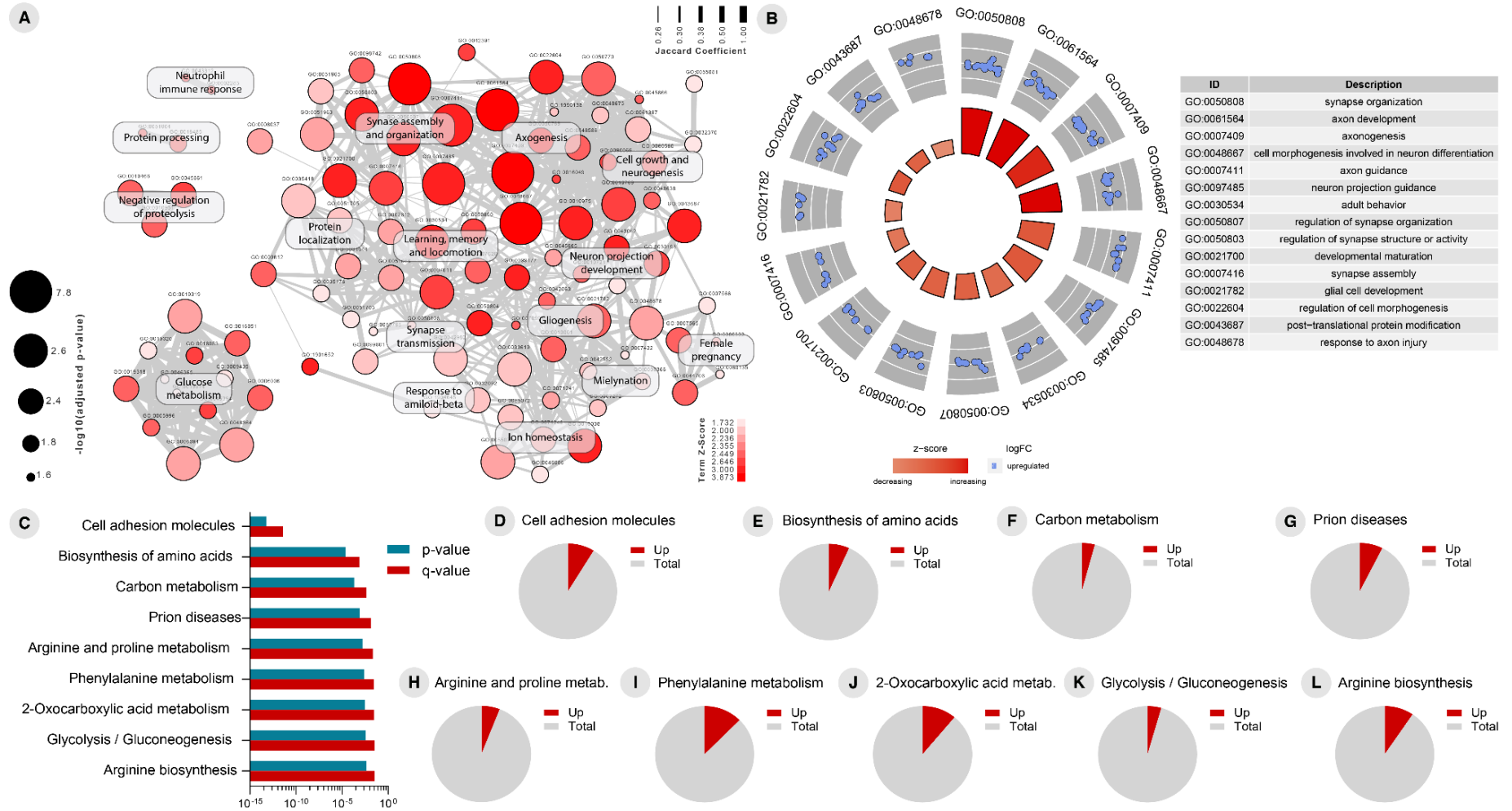


Table 1. Hyperparameters evaluated for the Machine Learning models used in the proposed framework.

Algorithm	Fixed Parameters	Iterated Parameters
Logistic Regression	solver: lbfgs max_iter: 250 penalty: l2	C: [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]
Naive Bayes	-	-
kNN	algorithm: ball_tree leaf_size: 50	n_neighbors: [1, 2, 3, 4, 5, 6, 7, 8, 9] p: [1, 2]
SVC	-	for kernels: [rbf, poly, sigmoid] C: [-4, -3, -2, -1, 0, 1, 2, 3] for kernel: linear gamma: [0.00001, 0.0001, 0.001, 0.01, 0.1] C: [0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]
Decision Trees	-	max_depth: [1, 2, 3, 4, 5, 6, 7, 8, 9] criterion: [gini, entropy]
Random Forest	-	max_depth: [3, 4, 5, 8, 10] n_estimators: [5, 20, 50, 100, 200, 500, 1000]
Gradient Boosting	-	max_depth: [3, 4, 5, 8, 10] learning_rate: [0.01, 0.05, 0.1, 0.2] n_estimators: [5, 20, 50, 100, 200, 500, 1000]
XGBoost	-	max_depth: [6, 7, 8] learning_rate: [0.01, 0.025, 0.05, 0.075, 0.1] n_estimators: [5, 20, 50, 100, 200, 500, 1000]
AdaBoost	-	learning_rate: [0.25, 0.5, 1.0, 1.25, 1.5] n_estimators: [20, 50, 100, 150, 200]

kNN: k-Nearest Neighbors; SVC: Support Vector Classifier

Table 2. Sample characteristics.

Characteristic	CU	A-	A+	T-	T+	N-	N+
# of individuals	318	60	50	52	58	67	43
Sex (% female)	50%	51.67%	48%	50%	50%	50.75%	48.84%
Age (y)	75.66 ± 5.22	75.37 ± 5.67	76.01 ± 4.67	73.87 ± 4.54	77.26 ± 5.32 ^{b***}	74.46 ± 4.66	77.52 ± 5.55 ^{c**}
Education (y)	15.73 ± 2.83	15.42 ± 2.68	16.1 ± 2.99	15.77 ± 2.77	15.69 ± 2.91	15.57 ± 3.1	15.98 ± 2.37
MMSE	29.08 ± 1.03	28.98 ± 1.1	29.2 ± 0.95	29.13 ± 0.93	29.03 ± 1.12	29.01 ± 1.05	29.19 ± 1.01
ADAS-cog11	6.42 ± 2.92	6.09 ± 2.91	6.81 ± 2.92	6.18 ± 2.85	6.64 ± 2.99	6.22 ± 2.69	6.73 ± 3.27
APOE e4 carriers (%)	24.55%	11.67%	40% ^{a***}	15.38%	32.76% ^{b*}	19.40%	32.56%

CU: Cognitively Unimpaired; A+: Amyloid-beta positive; A-: Amyloid-beta negative; T+: tau positive; T-: tau negative; N+: neurodegeneration positive; N-: neurodegeneration negative; y: year; MMSE: Mini-Mental State Exam; ADAS-cog: Alzheimer's Disease Assessment Scale-cognitive subscale. Statistical differences for numerical characteristics were tested using t test. Statistical differences for sex and APOE status were tested using Fisher's exact test. (*p < 0.05, **p ≤ 0.01, ***p ≤ 0.001). ^a significantly different from A-, ^b significantly different from T-, ^c significantly different from N-.

Table 3. AUC results for predicting T+ and N+ CU individuals using single A β isoforms or its ratios.

Prediction	A β_{1-38}	A β_{1-40}	A β_{1-42}	A $\beta_{1-42}/A\beta_{1-40}$	A $\beta_{1-42}/A\beta_{1-38}$	A $\beta_{1-40}/A\beta_{1-38}$
T+	0.811	0.811	0.529	0.693	0.682	0.484
N+	0.847	0.855	0.580	0.663	0.652	0.479

T+: tau positive; N+: neurodegeneration positive; A β : Amyloid-beta.

Additional file 1

Machine learning results for predicting tau pathology positivity (T+).

[w Additional file 1.docx](#)

Additional file 2

Machine learning results for predicting neurodegeneration positivity (N+).

[w Additional file 2.docx](#)

Additional file 3

Differentially expressed proteins (DEPs) in the cerebrospinal fluid (CSF) of cognitively unimpaired (CU) tau pathology positive (T+) compared to negative (T-) subjects.

[w Additional file 3.docx](#)

Additional file 4

Differentially expressed proteins (DEPs) in the cerebrospinal fluid (CSF) of cognitively unimpaired (CU) neurodegeneration positive (N+) compared to negative (T-) subjects.

[w Additional file 4.docx](#)

Capítulo 2

Serine/threonine kinase activity regulates brain glucose metabolism in Alzheimer's Disease

No capítulo 2 apresentamos o manuscrito a ser submetido ao periódico *eLife*.

O diagnóstico e o monitoramento da DA melhoraram muito devido ao papel fundamental da imagem PET. Desta forma, integrar imagens PET e dados ômicos pode fornecer novos *insights* sobre a fisiopatologia da DA. Para isso, neste trabalho utilizamos genes diferencialmente expressos (DEGs) para implementar uma abordagem de redução de dimensão baseada em Gene Ontology (GO) e engenharia reversa de redes transcricionais centradas em fatores de transcrição (TF). A partir disso, selecionamos *clusters* de GOs e unidades reguladoras de TF para serem integrados com imagens de FDG-PET utilizando modelos de regressão linear a nível de *voxel*. A partir deste trabalho, identificamos processos biológicos periféricos associados ao metabolismo de FDG-PET no cérebro de indivíduos em todo o espectro clínico da DA. Além disso, destacamos o potencial das assinaturas gênicas como novos biomarcadores da DA.

PARTE III

Na Parte III, são apresentadas a discussão e as conclusões finais desta tese. Além disso, são descritas as referências utilizadas no corpo principal do texto.

3. DISCUSSÃO

A DA é uma doença neurodegenerativa progressiva que não possui cura. A primeira alteração patológica detectável por biomarcadores é a diminuição dos níveis da proteína $A\beta_{1-42}$ no compartimento liquórico, indicativo de seu acúmulo em placas extracelulares no cérebro, que ocorre décadas antes da fase clínica da doença (DETURE; DICKSON, 2019). Estudos sugerem que o acúmulo de $A\beta$ é seguido por disfunção sináptica e o aumento da fosforilação e secreção de tau, uma proteína axonal que estabiliza os microtúbulos (PALMQVIST et al., 2019). Consequentemente, desenvolvem-se os NFTs intracelulares que são compostos de proteínas tau hiperfosforiladas (JACK; HOLTZMAN, 2013). Esta descrição define a hipótese da cascata amiloide (HARDY; HIGGINS, 1992), que é o modelo mais aceito da cascata fisiopatológica na DA.

A deposição extracelular de $A\beta$, gerada através da clivagem da APP por BACE1 e γ -secretase, em placas é a principal característica patológica da DA, e tem sido proposto como um dos principais eventos patogênicos da doença (SELKOE; HARDY, 2016). Atualmente, podemos medir a patologia $A\beta$ *in vivo* por meio de biomarcadores no LCR e de imagem PET. No entanto, os biomarcadores de amiloide, geralmente indexados pelo $A\beta_{1-42}$, não inferem *per se* se um indivíduo apresenta ou desenvolverá patologia de tau ou neurodegeneração. Tanto é que cerca de 30% dos indivíduos considerados positivos para amiloide não apresentam essas outras características patológicas da DA (AIZENSTEIN et al., 2008; JACK et al., 2009; PIKE et al., 2007).

A agregação da proteína tau hiperfosforilada é outra característica patológica chave na DA. Os marcadores tau total (t-tau) e p-tau foram propostos, juntamente

com a razão $A\beta_{1-42}/A\beta_{1-40}$ no LCR, como biomarcadores para definir biologicamente a DA (JACK et al., 2018). As concentrações de t-tau e p-tau no LCR refletem a neurodegeneração e patologia de tau na DA (ITOH et al., 2001; SKILLBÄCK et al., 2015).

A RM e níveis de t-tau no LCR têm sido comumente utilizados como biomarcadores para neurodegeneração relacionada à DA (FRISONI et al., 2010; JACK et al., 2018). A medida de t-tau no LCR foi proposta como um biomarcador de neurodegeneração, e pode ser utilizada como um marcador de intensidade da doença na DA, indicando que quanto mais alta sua concentração, mais rápida a progressão clínica do indivíduo na doença (ZETTERBERG, 2017).

Uma consequência esperada da hipótese da cascata amiloide é que a modulação da produção de níveis de $A\beta$ no cérebro deve prevenir os efeitos subsequentes na DA, retardando o curso da doença (CULLEN et al., 2022). No entanto, apesar das evidências que falam do papel do amiloide como condutor da doença, terapias anti amiloides são altamente questionáveis e dividem a comunidade científica, fato elucidado pela aprovação controversa do anticorpo monoclonal Aducanumab. Associados, esses fatores fazem com que o papel do amiloide na cadeia de eventos que culmina na DA se torne uma questão ainda sem resposta (CULLEN et al., 2022). Uma das possíveis explicações para isso, é que grande parte das pesquisas envolvendo amiloide utilizam apenas a isoforma $A\beta_{1-42}$. Cada vez mais temos evidências que mostram que diferentes isoformas de $A\beta$, especialmente aquelas com cadeias mais curtas que o $A\beta_{1-42}$, podem desempenhar um papel mais decisivo na patogênese da DA do que inicialmente se pensava (DUNYS; VALVERDE; CHECLER, 2018; WIRTHS; ZAMPAR, 2019). Nesse contexto,

mutações na presenilina podem perturbar a relação entre $A\beta_{1-42}$ e espécies mais curtas de $A\beta$ ao invés de afetarem apenas os níveis de $A\beta_{1-42}$ no cérebro (KELLEHER; SHEN, 2010; KUMAR-SINGH et al., 2006, p. 40; XIA et al., 2015, p.). Assim, fica evidente que atacar o processo amiloidogênico apenas na produção de $A\beta_{1-42}$ - ou a isoforma $A\beta_{1-42}$ diretamente - não é suficiente para interromper a trajetória da DA. É por conta disso que estudos mais recentes têm investigado a associação entre peptídeos de $A\beta$ de cadeia mais curta e as alterações relacionadas à DA, tentando entender o acúmulo de amiloide, particularmente no que se refere a terapias que alteram o curso da doença (CULLEN et al., 2022).

Desta forma, considerando que a deposição de placas de $A\beta$ seja o primeiro evento patológico da DA, ao invés de utilizarmos apenas o $A\beta_{1-42}$, propusemos que a combinação de isoformas de $A\beta$ com cadeias mais curtas (38, 40 e 42 aminoácidos), medidas no LCR, seriam capazes de prever os eventos subsequentes na DA. De fato, em nosso estudo, demonstramos que utilizando modelos de ML com a combinação de isoformas solúveis de $A\beta$ podemos prever a positividade de tau e neurodegeneração em indivíduos cognitivamente saudáveis. Ao considerarmos as isoformas isoladamente, o valor diagnóstico da $A\beta_{1-42}$ apresentou o menor poder preditivo. Já as isoformas com tamanho menor, apresentaram um desempenho consideravelmente superior. Interessantemente, a isoforma $A\beta_{1-38}$ foi a que teve a melhor performance para prever a positividade de tau e neurodegeneração.

Evidências recentes sugerem que níveis mais baixos de peptídeos $A\beta$ mais curtos ou uma proporção mais baixa de peptídeos $A\beta$ mais curtos para mais longos podem ser um fator importante na toxicidade de $A\beta$ (BLAIN et al., 2016; DUNYS; VALVERDE; CHECLER, 2018; MOORE et al., 2018). Em nosso estudo, mostramos

que pacientes cognitivamente saudáveis apresentam níveis de $A\beta_{1-42}/A\beta_{1-40}$ diminuídos no LCR (esperado para indivíduos no *continuum* biológico da DA) quando positivos para tau e neurodegeneração. Por outro lado, os níveis de $A\beta_{1-40}$ e $A\beta_{1-38}$ estão aumentados nesses indivíduos. No trabalho de Cullen, foi demonstrado que níveis mais elevados de $A\beta_{1-38}$ no LCR de indivíduos no *continuum* da DA estão negativamente associados ao declínio cognitivo e ao risco de desenvolver a DA (CULLEN et al., 2022). Desta forma, nossos resultados podem levar à interpretação de que, se níveis mais altos de $A\beta_{1-38}$ são neuroprotetores, e os indivíduos cognitivamente saudáveis, mas considerados positivos para tau e neurodegeneração indexados por p-tau e t-tau, apresentam valores maiores de $A\beta_{1-38}$ no LCR, taxas mais altas da clivagem de $A\beta_{1-42}$ em $A\beta_{1-38}$ (OKOCHI et al., 2013, p. 43), podem indicar um processo crucial que parece conduzir a patologia da tau e a neurodegeneração. Alternativamente, podemos hipotetizar a presença de um mecanismo de autodefesa do organismo que esteja tentando processar a isoforma $A\beta_{1-42}$ na tentativa de retardar os efeitos causados pela toxicidade já conhecida da formação de placas de $A\beta$. Nesse contexto, fica evidente que o potencial dessa isoforma para agregar informações na fase pré-clínica da doença permanece pouco explorado. De qualquer forma, não está claro se os níveis de $A\beta_{1-38}$ modulam, de fato, o desenvolvimento da patologia de tau em indivíduos que já atingiram os limiares para a positividade da patologia de $A\beta$ (CULLEN et al., 2022).

Em meio a tantos questionamentos relacionados à hipótese amiloide, cada vez fica mais evidente que outros processos biológicos também são críticos na progressão em direção aos sintomas clínicos da DA. Portanto, técnicas ômicas, como a proteômica, pode ser uma das escolhas para avaliar de forma rápida, extensiva e confiável as propriedades e a dinâmica do sistema. Neste contexto,

avaliamos quais processos biológicos poderiam estar relacionados a um aumento no poder de predição das isoformas de A β sobre os eventos patológicos da DA através de análises proteômicas do LCR. Com a análise proteômica, encontramos muitas proteínas diferencialmente expressas entre os indivíduos T+ ou (N)+, o que permitiu a determinação de processos biológicos e vias de sinalização significativamente enriquecidas nesses indivíduos. Com isso, em um estudo futuro, podemos utilizar essas proteínas diferencialmente expressas para melhorar o poder preditivo dos modelos de ML na detecção da positividade de tau e neurodegeneração.

Neste âmbito, identificamos a partir da rede de ontologia genética de termos enriquecidos construídos a partir de proteínas diferencialmente expressas de indivíduos cognitivamente saudáveis, que aqueles que são T+ apresentam um *cluster* de termos superregulados relacionados à resposta imune. Isso se torna ainda mais interessante uma vez que espécies patológicas de tau hiperfosforiladas podem ser secretadas extracelularmente, levando a disseminação progressiva da tauopatia. Esse evento é capaz de promover tanto uma ativação microglial quanto promover a reatividade astrocitária. Essas células gliais, por sua vez, liberam citocinas e/ou moderadores inflamatórios neurotóxicos, incluindo IL1 β ou TNF α . Além disso, por uma modulação das cinases de tau, a ativação microglial pode levar a uma exacerbação da patologia de tau. Por fim, ainda temos o envolvimento da microglia na propagação de tau pela liberação de tau exosomal uma vez que a tau é fagocitada por essas células (LAURENT; BUÉE; BLUM, 2018).

Ainda no contexto da resposta imune, além das patologias amiloide e tau, outra característica histológica da DA é a presença de astrócitos reativos e microglia ativada na proximidade das placas de amiloide. No cérebro saudável, os astrócitos

fornece energia aos neurônios, participam da função sináptica (como parte da sinapse tripartite), induzem a poda sináptica e a liberação de fatores neurotróficos, dentre outras funções que em conjunto garantem a homeostase do sistema nervoso central (SNC) (LIDDELOW; BARRES, 2015; PELLERIN et al., 1998; STOBART; ANDERSON, 2013). Durante condições neuroinflamatórias, fatores inflamatórios liberados pela microglia ativada, parecem favorecer a formação de um subconjunto neurotóxico de astrócitos reativos que perdem suas funções normais e sua capacidade de promover a formação de sinapses, causando a morte neuronal através da secreção de fatores prejudiciais (LIDDELOW et al., 2017; LIDDELOW; BARRES, 2017). Atualmente acredita-se que os astrócitos possam desenvolver diversos fenótipos de acordo com o microambiente ao qual eles estão inseridos, portanto, a presença de um ou outro evento patológico da DA (tau, amiloide e/ou neurodegeneração) pode ativar diferentes aspectos da resposta astrocitária (DE BASTIANI et al., 2022; ESCARTIN et al., 2021; FERRARI-SOUZA et al., 2022). No que diz respeito às tauopatias, o mal dobramento neuronal da tau é suficiente para induzir alterações morfológicas nos astrócitos impactando seu papel fisiológico. Essas células mudam para um perfil pró-inflamatório, conforme indicado pela regulação positiva da proteína glial fibrilar ácida (GFAP) acompanhado pela secreção de fatores pró-inflamatórios, que contribuem para a patogênese da DA (LAURENT et al., 2017; RODRÍGUEZ-ARELLANO et al., 2016). Desta forma, um ponto chave que pode ser explorado para melhorar o poder de modelos para predição da positividade de tau é a utilização de biomarcadores relacionados com neuroinflamação, como a GFAP, que está, inclusive, dentre as proteínas diferencialmente expressas que encontramos em nosso estudo.

Por outro lado, a rede de ontologia genética para indivíduos (N)+ em comparação com os (N)- não mostrou alteração em termos relacionados à resposta imune. De forma interessante, termos relacionados com a memória, aprendizagem e cognição apresentaram-se suprarregulados tanto em indivíduos T+ e (N)+, ainda que todos sejam cognitivamente saudáveis. Isso demonstra que, conforme esperado, alterações biológicas importantes - e mensuráveis - já estão em ação no cérebro antes da fase clínica da DA. Inclusive, os termos mais enriquecidos estão relacionados com organização sináptica, desenvolvimento axonal e axonogênese, o que significa que tanto a tau quanto a neurodegeneração acabam levando a alterações da função cerebral, o fenômeno mais típico da DA. Esses achados demonstram a importância da definição biológica da DA quando se visa o desenvolvimento de estratégias terapêuticas que previnam o processo de neurodegeneração e consequente perda cognitiva.

De fato, muitas doenças neurodegenerativas, dentre elas a DA, produzem alterações significativas na função cerebral, mesmo quando exames de imagem estrutural como TC ou RM não revelam anormalidades específicas (SILVERMAN, 2004). Neste sentido, uma ferramenta de neuroimagem promissora no diagnóstico da DA é o FDG-PET, pois com ele, é possível observar o consumo cerebral de glicose em estado de repouso ou durante uma tarefa. Essa medida vem sendo utilizada como um indicador da atividade sináptica e vários estudos têm demonstrado que as alterações metabólicas cerebrais precedem a manifestação clínica dos sintomas da DA (MARCUS; MENA; SUBRAMANIAM, 2014). Estudos utilizando FDG-PET fornecem não apenas evidências do processo de demência, mas também permitem a diferenciação dos tipos de transtornos demenciais e seu

estadiamento, uma vez que essas patologias apresentam padrões metabólicos regionais diferentes (BROWN et al., 2014).

De forma geral, a imagem PET permite a quantificação *in vivo* de processos metabólicos, receptores, transportadores e alterações patológicas, de forma não invasiva (ZIMMER et al., 2014). A imagem PET pode ser utilizada para identificar e quantificar placas de A β (KLUNK et al., 2004), tau (VILLEMAGNE et al., 2015), e neurodegeneração a partir do hipometabolismo de glicose (BRIER et al., 2016; JACK et al., 2018), enquanto a redução da razão A β_{1-42} /A β_{1-40} e o aumento dos níveis de t-tau e p-tau podem ser medidos no LCR. Apesar de exames de imagem e as medidas em LCR terem uma maior acurácia e permitirem uma visualização regional (no caso do PET), principalmente no contexto da DA, eles têm um custo relativamente alto e/ou exigem a necessidade de pessoal e instalações especializadas para sua obtenção, o que limita seu uso na prática clínica (WITTENBERG et al., 2019). Assim, há uma necessidade de métodos mais acessíveis e econômicos para identificar com precisão esses processos-chave da DA. Baseado nisso, o campo de biomarcadores nos últimos anos tem voltado os seus esforços para explorar exames sanguíneos como potencial valor para diagnóstico e estadiamento da DA. Com menor custo e maior acessibilidade, os biomarcadores sanguíneos da DA podem contornar as limitações dos exames de imagem e baseados em LCR.

Com base nisso, é de grande interesse encontrar biomarcadores sanguíneos capazes de refletir alterações metabólicas cerebrais. Por esse motivo, propusemos a implementação de duas estratégias baseadas em biologia de sistemas para integrar informações transcriptômicas (genes que estão alterados tanto no cérebro quanto no

sangue de pacientes com a DA) de grupos de genes com dados de neuroimagem. Primeiramente, *clusters* de termos de ontologia gênica agrupados por similaridade semântica foram associados a imagens FDG-PET. Após, utilizamos engenharia reversa de redes transcricionais para integrar unidades reguladoras de fatores de transcrição com imagens FDG-PET. O objetivo dessa nova metodologia de integração é ajudar a elucidar as bases biológicas e destacar novos biomarcadores potenciais da patologia da DA com o uso da biologia de sistemas. De fato, como resultado da aplicação dessa metodologia identificamos uma forte associação entre um *cluster* relacionado à regulação da atividade da proteína serina/treonina cinase e o sinal FDG-PET no cérebro. É importante mencionar que essa técnica de integração oferece informação de topografia, ou seja, indica quais áreas do cérebro temos uma melhor associação entre FDG-PET e os genes identificados. Também, mostramos que a unidade reguladora mais intimamente associada ao metabolismo da glicose no cérebro é a ZNF653.

Neste sentido, essa metodologia é uma nova ferramenta para integrar dados periféricos com imagem, permitindo a identificação de novos biomarcadores sanguíneos que se correlacionam com o metabolismo cerebral. Além disso, a integração de dados transcriptômicos do sangue e imagens cerebrais podem ser aplicadas usando várias modalidades de imagem, bem como adaptadas e usadas no contexto de diferentes doenças neurológicas.

Como mencionado anteriormente, o sistema AT(N) dicotomiza indivíduos usando biomarcadores que mapeiam as principais características fisiopatológicas da DA, ou seja, A β , tau e neurodegeneração. Esse sistema é flexível e adaptável e, conseqüentemente, aberto a evoluir dinamicamente à medida que nosso

conhecimento dos mecanismos fisiopatológicos subjacentes ao *continuum* DA se expande. Por conta disso, conforme evidências emergirem no futuro, novos biomarcadores, que reflitam aspectos complementares da DA, podem ser descobertos e fornecer um componente 'X' para expandir o sistema AT(N) para ATX(N), onde X tem sido proposto como novos biomarcadores para mecanismos fisiopatológicos adicionais (HAMPEL et al., 2021). Levando isso em consideração, nossa metodologia tem o poder de auxiliar na descoberta desses novos componentes, com isso, permitindo detectar anormalidades em vias biológicas críticas em pacientes com DA e que se manifestem também em compartimentos periféricos, o que tem o potencial de fornecer *insights* para o desenvolvimento de novos biomarcadores, inclusive sanguíneos, que quando apropriadamente validados podem ser úteis na prática clínica.

4. CONCLUSÃO

Esta tese destacou a potencial utilização de técnicas neurocomputacionais para a identificação de alterações funcionais na DA. Utilizando aprendizado de máquina, demonstramos que diferentes isoformas de A β predizem tanto a positividade para patologia de tau quanto para neurodegeneração em curso em indivíduos cognitivamente saudáveis. Também, mostramos que aqueles indivíduos positivos para patologia de tau apresentaram superregulação de proteínas relacionadas à resposta imune, o que nos levou à discussão sobre resposta neuro-inflamatória e a relação da patologia de tau com a microglia e astrócitos, reforçado pela proteína GFAP que se apresentou diferencialmente expressa no estudo. Subsequentemente, propusemos um método de integração de imagem PET e tecnologias ômicas, em específico, FDG-PET e transcriptômica do sangue. Nesse estudo, usamos genes diferencialmente expressos para implementar uma abordagem de redução de dimensão baseada em Gene Ontology (GO) e engenharia reversa de redes transcricionais centradas em fatores de transcrição (TF). Agrupamentos GO e unidades reguladoras de TF foram selecionados para serem integrados com imagens FDG-PET usando modelos de regressão linear. Com isso, identificamos processos biológicos periféricos associados ao metabolismo de FDG-PET no cérebro de indivíduos em todo o espectro clínico da DA. Por fim, o enriquecimento do cluster relacionado à atividade da proteína serina/treonina quinase e a unidade reguladora sob o TF ZNF653 destacam o potencial das assinaturas gênicas associadas a eles como novos biomarcadores da DA.

5. PERSPECTIVAS

Embora diferentes estudos tenham mostrado uma associação entre os marcadores sanguíneos e a DA até agora, nenhum estudo utilizou ML para determinar qual biomarcador de plasma melhor prevê a positividade de AT(N) em todo o espectro da DA. Desta forma, como proposta para trabalhos futuros, usaremos ML para determinar o desempenho de biomarcadores plasmáticos de A β , diferentes isoformas de p-tau, GFAP e NfL para identificar a patologia cerebral de A β , tau e neurodegeneração em todo o espectro da DA.

REFERÊNCIAS

- AIZENSTEIN, H. J. et al. Frequent amyloid deposition without significant cognitive impairment among the elderly. **Archives of Neurology**, v. 65, n. 11, p. 1509–1517, nov. 2008.
- ALZHEIMER'S & DEMENTIA. 2019 Alzheimer's disease facts and figures. **Alzheimer's & Dementia**, v. 15, n. 3, p. 321–387, 1 mar. 2019.
- BLAIN, J.-F. et al. Characterization of FRM-36143 as a new γ -secretase modulator for the potential treatment of familial Alzheimer's disease. **Alzheimer's Research & Therapy**, v. 8, p. 34, 30 ago. 2016.
- BLENNOW, K. et al. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. **Nature Reviews Neurology**, v. 6, n. 3, p. 131–144, mar. 2010.
- BLENNOW, K.; ZETTERBERG, H. Biomarkers for Alzheimer's disease: current status and prospects for the future. **Journal of Internal Medicine**, v. 284, n. 6, p. 643–663, 2018.
- BLENNOW, K.; ZETTERBERG, H.; FAGAN, A. M. Fluid Biomarkers in Alzheimer Disease. **Cold Spring Harbor Perspectives in Medicine**, v. 2, n. 9, set. 2012.
- BRIER, M. R. et al. Tau and A β imaging, CSF measures, and cognition in Alzheimer's disease. **Science Translational Medicine**, v. 8, n. 338, p. 338ra66, 11 maio 2016.
- BROWN, R. K. J. et al. Brain PET in Suspected Dementia: Patterns of Altered FDG Metabolism. **RadioGraphics**, v. 34, n. 3, p. 684–701, maio 2014.
- CHUN, W.; JOHNSON, G. V. W. The role of tau phosphorylation and cleavage in neuronal cell death. **Frontiers in Bioscience: A Journal and Virtual Library**, v. 12, p. 733–756, 1 jan. 2007.
- CRAS, P. et al. Senile plaque neurites in Alzheimer disease accumulate amyloid precursor protein. **Proceedings of the National Academy of Sciences**, v. 88, n. 17, p. 7552–7556, set. 1991.
- CULLEN, N. et al. Association of CSF A β 38 Levels With Risk of Alzheimer Disease–Related Decline. **Neurology**, v. 98, n. 9, p. e958–e967, 1 mar. 2022.
- CULLEN, N. C. et al. Plasma biomarkers of Alzheimer's disease improve prediction of cognitive decline in cognitively unimpaired elderly populations. **Nature Communications**, v. 12, n. 1, p. 3555, 11 jun. 2021.
- DE BASTIANI, M. A. et al. **Hippocampal GFAP-positive astrocyte responses to amyloid and tau pathologies**. [s.l.] Neuroscience, 28 fev. 2022. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/2022.02.25.481812>>. Acesso em: 29 abr. 2022.
- DETURE, M. A.; DICKSON, D. W. The neuropathological diagnosis of Alzheimer's disease. **Molecular Neurodegeneration**, v. 14, n. 1, p. 32, 2 ago. 2019.
- DICKSON, D. W. The Pathogenesis of Senile Plaques. **Journal of Neuro pathology & Experimental Neurology**, v. 56, n. 4, p. 321–339, 1 abr. 1997.
- DUGGER, B. N.; DICKSON, D. W. Pathology of Neurodegenerative Diseases. **Cold Spring Harbor Perspectives in Biology**, v. 9, n. 7, p. a028035, 7 jan. 2017.

DUNYS, J.; VALVERDE, A.; CHECLER, F. Are N- and C-terminally truncated A β species key pathological triggers in Alzheimer's disease? **The Journal of Biological Chemistry**, v. 293, n. 40, p. 15419–15428, 5 out. 2018.

ELLIS, K. A. et al. Advances in structural and molecular neuroimaging in Alzheimer's disease. **The Medical Journal of Australia**, v. 194, n. 4, p. S20-23, 21 fev. 2011.

ESCARTIN, C. et al. Reactive astrocyte nomenclature, definitions, and future directions. **Nature Neuroscience**, v. 24, n. 3, p. 312–325, mar. 2021.

FERRARI-SOUZA, J. P. et al. **Astrocyte biomarker signatures of amyloid- β and tau pathologies in Alzheimer's disease**. [s.l.] *Neurology*, 29 jan. 2022. Disponível em: <<http://medrxiv.org/lookup/doi/10.1101/2022.01.25.22269841>>. Acesso em: 29 abr. 2022.

FRISONI, G. B. et al. The clinical use of structural MRI in Alzheimer disease. **Nature Reviews. Neurology**, v. 6, n. 2, p. 67–77, fev. 2010.

G7. **The challenge of neurodegenerative diseases in an aging population**. [s.l.: s.n.]. Disponível em: <<http://www.academie-sciences.fr/en/Advice-Notes-and-Reports/g7-science-academies-statement-2017.html>>. Acesso em: 12 dez. 2017.

GBD 2017 US NEUROLOGICAL DISORDERS COLLABORATORS et al. Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study. **JAMA neurology**, v. 78, n. 2, p. 165–176, 1 fev. 2021.

GOOCH, C. L.; PRACT, E.; BORENSTEIN, A. R. The burden of neurological disease in the United States: A summary report and call to action. **Annals of Neurology**, v. 81, n. 4, p. 479–484, 1 abr. 2017.

HAASS, C. et al. Trafficking and proteolytic processing of APP. **Cold Spring Harbor Perspectives in Medicine**, v. 2, n. 5, p. a006270, maio 2012.

HAMEED, S. et al. Role of Fluid Biomarkers and PET Imaging in Early Diagnosis and its Clinical Implication in the Management of Alzheimer's Disease. **Journal of Alzheimer's Disease Reports**, v. 4, n. 1, p. 21–37, 2020.

HAMET, P.; TREMBLAY, J. Artificial intelligence in medicine. **Metabolism**, Insights Into the Future of Medicine: Technologies, Concepts, and Integration. v. 69, p. S36–S40, 1 abr. 2017.

HAMLEY, I. W. The amyloid beta peptide: a chemist's perspective. Role in Alzheimer's and fibrillization. **Chemical Reviews**, v. 112, n. 10, p. 5147–5192, 10 out. 2012.

HAMPEL, H. et al. Developing the ATX(N) classification for use across the Alzheimer disease continuum. **Nature Reviews. Neurology**, v. 17, n. 9, p. 580–589, set. 2021.

HARDY, J. A.; HIGGINS, G. A. Alzheimer's disease: the amyloid cascade hypothesis. **Science (New York, N.Y.)**, v. 256, n. 5054, p. 184–185, 10 abr. 1992.

ITOH, N. et al. Large-scale, multicenter study of cerebrospinal fluid tau protein phosphorylated at serine 199 for the antemortem diagnosis of Alzheimer's disease. **Annals of Neurology**, v. 50, n. 2, p. 150–156, ago. 2001.

JACK, C. R. et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. **Journal of magnetic resonance imaging: JMRI**, v. 27, n. 4, p. 685–691,

abr. 2008.

JACK, C. R. et al. Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. **Brain: A Journal of Neurology**, v. 132, n. Pt 5, p. 1355–1365, maio 2009.

JACK, C. R. et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. **Neurology**, v. 87, n. 5, p. 539–547, 2 ago. 2016.

JACK, C. R. et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. **Alzheimer's & Dementia: The Journal of the Alzheimer's Association**, v. 14, n. 4, p. 535–562, 2018.

JACK, C. R.; HOLTZMAN, D. M. Biomarker modeling of Alzheimer's disease. **Neuron**, v. 80, n. 6, p. 1347–1358, 18 dez. 2013.

KELLEHER, R. J.; SHEN, J. Genetics. Gamma-secretase and human disease. **Science (New York, N.Y.)**, v. 330, n. 6007, p. 1055–1056, 19 nov. 2010.

KIRSCHNER, M. W. The meaning of systems biology. **Cell**, v. 121, n. 4, p. 503–504, 20 maio 2005.

KLUNK, W. E. et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. **Annals of Neurology**, v. 55, n. 3, p. 306–319, mar. 2004.

KUMAR-SINGH, S. et al. Mean age-of-onset of familial Alzheimer disease caused by presenilin mutations correlates with both increased Aβ42 and decreased Aβ40. **Human Mutation**, v. 27, n. 7, p. 686–695, jul. 2006.

KUMMER, M. P.; HENEKA, M. T. Truncated and modified amyloid-beta species. **Alzheimer's Research & Therapy**, v. 6, n. 3, p. 28, 2014.

LAKHAN, S. E. Alzheimer Disease: Practice Essentials, Background, Anatomy. 5 out. 2016.

LAURENT, C. et al. Hippocampal T cell infiltration promotes neuroinflammation and cognitive decline in a mouse model of tauopathy. **Brain: A Journal of Neurology**, v. 140, n. 1, p. 184–200, jan. 2017.

LAURENT, C.; BUÉE, L.; BLUM, D. Tau and neuroinflammation: What impact for Alzheimer's Disease and Tauopathies? **Biomedical Journal**, v. 41, n. 1, p. 21–33, fev. 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 28 maio 2015.

LIDDELOW, S. A. et al. Neurotoxic reactive astrocytes are induced by activated microglia. **Nature**, v. 541, n. 7638, p. 481–487, 26 jan. 2017.

LIDDELOW, S. A.; BARRES, B. A. Reactive Astrocytes: Production, Function, and Therapeutic Potential. **Immunity**, v. 46, n. 6, p. 957–967, 20 jun. 2017.

LIDDELOW, S.; BARRES, B. SnapShot: Astrocytes in Health and Disease. **Cell**, v. 162, n. 5, p. 1170–1170.e1, 27 ago. 2015.

MARCUS, C.; MENA, E.; SUBRAMANIAM, R. M. Brain PET in the Diagnosis of Alzheimer's Disease. **Clinical nuclear medicine**, v. 39, n. 10, p. e413–e426, out. 2014.

- MESKO, B. The role of artificial intelligence in precision medicine. **Expert Review of Precision Medicine and Drug Development**, v. 2, n. 5, p. 239–241, 3 set. 2017.
- MOORE, B. D. et al. Short A β peptides attenuate A β 42 toxicity in vivo. **The Journal of Experimental Medicine**, v. 215, n. 1, p. 283–301, 2 jan. 2018.
- OKOCHI, M. et al. γ -secretase modulators and presenilin 1 mutants act differently on presenilin/ γ -secretase function to cleave A β 42 and A β 43. **Cell Reports**, v. 3, n. 1, p. 42–51, 31 jan. 2013.
- OLTVAI, Z. N.; BARABÁSI, A.-L. Systems biology. Life's complexity pyramid. **Science (New York, N.Y.)**, v. 298, n. 5594, p. 763–764, 25 out. 2002.
- PALMQVIST, S. et al. Cerebrospinal fluid and plasma biomarker trajectories with increasing amyloid deposition in Alzheimer's disease. **EMBO molecular medicine**, v. 11, n. 12, p. e11170, dez. 2019.
- PELLERIN, L. et al. Evidence supporting the existence of an activity-dependent astrocyte-neuron lactate shuttle. **Developmental Neuroscience**, v. 20, n. 4–5, p. 291–299, 1998.
- PERL, D. P. Neuropathology of Alzheimer's Disease. **The Mount Sinai Journal of Medicine, New York**, v. 77, n. 1, p. 32–42, 2010.
- PERRIN, R. J.; FAGAN, A. M.; HOLTZMAN, D. M. Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. **Nature**, v. 461, n. 7266, p. 916–922, 15 out. 2009.
- PIKE, K. E. et al. Beta-amyloid imaging and memory in non-demented individuals: evidence for preclinical Alzheimer's disease. **Brain: A Journal of Neurology**, v. 130, n. Pt 11, p. 2837–2844, nov. 2007.
- RODRÍGUEZ-ARELLANO, J. J. et al. Astrocytes in physiological aging and Alzheimer's disease. **Neuroscience**, v. 323, p. 170–182, 26 maio 2016.
- RUSSELL, S. J.; NORVIG, P.; DAVIS, E. **Artificial intelligence: a modern approach**. 3rd ed ed. Upper Saddle River: Prentice Hall, 2010.
- SELKOE, D. J. The molecular pathology of Alzheimer's disease. **Neuron**, v. 6, n. 4, p. 487–498, abr. 1991.
- SELKOE, D. J.; HARDY, J. The amyloid hypothesis of Alzheimer's disease at 25 years. **EMBO molecular medicine**, v. 8, n. 6, p. 595–608, jun. 2016.
- SERRANO-POZO, A. et al. Neuropathological alterations in Alzheimer disease. **Cold Spring Harbor Perspectives in Medicine**, v. 1, n. 1, p. a006189, set. 2011.
- SILVERMAN, D. H. S. Brain 18F-FDG PET in the diagnosis of neurodegenerative dementias: comparison with perfusion SPECT and with clinical evaluations lacking nuclear imaging. **Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine**, v. 45, n. 4, p. 594–607, abr. 2004.
- SKILLBÄCK, T. et al. Cerebrospinal fluid tau and amyloid- β 1-42 in patients with dementia. **Brain: A Journal of Neurology**, v. 138, n. Pt 9, p. 2716–2731, set. 2015.
- STOBART, J. L.; ANDERSON, C. M. Multifunctional role of astrocytes as gatekeepers of neuronal energy supply. **Frontiers in Cellular Neuroscience**, v. 7, p. 38, 2013.

SUH, Y.-H.; CHECLER, F. Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. **Pharmacological Reviews**, v. 54, n. 3, p. 469–525, set. 2002.

TARASOFF-CONWAY, J. M. et al. Clearance systems in the brain-implications for Alzheimer disease. **Nature Reviews. Neurology**, v. 11, n. 8, p. 457–470, ago. 2015.

VARGHESE, T. et al. A review of neuroimaging biomarkers of Alzheimer's disease. **Neurology Asia**, v. 18, n. 3, p. 239–248, 2013.

VERHEIJEN, J.; SLEEGERS, K. Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. **Trends in genetics: TIG**, v. 34, n. 6, p. 434–447, jun. 2018.

VILLEMAGNE, V. L. et al. Tau imaging: early progress and future directions. **The Lancet. Neurology**, v. 14, n. 1, p. 114–124, jan. 2015.

VILLOSLADA, P.; STEINMAN, L.; BARANZINI, S. E. Systems biology and its application to the understanding of neurological diseases. **Annals of Neurology**, v. 65, n. 2, p. 124–139, fev. 2009.

WALHOVD, K. B. et al. Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. **AJNR. American journal of neuroradiology**, v. 31, n. 2, p. 347–354, fev. 2010.

WALKER, L. C. A β Plaques. **Free Neuropathology**, v. 1, p. 31, 2020.

WANG, Z.-T. et al. Systems biology and gene networks in Alzheimer's disease. **Neuroscience and Biobehavioral Reviews**, v. 96, p. 31–44, jan. 2019.

WIRTHS, O.; ZAMPAR, S. Emerging roles of N- and C-terminally truncated A β species in Alzheimer's disease. **Expert Opinion on Therapeutic Targets**, v. 23, n. 12, p. 991–1004, dez. 2019.

WITTENBERG, R. et al. Economic impacts of introducing diagnostics for mild cognitive impairment Alzheimer's disease patients. **Alzheimer's & Dementia: Translational Research & Clinical Interventions**, v. 5, p. 382–387, 1 jan. 2019.

WOLZ, R. et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. **PloS One**, v. 6, n. 10, p. e25446, 2011.

XIA, D. et al. Presenilin-1 knockin mice reveal loss-of-function mechanism for familial Alzheimer's disease. **Neuron**, v. 85, n. 5, p. 967–981, 4 mar. 2015.

YANASE, D. et al. Brain FDG PET study of normal aging in Japanese: effect of atrophy correction. **European Journal of Nuclear Medicine and Molecular Imaging**, v. 32, n. 7, p. 794–805, jul. 2005.

ZETTERBERG, H. Review: Tau in biofluids – relation to pathology, imaging and clinical features. **Neuropathology and Applied Neurobiology**, v. 43, n. 3, p. 194–199, 2017.

ZIMMER, E. R. et al. MicroPET imaging and transgenic models: a blueprint for Alzheimer's disease clinical research. **Trends in Neurosciences**, v. 37, n. 11, p. 629–641, nov. 2014.